An integrated multi-level feature fusion framework for crowd behaviour prediction and analysis

Manu Yadakere Murthygowda[1], Ravikumar Guralamata Krishnegowda[2], Shashikala Salekoppalu Venkataramu[3]
Department of Computer Science and Engineering, BGS Institute of Technology, Adichunchanagiri University, B. G Nagara, India

Article Info

Article history:

Received Mar 14, 2022
Revised Feb 13, 2023
Accepted Feb 18, 2023

Keywords:

Convolution neural network
Crowd behavior analysis
Feature fusion
Multi level feature fusion
Surveillance

ABSTRACT

The uncontrolled outburst in population has led to crowd gatherings in various public places causing panic and disaster in certain unpleasant and extreme conditions. A study on the analysis of crowd accumulation has been carried out for various reasons that include management of crowd, design of a well-planned public space, the possibility for surveillance at every area and transportation systems. A lot of disasters also occurs due to uncontrollable crowd behaviour and poor crowd management. It could result in loss of property, fatalities or casualties. To avoid this, the conduct of a crowd of people has been studied and analyzed to control their movement and traffic. Hence, in this research work, integrated multi-level feature fusion (IMFF) framework is designed to predict the behaviour; further classification based on the local region is carried out to enhance the prediction. In the case of multi-level feature fusion; first level feature fusion utilizes the motion and appearance; second-level feature fusion utilizes the spatial connection and third-level utilizes the temporal connections. Further, the classification approach is integrated based on the local region is used to enhance the crowd behaviour prediction in terms of accuracy and faster. Moreover, performance evaluation is carried out considering the two distinctive datasets.

Corresponding Author:

Manu Yadakere Murthygowda
Department of Computer Science and Engineering, BGS Institute of Technology
Adichunchanagiri University
Mandya District, B. G Nagara, Karnataka 571448, India
Email: manuym@bgsit.ac.in

1. INTRODUCTION

Crowd analysis has been done widely by the use of surveillance videos and other means of perceiving. The analysis of crowd conduct is more difficult and challenging as compared to the movement analysis of any individual. Crowd conduct is essentially performed under sociological constraints. Crowd analysis and simulation has various applications that include control on the transportation system, management of crowd, crowd disaster management prevention, and well-planned public area. There are various uses foranalysing and studying crowd conduct although complex models have been used and implemented in the process [1]. The social and economic development that is occurring on a large scale causes an increase in the group gathering activities and the events that take place in relation. An analysis of a gathered mob is essential also to make evacuation plans in case of emergencies. Various physical attributes are considered while referring to mob analysis, individual presence is highly essential while considering this factor. The direction, movement and speed of the individual need to be contemplated. These physical attributes that are involved in crowd feature optimization are calculated using physics [2]. Simulation about crowd analysis is one of the most common and essential topics about computer graphics that are being used recently in terms of emergencies. This is mainly calculated using individual movement in a mob, the speed and movement of an individual are observed.

Physical attributes of an individual have to be considered which are termed crowd feature optimization [3]. Few of these factors which contribute physically dynamically evolve it has been influenced by the individual movement. The challenges that are posed have been addressed using a simulative model that is used by integrating physical attributes as well as physics. Firstly, the crowd feature optimization is calculated. Furthermore, the fused feature optimization is done which analyses the features for better analysis on the mob surveillance. This phase helps in exploring the features further which is used to enhance the crowd analysis. Crowd feature optimization refers to the strength of the person to exert force on other objects, in this case, it is mainly observed by the speed and direction in which the person proceeds [4], [5]. Automated analysis of crowd conduct has been observed and studied but it gaining more importance lately as the requirement for crowd gathering and maintaining is more crucial currently. The crowd conduct in any public area can be either structured or unstructured. The structured refers to the motion of the crowd in a consistent manner, where the movement, speed and direction remain constant forthetime at a particular area. Examples of these include a running marathon race, and a queue system of waiting.

An unstructured form of crowd conduct is an inconsistent, random motion of people in an area where the movement, speed and direction of individuals vary for time. Examples of unstructured crowd conduct include crowds at the railway station, bus stations, zoos and other public places [6]. There is also a lot of implication on security when concerning crowd analysis. Any human operating on surveillance could miss out on any hints that may lead to emergencies; therefore, it is more enhanced and accurate to use an automated methodology. There have been a lot of crowd disasters that have occurred previously such as the disaster that happened in the Love Parade 2010 in Germany. This disaster could have been avoided if there was an automated system for crowd analysis [7]. Recently there has been excessive use of surveillance camera systems fixed at every public location that have caused advancements in the digital imaging world where the information that is extracted from the video can be analysed and used further. The raw information that has been extracted to surveillance videos is processed for feature extraction whereas in this case, physical attributes are studied and analysed. Using this information, the various features that are needed for an automated crowd detection system is processed and passed into a machine learning model which helps in detecting, tracking and analysing objects [8].

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

- At first, the surveillance videos are used to extract the necessary features of individuals to learn the crowd conduct. These features are thoroughly studied considering the structure of the crowd.
- This research work develops an multi-level feature fusion (MFF) which exploits the different features at a different level to extract the useful information; first level feature fusion exploits the physical features, the second level of feature fusion exploits spatial connection and the third level of feature fusion exploits the temporal features. Further, local regions are exploited for classification to enhance crowd behaviour prediction.
- Performance evaluation is carried out considering the two distinctive popular data sets i.e. UMN dataset and the violent dataset considering the different parameters like accuracy, efficiency and classification error.
- Comparative analysis is carried out considering the various existing methodologies including deep learning-based methodologies and it is observed that MFF outperforms the other methodologies with the maximum value of evaluation parameter.

The organization of this research work is such that the section 1 discusses the previous crowd analysis, it emphasizes the background of crowd management systems with their feature extraction process along with the motivation and contribution to carry out this work. The section 2 phase consists of the already existing methodologies along with their shortcomings and various techniques that have been applied. The section 3 focuses on the development of the mathematical model for the feature extraction and fused feature process. The section 4 contains a comparative study on the features for a better evaluation. This ends with a conclusion stating the outcome of this research.

