

ORIGINAL

Crowd Behavior Analysis and Prediction using the Feature Fusion Framework

Análisis y predicción del comportamiento de las multitudes mediante el marco de fusión de características

Manu Yadakere Murthygowda¹  , Ravikumar Guralamata Krishnegowda¹  , Shashikala Salekoppalu Venkataramu²  

¹Adichunchanagiri University. BGS Institute of Technology, Department of Research and Development (R&D). Karnataka, India.

²Adichunchanagiri University. BGS Institute of Technology, Department Of Computer Science and Engineering (CS&E). Karnataka, India.

Cite as: Murthygowda MY, Krishnegowda RG, Venkataramu SS. Crowd Behavior Analysis and Prediction using the Feature Fusion Framework. Salud Cienc. Tecnol. 2022; 2(S2):251. <https://doi.org/10.56294/saludcyt2022251>

Submitted: 18-11-2022

Revised: 10-12-2022

Accepted: 24-12-2022

Published: 31-12-2022

Editor: Fasi Ahamad Shaik 

ABSTRACT

The increasing number of people is a major cause of disasters that occur due to overcrowding. The gatherings of crowds in public places are a source of panic, which results in disaster. An analytical study was performed on crowd management. This is highly essential for the design of a well-planned public space, the possibility of surveillance in every area, and transportation systems. The disasters that occur due to uncontrollable crowd behaviour involve loss of property, fatalities, or casualties. To avoid this, the crowd's behaviour was analysed. A MFF (multi-level feature fusion) framework was designed in this paper to predict behaviour. The first level of multi-level feature fusion employs motion and appearance, the second level employs spatial connections, and the third level employs temporal features. The combination of these characteristics aids in the exploitation of crowd behaviour. Furthermore, MFF was evaluated considering the web dataset, considering accuracy, precision, and recall as parameters. Comparative analysis was carried out with various existing methodologies with an accuracy of above 99 %.

Keywords: Crowd Behaviour Analysis; Multi-level Feature Fusion and Surveillance.

RESUMEN

El creciente número de personas es una de las principales causas de las catástrofes que se producen debido al hacinamiento. Las concentraciones de multitudes en lugares públicos son una fuente de pánico que desemboca en catástrofes. Se ha realizado un estudio analítico sobre la gestión de multitudes. Es esencial para el diseño de un espacio público bien planificado, la posibilidad de vigilancia en todas las zonas y los sistemas de transporte. Las catástrofes que se producen debido al comportamiento incontrolable de las multitudes implican pérdidas materiales, víctimas mortales o heridos. Para evitarlo, se analizó el comportamiento de la multitud. En este artículo se diseñó un marco MFF (fusión de características multinivel) para predecir el comportamiento. El primer nivel de fusión de características multinivel emplea el movimiento y la apariencia, el segundo las conexiones espaciales y el tercero las características temporales. La combinación de estas características ayuda a explotar el comportamiento de las multitudes. Además, MFF se evaluó teniendo en cuenta el conjunto de datos web, considerando como parámetros la exactitud, la precisión y el recuerdo. Se realizó un análisis comparativo con varias metodologías existentes con una precisión superior al 99 %.

Palabras clave: Análisis del Comportamiento de Multitudes; Fusión de Características Multinivel y Vigilancia.

INTRODUCTION

Crowd management is highly essential as it directly affects human lives. There are various situations in which massive crowd gatherings occur, such as games, religious gatherings, political events, concerts, society gatherings; educational events, and parties.⁽¹⁾ There are also occurrences of events that are unfamiliar. These events are distinguished and classified based on the objects that move in a crowded scenario. Patterns of motion that are detected in the surveillance videos, are known as "abnormal behaviour detection".⁽²⁾

In surveillance videos, a mechanism based on Multiview clustering is applied to crowd gatherings.⁽³⁾ A framework is proposed that is exclusive of parameters for the identification of structural properties. This mechanism was used to generate feature points for crowd conduct detection. Moreover, an analysis performed on the behaviour shows it can be categorized into subspaces. A methodology for classification is introduced in this proposed work to manage a crowd that has an imbalance. All the categories of temporal-spatial features are meant to share a relationship. Although there is a requirement for features in the dynamic sense for an efficient analysis of crowd.⁽⁴⁾

A framework of deep learning that involves semi-supervision is proposed for the examination of events that are abnormal in crowd gatherings. This deep learning mechanism aids in distinguishing crowd behaviour irregularities. This system detects the object point in surveillance videos of crowd masses automatically. There are efficient predictions that are performed using the concept of entropy reduction.⁽⁵⁾

