How to Cite:
Sundhar, S. V., Selvan, I. V. N., Prasad, S. S., Hariharan, S., & Agarwal, R. (2022). Identifying highly vulnerable mosquito breeding sites using machine learning and drone based aerial survey. *International Journal of Health Sciences*, *6*(S3), 209–221. https://doi.org/10.53730/ijhs.v6nS3.5263

Identifying Highly Vulnerable Mosquito Breeding Sites Using Machine Learning and Drone Based Aerial Survey

Shyam Sundhar V.
Department of Geography, University of Madras, Chennai, India
Email: Shyamsundhar35@gmail.com

Indhiya Selvan V. N
Department of Geography, University of Madras, Chennai, India
Email: indhiyaselvan@gmail.com

S. Sanjeevi Prasad
Department of Geography, University of Madras, Chennai, India
Email: Sanjeevi.geo.unom@gmail.com

Hariharan S.
M. Tech Geoinformatics, Department of Geography, University of Madras, Chennai, India
Email: hariharansnair77@gmail.com

Rashi Agarwal
Department of CSE, UIET, Chhatrapati Shahu Ji Maharaj University, Kanpur, India
Email: dr.rashiagrawal@gmail.com

Abstract---This study deals with Drone based Aerial Survey in analyzing and identification of highly vulnerable mosquito breeding sites at Buckingham canal Chepauk, Chennai. It is small section of Buckingham canal studied using the drone for capturing the images and furtherly images were processed for identifying the sites prone to breeding of mosquitos. Approach used here of capturing images and classifying the vulnerable sites, using a machine learning approach in for extracting features using the algorithm and later generating a Support Vector Machine to train and classify the images. Tensorflow object detection algorithm was used to detect the object and also generate the probability level. Tensorflow based supervised classification involves stacking multiple layers of neural network for a classification. Methods like back propagation invoked in the neural

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
***Corresponding author**: Sundhar, S. V.; Email: Shyamsundhar35@gmail.com*
Manuscript submitted: 18 Nov 2021, Manuscript revised: 09 Feb 2022, Accepted for publication: 27 March 2022

networks ensures the classification accuracy are increased. In this algorithm Single shot multi box detector (SSD) has been used which provides fast detection. Single shot multi box detector (SSD) method is based on a feed-forward convolutional network, followed by a non-maximum suppression step to produce final detections. The accuracy constraint of 70% was kept for qualifying as a potential site to negate ambiguities arising due to processing errors. Accuracy level was tested and found out to be 70% on a random test. Potential mosquito breeding sites have been identified and marked as high low and medium probability. The banks and region of water has sources identified as a high probability region especially due to poor waste disposal by the local inhabitants. The trained algorithm can identify sites even huge areas without much of human interference.

***Keywords*---**drone survey, GIS, health geography, machine learning, remote sensing.

Introduction

As per World Health Organization (WHO), mosquito is the most dreadful as it is the common causing agent for various life-threatening diseases like malaria, dengue and yellow fever, which has resulted in loss of life of millions of people across the globe. Mosquitos are also responsible for the transmission of other diseases like lymphatic filariasis and Japanese encephalitis. Malaria is found to be prevalent extensively in more than 90 countries, that is nearly forty per cent of the population liable to be infected. As the health and resource of human beings get degraded this has a cascading net effect in resulting of hampering social and economic wellbeing. As per the WHO reports nearly 500 million cases happen annually, which is 90% of them occur in Africa, and there are close to 2.7 million deaths annually. Is the deadliest mosquito transmitted disease with 2500 million exposed to infection and 20 are affected [16].Of all the mosquito borne disease malaria and dengue is the most life threatening and dreaded. Currently, around 100 countries are reported to have active transmission of Malaria and Dengue. In 2012, there were an estimated 207 million malaria cases, with more than one case per 1000 population in high-risk areas and 627000 deaths due to malaria. Dengue fever afflicts an estimated 100 million people per year [12]. The numbers have substantially increased to over 100 in 2012 [17]. Millennium Development Goal (MDG) has set the year 2015 as the time for stemming the rot and set the negative growth [13]. This target can only have achieved by accelerated implementation of the integrated vector management, which is a low-cost method that employs simultaneous implementation of chemical, biological and mechanical methods.

