International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 11, No.1 (January-2022)

http://dx.doi.org/10.12785/ijcds/110152

Fuzzy Hypergraph Modeling, Analysis and Prediction of Crimes

Dhanya P.M.[1], Ramkumar P.B.[2], Nirmaljith Cletus[3] and Paul Joy[4]

[1,3,4] *Department of Computer Science and Engineering, Rajagiri School of Engineering and Technology, Kerala, India.*
[2] *Department of Mathematics, Rajagiri School of Engineering and Technology, Kerala, India*

Received 07 August 2021, Revised 30 November 2021, Accepted 04 Jan 2022, Published 20 Jan 2022

Abstract: Hypergraphs are graphs in which more nodes are found in an edge as opposed to two nodes in a simple graph. In this work hypergraphs are created out of crime data and this is used to highlight areas with more crime. Various hypergraph morphological operations like dilation with respect to node, edge, erosion with respect to node, edge are applied which will result in crime data analysis. Moreover, the nodes and edges are fuzzified to make it a fuzzy hypergraph. This is pioneer work which models data using fuzzy hypergraph by applying morphological operations on crime data. Also this is a premier work in multilevel hypergraph. The aim of our work is the development of a novel prediction model which predicts crime behavior of a location using Lukasiewicz implication applied on a fuzzy multilevel hypergraph. Various parameters like proximity to ATMs, highways, shopping malls, railway stations, bus stations, literacy rate, urban/rural factor and the existing crime behavior of a location are considered for this crime prediction.

Keywords: Hypergraph, Dilation, Erosion, Morphology, Lukasiewicz

1. Introduction and Overview

With increasing population and diversity among the latitudes and longitudes all around the globe, it is important to understand the crimes and its nature in order to provide a better secure life for the people. Crime mapping is one of the effective ways to understand the major crime attributes like when and where the crime occurred. Such tools come helpful not only to the police officers but also to the public as they get an idea of a particular locality or a place.

This could really help in various ways like the equity of land, distribution of enforcement officers in the area, and also it could boost the sense of security among the public. If people have access to the data regarding crime that has happened in an area for a particular period of time, they could decide on how to make their residential or work-spaces secure and get a clear picture of the place too.

Since the beginning of mankind there has been crimes like theft, murder, cheating, sexual assaults etc. Crimes have always occurred around us in one way or another. One of the major reasons for the exponential growth in crime rates is due to the lack of knowledge regarding the crimes that are happening around us. The system which we developed provide crime information to the public and the respective authorities and thus help in reducing the number of crimes. This transparency helps in creating more self-reliance between community members and also the system will help in analyzing different types of crimes. For this our system has used multilevel fuzzy hypergraphs.

Hypergraph is a very effective mathematical model to analyze the data at a glance. It gives relevant information about the location and crimes based on a fitness value. Thus, based on this representation, we can retrieve and look upon certain features and get an idea on how secure the place is. Thus, people are being educated and made aware about their society. The statistical data helps in searching, sorting, visualizing and analyzing crime data based on our requirements. The tailored alerts based on the day or the location is another essential feature to make users updated about the locality and other vicinity.

This paper is organized with the following sections. Section 2 is the literature study where a number of works in crime prediction and hypergraphs are referred. Section 3 is the mathematical modeling of hypergraph used in this work. Various morphological operations on a hypergraph are discussed here. Section 4 shows how crime data can be mapped on to a two level fuzzy hypergraph. Section 5 and section 6 describes the architecture and design of the system developed. Section 7 is the implementation of the

E-mail address: dhanya_pm@rajagiritech.edu.in, ramkumar_pb@rajagiritech.edu.in, nirmaljithcletus@gmail.com, pauljoymaliakkals@gmail.com

prediction model developed followed by result analysis in section 8. Contributions of the work are given in section 9, followed by references.

2. Related Works

This section deals with some of the existing works in the field of crime analysis and also some works in the field of hypergraphs. The history of crime analysis [1] dates back to 1820s where a detective bureau consisting of many men were formed to analyze the crime patterns and solve criminal issues. Identifying crime zones [2] by k-means clustering and DB-Scan clustering method have considered only theft and robbery. The work has tracked the changes in crime rate from one year to the next and uses that to predict the future crime. Here crime prediction is done with an accuracy of 89%.

The crime analysis and prediction using data mining [3] is a method to identify the patterns and trends in crime. This system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. The crime detection in India [4] uses k-means clustering, Random Forest algorithm and Neural networks for classification and visualization using Google marker. This system is tried on a data set of crimes in the year 2001 to 2012 and accuracy is verified using WEKA tool.

The predictive modeling [5] of crime data set help the police and citizens to take necessary action to decrease the probability of occurrence of crime by studying the features that affect high crime rate. Crime prediction based on criminal hotspot [6] has found out patterns of criminal hotspots using Apriori algorithm and predicted crime types using Decision tree classifier and Naïve Bayes classifier. The data set used is Denver crime data set from two cities of US. Moreover, safe and dangerous neighborhoods are found based on age/sex. According to the findings, places with more female are safe places and those with more males are dangerous. Also places with middle aged people are safer than places with young people of age group 20-29. The system claims a prediction accuracy of 54%.

