Crowd Detection

Dr. Shripad Bhatlawande

shripad.bhatlawande@vit.edu

Nihar Chaudhari

nihar.chaudhari19@vit.edu

Unmesh Chaudhari

unmesh.chaudhari19@vit.edu

Deven Nalawade

deven.nalawade20@vit.edu

Department of Electronics and Telecommunication, Vishwakarma Institute of Technology, Pune, India, 411038

**Abstract- For visually impaired people, having their own independent mobility is a major challenge making it difficult for them to recognize and avoid crowded places. Identification of crowds and its analysis has been receiving a lot of attention recently, which has led to the development of a number of solutions for the same. Some of them work poorly whereas others are complex and expensive. Focusing over this issue, a system is proposed to assist the blind people to detect crowd.The proposed system includes the use of image preprocessing, SIFT feature extraction, normalization, dimensionality reduction, classification algorithms over the dataset along with performance evaluation metrics like accuracy, precision, recall and F-1 score to evaluate the model. The model has been trained using Decision tree, Random forest, SVM(Support Vector Machine), K Nearest Neighbour and Logistic Regression. The Random forest was able to detect crowded places with an accuracy score of 83.22% thus aiding the blind.**

**Keywords: Computer vision, Crowd detection, Machine learning.**

1. **Introduction**

Visual impairment or vision loss relates to complete or partial blindness. Though it does not directly cause death, it has a significant influence on the quality of life of affected individuals as it creates disability that lasts for lifetime. The current going covid-19 situation is a major issue. It is an individual’s responsibility to be safe over places. People with fully functional organs could easily manage to survive around the places as one could classify between a crowded and non crowded place just by looking around. But for a visually impaired person, it is difficult to understand the present surroundings as well as if it is safe to be around. In light of this circumstance, a method is being developed to assist visually impaired people in classifying their surroundings as crowded or non-crowded.

In this paper, the topics are arranged as follows: Literature survey is described in Section II. Section III describes methodology where the information about the algorithm followed, feature descriptors & classifiers is specified. Results, Conclusion and References are given in section IV, V and VI respectively.

1. **Literature Survey**

Crowd monitoring was done using image processing techniques [1] and closed-circuit television systems that supported both data collection and on-line monitoring of crowds led to an improved understanding of crowd behavior, improved design of the built environment and increased pedestrian safety.

By analyzing the motion instead of tracking subjects one by one [2], by using Motion History Image [3] and optical flow method, abnormal crowd behavior was detected. A mobile-based system [4] consisting of a server-side application connected to an IP surveillance camera(s) and mobile application with different user rights for receiving an alarm from the server-side application were developed. “Pfinder” [5] utilized a frequency of around 10 Hz for detection. The system that analyzed human motion during an image sequence for video-surveillance applications without using any 3D models was developed that traced the motion such as a human walking, with the help of visible boundaries [6].

A crowd segmentation algorithm [7] was developed that distributed an image into crowd and non-crowd regions. The main idea was to capture two key properties of crowds: on a narrow scale, its basic element should appear as if an individual's and on a larger scale, a crowd inherently contained repetitive appearance elements. Proposed method exploited this by building a pyramid of sliding windows and quantifying how “crowd-like” each level of the pyramid is using an underlying statistical model based on quantized SIFT features. An almost unsupervised learning method for crowd counting was proposed that divided a convolution layer into a grid of cells. Only the maximally activated neuron was allowed to update the filter within each cell. Almost 99.9% of the parameters were trained without using any labelled data, while the remaining 0.1 percent were pulled with supervision, which produced better results than other unsupervised approaches and was close to the accuracy of the supervised baseline. [8] Algorithms were developed to differentiate between crowded or non-crowded regions using VisDrone dataset with a variety of locations, environments, and lighting conditions [9], as well as tracking and counting using PETS2009 dataset [10] with efficiency of around 83.14%.

A wearable system for ease in mobility [11] was developed that used a camera which was able to be mounted overhead or chest. It assumed that the captured visual information was the user's pose. Along with it, an alignment technique was used that took a visual snippet of the current path from the user and corrected the current user’s pose to align with the corresponding pose in the training location. Audio information about the right direction was given as an output.