The word “drone” was derived from the name of ‘Male bee’. A drone is developed from different ductile, light and strong materials in to provide agility at a low mass. Drones can be fitted with LIDAR sensor, Navigation sensors, Various payload such as small tanks for pesticide and insecticide applicators, and so on. Drones are available in various configurations and each to suit various applications. The variants can range from launching manually to totally auto controlled. Drones can be simple to complex in construction. Drones are used widely and there's no limitation on their applications. Drones have been found to be useful in various spheres of work. In the last five years, drones have witnessed drastic growth and are in huge demand in India. Apart from being mostly used by the Law Enforcement Agencies (LEAs), drones are also specialized to locate victims during natural disasters. The Indian Railways has been using drones for inspecting and tracking progress of its projects [14]. It is also used for monitoring gas transmission lines by government owned largest natural gas processing and distribution company.

Since slums lack even the basic water connection, the inmates store water in containers which they mostly leave uncovered, becoming a breeding ground for mosquitoes. For the last few decades malaria has become endemic to Chennai. Close to 70 percent of malaria cases in Tamil Nadu are from Chennai alone [18]. The Anopheles stephensi mosquito, that is responsible for the transmission of malaria, breeds in clear water. Many breeding sites identified in 2011, such as wells, open over-head-tanks and sumps have been the cause for malaria breeding and transmission [18].

Materials and Methods

The geotagged images are resized. The training is carried out with images by resizing the images due to constraint of computer processing capabilities. Bulk resize of images are carried out online. BRExif extractor is used to recover the metadata from the images. The Tensor flow software is installed with the dependencies. Object detection algorithm is used in supervised Learning. The features are extracted. This feature is used to train the classifiers automatically. The images are used to train the software "Fig. 2". The images are tested and checked for accuracy. In case the accuracy is less than the threshold retrains the classifier. The trained images are used to classify the test images. Once the desired accuracy greater than 70% the images and the results are converted into a geojson file for better visualization. Tensor board is used for analyzing the metrics.

Table 1
Datasets

		Datasets	
Sl.No	*Data*	*Source*	*Scale / Resolution* (GSD)
1.	373 Geotagged Images covering 1344.705 sq. meters	Drones used to collect images of study area. (20MP) camera, focal length =24mm	0.24 Pixel / Cm 8 MB/Image

Tensor Flow is a framework developed by Google for performing Deep Learning models. Deep Learning is a subset of machine learning model that incorporates multi- layer neural networks. It is a open source library that can run on all computers and smartphones. It allows for instant creation of trained production models. The various dependencies used for completion of this project includes Python, Tensorboard and Wheel. ArcGIS was used for designing, managing solution through involvement and specialization of geospatial knowledge. It was also utilized for compiling, collaborating and production of map. ArcGIS provides an infrastructure for making maps using geographic information available. ArcMap is used for cartographic and editing tasks, as well as for spatial-based analysis. ArcCatalog application which is a part of ArcGIS is used to maintain Geospatial data and its corresponding metadata. It makes it easy to integrate relational databases, files, ArcMap documents, etc. The images, shapefiles, databases are categorized into separate folders and stored using ArcCatalog to avoid confusion and for effective execution and analysis.