Another proposed system, based on a deep learning model known as CNN (Convolutional Neural Network), is used to detect crowd irregular behaviour in crowd videos using information based on motion. The architecture used in this system helps to differentiate abnormal from normal behaviour.⁽⁶⁾

Crowd conduct analysis has become a necessity due to the disasters that follow at crowd gatherings by occurrence of unusual events such as stampedes, natural disasters, attacks, bombings and, most commonly, traffic jams that occur under normal conditions. These events need proper management in order to avoid these disasters. As a result, video surveillance is critical for extracting and analysing human behaviour. Moreover, a smart automated monitoring system could handle these issues efficiently. This system would require feature optimization and analysis for accurate crowd management.⁽⁷⁾

To analyse the unfamiliar events that occur in crowded scenarios, a framework based on deep learning methodology with a semi-supervised approach has been introduced. In crowded gatherings, this deep learning mechanism efficiently detects irregularities and differences. Additionally, automatic feature extraction of particular points and meaningful information about these features are provided. A concept based on entropy minimization is also proposed for effective and efficient predictions.⁽⁸⁾

There are also several implications on security when it comes to crowd analysis. Any human operating on surveillance may miss any hints that could lead to an emergency situation; therefore, using an automated methodology is more enhanced and accurate. There have been numerous crowd disasters in the past, such as the disaster that occurred during the Love Parade in Germany in 2010. This disaster could have been avoided if a crowd analysis system had been in place.⁽⁹⁾

Among other approaches, a data mining methodology based on WIFI sensing was used to analyse crowd behaviour in crowded conditions. In addition, for the collection of probe requests and the estimation of object patterns, a detailed and specific data analysis is performed. This study evaluates crowded conditions by using patterns based on spatial objects.⁽¹⁰⁾

However, crowd-based analysis is a difficult task that requires accurate and precise features for detection and other functions to be performed using surveillance videos. The extracted features are used in a crowd feature optimization process using a fused feature methodology to improve the efficiency of the result produced.

Motivation and contribution

The current situation regarding the crowd at social gatherings and public places is a cause of crowd calamities and disasters. This can be avoided by using an automated system for crowd analysis. The increasing need of individuals to maintain social distance among one another in public places also makes the development of these methodologies highly essential. For better control, the management of transportation systems, local stations, and the design of public spaces to be visited by the public for event gatherings necessitates an automated crowd management system. The absence of this system could result in a disaster such as a stampede, which has occurred in previous events around the world. The system proposed in this paper is a crowd analysis system that extracts physical features and creates a fused feature system for improved and accurate management. Moreover, considering the above motivation, this research work has the following contributions:

- At first, the surveillance videos are used to extract the necessary physical features of individuals to learn the crowd conduct. These features are thoroughly studied considering the structure of the crowd.
- This research work introduces MFF (Multi-level feature fusion) to exploit the crowd behaviour feature; multilevel feature comprises three distinctive levels of feature exploitation. The first level of feature fusion exploits the physical feature, the second level of feature fusion exploits the spatial features, and the third

level feature fusion exploits the temporal features.

- In order to evaluate the MFF performance, a web dataset was considered with evaluation parameters: accuracy, precision and recall.
- Moreover, performance analysis also includes the comparison with various algorithms; MFF outperforms the existing methodologies.

Each feature in this research work was cautiously evaluated to help study and enhance new features as well as learning the application of these features for the further part of this research.

The organization of this research work is such that the first section discusses the previous crowd analysis, it emphasizes on the background of crowd management systems with their feature extraction process, along with the motivation and contribution to carry out this work. The second phase includes the existing methodologies, as well as their shortcomings and various techniques that were used. The third section focuses on the development of a mathematical model for feature extraction and fused feature process. The fourth section contains a comparative study of the features for a better evaluation.

METHODS

In this section, the mathematical modelling of the crowd features was obtained using fused feature methodology. The methodology for crowd analysis uses physical features that are collectively termed as “crowd features”. These crowd features are used for calculation and optimization, which further result in fused features. The fused feature mechanism provides an elaborate and detailed description of every feature that is being used in this study. The proposed mechanism for feature fusion is based on motion features. These crowd feature optimizations include features such as direction, motion, and speed, among others, obtained from surveillance videos by detecting objects such as individuals in the crowd. Figure 1 shows multilevel feature fusion and how different levels of feature fusion will occur.

