

Original Paper

Automated Identification of Hookahs (Waterpipes) on Instagram: An Application in Feature Extraction Using Convolutional Neural Network and Support Vector Machine Classification

Youshan Zhang¹, MS; Jon-Patrick Allem², MA, PhD; Jennifer Beth Unger², PhD; Tess Boley Cruz², PhD

¹Department of Computer Science, Lehigh University, Bethlehem, PA, United States

²Keck School of Medicine of USC, Los Angeles, CA, United States

Corresponding Author:

Jon-Patrick Allem, MA, PhD

Keck School of Medicine of USC

2001 N Soto Street, 3rd Floor Mail

Los Angeles, CA, 90032

United States

Phone: 1 8586030812

Email: allem@usc.edu

Abstract

Background: Instagram, with millions of posts per day, can be used to inform public health surveillance targets and policies. However, current research relying on image-based data often relies on hand coding of images, which is time-consuming and costly, ultimately limiting the scope of the study. Current best practices in automated image classification (eg, support vector machine (SVM), backpropagation neural network, and artificial neural network) are limited in their capacity to accurately distinguish between objects within images.

Objective: This study aimed to demonstrate how a convolutional neural network (CNN) can be used to extract unique features within an image and how SVM can then be used to classify the image.

Methods: Images of waterpipes or hookah (an emerging tobacco product possessing similar harms to that of cigarettes) were collected from Instagram and used in the analyses (N=840). A CNN was used to extract unique features from images identified to contain waterpipes. An SVM classifier was built to distinguish between images with and without waterpipes. Methods for image classification were then compared to show how a CNN+SVM classifier could improve accuracy.

Results: As the number of validated training images increased, the total number of extracted features increased. In addition, as the number of features learned by the SVM classifier increased, the average level of accuracy increased. Overall, 99.5% (418/420) of images classified were correctly identified as either hookah or nonhookah images. This level of accuracy was an improvement over earlier methods that used SVM, CNN, or bag-of-features alone.

Conclusions: A CNN extracts more features of images, allowing an SVM classifier to be better informed, resulting in higher accuracy compared with methods that extract fewer features. Future research can use this method to grow the scope of image-based studies. The methods presented here might help detect increases in the popularity of certain tobacco products over time on social media. By taking images of waterpipes from Instagram, we place our methods in a context that can be utilized to inform health researchers analyzing social media to understand user experience with emerging tobacco products and inform public health surveillance targets and policies.

(*J Med Internet Res* 2018;20(11):e10513) doi: [10.2196/10513](https://doi.org/10.2196/10513)

KEYWORDS

convolutional neural network; feature extraction; image classification; Instagram; social media; support vector machine

Introduction

Instagram, with millions of posts per day [1], can be used to inform public health surveillance targets and policies. However,

this research relying on image-based data often relies on hand coding of images [2,3], which ultimately limits the scope of the study. Images from social media may be more useful than findings from text-based platforms alone (eg, Twitter and

Reddit) when attempting to understand health behaviors, for example, user experiences with emerging tobacco products [4]. While automated image classification is useful for large-scale image classification (eg, processing and assigning labels to millions of images), current best practices in automated image classification are limited in their capacity to accurately distinguish between objects within images [5-7]. Automated image classification has been used in supervised, unsupervised, and hybrid approaches in classifying data [8-10]. Compared with unsupervised methods, supervised methods can be divided into stages of training and testing. The training stage consists of training a classifier by images and its labels, for example, describing image content, such as a person, dog, elephant, etc. The testing stage predicts the labels of the test images (in a new set of images) by a trained classifier.

Prior research has focused on ways to overcome the methodological challenges of automated image classification such as low accuracy. For example, Perronnin et al improved the Fisher Kernel approach to extend the bag-of-visual-words, also called bag-of-features (BOF), for large-scale image classification using internet images from ImageNet and Flickr, which increased precision from 47.9% to 58.3% but did not improve accuracy [5]. Verma et al used the backpropagation neural network approach to classify large images with good accuracy (97.02%), but this approach could not identify multiple categories of an image [6]. To reduce the time and spatial complexity of images, Simonyan et al proposed 2 visualization techniques using deep convolutional networks (ConvNets) to classify artificial images [7]. They combined understandable visualizations of ConvNets, maximizing the scores of images within different classes with gradient-based ConvNets visualization generating the saliency map (also called features map, which can represent the influence of pixels in image on image classification results) of every image (corresponding to one class) to use a deconvolution (also called transpose of convolution, which performs upsampling tasks instead of downsample in convolutional layer) network to segment objects in the images [7].

These earlier approaches have moved automated image classification forward; however, there are still a number of significant limitations to overcome [11-13]. For example, the large number of images that need to be extracted to train a model requires great computational power. In addition, the BOF method cannot localize the objects within an image and cannot use visual word positions (eg, if a cup was in an image, BOF could not find its position) [14,15]. Support vector machine (SVM) have a limitation in showing the transparency of results, as the final model is difficult to visualize. Moreover, it is a challenge to choose a suitable kernel in kernel SVM [16-18]. A convolutional neural network (CNN), on the other hand, can improve the generalization of the algorithm and can solve nonlinear problems. While a CNN has high accuracy, to get

better results, the parameters should be fine-tuned (eg, input image size, patch size, and the number of convolutional layers), and network performance is hard to optimize [19,20].

The purpose of this study is to determine whether combining CNN and SVM can achieve higher accuracy in image classification compared with CNN or SVM alone. To this end, data from Instagram containing images of waterpipes, also known as hookah (an emerging tobacco product possessing similar harms to that of cigarettes), were examined. By taking data from Instagram, we place our methods in a context that can be utilized to later inform researchers in the health domain who wish to analyze social media to understand user experience with emerging tobacco products and inform public health surveillance targets and public policies [2-4,21-25].

Methods

Data Acquisition

Data used in this study comprised posts on Instagram between February 19, 2016 and May 19, 2016, in the United States that included the hashtag #hookah. A total of 820 images was used in this study. The ground truth was manually labeled (hookah and nonhookah images). To balance the data and classes, the training images included 420 images (210 hookah and 210 nonhookah images), and test images also included 420 images (210 hookah and 210 nonhookah images). Further details on data collection are described elsewhere [24]. MATLAB was used to classify images into 2 categories: images containing a waterpipe (hookah) and those not containing.

Convolutional Neural Network

Image features comprising 25 layers were extracted using AlexNet [26-28] (a well-trained CNN software). Figure 1 shows the architecture of AlexNet. Among these 25 layers, there are input and output layers, 7 rectified linear units (ReLU) layers, 2 normalization layers, 3 pooling layers, 2 dropout layers (drop), 1 softmax layer (prob), and 8 learnable weights layers, which contain 5 convolutional layers (conv) and 3 fully connected layers (fc) [26]. The input layer comprised 227×227-pixel images. The ReLU layer reduces the number of epochs to achieve the training error rate higher than traditional tanh units. The normalization layer (norm) increases the generalization and reduces the error rate. The pooling layers summarize the outputs of adjacent pooling units [29]. The dropout layer efficiently decreases the test errors [30], and both dropout layer and the softmax layer reduce the overfitting phenomenon, while the output layer is the categories of images. To extract the features, we fine-tuned the network by removing the last 2 layers of the original 25 layers, as all layers are not suitable for extracting the features. As the layers at the beginning of the network can only detect the edges of the images, we used the results of the fully connected layers to extract features.

