12-2020

A Convolutional Neural Network (CNN) based Pill Image Retrieval System

Ezekiel Adebayo Adewumi
eaa8771@rit.edu

A Convolutional Neural Network (CNN) based Pill Image Retrieval System

by

Ezekiel Adebayo Adewumi

A Graduate Capstone Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Data Analytics

Department of: Graduate Programs & Research

Rochester Institute of Technology

RIT Dubai

December 2020

RIT

Master of Science in

Data Analytics

Graduate Capstone Approval

Graduate Capstone Committee:

Name: Dr. Sanjay Modak **Date:**

Chair of committee

Name: Dr. Khalil Al Hussaeni **Date:**

Member of committee

ACKNOWLEDGMENTS

I would like to acknowledge the efforts of my supervisor Dr. Khalil Al Hussaeni alongside Mrs. Rema Amawi for their supervision and assistance during this research work.

I would also like to recognize and thank my parents for their unwavering support and motivation during the course of this project, in my studies, and all the extracurricular activities that I had to do while completing this research work.

Along with that, none of this would have been possible without the blessings and help from Almighty God who has kept me motivated and determined all this time to complete my project within the provided time. Lastly, I would like to appreciate the University, my fellow classmates, and all the members of staff as they have provided for us a platform, through which we can conduct our researches in the best possible manner.

ABSTRACT

Several works have been done in the area of image retrieval systems, and many are still trying to provide improvements for a better model for retrieving said images. Image segmentation using clustering techniques is one of the most used approaches. There are various clustering methods available, but the non-linear k-means clustering technique is the most used method. In the following research, a model of retrieving images using a non-linear classifier aided with a convolutional neural network is proposed. Both algorithms were exploited and paired in terms of feature extraction and classification. Comprehensive evaluations over a dataset containing over 7,000 pill images of 1,000 pill types obtained from the National Library of Medicine database demonstrate significant success during the data classification using the proposed model.

TABLE OF CONTENTS

ACKNOWLEDGMENTS ..... **I**

ABSTRACT ..... **II**

LIST OF FIGURES ..... **IV**

LIST OF TABLES ..... **V**

CHAPTER 1 ..... **1**

1.1 BACKGROUND OF THE PROBLEM **1**

1.2 MOTIVATION 3
1.3 PROBLEM STATEMENT 5
1.4 PROJECT DEFINITION AND GOALS 5
1.5 METHODOLOGY 6

CHAPTER 2 ..... **7**

LITERATURE REVIEW **7**

CHAPTER 3 ..... **10**

PROJECT DESCRIPTION **10**

3.1 IMAGE SEGMENTATION 10
3.2 CLASSIFICATION 12

CHAPTER 4 ..... **14**

PROJECT ANALYSIS **14**

4.1 DISCUSSION 15
4.2 EXPLORATORY DATA ANALYSIS 16
4.3 DATA UTILITY 21

5 CONCLUSIONS **24**

BIBLIOGRAPHY **25**

LIST OF FIGURES

Figure 1……..…….….10
Figure 2……..…….….11
Figure 3……..…….….11
Figure 4……...…….…12
Figure 5……14
Figure 6……...……..…15
Figure 7……..……..….18
Figure 8……...……..…19
Figure 9……..……..….19
Figure 10…….……..…20
Figure 11…….……..…20
Figure 12…….……..…21
Figure 13…….……..…22
Figure 14…….……..…22
Figure 15…….……..…23

LIST OF TABLES

Table 1 ……16

Table 2 ..…… 17

CHAPTER 1

1.1 BACKGROUND OF THE PROBLEM

Information retrieval describes the process of sourcing information from a storage system. The information obtained may be in text, image, sound, or metadata describing a database or data. Image information retrieval on a computer system enables a user to identify the searches from an extensive digital image database visually. The World Wide Web facilitates access to the largest database – the internet. Over the last few years, the population of individuals surfing the internet has tremendously increased; the increased internet presence owes mostly to an increase in mobile phones. Moreover, recent smartphones have computer functionalities that are comparable to those of the computers of yesterday. The increased dependency on smartphones for information retrieval has made mobile Informational retrieval a growing branch (Crestani, 2017), giving rise to the development of effective representative techniques for model building, indexing, retrieving, and information presentation.

Traditionally, metadata's inclusion, such as keywords, captioning, or title to an image, enables its retrieval. However, this manual approach consumes time, effort, and cost. With increasing internet activities, including social web applications, researches on Content-Based Information Retrieval (CBIR) has become prominent in the field of Information Retrieval. The CBIR is an automated image retrieval technique that can identify an image based on its features, such as its shape, texture, and color (Eakins, 1999) (Goodrum, 2000). Researches are ongoing to improve the effectiveness of the primitive (color and shape), logical (object identity), and abstraction (the importance of the scenery) levels of CBIR. The advances in image retrieval techniques have enabled the process to proffer solutions to a variety of needs. Moreover, image retrieval has found relevance in a wide field of applications, including medicine, law enforcement, and engineering. Automated pill image recognition remains a significant application of this technique to medicine.

The use of medicines is necessary for everyone for the prevention and treatment of illnesses and diseases. There is a high possibility of mistakes occurring while the health personnel prescribe, dispenses, or administers drugs. Makary & Daniel (Makary, 2016) argued that medical error ranks third on the major causes of death among hospital inpatients in the US. The WHO statistics revealed that approximately 1.3 million patients are lost to preventable medication blunders annually in the United States, a minimum of one death daily; the WHO also admitted medical error to be one of the top ten causes of death and disability (Eakins,

1999). Adverse Drug Effect (ADE) can also result in severe ailments, including Stevens-Johnson syndrome and Parkinson's syndrome (WHO, Medication Without Harm: Real-life stories, 2020). WHO statistics reported that health caregivers harm 4 of every ten patients globally (WHO, 10 facts on patient safety, 2020). In another survey, 39% are severe enough to cause injury to patients (Delgado, 2019). Consumers often find it challenging to identify pills and thereby run the risk of getting harm from consuming the wrong medication, underdose, or overdose. The frequent occasions are when pills are moved to different packaging containers, combined into a single container and when pills are shared into pillboxes for ease of administration.

Due to the lack of good data, medication error's global impact cannot be entirely determined (Makary, 2016) (Eakins, 1999). Nevertheless, the rate of error occurrence is believed to be similarly high globally though the impact is more devastating in low and middle-income countries compared to high-income countries (Eakins, 1999). Aside from the human loss, the financial implications of medication error are alarming: One-seventh of the Canadian budget goes to mitigating the effects of medication error (WHO, 10 facts on patient safety, 2020) and about one percent (US$ 42 billion) of the total global health spending (Eakins, 1999).

Therefore, to ensure a safe medication, it is imperative to avoid the common discrepancies involved in identifying a patient's medication. Efforts made in resolving pill identification difficulty includes the authorization of a distinctive appearance – from the combination of size, color, shape, and imprint - for every prescription medicine (Eakins, 1999) (Yu, 2014). An unidentified pill can, therefore, be cross-referenced by health practitioners against the database of prescription drugs. (Delgado, 2019) Pharmacists usually help their patients during a brown bag consultation with the drugs they bring in for identification. Manual search can be tedious, exhausting, and time-consuming, particularly when dealing with many pills with large generic variations (Delgado, 2019). Moreover, reading tiny imprints on small drugs can easily introduce human error (Yu, 2014). Alternatively, automated pill recognition techniques help to identify pill rather quickly, decrease the possibility of pill misidentification, and give visual assurance to the patient (Zeng, 2017). (Cunha, 2020) RxList Pill Identification Tool and (Cafasso, 2018) Healthline Pill Identifier are standard websites rendering pill identification services.

