applied sciences

MDPI

Article

CNN-Based Pill Image Recognition for Retrieval Systems

Khalil Al-Hussaeni [1,*], Ioannis Karamitsos [2], Ezekiel Adewumi [2] and Rema M. Amawi [3]

[1] Computing Sciences, Rochester Institute of Technology, Dubai 341055, United Arab Emirates
[2] Graduate and Research, Rochester Institute of Technology, Dubai 341055, United Arab Emirates
[3] Science and Liberal Arts, Rochester Institute of Technology, Dubai 341055, United Arab Emirates
* Correspondence: kxacad@rit.edu

Abstract: Medication should be consumed as prescribed with little to zero margins for errors, otherwise consequences could be fatal. Due to the pervasiveness of camera-equipped mobile devices, patients and practitioners can easily take photos of unidentified pills to avert erroneous prescriptions or consumption. This area of research goes under the umbrella of information retrieval and, more specifically, image retrieval or recognition. Several studies have been conducted in the area of image retrieval in order to propose accurate models, i.e., accurately matching an input image with stored ones. Recently, neural networks have been shown to be effective in identifying digital images. This study aims to provide an enhancement to image retrieval in terms of accuracy and efficiency through image segmentation and classification. This paper suggests three neural network (CNN) architectures: two models that are hybrid networks paired with a classification method (CNN+SVM and CNN+kNN) and one ResNet-50 network. We perform various preprocessing steps by using several detection techniques on the selected dataset. We conduct extensive experiments using a real-life dataset obtained from the National Library of Medicine database. The results demonstrate that our proposed model is capable of deriving an accuracy of 90.8%. We also provide a comparison of the above-mentioned three models with some existing methods, and we notice that our proposed CNN+kNN architecture improved the pill image retrieval accuracy by 10% compared to existing models.

Keywords: image recognition; pill information retrieval; CNN; CBIR; machine learning; convolutional neural networks

Citation: Al-Hussaeni, K.; Karamitsos, I.; Adewumi, E.; Amawi, R.M. CNN-Based Pill Image Recognition for Retrieval Systems. *Appl. Sci.* **2023**, *13*, 5050. https://doi.org/10.3390/app13085050

Academic Editor: Chien-Hung Yeh

Received: 4 February 2023
Revised: 12 April 2023
Accepted: 13 April 2023
Published: 18 April 2023

1. Introduction

Information Retrieval describes the process of sourcing information from a storage system. The retrieved information may be in the format of text, image, sound, or metadata describing a database or data. One interesting area is information retrieval from images, whereby an automated tool is used to identify objects in images. In this day and age, the increased dependency on smartphones has made informational retrieval from mobile phone photos a growing area of research [1].

Traditionally, metadata, such as keywords, captions, or image titles, have helped in information retrieval. However, this manual approach consumes time, effort, and costs. With increasing online activities, including social web applications, research on Content-Based Information Retrieval (CBIR) has become prominent in the field of Information Retrieval. CBIR is the field that describes automated image retrieval techniques that are capable of identifying images based on their "content", i.e., features embedded in the image, such as shape, texture, and color [2–4]. Research is still ongoing to improve the effectiveness of CBIR in terms of extracting primitive features (color and shape) and creating abstraction models to identify the level of relevance. The advances in image retrieval techniques have carved the path to applications in a variety of fields, including medicine, law enforcement, and engineering. Automated pill image recognition remains a significant application of CBIR in the field of medicine.

Given the criticality of medicine consumption, there is little to no room for errors, e.g., mistakenly prescribing or taking the wrong type of medicine. Yet, there is a high possibility of errors occurring while health personnel prescribe, dispense, or administer drugs. Makary and Daniel [5] argued that medical error ranks third on the major causes of death among hospital inpatients in the United States. WHO (the World Health Organization) statistics reveal that approximately 1.3 million patients die annually due to preventable medication blunders in the United States, which comes out to a minimum of one death per day. Moreover, WHO also admits medical error to be one of the top causes of injuries and avoidable harm [6]. Adverse Drug Effects (ADE) can also result in severe ailments, including Stevens–Johnson syndrome and Parkinson's disease [7]. WHO statistics report that health caregivers harm 4 out of every 10 patients globally [8], and Larios Delgado et al. [9] report that 39% of cases are severe enough to cause injury to patients.

Consumers often find it challenging to identify pills; consequently, they run into the risk of harming themselves from either consuming the wrong medication, underdosing, or overdosing. The risk of misidentifying pills is more prominent when pills are moved to different packaging containers, combined into a single container, or shared into pillboxes for ease of administration. Moreover, the financial implications of medication error are alarming: One-seventh of the Canadian budget and about 1% ($42 billion) of the total global health spending are spent on mitigating the effects of medication error [8].

To ensure safe medication consumption, each pill is made to have a distinctive appearance by having a unique combination of size, color, shape, and imprint [10]. An unidentified pill can, therefore, be cross-referenced by health practitioners against a database of prescription drugs. Pharmacists usually help their patients during a brown bag consultation with the drugs they bring in for identification. A manual search can be tedious, exhausting, and time-consuming, particularly when dealing with many pills with large generic variations. Moreover, reading tiny imprints on small drugs can easily introduce human error. Alternatively, automated pill recognition techniques can help identify pills rather quickly, decrease the possibility of pill misidentification, and provide visual assurance to the patient. Examples of such automation are the RxList Pill Identification Tool [11] and the Healthline Pill Identifier [12], which are web-based applications offering pill identification services.

The concept of identifying pills from images has been studied, particularly using deep neural networks, with promising results. However, unlike these studies, our proposed approach does not only use neural networks, but also incorporates the non-parametric classifier known as k-Nearest Neighbors (k-NN) [13,14]. The classifier k-NN is effective in developing arbitrary decision regions and can complete in polynomial time. Moreover, k-NN can obtain more convoluted decision boundaries than the usual mapping technique used in the prediction layer of a generic convolutional neural network. We summarize the contributions of our work as follows:

1. We investigate the challenging problem of image retrieval, specifically targeting pill images.
2. We develop an efficient image retrieval system based on deep learning and the k-Nearest Neighbor (k-NN) classifier.
3. We employ a real-life dataset of pill images to evaluate the proposed system against accuracy and runtime, as well as compare the results with relevant image retrieval systems from the literature.
4. Our proposed model increased the accuracy of identifying pills form images by 10% while maintaining the same runtime as comparable methods.

The paper is organized as follows. Section 2 surveys the literature for related work that has been conducted on information retrieval in general and information identification from images, specifically. The proposed method and architecture are discussed in Section 3. Experimental settings and results are detailed in Section 4. A discussion of the results and comparisons are presented in Section 5. Finally, the paper concludes in Section 6.

