Nicotine and Tobacco Research, 2026, **00**, 1–6
https://doi.org/10.1093/ntr/ntag125
Published: 9 July 2026
Original Investigation

Automatic detection of e-cigarette screens using object detection and vision language models

Hunter Morera, PhD[1,*], James Jun, MSCS[1], Charity Hilton, MSCS[1], Dianna King, MBA[1], Elizabeth L. Seaman Jones, PhD[2], Kristy Marynak, PhD[2], Elisha Crane, MPH[2]

[1]Georgia Tech Research Institute, 756 W Peachtree St NW, Atlanta, GA 30318, United States
[2]CDC Foundation, 600 Peachtree St NE #1000, Atlanta, GA 30308, United States
[*]Corresponding author: Hunter Morera, PhD, Georgia Tech Research Institute, 756 W Peachtree St NW, Atlanta, GA, 30318, United States (hunter.morera@gtri.gatech.edu).

Abstract

Introduction: **The e-cigarette market evolves faster than regulation and enforcement. In recent years, e-cigarette manufacturers have introduced vapes featuring not just addictive, flavored nicotine, but also technological features that appeal to youth. These products with digital screens have the potential to gamify the use of an addictive drug. To keep pace with these developments, automated tools capable of rapidly analyzing the online marketplace could inform regulatory enforcement.**

Methods: **The goal of this work is to automatically detect and classify e-cigarette devices with screens. To do this, we first use an object detection model to filter web-scraped images that contain e-cigarettes. Product images and descriptions are then provided to a vision language model (VLM) to determine if the e-cigarette contains a screen. We evaluated the pipeline with images and text descriptions from five e-cigarette online retailers and manually validated predictions.**

Results: **The object detection model was trained on approximately 7000 images and tested on 3920 images achieving an accuracy, precision, recall, and F1 score of 0.95, 0.98, 0.94, 0.96 respectively in identifying e-cigarette devices. The VLM was tested on 2401 images and was able to detect screens on e-cigarette devices with an accuracy of 0.92, precision of 0.96, recall of 0.87, and F1 score of 0.92.**

Conclusions: **Object detection models and VLMs can be combined to automatically detect and classify e-cigarettes with integrated screens. This method can enable rapid-response pipelines to identify youth-appealing features in emerging e-cigarette products and provide policymakers insights into their proliferation in the online marketplace.**

Implications: **New e-cigarettes with potentially youth-appealing technological features, such as integrated screens, continue to be introduced to the US market, challenging timely and effective tobacco product surveillance and regulatory efforts. Digital screens on e-cigarettes offer greater user control and make products that deliver addictive, flavored nicotine resemble entertainment devices, rather than tobacco products. Our automated screen detection pipeline provides a scalable, rapid-response tool for monitoring online e-cigarette retailers in near real-time, quantifying the proliferation of products with specific features. This pipeline could be adapted to accommodate a variety of new features for detection, here we use the presence of screens as a test example. Rapid identification and classification of new tobacco products available online could help inform interventions before the sale and use of such products becomes widespread.**

Keywords electronic cigarettes, medical surveillance, computer vision, machine learning, regulatory, vaping

Introduction

E-cigarettes are the most commonly used tobacco product among youth and young adults in the United States.[1] E-cigarettes continue to change rapidly, incorporating new features and characteristics that could increase their appeal and addictiveness among young people.[2,3] An emerging trend in recent years are so-called "smart" vapes—devices that deliver flavored, addictive nicotine that incorporate digital displays that can allow for built-in gaming systems, touchscreens, smartphone app connectivity, and a variety of other features. Devices incorporating these features rapidly emerged as top-selling brands in the US e-cigarette market, with national retail sales estimates placing Geek Bar Pulse and Raz as the third and fifth top-selling e-cigarettes, as of May 2025.[4]

Received: January 27, 2026. **Revised:** May 15, 2026. **Accepted:** May 22, 2026