1.2 MOTIVATION

The Pill Image Recognition Challenge organized by the National Institutes of Health (NLM) led to the Computational Photography Project for Pill Identification (C3PI) (NLM, 2016). The use of a high-resolution camera on smartphones combined with computer vision algorithms has proven to be an efficient means of retrieving image information (Zeng, 2017). Deep learning techniques are introduced into CBIR to enhance its capability to extract features (contents) from an input image to identify and retrieve similar images from a database (Bose, 2020). With deep models, high-level features can be extracted along with the low-level features, which the conventional CBIR cannot extract (Bose, 2020). Deep learning has been impressive in its competence to recognize objects (Krizhevsky, 2014), faces (Taigman, 2014), and to handle extensive learning problems (LeCun, 2015). Deep learning has also improved clinical workflows – enhancing the experience of both the caregivers and the patients (Delgado, 2019). A Convolutional Neural Network (CNN) is a deep technique for digital image retrieval. A CNN architecture comprises a sequence of interacting convolutional, pooling, and fully connected layers, which are stacked (Bose, 2020).

MobileDeepPill's (Zeng, 2017) is a multi-CNN architecture based on AlexNet developed by Krizhevsky (Krizhevsky, 2014). The approach merged pill localization using color, gradients, and shape measurement to determine comparisons between consumer and reference images. Wang et al. (Wang, 2010) use canny edge detection and GoogLeNet Inception Network classifier for image recognition. GoogleNet - shape model, color model, and feature model for detecting the pill's shape, color, and imprint, respectively. However, pill data capture was taken in a highly controlled environment than the NINJH dataset (Delgado, 2019).

Several other techniques have been developed for pill image recognition with different accuracy levels. C3PI techniques include using the color property and a support vector machine (SVM) learning algorithm (Guo P. S., 2017). The technique achieved 97.90% overall color classification accuracy. Despite this, the effectiveness of the technique is limited by factors such as the lighting condition, the camera resolution, and the pill and background color contrast (Guo P. S., 2017). Grigorescu et al. (Grigorescu, 2003) introduced Distance Set, a local descriptor. The technique examines distance sets between any given point and its k neighbor points on the pill shape contour. The technique's limitation includes distortion in reading the imprints due to noise, complex or irregular shapes. (Eakins, 1999) The Two-Step Sampling Distance Set (TSDS) improves the distance sets technique adding imprint and color properties

to the pill's shape. The technique recorded 93.64% accuracy from querying 12500 images. The deep Residual Network (ResNet) has been described as one of the best in computer vision (He & Zhang, 2016) with profound object detection and face recognition ability. The deep learning technique can achieve convincing results, even when training as much as thousands of layers (He & Zhang, 2016) (Yu, 2014). ResNet is a much deeper learning technique and therefore has a high comparative advantage over AlexNet, the VGG network, and GoogLeNet, which had just 5, 19, and 22 convolutional layers, respectively (Szegedy, 2015) (Simonyan, 2014). ResNet is a collection of smaller networks (Fung, 2017).

1.3 PROBLEM STATEMENT

We informally describe our problem as follows: a query image is passed through a set of 5 convolution layers which are then passed to two fully connected layers. The extracted features from these are then used in the classification layers where a kNN classifier is employed to handle the prediction more accurately and with less runtime, which then generates the final predicted class.

The proposed identification approach was considered because it employs the techniques which captures the dynamics of the information needed for fast and accurate segmentation of both controlled and less controlled pill images. It addresses the issues relating to reliable matching between an input image and a KNN region, because the matching needs to handle the large spatial variance of semantic regions present in the query image.

1.4 PROJECT DEFINITION AND GOALS

Definition: Given a query image Q, consisting of a set of distinct features (Q_i, Q_j, Q_k) representing shape, color and imprint respectively, we wish to produce feature vectors from this image and group them using classification to determine the distance between the center and each pixel of the query image for accurate retrieval.

In this work, we presented comprehensive evaluations over a large dataset with over 7000 pill images of 1000 pill types obtained from the National Library of Medicine database (NLM, 2016). The database used comprises of images designed for the Pill Image Recognition Challenge to well demonstrate the effectiveness of our framework. The major contributions are summarized as follows:

- We developed a novel CNN + KNN image retrieval system that effectively solves both classification and regression problems via the placement of the KNN classifier in between the fully connected feature layer and the output layer in order to handle the prediction more accurately and with less runtime.
- The developed model significantly outperforms some existing systems discussed such as ResNet-50 and CNN with SVM classifier. The advantage of using this model was that a classifier like kNN used alongside the proposed neural network appreciably increases the accuracy with low noise.

1.5 METHODOLOGY

The proposed method was evaluated using the pills images on the publicly available National Library of Medicine database (NLM, 2016). The database used comprises 7000 pill images of 1000 pill types, designed for the Pill Image Recognition Challenge. The images were classified into two categories; the reference images and the consumer images, this is shown in Figure 6. Pharmaceutical companies took reference images under regulated conditions, ensuring appropriate control over lighting and background. There are 2000 images classified as reference – front and back images for 1000 pills. The consumer images were taken to portray the types of images that pill users would send to an automated pill recognition system. The images vary in quality, focus and device type. It comprises of 5000 jpeg images – front and back images of 2500 pills.

CHAPTER 2

LITERATURE REVIEW

A huge amount of work has been done by researchers to improve the methods of information retrieval over the years which has led to the development of various models for information retrieval. These models have brought about significant improvements in all aspects of the retrieval process. Some of these models are reviewed and explained below.

Firstly, probabilistic information retrieval, using weighted indexing was introduced in the ’60s by Maron and Kuhns, (Maron, 1960). They introduced the first paper on probabilistic information retrieval, proposing the use of weighted indexing. Meaning, the information retrieval system foretells which documents in the collection would most probably be relevant to a user's search "term" after the documents are weighted. These stored documents were previously weighted by the probability that a user would search the collection using that particular term, thereby ranking the documents by the probability of relevance. The drawback to this model is that the user query's output is not a set of matched documents, rather a set of probabilities of relevance.

Next, a tree-like structure was examined by the authors in (Adel'son-Vel'skii, 1962) which was named Adel'son-Vel'skiy and Landis tree (AVL), the AVL tree was further examined by Foster in (Foster, 1965) which led to the conclusion that the AVL tree provides a good compromise between the two extremes of complete balancing and unrestricted growth. Foster reviewed the number of probes (requests sent to computer memory) used to determine the posting and retrieval time of files in the tree while examining the AVL technique. There were 11 retrieval probes and 16 probes for posted files. Foster aimed to determine how an average computer would handle the probes. Conclusively, this examination showed that the number of probes needed to post or retrieve files in an AVL tree was small enough for an average computer with as low as 1.0 GHz processor speed to handle. However, would have a low retrieval and file maintenance time.

The picture indexing and abstraction method proposed by Chang and Liu in (Chang, 1984) led to a paradigm shift in image retrieval. In (Chang, 1984), they improved on the work done by Foster in (Foster, 1965) by proposing a picture indexing and abstraction method. This method worked by constructing picture indices using an object and relational naming for identifying images in a tree-like structured database. The abstraction method was used to cluster and

classify stored images. The method used syntactic and semantic abstraction rules to ensure that only accurate images were retrieved. This method can be used to facilitate pictorial information retrieval, although the image database is tough to update due to its rigid structure.

On the method of automatic indexing to assign identifiers to documents and search requests, the authors of (Salton, 1971) proposed one of the most prominent advancements to retrieval through the development of the System for the Mechanical Analysis and Retrieval of Text (SMART). The SMART system provided researchers with a framework to experiment with their proposed methods to improve the search quality. SMART made use of fully automatic indexing methods to assign identifiers to documents and search requests. Another major advancement made in (Salton, 1971) was the possibility to collect related documents by similarity computations between stored items and incoming queries. This made it possible to rank the retrieved items in decreasing order of their similarity with the incoming queries. Also, In addition to SMART, the authors of (Robertson, 1976) carried out a series of experiments to improve the retrieval performance of documents. Robertson and Jones (Robertson, 1976) discovered that the retrieval time improved whenever a reoccurring search term was added to the information used to identify already stored documents. The significance of their research led to the discovery that the methodology would greatly improve the retrieval performance on small datasets.