Several sticks as well as navigation devices were developed using ultrasonic sensors [12-19], infrared sensors [20], RFID tags [21] and water sensors, along with a microcontroller, cameras [22], Bluetooth [23] as well as an android device to detect impediments or obstacles in the path that communicated physical location and object location; directed the user through sound output or text-to-speech [24] or vibration feedback guidance allowing the user to reach the destination even if diverted from the directed path. The range of detection varied from 2m to 4m.

A mini handheld device named Indriya was developed for the blind which could be used as a companion with a walking cane featuring vibratory and voice feedback for easy navigation and collision alert if any obstacles were detected within the range of 3m as well as an ability to differentiate between objects and humans with the help of Android and Internet of Things (IOT) support [25].

"NavGuide" [26], an electronic gadget was created to aid visually impaired persons in finding an obstacle-free path. It offered simplified information about the surrounding environment and derived priority information without generating any information overload. Prioritization was aided by vibration and audible feedback methods. As a low-power embedded system, the hardware included ultrasonic sensors, vibration motors, and a battery. A controlled, real-world test set was used to evaluate 70 blind persons.

The system was primarily focused on providing assistive gadgets for visually impaired people in order to create a hearable environment for them. Image and video processing that is suited for a blind user converts visual data into audio, haptic, or a combination of the two as feedback. The correct answer with more than 90% accuracy rate was obtained by using CNN image classifier. The CIFAR-10 dataset was used [27].

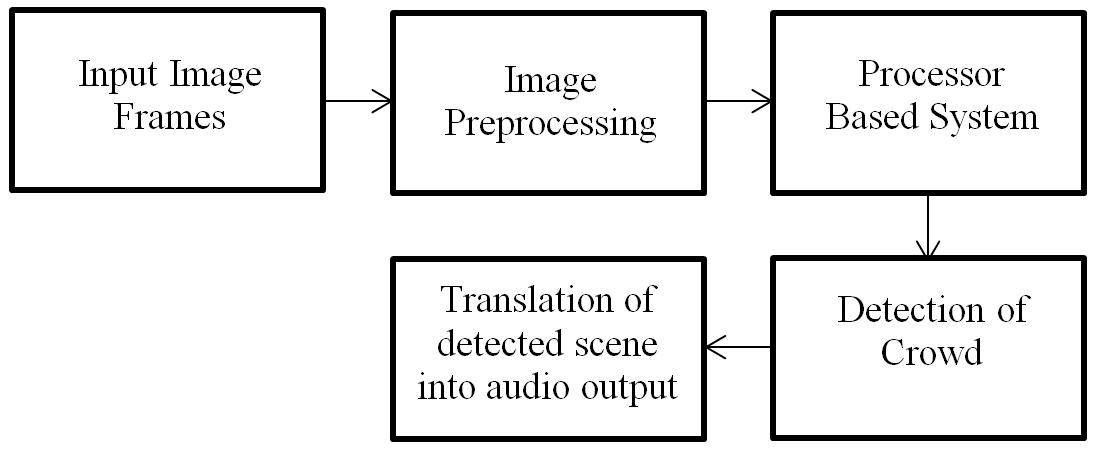
Using sensors like ultrasonic sensors, IR sensors and PIR sensors, Be My Third Eye [28] was developed that alerted the user by the beep sound of the buzzer if obstacles were detected. The PIR sensor generated a beep sound if motion was detected. It also consisted of GSM and GPS modules which could be used in case of emergency for location tracking and sending messages.

The dynamic scene understanding system was implemented on a head-mounted mobile device equipped with Intel’s Real sense camera along with a cell phone to understand the scenes in which pedestrians and vehicles can be detected [29].

A stereo vision-based system was proposed that assisted visually impaired people to get familiar with the unknown environments and whether it is safe by sensing the environment, segmenting the floor in 3D, merging local 2D grids with camera tracking, building a global occupancy 2D grid, reacting to close impediments, and generating haptic feedback with vibrations in a certain pattern were all used to achieve this. The authors analysed normal vectors and camera orientation generated from the depth and inertial data, respectively [30].

1. **Methodology**

The proposed system detects whether the surrounding is crowded or non crowded. Block diagram of the entire system is shown in Fig 1.



**Fig 1**: Proposed system

A monocular camera acquires a real time image of the surrounding and provides it to the image pre-processing stage. The processor based system works on the converted image, detects the class as crowded or non crowded and provides the output in audio format via earphone.