2. RELATED WORK

Crowd conduct analysis has become a necessity due to the disasters that follow at crowd gatherings by the occurrence of unusual events such as stampede, natural disasters, attacks, bombings and most commonly traffic jam that occurs under normal conditions. These events need proper management to avoid these disasters. Therefore, it is essential to use video surveillance for the extraction of human conduct analysis. These problems could be handled efficiently by the use of a smart automated monitoring system. This system would require feature optimization and analysis for accurate crowd management. Hence, several people in the research field have provided solutions for the above-stated problems and a few of these studies are presented below.

Motion information is used in the detection and analysis of human conduct in crowded environments. Evaluation based on optical flow by feature vectors is done through the information of angles and magnitude. This model proposed in [9] is used for the distinction between abnormal and normal behaviour of individuals in a mob. Mechanism for clustering based on a Multiview is proposed for the detection of groups that are coherent in public gatherings based on L-1 and L-2 norms [10]. The clustering mechanism that is introduced in this proposed work generates the weights of features that are determined by a group number by the use of a counting framework [11]. Proposes a clustering mechanism based on the density of spatial angularity for the segmentation of patterns in crowded situations. In this study, input trajectories are used based on computational information that involves both angular and spatial. The benchmarked dataset that is used for the testing of this model's performance is CUHK. Andit introduces a methodology based on data mining used the analysis the behaviour of the crowd in crowded conditions based on WIFI sensing [12]. In addition to this, a detailed and specific data analysis is performed for the collection of probe requests and estimation of object patterns. This study uses patterns based on spatial objects for the evaluation of crowded conditions. Crowd surveillance videos are studied and analysed from public gatherings using the architecture of deep neural networks [13]. Here, structures, object position and patterns are used in computation through a Spatio-temporal model for transfer. A prediction module based on point aware flow is also proposed which is used for the analysis of the consistency for object appearance. It introduces a framework based on the deep learning methodologies for the analysis of human conduct for events that are unusual based onafuzzy cognitive deep learning model [14].

The recognition of the cognitive and psychological characteristics is performed using this model for crowd behaviour. Visual behaviour is analysed in this paper using a five-factor model. A methodology for the detection of crowd behaviour is proposed using the phenomenon of collective motion [15]. Temporal-spatial features are analysed in this model by detection using the collective motion of objects. Also, a model that is intention aware is proposed for capturing intrinsic dynamics in the objects. A convolutional neural network architecture is proposed in [16] for counting objects present in massive public gatherings to analyse crowd conduct. The dataset used for performance testing is UCF-QNRF. Also, the analysis performed on the crowd is done on the estimation based on a density map [17]. A framework using deep learning methodology is introduced to have a semi-supervised approach to analyse the unfamiliar events that occur in crowded scenarios [18]. The irregularities and differences are efficiently spotted by applying the deep learning mechanism at crowded gatherings. 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. In the collective motion of the crowd gatherings is considered which makes tracking and analysis much easier [19].

The collective motion that is analysed in these video surveillances is mainly helpful to understand the behaviour of individuals which leads to a complete examination of the patterns in crowd behaviours. The study in [20], [21], shows the human behaviour methods of recognition as well as to object tracking. It is useful to track human motion on videos as object points that are more precise and accurate during evaluation. Emphasis on the crowded situations spotted at subways, railway stations, bus stations, metro stations and airports, where the density of the crowd is extremely intense. In these conditions object tracking among large gatherings is essential for accurate crowd analysis. However, crowd-based analysis is a challenging task and requires accurate and precise features for detection and other functions to be performed using surveillance videos [22]. The features that are extracted are used in a crowd feature optimization process applied to a fused feature methodology for better efficiency of the result produced.

3. PROPOSED METHODOLOGY

In this section, the mathematical modelling of the multi-level feature fusion forcrowdbehaviour; The methodology for crowd analysis uses multi-level features such as physical features that are collectively termed as crowd features. These crowd features are used for calculation and optimization which further result in infused features. The fused feature mechanism helps to provide an elaborate detailed description of every feature that is being used in this study. The proposed feature fusion mechanism works based on motion features. These crowd feature optimizations that are developed include a feature such as direction, motion, and speed. by detection of objects such as individuals in the crowd taken from surveillance videos.

3.1. Multilevel feature fusion

In this section of the research, we design and develop a multi-level feature fusion framework for crowd behaviour prediction; first level feature fusion utilizes the physical feature i.e. motion and physical appearances. Second level feature fusion utilizes the spatial based features and third level feature fusion utilizes the temporal feature description. Further, local region based region is used for classification; Figure 1 shows the proposed workflow.

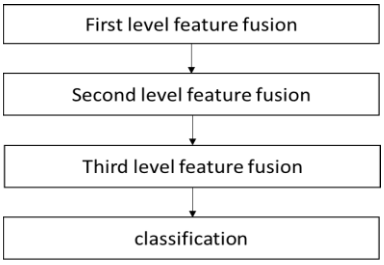
First level feature fusion
Second level feature fusion
Third level feature fusion
classification

Figure 1. Multi-level feature fusion workflow

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 $Q \in G^{i \times j}$ 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 Y is created for eliminating the inconsistencies between matrix $Q$and matrix Y. Therefore, the optimization for the inconsistencies is done as given (1):

$$\Xi_{Y,D} \|D\|_{\mathrm{J}} whereasb(Y) islessthanorequaltol, \; Q = Y + E \tag{1}$$

where $l << minimum(i,j)$which represents the subspace dimensions that are determined by a normalization function $\|D\|_J$ for the dataset that is given. A dimensionality matrix Y spatial matrix D is represented using the background and foreground of the video respectively. 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 based on the motion flow method. Therefore, considering the frame information, the Crowd Feature Objects is designed as (2).

$$\xi(h,k) = \left\{ \forall c(h,k) \in D_m | (F_h^2 + F_k^2)^{\frac{-1}{2}} isgreaterthanD_0 \right\} \tag{2}$$