To improve the retrieval efficiency and accuracy, Bonchev and Maria in (Stanchev, 1987) proposed a non-text based approach for retrieving images from an extensive image database via the use of fuzzy set techniques to define the descriptions of stored images and the distances between these images. Using this method, a user query is not a text term but rather an image. The user first provides an image to the database; after that, the image description and distance are calculated and used to retrieve a stored image. This approach is limited to having all stored images described and their distances measured in advance.

The use of colour histograms was explored by Wang et al. (Wang, 2010). They explained that Local Feature Regions (LFR) would be more effective in retrieving images. This is achieved by first extracting the feature points of an image. After that, the LFR color histogram is constructed, and the similarity between the colour images is computed using the colour histogram of LFR. The method aimed at reducing the limitation of color histograms, which lacked spatial information and were sensitive to intensity variation, colour distortion, and cropping. The evaluation outcome showed that the proposed method was more accurate and

efficient in retrieving the user-queried images. The disadvantage of this method is that in the case of inaccurate segmentation, the system may partition an object into several incorrect regions. Next in the field of colour histograms, Lee et al. (Lee, 2012) utilized Wang et al.'s color histogram approach (Wang, 2010) and proposed an automatic pill recognition system based on pill imprint, which encompasses three features: shape, color, and texture. Lee et al (Lee, 2012) extracted feature vectors based on edge localization and invariant moments of tablets as an identifier to match and retrieve images between illegal and legal drugs. The experimental results showed 73% matching accuracy over the 13,000 legal drug pill images.

Neural network architecture is a pattern-recognition construct, which can detect and classify patterns in the input (LeCun, 2015). In recent times, information retrieval research exploits deep learning algorithms that tend to function like the human brain has been explored by numerous researchers. Neural Networks extract features (region of interest (Wang, 2010)) from the input data that can influence the output. The success rate achieved by convolutional neural networks during the ImageNet challenge highlighted the neural network capability in information retrieval (Ribera, 2017). Wang et al. (Wang, 2010) used GoogleNet models to extract color, shape, texture, and imprint features on a pill image. Therefore, the use of CNN has been considered basic to computer vision (Razavian, 2014) (Wang, 2010). MobileDeepPill built on AlexNet architecture attained 73:7% and 95:6% Top-1 and Top-5 accuracy respectively in two-side pill image recognition (Zeng, 2017), an advancement over 62.5% Top-1 and 83.0% Top-5 accuracy recorded by Krizhevsky's AlexNet, the winner of the NLM competition (Krizhevsky, 2014). Back-propagation of data to previous layers is a considerable challenge in training data in CNN, resulting in the problem of vanishing gradient and reduced performance. (Szegedy, 2015) Unlike the less effective technique of adding an auxiliary loss middle layer, ResNet introduced an identity shortcut connection. Delgado et al. in (Delgado, 2019), compared the accuracy of four different CNN architectures on NLM dataset, the research proved ResNet50 to be the most accurate with 77.0% Top-1 and 95.3% Top-5 others are MobileNet (77.1% Top-1, 94.4% Top-5), SqueezeNet (56.2% Top-1, 83.2% Top-5), InceptionV3 (76.3% Top-1, 94. Top-58). In (He & Zhang, 2016), a 1001-layer deep ResNet was trained which achieved higher accuracy than ResNet with fewer layers.

Geradts and Bijhold (Geradts, 2002) also proposed a method that can be used to retrieve information in a forensics image database. A traditional forensics image database is made of images of fingerprints, faces, shoeprints, handwriting, and more. The method was proposed to

highlight the significance of image retrieval in the forensics world, specifically by allowing a user to search the database for features such as texture, shapes, and colour distribution.

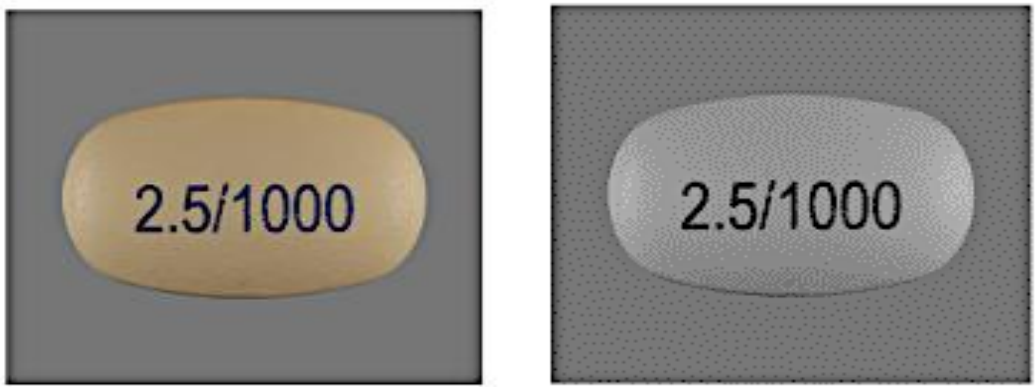
CHAPTER 3

PROJECT DESCRIPTION

A description of the preprocessing technique, the classification techniques, and the proposed convolutional neural network architecture used is provided in this section. This is followed by an explanation of how the proposed neural network is fused with the proposed classification technique.

3.1 IMAGE SEGMENTATION

A combination of image segmentation processes is applied to the pill image in order to make up for the color distortion and identify relevant information within the pill images. The first conversion is to grayscale format. This preprocessing method is used because a grayscale image regulates the intensity of the red, green and blue (RGB) components in an image, and so it is essential to denote a single intensity value for each pixel. The visibility of the pill in a colored image with its grayscale intensity image is shown in Figure 1.



(a) (b)

Figure 1 (a) Colored pill image; (b) Grayscale pill image.

i. ***Color detection*:** Gaussian filter is applied to the greyscale images to blur the images, removing unwanted details and noise. Also, a mean filter is applied to the output of the Gaussian filter to smoothen the images. Thereafter the Histogram equalization is then used to enhance the output filter images, the color contrast, and to extract the colors

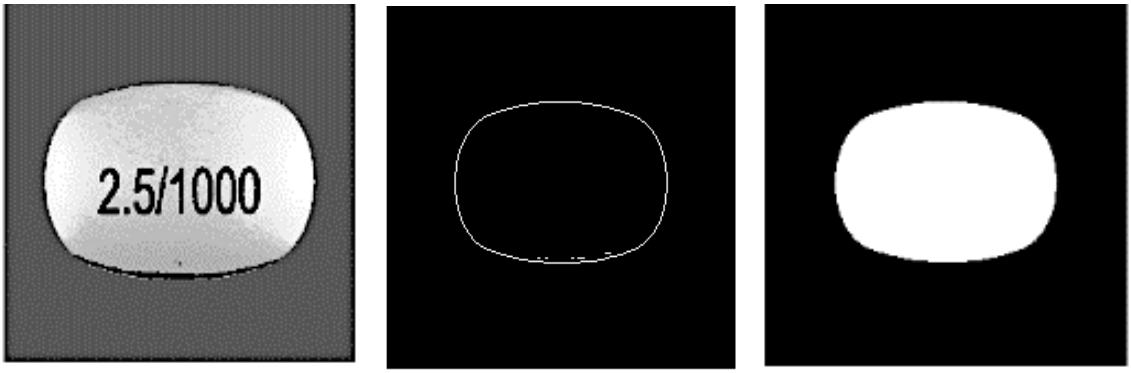
based on its measurement. The visibility of the pill after enacting the above filters and equalization is shown in Figure 2.

ii. ***Shape detection and extraction*:** To further enhance the extraction of the pill images. By using Sobel filtering on the images, we are able to refine these images in order to reveal the edges and the boundary lines of the drug pills. The visibility of the pill after enacting the above filters is shown in Figure 3.

iii. ***Imprint extraction:*** The use of a canny edge detector allowed all the edges in the image to be determined, next, a dilution operation was performed to soften it. Clear imprint on the image is finally revealed after applying Scale Invariant Feature Transform (SIFT) and Multi-scale Local Binary Pattern (MLBP) descriptors. The visibility of the pill after enacting the above convolutional dilution is shown are shown in Figure 4.