Figure 1. The architecture of AlexNet, which comprised 25 layers.

Support Vector Machine

SVM, a supervised learning model with algorithms that analyze data for classification, has been used to predict the categories of objects in images [18,31]. Our proposed method goes beyond earlier research as the input (feature vectors) was based on the results of the CNN, which can boost accuracy and save time. AlexNet was used to extract features, and those features were used to then train the SVM classifier, requiring only minutes to train all images, thereby saving time [11]. Once the SVM classifier was trained using the feature vectors, the categories of images were predicted.

Analytical Approach

First, we classified images into 2 categories—hookah and nonhookah images—and labeled accordingly. Figure 2 shows the classification scheme, for example, the input image dimension is $227 \times 227 \times 3$ pixels, and the output of the CNN is the $4096 \times 1 \times 1$ feature maps, which are used to train the SVM classifier; then, the classifier is used to predict the categories (hookah vs nonhookah) of test images. The hookah images contain a waterpipe, and the nonhookah images do not contain a waterpipe (Figure 3). Next, we divided image sets into training and test images; the training images were used to extract and learn the features ($n=420$, randomly selected), while the test images were used to calculate the accuracy of the method ($n=420$, randomly selected). To extract features of the images,

the dimension of the input images was made uniform, for example, the image size was 227×227 , as the image dimensions of 227×227 are the default of AlexNet. If an image is larger or smaller, we resized the dimensions of the input image to 227×227 . We loaded the pretrained CNN by utilizing AlexNet [26], which has been trained by >1 million images. As discussed above, AlexNet was fine-tuned in our method, for example, we removed the last 2 layers of the AlexNet and used the data of the final fully connected layer. Based on the data of the last fully connected layer, we computed the features of the training and test images based on the CNN. Then, the class labels were extracted from the training and test image sets.

To optimize the SVM classifier, we automatically optimized hyperparameters (such as learning rate, the number of layers in the CNN, and mini-batch size) of the waterpipe features vector, and based on the optimized results, we arrived at an optimized SVM classifier (Figure 2) [18,32,33]. The input images dimension is $227 \times 227 \times 3$. The output of the CNN was $4096 \times 1 \times 1$ features maps of 2 image classes. These features were trained by the SVM classifier, and the trained classifier was later used to predict the categories of test images. We then assessed the performance of the SVM classifier by using the test images and increased the number of images to improve accuracy (the number of images increased from 42 to 420; Figures 3 and 4). Features of the waterpipes in the yellow box were extracted to train the SVM classifier. Based on the trained classifier, we predicted the classes of new images.

Figure 2. The scheme of our method. SVM: support vector machine; CNN: convolutional neural network.

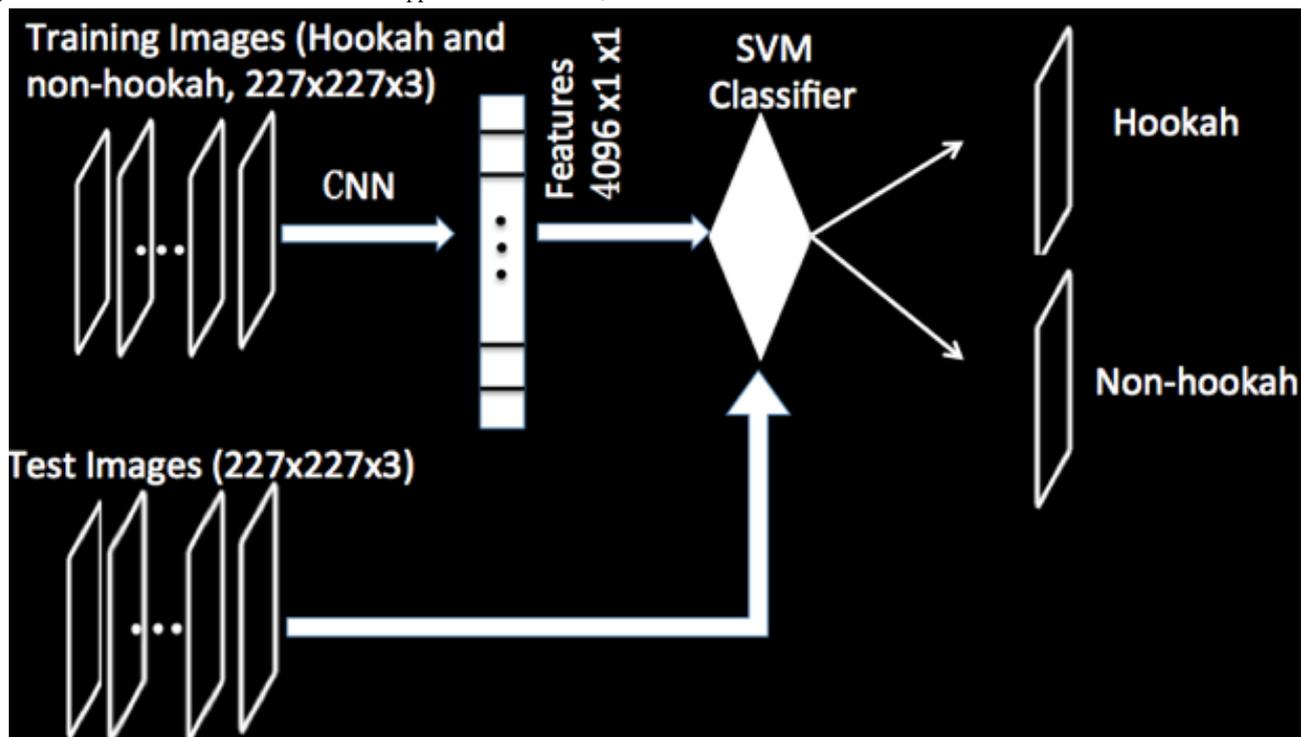


Figure 3. Examples of images with waterpipes (left) and without waterpipes (right).