In (2), D_0 is the threshold coefficient. This value remains constant and the foreground frame D_m has pixels that are denoted as $c(h,k)$. Motion heat maps are computed by the use of gradient maps. However, computed areas that are linked are firstly segmented after which they are highlighted with various colours. The heat maps give the exact location of the objects that are present at large public gatherings. A technique for filtering maybe used because only the moderate object points are retained whereas the other objects points are eliminated. The eliminated features are usually with zero dimensionality. The crowd feature optimization is used to evaluate the object points from the video frames to detect objects and generate feature fusion. For the generation of features such as shapes and motion, temporal and spatial domains are used. We use the crowd feature optimization to help distinguish between various gradients such as F_h, F_k, F_pthat are obtained through objects for object point (h,k,p). Therefore, these object points $b(h,k,p)$have a magnitude considering the temporal $\Phi(h,k,p)$ and spatial $\tau(h,k,p)$ whose directions are as (3).

$$\begin{aligned} & b(h,k,p) = \left(F_h^2 + F_k^2 + F_p^2\right)^{-1/2} \\ & \tau(h,k,p) = \cot(F_k . (F_h)^{-1}) \\ & \Phi\ (h,k,p) \ = \cot\ (\mathrm{F_p}.(\sqrt{F_h^2 + F_k^2})^{-1}) \end{aligned} \tag{3}$$

A split mechanism known as container split is used for splitting the frames among temporal and spatial containers. $\delta\Phi$ is the size of the temporal container and $\delta\tau$ is the size of the spatial container. After which the magnitude of the object point $b(h,k,p)$ is computed and normalized by the (4) angle $\emptyset$.

$$\emptyset = \delta\Phi(\cos\tau - \cos(\tau + \Delta\tau)) \tag{4}$$

If the object has a surrounding where a point is located then that point is known as (h', k', p'). Furthermore, the inconsistencies that are present between the object and point are referred to as $(\delta h, \delta k, \delta p)$. After this evaluation of the magnitude, there are histogram gradients Z_a that is defined by the (5).

$$Z_a(\overline{\Phi},\overline{\tau}) = \frac{1}{\emptyset}\, b(h', k', p')e^{-(\Delta h^2+\Delta k^2+\Delta p^2)} \tag{5}$$

The coefficient of spatial container τ is denoted as $\overline{\tau}$ and the coefficient of temporal container Φ is represented as $\overline{\Phi}$. Considering the video frames from the surveillance, a more dominant direction is taken into consideration by the selection of a spatial angle and temporal angle (τ,Φ). At this angle the magnitude of histogram gradient Z_a is maximum. Furthermore, there exists a rotational matrix T which is used for the evaluation of rotated objects in the environment, this is calculated by the (6).

$$\mathrm{T} = \begin{bmatrix} \cos\tau\cos\Phi & -sin\Phi & -cos\tau sin\Phi \\ sin\tau cos\Phi & \cos\tau & -sin\tau sin\Phi \\ \sin\Phi & 0 & \cos\Phi \end{bmatrix} \tag{6}$$

The magnitude of histogram gradient Z_a can be further computed using (7) considering the neighbourhood of the object in the spatial domain. Here, a constant rotational property is followed by the object as given:

$$Z_R(\overline{\tau}) = \sqrt{F_h^2 + F_k^2} \tag{7}$$

the gradients in the histogram of the spatial domain are done through computation of spatial container being split again, where the container size is defined as Δτ. Furthermore, by the use of (4) a new spatial angle can be obtained, the index value of this new angle is represented by $\overline{\tau}$. Two adjacent frames are used to compute the gradients of histogram based on an optical flow motion $\{w, s\}$. Where *w* and *s* are used to denote the horizontal and vertical optical flow. Similarly, there is a splitting of the temporal container where the size of the temporal container is defined as ΔΦ for the computation of optical motion flow. In this case, there is a re-computation of the temporal angle Φ as given (8).

$$\mathit{\Phi} = \cot\,(w(h,k).\,(s(h,j)^{-1}) \tag{8}$$

The index value is $\overline{\Phi}$ for the temporal angle Φ for the temporal container that is separated. Furthermore, the magnitude of histogram gradient is computed for the temporal angle $\overline{\Phi}$ and is given by (9):

$$Z_\chi\,(\overline{\Phi}) = \sqrt{w(h,k)^2 + s(h,k)^2} \tag{9}$$
$$Z_\Psi = [Z_a + Z_R + Z\chi]$$

the histogram gradients that are computed in this section comprise of three types and are evaluated using the fused feature methodology, the gradients that are obtained by considering the spatial angle and the gradients evaluated by temporal angle are defined as $Z_a + Z_R + Z\chi$. The three gradients that are obtained are summed up together as Z_Ψ. The sum of these gradients Z_Ψ is used to distinguish between abnormal and normal behaviour in crowd gatherings.

3.2. Second level feature fusion

In this section, the crowd features are mathematically represented in which the model that is 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 $J = \{j_1, j_2, \dots \dots . j_\varpi\}$in a dataset that is (10).

$$\Xi_{Q_c}E\,(Q_c, Y) \triangleq 2\varpi^{-1}\textstyle\sum_{m=1}^{\varpi}\{\left\|h_m - Q_c\rho_m\right\|_2^2 + \mathit{\Theta}\|\rho_m\| \tag{10}$$

Considering (13), the notation ϖ denotes the total number of crowd features that are trained whereas the Θ notation represents the factor that is controlling. $Y = \{\rho_1, \rho_2, \dots \dots \rho_\varpi\}$ is a set of all the features that are implied considering the total crowd features $H = \{h_1, h_2, \dots, h_\varpi\}$. Although, the coefficients Q_c and Y is used for normalizing the model of the data which has been shown as $\left|\left|h_m - Q_c\rho_m\right|\right|_2^2$ and $||\rho_m||_1$ represents regularization functions. Furthermore, $\mathbb{L}_1$ is the normalization that is used for the representation of the features. A multiplication operation is applied to the coefficient Q_c and the representation of the feature is done by using the coefficient ρ_m. On considering the above analysis, the features are constrained using a closed set and given by (11),

$$Q_c \triangleq \left\{Q_c \in G^{a\times b} \ \ such\ that \ \ \forall \text{ç} = 1, \dots \dots, \varpi,\ q_{\text{ç}}^K\ q_{\text{ç}}\ is\ less\ than\ or\ equal\ to\ 1\right\} \tag{11}$$

consider, a to represent the size of the feature and b denotes the weight of the feature size. In this case, ç *-th*of the column of the features that have been represented by $q_{\text{ç}}$. The weight of the feature size should be considered small. Therefore, the function of the object is defined by (12).

