HHS Public Access
Author manuscript
Lancet Digit Health. Author manuscript; available in PMC 2025 April 29.

Published in final edited form as:
Lancet Digit Health. 2024 July ; 6(7): e500–e506. doi:10.1016/S2589-7500(24)00094-3.

Automated cooling tower detection through deep learning for Legionnaires' disease outbreak investigations: a model development and validation study

Karen K Wong,
Centers for Disease Control and Prevention, Atlanta, GA, USA; University of California, Berkeley, CA, USA

Thaddeus Segura,
University of California, Berkeley, CA, USA

Gunnar Mein,
University of California, Berkeley, CA, USA

Jia Lu,
University of California, Berkeley, CA, USA

Elizabeth J Hannapel,
Centers for Disease Control and Prevention, Atlanta, GA, USA

Jasen M Kunz,
Centers for Disease Control and Prevention, Atlanta, GA, USA

Troy Ritter,
Centers for Disease Control and Prevention, Atlanta, GA, USA

Jessica C Smith,
Centers for Disease Control and Prevention, Atlanta, GA, USA

Alberto Todeschini,
University of California, Berkeley, CA, USA

Fred Nugen,
University of California, Berkeley, CA, USA

Chris Edens

This is an Open Access article under the CC BY-NC-ND 4.0 license.

Correspondence to: Dr Karen K Wong, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA, kwong.cdc@gmail.com.

Contributors

KKW and CE conceptualised the study. KKW, EJH, JCS, JMK, TR, and CE annotated and verified the raw training data. JL, GM, TS, and KKW synthesised additional training data. TS, GM, JL, KKW, FN, and AT contributed to the methods. CE, EJH, JCS, JMK, and TR provided user feedback. TS, GM, JL, and KKW wrote the code for the tool. EJH, JCS, CE, and KKW manually reviewed imagery for the speed test. KKW drafted the manuscript. All authors critically reviewed, edited, and approved the manuscript. KKW, JL, GM, and TS verified the raw data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

FN owns stock in Google and Amazon. All other authors declare no competing interests. The findings and conclusions of this study are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Centers for Disease Control and Prevention, Atlanta, GA, USA

Summary

Background—Cooling towers containing *Legionella* spp are a high-risk source of Legionnaires' disease outbreaks. Manually locating cooling towers from aerial imagery during outbreak investigations requires expertise, is labour intensive, and can be prone to errors. We aimed to train a deep learning computer vision model to automatically detect cooling towers that are aerially visible.

Methods—Between Jan 1 and 31, 2021, we extracted satellite view images of Philadelphia (PN, USA) and New York state (NY, USA) from Google Maps and annotated cooling towers to create training datasets. We augmented training data with synthetic data and model-assisted labelling of additional cities. Using 2051 images containing 7292 cooling towers, we trained a two-stage model using YOLOv5, a model that detects objects in images, and EfficientNet-b5, a model that classifies images. We assessed the primary outcomes of sensitivity and positive predictive value (PPV) of the model against manual labelling on test datasets of 548 images, including from two cities not seen in training (Boston [MA, USA] and Athens [GA, USA]). We compared the search speed of the model with that of manual searching by four epidemiologists.

Findings—The model identified visible cooling towers with 95·1% sensitivity (95% CI 94·0–96·1) and a PPV of 90·1% (95% CI 90·0–90·2) in New York City and Philadelphia. In Boston, sensitivity was 91·6% (89·2–93·7) and PPV was 80·8% (80·5–81·2). In Athens, sensitivity was 86·9% (75·8–94·2) and PPV was 85·5% (84·2–86·7). For an area of New York City encompassing 45 blocks (0·26 square miles), the model searched more than 600 times faster (7·6 s; 351 potential cooling towers identified) than did human investigators (mean 83·75 min [SD 29·5]; mean 310·8 cooling towers [42·2]).

Interpretation—The model could be used to accelerate investigation and source control during outbreaks of Legionnaires' disease through the identification of cooling towers from aerial imagery, potentially preventing additional disease spread. The model has already been used by public health teams for outbreak investigations and to initialise cooling tower registries, which are considered best practice for preventing and responding to outbreaks of Legionnaires' disease.

Funding—None.

Introduction

Legionnaires' disease is a severe pneumonia caused by *Legionella* bacteria with a case fatality ratio of 8–12%.[1] Nearly 9000 cases of Legionnaires' disease were reported in the USA in 2019, but the true number of cases might be up to 2·7 times higher than the reported value.[2,3] Cases in the USA are reported most frequently in the East North Central, Middle Atlantic, and New England US Census Bureau regional divisions.[4] Community outbreaks of Legionnaires' disease have been associated with cooling towers,[5–7] which cool the air and water systems of buildings through evaporative heat transfer. These towers, which are often located on the rooftops of buildings and have large fans, can harbour *Legionella* bacteria, especially when inadequately maintained.[8] The fans can aerosolise water containing *Legionella* spp, spreading the bacteria across large distances into the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

surrounding area.[5] Differences in environmental risk factors, including increased risks associated with inadequately maintained cooling towers, might contribute to racial and socioeconomic disparities in Legionnaires' disease incidence.[9]