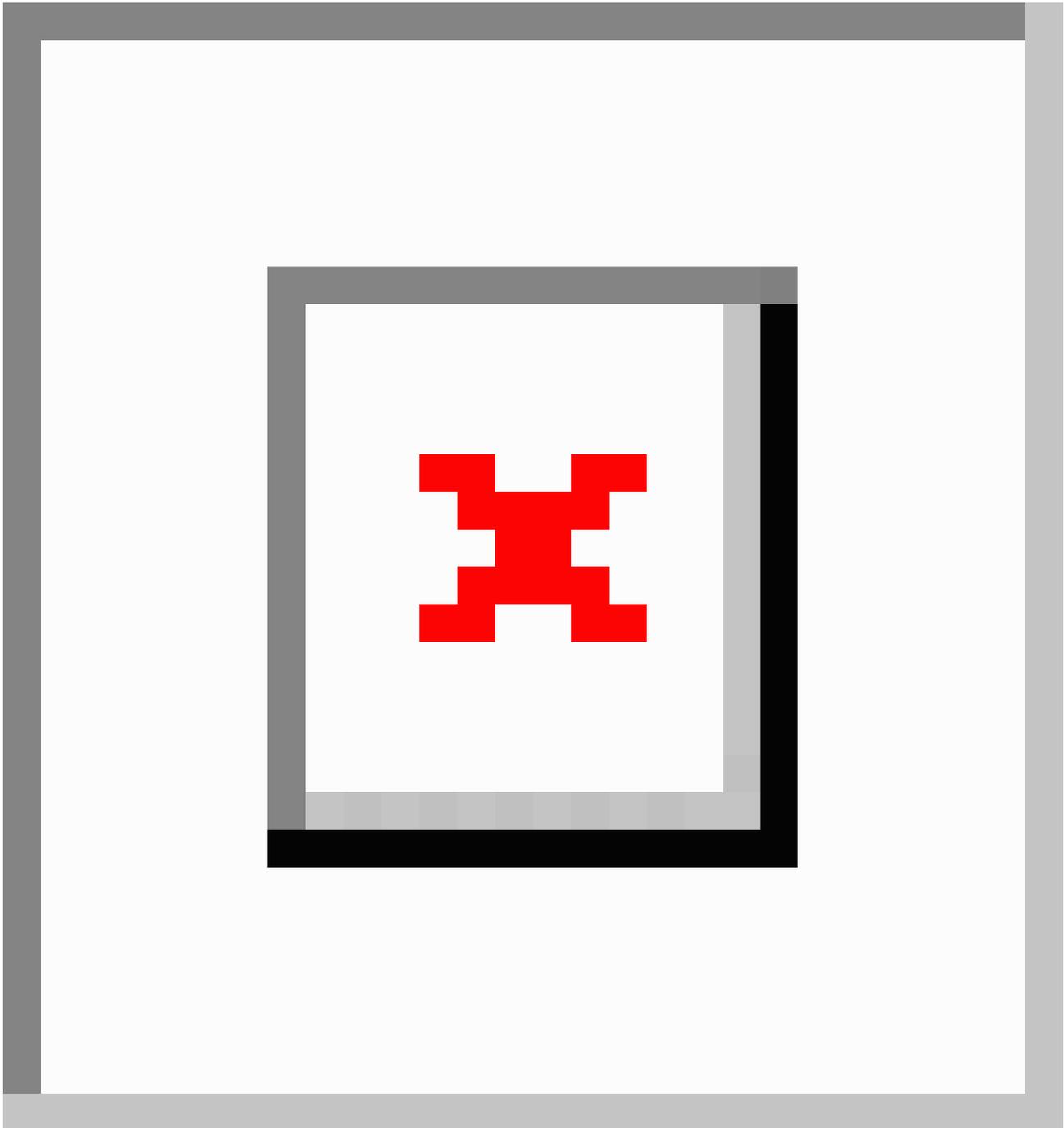


Figure 4. The localization of waterpipes in one image.



Results

Feature Extraction

Results demonstrated that hookah features could be extracted by the CNN, with image categories classified by the SVM, maintaining a high level of accuracy: highest, 99.5% (418/420). Figure 5 shows the features that were extracted from the first convolutional layer; this layer can only detect the edges and blobs, while more features were extracted from the remaining convolutional layers. The original hookah image is on the left. The feature images (right) contains a montage of 96 images, which can reflect the processing of extracting features. Figure 6 shows the feature vectors of the 420 training images, with range -20 to 20; the majority of feature vectors are located

between -10 and 10. The x-axis is the image features vector with 4096 total feature vectors. The y-axis is the range of the features with the range between -20 and 20. Figure 7 presents the histogram of the feature vectors. The maximum number of features was between -2 and 2; this interval reflected the most important features of the hookah images. Figure 8 shows the relationship between the function evaluations and the minimum objective. When the function evaluation was 25, the error between the minimum objective and the estimated minimum objective was the highest. Function evaluations demonstrated how many times to evaluate the optimized output. The minimum objective was the minimum observed value of the objective function; it is the smallest overall observation point if there are coupled constraints or evaluation errors.

Figure 5. The extracted features of the first layer using the convolutional neural network.

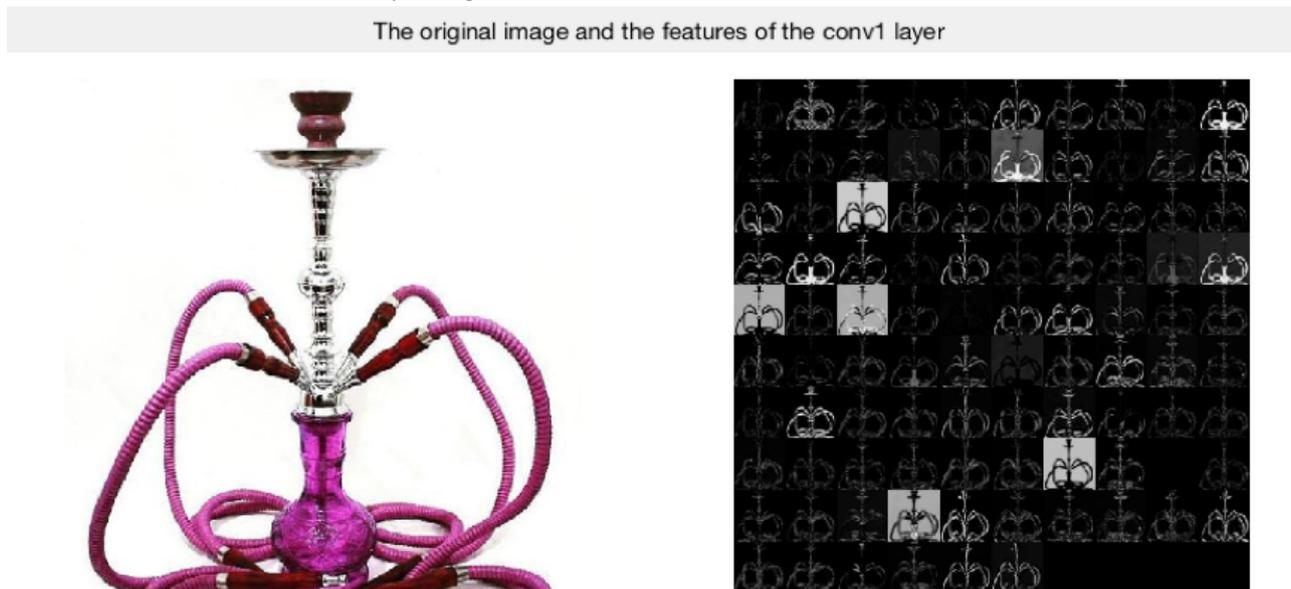


Figure 6. The features of the total image sets (420 images).



Figure 7. The histogram of the features. The interval of -2 and 2 contains the maximum number of features.