$$\min_{Q_c} E\,(Q_c, Y) + \partial \textstyle\sum_{\text{ç}=1}^{\varpi} M\left(\hat{\rho}_{\text{ç}}\right) \tag{12}$$

In the factor of balance is denoted as ∂ and $\sum_{\text{ç}=1}^{\varpi} M\left(\hat{\rho}_{\text{ç}}\right)$ is the function that has been used for the weight of the feature to be constrained. The pointer function has been defined in the given (13).

$$M\left(\hat{\rho}_{\text{ç}}\right) = \begin{cases} \qquad\quad 1 & otherwise \\ 0 \qquad\quad if\ \left\|\hat{\rho}_{\text{ç}}\right\|_1 & is\ less\ than\ or\ equal\ to\ \vartheta \end{cases} \tag{13}$$

considering the (16), ϑ is the factor for constraining that has a value that is nearing zero. $\hat{\rho}_{\text{ç}} = \{\rho_{1,\text{ç}}, \dots \dots., \rho_{\varpi,\text{ç}}\}$. Here, ç is under a range $[1, x]$. In this case, the pointer function is given as $\mathrm{M}\left(\hat{\rho}_{\text{ç}}\right)$ which is used to denote the total count of points in an object present in a frame. Therefore, the construction of the weight of the feature is done by consideration of all the crowd features $H = h_1, h_2, h_3, ..h_\varpi$ by the use of the fused feature for crowd feature optimization.

$$\underset{\varPsi_m}{\Xi}\ \|h_m - Q_c\,\varPsi_m\|_2^2 \qquad such\ that\ \|\,\varPsi_m\|_1\ Less\ than\ or\ equal\ to\ K_c \tag{14}$$

$\varPsi_m$ is represented as the weight of the feature coefficient for h_m and K_c is used to denote the parameter of the threshold deployed for the weight of feature constraints. Therefore, this model is constructed by the use of crowd feature optimization using fused feature methodology. This mechanism is used in training the data by using the fused feature methodology. The first phase of the crowd feature optimization is performed in the spatial domain.

3.3. Third level feature fusion

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 feature for the kth frame has been present in the mth block is represented as $H_{m,k} = \{k_1, k_2, \dots. k_\varpi\}$. Here, the crowd feature optimizer is defined as by $h_{\text{ç}}$, ç $\in [1, \varpi]$. The total number of optimizers used are denoted as the variable ϖ. Here, the set of features is evaluated by the use of a mechanism based on optical flow. This is represented as $Y_{m,k} = \{\rho_1, \dots \dots, \rho_\varpi\}$ and the set of features is given by x. Hence, a matrix of feature and coefficients is formed $\varpi \times x$ and the x *-th*column of the matrix has been expressed as ρ_x. Therefore, the set of columns is expressed as $\acute{a}_{m,lk} = \{\hat{\rho}_1, \dots \dots \dots., \hat{\rho}_x\}$. Then, the features of the motion are given as (15):

$$\text{Ø}_{m,k} = \{\max(\hat{\rho}_1), \dots \dots \dots., \max(\hat{\rho}_x)\} \tag{15}$$

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 have been denoted as k and every frame is divided into c blocks. Considering a feature of motion that is set for a specific m -*th*block represented by $\emptyset m = \{\emptyset_{m,1}, \dots \dots \emptyset_{m,k}\}$. On considering every frame of the$m$ -*th*blockbased on the motion is given as $Q_{\mathbb{R}m}$. Hence (16).

$$\underset{Q_{\mathbb{R}m}}{\Xi} \|\emptyset_m - Q_{\mathbb{R}m}\mathcal{A}_m\|_2^2 + \mathbb{R}\|\mathcal{A}_m\|_1 \tag{16}$$

In this case, temporal associations $Q_{\mathbb{R}m}$ are represented as $Q_{\mathbb{R}m} \in G^{e\times f}$, in which the size of the temporal association is denoted by the variable a. The parameter of the feature is denoted as $\mathbb{Y}_m$ for every frame in the *mth*block, such that $\mathcal{A}\quad_m = \{a_1, \dots \dots, a_k\}$. Therefore, the set for training the features on considering every block in the second phase is denoted as $\emptyset_m$. Hence, the set of feature parameters$\mathcal{A}\quad_m = \{a_1, \dots \dots, a_k\}$ respectively, is used for the classification in the coming steps of the process. Class labels are being used from the index values of the block. In this case, the features for motion are considered in the -*th*block of the testing phase and are given as $\emptyset'_m$. In this case, m $\in [1, c]$. Furthermore, the set of features that are used for testing is given by considering frame tests *k* which is denoted as Ø' and is given by the (17),

$$\underset{\mathcal{A}'_m}{\Xi} \|\emptyset'_m - Q_{\mathbb{R}m}\mathcal{A}'_m\|_2^2 \qquad such\ that\ \|\mathcal{A}'_m\|_1\ is\ less\ than\ or\ equal\ to\ K_m \tag{17}$$

in the (17), 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 $\|\emptyset'_m - Q_{\mathbb{R}m}\mathcal{A}'_m\|_2^2$whereas $\|\mathcal{A}'_m\|_1$ is used to represent the regularization term of the temporal domain. In this case, the block representations are given as $\mathcal{A}' = \{\mathcal{A}'_1, \dots \dots, \mathcal{A}'_c\}$. There is also a distinction that has to be done between abnormal and normal conduct by the use of classifiers that is given as $\mathcal{A}'_m \in G^{e\times f}$. This is the performance of the crowd feature optimizer using fused feature methodology in the temporal domain.