2. Related Work

A huge amount of work has been done by researchers over the years to improve information retrieval from stored data. The literature contains various models that significantly contributed to this area of research. In this section, we survey some of the most prominent models and approaches in the area of information retrieval by briefly going over the history and then focusing on work in pill image identification.

Probabilistic information retrieval using weighted indexing was introduced in the 1960s by Maron and Kuhns [15]. The authors of [16] proposed to store and organize information using a tree-like structure, called the Adel'son-Vel'skiy and Landis (AVL) tree. Chang and Liu [17] improved the work done by Foster [18] by proposing a picture indexing and abstraction method, which led to a paradigm shift in image retrieval. Salton and Lesk [19,20] proposed one of the most prominent advancements to retrieval by developing a method called System for the Mechanical Analysis and Retrieval of Text (SMART). Rabitti and Stanchev [21] proposed a non-text based approach for retrieving images from an extensive image database.

The use of color histograms was explored by Wang et al. [22]. They explained that Local Feature Regions (LFR) would be more effective in retrieving images. On the same note of color histograms, Lee et al. [23] utilized Wang et al.'s color histogram approach to propose an automatic pill recognition system based on pill imprint, which encompassed three features: shape, color, and texture. Lee et al. extracted feature vectors based on edge localization and invariant moments of tablets as an identifier. Their experimental results showed 73% matching accuracy over a dataset of 13,000 legal drug pill images.

Deep learning techniques have been introduced in Content-Based Information Retrieval (CBIR). Such techniques are used to enhance feature extraction from input images in order to identify and retrieve similar images from a database [24]. Deep learning has been impressive in its competence in recognizing objects [25,26] and faces [27], and in handling extensive learning problems [28]. Deep learning has also improved clinical workflows by enhancing the experience of both the caregivers and patients [9]. Several deep learning-based models exist in the literature, such as the Convolutional Neural Network (CNN) [24], GoogLeNet [29], AlexNet [30], and the Residual Network (ResNet) [31]. A Convolutional Neural Network (CNN) is a deep learning technique for digital image retrieval. A CNN architecture comprises a sequence of interacting convolutional, pooling, and fully connected layers [24].

Several techniques have been developed for pill image recognition with different accuracy levels and limitations [32,33]. MobileDeepPill [34] is a CNN architecture that integrates pill color, gradients, and shape measurements to compare between consumer and reference images. Guo et al. [35] used a support vector machine (SVM) to study the color property, wherein they achieved a 97.90% overall color classification accuracy. However, the effectiveness of Guo et al.'s technique is limited by certain factors such as the lighting condition, the camera resolution, and the pill and background color contrast. Some pill recognition techniques have been developed to identify a pill based on only a subset of shape, color, and imprint, such as the works in [36–38]. The work in [39] identified pills that had only one of four pre-identified colors and classes.

Recently, Kwon et al. [40], Holtkötter et al. [41], and other similar works proposed neural network-based methods to detect pills from images. Unlike our method that aims to identify a single pill from a pill image, the approaches in [40,42,43] focused on identifying a pill from an image containing a group of pills. The studies in [41,44] aimed to detect the presence of pills in an image of a blister to track oral pill intake. The work by Nguyen et al. [45] utilized external help, namely, extracted information from prescriptions, to learn the potential associations between pills. While the problems in these studies differ from ours, we believe that our study complements the body of work in the literature by proposing an accurate *and* efficient method for identifying pills.

3. Methodology

This study aims to enhance image retrieval accuracy and efficiency to minimize clinical errors when prescribing drugs, particularly pills. Overall, the scenario is as follows: A medical practitioner or patient takes a photo of an unidentified pill. Then, the pill photo (query image) is sent to our proposed system for identification against an existing database of pill images. The challenge lies in the fact that photos may be captured in less-than-ideal environments. For example, the photo could be captured using a low-quality camera, in a room with not enough lighting, from different angles, or with a noisy background.

In order to tackle the above challenge, we propose an image retrieval approach based on two steps: (1) a preprocessing phase with features extraction and (2) classification. The overall proposed methodology is illustrated in Figure 1.

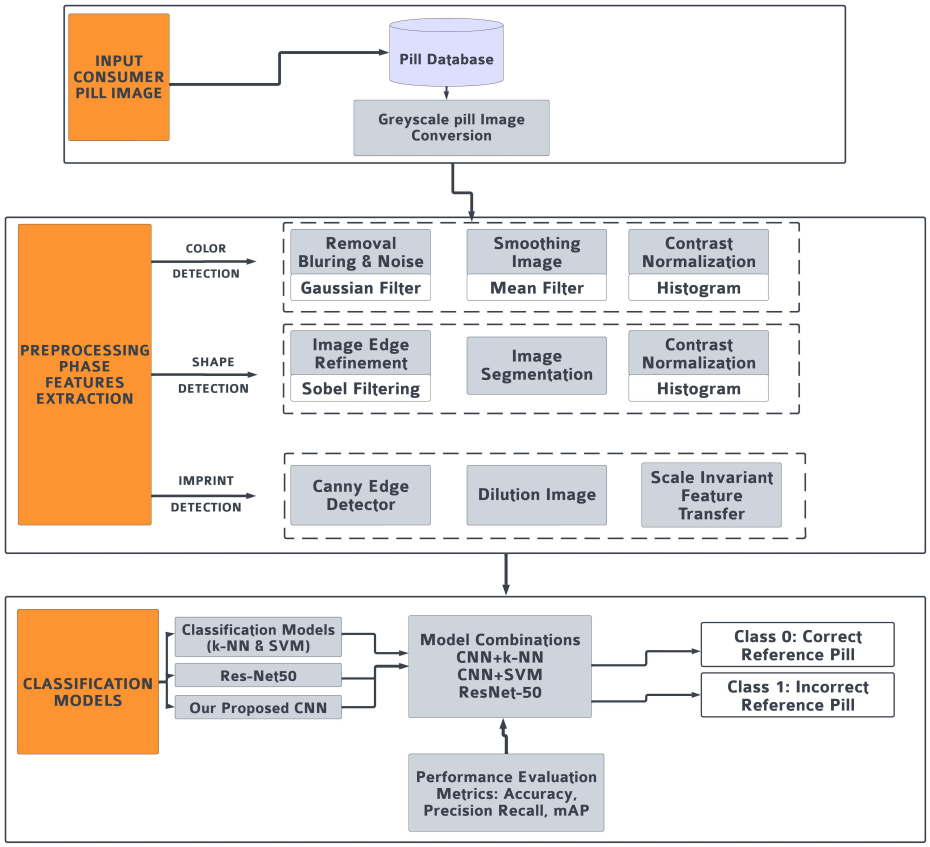


Figure 1. Proposed Approach.

3.1. Preprocessing and Features Extraction

An input pill image (query image) undergoes a series of preprocessing processes in order to make up for the color distortion and identify relevant information. The overall preprocess shows 3 main steps for the detection and extraction of color, shape, and imprint.