Responding to a community outbreak of Legionnaires' disease requires the identification of cooling towers in close proximity to the outbreak; an environmental assessment; testing for *Legionella* spp; and cleaning, remediating, and re-testing of affected towers. It is crucial that a multidisciplinary outbreak response is coordinated as quickly as possible because an implicated cooling tower might continue to cause infections and deaths through the aerosolisation of contaminated water, especially in densely populated areas.[6] Cooling towers are often visible from aerial imagery, but identifying them requires time and expertise. This process can be labour intensive, and in dense urban areas, multiple buildings with one or more cooling towers might be located within the outbreak search zone.[7] A small block or census tract can take 2–3 h for experienced personnel to examine manually by visual inspection of online aerial imagery, and scanning a large city can take several months.[10] Some jurisdictions, such as New York City (NY, USA), maintain registries of cooling towers that help outbreak investigators to locate cooling towers quickly when a community cluster of Legionnaires' disease is identified. However, most cities in the USA do not maintain cooling tower registries and must rely on manual identification of cooling towers from aerial imagery during an outbreak investigation. Automating this task would shorten the time to cooling tower identification and remediation, potentially ending Legionnaires' disease outbreaks more quickly.

Computer vision refers to the use of computers to interpret digital images, usually to mimic a task that would traditionally require human visual interpretation. Computer vision models can accelerate manual processes, such as interpreting radiographs.[11,12] Certain models have been used to identify objects from satellite imagery, such as ore mines, wildfire smoke plumes, brick kilns, and cooling systems.[13–16] In this study, we aimed to train a computer vision model to automatically detect cooling towers that are visible from aerial imagery to improve and accelerate the response to and prevention of Legionnaires' disease outbreaks.

Methods

Model training data

The full dataset of images used to train models comprised: manually annotated images from New York state (including New York City, NY, USA) and Philadelphia (PA, USA); synthetically generated images; and images from additional locations annotated by a preliminary model that were manually verified (model-assisted labelling). Addresses of cooling towers were publicly available for New York City and New York state;[17,18] addresses of cooling towers from Philadelphia, verified in 2019, were obtained from the Philadelphia Department of Public Health. We required the geographical boundaries of individual cooling towers, which have higher geographical specificity than an address. To do this, between Jan 1 and 31, 2021, we first extracted satellite view image tiles of addresses with cooling towers from Google Maps (Google, Mountain View, CA, USA). Image tiles were 1280 × 1280 pixels and covered approximately 21 km^2 each, although resolution varied by location, image date, and zoom level. A team (KKW, CE, EJH, JCS, JMK, and TR) at the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Centers for Disease Control and Prevention (Atlanta, GA, USA) with expertise in cooling tower identification manually annotated all visible cooling towers within each image tile with a bounding box. For sets of multiple towers, each visible cooling tower was annotated separately.

To supplement the original training data with additional synthetic training data, we extracted images for Atlanta (GA, USA), Chicago (IL, USA), and Houston (TX, USA) with no visible cooling towers to use as background images on which to place cooling towers for synthetic data generation. These locations were selected for their visual diversity on aerial imagery, with the aim of improving the model's performance in a variety of landscapes. Using the cut-and-paste method,[19] we synthesised 119 additional training images by pasting between one and three cooling towers onto these background images. The cooling towers were blended with background images using Gaussian blurring of edges (radius 10 pixels; appendix p 3).

After training the model with registry-derived and synthetic data, we used the preliminary model to annotate additional training data. The model predictions were manually reviewed and adjusted to create verified training data for the following cities: Seattle (WA, USA), Las Vegas (NV, USA), Baton Rouge (LA, USA), and Irvine (CA, USA). The final training data contained 2051 images with 7292 labelled cooling towers (appendix p 1). The object detection model was trained using image tiles showing one or several buildings. The image classification model was trained on objects identified by the object detection model and cropped from the image tiles.

Test data comprised manually annotated images from New York state (including New York City) and Philadelphia that had been randomly excluded from the training dataset and two unseen locations (Boston [MA, USA] and Athens [GA, USA]). 378 images were included from New York state and Philadelphia (1700 cooling towers), 140 from Boston (610 cooling towers), and 30 from Athens (61 cooling towers).

No ethical approval was required for this study.

Development and validation of a two-stage model

The cooling tower identification tool used a two-stage deep learning model (figure 1). The first stage used YOLOv5 (Ultralytics, Los Angeles, CA, USA),[20] an open-source object detection model, to identify potential cooling towers from the image tile and assign a probability that the detected object was a cooling tower. The search area was divided into image tiles that overlapped by 5%, and each image tile was evaluated by the model for the presence of cooling towers. The first stage of the model was intended to be sensitive, and ranges were determined for low, intermediate, and high probability of detecting a cooling tower by examining the ratios of true positive to false positive detections across probability thresholds. Objects with low predicted probability of being a cooling tower were identified as not cooling towers. Objects with intermediate predicted probability of being a cooling tower were evaluated further in the second stage of the model. Objects with high predicted probability of being a cooling tower bypassed the second stage of the model and were identified as cooling towers.

Author Manuscript

The second stage used EfficientNet-b5, an image classification model.[21] Objects predicted by the YOLOv5 stage to have intermediate probability of being cooling towers were automatically cropped from the image tile using the coordinates outputted by YOLOv5, resized, and normalised. The dataset of cropped objects was divided into 70% training, 20% validation, and 10% test datasets for the EfficientNet-b5 model, which classified objects as cooling towers or not cooling towers. This stage was trained to distinguish true positives from false positives from the first stage, thus increasing the model's positive predictive value (PPV) for detecting cooling towers.