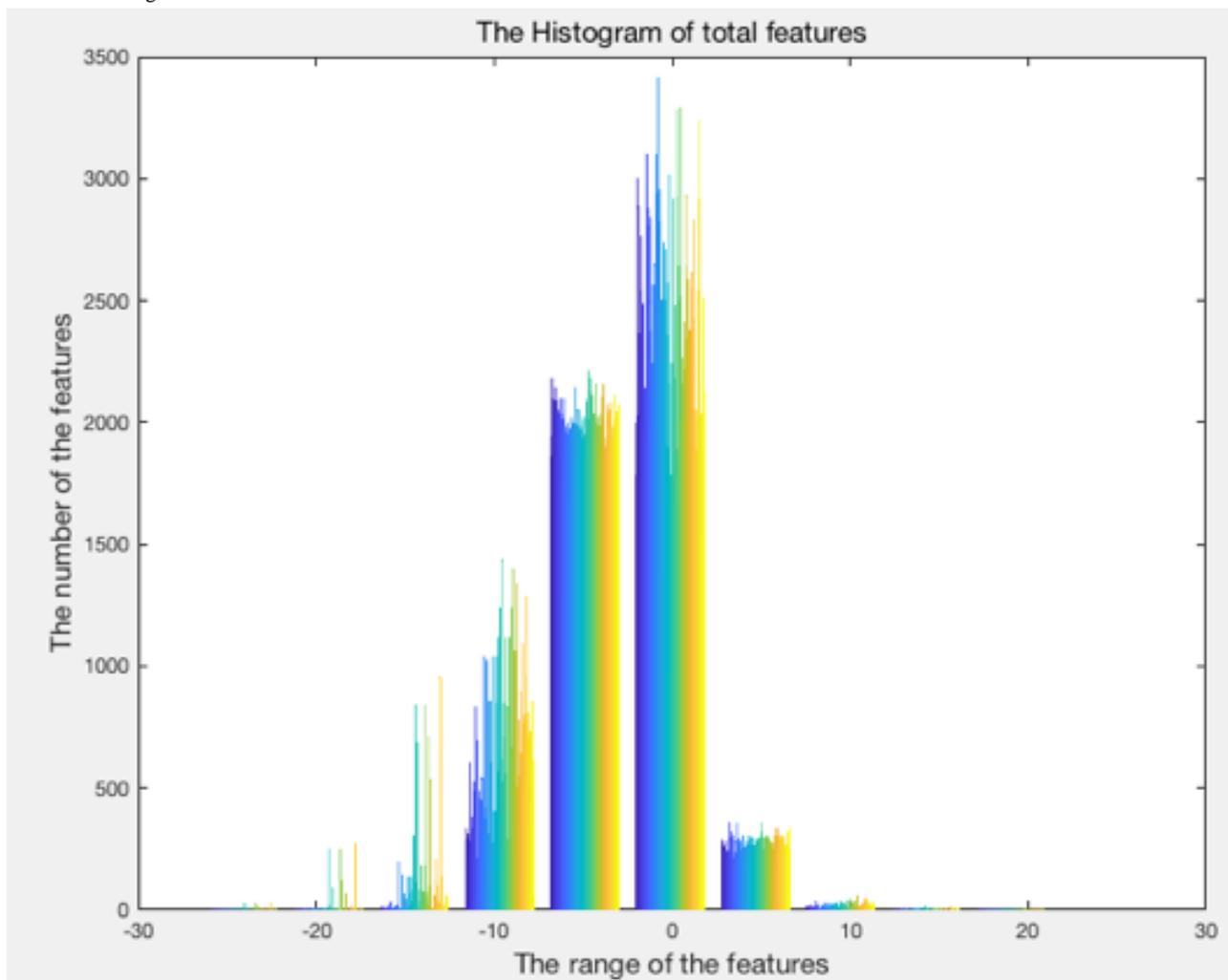
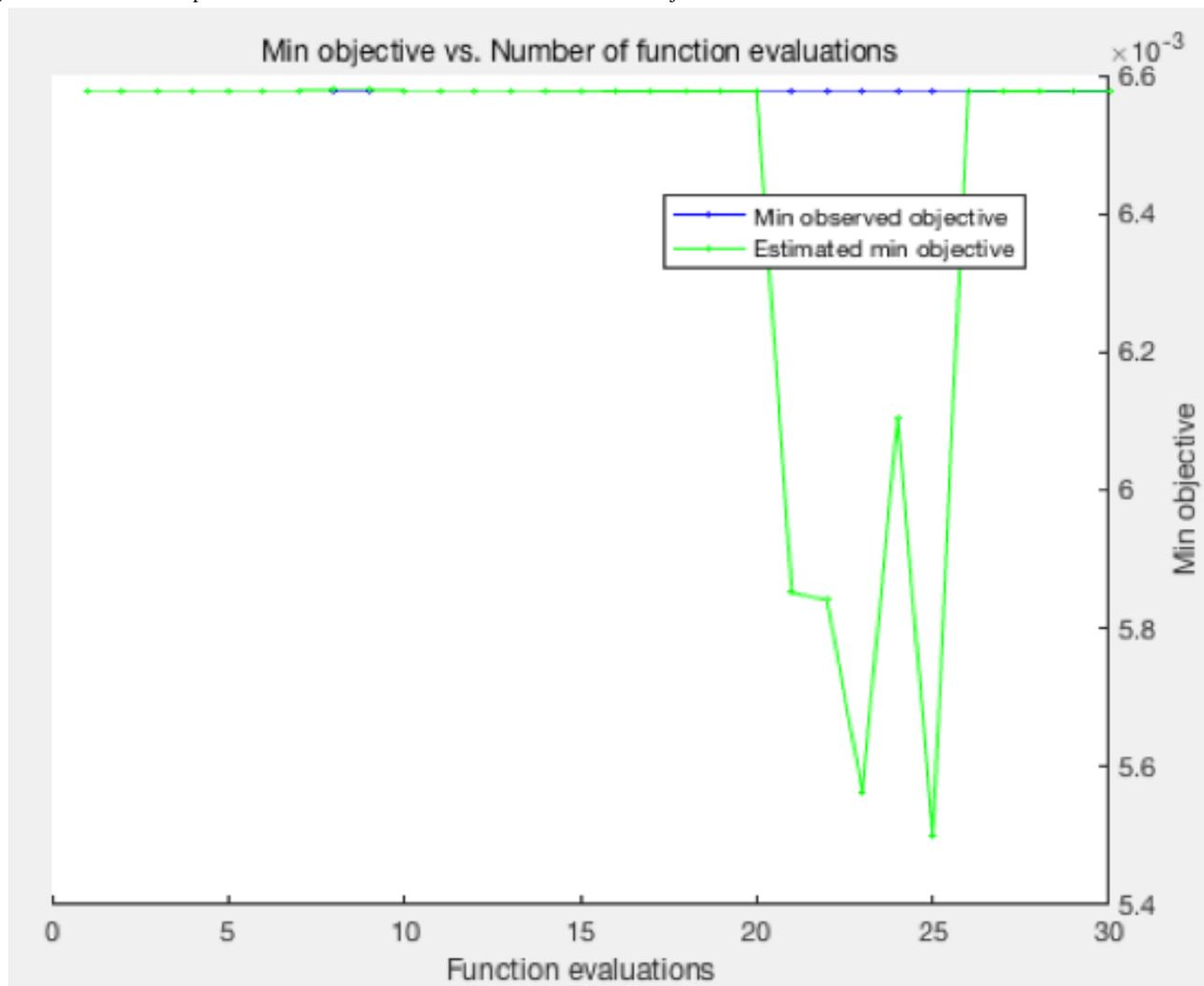


Figure 8. The relationship between the function evaluations and the minimum objective.

The estimation of minimum objective functions can show the difference between the estimated (optimized) minimum objective and real minimum objective. The minimum objective and the estimated minimum objective are similar; however, there are differences across certain function evaluations. The maximum proportion of error is $<.01$, which is acceptable [19]. Based on the optimized SVM classifier, we evaluated the performance of our method by the test images.

Test Image Classification Results

Figure 9 presents the learning curve showing the relationship between the percentage of validated images (eg, the training images, excluding the test images) and the average level of accuracy of the method. From the chart, with increase in the percentage of validated images, some of the accuracies boosted significantly. For example, from 80% (336/420), the accuracies increase faster than previous percentages, demonstrating that more training images are beneficial in predicting results. As the number of validated training images increased, the total number of extracted features increased: For 1 image, we can extract 4096 features; therefore, with the number of the validated training images (n) increased, the total number of extracted features can increase into $n \times 4096$. In addition, as the number of features learned by the SVM classifier increased, the average

level of accuracy increased. The number of validated images was equal to the percentage \times the number of the training set; for example, if 10%, then the validated images = $10\% \times 420 = 42$.

