Proceedings of the 2st IEEE International
Conference on Micro/Nano Sensors for AI, Healthcare and Robotics
Oct.31- Nov.2, 2019, Shenzhen, China

Deep Learning with Hyperspectral and Normal Camera Images for Automated Recognition of Orally-administered Drugs

Tejal Gala[1], Yanwen Xiong[2], Min Hubbard[2], Winn Hong[2], John D. Mai, Ph.D.[2]

1. Dept. of Bioengineering, *University of California, Berkeley*, USA. {tejalgala@berkeley.edu}
2. Alfred E. Mann Institute for Biomedical Engineering, *University of Southern California*, USA. {yanwenxi@usc.edu; min.hubbard@usc.edu; winnhong@usc.edu; johnmai@usc.edu}

***Abstract*—Patient compliance during drug trials and adherence to treatment regimens after a medical diagnosis are known pervasive problems in the practice of medicine. Any practical solution to this problem will require an easy method to identify and to verify the administration of orally-ingested drugs. Deep learning algorithms were applied to images of drugs in pill form. These images were taken using both a smart phone camera and using a hyperspectral imager based on a low-cost CMOS camera. As a proof-of-concept demonstration, 1,788 images were taken using a normal CMOS camera of four common pill types. The images of acetaminophen, acetylsalicylic acid and ibuprofen were taken using various backgrounds, image angles, and lighting conditions. The results show over 90% accuracy when the convolutional neural network is trained and tested using only normal camera images. The results improved to 100% when trained and tested using 4 baseline "datacubes" taken with a low-cost hyperspectral camera solution; however, due to matrix dimensional differences, a 1D CNN was used in this case, while a 2D CNN was used with the normal camera images. Each hyperspectral cube included information from effectively 31 wavebands. With more hyperspectral images to expand the drug training set, this approach would be promising for daily use to quickly identify similar pills in the clinical or home environment as well as in smart phone apps to remotely monitor patient compliance to a drug-based treatment regimen.**

***Keywords*—hyperspectral, deep learning, drug identification, patient compliance, low cost**

I. Introduction

Former US Surgeon General, C. Everett Coop is quoted as saying "Drugs don't work in patients who don't take them," and the economic and health care burden of this non adherence is supported by nationwide and individual hospital studies. For example, Rosen, et al., reported that the 30-day hospital readmission rates for patients with low or medium medication adherence was more than 2.5 times higher than for patients with high adherence [1]. It has been estimated that non adherence to a medication-based course of treatment could be responsible for up to 50% of treatment failures and up to 25% of all hospitalizations annually in the US. This translates into approximately $100 billion to $300 billion in additional health care costs annually [2].

There are existing programs for identifying pills based on camera images. WebMD allows users to enter the shape, color, and/or imprint on the pill and will identify it based on those three characteristics [3]. The NIH has a program called *Pillbox* and Drugs.com offers a program called *Pill Identification Wizard*, both of which similarly require users to enter identifying information about the pill rather than a picture [4, 5]. In terms of existing smartphone applications, Drugs.com has the Drugs.com Medication Guide, which allows users to "look up drug information, identify pills, check interactions, and set up personal medication records" [6]. More than the number of mobile applications for identifying pills is the number of mobile applications for reminding people to take their medications. Among the pill reminder mobile applications are *Round Health* by Circadian Design, *Mango Health* by Mango Health, and *Pill Reminder – All in One* by Sergio Licea [7, 8, 9]. These applications do not involve identifying pills using the phone camera. The iOS application *Drug ID App* by Rene Castaneda does attempt to recognize pills based on an image database sourced from Cerner and using only the phone camera, but after taking the picture, the user is prompted to optionally enter the imprint, shape, and color of the pill [10].

The goal of this project is to develop a user-friendly smartphone app that can be used by patients and clinicians to track and verify adherence to a medical treatment regime requiring the routine ingestion of pills. The technical premise of this research is to investigate whether the accuracy of image recognition results of drugs in a pill form can be enhanced using a hyperspectral imager built around a low-cost CMOS imager. The first step is to compare the accuracy of a proven convolutional network architecture, VGG-16, in identifying various pills types from various camera images taken at different lighting, background, and angles versus hyperspectral images under similar variables. Furthermore, we hope to extract wavelength information from this deep learning algorithm and correlate it with known characteristic chemical peaks.