(a) (b) (c)

Figure 2 (a) Gaussian filtered pill image; (b) Mean pill filtered pill image; (c) Applied Local histogram equalized pill image.



(a) (b) (c)

Figure 3 (a) Applied Local histogram equalized pill image (b) Sobel filtered pill image; (c) Segmented shape

(a) (b) (c)

Figure 4 (a) Segmented imprint using canny edge detector; (b) Diluted pill image; (c) Final segmented pill imprint.

3.2 CLASSIFICATION

In order to make the neural network integrate well with the proposed classification techniques, the images are clustered and stored based on the features of each pill; color by means of k-clustering. For comparison the trained CNN is also tested with the techniques stated below:

i. ***k-nearest neighbors (kNN)*** (Guo & Wang, 2003)***:*** it is an algorithm that can be used for classification. The idea behind the method is that it assumes that similar things exist in close proximity. The method takes into consideration the data points of each image in the dataset using Euclidean distance in other to group them together. So, when an image is applied to the model, the input image will be converted to a feature vector. Thereafter, the image will be used to construct a color histogram to classifier the color of the pills and then stored under a class label extracted from the image path. The simple Euclidean Distance formula described this as:

$$\text{i. } d(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \qquad \text{(i)}$$

ii. ***Support vector machines (SVM)*** (Vapnik, 1998)***:*** It is an algorithm that can also be used for classification. The major stand out feature is that it extracts features from an image and then segregate them into classes with hyperplanes. Then the model will select the hyperplane with the best classifier to decide which hyperplane has the lowest

classification error and least match. In other words, it segregates the classes and chooses the most accurate hyperplane.

iii. ***Residual network*** (He & Zhang, 2016)***:*** This is also known as the resnet50 technique is a model that can be used as a final identifier in a convolutional layout. It can accommodate more than 50 layers and also be used to classify and extract features in an image. This technique makes uses a skip connection which helps to add the output from an earlier layer to a later layer without losing the image gradient.

CHAPTER 4

PROJECT ANALYSIS

The concept of identifying pills images has been studied previously, particularly on deep neural networks. However, unlike the above-mentioned related techniques in the related works, the proposed approach is not performed only on the neural network but aided with a non-linear classifier (Guo & Wang, 2003). The proposed method was constructed on a technique (Guo & Wang, 2003) that has been extensively used for supervised machine learning to solve both classification and regression problems. The classifier which is commonly known as k-nearest neighbors (kNN) was placed in between the fully connected feature layer and the output layer. This was done so that the classifier can handle the prediction more accurately and with less runtime. According to Murphy (Murphy, 1990), kNN is effective in developing arbitrary decision regions, the classification is specific to its trained data and can be completed in polynomial time. The incentive of using kNN is that it is a non-linear method that obtains more convoluted decision boundaries than the usual mapping technique used in the prediction layer of a generic convolutional neural network.

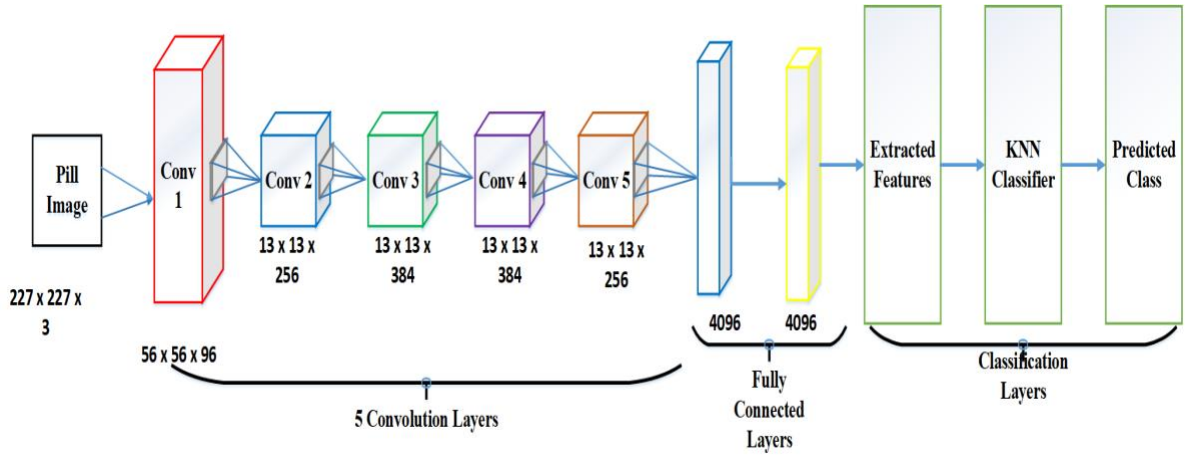
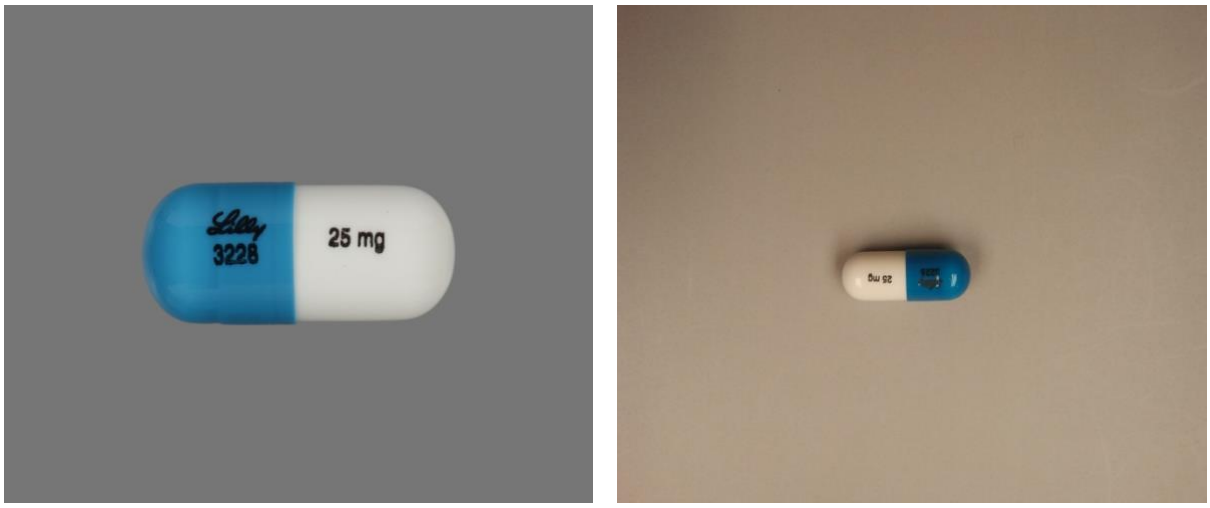


Figure 5. The Proposed Model

The CNN architecture used has 25 layers: one input layer, one output layer, and 23 hidden layers. The hidden layers consist of five convolutional layers, three max-pooling layers, seven rectified linear unit (ReLU) layers, three fully connected layers, two cross-channel normalization layers, two dropout layers, and one softmax layer. Furthermore, both algorithms were both exploited and paired in terms of feature extraction and classification.

The dimension of the input image fed into AlexNet's is 227 x 227 x 3. This input image was fed into (C1), the first convolutional layer. The first layer C1 has 96 kernels and a stride size of 4 x 4 which was then used in extracting the feature from each image (shape, color, and imprint). The extracted feature values from these images were then converted into feature vectors. Thereafter kNN was used in classifying the nearest feature vectors based on the proximity. This was done using the Euclidean (i) to determine the distance between the center and each pixel of each pill images.



(a) (b)

Figure 6 (a) Reference quality image; (b) Consumer quality image.