The crime prediction system using machine learning [7] predicts crime which will happen in different locations using K-neighbors classifier, Gaussian Naive Bayes, Multinomial Naive Bayes, Support Vector Classifier, Decision tree etc., out of which K-neighbors showed the maximum accuracy. We have considered some projects like Trulia, CrimeView, RAIDS Online, RiskAhead and SpotCrime.com.

CrimeView[8] is a crime analysis, mapping and reporting software. It is designed for the detailed study of patterns of crime as they relate to geography and time. The first, working CrimeView application was installed at Indio Police Department in California in August 1996.

RAIDS Online[9] is a free public crime map developed by BAIR Analytics. It aims to reduce information requests and improve trust between law enforcement entities and their public with data transparency and accuracy. The map can easily search for information and view nearby crime activity. Users can choose different data layers to see how demographics and socio-economic factors affect crime. Citizens can sign up for reports to help their neighborhood stay safe. RiskAhead is a free, multilingual, app-based risk assessment program [10] and crime map. One can easily and quickly report any incident with the questions and answers dialogue in the application and also get a quick overview of your destination or neighborhood by using the provided HeatMap functionality on RiskAhead's Crime Map. A variety of different incident reports can be found in the map, e.g. light-severe concentration of crimes, natural disasters and medical emergencies. There are features for quick overview of the area, reward points for incident reporting, safety of travelling women etc. Crime prediction works[11] using Naïve Bayes, Decision trees, Apriori algorithm, Random forest, Support Vector Machines, Neural networks, Linear Discriminant Analysis etc. are being compared in a detailed survey. They also discussed about ensemble methods. Most of the crime prediction methods are aiming at helping people in becoming more vigilant while traveling through crime prone areas. A tool for visualization [12] using R software for crimes based on the UK police data set is developed where the crime frequency report is shown using various charts and also a crime prediction using K-nearest neighbor classifier where the distance between the testing and training areas is taken as the distance factor while the Naïve Bayes classifier developed used the latitude, longitude and date attributes of the crime.

Survey [13] made on various NLP based methods identifies suspicious behaviors based on social network feeds and mobile chats, crime patterns and evidence based methods where crime location and evidences from the location are given importance, spatial and geo-location based methods where geographical features of the location are analyzed, prisoner based methods or crime reduction methods which deals with the mental health of the prisoners. This system is mainly focusing on further crime reduction by separating them from hazardous criminals and communication based methods which basically identifies crime organizations and crime leaders by tracking their mobile networks of communication. A system [14] which predicts crime in busy parts of the city, subways, pick-up, drop off locations, airports, parks etc. was also developed.

Since the system developed is based on hypergraphs, we will have an insight in to hypergraphs and the works done so far in that field. Hypergraphs[15] are very well explained and properties defined, where the authors have gone through basic structures, dualities, hypergraph similarities and hypergraph metrics with illustrative examples. Construction of hypergraphs from a formal context, its derivations, dilations and anti-dilations [16] are also developed. Random walk through hypergraph, Normalized hypergraph cut [17], spectral partitioning etc. are having applications in the field of natural language processing, computer networks etc. Hypergraphs are also used in the field of image processing[18], where the authors have used it for edge detection of images.

3. Mathematical Modeling

Hypergraphs [19] consists of hyperedges, where one hyperedge has more number of nodes.

The same is shown in Fig. 1.

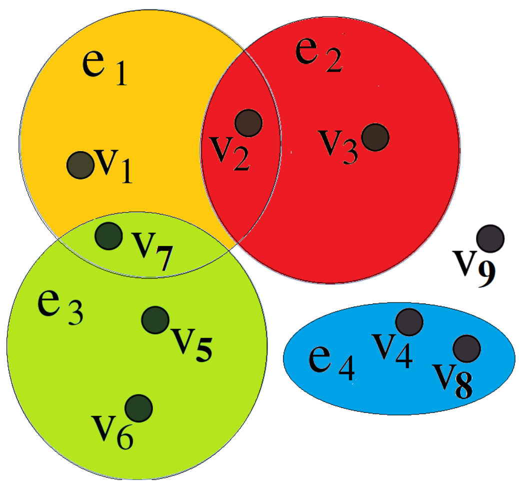


Figure 1. Hypergraph

From the figure we can see for the hypergraph From the figure we can see for the hypergraph $H = \{H_v, H_e\}$, the nodes v_1, v_2 and v_7 are parts of hyperedge e_1, while nodes v_2 and v_3 are parts of hyperedge e_3. Also nodes v_5, v_6, v_7 are parts of hyperedge e_3, while nodes v_4, v_8 are parts of hyperedge e_4. Node v_9 is not part of any of the hyperedges. In this work we are dealing with four morphological operations for which we need to define a sub hypergraph X of H. A sub hypergraph X is defined as $\{X \in H/\{X_v \in H_v; X_e \in H_e\}\}$ where $X_v$and X_e are the nodes and hyperedges of the sub hypergraph X and H_v and H_e are the nodes and hyperedges of the parent hypergraph H. Let us take X consisting of edge e_2 and the nodes within it as shown in Fig. 2(a). Dilation w.r.to node $\delta^n(X^e)$ [20] is a morphological operation on H, where it retrieves all the nodes in edges of X as shown in Fig. 2(b).