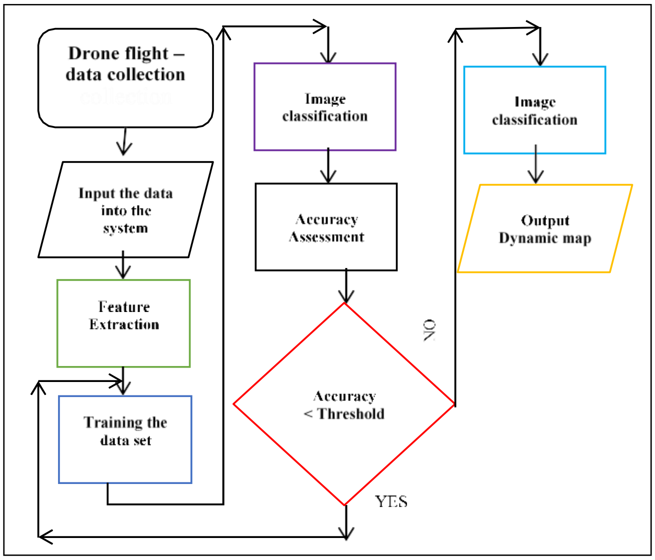

Figure 1. Methodology Flow Chart

Results and Discussions

Calibration
ESC Calibration

Electronic Speed Controllers (ESC) control the spinning of the motors at the speed requested by the autopilot. Almost all ESCs must be calibrated so that they record the minimum and maximum pwm values that the flight controller will send. ESCs emit musical tones, the regular number of beeps indicating your battery's cell count (i.e. 3 for 3S, 4 for 4S) and then an additional two beeps to indicate that the maximum throttle has been captured, and a long tone to indicate that the minimum throttle has been captured to complete the calibration.

Flight Control Calibration

RC transmitters are used to control vehicle movement and orientation; Mission planner software is used for the Radio control calibration. Selection of appropriate com port and 115200 as the band rate. Radio Control calibration is done to select the channels of operation. In this the controls of the transmitter channel is mapped to relevant action like throttle, fpitch. Yaw, flight mode, optional. Flight mode selection was done to select the various flight modes available like Stabilize, Alt Hold, Loiter, RTL (Return-to- Launch), Auto etc.

The compass is calibrated to a set of data points to correctly orient it to the earth's magnetic field. The Initial setup in mission planner is done first, followed by the mandatory hardware and compass. During the calibration, the drone is rotated in circular motions in order to align the system with the Earth's magnetic North. The drone is rotated in various ways to collect different array of data points around it. On successful completion, the image on Mission planner will glow with all the points that was collected and will commence to build set of data points around your drone.

Code for integrating raspberry PI with camera

```
>>import sys, os, time

>>image_num=1

>>print ("ready to take pic")

>>while True:

strImage=str(image_num)
os.system("fswebcam/home/pi/Mainproject/image"+strImage+".jpg")
>>image_num=image_num+1

>>print ("Captured")

>> time.sleep(1)
```

Image capture

The geotagged images captured are stored onboard. The same is downloaded and further processed. Average size of each image is approximately 8MB. This needs to be reduced for further processing. The reduced im- age size is used to train the object detection deep learning algorithm based on TensorFlow API ".

Resized image

The RGB images received from the drone are beyond the processing capability of the object identification algorithm. Hence, they needed to be resized using the Bulk Resize software the image sizes are reduced to approx. 800kb.A reduction of 1/10 of the overall image size.

Processed images

The images were trained using the convolutional neural networks. Initially the network was trained using 48 images and 500 steps. The test images were restricted to 5 images. As the accuracy was found to be below 20% the number of training images was increased to 100 images and the number of steps was increased to 2000. The accuracy was increased and there were no false positives but there was instance were the algorithm failed to detect a vulnerable site. Using a 12 GB RAM the system converged in 12 hours of processing. The final classified image sample is given below “Fig.2”.