Overall, 99.5% (418/420) of images classified were correctly identified as either hookah or nonhookah images (Figure 10). The first 2 green squares show the number of the test images and the percentage of the correct image classifications. For example, there were 208 images correctly classified as hookah, and this number accounted for 49.5% (208/420) of all test images. Similarly, 210 images were properly classified as nonhookah, and this accounted for 50% (210/420) of all test images. In the first row, all nonhookah images were correctly classified as such. In the second row, there were 2 hookah images incorrectly classified as nonhookah images, representing 0.5% (2/420) of all data. In the first row, 100% (208/208) of hookah images were correctly classified. In the second row, 99.1% (210/212) were correctly classified as hookah images. In the first column, 99% (208/210) were correctly classified as hookah images, and 0.1% (2/210) were correctly classified as nonhookah images. In the second column, out of 210 nonhookah images, 100% (210/210) were correctly classified as nonhookah images, and all images were correctly classified as hookah images.

Figure 9. The learning curve showing the line graph of the accuracy of the classifier with a different number of validated images.

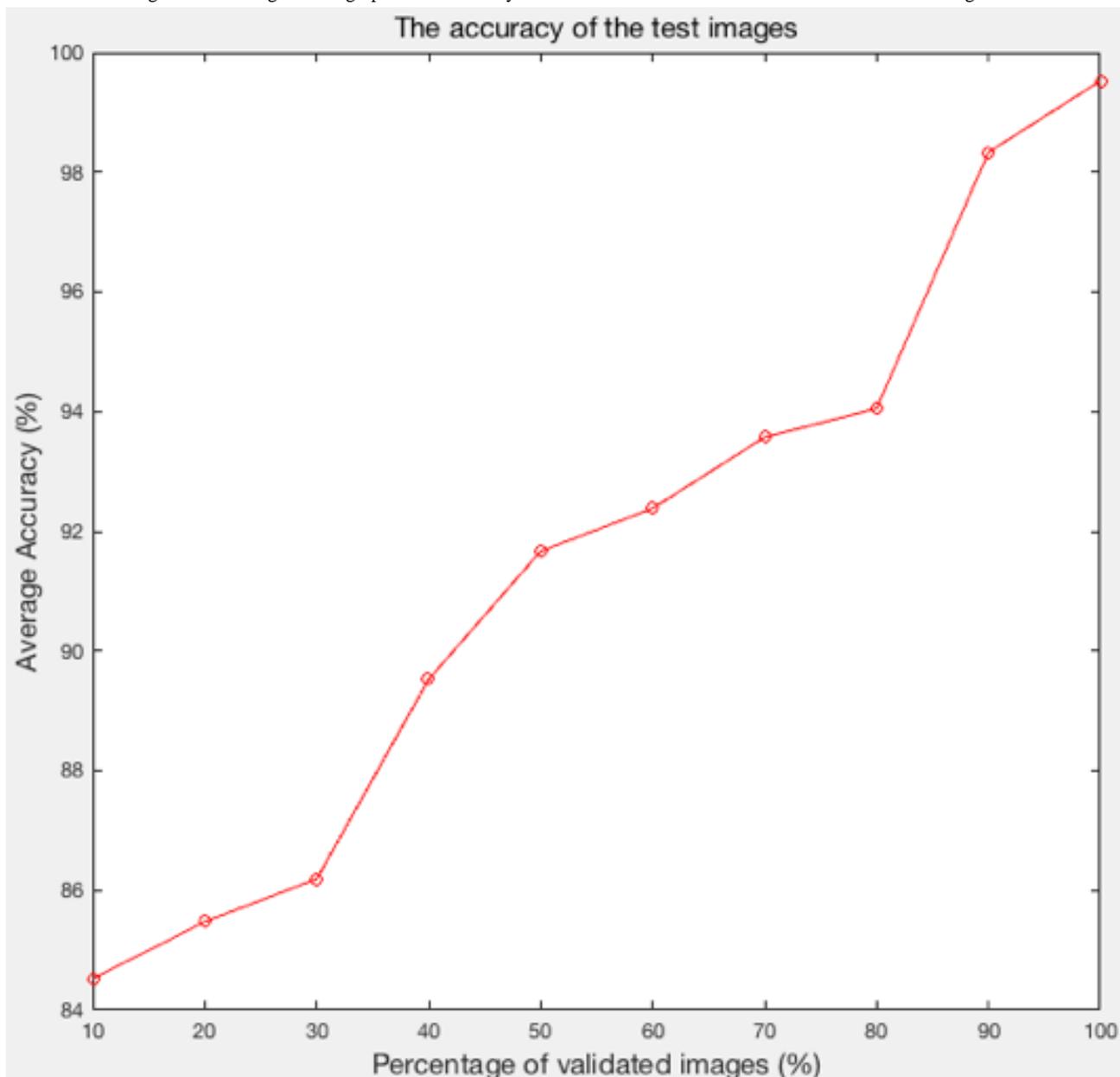
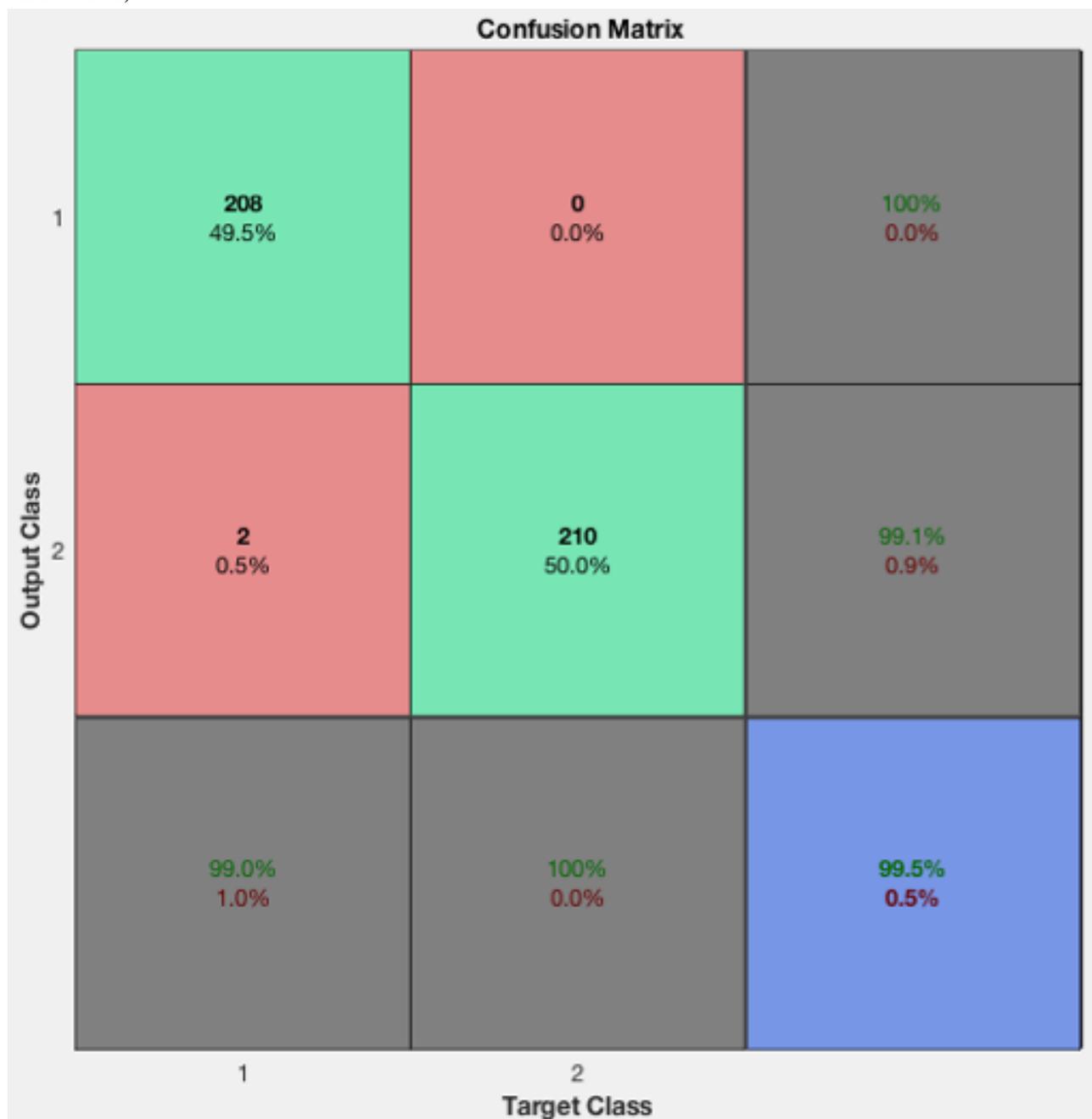