II. Theory

A. Deep Learning Algorithm

VGG-16 was selected as the basic deep learning algorithm for these experiments. VGG-16 is a proven, highly accurate, image recognition algorithm based on a convolutional neural network (CNN) architecture and previously verified on ILSVRC classification and localization tasks [11, 12]. In order to run the

978-1-7281-6652-0/19/$31.00 © 2019 IEEE

Authorized licensed use limited to: Atmiya University. Downloaded on May 03,2025 at 09:35:46 UTC from IEEE Xplore. Restrictions apply.

CNN within the 16 GB of RAM available on our desktop computer workstations in a reasonable amount of time, transfer learning was used. This made the VGG-16 code extremely efficient. In fact, after transfer learning, it was much more efficient that the smallervggnet.py code (discussed in the next section). The transfer learning with VGG-16 took less than 1 minute to process the image training data while smallervggneet.py took approximately 90 minutes to train.

Smallervggnet.py was implemented in the mobile app because of CoreML restrictions on the transfer learning. The Smallervggnet.py use-case outlined in [16] was adapted to our experimental situation. The approximately 1,834 (46 from the internet and 1,788 using a smartphone camera) pill images were resized from their original resolution down to a 96 pixel x 96 pixel x 3 data cube (where 3 is the RGB component of the image) in order to work correctly with the input matrix into the CNN. Thus, each image was scaled down to a [96,96,3] matrix.

These smallervggnet.py results were then compared against results when transfer learning with VGG-16 is applied. The architecture of smallervggnet.py resembles that of VGG-16 but it has fewer layers. VGG16 has a maximum of 13 convolutional layers and 3 dense layers while smallervggnet.py has 5 convolutional layers and 2 dense layers. Summarizing, the pre-built and pre-trained (e.g. trained on a larger generic image dataset) VGG-16 was then trained on our pill dataset in an effort to transfer its knowledge to this smaller dataset. This is done by freezing the early CNN layers and only training the last few layers which are used to make the prediction about the type of pill, in this case. A total of only 7 layers were required for VGG16 with transfer learning. This is done under the assumption that VGG16 is extracting general features applicable to all images, e.g. edges, shapes, and gradients, from the early CNN layers and then identifies specific features from the later pill layers such as markings and colors. In general, transfer learn with VGG produced more accurate results than smallervggnet.py. These features can be easily expanded, in later practice, because the Prescriber's Digital Reference® contains information about the specific colored dyes, pill shapes, and markings for all FDA-approved drugs in the USA [15].

B. Hyperspectral data transform

Traditional hyperspectral imaging usually requires expensive specialized cameras that are basically imaging spectrometers. AMI-USC has developed a low-cost hyperspectral system with COTS hardware and software components that is used to acquire images and then to unmix the spectral components. The Light Source in Figure 1 is an array of 6 light emitting diodes (LEDs) with up to 6 different spectral bands. The sequential illumination of the target sample is coordinated and controlled by a personal computer in the block labeled as Controller. The Detector block is a low-cost, CMOS digital camera. The imager has a polarizer when used with visible wavelength imagers. The processor runs the phasor analysis software which is based on the HySP software originally developed and previously presented in [13] and [14].

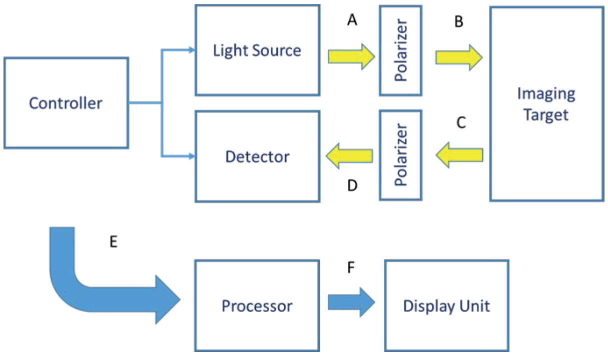


Figure 1. Block diagram illustrating the components of the low cost hyperspectral imager.