Both model stages were trained using transfer learning, whereby the models were pretrained on their task using a large dataset unrelated to cooling towers, then fine-tuned for the task of identifying cooling towers. YOLOv5 was trained to detect objects using the COCO 2017 object detection dataset, which contains more than 200 000 images and 80 object categories showing objects in their surrounding context.[22] EfficientNet-b5 was trained to classify images using ImageNet, a dataset of more than 3 million images grouped into thousands of classes.[23]

We evaluated sensitivity (recall), PPV (precision), and F1 (the harmonic mean of sensitivity and PPV) of the two-stage model by its ability to detect individual cooling towers—eg, for a single building with two cooling towers, detection of both towers would be required for perfect performance of the model. Since the process of locating cooling tower addresses for Legionnaires' disease investigation does not require highly precise locations of individual cooling towers within the bounds of an address, any overlap between the model's predicted bounding box for a cooling tower and a true bounding box was considered a positive detection. Detections that did not overlap with any true annotations were considered to be false positives. We used the PyTorch implementation for training and testing of both model stages. Training was conducted on Google CoLab (Google) with a Tesla V100-SXM2-16GB graphics processing unit (Tesla, Austin, TX, USA).

Comparison of search speeds

We compared the search time of the model on an Amazon Web Services g4dn.xlarge instance (Amazon, Seattle, WA, USA; on-demand rate US$0·526 per h as of Jan 23, 2024[24]) with that of four independent epidemiologists with expertise in the identification of cooling towers. A search area was defined that encompassed 45 dense urban city blocks (0·26 square miles) in New York City, the boundaries of which were 50th Street, 59th Street, 2nd Avenue, and 5th Avenue. Epidemiologists were instructed to conduct their manual search in the same way that they typically would for an outbreak investigation and annotate all cooling towers using Google My Maps (Google).

Role of the funding source

There was no funding source for this study.

Results

For the YOLOv5 object detection stage, we examined the distribution of prediction probabilities and evaluated detected objects with a minimum prediction probability of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

0·25 as potential cooling towers. Among 1700 cooling towers from New York City and Philadelphia test images, we compared the ratio of true positives to false positives across prediction probabilities and found that the ratio of true positives to false positives increased at probabilities of 0·65 or higher (figure 2). Thus, we defined cooling tower detection probability thresholds of lower than 0·25 as low, 0·25 to less than 0·65 as intermediate, and 0·65 or higher as high.

The YOLOv5 model identified 790 potential objects from New York City and Philadelphia image tiles that it classified as intermediate probability of being cooling towers. These objects comprised 563 true positive and 227 false positive detections and were used to fine-tune the EfficientNet-b5 classification model.

The baseline single-stage YOLOv5 model identified 1547 of 1700 annotated cooling towers (91·0% sensitivity [95% CI 89·5–92·3]) with a PPV of 90·5% (1547 of 1709; 95% CI 89·5–91·5) when tested on images of New York City and Philadelphia not seen during model training. The best-performing model, which was trained on expanded data from synthetic data generation and images from additional cities and incorporated the second-stage classification model, identified aerially visible cooling towers with 95·1% sensitivity (1582 of 1663; 94·0–96·1) and a PPV of 90·1% (1582 of 1756; 90·0–90·2) in New York and Philadelphia (F1 92·5%). Among unseen locations, the model identified 578 of 631 cooling towers (sensitivity 91·6% [89·2–93·7]) in Boston and 53 of 61 cooling towers (sensitivity 86·9% [75·8–94·2]) in Athens, and the PPV was 80·8% (578/715; 80·5–81·2) in Boston and 85·5% (53/62; 84·2–86·7) in Athens (figure 3). Specifications and evaluation metrics for the model are shown in the appendix (p 2).

False positive detections commonly included other types of air conditioning unit and objects with a radially symmetrical component, such as architectural features, water towers, and patio umbrellas (figure 4). However, the model also identified cooling towers that had been missed in the initial annotation of the data, making these erroneous false positive detections. Cooling towers missed by the model often had no obvious characteristics that would explain the model's inability to detect them; however, some missed towers were partially obscured, appeared blurry, or had a fan that did not have a radially symmetrical appearance due to shadows (figure 4).

The complete two-stage model completed the search of a predefined area of New York City in 7·6 s and identified 351 potential cooling towers. Independent manual searches of the same area conducted by four epidemiologists took a mean duration of 83 min and 45 s (SD 29·5 min, range 55–125 min) and identified a mean of 310·8 cooling towers (SD 42·2, range 269–375; table).

Discussion

We trained a computer vision model to identify cooling towers visible from aerial imagery, which could accelerate outbreak investigation and response for Legionnaires' disease. This model was trained to be highly sensitive, because it requires less effort for an investigator to verify a proposed cooling tower detection rather than identify a cooling tower from

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

an unannotated image. However, because of the second classifier stage of the model, the sensitive model achieved a PPV of 90·0%. The baseline model, trained only on images from New York and Philadelphia, had more than 90% sensitivity and PPV in those locations. We generated synthetic data and used model-assisted labelling of additional cities to increase the volume and variety of the model's training data. Expansion of the training data and addition of a secondary classification model to increase PPV improved the model compared with baseline performance.