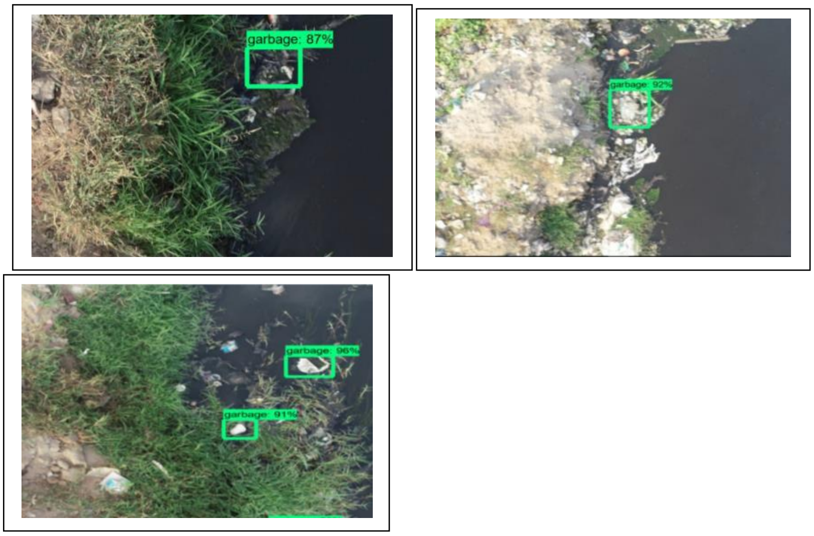

Figure 2. Methodology Flow Chart

Metrics and inference

The summary of the metrics for processing indicates the total losses when classifying the images trained at lesser number of steps (500) is much higher when compared to the training done at 2000 steps “Tab. 2”. It shows the losses decrease appreciably with increase in number of steps. Classification loss pertains to the error level in prediction and localization refers to the positioning the box correctly with fine adjustments. The clone loss refers to the loss occurring while training the images over multiple GPU (graphic processor unit). Since in this case only one GPU was used the clone loss is same as the total loss. A way to obtain this is to add a regularization term to the loss function. In this case the training with 500 steps yielded a regularization loss of 0.548 and the training with 2000 steps has a loss of 0.5468. Regularization loss is an additional loss generated by the regularization function. Since this adds to the total cost function it should be less. On analysis the total loss it is observed that it is preferable to train at higher number of steps.

Table 2
Tensorflow loss Metrics

Steps	Datasets		
	Classification loss	Localization Loss	Total Loss
500	5.0	2.0	7.0
2000	4.2	1.0	5.5

LabelImg

LabelImg is a graphical image annotation tool and label object bounding boxes in images. This has been used to annotate the training sets. The extracted label samples are listed below.

Table 3
Input Annotation

filename	width	height	class	xmin	ymin	xmax	ymax
1	1642	1094	breeding site	923	810	1064	1009
2	1642	1094	breeding site	1001	468	1113	617
3	1642	1094	breeding site	972	622	1106	795
4	1642	1094	breeding site	945	363	1032	490
5	1642	1094	breeding site	844	537	1032	793
6	1642	1094	breeding site	937	178	1040	346
7	1642	1094	breeding site	836	289	937	506
8	1642	1094	breeding site	112	755	224	869
9	1642	1094	breeding site	775	122	897	222
10	1642	1094	breeding site	99	291	214	338
11	1642	1094	breeding site	127	374	200	477
12	1642	1094	breeding site	60	501	175	601
13	1642	1094	breeding site	40	149	137	275
14	1642	1094	breeding site	747	293	777	328
15	1642	1094	breeding site	338	406	384	474
16	1642	1094	breeding site	5	283	44	322
17	1642	1094	breeding site	322	607	373	664
18	1642	1094	breeding site	62	547	216	662
19	1642	1094	breeding site	66	430	176	511

Table 4
Output Annotation

filename	width	height	class	xmin	ymin	xmax	ymax
1	1642	1094	breeding site	923	810	1064	1009
2	1642	1094	breeding site	1001	468	1113	617
3	1642	1094	breeding site	972	622	1106	795
4	1642	1094	breeding site	945	363	1032	490
5	1642	1094	breeding site	844	537	1032	793
6	1642	1094	breeding site	937	178	1040	346

7	1642	1094	breeding site	836	289	937	506
8	1642	1094	breeding site	112	755	224	869
9	1642	1094	breeding site	775	122	897	222
10	1642	1094	breeding site	99	291	214	338
11	1642	1094	breeding site	127	374	200	477
12	1642	1094	breeding site	60	501	175	601
13	1642	1094	breeding site	40	149	137	275
14	1642	1094	breeding site	747	293	777	328
15	1642	1094	breeding site	338	406	384	474
16	1642	1094	breeding site	5	283	44	322
17	1642	1094	breeding site	322	607	373	664

Probability frequency distribution

The probability percentage “Tab. 5” for each vulnerable location identified as per the algorithm is listed below. The values are classified and tabulated for visualization “Fig. 3”.