We used the algorithm presented by Cutrale et al. in [13] and [14], and available for academic use as HySp [17], to quickly analyze the hyperspectral data via the G-S plots of the Fourier coefficients of the normalized spectra, where

$$z(n) = G(n) + iS(n) \quad (1)$$

$$G(n) = \frac{\sum_{\lambda s}^{\lambda f} I(\lambda)\cos(n\omega\lambda)\Delta\lambda}{\sum_{\lambda s}^{\lambda f} I(\lambda)\Delta\lambda} \quad (2)$$

And

$$S(n) = \frac{\sum_{\lambda s}^{\lambda f} I(\lambda)\sin(n\omega\lambda)\Delta\lambda}{\sum_{\lambda s}^{\lambda f} I(\lambda)\Delta\lambda} \quad (3)$$

Where λs and λf are the starting and ending wavelengths of bands of interest, respectively; I is the intensity, ω= 2π/τs, where τs is the number of spectral channels (32 in our case) and n is the harmonic (usually chosen to be either n = 1 or 2, consistently).

A pseudo-inverse method, as illustrated in Figure 2, is used to reconstruct a hyperspectral cube from the digital images. In Stage 1, a CMOS camera is used to capture images of from the ColorChecker® standard (X-Rite Passport Model# MSCCP, USA). A transformation matrix **T** is constructed by a generalized pseudo-inverse method based on singular value decomposition (SVD) where

$$\mathbf{T} = \mathbf{R} \text{ x } \mathit{PINV}(\mathbf{D}) \quad (4)$$

$$\mathbf{T} = \mathbf{RD}^{+} \text{ (least-squares solution for } \mathbf{RD} - \mathbf{T}\text{)} \quad (5)$$

$$\mathbf{T} = \mathbf{RD}^{+} = \mathbf{R}(\mathbf{D}^{T}\mathbf{D})^{-1}\mathbf{D}^{T} \quad (6)$$

Where the matrix **R** contains spectral reflectance factors of the calibration samples, *PINV()* is the pseudo inverse function, and the matrix **D** are the corresponding camera signals of the calibration samples. Then, the predicted spectral reflectance factor **R** can be calculated using matrix multiplication for both the calibration (Stage 1) and verification (Stage 2) targets.

$$\mathbf{R} = \mathbf{T} \text{ x } \mathbf{D} \quad (4)$$

Authorized licensed use limited to: Atmiya University. Downloaded on May 03,2025 at 09:35:46 UTC from IEEE Xplore. Restrictions apply.

This method has the advantage that the camera spectral sensitivity does not need to be known a priori.

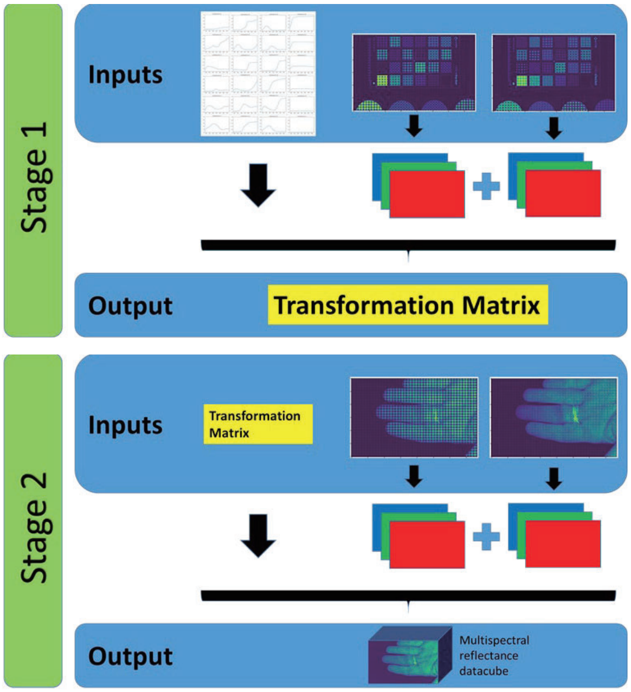
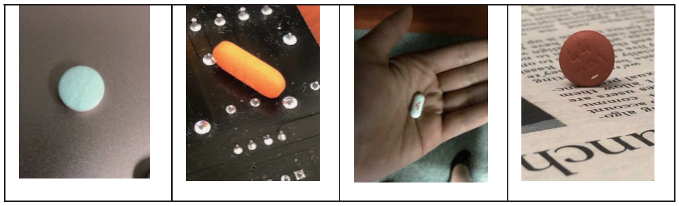