Before a pill image undergoes the segmentation steps in Figure 1, the image is converted to a grayscale format. This preprocessing step is used to regulate the intensity of the red, green, and blue (RGB) components in the image. Therefore, it is essential to denote a single intensity value for each pixel. Figure 2 shows an example of a raw input pill image (Figure 2a) and its grayscale version (Figure 2b). We note that all the pill images shown in the various figures in this paper are from the National Library of Medicine (NLM) pill image dataset [46].

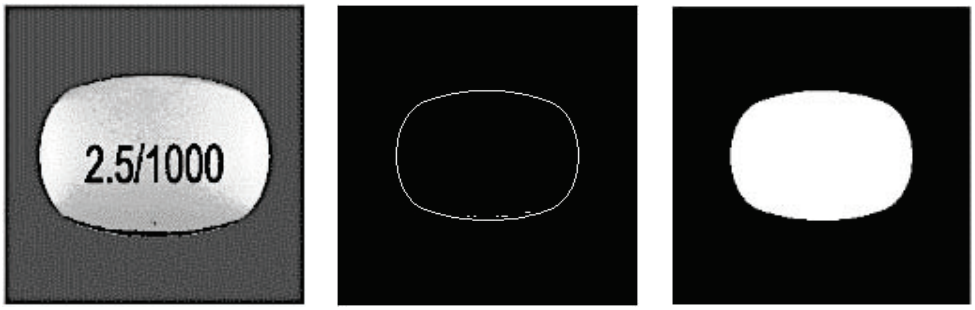
(**a**) (**b**)

Figure 2. Colored vs. grayscale intensity pill image: (**a**) Raw input pill image; (**b**) pill image after grayscale conversion.

For color detection, a Gaussian filter is applied to the greyscale image to blur the image, thus removing unwanted details and noise. After that, a mean filter is applied to the output of the Gaussian filter to smooth the image. Next, histogram equalization is used to enhance the color contrast and to extract the colors. Figure 3 visualizes the color detection process of the same pill image in Figure 2b. For shape detection and extraction, we use Sobel filtering on the greyscale image to refine the image, which helps reveal the edges and the boundary lines of the drug pill. Figure 4 visualizes the shape detection and extraction process of the same pill image in Figure 2b. Lastly, for imprint extraction, we apply a Canny edge detector to determine all the edges in the image, followed by a dilution operation to soften the image. Clear imprint is finally revealed after applying Scale Invariant Feature Transform (SIFT) and Multi-Scale Local Binary Pattern (MLBP) descriptors. Figure 5 visualizes the imprint extraction process of the same pill image in Figure 2b.

(**a**) (**b**) (**c**)

Figure 3. Pill image after applying various filters for color detection: (**a**) Gaussian filter. (**b**) Mean filter. (**c**) Local histogram equalization filter.



(**a**) (**b**) (**c**)

Figure 4. Pill image after applying various filters for shape detection: (**a**) Local histogram equalization filter. (**b**) Sobel filter. (**c**) Segmented shape.

(**a**) (**b**) (**c**)

Figure 5. Pill image imprint extraction process: (**a**) Canny edge detector. (**b**) Dilution. (**c**) Scale Invariant Feature Transform.

3.2. Proposed CNN Architecture

The first step constructs a Convolutional Neural Network (CNN) to extract the query image features, namely, shape, color, and imprint. The second step uses a classifier to match the extracted query image features with those of an existing pill image. The overall proposed architecture is illustrated in Figure 6.

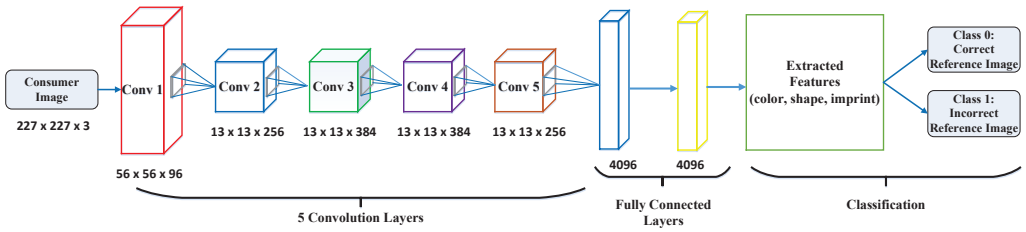


Figure 6. An overview of the proposed CNN model architecture.

The first layer of the proposed CNN network is responsible for accepting the input pill image. In our case, the input layer accepts RGB images of size 227 × 227 pixels. After that, they are fed into the CNN model, which processes them as follows:

- Pill images go through one convolutional layer (Conv1) with 56 × 56 × 96, which means that the input to the layer is a pill image with a height and width of 56 pixels and that has 96 color channels.
- The resulting tensors (images) go through four additional convolution layers with a smaller height and width (13 pixels) than the previous layer, and the number of input channels is increased to 256 color channels.
- The resulting feature maps are converted to a fully connected (FC) layer of 4096 neurons, which is connected to a second fully connected layer of 4096 neurons.
- Then, the extracted features (color, shape, and imprint) are fed into the classification layers; we then employ a k-NN classifier to handle the prediction more accurately and with less runtime.
- Finally, the classification layers output a predicted class, i.e., a matched set of images from a stored database. For more details on the CNN layers and processing, we refer the reader to [30,47].

We also note that we use the terms "prediction" and "identification" interchangeably throughout the paper.

3.3. Classification

After extracting the features of an input raw pill image, the next step is to predict the pill type using a classifier. Classification is a supervised machine learning technique where the class (pill type) to be predicted is known in advance. Several classifiers exist

in the literature, though the vary in accuracy and efficiency. Below are some of the most prominent classifiers.

1. k-Nearest Neighbors (k-NN) [13,14] is a non-parametric classifier that assumes similar objects (i.e., data points) are usually "closer" to one another in comparison to dissimilar objects. k-NN measures similarity between data points using distance metrics. One of the most common distance metrics is the Euclidean distance and is measured by the following function:

$$d(X, Y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}, \quad (1)$$

   where X and Y are two data points in the n-dimensional space, and x_i and y_i are Euclidean vectors from the point of origin. When our proposed model receives a query image, the model converts the image to feature vectors, which the classifier will use to predict the pill type in the query image. We set k to 5 for all our experimental analysis in Section 4.
2. The Support Vector Machine (SVM) [48] is a classifier that, when given a set of input objects, creates an imaginary wall that separates dissimilar objects. This imaginary wall is called a *hyperplane,* because it can separate data points represented in spaces beyond three-dimensional. Given a set of input data points, there are several potential hyperplanes that the SVM can create. The SVM creates the best separation between the data points, i.e., it only keeps the hyperplane that minimizes the classification error.
3. Residual Network [31], or otherwise known as ResNet, is a neural network-based model that can be used as a final identifier in a convolutional layout. ResNet can accommodate more than 50 layers and be used to classify and extract features in an image. This technique makes use of skip connections to reduce the training error and help add the output of earlier layers to later layers without losing the image quality.