4.1 DISCUSSION

During comparison, the experimental results from trained CNN which was also tested on Support Vector Machine (SVM) alongside resNet50 for classification shows that the proposed model outperforms ResNet-50 and CNN with SVM classifier. Although all the three classification techniques performed well, the advantage of using this model was that a classifier like kNN used alongside the proposed neural network appreciably increases the accuracy with low noise. Also, the slightly higher segmentation and identification rates achieved by the proposed techniques indicated the features were more accurately classified when compared to the use of other techniques

While we recorded significant success during the classification of data using the proposed model, some drawbacks were encountered due to the type of available data. The understatement on the shape of the pills due to conflicting light conditions, placement of the pills and the distance from the camera used in the consumer quality images led to some minor

difficulties during the shape extraction from the images. This highlights that the proposed model doesn't perform efficiently as expected with high noise.

However, with a supervised mode of taking the pills images the model performs very well as expected. Also, we were able to expand our experience with using convolutional neural network and other classifiers. The performance and efficiency of the proposed model was impressive and would be suitable for implementation in applications for real time pill recognition.

4.2 EXPLORATORY DATA ANALYSIS

In this section, we discuss the exploratory data analysis where we investigated the ground truth table data to derive the data summary and gain more insights into the data and image information.

To uncover patterns, find irregularities, and generate sample summaries, we had to perform some investigations on the data. This was done by summarizing the data and gathering as many insights from it before training it.

I. **Summary:** We decided to investigate the ground truth table data to derive the data summary and gain more insights into the data: Our investigation showed that the dataset is a categorical data type which comprises of 10,000 observations and 2 characteristics with 3000 missing values. We also describe the first 5 observations of the dataset which can be seen in the Table 1.

	Reference quality image	Consumer quality image
1	00093-7155-98_PART_1_OF_1_CHAL10_SF_4A21A50D.jpg	1479.jpg
2	53489-0146-01_PART_1_OF_1_CHAL10_SF_8021C06E.jpg	3227.jpg
3	60505-1308-01_PART_1_OF_1_CHAL10_SF_4F21A7DD.jpg	2505.jpg
4	00781-5184-01_PART_1_OF_1_CHAL10_SF_6E21B76D.jpg	302.jpg
5	00172-5728-60_PART_1_OF_1_CHAL10_SF_5821AC5D.jpg	4076.jpg

Table 1. First 5 observations of the dataset

II. **Metadata**: After investigating the data summary, we went ahead to gain more insight into the image information. We were able to discover the information embedded within the pill images as seen in Table 2. It was discovered that the images were all taken in 2015, shot in a 24 bit-depth jpeg format, with TrueColor color type. The major differences; were the camera types, image sizes, and positioning. All the reference pill images were taken in a centered position while the consumer pill images were taken in a co-sited position.

	Reference quality image	**Consumer quality image**
Format	jpeg	jpeg
Width	2400	4416
Height	1600	3312
XResolution	72	180
YResolution	72	180
ColorType	TrueColor	TrueColor
BitDepth	24	24
YCbCrPositioning	Centered	Co-sited

Table 2. Metadata of a Reference image and of a Consumer quality image.

III. **Visualization:** Figure 7 below presents a visual representation of the data points of each pill via a scatter plot of pixel values for the dataset. Here, the position of each dot on the graph indicates the values for a single data point. The locations where we had more clusters were identified below as well as the locations of the outliers.

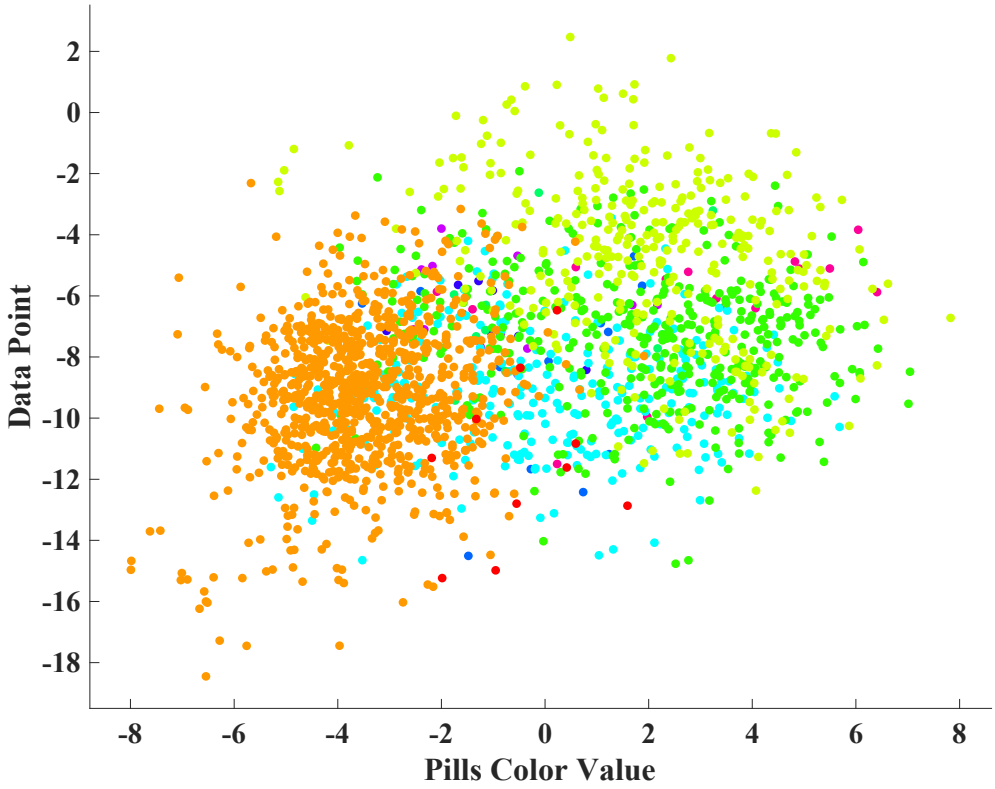


Figure 7. Scatter plot showing the data points of each pills.

Experiments were carried out using MATLAB R2018 on an Intel Xeon Gold 6140 @ 2.30GHz, 23 GB of RAM. After implementing the experiment using the aforementioned dataset, the performance and efficiency measures were recorded. The measures are described in (ii) – (iv) as follows:

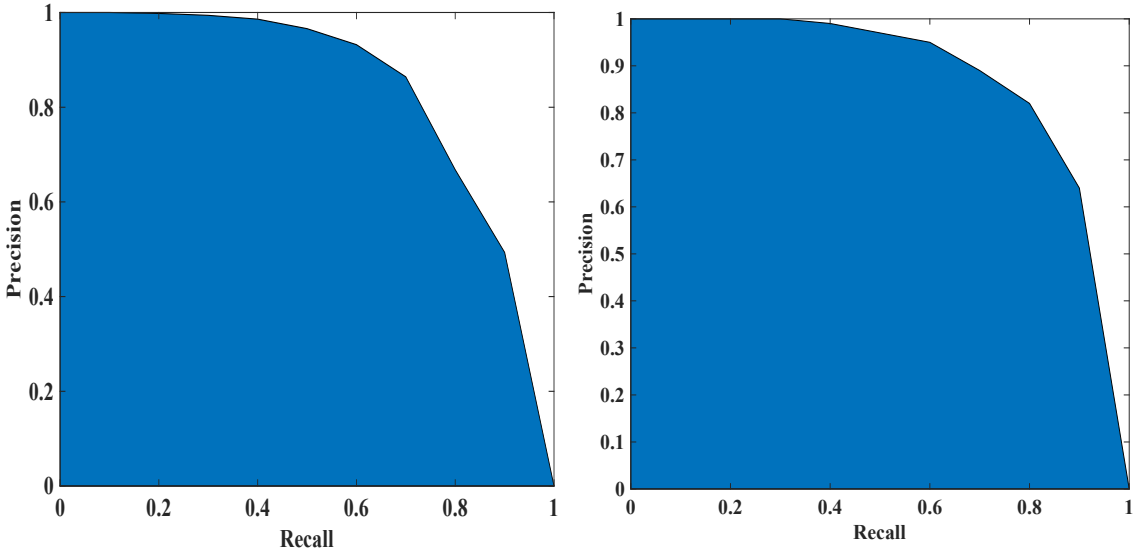
$$\text{Recall} = \frac{(TP)}{(TP+ FN)} \qquad \text{(ii)}$$

$$\text{Precision} = \frac{(TP)}{(TP+ FP)} \qquad \text{(iii)}$$

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad \text{(iv)}$$

Where, TP = true positive, TN = true negative, FP = false positive, and FN = false negative. The Recall measure indicated the percentage of the significant total result that was accurately classified by the methods used, while the Precision measure indicated the percentage of the significant result produced by the methods. A precision-recall curve was plotted for the different thresholds, Figure 8 shows the precision-recall curve of the ResNet-50 with the mean average precision (MAP) of 80% and the precision-recall curve of the proposed CNN integrated with support vector machine (SVM) with the MAP of 86%. Similarly, Figure 9

illustrated the precision-recall curve for CNN having KNN as a classifier and the mean precision accuracy (MAP) for this is 91%.