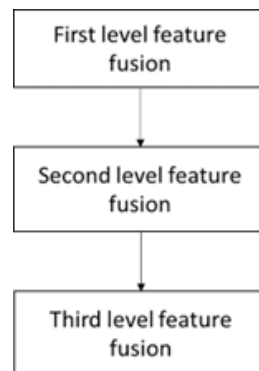


Figure 1. Multilevel Feature Fusion

Assuming a video frame of a crowd surveillance video is considered from a dataset is represented in the form of a matrix row such that it is a vector quantity. This is then represented as:

$$Q \in \mathcal{G}^{i \times j}$$

In which the total number of frames in the crowd videos along with the total number of pixels that are available in each frame is denoted as the coefficient i and variable j respectively. This is used to obtain a mathematical model of low dimensionality. A matrix is created for eliminating the inconsistencies between matrix Q and matrix Y . Therefore, the optimization for the inconsistencies was done as given below; it E was also used for the minimization parameter.

$$\mathbb{E}_{y,D} \|D\|, \text{whereas } \mathcal{B}(Y) \text{ is less than or equal to } \ell, Q = Y + E \quad (1)$$

Consider a video frame m , where the foreground frame D_m has a mapping size that is rearranged to the actual size of the frame. Considering the information of the foreground frame of i and j for generation of feature weights $F_h = \lambda_h(D_m)$ and $F_k = \lambda_k(D_m)$ respectively. This is performed on the basis of the motion flow method. Therefore, considering the frame information, the Crowd Feature Objects was designed as follows:

$$\xi(h, k) = \left\{ \forall c(h, k) \in \mathcal{D}_m \mid (F_h^2 + F_k^2)^{\frac{-1}{2}} > \mathcal{D}_0 \right\} \quad (2).$$

In equation 2, $\mathcal{D}_0$ is the threshold coefficient. This value remains constant and the foreground frame $\mathcal{D}_m$ has pixels that are denoted as $c(h, k)$.

Second Level Feature Fusion

In this section, the crowd features are mathematically represented, in which the model we proposed consists of a fused feature methodology based on the available data and the mechanisms that are available for the performance of encoding and decoding. In this case, the representation of the feature is done considering the crowd feature optimization $J = \{j_1, j_2, \dots, j_\omega\}$ in a dataset that is

$$\Xi_{Q_c} \mathcal{E}(Q_c, Y) \triangleq 2\varpi^{-1} \sum_{m=1}^{\varpi} \left\{ \|h_m - Q_c \rho_m\|_2^2 + \Theta \|\rho_m\|_1 \right\} \quad (3)$$

Considering the equation 3, the notation ω denotes the total number of crowd features that are trained whereas the notation represents the factor that is controlling. $Y = \{p_1, p_2, \dots, p_\omega\}$ is a set of all the features that are implied considering the total crowd features $H = \{h_1, h_2, \dots, h_\omega\}$. Although, the coefficients Q_c and Y are used for normalizing the model of the data which has been shown as represents regularization functions:

$$\|h_m - Q_c \rho_m\|_2^2 \text{ and } \|\rho_m\|_1$$

Third Level Feature Fusion

In the third level phase of crowd feature optimization is performed in the temporal domain while considering the fused feature methodology. Furthermore, the feature optimization performed in the first phase while considering the spatial domain is carried forward in the second phase.

The set of features for the k th frame has been present in the m th block is represented as $H_{m,k} = \{k_1, k_2, \dots, k_\omega\}$. Here, the crowd feature optimizer is defined as by $h_{c,c} \in [1, \omega]$. The total number of optimizers used are denoted as the variable ω . Here, the set of features are evaluated by the use of a mechanism based on optical flow. This is represented as $Y_{m,k} = \{p_1, \dots, p_\omega\}$ and the set of features is given by x . Hence, a matrix of feature and coefficients is formed $\omega \times x$ and the x th column of the matrix has been expressed as $p_{x,x}$. Therefore, the set of columns is expressed as $a_{m,lk} = \{p_1, \dots, p_x\}$. Then, the features of the motion are given as

$$\emptyset_{m,k} = \{\max(\hat{\rho}_1), \dots, \max(\hat{\rho}_x)\} \quad (4)$$