3.4. Classification based on the local region

In this section, the classification performed on the available data samples is being discussed. We consider each frame in this sub-region as a class and use a classifier to handle a problem based on multi-class. The already existing models use the mechanism of clustering for performing classification. However, the radius is different for every block and sometimes the classification for these blocks can be false. Although, if the classifier can provide the classification for every sub-region then this problem can be avoided. Therefore, the classification used on the crowd feature optimization along with fused feature methodology gives high accuracy considering the analysis of the behaviour. However, the classifier gives an output for every block is given by the (18),

$$probablity(i = 1|h) \approx C_{Y,\mathbb{R}}(f) \equiv (1 + \vartheta(Yf + \mathbb{R}))^{-1} \tag{18}$$

in the class labels are denoted as a variable i and $h$denoted the sample features. In this case, the function $f = f(h)$ is used for the estimation of the labels for the sample features h and a sigmoid function is used for the estimation of the probability $probablity(i = 1|h)$. In this case (19).

$$\underset{Z=Y,\mathbb{R}}{\Xi} J(Z) = -\sum_{m=1}^{X}\left(k_m log(Z_m)\right) + (1 - K_m)log(1 - K_m) \tag{19}$$

Consider,

$$\begin{aligned} K_m &= C_{Y,\mathbb{R}}(f_m) \\ K_m &= \begin{cases} if\ I_m\ is\ equal\ to\ unit\ \ (X'+1).(X'+2)^{-1} \\ if\ I_m is\ equal\ to\ negative\ unit 1(X''+2)^{-1} \end{cases} \end{aligned} \tag{20}$$

Considering (20), the count of samples used in training is denoted as X and the count of positive I_m is given as $X'$and count of negative I_m is given as $X''$and f_m is demonstrated for the prediction by $f(h_m)$. Considering the testing phase, the probability of the output is used to compute the abnormal conduct. In this case, the block score for the testing of abnormality is given by (21):

$$\wp_\varphi = 1 - probability(i = m|h) \tag{21}$$

considering (21), $\wp_\varphi$ is used to represent the abnormality of the testing. Therefore, each block respectively that is labelled and defined as $m \in [1, c]$. Here, $\wp_{score}$ provides a higher threshold value than the block and is recognized as a block of abnormal conduct. Therefore, there are various samples for testing that could have various values of threshold. This is the performance of the efficiency that is enhanced based on crowd feature optimization along with fused feature methodology.

4. PERFORMANCE EVALUATION

This section emphasizes experimental results that are obtained by the use of the crowd feature optimization (CFO) 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 is done through various visual analytics methods such as heat maps, and video frames. The analysis is performed on various patterns of crowd behaviour with different structures.

Motion features are mainly used for estimation of the tracking and detection system that is 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 is measured using various performance metrics such as precision, accuracy, area under the curve and recall, these metrics help in performing a comparative analysis. A comparative analysis is performed considering various datasets to which different algorithms are applied and these metrics evaluations are considered to find the more efficiently working model. The experimental results are demonstrated which indicate the performance efficiency of the fused features also considering the structure features, shapes with motion features. Elimination or dropout of unwanted features is also done while the required features are fused.

4.1. Dataset details

The proposed fused feature methodology that is applied to crowd features is tested on three different types of datasets: Violent-flow (VF) Dataset [23] and unusual crowd activity (UMN) Dataset [24]. The sub-sections stated below give a detailed description of these datasets. These video frames are depicted in the form of optical flow as well as heat maps for more accurate and precise identification and analysis. The first example consists of the video frame, optical flow, heat map and abnormal image of a crowded mix for analysis. The second example consists of the video frame, optical flow, heat map and abnormal image of a crowd present in a courtyard for analysis [25], [26]. The third example consists of the video frame, optical flow, heat map and abnormal image of a crowd present in a corridor for analysis. The result of analysing these video frames in this dataset are listed in this section.

4.1.1. Unusual crowd activity (UMN) data set

The unusual crowd activity (UMN) dataset consists of crowd scenarios that are of escape events taken from both outdoor and indoor scenarios. Each video that has the presence of a crowd initially starts with casual normal activities which later escalates into abnormal activities and ends in escape behaviour. The video frames are used to extract information that is categorized into two classes, where the first-class represents normal behaviour and the second class represents abnormal behaviour. The data utilization in this dataset constitutes 80% for training purposes and 20% for testing.

The basic video frames are extracted and then the video frame in this dataset is represented in the form of the motion heat map as well as optical flow portraying abnormal behaviour. The image datasets can be considered from courtyards, crowds and corridors which are further analyzed. Here fused feature and optimization technique can distinguish normal behaviour from abnormal behaviour by tracking individuals at public gatherings. Our model’s accuracy results that are obtained by the corridor video frame from the UMN dataset. The accuracy value obtained as experimental results is found to be 97.22%. The efficiency of the corridor video frame is depicted whose value is 96.67% and the classification error which has an accurate value of 0.02778. These are the calculated metrics values for the crowd frames in the UMN dataset. And we also represented the accuracy that is obtained through various traditional models with the use of violent-flow datasets.

The results obtained from simulation are received by using that dataset 80% for training and 20% for testing. The above figure uses the analysis techniques for crowd behaviour including irregularity-aware semi-supervised deep learning model (IASSLM), trajectory-based anomaly detection model (TADM), hybrid anomaly model (HAM), and integrated multi-level feature fusion (IMFF) using violent flow dataset. In this figure, the blue bar is used to represent the IASSLM algorithm, the orange bar is used to represent the TADM algorithm, the grey bar is used to represent the HAM algorithm and lastly, the yellow bar is used to represent the IMFF algorithm. Table 1 shows accuracy comparison for different methodologies which we considered and also highest accuracy for our model.

Table 1. Accuracy comparision for different methodologies

Methodology	Accuracy
Histogram of optical flow (HOF)	58.53
Oriented violent flow (OViF)	76.80
Histogram of gradients (HOG)	57.43
Combination of HOG and HOF (HNF)	56.52
Histogram of oriented tracklets (HOT)	82.30
Local binary tracklets (LBT)	81.90
Local trinary patterns (LTP)	71.53
Dense trajectories	78.21
Violent flow descriptor	81.3
3DCNN+SVM	90.60
ViF+Deep neural network	90.17
IMFF Model	99.56

Our model gives the exact accuracies of these algorithms making the distinction between them even more precise. Here 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. Figure 2 shows Video Frame, Optical Flow, Heat Map, and Abnormal activity Image of UMN Dataset which we considered.