Figure 10. The confusion matrix of the test images (columns 1 and 2 are the hookah and nonhookah categories, respectively, column 3 is the accuracy of classified results).

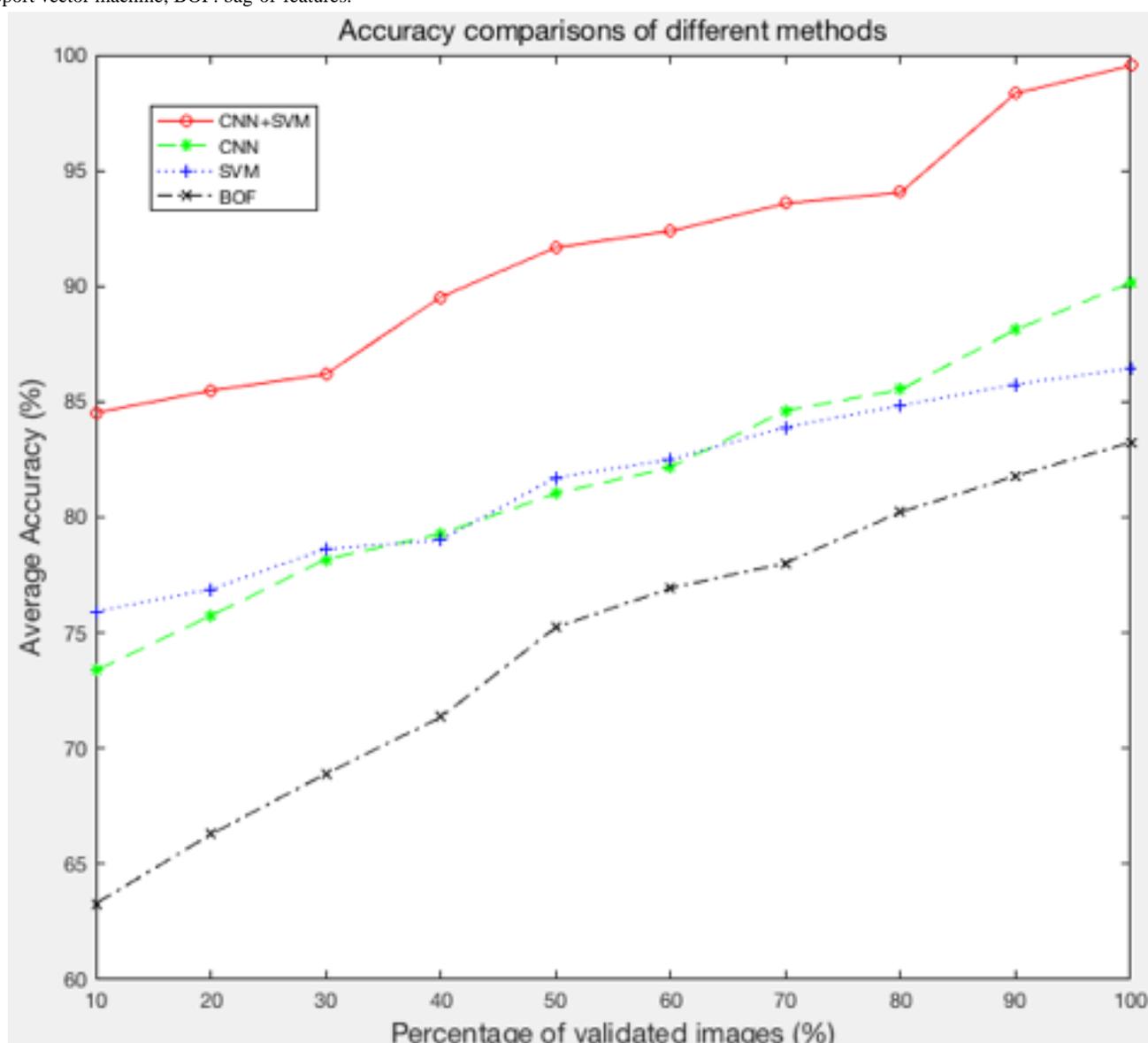


Comparison With Other Methods

We compared our method with CNN, SVM, and BOF [12,34]. For SVM and BOF, the input was the original image (raw pixel

values). Figure 11 shows how the accuracy of various models can be improved as a function of the size of the training data. Our method (CNN+SVM) had the highest accuracy, 99.5% (418/420), compared with other models (CNN, SVM, and BOF).

Figure 11. The prediction accuracy of different methods with different percentages of validated images. CNN: convolutional neural network; SVM: support vector machine; BOF: bag-of-features.



Discussion

Principal Findings

This study showed that the use of a CNN to extract features and SVM to classify images results in higher accuracy in automated image classification compared with using CNN or SVM alone. One crucial advantage of our pipelined approach is that we extracted sufficient features (4096 features from each image representing the details of each image) from a pretrained CNN model (AlexNet), taking advantage of SVM to train the features, saving time. Compared with earlier work using CNN, SVM, and BOF, our method improves accuracy when the number of training images is increased with accuracy reaching 99.5% (418/420), illustrating that our method is suitable for distinct images-like waterpipes.

The methods presented here could help detect increases in the popularity of certain tobacco products over time on social media. By identifying waterpipes in images from Instagram, we can identify Instagram users who may need tobacco-related

education to curb hookah use. Instagram may be used to bolster the reach and delivery of health information that communicates the risk of hookah use [35-38]. Earlier research used Instagram images to capture and describe the context in which individuals use and are marketed tobacco products [3,24,25]. For example, the analysis of Instagram data on electronic cigarettes demonstrated that a majority of images were either individuals showing their favorite combinations of products (eg, type of electronic cigarette device and flavored juice) or people performing tricks with the products (eg, blowing a large aerosol cloud in competition with others) [25], demonstrating how and why people use this tobacco product. Previous analyses of hookah-related posts to social media websites provide information about hookah-related contexts, including the importance of stylized waterpipes, use of hookah in social settings, copromotion with alcohol [24], and primarily positive user experiences [39-41].