Table 5
Vulnerable sites

Sl. No	Latitude	Longitude	Probability
0	13.06612	80.28197	82
1	13.06612	80.28197	86
2	13.06611	80.28193	90
3	13.06606	80.28191	85
4	13.06619	80.28201	73
5	13.0662	80.28203	71
6	13.06621	80.28204	91
7	13.06622	80.28204	98
8	13.06624	80.28206	81
9	13.06629	80.28212	90
10	13.06628	80.28211	92
11	13.06626	80.2821	75
12	13.06625	80.28209	82
13	13.06618	80.28201	98
14	13.06606	80.28192	78
15	13.06604	80.28189	98
16	13.06603	80.28186	73
17	13.06602	80.28184	76
18	13.06602	80.28183	80
19	13.066	80.28182	86
20	13.06599	80.28183	80
21	13.06598	80.28182	76
22	13.06597	80.28183	89
23	13.06596	80.28182	82
24	13.06595	80.28181	82
25	13.06593	80.2818	92
26	13.06592	80.28172	91
27	13.06591	80.28171	82

28	13.06591	80.28172	84
29	13.0659	80.2817	97
30	13.0659	80.28169	91
31	13.06589	80.28168	93
32	13.06589	80.28167	95
33	13.0658	80.28165	88
34	13.06583	80.28168	76
35	13.06584	80.28172	82
36	13.06583	80.28173	97
37	13.06581	80.28171	72
38	13.06586	80.28176	83
39	13.06588	80.28178	90
40	13.06589	80.28179	71
41	13.0659	80.28179	81
42	13.06591	80.2818	86
43	13.06591	80.28181	77
44	13.06593	80.28183	95
45	13.06592	80.28171	84
46	13.06595	80.28175	76
47	13.06606	80.28193	74
48	13.06607	80.28193	81
49	13.06609	80.28196	97
50	13.06612	80.28198	98
51	13.06603	80.28185	78
52	13.06605	80.28188	78
53	13.06609	80.28195	79
54	13.06611	80.28198	99
55	13.06617	80.28203	95
56	13.06618	80.28202	95
57	13.06617	80.282	94
58	13.06616	80.282	88
59	13.06602	80.28181	75
60	13.06597	80.28181	70
61	13.06588	80.28168	89
62	13.06587	80.28168	80
63	13.06587	80.28166	80
64	13.06586	80.28165	73
65	13.06581	80.28169	98
66	13.06582	80.28172	90
67	13.06585	80.28176	73
68	13.06586	80.28178	77
69	13.06591	80.28182	94
70	13.06595	80.28187	70
71	13.06603	80.28195	93
72	13.0661	80.282	85
73	13.06622	80.28208	82
74	13.06619	80.28204	89

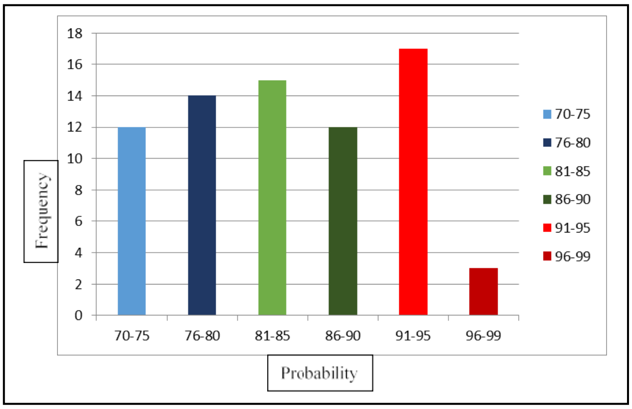

Figure 3. Probability Frequency distribution

Potential mosquito breeding sites have been identified and marked as high, medium and low probability “Fig. 8”. The sites have been classified 70% – 80%, 81% – 80%, 91% -99% of probability as high, medium and low probability respectively. The probability detection from the drone images infers that the banks and region of water has sources identified as a high probability region especially due to poor waste disposal by the local inhabitants. The garbage also can lead to clogging of the canal and disturb the natural flow of water, due to stagnation of water it creates ideal sites for breeding of mosquitoes and be a cause for health-related issues.