*Figure 2. Block diagrams illustrating the two stage "pseudo-inverse" method used to reconstruct a multispectral reflectance datacube from a series of camera images. In Stage 1, a color standard is imaged under a sequence of different lighting conditions in order to obtain their spectral reflectance factors, which is then used to solve for the transformation matrix **T**. In Stage 2, the transformation matrix **T** is used to recover the spectral information from the target human limb, under the same lighting sequence.*

III. Method

The normal CMOS images were taken using an iPhone X camera. This camera takes 12 megapixel images with a f/1.8 aperture and built-in optical image stabilization. As a simple proof-of-concept to quickly verify the accuracy of our approach, only four different pill types were initially trained and tested on the CNN. The pills are common over-the-counter headache and inflammation reducing medicines and are (1) Bayer® 350 mg aspirin (acetylsalicylic acid, NDC 0280-2000-10), (2) Tylenol® 500 mg (acetaminophen, NDC 50580-449-10), (3) Motrin® 200 mg (ibuprofen, NDC 50580-230-09) and (4) generic ibuprofen 200 mg (PhysiciansCare Model #90015). Approximately 500 images of each pill type were taken under various lighting conditions, angles, distance from the camera (in and out of focus), and backgrounds. Sample images are presented in Figure 3. Approximately 400 images of each pill type were used for training of the CNN and approximately 100 images of each pill type were reserved for testing.



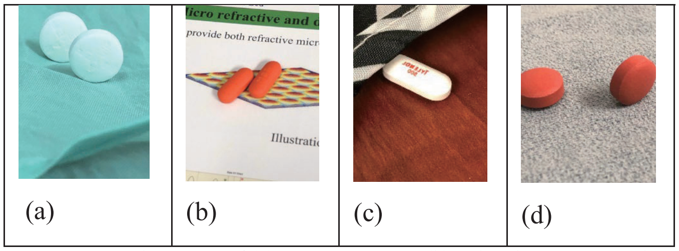


Figure 3. Sample images of (a) Bayer® aspirin, (b) Tylenol® acetaminophen, (c) Motrin® ibuprofen, and (d) generic ibuprofen used for training and testing of the 2D CNN.

The custom hyperspectral imager is built around an Allegra 174C (Imaging Solutions Group, Model LW-AL-IMX174C-USB3, NY, USA) camera with a 35mm lens (Navitar, NY, USA). This camera uses a CMOS imaging chip (Sony IMX174, Japan) capable of taking up to 150 frames per second at 10-bit resolution. Each pixel, on this 1/1.2-inch size imaging chip, is 5.86 microns yielding a 2.35-megapixel image. This camera is synchronized with a custom six LED illuminator (LED illumination peaks at 447nm, 530 nm, 627 nm, 590 nm, and a white light LED at a color temperature of 6500K) that is used with phasor analysis (e.g. HySp software, PhaseSpec, USA) to extract the 31 wavelength bands, as discussed previously in section II.B. Sample hyperspectral images of Motrin® and Tylenol® along with their phasor representation are presented in Figure 4 below.