Use of "smart" vapes appears prominent among US youth and young adults who use e-cigarettes. According to the Tobacco Epidemic Evaluation Network (TEEN+) Study Wave 6, fielded in early 2025, an estimated 15.2% of youth and young adults use e-cigarettes, of whom 37.3% report using a "smart" vape in the past 30 days. Among this group, four-in-five (81.7%) youth ages 13–17 and three-in-five (61.4%) young adults ages 18–27 reported use of Geek Bar in the past month.[5] Youth and young adults reported that the devices they used in the past month included indicators for device settings, e-liquid levels, and puffs consumed; games; and Bluetooth connectivity.[5]

"Smart" vapes pose a unique concern for public health, potentially tapping shared addiction mechanisms present in both nicotine and video games.[6,7] These sophisticated devices often include interactive, on-device LED screens with interfaces which allow users to view and customize settings and engage with features explicitly designed to encourage repeated use.[7] Such features draw from addiction mechanisms present in video games, incorporating gamification elements such as usage rewards and progress tracking to reinforce addictive behaviors.[7,8] The youth appeal of these devices is amplified by marketing strategies that blur the lines between tobacco products and consumer entertainment. Manufacturers have long produced e-cigarettes designed to resemble toys, popular video game consoles, food items, and popular cartoon characters that specifically attract young users, resulting in warnings and civil money complaints from the U.S. Food and Drug Administration.[7–9] Despite enforcement actions, e-cigarette manufacturers have continued to develop and release new e-cigarettes replicating and building upon these device features. The regulatory challenges posed by rapid e-cigarette diversification underscore the need for innovative approaches to monitoring product marketing and design.

Traditional surveillance methods struggle to keep pace with the volume and speed of tobacco product diversification—a phenomenon exemplified by, but not limited to, the introduction of vapes with digital screens. Surveys offer insights on behaviors and attitudes but may struggle to capture issues emerging shortly before or during fielding, a particular challenge in the rapidly changing e-cigarette market. Additionally, there is a time lag in collecting analyzing and disseminating results from surveys even when information on emerging issues is captured. Point-of-sale (POS) retail sales scanner data can provide near real-time sales estimates but may contain incomplete data and are subject to varying time lags in data delivery depending on vendor production and licensing schedules. Point-of-sale scanner data product attributes are provided directly by the vendor and usually do not track novel features of interest to public health, such as on-device video games or flavor intensity toggles. Researchers manually conduct additional, time-intensive internet investigations to track these novel device features or backfill incomplete variable data from the vendor, followed by extensive analysis to produce interpretable results. Furthermore, POS retail sales scanner data do not sample from online retailers, with estimates placing the size of the online marketplace at approximately 19% of total e-cigarette sales in 2021.[10] Vape and tobacco specialty shop data are also not included in POS data, further challenging surveillance and regulatory efforts. Because many online vendors offer products wholesale to vape shops, understanding the online marketplace could provide greater insight into vape shop offerings.[11]

Rapidly monitoring the development and proliferation of devices containing youth-appealing features in an automated, systematic, and replicable manner is thus vital for informing efforts to protect public health. Automated artificial intelligence (AI) methods, particularly object detection and vision-language models (VLMs), offer a promising solution. By leveraging rapidly and regularly collected data scraped from the web, combined with image and text analysis via VLMs, we endeavored to demonstrate that AI pipelines can systematically and efficiently identify emerging product features such as the presence of on-device screens, a novel feature present in most "smart" vapes and which serves as the interface for many of these devices' potentially youth-appealing interactivity, over thousands of e-cigarettes, providing timely data to inform policymakers and regulatory agencies. Prior work has applied object detection to tobacco surveillance research questions.[12–15] However, this paper is unique in its goal to add VLM technology to the tobacco retail data surveillance pipeline. We present an application of VLMs to identify the presence of on-device screens from thousands of e-cigarettes sold by online tobacco retailers as a case study for the application of these methods to tobacco product surveillance.