Cooling tower registries remain the reference standard as an accurate record of cooling towers, since computer vision models cannot detect towers that are not visible from aerial imagery. The maintenance of such registries is considered to be one of the best practices in preventing and responding to Legionnaires' disease outbreaks caused by cooling towers,[25] and registering and monitoring of cooling towers was a key recommendation from a National Academies of Sciences, Engineering, and Medicine report[26] on *Legionella* because it provides a demonstrable public health benefit. The predictions of a computer vision model that automatically identifies cooling towers can initialise a cooling tower registry that can subsequently be verified manually; this work is under way in two US jurisdictions that have adapted our model for use with locally acquired imagery data.[27,28] The model can scale to detect cooling towers over large geographical areas, and it can be run repeatedly as updated aerial imagery becomes available. Where registries are not maintained, the model makes it possible for jurisdictions without expertise in cooling tower identification to locate potential cooling towers within a targeted geographical scope during an outbreak investigation within seconds.

Between April 1, 2021, and Feb 29, 2024, our model has assisted at least 24 Legionnaires' disease outbreak investigations across 12 US states, where it has eliminated hours of manual cooling tower identification. Although the model does not eliminate the need for human review, it accelerates the manual process of identifying cooling towers by allowing experts to review and validate model predictions quickly. In a direct comparison of search time for a small urban area, the model was more than 600 times faster than were human investigators; these time savings are likely to be even greater when applied to larger search tasks that need to account for human fatigue. By quickly scanning satellite imagery and exporting relevant location data, this tool allows subject matter experts to verify results quickly and focus on other investigation tasks. The code is publicly available and open-source, and jurisdictions with machine learning expertise and computing resources can customise and implement the model for Legionnaires' disease outbreak investigations and cooling tower registry creation. Adapting this model to create an application for wider deployment, training, and customisation would be helpful to jurisdictions without machine learning expertise. Multiple jurisdictions are adapting the model to use private image data to initialise cooling tower registries, reflecting the robustness of the object detection algorithm. There are also applications beyond public health; for example, Barth and colleagues trained a similar model to quantify cooling capacities from aerial imagery.[16]

Our study had certain limitations. The model is unable to identify cooling towers with side-mounted fans or cooling towers that might be located under a covering structure. It is possible that rare or new cooling tower designs will not be identified by the model. The test

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

performance of the model in New York state, including New York City, and Philadelphia might not be generalisable to other locations, since images from those locations were used predominantly in training and there might be overlap between training images and testing images at the periphery of the tiles. Although we aimed to create a visually diverse set of images for training of the model, it is possible that the model might not perform well in certain locations due to visual characteristics that are unfamiliar to the model, and the model has not been evaluated in locations outside of the USA. The sensitivity of the model was lowest in Athens, a location that was not included in the training data and that is visually distinct from the mostly dense urban areas used in model training. However, jurisdictions could fine-tune the model on imagery for specific geographical regions, which should improve performance. The model was sensitive to lighting conditions at the time of aerial photography. Slanted shadows, such as those cast by buildings in dense urban areas, affected its performance. Future work to expand the training data, a task that can be assisted by the model, could increase the visual diversity of the training data and is likely to improve model performance. The usefulness of the tool depends on how accurately available aerial images reflect the current environment; however, this is also a limitation for manual cooling tower identification. The choices made in evaluating model results reflect its intended use as a tool to identify addresses that host cooling towers, and other uses, such as counting and pinpointing the location of individual cooling towers, might benefit from tuning different evaluation metrics.

We report the development of a computer vision model to allow rapid detection of cooling towers during Legionnaire's disease outbreaks. This study shows that automated cooling tower detection could be a useful tool for outbreak investigators, and also underscores the potential for use of deep learning on aerial imagery to identify other sources of environmental exposures during outbreak investigations. The high accuracy of the model, the rapid performance gains from additional training data, and the low operating costs make this a tool that can be implemented immediately to accelerate outbreak responses for Legionnaires' disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Lauren Finn and Emily Scalzo, from the Philadelphia Department of Health, and Melissa Smith, Nina D Dutton, Jeri Anderson, and Edward Gregory, from the Centers for Disease Control and Prevention, for sharing the Philadelphia cooling tower data. Amazon Web Services provided cloud computing credits for educational coursework to authors KKW, TS, GM, and JL; Amazon Web Services had no role in the design or execution of the study.

Data sharing

Data on cooling tower locations are publicly available for New York City at https://data.cityofnewyork.us/Health/DOHMH-Cooling-Tower-Data/y4fw-iqfr and for New York state at https://health.data.ny.gov/Health/New-York-State-Cooling-Tower-Registry-Weekly-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Extra/unmf-baqa. Model code and weights are available at https://github.com/TowerScout/TowerScout.