Dilation w.r.to edge $\delta^e(X^n)$ is a morphological operation ,where it retrieves all the hyperedges that contains nodes of X. This shown in Fig. 3. Erosion w.r.to edge $\varepsilon^e(X^n)$ [21],[22] is defined as the set of edges which contains only nodes of X,which is shown in Fig. 4(a) and Erosion w.r.to node is the set of nodes which are present only in X^e and not in $X^{e'}$ which can be seen in Fig. 4(b).

4. Crime data and fuzzy hypergraph

From the data set obtained from kalamassery Police station, India, a fuzzy hypergraph[21][23] can be modeled from the crime data as shown in Fig. 5 with an outer

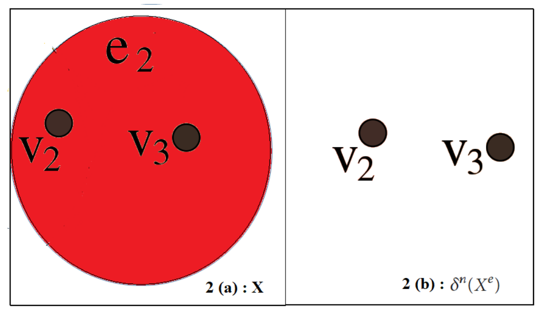


Figure 2. Sub hypergraph and dilation w.r.to node

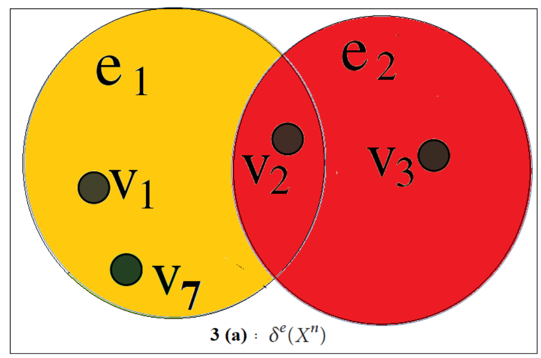


Figure 3. Dilation w.r.to edge

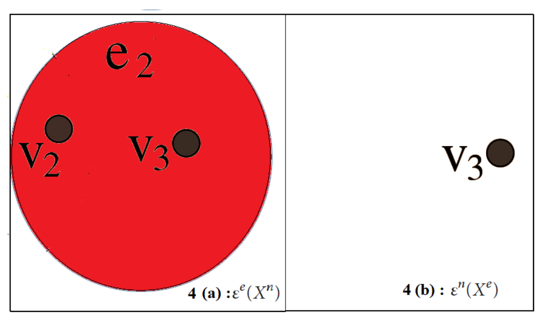


Figure 4. Erosions

hypergraph $H_{TP} = \{T^e, P^n, P_{cl}\}$ where T^e is the set of crime types $=\{T_1, T_2, ...T_m\}$ and $P^n = \{P_{cl1}, P_{cl2}...P_{cln}\}$ is the set of police stations. Also $P_{cl} = \{C^e, l^n\}$ is the inner fuzzy hypergraph where $C^e = \{C_1, C_2....C_r\}$ is the set of crimes and $l^n = \{l_1, l_2, ...l_s\}$ is the set of criminals involved in a crime.

Here, in crime-location hypergraph as given in Figs. 5 and 6, crime is taken as the hyperedge and location is taken as the node. Nodes and hyperedges are given fuzzy values. The fuzzy value of crime has been formulated using the number of occurrences of that crime and also the number of locations in which the crime has occurred. The fuzzy value of a location has been formulated using the number of crimes in that location and the number of types of crime in that location. A higher fuzzy value means the crime type has more chances of occurring or the location is prone to

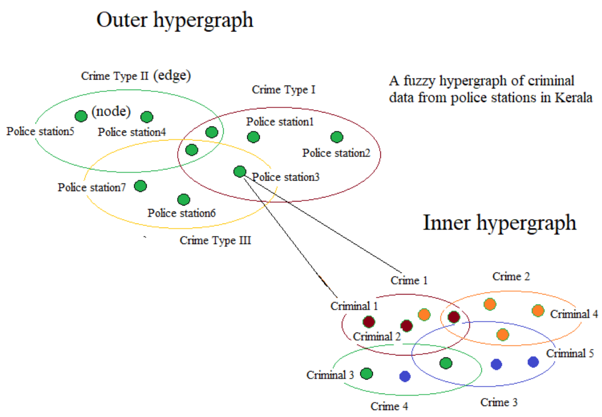


Figure 5. Two-level crime hypergraph

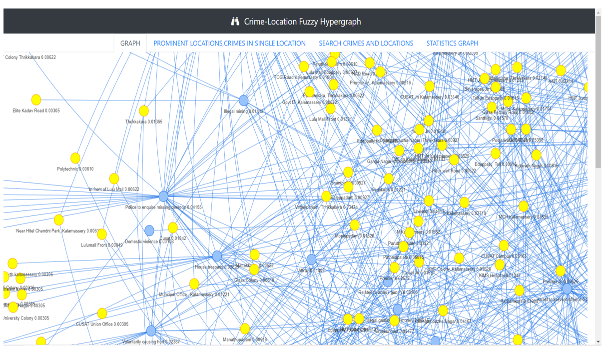


Figure 6. Crime-hypergraph

more number and types of crimes.