(a) (b)

Figure 8 (a) Precision Recall curve of ResNet-50, (b) Precision Recall curve of CNN+SVM.

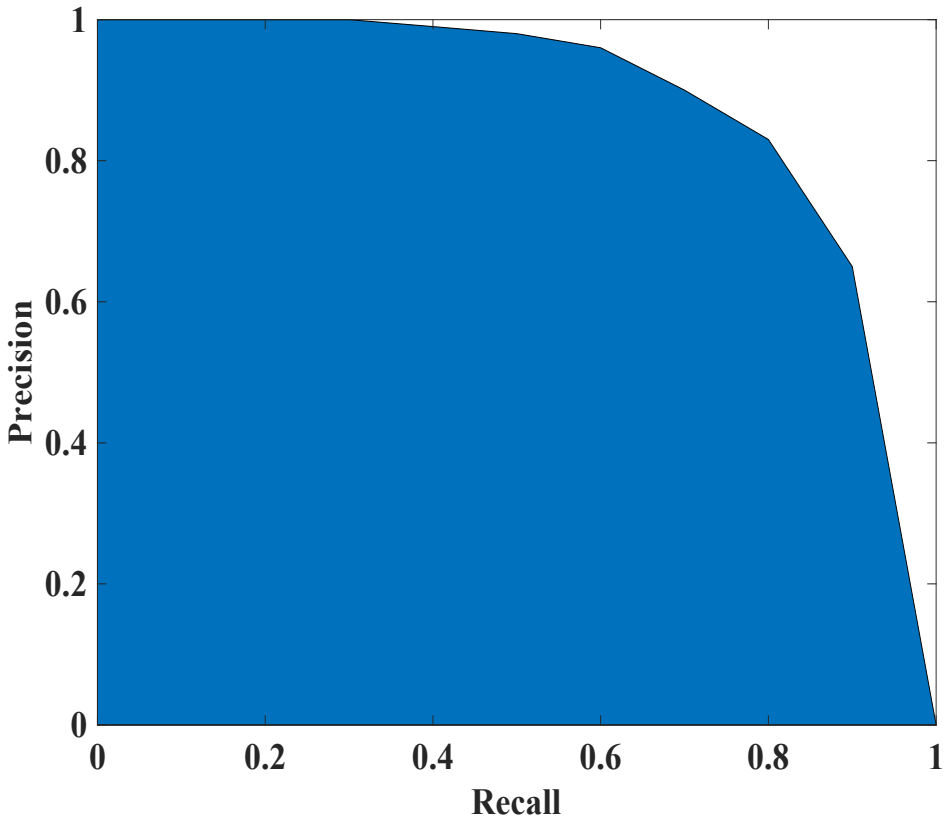
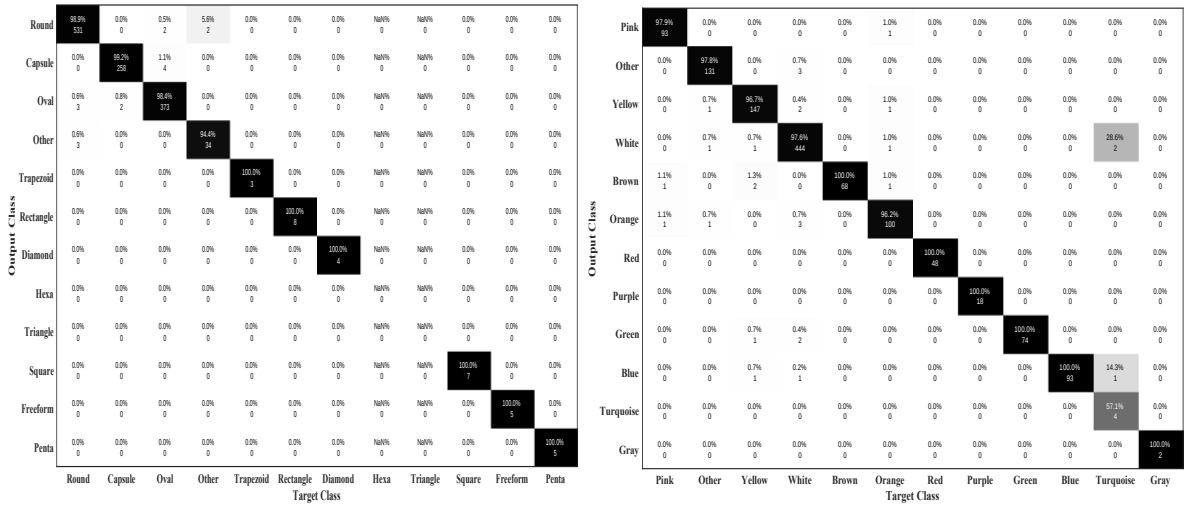


Figure 9. Precision Recall curve of CNN+KNN

The accuracy measure indicated the degree of consistency of the segmented pill images. The results obtained from the classification of the pills images in Figure 12 also show the visual results using the proposed technique on the NLM database.



(a) (b)

Figure 10 (a) Confusion matrix of predictions grouped by pill shapes; (b) Confusion matrix of predictions grouped by pill colors.

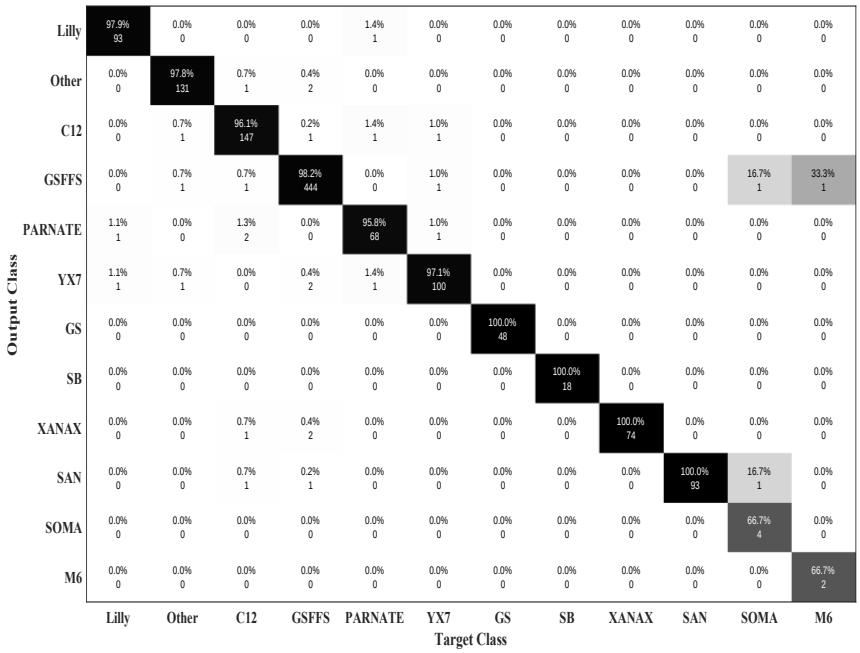


Figure 11. Confusion matrices of predictions grouped by pill imprints.