Earlier studies using image-based data provided timely information from a novel data source; however, their methods

relied on hand coding of images—a process requiring time, expertise, and sample sizes small enough to reasonably code by hand, ultimately limiting the scope of the work. The findings from this study showed how automated image classification could be used to overcome such limitations. In addition, the methods from this study can help researchers in tobacco control identify what proportion of viewers on a social media site are interested in certain products; such methods may be crucial to document the every changing tobacco landscape.

Limitations

The findings from this study should be considered with several limitations in mind, including the fact that our task was a simple binary classification (hookah vs nonhookah), which may result in high accuracy. To eliminate the problem of overfitting, we used ReLU, softmax, dropout layers in a CNN, and utilized several different training datasets (the number of datasets is different, which increased from 42 to 420; [Figure 11](#)). The

methods developed in this study were only applied in the context of images from Instagram that focused on waterpipes and should be applied in more categories and other contexts in the future. While we had high accuracy in classification, accuracy could be improved with better input features from the CNN model. In the future, researchers should try to enlarge the sets of training images to extract specific features of an image, which may achieve higher accuracy with less computation power.

Conclusions

Findings demonstrated that by combining CNN and SVM to classify images resulted in 99.5% (418/420) accuracy in image classification, which is an improvement over earlier method using SVM, CNN, or BOF alone. A CNN extracts more features of the images, allowing the SVM classifier to be better informed, which results in higher accuracy compared with methods that extract fewer features. Future research can use our method to reduce computational time in identifying objects in images.

Acknowledgments

Research reported in this publication was supported by Grant # P50CA180905 from the National Cancer Institute and the Food and Drug Administration (FDA) Center for Tobacco Products. The National Institutes of Health or FDA had no role in study design, collection, analysis, and interpretation of data, writing the report, and the decision to submit the report for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or FDA.

Conflicts of Interest

None declared.

Authors' Contributions

YZ and JPA conceived of the study and analyzed the data. YZ and JPA drafted the initial manuscript. JBU and TBC received funding for the study. JBU and TBC revised the manuscript for important intellectual content.

References

1. Instagram Business Team. Instagram.com. 2016 Sep 22. Hitting 500,000 Advertisers URL: <https://business.instagram.com/blog/500000-advertisers/> [accessed 2018-09-26] [[WebCite Cache ID 72j5U9uDE](#)]
2. Allem JP, Cruz TB, Unger JB, Toruno R, Herrera J, Kirkpatrick MG. Return of cartoon to market e-cigarette-related products. *Tob Control* 2018 Jul 26. [doi: [10.1136/tobaccocontrol-2018-054437](https://doi.org/10.1136/tobaccocontrol-2018-054437)] [Medline: [30049688](#)]
3. Allem JP, Escobedo P, Chu KH, Boley TC, Unger JB. Images of Little Cigars and Cigarillos on Instagram Identified by the Hashtag #swisher: Thematic Analysis. *J Med Internet Res* 2017 Jul 14;19(7):e255 [[FREE Full text](#)] [doi: [10.2196/jmir.7634](https://doi.org/10.2196/jmir.7634)] [Medline: [28710057](#)]
4. Allem JP, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related posts on Twitter. *Drug Alcohol Depend* 2018 Dec 01;190:1-5. [doi: [10.1016/j.drugalcdep.2018.05.018](https://doi.org/10.1016/j.drugalcdep.2018.05.018)] [Medline: [29958115](#)]
5. Perronnin F, Sánchez J, Mensink T. Improving the fisher kernel for large-scale image classification. In: *Computer Vision-ECCV. 2010 Presented at: European conference on computer vision; September 5, 2010; Springer, Berlin, Heidelberg* p. 143-156.
6. Verma K, Verma LK, Tripathi P. Image Classification using Backpropagation Algorithm. *Journal of Computer Science* 2014;1:7-15.
7. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps.: arXiv:1312.6034v2; 2014. URL: <https://arxiv.org/pdf/1312.6034.pdf> [accessed 2018-10-30] [[WebCite Cache ID 73YkRWIDb](#)]
8. Orlov N, Shamir L, Macura T, Johnston J, Eckley DM, Goldberg IG. WND-CHARM: Multi-purpose image classification using compound image transforms. *Pattern Recognit Lett* 2008 Jan;29(11):1684-1693 [[FREE Full text](#)] [Medline: [18958301](#)]
9. Dehkordy HI. Automated Image Classification Via Unsupervised Feature Learning by K-Means. Indiana University-Purdue University Indianapolis 2015:1-80 [[FREE Full text](#)]