Here fuzzy value of a crime is calculated as:-

$$C_f = ((C_n/T_c) * w_1 + (C_l/T_l) * w_2) \quad (1)$$

where C_f = fuzzy value of a crime, C_n= total number of occurrences of a particular crime, T_c= total number of crimes, C_l =Total number of locations in which a particular crime occurred, T_l= total number of locations. w_1 and w_2 represents the weights. Since our interest is in the crimes which occur in more number of locations, we have taken $w_2 > w_1$. i.e., more weightage is given if a crime is occurring in more number of places. So w_1 is in the interval $0.1 \leq w_1 \leq 0.4$ and w_2 is in the interval $0.6 \leq w_2 \leq 1.0$.

Similarly, fuzzy value for a location is calculated as

$$L_f = ((T_{cl}/T_c) * \theta_1 + (C_{tl}/C_t) * \theta_2) \quad (2)$$

where L_f = fuzzy value of a location , T_{cl}=total number of crimes at a particular location, T_c= total number of crimes, C_{tl}=total number of types of crimes at a particular location, C_t= total number of crime types. θ_1 and θ_2 are the weights. Here $\theta_1 > \theta_2$, since we are giving more weight if a location is having more number of crimes. Here θ_1 is in the interval $0.6 \leq \theta_1 \leq 1.0$ and θ_2 is in the interval $0.1 \leq \theta_2 \leq 0.4$. Depending on the fuzzy value (α, β) cut is applied to

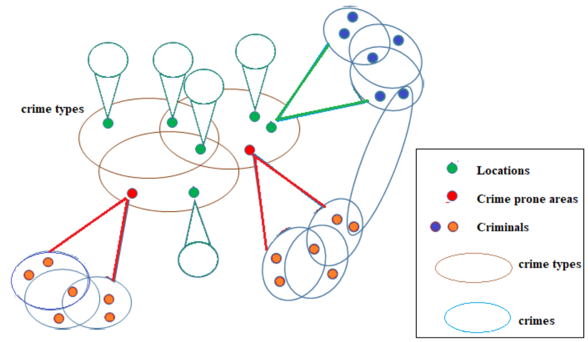


Figure 7. parent-hypergraph

the fuzzy hypergraph and sub hypergraph X_{TP} is generated which gives the set of edges with fuzzy value $> \theta$(where $\theta = 0.7$);
Now sub-hypergraph $X_{TP} = \{XT^e, XP^n, XP_{cl}\}$. where XT^e is the set of hyperedges (crime types), XP^n is the set of nodes (police stations) and XP_{cl} is the set of inner level hypergraphs (crimes and criminals). Now $XP_{cl} = \{XC^e, XI^n)\}$ is the inner level fuzzy hypergraph of sub-hypergraph X, where $XC^e = \{XC^1, XC^2, ...XC^r\}$ is the set of crimes and $XI^n = \{XI^1, XI^2....XI^n\}$ is the set of criminals involved in crimes of sub-hypergraph X. We have defined morphological dilation, erosion on this hypergraphs. So let the parent hypergraph be as given in Fig. 7. On calculating the fuzzy values and applying (α, β) cut, we get the sub-hypergraph as given in Fig. 8 which is the inner level hypergraph. Now morphological dilation, erosion [24] are applied on both inner and outer levels to get various results. This sub-hypergraph also show the crime prone areas.

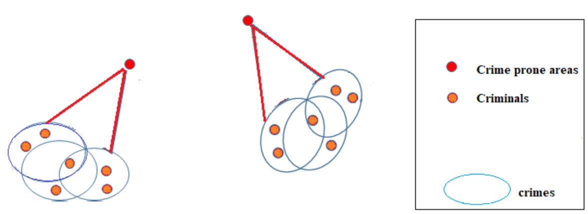


Figure 8. sub-hypergraph

Since the crime fuzzy hypergraph is a two-level one with an outer level and inner level hypergraphs, morphological operations like dilation and erosion can be applied on both the levels. Generally a dilation w.r.to edge[24] is defined as $\delta^e(X^n)$ which retrieves the set of edges in parent hypergraph where at least one node in sub-hypergraph is present. Applying this principle to the crime fuzzy sub-hypergraph on the inner level, we get the inner level dilation w.r.to the