Methods

The dataset for this study was collected by scraping five online e-cigarette retailers: myvaporstore.com, csvape.com, vape.com, mipod.com, and vapesourcing.com, between September and October 2024. These websites were selected based on their consistent traffic and ease of product scraping. To scrape these websites, we created custom Python scripts using the Selenium library[16] and saved product images and descriptions for use in our machine learning pipeline. A sample of images is shown in Figure 1.

We created a two-phase pipeline for detecting e-cigarette images and then identifying e-cigarette devices with screens. A visual representation of the pipeline can be seen in Figure 2. The first phase of our pipeline automatically separates images of e-cigarettes from other products, such as e-cigarette liquids, atomizers, and other accessories. Our custom scrapers did not discriminate between product types, resulting in a dataset that contained many non-e-cigarette products. To enable the isolation of e-cigarette images for downstream processing, we used an object detection network that generates bounding boxes with pixel coordinates and associated class labels.[17] Specifically, we deployed a well-known object detection network known as YOLO, (You Only Look Once). This object detection network has shown superior accuracy and speed of processing when evaluated on benchmark datasets[18] Here, we used YOLO version 8 (YOLOv8) due to its stability and open-source codebase provided by Ultralytics.[19] To be able to detect and classify e-cigarette devices, the network had to be trained on labeled images. Most of these images came from an open source dataset[20] consisting of 5700 images of e-cigarettes and cigarettes already labeled with bounding boxes. To augment the data, our team manually labeled images scraped from myvaporstore.com and csvape.com using the Computer Vision Annotation Tool.[21] The addition of this data brought the total number of labeled images to 7375 with a total number of labeled e-cigarettes to 10 145 as some images contain more than one e-cigarette.

Figure 1 Sample products scraped from online stores. From left to right: e-cigarette liquid (vape.com), atomizer (vapesourcing.com), and e-cigarette with screen (csvape.com). Images are included for informational purposes and do not constitute an endorsement by the CDC Foundation.

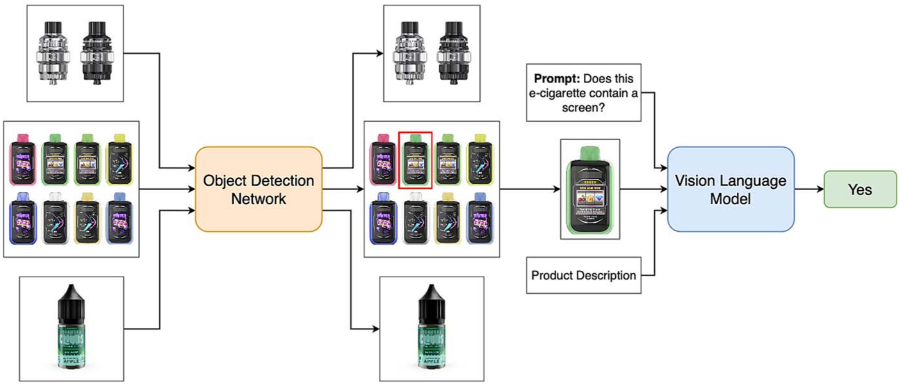

Figure 2 A visual representation of our automated pipeline. Scraped images are first passed through a trained object detection network to detect e-cigarette products. The image is then cropped around the most confidently predicted e-cigarette in each image and passed to a VLM. This model takes the image along with the product description and determines if the e-cigarette contains a screen. Images are included for informational purposes and do not constitute an endorsement by the CDC Foundation. VLM, vision language model.

Table 1 Summary of data splits for each model.

Model	Train images	Validation images	Test images
Object detection hyperparameter tuning	6153	1222	0
Final object detection model	7375	0	3920
VLM	0	0	2401

VLM, visual language model.

To train a YOLOv8 model for e-cigarette detection, first, the dataset was partitioned into a training set of 6153 images and validation set of 1222 to conduct hyperparameter optimization. Following the identification of optimal hyperparameters, a final YOLOv8 model was initialized with COCO pre-trained weights and trained for 100 epochs. This final iteration utilized the comprehensive dataset ($N = 7375$), merging the training and validation subsets to maximize the feature space for the resulting e-cigarette detection model. The final model was tested using 3920 unseen images from mipod.com, myvaporstore.com, and vapesourcing.com. Of the 3920 test images, 2401 images were predicted to contain an e-cigarette device. These images were then cropped using the bounding box of the most confidently predicted e-cigarette by the network for further processing. A summary of the data splits can be seen in Table 1.