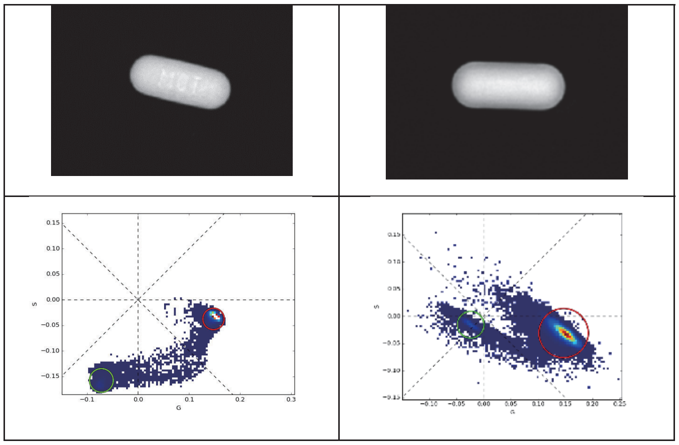


Figure 4. Sample hyperspectral (false color) images of Motrin® (left column) and Tylenol® (right column). The distinctly different G-S plots for each pill type are presented below the respective pills, although the false-color images look similar.

Since most of the hyperspectral images for each pill would look similar, irrespective of the pill orientation, background, or illumination; instead, we injected a Gaussian noise matrix in order to grow the data set to 1,000 “noisy” RGB images of each pill type. The Python function “numpy.random.normal” was used to generate the noisy array of pill images.

Authorized licensed use limited to: Atmiya University. Downloaded on May 03,2025 at 09:35:46 UTC from IEEE Xplore. Restrictions apply.

IV. Results

A. Smallervggnet.py and transfer learning with normal color images

The classification accuracy from Smallervggnet.py is presented in Figure 5. After 100 epochs, the training accuracy is approximately 90% when classifying the 2,000 images in the training set into the four different pill types. While the validation accuracy drops slightly to 85% when the trained Smallervggnet.py is tested against the remaining 400 pill images. Compare this to the transfer learning results presented in Figure 6.

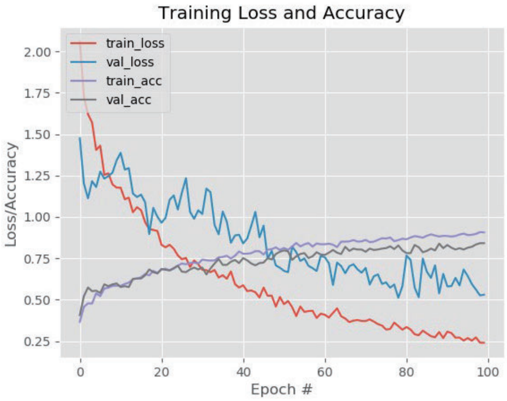


Figure 5. Classification accuracy using Smallervggnet and 1600 pill images for training and 400 pill images for validation.

When transfer learning is applied (with VGG-16), the results dramatically improve and the computing time decreased. The training accuracy increased to 100% while the validation accuracy improved to above 90%.

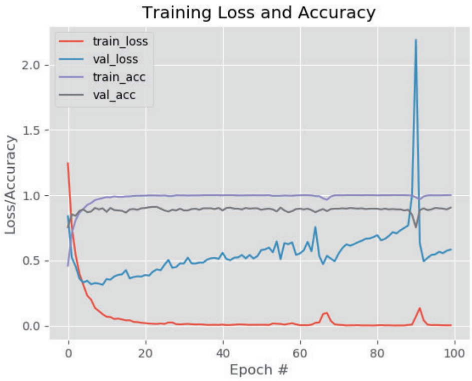


Figure 6. Classification accuracy using transfer learning with VGG-16 and 1600 pill images for training and 400 pill images for validation.

Sample results of the Smallervggnet.py implemented on an iOS smartphone and transfer learning with VGG-16 implemented on a desktop computer are presented in Figure 7.