References

1. Phin N, Parry-Ford F, Harrison T, et al. Epidemiology and clinical management of Legionnaires' disease. Lancet Infect Dis 2014; 14: 1011–21. [PubMed: 24970283]
2. Centers for Disease Control and Prevention. Legionella: history, burden, and trends. 2021. https://www.cdc.gov/legionella/about/history.html (accessed Jan 23, 2023).
3. Collier SA, Deng L, Adam EA, et al. Estimate of burden and direct healthcare cost of infectious waterborne disease in the United States. Emerg Infect Dis 2021; 27: 140–49. [PubMed: 33350905]
4. Barskey A, Lee S, Hannapel E, et al. Legionnaires' disease surveillance summary report, United States. 2022. https://www.cdc.gov/legionella/health-depts/surv-reporting/2018-19-surv-report-508.pdf (accessed Feb 26, 2024).
5. Walser SM, Gerstner DG, Brenner B, Höller C, Liebl B, Herr CEW. Assessing the environmental health relevance of cooling towers – a systematic review of legionellosis outbreaks. Int J Hyg Environ Health 2014; 217: 145–54. [PubMed: 24100053]
6. Jansà JM, Caylà JA, Ferrer D, et al. An outbreak of Legionnaires' disease in an inner city district: importance of the first 24 hours in the investigation. Int J Tuberc Lung Dis 2002; 6: 831–38. [PubMed: 12234140]
7. Weiss D, Boyd C, Rakeman JL, et al. A large community outbreak of Legionnaires' disease associated with a cooling tower in New York City, 2015. Public Health Rep 2017; 132: 241–50. [PubMed: 28141970]
8. Logan-Jackson AR, Rose JB. Water age effects on the occurrence and concentration of Legionella species in the distribution system, premise plumbing, and the cooling towers. Microorganisms 2021; 10: 81. [PubMed: 35056530]
9. Hunter CM, Salandy SW, Smith JC, Edens C, Hubbard B. Racial disparities in incidence of Legionnaires' disease and social determinants of health: a narrative review. Public Health Rep 2022; 137: 660–71. [PubMed: 34185609]
10. Fontenot J, Yu A, Rutschmann S, Jain S. Cooling Tower Identification 101: how to use satellite imagery to identify cooling towers. 2021. https://www.cdph.ca.gov/Programs/CID/DCDC/CDPHDocumentLibrary/CoolingTowerIDGuideProtocol.pdf (accessed March 6, 2023).
11. Vail TP. Deep learning dramatically reduces the work associated with image cataloguing and analysis: commentary on an article by Pouria Rouzrokh, MD, MPH, MHPE, et al.: "Applying deep learning to establish a total hip arthroplasty radiography registry. a stepwise approach". J Bone Joint Surg Am 2022; 104: e82. [PubMed: 36129678]
12. Rouzrokh P, Khosravi B, Johnson QJ, et al. Applying deep learning to establish a total hip arthroplasty radiography registry: a stepwise approach. J Bone Joint Surg Am 2022; 104: 1649–58. [PubMed: 35866648]
13. Balaniuk R, Isupova O, Reece S. Mining and tailings dam detection in satellite imagery using deep learning. Sensors (Basel) 2020; 20: 6936. [PubMed: 33291634]
14. Larsen A, Hanigan I, Reich BJ, et al. A deep learning approach to identify smoke plumes in satellite imagery in near-real time for health risk communication. J Expo Sci Environ Epidemiol 2021; 31: 170–76. [PubMed: 32719441]
15. Lee J, Brooks NR, Tajwar F, et al. Scalable deep learning to identify brick kilns and aid regulatory capacity. Proc Natl Acad Sci USA 2021; 118: e2018863118. [PubMed: 33888583]
16. Barth F, Schüppler S, Menberg K, Blum P. Estimating cooling capacities from aerial images using convolutional neural networks. Appl Energy 2023; 349: 121561.
17. NYC OpenData. NYC Cooling Tower registrations. https://data.cityofnewyork.us/Health/DOHMH-Cooling-Tower-Data/y4fw-iqfr (accessed Jan 23, 2023).
18. New York State Department of Health. New York State Cooling Tower Registry. https://health.data.ny.gov/Health/New-York-State-Cooling-Tower-Registry-Weekly-Extra/24a4-muw7/data (accessed Jan 23, 2023).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

19. Dwibedi D, Misra I, Hebert M. Cut, paste and learn: surprisingly easy synthesis for instance detection. https://ieeexplore.ieee.org/document/8237408 (accessed May 22, 2024).
20. Jocher G YOLOv5 byUltralytics (Version 7.0) [Computer software]. https://github.com/ultralytics/yolov5/blob/master/CITATION.cff. 2020. 10.5281/zenodo.3908559 (accessed Feb 29, 2024).
21. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. arXiv 2019; published online May 28. http://arxiv.org/abs/1905.11946 (preprint).
22. Lin T-Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context. 2014. 10.1007/978-3-319-10602-1_48 (accessed Feb 29, 2024).
23. Deng J, Dong W, Socher R, Li L-J, Li K, Li F-F. ImageNet: a large-scale hierarchical image database. https://ieeexplore.ieee.org/document/5206848 (accessed Feb 29, 2024).
24. Amazon Web Services. Amazon EC2 on-demand pricing. 2024. https://aws.amazon.com/ec2/pricing/on-demand/ (accessed Jan 22, 2024).
25. Paschke A, Schaible UE, Hein W. Legionella transmission through cooling towers: towards better control and research of a neglected pathogen. Lancet Respir Med 2019; 7: 378–80. [PubMed: 30930109]
26. National Academies of Sciences, Engineering, and Medicine. Management of Legionella in water systems. Washington, DC: National Academies Press, 2019.
27. County of Los Angeles GIS Day. Using aerial imagery and machine learning to identify cooling towers in LA county. 2022. https://www.youtube.com/watch?v=YGxJl0jB3uM (accessed Jan 21, 2024).
28. Neemann E Results update: detecting cooling towers with computer vision tools. Utah Geospatial Resource Center, 2023. https://gis.utah.gov/cooling-tower-update/ (accessed Feb 26, 2024).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Research in context