**Data Collection and Preprocessing:**

A custom dataset is created with 6108 labeled images. 80% of the images are from the internet [31][32] while 20% are captured from a smartphone [Redmi note 10 pro - 64mp (f/1.9) at resolution of 2400 x 1080]. The dataset is divided into two classes namely Crowded and Non crowded.

Images of the crowd class consists of 3 or more persons or even a mob in the frame. 3100 images belong to this class. Sample image is shown in Fig 2.



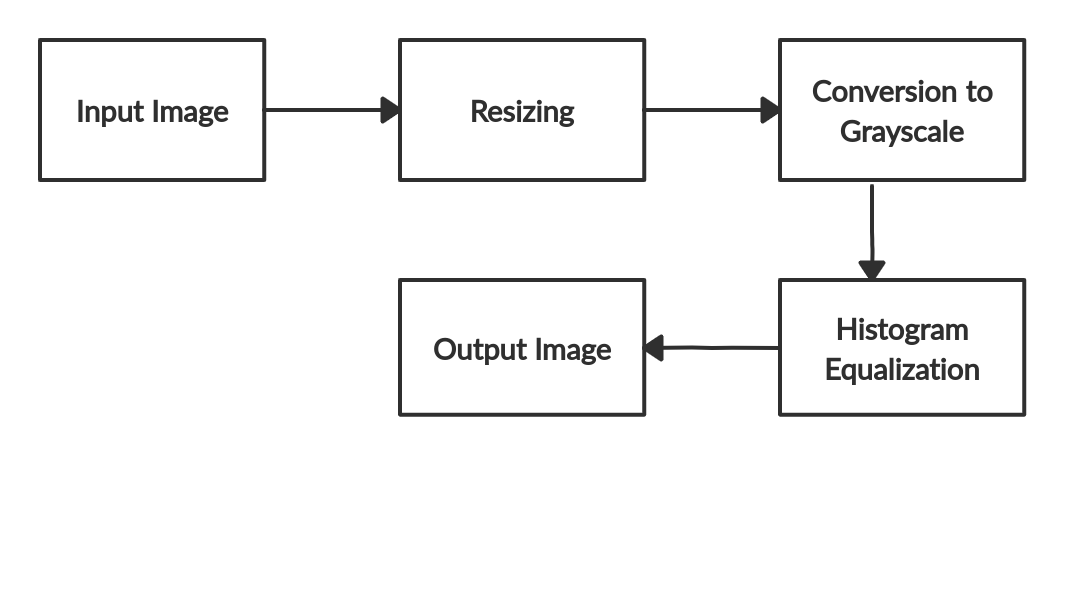
**Fig 2**: Sample image belonging to the crowded class

Images of a non crowded class consists of 1 or 2 persons in the frame. Images of empty roads are also classified as non crowded. 3108 images belong to this class. Sample image is shown in Fig 3.