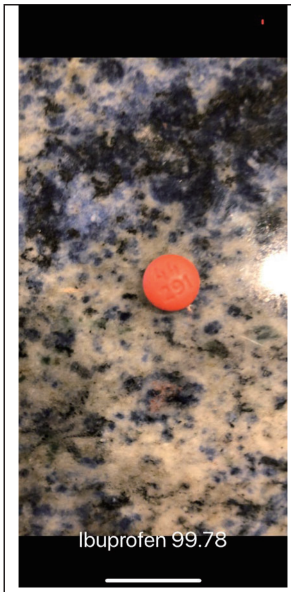
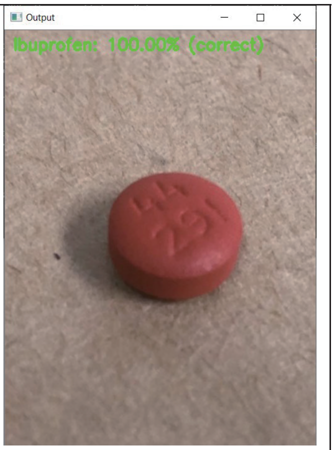


Figure 7. (a) The trained Smallervggnet.py was ported to an iOS smartphone and a sample test image with the predicted pill type and percentage certainty (99.78%) is captured in the image above.



Ibuprofen: 100.00% (correct)

Figure 7. (b) The trained CNN (VGG16 + transfer learning) is tested on an image above imported into a personal computer and the predicted pill type and percentage certainty (100%) is captured in the image above.

These CNN results with normal RGB images of pills are now used as a baseline in our comparison against the hyperspectral image processing results.

B. Hyperspectral Data and a 1D CNN

For hyperspectral image analysis, the "image" that is run through the CNN has only two dimensions. Smallervggnet.py contains a two dimensional convolutional neural network (2D CNN). 2D CNNs require their input images to have a third dimension. The image starts out as (600,960) pixels, is converted to an RGB image of (600,960,3), is resized to (60,96,3), and finally is converted to an image cube of (60,96,31). In order to compare the relative significance of each channel, 31 different models are created and trained. Each of these 31 models is trained on only one channel of the cube. Thus, each input to the model has size (60, 96) - only two dimensions. Thus, a 2D CNN cannot be used, as in the case of the normal RGB images. A custom program adapted from smallervggnet.py called Hyperspec.py contains a one dimensional convolutional neural network (1D CNN) and was used instead. 1D CNNs require their inputs to have two dimensions.

Authorized licensed use limited to: Atmiya University. Downloaded on May 03,2025 at 09:35:46 UTC from IEEE Xplore. Restrictions apply.

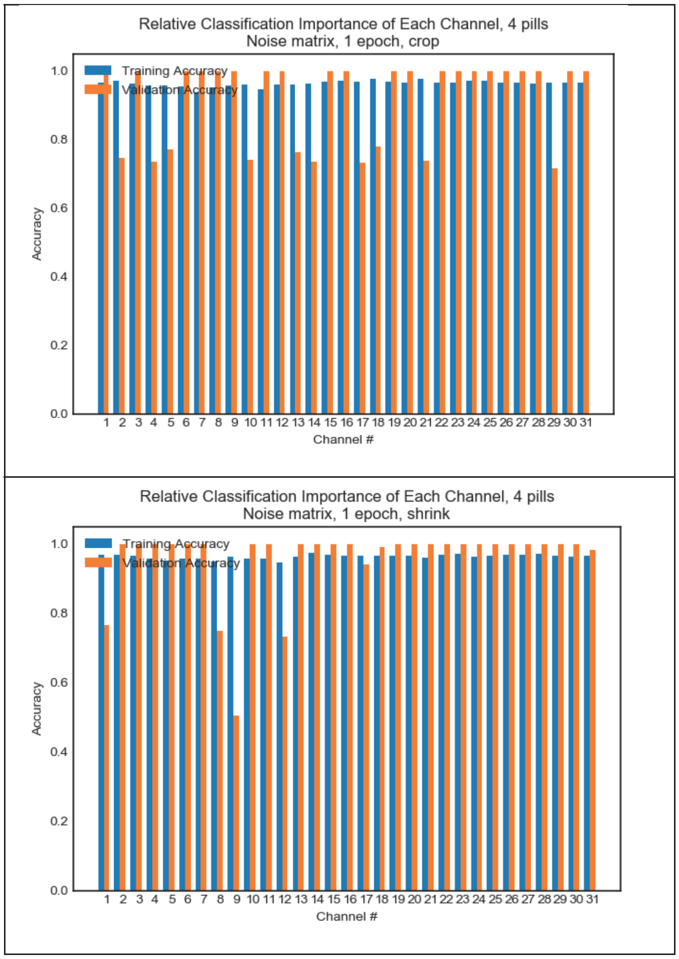


Figure 8. Comparison of effects of pre-processsing methods on the relative results from the 1D CNN. (top) Each image is automatically cropped to 225 x 300 pixels in the data set. (bottom) Each complete image is scaled to 225 x 300 pixels in the data set.