4.3 DATA UTILITY

Figure 14 shows the top-1 and top-5 accuracy rates of the different classification techniques using the NLM database. The Top 1 accuracy for the proposed CNN+KNN comes out to be 81% while this accuracy for ResNet-50 and CNN+SVM is 70% and 78% respectively. While top-5 search accuracy of the CNN+KNN is 95% and this accuracy for ResNet-50 and CNN+SVM is 86% and 92% respectively. Figure 13 shows that CNN+KNN achieves better top k accuracy than CNN+SVM and ResNet-50. From the CMC plot we can see the top 1 accuracy of CNN+KNN is 81% while the top-5 accuracy is 95%.

Figure 12. The visual results using the proposed technique on NLM database

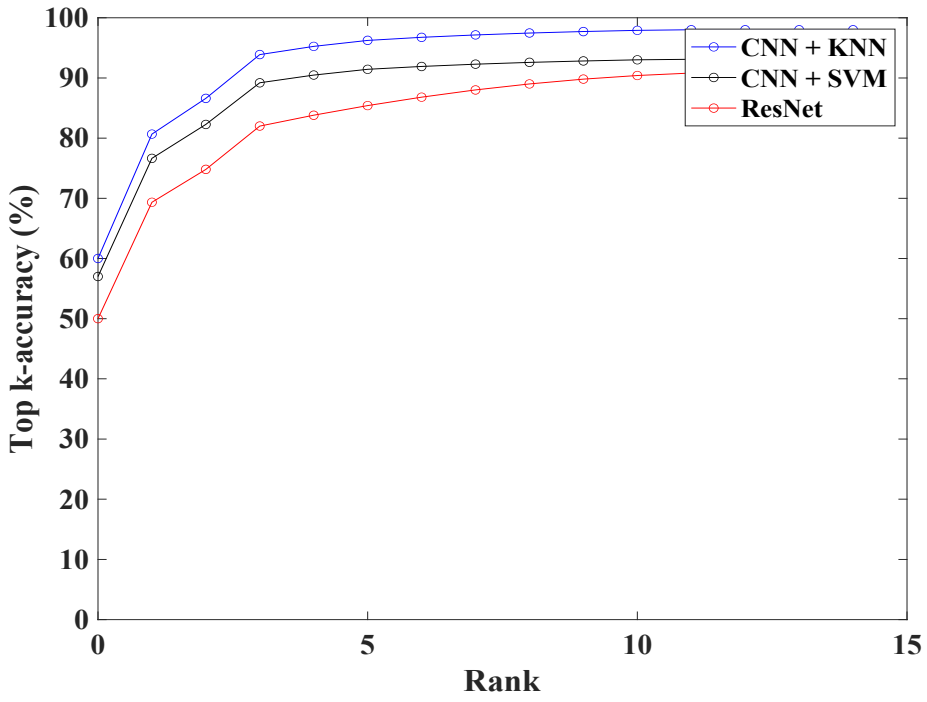
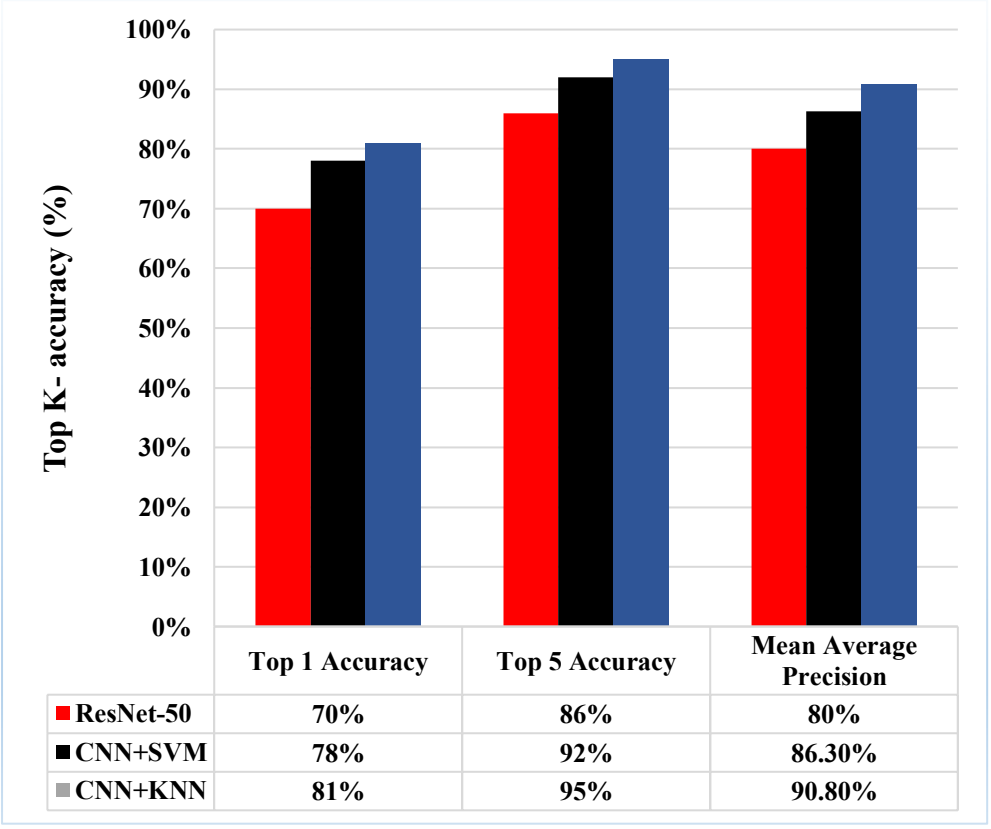


Figure 13. Cumulative Match Curves (CMC).



	Top 1 Accuracy	Top 5 Accuracy	Mean Average Precision
ResNet-50	70%	86%	80%
CNN+SVM	78%	92%	86.30%
CNN+KNN	81%	95%	90.80%

Figure 14. Top k Accuracies of different methods

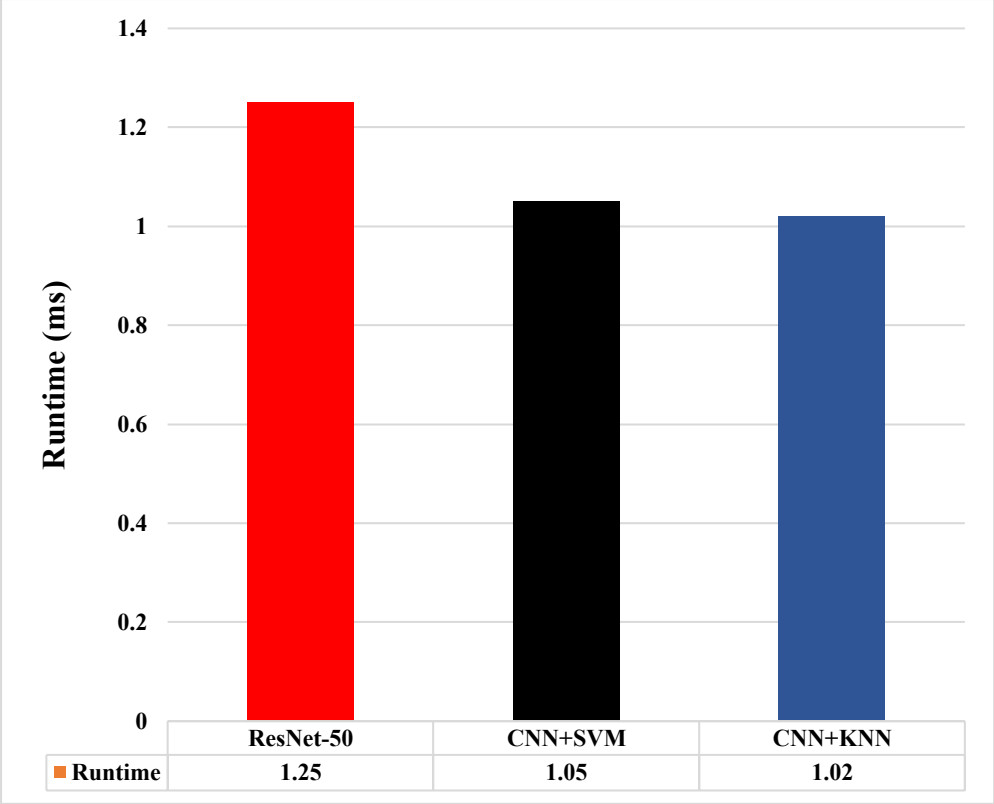


Figure 15. Runtime Performance Evaluation

Figure 15 shows the performance runtime of the techniques using the NLM database. The runtime of CNN + KNN is 1.02ms on CPU which is better than CNN+SVM and ResNet-50. The performance of the proposed system was also evaluated on the basis of precision and recall curves as defined below.