Evidence before this study

During community outbreaks of Legionnaires' disease, identification of cooling towers as potential sources of Legionella spp typically relies on manual identification of cooling towers from aerial imagery, which requires time and expertise, or from cooling tower registries, which are often not available. Manually identifying towers delays the time to cooling tower remediation, a necessary intervention to prevent additional infections. We searched PubMed and Google Scholar from database inception to Feb 24, 2024, without language restrictions, for any publications using the search term "cooling tower" in combination with "aerial" or "satellite". We excluded results related to power plant cooling towers. Several reports described Legionnaires' disease outbreak investigations and mentioned the identification of cooling towers from aerial or satellite imagery through manual inspection, and one protocol provided a stepby-step guide to the manual identification of cooling towers on satellite images. One study used deep learning to detect aircooled condensers and cooling towers from aerial imagery to quantify cooling capacities, comparing some of the results with those available for the model presented in this study.

Added value of this study

We developed an accurate and scalable computer vision model using deep learning that automatically detects cooling towers from aerial imagery. We report sensitivity and specificity of the model, setting a benchmark for future work, and we estimate time savings from the model compared with manual cooling tower identification.

Implications of all the available evidence

Deep learning models can accelerate public health tasks that have traditionally relied on manual review by experts. This cooling tower detection tool is specifically designed to augment the workflows of Legionnaires' disease investigators. Although automated cooling tower detection with computer vision models does not eliminate the need for human review, it is a cost-effective and time-saving method of accelerating investigations following outbreaks of Legionnaires' disease and can enable the maintenance of cooling tower registries for future outbreak investigations and epidemiological studies.

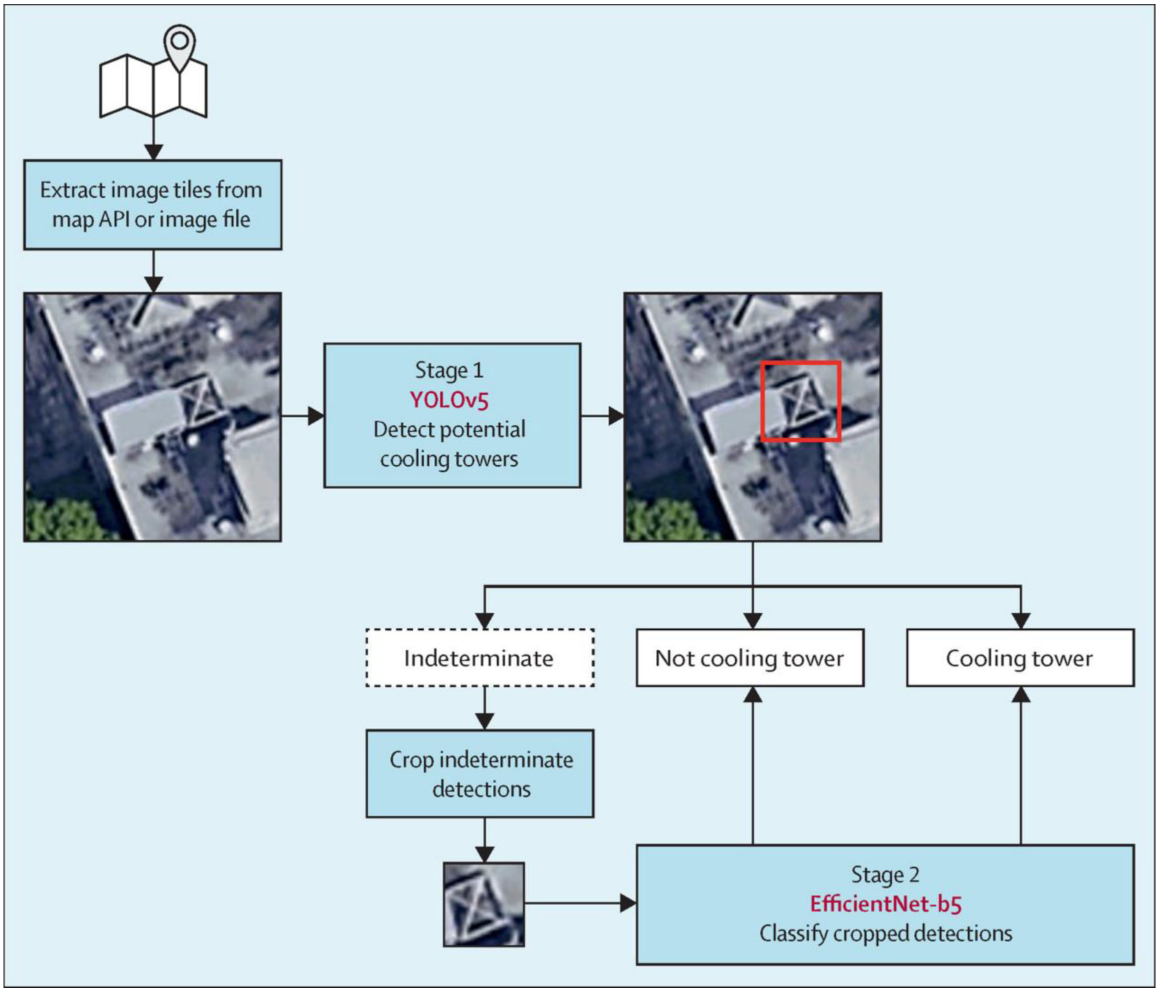

Figure 1: Two-stage model

The two-stage model uses YOLOv5, an open-source object detection model, and EfficientNet-b5, a model that classifies images, to identify cooling towers on satellite view images. The red rectangle shows a potential cooling tower identified by YOLOv5 during stage 1. If the object has indeterminate probability of being a cooling tower, it is cropped and evaluated by EfficientNet-b5 during stage 2. API=application programming interface. Images extracted between Jan 1 and 31, 2021. Maps data: Google, Imagery © 2021 Bluesky, CNES/Airbus, Maxar Technologies, Sanborn.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