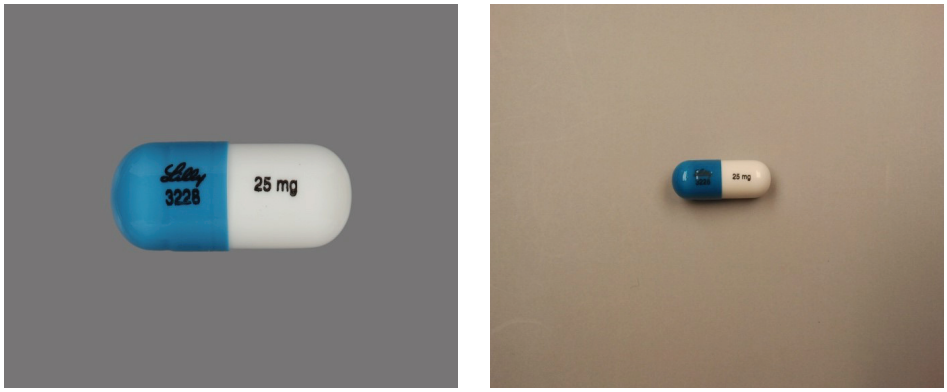
4. Results

The proposed model was implemented using MATLAB R2018 on an Ubuntu virtual machine with 100 GB of HDD, 24 GB of RAM, and 6 CPUs at 2.5 GHz. After that, we designed a set of experiments to evaluate the performance of our proposed model in terms of accuracy (percentage of correctly predicted pill types) and efficiency (runtime until completion).

4.1. Dataset

The proposed method was evaluated using pill images from the publicly available National Library of Medicine (NLM) dataset [46]. The NLM dataset comprises 7000 pill images from 1000 unique pills. Each pill image is categorized either as a *reference* image or as a *consumer* image. Figure 7 illustrates these two categories; Figure 7a shows a sample pill in a reference image; and Figure 7b shows the same pill in a consumer image. Reference images were taken under regulated conditions, thereby ensuring appropriate control over lighting and background. The NLM dataset contains 2000 reference images that belong to 1000 unique pills (each of which has a front and back image). On the other hand, consumer images were taken in such a way to mimic the quality of images that users would capture using their mobile phone cameras. That is, consumer images vary in quality, focus, and device types. The NLM dataset comprises 5000 consumer images, where each of the 1000 unique reference image has 5 associated consumer images.

Table 1 summarizes the metadata of the reference and consumer images, respectively. Images were shot in a 24 bit-depth jpeg format with a TrueColor color type. The major differences between the reference images and consumer images lie in the camera types, image sizes, and positioning. All the reference images were taken in a centered position, whereas the consumer images were taken in a co-sited position.



(**a**) (**b**)

Figure 7. Sample pill image: (**a**) Reference version. (**b**) Consumer version.

Table 1. Metadata of reference and consumer images.

Features	Reference Image	Consumer 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

4.2. Performance Analysis

Given an input pill image taken by a consumer, we wished to evaluate how accurate our model was at identifying the corresponding reference pill image based on pill shape, color, and imprint.

Figure 8 visually showcases the result of applying our proposed pill image recognition model using the pill images in the NLM dataset. Each object in Figure 8 is a pill. Matched pills (consumer image and its corresponding reference image) were put next to each other. The objective of this figure is to visually demonstrate the overall accuracy of the proposed model. In the remainder of this section, we will use widely adopted accuracy metrics, namely, mean Average Precision (mAP), confusion matrices to measure True Positives, and Top-k Accuracy. Moreover, we compared our model and labeled *CNN+kNN* with *CNN+SVM* and *ResNet-50* [31].

The above-mentioned three accuracy metrics are based on the notions of Precision and Recall. Precision measures the fraction of correct identifications among all positive identifications. Recall measures the fraction of correct identifications among all the dataset's actual positives. The term "positive" refers to a target class; in this case, a pill. Below are the equations for Precision, Recall, and Accuracy:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}, \tag{4}$$

where TP (True Positive) is the number of correct predictions of a target class, FN (False Negative) is the number of wrong predictions of a target class, FP (False Positive) is the

number of wrong predictions of a non-target class, and *TN* (True Negative) is the number of correct predictions of a target class.

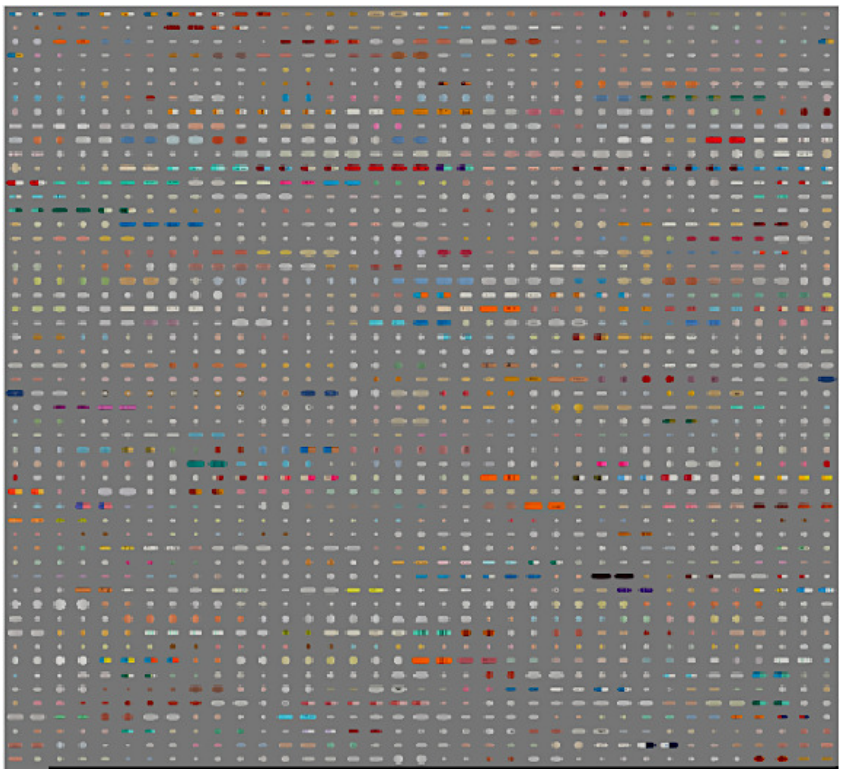


Figure 8. Matching the NLM dataset consumer pill images with reference pill images using the proposed model.