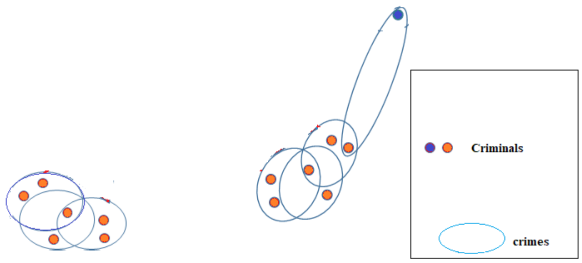


Figure 9. inner level dilation w.r.to edge

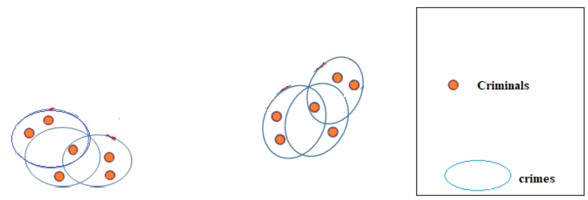


Figure 11. inner level erosion w.r.to edge

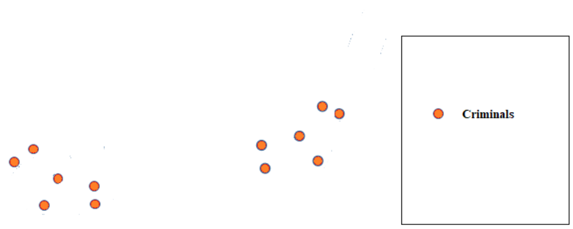


Figure 10. inner level dilation w.r.to node

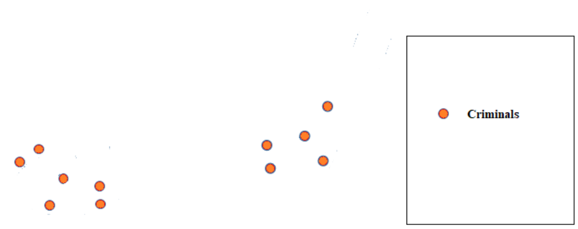


Figure 12. inner level erosion w.r.to node

hyperedge as

$$\delta_i^e(XI^n) = \{XC^e / XI^n \in XC^e\} \tag{3}$$

where XC^e is the set of hyperedges where at least one XI^n is present. In the scenario of crimes, it is the set of all crimes in which at least one criminal in the set XI^n is involved. More precisely it can be explained as the set of all crimes in which criminals of crime prone areas are involved and is shown in the Fig. 9. As per (3), and Fig. 9, we can see that some extra crimes other than those in crime prone areas are also retrieved since criminals in crime prone areas are also involved in that crime. This graph operation has given a very dilated result.

Similarly dilation w.r.to node $\delta^n(X^e)$ retrieves the set of nodes in parent hypergraph which are present in all edges of the sub-hypergraph. Applying this principle to crime hypergraph in the inner level, we get inner level dilation w.r.to node as

$$\delta_i^n(XC^e) = \{XI^n / XI^n \in XC^e\} \tag{4}$$

Where it retrieves all nodes present in all edges in XC^e. In the scenario of crimes, it is the set of all criminals XI^n involved in crimes XC^e of crime prone areas. This is shown in Fig. 10. As per (4), Fig. 10 shows all criminals in crime prone areas.

Just as we have defined both dilations on inner level, we can have erosions also on the inner level. Erosion w.r.to edge $\varepsilon^e(X^n)$ is generally defined the set of edges in parent hypergraph which has only nodes in sub-hypergraph. All edges with overlapping nodes with the parent hypergraph are skipped. Inner level morphological erosion w.r.to hyperedge is

$$\varepsilon_i^e(XI^n) = \{XC^e / XC^e \in T^e\} \tag{5}$$

and

$$\forall XI^n / XI^n \in XC^e \cap XI^{n'} \notin XC^e \tag{6}$$

This retrieves all crimes in which only criminals of crime prone areas are involved. Equations (5) and (6) shows what all edges are to included and excluded in the result. The output of this operation is shown in Fig. 11.

Now let us see how inner level erosion w.r.to node is done. Erosion w.r.to node $\varepsilon^n(X^e)$ is generally defined as the set of nodes which are part of sub-hypergraph X and not overlap with any edges of the parent hypergraph. Thus inner level erosion w.r.to node is

$$\varepsilon_i^n(XC^e) = \{XI^n / XI^n \in XC^e \cap XI^n \in XC^e\} \tag{7}$$

As per (7), criminals involved in crimes of crime prone areas only and not involved in any other crimes are retrieved. The same is shown in Fig. 12.

Now let us check the results of outer level morphological operations on the hypergraphs. Outer level dilation w.r.to edge is defined as

$$\delta_o^e(XP^n) = \{T^e / XP^n \in T^e\} \tag{8}$$

According to (8), this will retrieve all crime types from the parent hypergraph, where at least one police stations in crime prone areas are included as in Fig. 13.

Outer level dilation w.r.to node is defined as the police

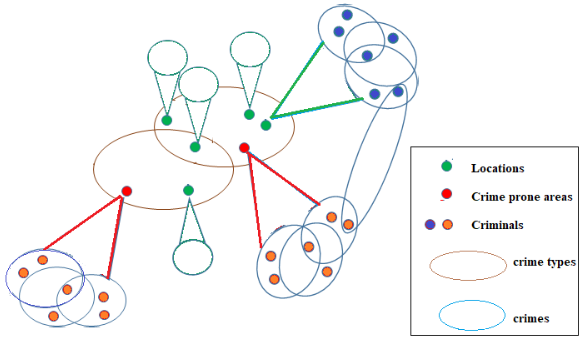


Figure 13. outer level dilation w.r.to edge

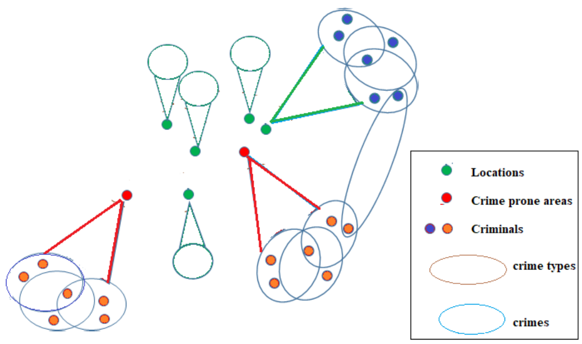


Figure 14. outer level dilation w.r.to node

stations which are coming under the same crime types where crime prone areas are involved, which is demonstrated in Fig. 14. and can be represented as (9).

$$\delta_o^n(XT^e) = \{P^n / P^n \in XT^e\} \tag{9}$$

Similarly erosions can be defined on the outer level hypergraph, where outer level erosion w.r.to edge is shown as

$$\varepsilon_o^e(XP^n) = \{X^e / XP^n \in X^e; XP^n \notin X^{e'}\} \tag{10}$$

Thus (10) retrieves all crime types from the parent hypergraph which consists of only police stations in crime prone areas. Also outer level erosion w.r.to node is defined as

$$\varepsilon_o^n(XP^e) = \{XP^n / XP^n \in X^e; XP^n \notin X^{e'}\} \tag{11}$$

This retrieves all police stations which are part of crime types of crime prone areas and not overlapping with any non-crime types in which non crime prone areas are located.