For every frame in each of the blocks, the features of the motion have been generated for obtaining representation at the frame level. Furthermore, all the frames that have been used in the training of features has been denoted as k and every frame is divided into c blocks. Furthermore, the set of features that are used for testing are given by considering frame tests k which is denoted as $\emptyset'$ and is given by equation 5.

$$\Xi_{Y'_m} \|\emptyset'_m - Q_{\mathbb{R}m} Y'_m\|_2^2 \text{ such that } \|Y'_m\|_1 \leq \mathcal{K}_m \quad (5)$$

In that equation, $\mathcal{K}_m$ is denoted to represent the upper limit for the set of control features. Hence, the model of the data by considering the temporal domain is denoted as:

$$\|\emptyset'_m - Q_{\mathbb{R}m} Y'_m\|_2^2$$

Whereas: $\|Y'_m\|_1$ It's used to represent the regularization term of the temporal domain.

In this case, the block representations are given as $Y'_m = \{y'_1, \dots, y'_c\}$.

There is also a distinction that has to be done between an abnormal and normal conduct by the use of classifiers that is given as $Y_m \in G^{\text{exf}}$. This is the performance of the crowd feature optimizer using fused feature methodology in the temporal domain.

RESULTS AND DISCUSSION

Performance evaluation

This section emphasizes on experimental results that were obtained by the use of multi-level feature fusion framework process performed with a fused feature methodology through video surveillance of crowd gatherings. The fused feature methodology aims at improving the efficiency of the extracted features by considering the evaluation of every particular feature that is extracted. The methodology proposed generates clear and specific descriptions of the features that are used for crowd pattern analysis. The analysis performed on these features were done through various visual analytics methods such as heat maps, video frames, etc. The analysis was performed on various patterns of crowd behaviour with different structures.

Motion features were mainly used for estimation of the tracking and detection system that was used further. These motion features are the physical attributes that are collectively termed as crowd feature optimization used on various types of datasets that represent different patterns of crowd gatherings. Performance of the crowd features and fused feature methodology was measured using multiple performance metrics such as precision, accuracy, area under the curve and recall, these metrics help in performing a comparative analysis. A comparative analysis was performed considering several datasets to which different algorithms were applied and these metrics evaluations were considered to find the more efficiently working model. The experimental results demonstrated performance efficiency of the fused features also considering the structure features, shapes along with motion features. Elimination or dropout of unwanted features are also done while the required features are fused.

Dataset Details

The performance of the proposed system using fused feature methodology was tested using a web dataset that is gathered through online sources such as Getty Images and ThoughtEquity. The web dataset consists of videos in urban areas which are of high quality, they also include documentary videos. This dataset includes both categories of abnormal as well as normal videos. The normal category of videos is 12 in number whereas the abnormal category of videos is 9 in number. Few examples of normal category of crowd videos include a pedestrian walking in the park and abnormal category of crowd videos include crowd protesting, fighting, and trying to escape from attacks. Figure 2 shows a comparative analysis using different algorithms for web data sets where multiple crowds in various environments and also different frames can be seen. Figure 3 shows results of accuracy by considering the different corridor scenarios for various video frames.



Figure 2. Comparative Analysis using various algorithms for web data set

Comparative Analysis using Various Algorithms for Web dataset

Figure 3 also represents the accuracy of several models that are compared using different crowd analysis methodologies. Namely these analysis techniques for crowd behavior include Irregularity-Aware Semi-Supervised Deep Learning Model (IASSLM)⁽¹¹⁾, Convolutional Neural Networks (CNN)⁽¹²⁾, WideResNet ⁽¹³⁾ and IMFF model using web datasets. The highest accuracy recorded was with the MFF algorithm (99,36 % - Table 1). This is the highest recorded accuracy of the MFF algorithm which is of great marginal value compared to the other crowd analysis methodologies that were used.