For a classifier to correctly predict a target class, the classifier must find an "acceptable" match between a query image (consumer pill image) and a target image (reference pill image). This match is numerically defined by a *threshold* that measures the fraction of the overlapping area between the query image and the target image. Based on this threshold, the values of Precision and Recall vary.

Another performance metric that is commonly used in the evaluation of information retrieval and object detection systems is the mean Average Precision (mAP). The mAP metric measures the average precision values of a classifier across different Recall values. A higher mAP score indicates better performance of the model in retrieving relevant information or detecting objects accurately. The mAP incorporates the trade-off between Precision and Recall, and it considers both false positives (FP) and false negatives (FN). This measurement provides a broader understanding of the classifier accuracy in identifying pills. The mAP is calculated as follows:

- For each object class, calculate the Average Precision (*AP*) as:

$$AP = \frac{1}{n} * \sum_{k=1}^{n} Precision\_at\_each\_k\text{-}object \tag{5}$$

  where (n) is the total number of relevant items in the dataset for the given object class, and the Precision at each relevant k-object is the Precision calculated at the position of the relevant item in the ranked list of predicted items.
- Calculate the *mAP* as the mean of the *AP* scores for all object classes:

$$mAP = \frac{1}{N} * \sum_{k=1}^{N} AP_k \tag{6}$$

where (N) is the total number of object classes in the dataset.

Given the pill images in the NLM dataset, we calculated the *mAP* performance metric for *ResNet-50*, *CNN+SVM*, and *CNN+kNN* (our proposed model), for a $0.1 \leq$ threshold ≤ 0.9. The mAP metric comparison is illustrated in Table 2. Moreover, we plotted the Precision–Recall curves of the three models. A Precision–Recall curve of a prediction model visualizes the accuracy of the model. The larger the area under the curve is, the better the prediction quality (reflecting both good Prediction and Recall). Figure 9 shows three Prediction–Recall curves for the above-mentioned three models, respectively.

Table 2. mAP Comparison Models.

Models	Mean Average Precision (mAP)
ResNet-50	80%
CNN+SVM	86.3%
CNN+kNN (our proposed model)	90.8%

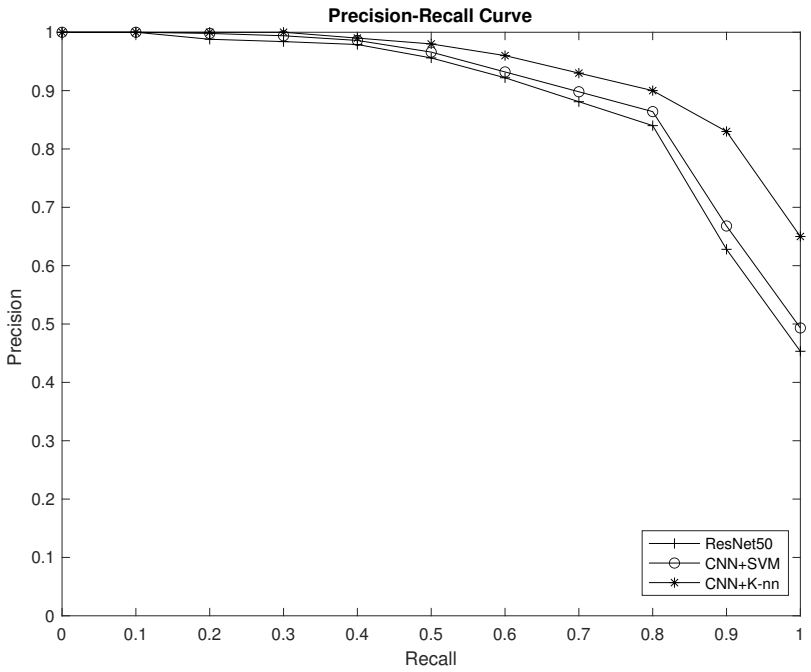


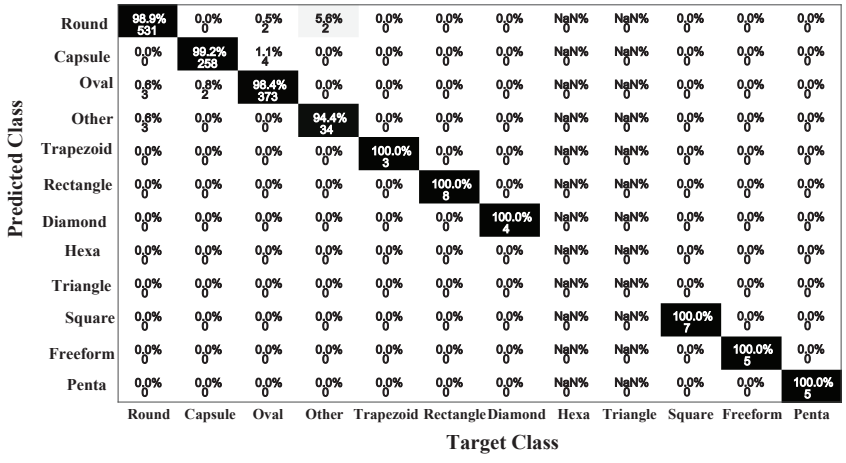
Figure 9. Precision–Recall curves of three pill identification models: *ResNet-50*, *CNN+SVM*, and *CNN+kNN*.

The Precision in Equation (2) is computed upon determining the TP and FP values of a target class in a prediction task. These values can be determined after running the prediction model against a dataset of images. For example, if a query image contains a round pill shape, then is the model able to predict that the shape of the query image is in fact round? In other words, we would like to know how many round-shaped pills the model successfully predicted as round-shaped (as opposed to any other shape).

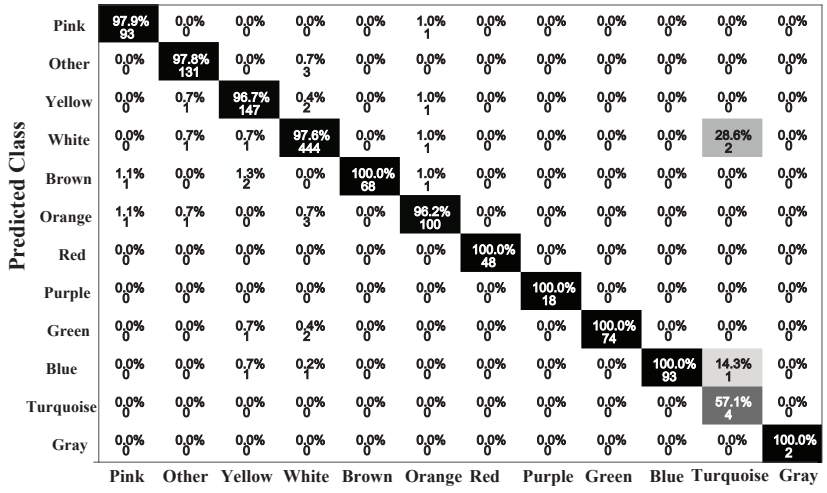
In the above example, the target class was *Round*, which belongs to the "shape" property of a pill. Our model extracts three features from any query pill image: shape, color, and imprint (see Figure 6).