The second phase of our pipeline relied on VLMs, a type of AI that combines large language models with computer vision models to process both images and text simultaneously.[22] Vision language models can be used for many different tasks such as describing what is contained in an image (image captioning), understanding and answering questions about an image (visual question answering), and translating visual information into text or text into visuals.[23] These models are pre-trained on vast amounts of unstructured data from the internet and can be used without tuning to a specific problem domain.[24] In our pipeline, we used a VLM to process the cropped images from the object detection model along with the description scraped from the website (if available) to determine if the e-cigarette device contains a screen. The VLM we chose for this part of the pipeline was the InternVL3-38B model[25,26] selected because it is open-source, cost-free, and performs competitively on various benchmark datasets. We tested the model with a zero-shot learning approach in which no examples were provided to the model prior to testing. To keep results consistent, we used a very low temperature (0.05), a model parameter which controls how creative or predictable its responses are. We asked the model the

Table 2 Evaluation metrics for two phase pipeline.

Metric	Phase 1—Object detection (*N* = 3920)	Phase 2—VLM (*N* = 2401)
Accuracy	0.95	0.92
Precision	0.98	0.96
Recall	0.94	0.87
F1	0.96	0.92

*Note: Accuracy is calculated as (True Positives + True Negatives)/(Total samples), Precision is calculated as True Positives/(True Positives + False Positives), Recall is calculated as True Positives/(True Positives + False Negatives), and F1 score is calculated as (2 * precision * recall)/(precision + recall).* VLM, visual language model.

Table 3 Cross-tabulation results for pipeline phase 1, objective detection network, for predicting the presence of e-cigarettes in web-scraped images.

N = 3920	True e-cigarette	True no e-cigarette
Predicted e-cigarette	2354 (60.1%)	47 (1.2%)
Predicted no e-cigarette	155 (4.0%)	1364 (34.8%)

same question for all images and descriptions as follows: "The following text consists of the name and a description for the e-cigarette in the image. Based on the text or the image, is there a screen on the device? Answer either yes or no". Each of the 2401 responses were validated manually by our team using the same information as the VLM. This involved personally investigating the images and descriptions provided to the VLM and manually validating the presence of a screen against the output of the VLM.

Results

We evaluated the two phases of our pipeline separately. The object detection component of our pipeline was able to detect e-cigarettes with an accuracy of 0.95, precision of 0.98, recall of 0.94, and F1 score (harmonic mean of precision and recall, balancing both, ranges 0–1; 1 is perfect) of 0.96 on the test set of 3920 images (Table 2). The object detection model had a false positive (FP) rate of 1.2% and a false negative (FN) rate of 4.0% (Table 3).

Of the 3920 images processed by the object detection network, 2401 (61.3%) were predicted to contain an e-cigarette and provided to the VLM for screen detection. The VLM was able to predict the presence of a screen with an accuracy of 0.92, precision of 0.96, recall of 0.87, and F1 score of 0.92. The 2401 cropped images were individually validated; however, some of these did not contain an e-cigarette device due to the performance of the object detection model resulting in 40 false positives (Table 3). Images that were not e-cigarette devices (atomizers, juices, etc.) were said to have a ground truth of "no screen". The VLM had a false positive rate of 1.7% and a false negative rate of 6.2% (Table 4). The table also shows that of the products processed, 1194 (49.7%) contained a screen, of which our model successfully classified 1044 (87.4%) of them as containing a screen.

Table 4 Cross-tabulation results for pipeline phase 2, VLM, for predicting the presence of screens in e-cigarette images.

N = 2401	True screen	True no-screen
Predicted screen	1044 (43.5%)	40 (1.7%)
Predicted no-screen	150 (6.2%)	1167 (48.6%)

VLM, visual language model.