In this section we have successfully defined eight different types of morphological operations within which, four are on the outer level hypergraph and four are on the inner level hypergraph. Out of four operations in one level, two of them are dilations and the rest two are erosions. Out of two dilations in one level one is for retrieving nodes and the other is for retrieving edges.

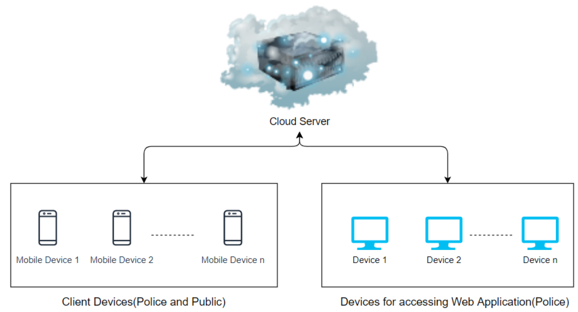


Figure 15. System Architecture

5. SYSTEM ARCHITECTURE

A. *Cloud Server: Database*

The database used for the project is a real-time database and is hosted on firebase. Firebase is aobile and is a web application development platform developed by Firebase, Google. This Database is a NoSQL cloud database and data is stored as JSON and synchronized in real time to every connected client. It also provides cross-platform functionalities that are required for our project as the project is developed for both android and web. Firebase console provides a set of features such as authentication and data transfer analysis features which help us in the project development. Features like real time-database, cloud messaging and authentication were used for the project development. To increases the scale of the database, it can be upgraded using various plans available on the firebase.

B. *Client Device: Mobile Application*

The front-end is an android application that works on both the data received and data written. The minimum version required is Android 6.0 marshmallow. Users can use the crime mapping application by installing the app on their android mobile. Multiple mobile users can access the functionalities of the app at the same time.

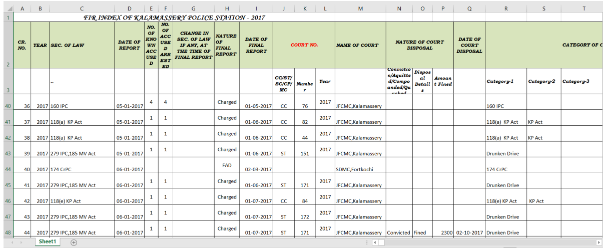
C. *Web Application*

As in Fig.15, the front-end of web application contains scripts used to take snapshot of Node, Edge, CrimeDataDetail and Location from the database and html code to render the data. The hypergraph rendered by the front end is done with the help of vis.js. The searching crime and location has an auto suggest feature provided by scripts. The statistical graphs are generated by google charts.

6. Design of the System

A. *Firebase*

The Firebase Real time Database is a cloud-hosted database. Data is stored as JSON and synchronized in real time to every connected client. When you build cross-platform apps with our iOS, Android, and JavaScript SDKs,



FIR INDEX OF KALAMASSERY POLICE STATION - 2017

CR. NO.	YEAR	SEC. OF LAW	DATE OF REPORT	NO. OF KNOWN ACCUSED	NO. OF ACCUSED ARRESTED	CHANGE IN SEC. OF LAW IF ANY, AT THE TIME OF FINAL REPORT	NATURE OF FINAL REPORT	DATE OF FINAL REPORT	COURT NO.			NAME OF COURT	NATURE OF COURT DISPOSAL			DATE OF COURT DISPOSAL	CATEGORY OF C		
		..							CC/ST/SC/CP/MC	Number	Year		Conviction/Aquitted/Compounded/Quashed	Disposal Details	Amount Fined		Category-1	Category-2	Category-3
36	2017	160 IPC	05-01-2017	4	4		Charged	01-05-2017	CC	76	2017	JFCMC,Kalamassery					160 IPC		
37	2017	118(a) KP Act	05-01-2017	1	1		Charged	01-06-2017	CC	82	2017	JFCMC,Kalamassery					118(a) KP Act	KP Act	
38	2017	118(a) KP Act	05-01-2017	1	1		Charged	01-06-2017	CC	44	2017	JFCMC,Kalamassery					118(a) KP Act	KP Act	
39	2017	279 IPC,185 MV Act	05-01-2017	1	1		Charged	01-06-2017	ST	151	2017	JFCMC,Kalamassery					Drunken Drive		
40	2017	174 CrPC	06-01-2017				FAD	02-03-2017				SDMC,Fortkochi					174 CrPC		
41	2017	279 IPC,185 MV Act	06-01-2017	1	1		Charged	01-06-2017	ST	171	2017	JFCMC,Kalamassery					Drunken Drive		
42	2017	118(e) KP Act	06-01-2017	1	1		Charged	01-07-2017	CC	84	2017	JFCMC,Kalamassery					118(e) KP Act	KP Act	
43	2017	279 IPC,185 MV Act	06-01-2017	1	1		Charged	01-07-2017	ST	172	2017	JFCMC,Kalamassery					Drunken Drive		
44	2017	279 IPC,185 MV Act	06-01-2017	1	1		Charged	01-07-2017	ST	171	2017	JFCMC,Kalamassery	Convicted	Fined	2300	02-10-2017	Drunken Drive		