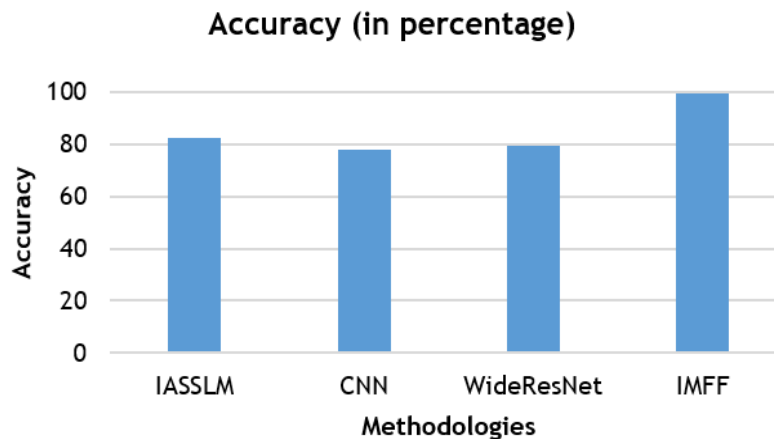


Figure 3. Accuracy results considering the corridor scenario

Methodologies	Accuracy
Cognition (ConvLSTM)	65,81
Social Force Model (SFM)	81,11
Force Field	85,15
Two-Stream CNN	90,51
Two Stream Deep VGG	94,59
LPT (Lagrangian Particle Trajectories)	95,66
CDM(Cognitive Deep Model)	95,67
MFF Model	99,36

Table 1 shows the accuracy results of all the algorithms that perform crowd analysis using web dataset. Using this web dataset, various algorithms such as the Social Force Model (SFM)⁽¹⁴⁾, Lagrangian Particle Trajectories⁽¹⁴⁾, Two-Stream CNN⁽¹⁴⁾, Cognitive Deep Model⁽¹⁴⁾, Two Stream Deep VGG⁽¹⁴⁾, Force Field⁽¹⁴⁾, Cognition (ConvLSTM)⁽¹⁴⁾, and MFF were also compared.

We also compared the precision and recall of the FCDLF mechanism with the proposed MFF mechanism (Table 2). Moreover, it is observed that MFF has a 99,82 % precision in comparison with 94,5 % of FCDLF. Furthermore, MFF observed a recall of 98,89 % in comparison with 93,35 % of FCDLF.

Methodologies	FCDLF ⁽¹⁵⁾	MFF
Precision	94,5	99,82
Recall	93,35	98,89
Accuracy	95,67	99,36

Table 1 also gives the exact accuracies of these algorithms making the distinction between them even more precise. The highest accuracy recorded is by the IMFF algorithm which is 99,56 %. This is the highest recorded accuracy of the IMFF algorithm has a large marginal value in comparison to the other used crowd analysis methodologies.⁽¹⁵⁾ Figure 4 and Figure 5 represents the efficiency and classification error by considering the courtyard data frame.

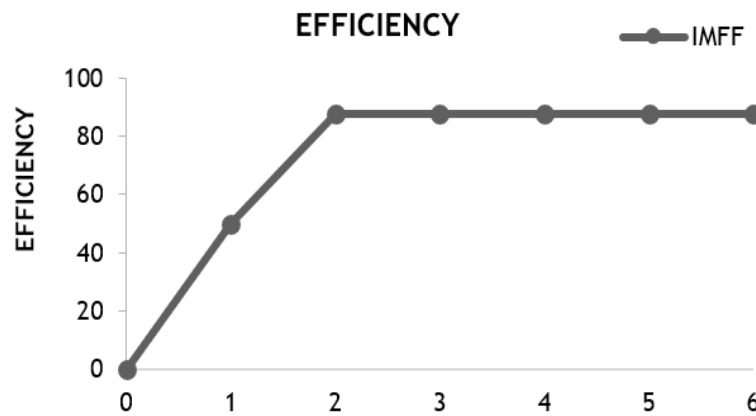


Figure 4. Efficiency considering the courtyard data frame



Figure 5. Classification error considering the courtyard data frame

CONCLUSION

The current situation regarding the crowd at social gatherings and public places is a cause of crowd calamities and disasters. This can be avoided by using an automated system for crowd analysis. The increasing need of individuals maintaining social distance among one another in public places also make the development of these methodologies highly essential. Our research work introduces the multi-level feature fusion to predict the crowd behaviour. Multi-level feature fusion also known as MFF comprises three different levels. The first level of feature fusion focuses on the physical features, the second level of feature fusion focuses on the spatial features, and the third level feature fusion focuses on the temporal feature fusion. Moreover, MFF (Multi-level feature fusion) exploits the crowd feature and enhances the prediction. MFF was evaluated considering the Web dataset in terms of accuracy, precision and recall. Further comparative analysis was carried out with the existing model with an accuracy of 99,36 %, a precision score of 99,82 % and a recall value of 98,89 %. Although MFF outperforms the existing model, there are still several research gaps since crowd behaviour is highly unpredictable and individual, requiring the use of multiple datasets for further evaluation.