Discussion

E-cigarette manufacturers continue to diversify their products rapidly, introducing e-cigarettes and features that heighten youth appeal and undermine public health surveillance and prevention efforts. Traditional monitoring methods struggle to match the pace or scale of this market, leaving regulators at a disadvantage. Agencies can leverage automated data collection methods and AI to capture and process the vast number of e-cigarette products that enter the market each day to better inform timely, evidence-based regulations.

In this paper, we developed and evaluated a novel method for automatic detection of screens on e-cigarette devices using a two-phase automated pipeline that integrates object detection networks with VLMs to detect and classify e-cigarette devices containing screens. Our approach achieved a high level of performance with 95% accuracy in identifying e-cigarette devices and 92% accuracy in detecting the presence of screens. Of all objects which were designated as e-cigarettes and provided to the VLM for screen detection, 1194 (49.7%) included screens, of which 1044 (87.4%) were correctly identified by the VLM. Our method of combining visual and textual product information provides a scalable and efficient framework for monitoring emerging features in the online e-cigarette market. This automated process can be integrated into existing surveillance workflows. Right now, when POS retail scanner data is purchased, it is often missing data for product features and categories, requiring manual review of products and coding of missing data. Fully automated VLM models could be leveraged to help fill in missing tobacco product feature and category data prior to data analysis. This would streamline the data analysis and interpretation process and ensure that data can be shared more rapidly.

We observed slightly higher false negative than false positive rates in both phases of our pipeline, for the object detection model the FN rate was 4.0% vs. FP 1.2%, and for the VLM the FN rate was 6.2% vs. FP 1.7%. For our object detection model, these false negatives appear to come from the large variety of shapes and sizes of e-cigarette devices, causing some to be classified as other products such as vape juices or accessories. In the VLM part of the pipeline, the apparent cause of the higher false negative rates was the many devices with very small screens such as digital battery indicators that were not mentioned in the product description. Increasing the number and variety of e-cigarettes in the training set could result in better performance of our object detection model and reduce these false negatives. Future research could also involve a "few-shot" learning technique to increase performance of the VLMs' ability to detect screens. This involves providing some difficult examples to the VLM in the prompt along with the reasoning behind the human decision before prompting it for answers for the remaining examples.

Our work has several limitations to note. First, the dataset was derived from a select group of online e-cigarette retailers and may not fully capture the diversity of products available in other retail channels, such as brick-and-mortar retail stores, physical tobacco specialty stores, or social media marketplaces. Second, our pipeline relied on product images and descriptions provided by the retailers, which may not always be accurate or complete; mislabeled or misleading product listings may bias results. Third, this work focused specifically on the detection of integrated screens; additional studies are needed to extend this approach to other emerging features that contribute to youth appeal, such as novel flavors or other form factors. Additionally, sales volumes and purchaser demographics are unavailable for all products included in the sample, an inherent limitation of online web scraping. Finally, data are scraped from what is available online and rely on the accuracy of posted text and images. To mitigate the potential that pipelines rely on false or misleading information about products posted online, it may be necessary to periodically review a subset of products.

This work highlights the broader potential of AI to supplement and enhance tobacco product surveillance. While web scraped data does not provide purchaser demographic information (and thus cannot be attributed directly to youth) or associated sales volumes, they serve as an important and rapidly obtainable source of marketplace availability and provide rich feature data which can be used in combination with traditional surveillance methods to enhance monitoring efforts. Automated pipelines can shift regulatory efforts from a reactive response to proactive market intelligence, easing the burden on public health officials and regulators while providing real-time evidence to inform restriction and enforcement policies.

Funding

This work was supported by the *Bloomberg Initiative to Reduce Tobacco Use* through the CDC Foundation with a grant from Bloomberg Philanthropies.

Declaration of interests

None declared.