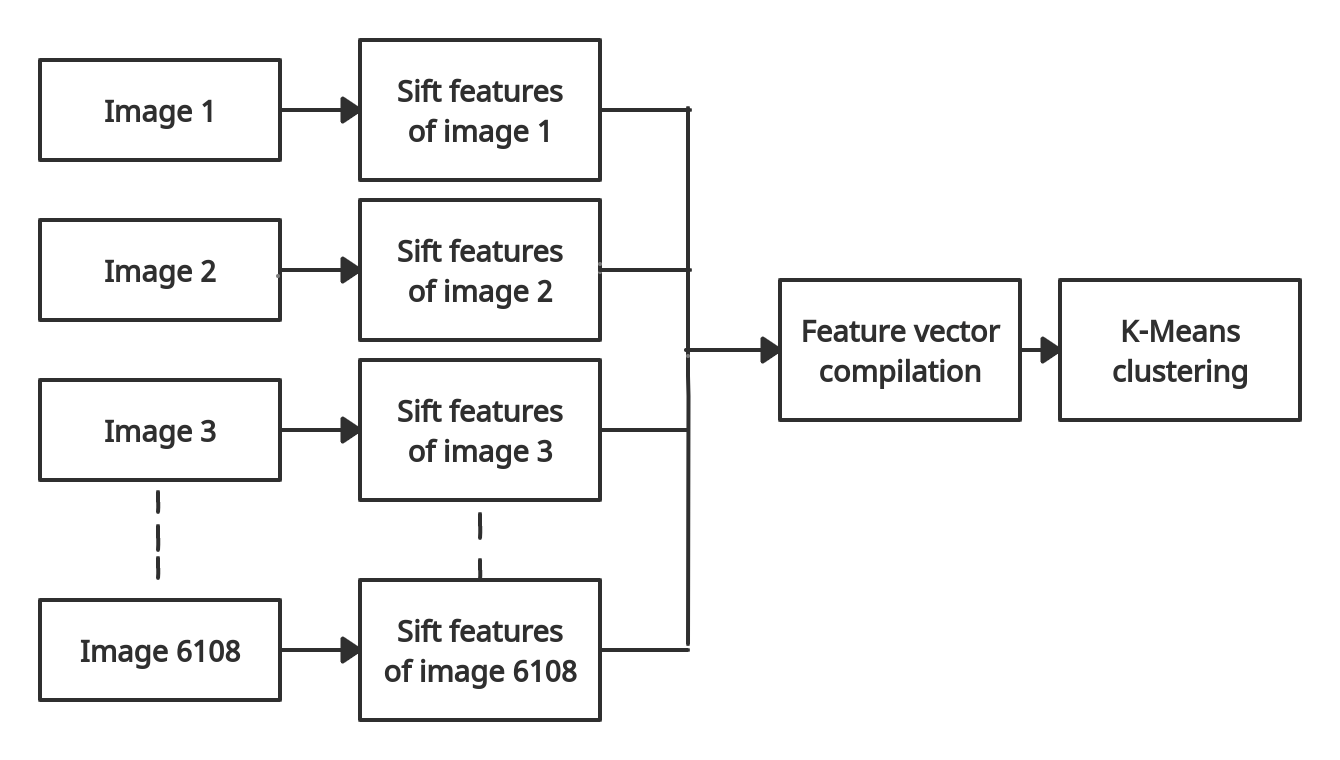


**Fig 3**: Sample image belonging to the non-crowded class

Preprocessing is performed to enhance the quality of the images. All images are resized to 100 x 300 pixels and converted to grayscale. Histogram equalization method is used to enhance unusual dark images by increasing the contrast. Image pre-processing is shown in Fig. 4.

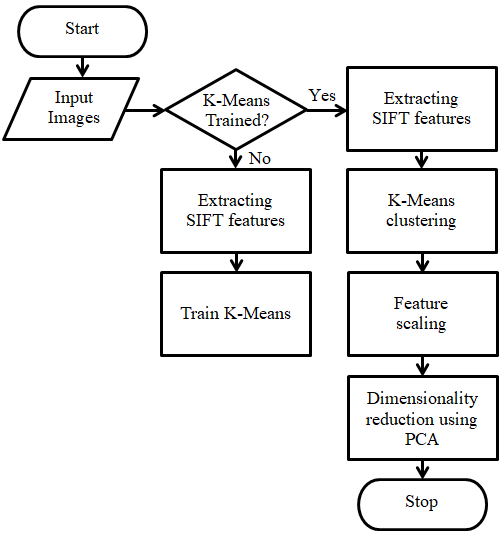
  
**Fig.4**: Image Pre-Processing

The Scale Invariant Feature Transform (SIFT) feature descriptor has been used for the feature extraction. It provided a feature vector of size 1630084 x 128 for 6108 images. Block diagram of Sift feature extraction is shown in Fig 5.



**Fig 5**:Sift feature extraction

The feature vector was fitted into K-Means clustering with K=5 and compacted into a 5 bin histogram. The pre-trained K-Means model reduced the size of feature vectors to 6109 x K. Principal Component Analysis (PCA) was used to further reduce the acquired feature vector of 6109 x K to 6109 x 3. Flow chart of feature extraction and dimensionality reduction is shown in Fig 6.