10. Sun J, Yang J, Zhang C, Yun W, Qu J. Automatic remotely sensed image classification in a grid environment based on the maximum likelihood method. *Mathematical and Computer Modelling* 2013 Aug;58(3-4):573-581. [doi: [10.1016/j.mcm.2011.10.063](https://doi.org/10.1016/j.mcm.2011.10.063)]
11. Chapelle O, Haffner P, Vapnik VN. Support vector machines for histogram-based image classification. *IEEE Trans Neural Netw* 1999;10(5):1055-1064. [doi: [10.1109/72.788646](https://doi.org/10.1109/72.788646)] [Medline: [18252608](https://pubmed.ncbi.nlm.nih.gov/18252608/)]
12. Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification. In: *European Conference on Computer Vision*. 2006 Presented at: European conference on computer vision; May 7, 2006; Springer, Berlin, Heidelberg p. 490-503.
13. Cireşan D, Meier U, Masci J, Schmidhuber J. Multi-column deep neural network for traffic sign classification. *Neural Netw* 2012 Aug;32:333-338. [doi: [10.1016/j.neunet.2012.02.023](https://doi.org/10.1016/j.neunet.2012.02.023)] [Medline: [22386783](https://pubmed.ncbi.nlm.nih.gov/22386783/)]
14. Jiang YG, Ngo CW CW, Yang J. Towards optimal bag-of-features for object categorization semantic video retrieval. 2007 Presented at: *Proceedings of the 6th ACM international conference on Image and video retrieval*; July 9, 2007; Amsterdam, Netherlands p. 494-501.
15. Marszaek M, Schmid C. Spatial weighting for bag-of-features. 2006 Presented at: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; June 17-22, 2006; New York, New York, United States p. 2118-2125.
16. Auria L, Moro R. Support Vector Machines (SVM) as a Technique for Solvency Analysis. 2007 Presented at: *DIW Diskussionspapiere - DIW Berlin Publikationen von Forscherinnen und Forschern des DIW Berlin*; August 2008; Berlin, Germany.
17. Burges CJ. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery* 1998;2(2):121-167.
18. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 2000 Oct;16(10):906-914. [Medline: [11120680](https://pubmed.ncbi.nlm.nih.gov/11120680/)]
19. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556* 2015:1-14.
20. Vedaldi A, Lenc K. Matconvnet: Convolutional neural networks for matlab. 2015 Presented at: *Proceedings of the 23rd ACM international conference on Multimedia*; October 26 - 30, 2015; Brisbane, Australia p. 689-692.
21. Ayers JW, Leas EC, Allem JP, Benton A, Dredze M, Althouse BM, et al. Why do people use electronic nicotine delivery systems (electronic cigarettes)? A content analysis of Twitter, 2012-2015. *PLoS One* 2017;12(3):e0170702 [FREE Full text] [doi: [10.1371/journal.pone.0170702](https://doi.org/10.1371/journal.pone.0170702)] [Medline: [28248987](https://pubmed.ncbi.nlm.nih.gov/28248987/)]
22. Allem JP, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends. *JMIR Public Health Surveill* 2017 Dec 20;3(4):e98 [FREE Full text] [doi: [10.2196/publichealth.8641](https://doi.org/10.2196/publichealth.8641)] [Medline: [29263018](https://pubmed.ncbi.nlm.nih.gov/29263018/)]
23. Allem JP, Ramanujam J, Lerman K, Chu KH, Boley Cruz T, Unger JB. Identifying Sentiment of Hookah-Related Posts on Twitter. *JMIR Public Health Surveill* 2017 Oct 18;3(4):e74 [FREE Full text] [doi: [10.2196/publichealth.8133](https://doi.org/10.2196/publichealth.8133)] [Medline: [29046267](https://pubmed.ncbi.nlm.nih.gov/29046267/)]
24. Allem JP, Chu KH, Cruz TB, Unger JB. Waterpipe Promotion and Use on Instagram: #Hookah. *Nicotine Tob Res* 2017 Oct 01;19(10):1248-1252. [doi: [10.1093/ntr/ntw329](https://doi.org/10.1093/ntr/ntw329)] [Medline: [28077449](https://pubmed.ncbi.nlm.nih.gov/28077449/)]
25. Chu KH, Allem JP, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. *Tob Control* 2016 Dec;26(5):575-578 [FREE Full text] [doi: [10.1136/tobaccocontrol-2016-053052](https://doi.org/10.1136/tobaccocontrol-2016-053052)] [Medline: [27660111](https://pubmed.ncbi.nlm.nih.gov/27660111/)]
26. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. 2010 Presented at: *Proceedings of the 27th international conference on machine learning (ICML-10)*; June 21-24, 2010; Haifa, Israel p. 807-814.
27. Ballester P, de Araújo RM. On the Performance of GoogLeNet and AlexNet Applied to Sketches. 2016 Presented at: *Thirtieth AAAI Conference on Artificial Intelligence*; February 12 - 17, 2016; Phoenix, Arizona.
28. Yu W, Yank K, Bai Y, Xiao T, Yao H, Rui Y. Visualizing and comparing AlexNet and VGG using deconvolutional layers. 2016 Presented at: *ICML'16 Proceedings of the 33rd International Conference on International Conference on Machine Learning*; June 19 - 24, 2016; New York, NY, United States.
29. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Communications of the ACM* 2017;60(6):84-90.
30. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580* 2012:1-18.
31. Hua S, Sun Z. Support vector machine approach for protein subcellular localization prediction. *Bioinformatics* 2001 Aug;17(8):721-728. [Medline: [11524373](https://pubmed.ncbi.nlm.nih.gov/11524373/)]
32. Escalera S, Pujol O, Radeva P. On the decoding process in ternary error-correcting output codes. *IEEE Trans Pattern Anal Mach Intell* 2010 Jan;32(1):120-134. [doi: [10.1109/TPAMI.2008.266](https://doi.org/10.1109/TPAMI.2008.266)] [Medline: [19926903](https://pubmed.ncbi.nlm.nih.gov/19926903/)]
33. Escalera S, Pujol O, Radeva P. Separability of ternary codes for sparse designs of error-correcting output codes. *Pattern Recognition Letters* 2009;30(3):285-297.
34. Jágou H, Douze M, Schmid C. Improving Bag-of-Features for Large Scale Image Search. *Int J Comput Vis* 2009 Aug 11;87(3):316-336. [doi: [10.1007/s11263-009-0285-2](https://doi.org/10.1007/s11263-009-0285-2)]

35. Pechmann C, Pan L, Delucchi K, Lakon CM, Prochaska JJ. Development of a Twitter-based intervention for smoking cessation that encourages high-quality social media interactions via automessages. *J Med Internet Res* 2015;17(2):e50 [FREE Full text] [doi: [10.2196/jmir.3772](https://doi.org/10.2196/jmir.3772)] [Medline: [25707037](https://pubmed.ncbi.nlm.nih.gov/25707037/)]
36. Pechmann C, Delucchi K, Lakon CM, Prochaska JJ. Randomised controlled trial evaluation of Tweet2Quit: a social network quit-smoking intervention. *Tob Control* 2017 Dec;26(2):188-194 [FREE Full text] [doi: [10.1136/tobaccocontrol-2015-052768](https://doi.org/10.1136/tobaccocontrol-2015-052768)] [Medline: [26928205](https://pubmed.ncbi.nlm.nih.gov/26928205/)]
37. Allem JP, Escobedo P, Chu KH, Soto DW, Cruz TB, Unger JB. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. *Tob Control* 2017 Dec;26(2):226-229 [FREE Full text] [doi: [10.1136/tobaccocontrol-2015-052757](https://doi.org/10.1136/tobaccocontrol-2015-052757)] [Medline: [26956467](https://pubmed.ncbi.nlm.nih.gov/26956467/)]
38. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. *Addict Behav* 2017 Oct;73:81-93. [doi: [10.1016/j.addbeh.2017.05.002](https://doi.org/10.1016/j.addbeh.2017.05.002)] [Medline: [28499259](https://pubmed.ncbi.nlm.nih.gov/28499259/)]
39. Chen AT, Zhu SH, Conway M. Combining Text Mining and Data Visualization Techniques to Understand Consumer Experiences of Electronic Cigarettes and Hookah in Online Forums. *Online J Public Health Inform* 2015;7(1):e117.
40. Krauss M, Sowles SJ, Moreno M, Zewdie K, Gruzca RA, Bierut LJ, et al. Hookah-Related Twitter Chatter: A Content Analysis. *Prev Chronic Dis* 2015 Jul 30;12:E121 [FREE Full text] [doi: [10.5888/pcd12.150140](https://doi.org/10.5888/pcd12.150140)] [Medline: [26226068](https://pubmed.ncbi.nlm.nih.gov/26226068/)]
41. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *J Med Internet Res* 2013 Aug 29;15(8):e174 [FREE Full text] [doi: [10.2196/jmir.2534](https://doi.org/10.2196/jmir.2534)] [Medline: [23989137](https://pubmed.ncbi.nlm.nih.gov/23989137/)]

Abbreviations

BOF: bag-of-features
CNN: convolutional neural network
ConvNets: convolutional networks
FDA: Food and Drug Administration
ReLU: rectified linear units
SVM: support vector machine

Edited by G Eysenbach; submitted 27.03.18; peer-reviewed by M Paul, A Benis; comments to author 05.06.18; revised version received 30.07.18; accepted 07.08.18; published 21.11.18

Please cite as:

Zhang Y, Allem JP, Unger JB, Boley Cruz T

Automated Identification of Hookahs (Waterpipes) on Instagram: An Application in Feature Extraction Using Convolutional Neural Network and Support Vector Machine Classification

J Med Internet Res 2018;20(11):e10513

URL: <http://www.jmir.org/2018/11/e10513/>

doi: [10.2196/10513](https://doi.org/10.2196/10513)

PMID: [30452385](https://pubmed.ncbi.nlm.nih.gov/30452385/)

©Youshan Zhang, Jon-Patrick Allem, Jennifer Beth Unger, Tess Boley Cruz. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 21.11.2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.