Author contributions

Hunter Morera (Conceptualization [lead], Formal analysis [lead], Investigation [lead], Methodology [lead], Software [lead], Validation [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [lead]), James Jun (Conceptualization [supporting], Investigation [supporting], Methodology [supporting], Validation [supporting], Writing—original draft [equal], Writing—review & editing [equal]), Charity Hilton (Conceptualization [equal], Data curation [lead], Funding acquisition [lead], Investigation [supporting], Project administration [lead], Supervision [lead], Writing—original draft [supporting], Writing—review & editing [supporting]), Dianna King (Conceptualization [supporting], Data curation [supporting], Funding acquisition [equal], Project administration [equal], Supervision [equal], Writing—review & editing [supporting]), Elizabeth L. Seaman Jones (Writing—original draft [equal], Writing—review & editing [equal]), Kristy Marynak (Writing—original draft [equal], Writing—review & editing [equal]), and Elisha Crane (Conceptualization [equal], Investigation [equal], Methodology [equal], Supervision [equal], Writing—original draft [equal], Writing—review & editing [equal])

Data availability

The data used in this paper is available upon request to the corresponding author.

CDC Foundation disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC Foundation.

References

1. Park-Lee E, Jamal A, Cowan H, *et al.* Notes from the field: e-cigarette and nicotine pouch use among middle and high school students—United States, 2024. *MMWR Morb Mortal Wkly Rep*. 2024;73:774-778. https://doi.org/10.15585/mmwr.mm7335a3.
2. Wong M, Talbot P. Pac-Man on a vape: electronic cigarettes that target youth as handheld multimedia and gaming devices. *Tob Control*. 2024;34:849-851. https://doi.org/10.1136/tc-2024-058794 Published online.
3. Marynak K, Crane E, VanFrank B. Vaping. *JAMA*. 2025;333:536-537. https://doi.org/10.1001/jama.2024.25255.
4. CDC Foundation. *Monitoring U.S. E-Cigarette Sales: National Trends Data Brief (Issue 31)*. CDC Foundation; 2025: Accessed September 30, 2025. https://tobaccomonitoring.org/wp-content/uploads/2025/08/National-E-Cigarette-Sales-Data-Brief_05.18.25_.pdf.
5. CDC Foundation. Monitoring tobacco product use among youth and young adults in the U.S. In: *TEEN+ Data Snapshot (Issue 2): Smart Vapes.* Atlanta, GA: CDC Foundation; December 2025. Accessed June 10, 2026. https://tobaccomonitoring.org/wp-content/uploads/2025/12/TEEN-Data-Snapshot-Issue-2.pdf.
6. Unger JB, Vassey J, Soto DW, Galimov A. Vaping devices with video games. *Subst Use Misuse*. 2024;59:2149-2150 Accessed September 30, 2025. https://www.tandfonline.com/doi/abs/10.1080/10826084.2024.2392547.
7. Flayelle M, Brevers D, King DL, Maurage P, Perales JC, Billieux J. A taxonomy of technology design features that promote potentially addictive online behaviours. *Nat Rev Psychol*. 2023;2:136-150. https://doi.org/10.1038/s44159-023-00153-4.
8. Tatum Z, Leventhal A, Wipfli HL. Playtime: vaping devices designed as cartoons and toys may appeal to kids. *Tob Control*. 2024;33:693-694. https://doi.org/10.1136/tc-2022-057908.
9. Center for Tobacco Products. *Retailers Warned to Stop Selling Illegal e-Cigarettes Resembling Youth-Appealing Characters, School Supplies, Toys, and Drinks.* Silver Spring, MD: US Food and Drug Administration; February 29, 2024. Accessed