**Fig 6:** Feature extraction & reduction

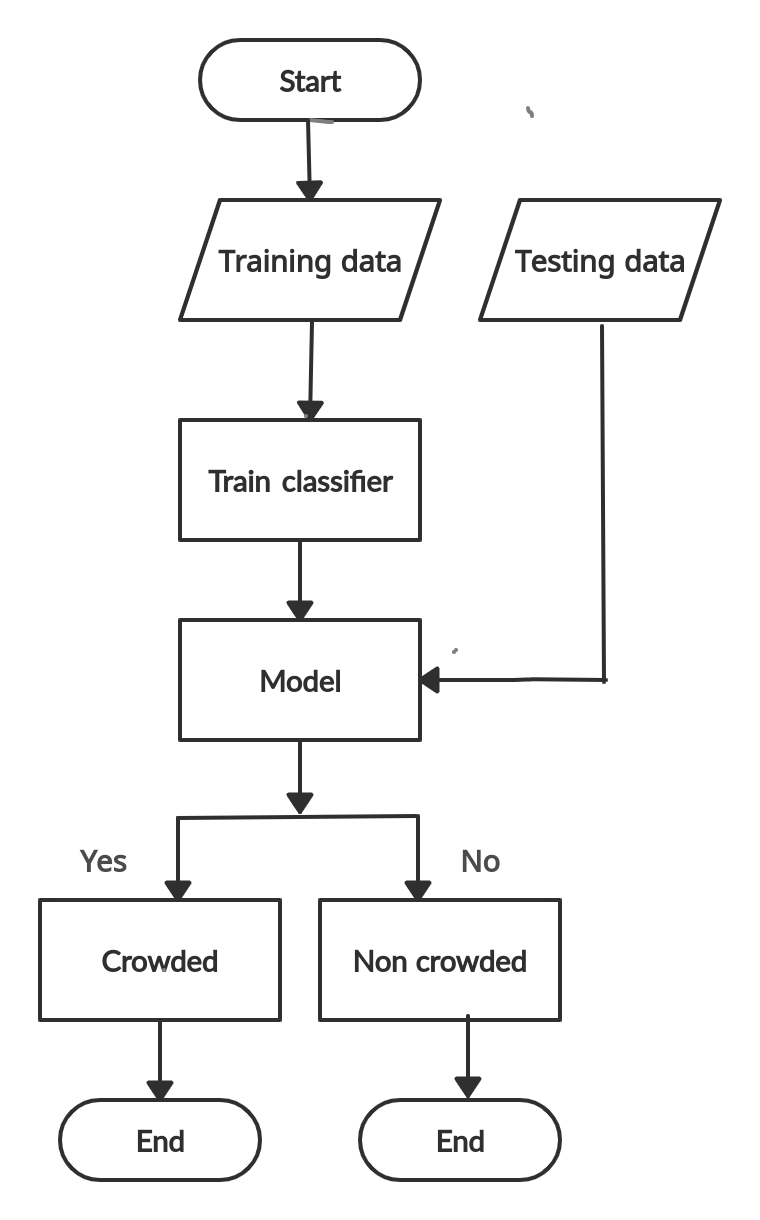
Algorithm 1 shows feature extraction and dimensionality reduction process.

|  |
| --- |
| **Algorithm 1: Feature extraction and Dimensionality reduction** |
| **Input:** Raw Image data ( 1630084 x 128 )  **Output:** Dimensionally Reduced feature vectors (6109 x 3) |
| *Initialization:*   1. **for** all images **do** 2. Extract the features using SIFT 3. Apply K-Means. ( K=5 ) 4. Feature scaling. 5. Append to feature vectors 6. **end for** 7. Reduce the large set of variables to required variables using PCA. 8. **return** vector of dimensionality reduced features. |

The dataset is split in training and testing sets and training dataset is fed to the classifiers.Seven classification models have been used i.e Decision Tree, Support Vector Machine (SVC with linear kernel,SVC with polynomial kernel, SVC with rbf kernel), K-Nearest Neighbours (KNN), Random Forest and Logistic regression.The flow chart of prediction is shown in Fig 7.

Algorithm 2 shows the classification process for crowd detection.

|  |
| --- |
| **Algorithm 2: For prediction of crowd** |
| **Input:** Training and testing data  **Output:** Prediction in audio format |
| *Initialization:*  Classifiers = Decision tree, SVM, random Forest, KNN, Logistic Regression   1. **for**  each C in classifiers **do** 2. split the data 3. model = classifier (data) 4. accuracy = evaluate (model) 5. prediction = model.predict (data) 6. **end for** 7. **for** prediction of each model **do** 8. **if** prediction = 1 9. audio\_output = non-crowded 10. **else** 11. audio\_output = crowded 12. **end for** 13. **return** prediction in audio format |



**Fig 7:** Training and testing

The first classifier used is Decision tree. The number of splits the tree makes before reaching a prediction is measured by tree depth. The value of tree depth in this system is considered as 13. The equation for entropy of the Decision tree is given by eq.1.