REFERENCES

1. Helbing D, Mukerji P. Crowd disasters as systemic failures: Analysis of the love parade disaster. EPJ Data Sci. 2012;1(1):7. <http://doi.org/10.1140/epjds7>.
2. Shao J, Loy CC, Kang K, Wang X. Slicing convolutional neural network for crowd video understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 26-30; Las Vegas, NV, USA. IEEE; 2016. p. 5620-5628. <http://doi.org/10.1109/CVPR.2016.606>.
3. Wang Q, Chen M, Nie F, Li X. Detecting coherent groups in crowd scenes by multiview clustering. IEEE Trans Pattern Anal Mach Intell. 2020;42(1):46-58. <http://doi.org/10.1109/TPAMI.2018.2875002>.

4. Deng C, Kang X, Zhu Z, Wu S. Behavior recognition based on category subspace in crowded videos. IEEE Access. 2020;8:222599-222610. <http://doi.org/10.1109/ACCESS.2020.3043412>.
5. Aljaloud AS, Ullah H. IA-SSLM: Irregularity-Aware Semi-Supervised Deep Learning Model for Analyzing Unusual Events in Crowds. IEEE Access. 2021;9:73327-73334. <http://doi.org/10.1109/ACCESS.2021.3081050>.
6. Direkoglu C. Abnormal crowd behavior detection using motion information images and convolutional neural networks. IEEE Access. 2020;8:80408-80416. <http://doi.org/10.1109/ACCESS.2020.2990355>.
7. Xu M, Li C, Lv P, Chen W, Deng Z, Zhou B, Manocha D. Emotion-based crowd simulation model based on physical strength consumption for emergency scenarios. IEEE Trans Intell Transp Syst. 2020;1-15. <http://doi.org/10.1109/TITS.2020.3000607>.
8. Zhu F, Wang X, Yu N. Crowd tracking with dynamic evolution of group structures. In: Proceedings of the European Conference on Computer Vision (ECCV); 2014 Sep 6-12; Zurich, Switzerland. Springer; 2014. p. 139-154.
9. Güler MS, Department of Information Systems. Automated crowd behavior analysis for video surveillance applications [master's thesis]. Istanbul, Turkey: Bogazici University; 2012. p. 73.
10. Zhou Y, Lau BPL, Koh Z, Yuen C, Ng BKK. Understanding Crowd Behaviors in a Social Event by Passive WiFi Sensing and Data Mining. IEEE Internet of Things Journal. 2020;7(5):4442-4454. <http://doi.org/10.1109/JIOT.2020.2972062>.
11. Aljaloud AS, Ullah H. IA-SSLM: Irregularity-Aware Semi-Supervised Deep Learning Model for Analyzing Unusual Events in Crowds. IEEE Access. 2021;9:73327-73334. <http://doi.org/10.1109/ACCESS.2021.3081050>.
12. Fradi H, Luvison B, Pham QC. Crowd Behavior Analysis Using Local Mid-Level Visual Descriptors. IEEE Transactions on Circuits and Systems for Video Technology. 2017;27(3):589-602.
13. Zagoruyko S, Komodakis N. Wide residual networks. arXiv preprint arXiv:1605.07146. 2016. <https://doi.org/10.48550/arXiv.1605.07146>.
14. Varghese E, Thampi SM, Berretti S. A Psychologically Inspired Fuzzy Cognitive Deep Learning Framework to Predict Crowd Behavior. IEEE Transactions on Affective Computing. <http://doi.org/10.1109/TAFFC.2020.2987021>.
15. Mondal S, Roy A, Mandal S. A supervised trajectory anomaly detection using velocity and path deviation. In: Proceedings of International Conference on Frontiers in Computing and Systems. Springer; 2021. p. 777-784.

CONFLICTS OF INTEREST

None.

FINANCING

None.

AUTHORSHIP CONTRIBUTION

Conceptualization: Manu Yadakere Murthygowda, Ravikumar Guralamata Krishnegowda, Shashikala Salekoppalu Venkataramu.

Methodology: Manu Yadakere Murthygowda, Ravikumar Guralamata Krishnegowda, Shashikala Salekoppalu Venkataramu.

Writing - Original Draft: Manu Yadakere Murthygowda, Ravikumar Guralamata Krishnegowda, Shashikala Salekoppalu Venkataramu.

Writing - Review & Editing: Manu Yadakere Murthygowda, Ravikumar Guralamata Krishnegowda, Shashikala Salekoppalu Venkataramu.