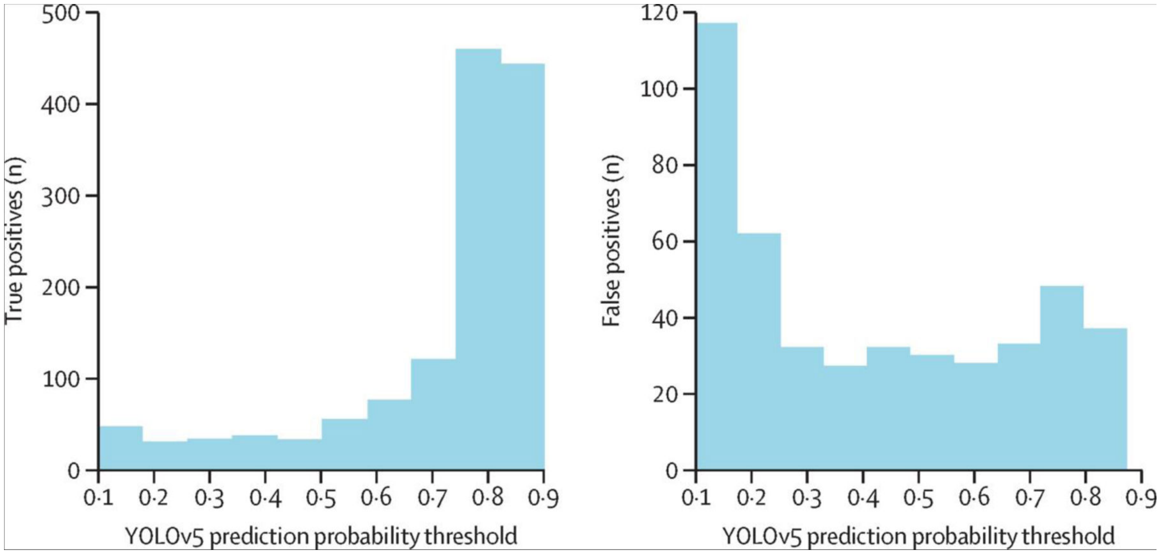

Figure 2: Distribution of true positives and false positives by YOLOv$_5$ prediction probability threshold