Figure 16. Excel sheet from Police Station

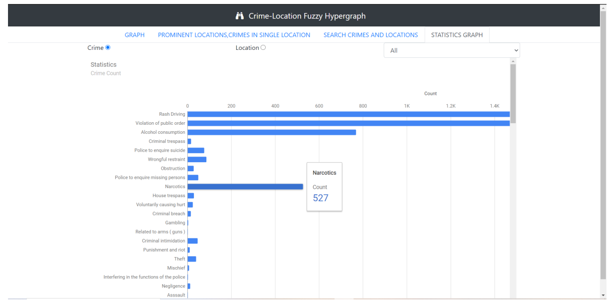


Figure 17. Crime Statistics

all of clients share one Real time Database instance and automatically receive updates with the newest data. NoSQL cloud database is used for data storage. Data is synced across all clients in real time, and remains available when your app goes offline. Data is passed and received through JSON packets.

B. Police

The data given by the Police department as shown in Fig. 16, is in the form of excel sheet and is first ordered based on date of occurrence. Irrelevant data is removed and data sheet is converted to csv format. Using csv to JSON converter, file is converted to required format to store it in the database. Dataset is imported to the realtime database named "crimemapping".

C. Features of the system

The system is divided in to different sections like a) Crime mapping - which map and display the crimes that have happened on to Google maps based on date and crime type filters b) Crime Reporting which is an option for the users to report the crimes to the police c) Statistics- which gives the statistics of crimes based on date, crime type, locality etc. d) Notification- where the police officials can notify the public about various important messages e) hypergraph - displays the fuzzy hypergraph view of the crimes and locations f) A prediction model based on Lukasiewicz implication on the multilevel fuzzy hypergraph.

The crime statistics feature as shown in Fig .17 shows the graphical statistics of crimes in the form of a bar graph. When the user selects location option graphical representation of the number of occurrences of crimes at different locations is displayed as in Fig. 18 and when the crime option, the graphical representation of the number of occurrences of all different types of crimes will be shown as in Fig. 19. There is also a filter to display the statistics of various years.

D. Prominent Locations And Crimes In Single Location

The Fig. 20 shows the crimes which occurred only in one location and the locations with the maximum variety of crimes. The first one are the edges in the inner level hypergraph which are connected to only one node in the outer level hypergraph. The later is the nodes in the outer levels which are part of maximum edges in the outer level hypergraph.

E. Statistics

The application user can view the statistics of the crimes based on 4 types:

- Crime type vs Total number of crimes.
- Crime type vs number of crime 2016.
- Crime type vs number of crime 2017.
- Crime type vs number of crime 2018.
- Place vs number of crime.

F. Notification

Notification feature notifies the application users through the android notification. Notification works as a service and hence will work even if the application is closed. Users are notified using two methods: a) Users are notified if the current system time and date on their mobile is equal to the time and date of the predefined set of notifications available on the database. Example- World Environment day or possible crime that can occur on a particular festival season like pick-pocketing etc. b) Users are also notified in real-time using google cloud messaging from the firebase console.

7. Prediction of Crime Prone Area

Various parameters are considered for predicting the crime in a particular location. In fact there are many factors which can influence the occurrence of a crime. Some parameters play a significant role in the crime rate. This system which we developed mainly focus on geographic and socio-economic parameters namely proximity to ATMs which we will denote as p_a, proximity to highways which is here denoted by the symbol p_h, proximity to rail way station which is denoted by p_r. proximity to bus station which is denoted as p_b, proximity to malls denoted by p_m and also density of population denoted by d. All these parameters increases the crime rate. The rate of education denoted by e, urban/rural classification shown with u also affects the crime rate. Considering these parameters, the Lukasiewicz implication can be applied to find the fuzzy value of a location.