September 30, 2025. https://www.fda.gov/tobacco-products/ctp-newsroom/retailers-warned-stop-selling-illegal-e-cigarettes-resembling-youth-appealing-characters-school.
10. Seaman EL, Ali FRM, Schillo BA, Vallone DM, King BA. Different times call for different measures: using retail sales to monitor the tobacco product landscape. *Am J Prev Med*. 2022;63:e99-e102. https://doi.org/10.1016/j.amepre.2022.03.028.
11. Source I. *How to Start a Profitable Vape Wholesale Business.* Jacksonville, FL: Inventory Source website; July 2025. Accessed November 11, 2025. https://www.inventorysource.com/how-to-start-profitable-vape-wholesale/.
12. Vassey J, Kennedy CJ, Chang HCH, Smith AS, Unger JB. Scalable surveillance of e-cigarette products on Instagram and TikTok using computer vision. *Nicotine Tob Res*. 2023;26:552-560. https://doi.org/10.1093/ntr/ntad224.
13. Murthy D, Ouellette RR, Anand T, *et al.* Using computer vision to detect e-cigarette content in TikTok videos. *Nicotine Tob Res*. 2024;26:S36-S42. https://doi.org/10.1093/ntr/ntad184.
14. Kong G, Keshari S, Ouellette RR, Lee J, Murthy D. Computer vision models for detecting large cigars on social media. *Nicotine Tob Res*. 2025;28:873-876. https://doi.org/10.1093/ntr/ntaf206.
15. Kennedy CJ, Gajjar B, Chang HCH, Unger JB, Vassey J. Demographic trends in e-cigarette social media marketing: perceiving gender presentation and facial age via computer vision. *Nicotine Tob Res*. Published online March 7. 2025;27:2063-2071. https://doi.org/10.1093/ntr/ntaf057.
16. Selenium Python Bindings [Computer Software]. *Version 4.22.0.* Online: The Selenium Project; 2024: Accessed September 20, 2024. https://selenium-python.readthedocs.io/.
17. Hsu CL, Chen YW. A comprehensive survey of machine learning techniques and models for object detection. *Sensors (Basel)*. 2024;25:214. https://doi.org/10.3390/s25010214.
18. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Washington, DC: IEEE Computer Society; 2016:779-788. https://doi.org/10.1109/CVPR.2016.91.
19. Jocher G, Chaurasia A, Qiu J. *Ultralytics YOLOv8 [Computer Software]. Version 8.2.0.* Online: Ultralytics; 2023: Accessed August 4, 2024. https://github.com/ultralytics/ultralytics.
20. Takoyati. Cigarette & E-cigarette detection dataset [data set]. *Roboflow*. 2025; Accessed May 8, 2024. https://universe.roboflow.com/takoyati/cigarette-e-cigarette-detection.
21. CVAT.ai. *CVAT: Annotate Better with CVAT, the Industry-Leading Data Engine for Machine Learning. Version [computer software]. Version 2.14.0.* Online: GitHub; 2024: Accessed November 19, 2024. https://github.com/cvat-ai/cvat.
22. IBM. Vision-Language Models: An Introduction. Armonk, NY: IBM Corporation; May 20, 2024. Updated May 20, 2024. Accessed July 9, 2025. https://www.ibm.com/think/topics/vision-language-models.
23. Wu J, Gan W, Chen Z, Wan S, Yu PS. Multimodal large language models: a survey. In: He J, Palpanas T, Hu X, eds. *2023 IEEE International Conference on Big Data (Big Data)*. Piscataway, NJ: IEEE; 2023:1-10. https://doi.org/10.1109/BigData59044.2023.10386743.
24. Radford A, Kim JW, Hallacy C, *et al.* Learning transferable visual models from natural language supervision. In: Meila M, Zhang T, eds. *Proceedings of the 38th International Conference on Machine Learning (ICML).* Cambridge, MA: PMLR; 2021:8748-8763. https://proceedings.mlr.press/v139/radford21a.html.
25. Chen Z, Wang W, Cao Y, *et al.* Expanding performance boundaries of open-source multimodal models with model, data, and test-time scaling. preprint. Posted online 2024. arXiv:2412.05271.
26. Wang W, Chen Z, Wang W, *et al.* Enhancing the reasoning ability of multimodal large language models via mixed preference optimization. preprint. Posted online 2024. arXiv:2411.10442.

Nicotine and Tobacco Research, 2026, 00, 1–6
https://doi.org/10.1093/ntr/ntag125
Original Investigation