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

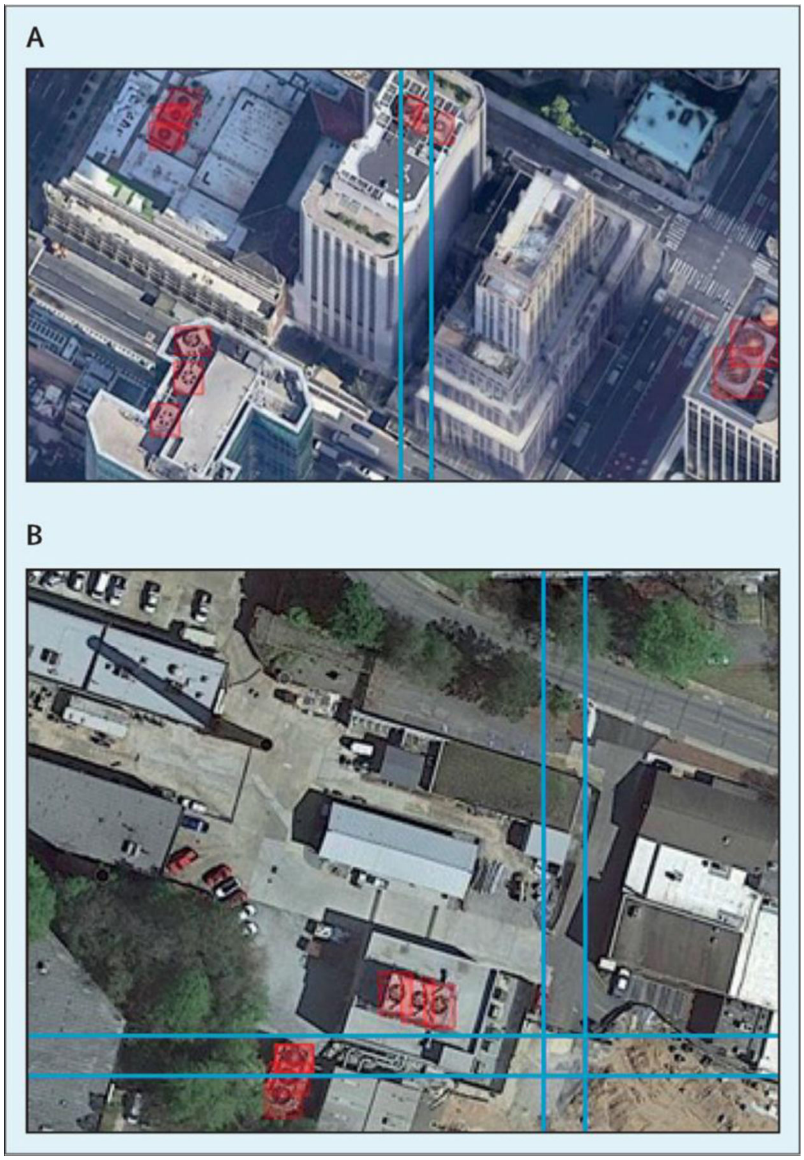

Figure 3: Examples of predicted cooling towers in New York City and Athens

(A) Cooling towers predicted by the model in New York City, a city used in training. Google, Imagery © 2021 Bluesky, CNES/Airbus, Maxar Technologies, Sanborn. (B) Cooling towers predicted by the model in Athens (GA, USA), a city unseen in training. Google, Imagery © 2021 Maxar Technologies, U.S. Geological Survey. Red rectangles show predicted cooling towers. Blue lines that extend across the full height and width of the images show boundaries of overlapping image tiles in the search area. Images extracted between Jan 1 and 31, 2021.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

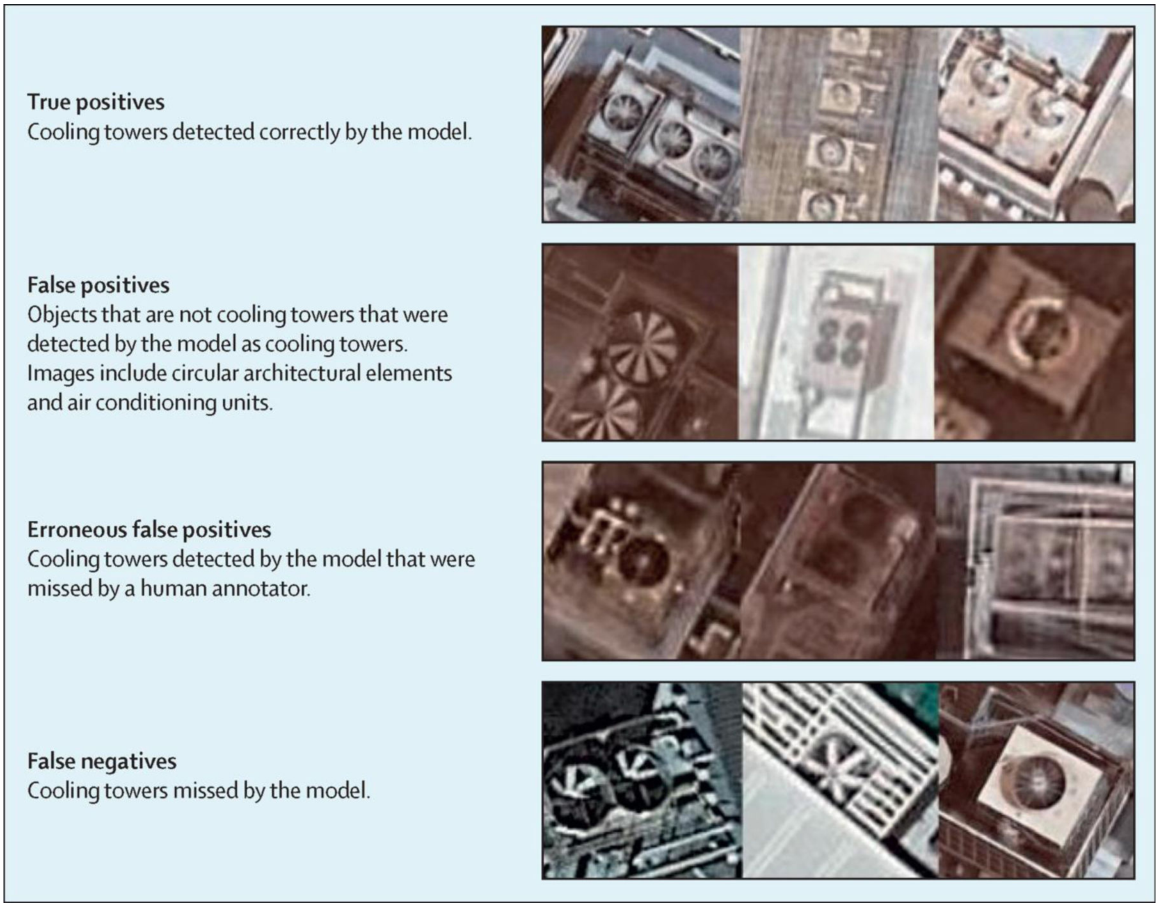

Figure 4: Examples of model predictions

Images extracted between Jan 1 and 31, 2021. Maps data: Google, Imagery © 2021 Bluesky, CNES/Airbus, Maxar Technologies, Sanborn.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Table:

Comparison of human and model search times and number of cooling towers identified

	Model	Epidemiologist 1	Epidemiologist 2	Epidemiologist 3	Epidemiologist 4
Search time	7·6 s	55 min	75 min	80 min	125 min
Cooling towers detected, n	351	302	297	269	375

The test was done for images covering an area of 45 blocks (0·26 square miles) in New York City (NY, USA).