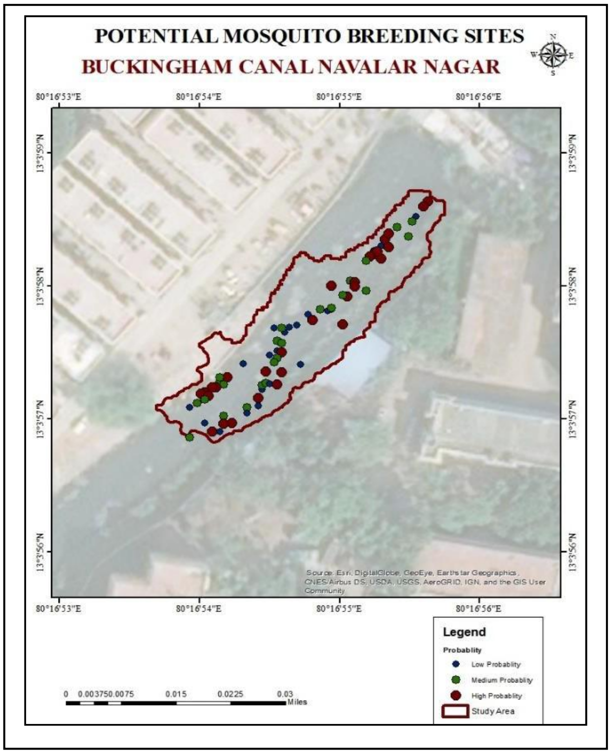

Figure 4. Vulnerability Profile of Study Area

The approach used here of capturing images and classifying the vulnerable sites was similarly done by Suduwella et.al (2017) using a machine learning approach in for extracting features using the HOG algorithm and later generating a Support Vector Machine to train and classify the imag es. The methodology adopted in this research work is similar to it was done using Tensor flow object detection algorithm. The algorithm is able to detect the object and also generate the probability level which is not available in the methods used by Suduwella et.al (2017) [5]. The vulnerable sites identified are classified on the basis of probability as high, medium and low. The majority of the areas are in the medium probability region. The sites are plotted in the Geojson map for better visualization. This is an open standard format for displaying the spatial and attributes data with various base maps. Topology cannot be included which needs TopoJSON.

The accuracy constraint of 70% was kept for qualifying as a potential site to negate ambiguities arising due to processing errors. Accuracy level was tested and found out to be 70% on a random test. Tensorflow based supervised classification involves stacking multiple layers of neural network for a classification. Methods like back propagation invoked in the neural networks ensures the classification accuracy are increased. In this algorithm Single shot multi box detector (SSD) has been used which provides fast detection. SSD method is based on a feed-forward convolutional network, followed by a non-maximum suppression step to produce final detections. A feed -forward network defines a mapping $y = f(x; \theta)$ and learns the value of the parameters θ that result in the best function approximation. Unlike recurrent networks there are no loops. Non maximum suppression is used to provide single bounding box per object identified and prevent the formation of multiple boxes around the detected object due to inability of the detection algorithm to localize the region of interest [15].

Conclusion

Over the last two decades, the scientific and technological progress has been substantial. This growth has been made by paying a huge price in terms of environmental degradation. The waste generated by various man-made activities get indiscriminately disposed of untreated. The of treating waste is only understood when the negative impacts of these activities tend to affect the health and lives of humans. These kind of garbage and waste piles tend to be the ideal breeding sites for the mosquitos. There needs to be an efficient system for identifying such sites and taking remedial actions. The challenge occurs in identifying such sites in a vast geographic terrain with minimal manpower. So, the existing modern technologies need to be leveraged to solve the problem.

In the present study area, a small section of Buckingham canal was studied using the drone for capturing the images and the images processed for identifying the sites prone to breeding of mosquitos. The points have been identified and marked as per the probability of occurrence. It is observed that high number of sites is identified especially along the banks of the canal. Since the area is small it can be visibly observed but when the area is large it is not possible to identify the potential sites only by observation. Hence a machine learning approach needs to be done to train and identify sites. The trained algorithm can identify sites even huge areas without much of human interference. The high-resolution satellite images are available with resolution of .3 meters but the same is very costly and is subject to atmospheric absorption and cloud cover distortions. The image downloaded from drones are of high resolution and is managed by the user and hence cost effective. The user can take multiple data acquisitions without incurring additional cost.