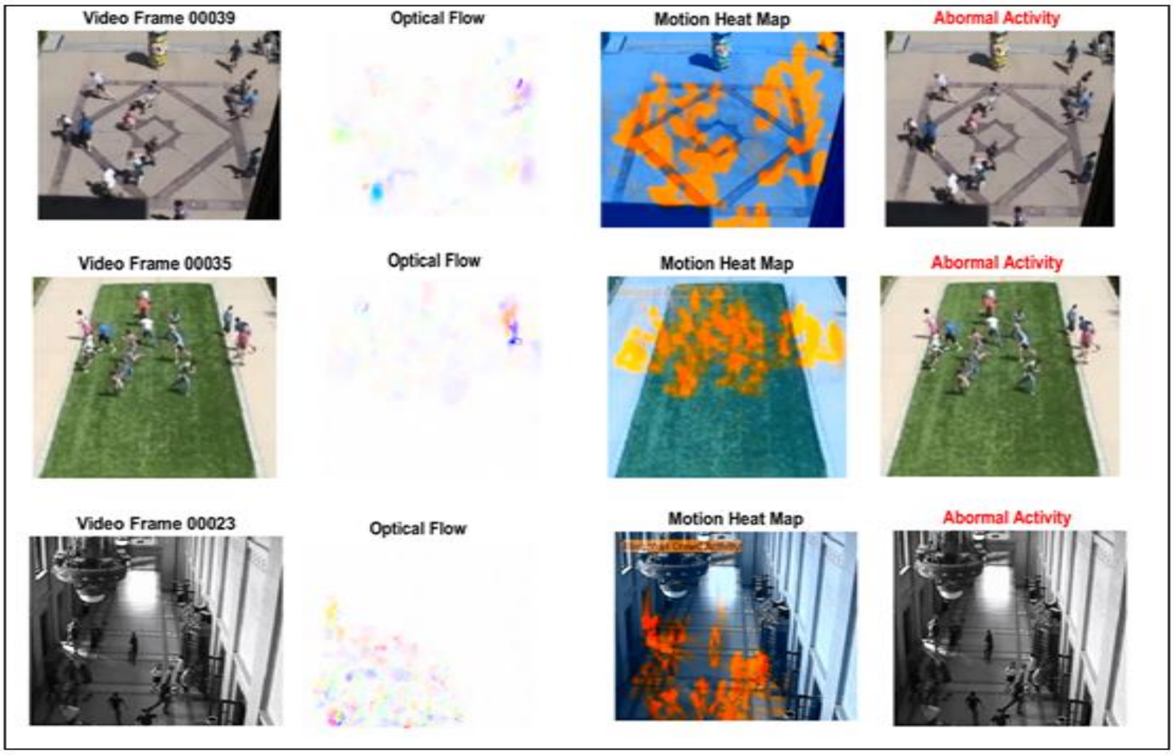

Figure 2. Video frame, optical flow, heat map and abnormal activity image of UMN dataset

Figure 3 shows Accuracy results considering the corridor scenario, Figure 4 shows efficiency results for the same corridor scenario, Figure 5 shows Accuracy results considering the courtyard scenario, Figure 6 shows efficiency results for the same courtyard scenario, Figure 7 shows classification error considering the corridor scenario Figure 8 shows classification error considering the courtyard data set scenario. Figure 9 shows the accuracy for different TADM, HAM, IASSLM, and IMFF models and this model gives the exact accuracies of these algorithms making the distinction between them even more precise. Here highest accuracy recorded is by the IMFF algorithm which is 99.56%. Table 2 represents the accuracy results of all the algorithms that perform crowd analysis using the violent-flow dataset. This dataset is used for the comparative analysis with various algorithms for the analysis on crowd behaviour that include oriented violent flow (OViF), cognitive deep model, local trinary patterns (LTP), local binary tracklets (LBT), 3DCNN+SVM, histogram of optical flow (HOF), histogram of gradients (HOG), combination of HOG and HOF (HNF), dense trajectories, histogram of oriented tracklets (HOT), ViF+deep neural network, and IMFF model. The highest accuracy in the comparative study of these algorithms for crowd analysis is found to be the IMFF model with an accuracy value of 99.56%. Moreover, Table 2 shows the comparison of FCDLF with IMFF in terms of precision and recall; furthermore, FDLCF observes precision and recall of 93 and 92 whereas IMFF observed 98.85 and 99.56 which is marginal improvisation.