(1)

Where E, x and f denote the entropy, the number of classes and frequency of classes respectively.The problem of overfitting results in poor accuracy for the decision tree.

Random Forest combines a number of decision trees on different subsets of a dataset and takes the averages of the results to increase the dataset's predicted accuracy.Default hyperparameters have been considered.The equation for Random Forest is given by eq. 2

(2)

where, c is the entropy of all trees in forest and T is number of trees.Default hyperparameters were used for this classifier.

K-Nearest Neighbours (KNN) algorithm uses data and differentiates new data points based on similarity factors. Classification is done by a majority of votes obtained with the help of its neighbours.The default value of K has been taken as 5 (K=5). The equation for KNN is given by eq.3

(3)

where, the distance between j and k is given by the square root of summation of distance of every neighbor from the unknown point. This distance is used as a probability factor to assign an unknown point to a particular task.

Classification in SVM is carried out by finding the hyperplane which differentiates the two classes.SVM classification has been done using three kernels namely linear,radial basis function(rbf) and polynomial.The equation for the hyperplane is given by eq.4

(4)

where w is vector normal to the hyperplane, x is the data point and b is the bias.

Logistic Regression uses a logistic function to model the dependent variable. The hyperparameter used is random state whose value has been taken as 0.The equation of Logistic regression is given by eq.5

(5)

where P is the probability of success and x is a data point.

1. **Results**

The testing accuracies achieved by the classifiers Decision Tree, Support Vector Machine (SVM with linear kernel,SVM with polynomial kernel, SVM with rbf kernel), K-Nearest Neighbours (KNN), Random Forest and Logistic regression are 76.84% , 82.65% ,79.95%, 83.38% ,81.99% , 80.36% and 82.40% respectively. Table 1 represents the training and testing accuracies evaluated from the above listed models.

**Table 1:** classifiertraining and testing accuracies

|  |  |  |
| --- | --- | --- |
| **Classifier** | **Training Accuracy(%)** | **Testing Accuracy(%)** |
| **Decision Tree** | **93.34** | **76.84** |
| **SVM(rbf)** | **80.09** | **83.38** |
| **SVM(linear)** | **79.94** | **82.65** |
| **SVM(poly)** | **77.92** | **79.95** |
| **Random Forest** | **99.87** | **81.99** |
| **KNN** | **83.48** | **80.36** |
| **Logistic Regression** | **79.78** | **82.40** |

Table 1 interprets that SVM with rbf kernel and linear kernel provide better results compared to other classifiers as they are good in classifying non-linear data.

The time complexities for the classification models are O(n\*log(n)\*d) for decision tree, O(n²) for SVM, O(n\*log(n)\*d\*k) for random forest, O(knd) for KNN , and O(nd) for logistic regression assuming n as number of training examples, d as dimensions of the data, and k as number of neighbors.

Other performance metrics evaluated are the sensitivity, specificity, precision, recall and f-1 score. By comparing these values for all classifiers, SVM with polynomial kernel had the highest precision value of 0.85 while decision tree had the least with 0.75. SVM with rbf & linear kernels had the highest recall values as well as an f-1 score of 0.82. Decision tree provides the least values i.e. 0.76 for all the metrics as observed in Table 2.

**Table 2:** Performance evaluation metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classifier** | **Sensitivity** | **Specificity** | **Precision** | **Recall** | **F-1 score** |
| **Decision Tree** | **0.76** | **0.76** | **0.76** | **0.76** | **0.76** |
| **SVM**  **(rbf)** | **0.82** | **0.84** | **0.83** | **0.82** | **0.82** |
| **SVM**  **(linear)** | **0.82** | **0.82** | **0.82** | **0.82** | **0.82** |
| **SVM**  **(poly)** | **0.71** | **0.88** | **0.85** | **0.71** | **0.77** |
| **Random Forest** | **0.81** | **0.82** | **0.81** | **0.81** | **0.81** |
| **KNN** | **0.80** | **0.80** | **0.79** | **0.80** | **0.79** |
| **Logistic Regression** | **0.82** | **0.81** | **0.81** | **0.82** | **0.82** |

The classifiers were used on 100 random images of both the crowded and non crowded classes equal in number.