- Awareness among the residents about the ill effects of mosquito breeding disposal
- Municipal authorities to allot additional resources to such highly vulnerable mosquito breeding locations
- Garbage arrestors to be constructed to prevent the garages from draining into the sea and creating more vulner- able mosquito breeding locations.

Acknowledgments

We are grateful to two anonymous reviewers for their valuable comments on the earlier version of this paper.

References

1. Amarasinghe, A., Suduwella, C., Niroshan, L., Elvitigala, C., De Zoysa, K., & Keppetiyagama, C. (2017, September). Suppressing dengue via a drone system. In 2017 Seventeenth International Conference on Advances in ICT for Emerging Regions (ICTer) (pp. 1-7). IEEE.
2. Rani, A., Gupta, A., Nagpal, B. N., & Mehta, S. S. (2018). Mosquito borne diseases and sanitation in Ghaziabad district, Uttar Pradesh, India. Int J Mosquito Res, 5(5), 25-30.
3. Agarwal, A., Chaudhuri, U., Chaudhuri, S., & Seetharaman, G. (2014, July). Detection of potential mosquito breeding sites based on community sourced geotagged images. In Geospatial InfoFusion and Video Analytics IV; and Motion Imagery for ISR and Situational Awareness II (Vol. 9089, p. 90890M). International Society for Optics and Photonics.
4. Annadurai, K., Danasekaran, R., Mani, G., & Ramasamy, J. (2015). Mosquito menace: A major threat in modern era. Medical Journal of Dr. DY Patil University, 8(3), 414-414.
5. Amarasinghe, A., Suduwella, C., Elvitigala, C., Niroshan, L., Amaraweera, R. J., Gunawardana, K., ... & Keppetiyagama, C. (2017, November). A machine learning approach for identifying mosquito breeding sites via drone images. In Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems (pp. 1-2).
6. Kim, B. K., Kang, H. S., & Park, S. O. (2016). Drone classification using convolutional neural networks with merged Doppler images. IEEE Geoscience and Remote Sensing Letters, 14(1), 38-42.
7. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91-110.
8. Sreeram, S., & Shanmugam, L. (2018, June). Autonomous robotic system based environmental assessment and dengue hot-spot identification. In 2018 IEEE International Conference on Environment and Electrical Engineering

and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe) (pp. 1-6). IEEE.

9. Mehra, M., Bagri, A., Jiang, X., & Ortiz, J. (2016, June). Image analysis for identifying mosquito breeding grounds. In 2016 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops) (pp. 1-6). IEEE.
10. Renwick, J. D., Klein, L. J., & Hamann, H. F. (2016, December). Drone-based reconstruction for 3D geospatial data processing. In 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT) (pp. 729-734). IEEE.
11. Suduwella, C., Amarasinghe, A., Niroshan, L., Elvitigala, C., De Zoysa, K., & Keppetiyagama, C. (2017, June). Identifying mosquito breeding sites via drone images. In Proceedings of the 3rd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications (pp. 27-30).
12. “Handbook for Integrated Vector Management.” World Health Organization. World Health Organization. Accessed January 15, 2022.http://apps.who.int/iris/bitstream/handle/10665/
13. Vo, X. T., & Jo, K. H. (2021). Accurate Bounding Box Prediction for Single-Shot Object Detection. IEEE Transactions on Industrial Informatics.
14. “Executive Summary.” World Health Organization. World Health Organization, July 29, 2013. https://www.who.int/whr/1996/media_centre/executive_summary1/en/.
15. Kumar, D. S., Andimuthu, R., Rajan, R., & Venkatesan, M. S. (2014). Spatial trend, environmental and socioeconomic factors associated with malaria prevalence in Chennai. Malaria Journal, 13(1), 1-9.