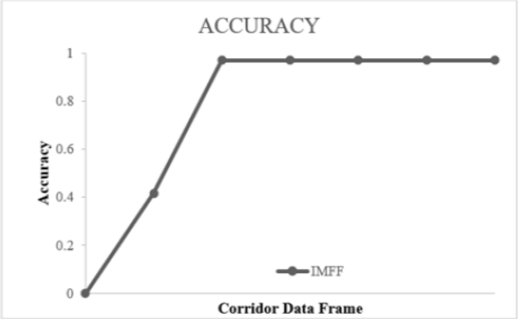

Figure 3. Accuracy results for the corridor scenario

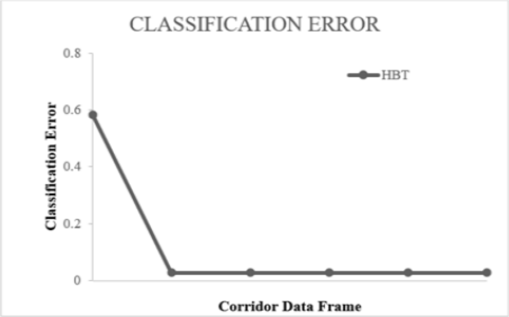

Figure 4. Efficiency results for the corridos scenario

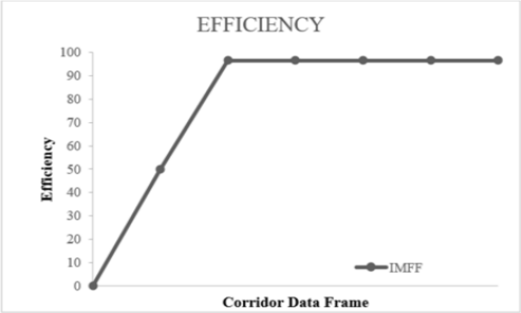

Figure 5. Accuracy results for the courtyard

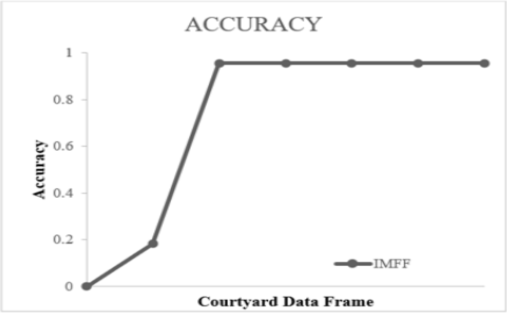

Figure 6. Efficiency results for the courtyyard

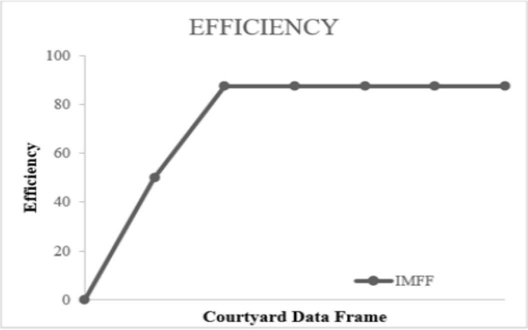

Figure 7. Classification error considering the corridor

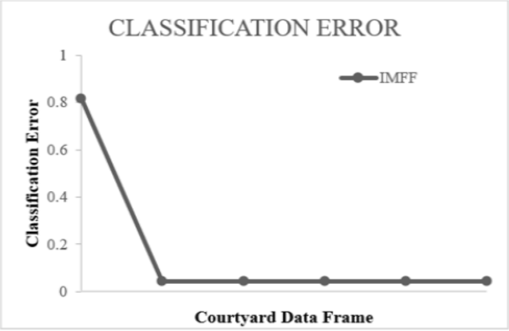

Figure 8. Classification error considering the courtyard

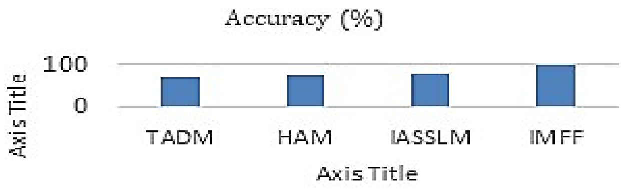

Figure 9. Accuracy for different models

Table 2. Precision and recall comparison

Methodologies	FCDLF	IMFF
Precision	93	98.75
Recall	92	98.85
Accuracy	92.9	99.56

5. CONCLUSION

The significance and importance of analyzing crowd data are implicated in this research work. Various crowd gatherings are considered and their behaviour which is the crowd conduct has to be studied for extraction of necessary features. Hence, considering the importance of crowd features, this research work introduces IMFF framework which tends to exploits the various features to enhance the prediction. IMFF comprises the three-level feature fusion and integration of classification using local region; the first level of feature fusion exploits the physical features like appearance, second and third level exploits the spatial and temporal in a respective manner. Further classification approach based on the local region is introduced for enhancement. IMFF is evaluated considering the UMN and violent flow dataset considering the different parameter; in the case of the UMN dataset accuracy, efficiency and classification error is evaluated with a different data frame. Further, in the case of Violent flow dataset accuracy precision and recall is considered; also comparative analysis is carried out with different existing models and it is observed that the proposed IMFF outperforms existing methodologies with the accuracy of 99%. Although IMFF simply outperforms the existing model, there is still several research gap as crowd behaviour is highly unpredictable and based on individuals. Hence other aspects can be looked for future analysis such as individual personnel mobility patterns to analyze the crowd behaviour.

REFERENCES

[1] R. Javid, M. M. Riaz, A. Ghafoor and N. I. Rao, "Direction, Velocity, Merging Probabilities and Shape Descriptors for Crowd Behavior Analysis," in IEEE Access, vol. 7, pp. 102561-102568, 2019, doi: 10.1109/ACCESS.2019.2929242.

[2] Y. Zhang, L. Qin, R. Ji, H. Yao, and Q. Huang, "Social attribute-aware force model: exploiting richness of interaction for abnormal crowd detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 7, pp. 1231–1245, 2015, doi: 10.1109/tcsvt.2014.2355711.

[3] M. Xu *et al.*, "Crowd simulation model integrating "physiology-psychology-physics" factors," *arXiv:1801.00216v1 [cs.MA]* 31 Dec 2017.

[4] T. Yin, L. Hoyet, M. Christie, M.-P. Cani and J. Pettré, "The One-Man-Crowd: Single User Generation of Crowd Motions Using Virtual Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2245-2255, May 2022, doi: 10.1109/TVCG.2022.3150507.

[5] Y. Li, "A Deep Spatiotemporal Perspective for Understanding Crowd Behavior," *IEEE Transactions on Multimedia*, vol. 20, no. 12, pp. 3289-3297, Dec. 2018, doi: 10.1109/TMM.2018.2834873.

[6] Z. Li, X. Yang, C. Wang, K. Ma, and C. Jiang, "Crowd-Learning: A Behavior-Based Verification Method in Software-Defined Vehicular Networks with MEC Framework," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1622-1639, 15 Jan.15, 2022, doi: 10.1109/JIOT.2021.3107581.

[7] H. Fradi, B. Levison, and Q. C. Pham, "Crowd behavior analysis using local mid-level visual descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 3, pp. 589-602, March 2017, doi: 10.1109/TCSVT.2016.2615443.

[8] L. Ge, Z. Yang, and W. Ji, "Crowd evolution method based on intelligence level clustering," *International Journal of Crowd Science*, vol. 5, no. 2, pp. 204-215, Aug. 2021, doi: 10.1108/IJCS-03-2021-0010.

[9] A. K. Pai, A. K. Karunakar, and U. Raghavendra, "Scene-independent motion pattern segmentation in crowded video scenes using spatio-angular density-based clustering," in *IEEE Access*, vol. 8, pp. 145984-145994, 2020, doi: 10.1109/ACCESS.2020.3015375.

[10] R. Wang, Q. Yu, B. Alzahrani, A. Barnawi, A. Alhindi, and M. Zhao, "The Limo-Powered Crowd Monitoring System: Deep Life Modeling for Dynamic Crowd With Edge-Based Information Cognition," *IEEE Sensors Journal*, vol. 22, no. 18, pp. 17666-17676, 15 Sept.15, 2022, doi: 10.1109/JSEN.2021.3080917.

[11] L. Chai, Y. Liu, W. Liu, G. Han, and S. He, "CrowdGAN: identity-free interactive crowd video generation and beyond," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 2856-2871, 2020, doi: 10.1109/TPAMI.2020.3043372.