Analysis of all the above metrics recall, precision,F-1 score, sensitivity and specificity suggests that this system can give cognition about people around to the visually blind person.

The system is implemented using HP 14s laptop with AMD Ryzen 3500u processor (Quad-core, 8GB ram) along withAnaconda v4.10.3 (Jupyter Notebook) software for development and testing of the code.

1. **Conclusion**

The paper proposes a system which detects crowds to help the visually impaired people walk outdoors safely. It presents a comparative study of the existing work in the field of crowd detection for visually impaired people that requires additional devices, which adds to the load of visually impaired people. The proposed model classifies a place as crowded or non-crowded with minimal hardware.

This study can help determine the density of people in pandemic situations like Covid-19, when social distancing is required. The accuracy of the system can be improved further by increasing the size of the dataset.

Although this system works with an accuracy that is good enough to classify crowds, it will not be able to provide the exact information about the distance at which the crowd is present .This gives a future direction to this work wherein the user is informed about his distance to the crowd.

1. **References:**
2. Davies, Anthony C., Jia Hong Yin, and Sergio A. Velastin. "Crowd monitoring using image processing." *Electronics & Communication Engineering Journal* 7, no. 1 (1995): 37-47.
3. Ihaddadene, Nacim, and Chabane Djeraba. "Real-time crowd motion analysis." In *2008 19th International Conference on Pattern Recognition*, pp. 1-4. IEEE, 2008.
4. Lahiri, Shubham, Nikhil Jyoti, Sohil Pyati, and Jaya Dewan. "Abnormal crowd behavior detection using image processing." In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1-5. IEEE, 2018.
5. Shalash, Wafaa Mohib, Azzah Abdullah AlZahrani, and Seham Hamad Al-Nufaii. "Crowd Detection Management System." In *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1-8. IEEE, 2019.
6. Wren, Christopher Richard, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland. "Pfinder: Real-time tracking of the human body." *IEEE Transactions on pattern analysis and machine intelligence* 19, no. 7 (1997): 780-785.
7. Ricquebourg, Yann, and Patrick Bouthemy. "Real-time tracking of moving persons by exploiting spatio-temporal image slices." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, no. 8 (2000): 797-808.
8. Arandjelovic, Ognjen. "Crowd detection from still images." In *BMVC 2008: Proceedings of the British machine vision association conference 2008*, pp. 1-10. BMVA Press, 2008.
9. Sam, Deepak Babu, Neeraj N. Sajjan, Himanshu Maurya, and R. Venkatesh Babu. "Almost unsupervised learning for dense crowd counting." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 8868-8875. 2019.
10. Castellano, Giovanna, Ciro Castiello, Corrado Mencar, and Gennaro Vessio. "Crowd detection for drone safe landing through fully-convolutional neural networks." In *International conference on current trends in theory and practice of informatics*, pp. 301-312. Springer, Cham, 2020.
11. Aladren, Aitor, Gonzalo López-Nicolás, Luis Puig, and Josechu J. Guerrero. "Navigation assistance for the visually impaired using RGB-D sensor with range expansion." *IEEE Systems Journal* 10, no. 3 (2014): 922-932.
12. Salim, Sohail, Othman O. Khalifa, Farah Abdul Rahman, and Adidah Lajis. "Crowd Detection And Tracking In Surveillance Video Sequences." In *2019 IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, pp. 1-6. IEEE, 2019.
13. Prattico, Flavio, Carmelo Cera, and Filippo Petroni. "A new hybrid infrared-ultrasonic electronic travel aid for blind people." *Sensors and Actuators A: Physical* 201 (2013): 363-370.
14. Alma, A., S. Nithyashree, Podili Alekhya, S. N. Ramya, and Lovee Jain. "Blind Guide-an Outdoor Navigation Application for Visually Impaired People." *International Journal of Advances in Electronics and Computer Science* (2016): 102-106