In Figure 8, each hyperspectral channel is approximately 10 nm wide. So channel 1 is 400 nm to 410 nm in bandwidth. Similarly, channel 10 is 500 nm to 510 nm. This result also gives us some insight into the relative importance of each wavelength band from 400 to 700 nm. This also indirectly yields information about the reflected chemical spectral peaks of the pill components with respect to how the CNN weights the importance of these peaks as a unique signature of the pill. Note that if we remove the wavelength information from channels 1, 8, 9, 12, and 31 from the automatically shrunken hypercube test case, we find that the resulting validation accuracy increases to almost 100%.

V. Conclusions

A preliminary comparison was made between CNN apps trained to recognize normal camera images and hyperspectral images from a custom, low-cost CMOS camera setup. Using transfer learning and a popular 2D CNN, VGG-16, we achieved accuracies above 90% when identifying four different types of over-the-counter analgesic medications. Due to matrix size differences, a 1D CNN was used to identify hyperspectral images. We achieved 100% accurate identification of the same four different types of pills. The next steps are (1) to expand the training set of different pill types using the process outlined in this paper and (2) retrain the 1D CNN only on HySP output only, as compared to the complete 31 waveband hyperspectral hypercube. For example, a database that only includes the top 200 most common pills would cover more than 1 billion drug prescriptions in the US alone. The 1D CNN based on HySP output should further reduce the training time required while possibly maintaining the high accuracy rate achieved with the limited hyperspectral data presented in this paper.

Acknowledgment

This work is wholly funded by the Alfred E. Mann Institute for Biomedical Engineering at the University of Southern California (AMI-USC) as part of the AMI Summer Internship Program.

References

[1] O.Z. Rosen, et al. “Medication adherence as a predictor of 30-day hospital readmissions.” *Patient preference and adherence* vol. 11 801-810. 20 Apr. 2017, doi:10.2147/PPA.S125672

[2] J. Kim, et al. “Medication Adherence: The elephant in the room,” *US Pharm.* 2018;43(1)30-34.

[3] https://www.webmd.com/pill-identification/default.htm

[4] https://pillbox.nlm.nih.gov

[5] https://www.drugs.com/pill_identification.html

[6] https://apps.apple.com/us/app/drugs-com-medication-guide/id599471042

[7] https://apps.apple.com/us/app/round-health/id1059591124

[8] https://www.mangohealth.com

[9] https://apps.apple.com/us/app/pill-reminder-all-in-one/id816347839

[10] https://apps.apple.com/us/app/drug-id-app/id1372681668

[11] https://arxiv.org/pdf/1409.1556v6.pdf

[12] https://keras.io/applications/#vgg16

[13] F. Cutrale, V. Trivedi, L.A. Trinh, C.L. Chiu, J.M. Choi, M.S. Artiga, S.E. Fraser, “Hyperspectral phasor analysis enables multiplexed 5D in vivo imaging,” *Nature Methods* 14, 149-152 (2017).

[14] W. Shi, E.S. Koo, L.A. Trinh, S.E. Fraser, F. Cutrale, “Enhancing visualization of hyperspectral data with Phasor-Maps,” *Molecular Biology of the Cell* 28, (2017).

[15] https://www.pdr.net/

[16] https://www.pyimagesearch.com/2018/04/16/keras-and-convolutional-neural-networks-cnns/

[17] http://bioimaging.usc.edu/software.html#HySP

Authorized licensed use limited to: Atmiya University. Downloaded on May 03,2025 at 09:35:46 UTC from IEEE Xplore. Restrictions apply.