[12] A. Mehmood, "Efficient anomaly detection in crowd videos using pre-trained 2D convolutional neural networks," *IEEE Access*, vol. 9, pp. 138283-138295, 2021, doi: 10.1109/ACCESS.2021.3118009.

[13] X. Li, M. Chen, and Q. Wang, "Quantifying and detecting collective motion in crowd scenes," in *IEEE Transactions on Image Processing*, vol. 29, pp. 5571-5583, 2020, doi: 10.1109/TIP.2020.2985284.

[14] R. Alotaibi, B. Alzahrani, R. Wang, T. Alafif, A. Barnawi, and L. Hu, "Performance comparison and analysis for large-scale crowd counting based on convolutional neural networks," in *IEEE Access*, vol. 8, pp. 204425-204432, 2020, doi: 10.1109/ACCESS.2020.3037395.

[15] S. Guo, Q. Bai, S. Gao, Y. Zhang, and A. Li, "An Analysis Method of Crowd Abnormal Behavior for Video Service Robot," *IEEE Access*, vol. 7, pp. 169577-169585, 2019, doi: 10.1109/ACCESS.2019.2954544

[16] Q. Wang, J. Gao, W. Lin, and Y. Yuan, "Learning from synthetic data for crowd counting in the wild," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2019, pp. 8198–8207, doi: 10.1109/CVPR.2019.00839.

[17] K.-J. Hsiao, K. S. Xu, J. Calder, and A. O. Hero, "Multicriteria similarity-based anomaly detection using Pareto depth analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1307–1321, Jun. 2016, doi: 10.1109/TNNLS.2015.2466686.

[18] S. Mondal, A. Roy, and S. Mandal, "A supervised trajectory anomaly detection using velocity and path deviation," in *Proceedings of International Conference on Frontiers in Computing and Systems*. Springer, 2021, pp. 777-784, doi: 10.1007/978-981-15-7834-2_72.

[19] M. Xu *et al.*, “Emotion-based crowd simulation model based on physical strength consumption for emergency scenarios,” *IEEE Transactions on Intelligent Transportation Systems,* pp. 1–15, 2020, doi: 10.1109/tits.2020.3000607.

[20] D. Helbing and P. Mukerji, “Crowd disasters as systemic failures: Analysis of the love parade disaster,” *EPJ Data Science*, vol. 1, no. 1, 2012, Art. no. 7, doi: 10.1140/epjds7.

[21] J. Shao, C. C. Loy, K. Kang, and X. Wang, “Slicing convolutional neural network for crowd video understanding,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* Las Vegas, NV, USA, Jun. 2016, pp. 5620-5628, doi: 10.1109/CVPR.2016.606.

[22] Q. Wang, M. Chen, F. Nie, and X. Li, “Detecting coherent groups in crowd scenes by multiview clustering,” *IEEE Transactions on Pattern Analysis and Adaptive Intelligence,* vol. 42, no. 1, pp. 46-58, 1 Jan. 2020, doi: 10.1109/TPAMI.2018.2875002.

[23] A. S. Aljaloud and H. Ullah, “IA-SSLM: irregularity-aware semi-supervised deep learning model for analyzing unusual events in crowds,” in *IEEE Access*, vol. 9, pp. 73327-73334, 2021, doi: 10.1109/ACCESS.2021.3081050.

[24] Y. Zhou, B. P. L. Lau, Z. Koh, C. Yuen, and B. K. K. Ng, “Understanding crowd behaviors in a social event by passive WiFi sensing and data mining,” *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4442-4454, May 2020, doi: 10.1109/JIOT.2020.2972062.

[25] Z. Ilyas, Z. Aziz, T. Qasim, N. Bhatti, and M. F. Hayat, “A hybrid deep network-based approach for crowd anomaly detection,” *Multimedia Tools Applications*, vol. 61, pp. 1-15, Mar. 2021, doi: 10.1007/s11042-021-10785-4.

[26] E. Varghese, S. M. Thampi, and S. Berretti, “A psychologically inspired fuzzy cognitive deep learning framework to predict crowd behavior,” in *IEEE Transactions on Affective Computing,* vol. 80, pp. 24053-24067, 2021, doi: 10.1109/TAFFC.2020.2987021.

BIOGRAPHIES OF AUTHORS

Manu Yadakere Murthygowda received his Bachelor degree from BGSIT, VTU during the year 2011, M.Tech. from BGSIT, VTU during the year 2015. Currently pursuing Ph.D. degree in ACU. He is havingmore than 8 years of Professional experience which includes Software Industry and teachingexperience. His areas of interests are big data security, cloud computing, AI and machine learning. He has published and presented papers in National and International conferences andjournals. He contributed as a reviewer for four IEEE International Conferences. He received Academic Excellence Award in the year from Institute of Scholars in the year 2021. He can be contacted at email: manuym@bgsit.ac.in.

Ravikumar Guralamata Krishnegowda received his Bachelor degree from Bangalore University during the year1996, M. Tech from Karnataka Regional Engineering College Surthkal (NITK) during the year2000 and Ph.D. from Dr MGR University, Chennai. He is working as a Professor and ResearchHead in Department of Computer Science Engineering, BGSIT. He had worked with iGATEGlobal solutions Bangalore, Wipro and also has worked with SJBIT as Prof and HOD of Dept ofCSE and ISE, having more than 20 years of Professional experience which includes Software Industry and teaching experience. His are a sofinterestsare data warehouse and business intelligence, multimedia, databases, AI, machine learning. He has published and presentedpapersin National and International conferences andjournals. He can be contacted at email: ravikumargk@bgsit.ac.in.

Shashikala Salekoppalu Venkataramu received her B.E degree in Computer Science from Mysore University, Karnataka in 1990 and M. Tech in Computer Science and Engineering from Visvesvaraya Technological University in 2005, Ph.D. degree from VTU. She is a Professor and Head in the department of computer science at B.G.S Institute of Technology; B.G.Nagar, Mandya and having totally 25 years of teaching experience. She has presented many papers in National level conferences and also presented many papers in international conferences; she has published 5 papers in International Journals. Her main research interests include computer networks, big data and data mining. She can be contacted at email: shashikala7@bgsit.ac.in.