We also created confusion matrices for this model's predictions on the predicted classes. The confusion matrices show how accurate the model predicts each class and also evaluates the performance of the model. Figure 10 shows the confusion matrix of the proposed CNN integrated with kNN classifiers for the different colors of the pill drugs and the different shapes of the pill drugs. Similarly, Figure 11, shows the confusion matrix of CNN for the different imprints of the pill drugs. From these matrices we can see how accurate the model for each individual class is either on the basis of color or shape or imprint classes.

5 CONCLUSIONS

This paper proposed a hybrid method of image retrieval on a technique used for supervised machine learning to solve both classification and regression problems. The kNN classifier which is effective in developing arbitrary decision regions was placed in between the fully connected feature layer and the output layer to handle prediction accurately with less runtime. The results of the proposed method were compared with ResNet-50, CNN with SVM classifier. The results show that a classifier like kNN used alongside the proposed neural network appreciably increases the accuracy with low noise. In the future, we can introduce a more efficient method to tackle some drawbacks we encountered irrespective of the type of available data. This will help alleviate the underestimation of the shape of the pills due to conflicting light conditions which led to some minor difficulties during the shape extraction from the images.

BIBLIOGRAPHY

Adel'son-Vel'skii, G. M. (1962). An algorithm for organization of information. *In Doklady Akademii Nauk (Vol. 146, No. 2). Russian Academy of Sciences*, 263-266.

Bose, S. M. (2020). *CBIR using features derived by Deep Learning*. Retrieved November 15, 2020, from arXiv: https://arxiv.org/abs/2002.07877v1

Cafasso, J. (2018, September 18). *Medication Safety: Pill Identification, Storage, and More*. Retrieved September 26, 2020, from Healthline: https://www.healthline.com/health/pill-identification

Chang, S. K. (1984). Picture indexing and abstraction techniques for pictorial databases. *IEEE Transactions on Pattern Analysis and Machine Intelligence, (4)*, 475-484.

Chen, Z. Y. (2015). Accurate system for automatic pill recognition using imprint information. *IET Image Processing, 9(12)*, 1039-1047.

Crestani, F. M. (2017). Mobile information retrieval. *Springer International Publishing*.

Cunha, P. J. (2020). *Pill Identifier (Pill Finder Wizard)*. Retrieved September 26, 2020, from Pill Identification Tool Using Drug Pictures, Color, Shape, Number, or Imprint, Rxlist: https://www.rxlist.com/pill-identification-tool/article.htm.

Delgado, N. L. (2019). Fast and accurate medication identification. *NPJ digital medicine, 2(1)*, 1-9.

Eakins, J. G. (1999). *Content-based Image Retrieval*. Retrieved November 15, 2020, from Inf.fu-berlin.de: http://www.inf.fu-berlin.de/lehre/WS00/webIS/reader/WebIR/imageRetrievalOverview.pdf

FDA. (2020). *Code of Federal Regulations Title 21*. Retrieved September 26, 2020, from U.S. Food and Drug Administration: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=206.10.

Foster, C. C. (1965, August). Information retrieval: information storage and retrieval using AVL trees. *In Proceedings of the 1965 20th national conference*, 192-205.

Fung, V. (2017, July 15). *An Overview of ResNet and its Variants*. Retrieved September 26, 2020, from Towards Data Science: https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035

Geradts, Z. B. (2002). Content based information retrieval in forensic image databases. *Journal of Forensic Science, 47(2)*, 285-292.

Goodrum, A. A. (2000). Image information retrieval: An overview of current research. *Informing Science, 3(2)*, 63-66.

Grigorescu, C. &. (2003). Distance sets for shape filters and shape recognition. *IEEE transactions on image processing, 12(10)*, 1274-1286.

Guo, G., & Wang, H. B. (2003). KNN model-based approach in classification. *In OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, 986-996.

Guo, P. S. (2017). Color Feature-based Pillbox Image Color Recognition. *In VISIGRAPP (4: VISAPP)* , 188-194.

He, K., & Zhang, X. R. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

He, K., & Zhang, X. R. (2016). Identity mappings in deep residual networks. *In European conference on computer vision Springer, Cham*, 630-645.

Krizhevsky, A. S. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research, 15(1)*, 1929-1958.

LeCun, Y. B. (2015). Deep learning. *Nature, 521(7553)*, 436-444.

Lee, Y. B. (2012). Pill-ID: Matching and retrieval of drug pill images. *Pattern Recognition Letters, 33(7)*, 904-910.

Makary, M. A. (2016). Medical error—the third leading cause of death in the US. . *BMJ*, 353.

Maron, M. E. (1960). On relevance, probabilistic indexing and information retrieval. *Journal of the ACM (JACM), 7(3)*, 216-244.

Murphy, O. J. (1990). Nearest neighbor pattern classification perceptron. *Proceedings of the IEEE, 78(10)*, 1595-1598.

NLM. (2016). *Pill identification challenge*. Retrieved from National Library of Medicine: https://www.nlm.nih.gov/databases/download/pill_image.html

Razavian, A. A. (2014). CNN features off-the-shelf: an astounding baseline for recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 806-813.

Ribera, J. W. (2017, April). Pill recognition using minimal labeled data. *In 2017 IEEE Third International Conference on Multimedia Big Data (BigMM). IEEE*, 346-353.

Robertson, S. E. (1976). Relevance weighting of search terms. *Journal of the American Society for Information science, 27(3)*, 129-146.

Salton, G. (1971). The SMART system. Retrieval Results and Future Plans. *Experiments in Automatic Document Retrieval.*

Simonyan, K. Z. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*, 1409-1556.

Stanchev, P. R. (1987). An approach to image retrieval from large image databases. *In Proceedings of the 10th annual international ACM SIGIR conference on Research and development in information retrieval*, 284-295.

Szegedy, C. L. (2015). Going deeper with convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9.

Taigman, Y. Y. (2014). Deepface: Closing the gap to human-level performance in face verification. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 1701-1708.

Vapnik, V. (1998). *Statistical learning theory.* Boston, USA: 1st EdWiley.

Wang, X. Y. (2010). Robust image retrieval based on color histogram of local feature regions. *Multimedia Tools and Applications, 49(2)*, 323-345.

WHO. (2020). *10 facts on patient safety*. Retrieved September 26, 2020, from World Health Organization: https://www.who.int/features/factfiles/patient_safety/en/.

WHO. (2020). *Medication Without Harm: Real-life stories*. Retrieved September 26, 2020, from World Health Organization: https://www.who.int/patientsafety/medication-safety/photostory/en/

Yu, J. C. (2014). Pill recognition using imprint information by two-step sampling distance sets. . *In 2014 22nd International Conference on Pattern Recognition. IEEE*, 3156-3161.

Zeng, X. C. (2017, June). MobileDeepPill: A small-footprint mobile deep learning system for recognizing unconstrained pill images. *In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, 56-67.