To help us evaluate the performance of our model in terms of correctly identifying target classes, we constructed a confusion matrix for each pill feature. Figure 10 represents three confusion matrices for shape, color, and imprint, respectively. The x-axis and y-axis arbitrarily list values (or target classes) of a specific feature (e.g., *Round*, *Capsule* and *Oval* are values of the shape property). The x-axis represents the known target class, whereas the y-axis represents the predicted class. An intersection between any pair of values (c_x, c_y)

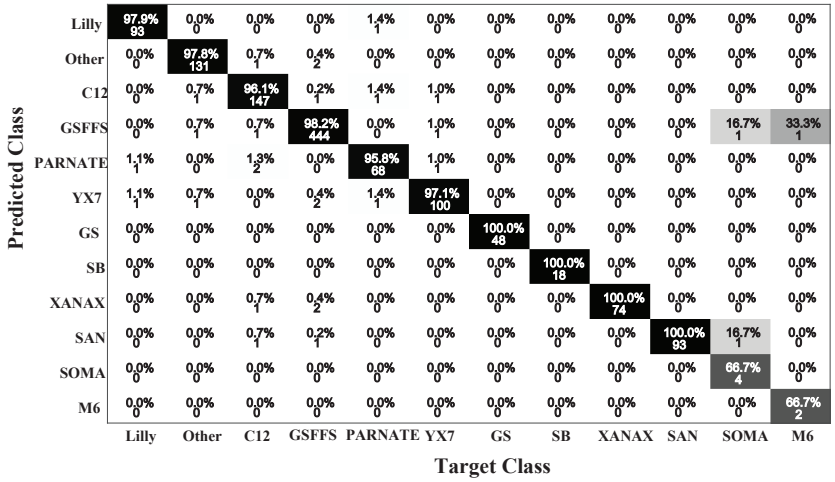
on the x-axis and y-axis, respectively, represents the number of times (or percentage) that the model predicted c_y given a pill of a target class of c_x.



(**a**)



(**b**)



(**c**)

Figure 10. Confusion matrices reporting our model accuracies w.r.t. predicting pill: (**a**) shape, (**b**) colors, and (**c**) imprints.

For example, Figure 10 demonstrates the ability of the proposed model to understand the shape of the query pill. Looking at the *Round* target class on the x-axis, the model predicted *Round* as *Round* 98.9% of the times, but predicted *Round* as *Oval* 0.6% of the times.

The Top-k Accuracy metric considers a model's prediction to be correct if the target class exists among the top k predictions. For example, given a query pill image I_q from the NLM dataset and k = 5, if the model matches I_q with an ordered set of n images $\langle I_1, I_2, I_3, \ldots, I_n \rangle$, where $I_q \in \{I_1, I_2, I_3, I_4, I_5\}$, then the Top-k Accuracy metric considers this match to be a correct prediction. It stands to reason that, as k increases, the Top-k Accuracy is expected to increase, because the metric considers a larger pool of potential matches. For k = 1 and 5, Table 3 reports the results of the Top-k Accuracy of the four related models.

Table 3. Top-k Accuracy in Related Models (%).

Method	Top-1	Top-5
SqueezeNet [49]	49.0%	76.8%
AlexNet [30]	62.5%	83.0%
ResNet-50 [31]	71.7%	85.5%
MobileNet [50]	71.7%	92.0%
MobileDeepPill [34]	73.7%	95.6%
InceptionV3 [51]	74.4%	93.3%
CNN+SVM (our suggested model)	76.5%	92.0%
CNN+kNN (our proposed model)	80.5%	96.1%

Based on the evaluations performed by Larios Delgado [9], *ResNet-50* [31] performed the best among all the other models. Next, we performed further comparisons between *ResNet-50*, *CNN+SVM*, and *CNN+kNN* (our proposed model). Table 4 varies k in the Top-k Accuracy measure between $1 \leq k \leq 15$, and reports the finding for *ResNet-50*, *CNN+SVM*, and *CNN+kNN*.

Table 4. Top-k Accuracy Comparison (%).

k-Value	ResNet-50	CNN+SVM	CNN+kNN
k = 1	71.7%	76.5%	80.5%
k = 2	74.5%	82.1%	86.1%
k = 3	82.0%	89.0%	94.0%
k = 4	83.0%	90.5%	95.5%
k = 5	85.5%	92.0%	96.1%
k = 6	86.5%	92.2%	96.5%
k = 7	88.0%	92.5%	97.0%
k = 8	89.0%	92.7%	97.5%
k = 9	89.8%	93.0%	97.8%
k = 10	90.5%	93.0%	97.9%
k = 11	90.8%	93.0%	98.0%
k = 12	91.0%	93.0%	98.0%
k = 13	91.0%	93.0%	98.0%
k = 14	91.0%	93.0%	98.0%
k = 15	91.0%	93.0%	98.0%

We noticed that, as k increased, the Top-k Accuracy result in Table 4 also increased, since the set of potential matched images expanded. Our proposed *CNN+kNN* model maintained a consistently higher Top-k Accuracy result across all values of k, followed by *CNN+SVM* and *ResNet-50*.

4.3. Efficiency

We would like to evaluate the performance of our proposed model in terms of runtime and compare it to similar models. Runtime is measured as the start from the moment

the user submits a query pill image to the moment the model returns matched images. Runtime is averaged over all the NLM dataset images. Table 5 reports the total execution time of each of the three models, *ResNet-50*, *CNN+SVM*, and *CNN+kNN*, in milliseconds. All the three models achieved nearly the same runtime.

Table 5. Runtime Comparison (ms).

ResNet-50 (Best Performing (Table 3)	CNN+SVM	CNN+kNN (Our Proposed Model)
1.25 ms	1.05 ms	1.02 ms

5. Discussion

The experimental results in Section 4 suggest that our proposed *CNN+kNN* model architecture outperformed the closely-related *ResNet-50* model and the *CNN+SVM* model. Although all three classification techniques (k-NN, SVM, and ResNet-50) performed well, the proposed *CNN+kNN* model was able to achieve the highest accuracy.

Table 2 compares the mAP values of *ResNet-50*, *CNN+SVM*, and *CNN+kNN* using the NLM dataset. With regard to the mAP values, applying the *CNN+kNN* successfully increased the prediction Precision from 80% to 91%. This finding implies that, if the *CNN+kNN* model predicts a target class (i.e., finds a pill type of a query pill image), then there is a 91% chance that the prediction will be correct.