15. Radhika, R., Payal G. Pai, S. Rakshitha, and Rampur Srinath. "Implementation of smart stick for obstacle detection and navigation." *International Journal of Latest Research in Engineering and Technology* 2, no. 5 (2016): 45-50.
16. Gbenga, Dada Emmanuel, Arhyel Ibrahim Shani, and Adebimpe Lateef Adekunle. "Smart Walking Stick for visually impaired people using ultrasonic sensors and Arduino." *International Journal of Engineering and Technology* 9, no. 5 (2017): 3435-3447.
17. Sharma, Himanshu, Meenakshi Tripathi, Amit Kumar, and Manoj Singh Gaur. "Embedded assistive stick for visually impaired persons." In *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-6. IEEE, 2018.
18. A. Satam, Ihab & Al-Hamadani, Mokhaled & Ahmed, Alaa. (2019). Design and Implement A Smart Blind Stick. Journal of Advanced Research in Dynamical and Control Systems. 11. 42-47.
19. Ahmed, Wail Motwakil Idress, and Dr Eltahir Mohamed Hussein. "Design and Implementation of Eye Stick for Blind People." *International Journal of Engineering, Applied and Management Sciences Paradigms* 45, no. 01.
20. Taraniya, I., Bhaskar Reddy PV, Varshini Chaithra, Yeddula Divyasri, and Nalam Lakshmi Raviteja. "SMART BLIND STICK PRO." *International Journal of Advanced Research in Computer Science* 11 (2020).
21. Kanwal, Nadia, Erkan Bostanci, Keith Currie, and Adrian F. Clark. "A navigation system for the visually impaired: a fusion of vision and depth sensor." *Applied bionics and biomechanics* 2015 (2015).
22. Kher Chaitrali, S., A. Dabhade Yogita, K. Kadam Snehal, D. Dhamdhere Swati, and V. Deshpande Aarti. "An intelligent walking stick for the blind." *International Journal of Engineering Research and General Science* 3, no. 1 (2015): 1057-1062.
23. TAKIZAWA, Hotaka, Akira KITAGAWA, and Mayumi AOYAGI. "Stereovision cane system: obstacle detection and seat recognition for the visually impaired." *IIEEJ Transactions on Image Electronics and Visual Computing* 6, no. 2 (2018): 74-81.
24. Sharma, Sharang, Manind Gupta, Amit Kumar, Meenakshi Tripathi, and Manoj Singh Gaur. "Multiple distance sensors based smart stick for visually impaired people." In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 1-5. IEEE, 2017.
25. Kumar, Mithiles, Faysal Kabir, and Sahadev Roy. "Low cost smart stick for blind and partially sighted people." *International Journal of Advanced Engineering and Management* 2, no. 3 (2017): 65-68.
26. Kallara, Sujith B., Mitu Raj, Roshan Raju, Nidhin Jacob Mathew, V. R. Padmaprabha, and D. S. Divya. "Indriya—a smart guidance system for the visually impaired." In *2017 International Conference on Inventive Computing and Informatics (ICICI)*, pp. 26-29. IEEE, 2017.
27. Patil, Kailas, Qaidjohar Jawadwala, and Felix Che Shu. "Design and construction of electronic aid for visually impaired people." *IEEE Transactions on Human-Machine Systems* 48, no. 2 (2018): 172-182.
28. Kumar, Ashwani, and Ankush Chourasia. "Blind navigation system using artificial intelligence." *International research journal of engineering and technology (IRJET)* 5 (2018).
29. Suryawanshi, Rahul, Manisha Valecha, Jyoti Tejwani, Bhavna Jhamtani, and Mannat Doultani. "Be My Third Eye-A Smart Electronic Blind Stick with Goggles." (2019).
30. Tian, Shishun, Minghuo Zheng, Wenbin Zou, Xia Li, and Lu Zhang. "Dynamic crosswalk scene understanding for the visually impaired." *IEEE transactions on neural systems and rehabilitation engineering* 29 (2021): 1478-1486.
31. Díaz-Toro, Andrés A., Sixto E. Campaña-Bastidas, and Eduardo F. Caicedo-Bravo. "Vision-Based System for Assisting Blind People to Wander Unknown Environments in a Safe Way." *Journal of Sensors* 2021 (2021).
32. [https://www.alamy.com/stock-photo/crowded-street-india.htm](https://www.alamy.com/stock-photo/crowded-street-india.html)l
33. <https://www.shutterstock.com/search/india+crowded+street>