In Figure 9, we compared between the three models to evaluate the "goodness" of the prediction model. Each line in the figure represents the model's Precision–Recall curve. A larger area under the curve implies better Precision and Recall. That is, the more the curve pushes to the top-right corner of the plot, the better the model is. Out of all the three models, *CNN+kNN* had the largest area, thus implying a higher prediction quality.

Figure 10 provides us with an idea of how well (or bad) the model understands the different features of a pill in order to make a decision about the pill type. Figure 10a suggests that the model was successfully able to differentiate between all distinct shapes. However, Figure 10b,c suggest that the model struggled in making a decision when the pill color was *Turquoise* and imprint was *SOMA* and *M6*, respectively. The low Precision values for these target classes (e.g., 57.1% for predicting *Turquoise*) may be due to the low number of training images with pills having these target classes.

Table 4 suggests that using a Convolutional Neural Network with the k-Nearest Neighbor classifier improves pill identification accuracy by about 10%, which is considered a significant improvement. For sensitive applications or diseases, we suggest considering Top-1 Accuracy. If Top-5 Accuracy is to be considered, albeit at 96% accuracy, we suggest consulting an expert for confirmation.

The runtime experiment, as reported in Table 5, suggests that using the k-NN classifier does not compromise the efficiency of the overall image retrieval system. This finding implies that a pill image retrieval system's accuracy can be improved without compromising its runtime.

Although our experiments suggest higher pill image retrieval success than comparable methods, we encountered some challenges pertaining to consumer image quality. Conflicting light conditions, placement of the pills, and the distance from the camera used in the consumer images negatively impacted the shape extraction process. Thus, the presence of high noise in images may incur high classification error if the image retrieval model is not equipped with adequate filters to account for such noise.

Lastly, we would like to mention that, for the performance evaluation of our proposed model, we used the NLM dataset [46]. This dataset was published by the National Institutes of Health for an open research competition, and it has since been widely used by various seminal exiting works in the area of pill identification from images. For the sake of performance comparison with existing studies in the same area (see Table 3, we adopted

the NLM dataset in our evaluations. That said, we plan on using more datasets in our future work that builds on this study to further validate our findings.

6. Conclusions

The impact of consuming the wrong medication can be lethal. This paper proposes a method for identifying pills from images. The proposed method studies the impact of combining widely-known classifiers, namely k-NN and SVM, with neural networks. The classifier is placed between the fully connected feature layer and the output layer to handle prediction. Experimental evaluation was conducted on a real-life dataset called the NLM dataset, and results were compared with those obtained from comparable models. We have examined three deep learning models for the classification of pill images; two are hybrid models (a combination of proposed CNN with SVM and k-NN classifiers), and the third is the ResNet-50 model. Results show that using the k-NN classifier in a Convolutional Neural Network architecture (our proposed model) increased pill identification accuracy by around 10% while maintaining almost the same runtime as in the compared methods (nearly 1 ms per execution).

For future work, the proposed method can be improved to account for some inherent drawbacks in consumer-grade pill images. For example, the model may not be able to accurately determine the shape of a pill if the pill image was taken under conflicting lighting conditions. One naïve solution to this problem could be constructing a 3D model of the query pill by having more than one image showing the pill from multiple angles.

Author Contributions: Conceptualization, E.A. and R.M.A.; Methodology, K.A.-H. and I.K.; Software, E.A.; Validation, K.A.-H., E.A. and I.K.; Formal analysis, K.A.-H., E.A. and I.K.; Investigation, E.A.; Resources, K.A.-H. and R.M.A.; Data curation, E.A.; Writing—original draft, E.A. and K.A.-H. and R.M.A.; Writing—review and editing, K.A.-H., I.K. and R.M.A.; Visualization, E.A. and R.M.A.; Supervision, K.A.-H. and R.M.A.; Project administration, K.A.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported in part by the DSO-RIT Dubai Research Fund (2022-23-1003) from the Rochester Institute of Technology—Dubai.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the pill images used in this study are publicly available through the National Library of Medicine [46].

Acknowledgments: The authors would like to thank the TDRA-ICT Fund for providing high-end computing machines for our experiments through the Digital Transformation Lab at Rochester Institute of Technology—Dubai (RIT-Dubai).

Conflicts of Interest: The authors have no competing interests to declare that are relevant to the content of this article.

Abbreviations

The following abbreviations are used in this manuscript:

CBIR	Content-Based Information Retrieval
WHO	World Health Organization
CNN	Convolutional Neural Network
ResNet	Residual Network
SVM	Support Vector Machine
NLM	National Library of Medicine
NIH	National Institutes of Health
k-NN	k-Nearest Neighbor

References

1. Crestani, F.; Mizzaro, S.; Scagnetto, I. *Mobile Information Retrieval*, 1st ed.; Springer International Publishing: Cham, Switzerland, 2017.
2. Celik, C.; Bilge, H.S. Content based image retrieval with sparse representations and local feature descriptors: A comparative study. *Pattern Recognit.* **2017**, *68*, 1–13. [CrossRef]
3. Madduri, A. Content based Image Retrieval System using Local Feature Extraction Techniques. *Int. J. Comput. Appl.* **2021**, *183*, 16–20. [CrossRef]
4. Dubey, S.R. A Decade Survey of Content Based Image Retrieval Using Deep Learning. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 2687–2704. [CrossRef]
5. Makary, M.A.; Daniel, M. Medical error—the third leading cause of death in the US. *BMJ* **2016**, *353*, i2139. [CrossRef] [PubMed]
6. World Health Assembly. *Patient Safety: Global Action on Patient Safety: Report by the Director-Genera*; World Health Assembly: Geneva, Switzerland, 2019; p. 8.
7. WHO. *Medication Without Harm: Real-Life Stories*; WHO: Geneva, Switzerland, 2017.
8. WHO. 10 Facts on Patient Safety. 2019. Available online: https://www.who.int/news-room/photo-story/photo-story-detail/10-facts-on-patient-safety (accessed on 31 July 2022).
9. Larios Delgado, N.; Usuyama, N.; Hall, A.K.; Hazen, R.J.; Ma, M.; Sahu, S.; Lundin, J. Fast and accurate medication identification. *NPJ Digit. Med.* **2019**, *2*, 10. [CrossRef]
10. Yu, J.; Chen, Z.; Kamata, S.i. Pill Recognition Using Imprint Information by Two-Step Sampling Distance Sets. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 3156–3161. [CrossRef]
11. RxList. Pill Identifier (Pill Finder Wizard). 2022. Available online: https://www.rxlist.com/pill-identification-tool/article.htm (accessed on 31 July 2022).
12. Healthline. Medication Safety: Pill Identification, Storage, and More. 2021. Available online: https://www.healthline.com/health/pill-identification (accessed on 31 July 2022).
13. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN Model-based Approach in Classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 986–996, Volume 2888. [CrossRef]
14. Altman, N.S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am. Stat.* **1992**, *46*, 175–185.
15. Maron, M.E.; Kuhns, J.L. On Relevance, Probabilistic Indexing and Information Retrieval. *J. ACM* **1960**, *7*, 216–244. [CrossRef]
16. Adel'son-Vel'skii, G.M.; Landis, E.M. An algorithm for organization of information. *Dokl. Akad. Nauk.* **1962**, *146*, 263–266.
17. Chang, S.K.; Liu, S.H. Picture indexing and abstraction techniques for pictorial databases. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *4*, 475–484. [CrossRef]
18. Foster, C.C. Information Retrieval: Information Storage and Retrieval Using AVL Trees. In Proceedings of the 1965 20th National Conference, Cleveland, OH, USA, 24–26 August 1965; Association for Computing Machinery: New York, NY, USA, 1965; pp. 192–205. [CrossRef]
19. Salton, G.; Lesk, M.E. The SMART Automatic Document Retrieval Systems-an Illustration. *Commun. ACM* **1965**, *8*, 391–398. [CrossRef]
20. Salton, G. *The SMART Retrieval System-Experiments in Automatic Document Processing*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1971.
21. Rabitti, F.; Stanchev, P. 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, New Orleans, LA, USA, 3–5 June 1987; Association for Computing Machinery: New York, NY, USA, 1987; pp. 284–295. [CrossRef]
22. Wang, X.Y.; Wu, J.F.; Yang, H.Y. Robust image retrieval based on color histogram of local feature regions. *Multimed. Tools Appl.* **2010**, *49*, 323–345. [CrossRef]
23. Lee, Y.B.; Park, U.; Jain, A.K.; Lee, S.W. Pill-ID: Matching and retrieval of drug pill images. *Pattern Recognit. Lett.* **2012**, *33*, 904–910. [CrossRef]
24. Maji, S.; Bose, S. CBIR using features derived by deep learning. *ACM/IMS Trans. Data Sci.* **2021**, *2*, 1–24. [CrossRef]
25. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
26. Razavian, A.S.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 512–519. [CrossRef]
27. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708. [CrossRef]
28. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
29. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]

30. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
32. Wang, Y.; Ribera, J.; Liu, C.; Yarlagadda, S.; Zhu, F. Pill Recognition Using Minimal Labeled Data. In Proceedings of the IEEE Third International Conference on Multimedia Big Data (BigMM), Laguna Hills, CA, USA, 19–21 April 2017; pp. 346–353. [CrossRef]
33. Ou, Y.Y.; Tsai, A.C.; Zhou, X.P.; Wang, J.F. Automatic drug pills detection based on enhanced feature pyramid network and convolution neural networks. *IET Comput. Vis.* **2020**, *14*, 9–17. [CrossRef]
34. Zeng, X.; Cao, K.; Zhang, M. 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, Niagara Falls, NY, USA, 19–23 June 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 56–67. [CrossRef]
35. Guo, P.; Stanley, R.; Cole, J.G.; Hagerty, J.; Stoecker, W. Color Feature-based Pillbox Image Color Recognition. In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications—Volume 4: VISAPP, (VISIGRAPP 2017), Porto, Portugal, 27 February–1 March 2017; pp. 188–194. [CrossRef]
36. Cordeiro, L.S.; Lima, J.S.; Rocha Ribeiro, A.I.; Bezerra, F.N.; Rebouças Filho, P.P.; Rocha Neto, A.R. Pill Image Classification using Machine Learning. In Proceedings of the 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), Salvador, Brazil, 15–18 October 2019; pp. 556–561. [CrossRef]
37. Suksawatchon, U.; Srikamdee, S.; Suksawatchon, J.; Werapan, W. Shape Recognition Using Unconstrained Pill Images Based on Deep Convolution Network. In Proceedings of the 2022 6th International Conference on Information Technology (InCIT), Nonthaburi, Thailand, 10–11 November 2022; pp. 309–313. [CrossRef]
38. Proma, T.P.; Hossan, M.Z.; Amin, M.A. *Medicine Recognition from Colors and Text*; Association for Computing Machinery: New York, NY, USA, 2019; ICGSP '19. [CrossRef]
39. Swastika, W.; Prilianti, K.; Stefanus, A.; Setiawan, H.; Arfianto, A.Z.; Santosa, A.W.B.; Rahmat, M.B.; Setiawan, E. Preliminary Study of Multi Convolution Neural Network-Based Model To Identify Pills Image Using Classification Rules. In Proceedings of the 2019 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, Indonesia, 28–29 August 2019; pp. 376–380. [CrossRef]
40. Kwon, H.J.; Kim, H.G.; Lee, S.H. Pill Detection Model for Medicine Inspection Based on Deep Learning. *Chemosensors* **2022**, *10*, 4. [CrossRef]
41. Holtkötter, J.; Amaral, R.; Almeida, R.; Jácome, C.; Cardoso, R.; Pereira, A.; Pereira, M.; Chon, K.H.; Fonseca, J.A. Development and Validation of a Digital Image Processing-Based Pill Detection Tool for an Oral Medication Self-Monitoring System. *Sensors* **2022**, *22*, 2958. [CrossRef]
42. Nguyen, A.D.; Pham, H.H.; Trung, H.T.; Nguyen, Q.V.H.; Truong, T.N.; Nguyen, P.L. High Accurate and Explainable Multi-Pill Detection Framework with Graph Neural Network-Assisted Multimodal Data Fusion. *arXiv* **2023**, arXiv:2303.09782.
43. Chang, W.J.; Chen, L.B.; Hsu, C.H.; Lin, C.P.; Yang, T.C. A Deep Learning-Based Intelligent Medicine Recognition System for Chronic Patients. *IEEE Access* **2019**, *7*, 44441–44458. [CrossRef]
44. Ting, H.W.; Chung, S.L.; Chen, C.F.; Chiu, H.Y.; Hsieh, Y.W. A drug identification model developed using deep learning technologies: Experience of a medical center in Taiwan. *BMC Health Serv. Res.* **2019**, *20*, 312. [CrossRef] [PubMed]
45. Nguyen, A.D.; Nguyen, T.D.; Pham, H.H.; Nguyen, T.H.; Nguyen, P.L. Image-based Contextual Pill Recognition with Medical Knowledge Graph Assistance. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Ho Chi Minh City, Vietnam, 28–30 November 2022.
46. National Library of Medicine. Pill Identification Challenge. 2016. Available online: https://www.nlm.nih.gov/databases/download/pill_image.html (accessed on 31 July 2022).
47. Stanford. CS231n: Convolutional Neural Networks for Visual Recognition. 2022. Available online: http://cs231n.stanford.edu/ (accessed on 31 July 2022).
48. Vapnik, V.N. *Statistical Learning Theory*; Wiley-Interscience: Hoboken, NJ, USA, 1998.
49. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *arXiv* **2016**, arXiv:1602.07360.
50. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
51. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]