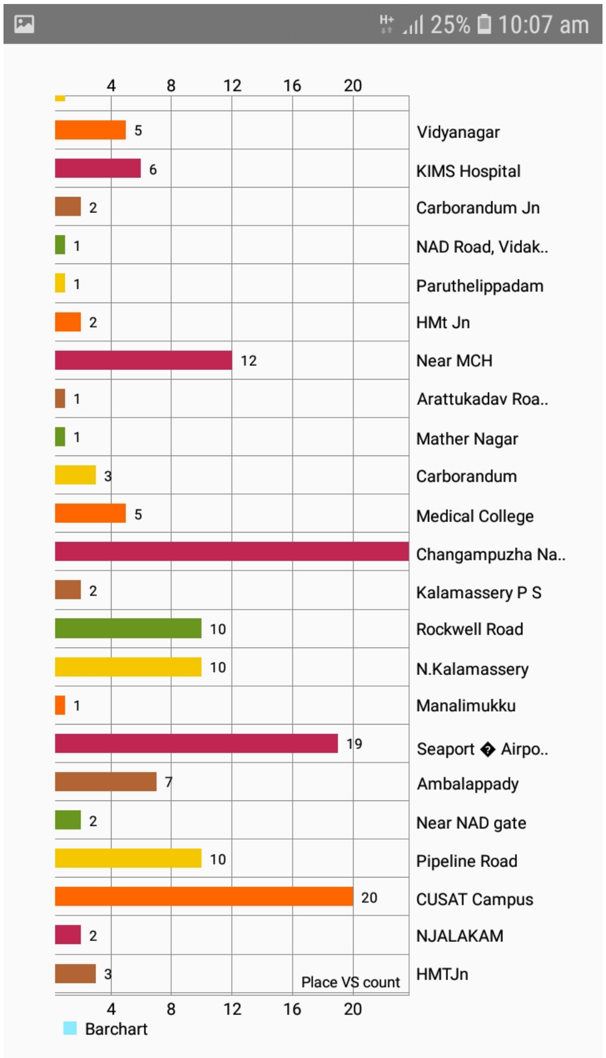


Figure 18. Place vs Number of Crimes

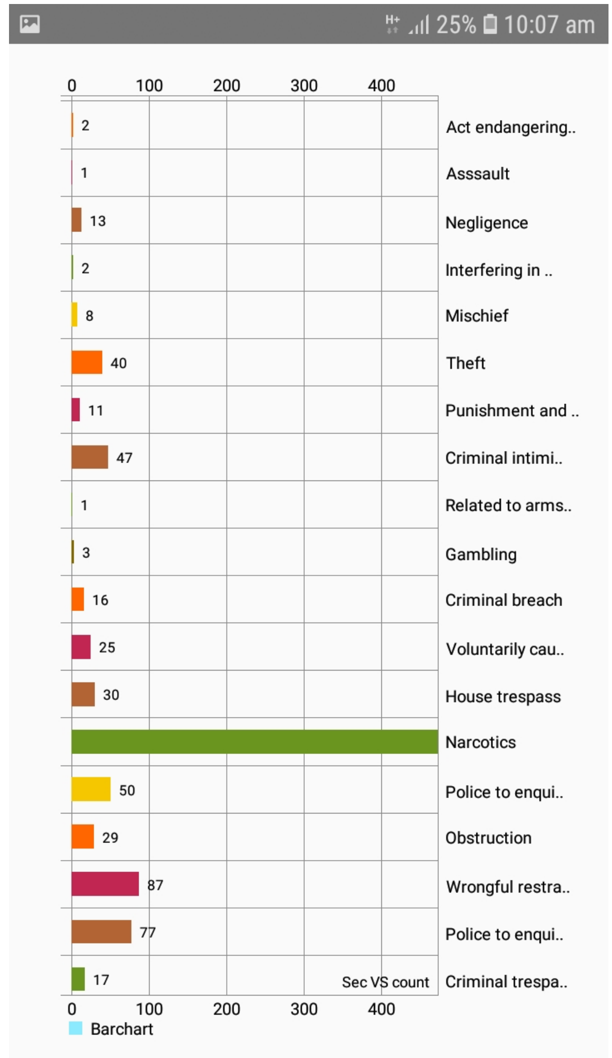


Figure 19. Crime type vs Number of Crimes

Lukasiewicz implication for a location with proximity to ATM p_a and proximity to highway p_h can be written as

$$f_{\rightarrow}(p_a, p_h) = min\{1, 1 - p_a + p_h\} \tag{12}$$

Similarly $f_{\rightarrow}(p_r, p_b) = \min\{1, 1 - p_r + p_b\}$ where p_r is the proximity to railway station and p_b is the proximity to bus station. Also $f_{\rightarrow}(p_m, d) = \min\{1, 1 - pm + d\}$ where p_m is the proximity to malls and d is the population density. The system also considered the parameters e which represents literacy rate and parameter u which represents urban factor of the area and found $f_{\rightarrow}(e, u) = \min\{1, 1 - e + u\}$. This is continued to the next level as given in Fig.21 and we found $f_{\rightarrow}(p_a, p_h, p_r, p_b) = \min\{1, 1 - f_{\rightarrow}(p_a, p_h) + f_{\rightarrow}(p_r, p_b)$. This is repeated until we find $f \rightarrow (p_a, p_h, p_r, p_b, p_m, d, e, u)$ which is the Lukasiewicz implication of fuzzy value of a location considering all the above parameters.

The algorithm 1 is the training phase where for all crime prone areas in the data set, Lukasiewicz implication is applied and a fuzzy value is calculated based on 8 parameters of each location. Algorithm 2 is the testing phase where for each new location in S2, we calculate the Lukasiewicz implication and find its k nearest neighbors and assign it to the class based on majority neighbors.

Predicting crime rate Similarly crime rate in a particular year is predicted using the previous years data by two methods

- Prediction-by-means where given year as x and crime rate as y, over many years, crime rate of a future year is predicted as $y(x)$ = the mean of all y.
- Legrange's interpolation- Let x_0, x_1, x_2 .... be the years and y_0, y_1, y_2.... be the respective crime rates obtained by the morphological operations applied earlier then $y(x) = (((x - x_1)(x - x_2)....)/(x_0 - x_1)(x_0 - x_2.....)y_0 +$

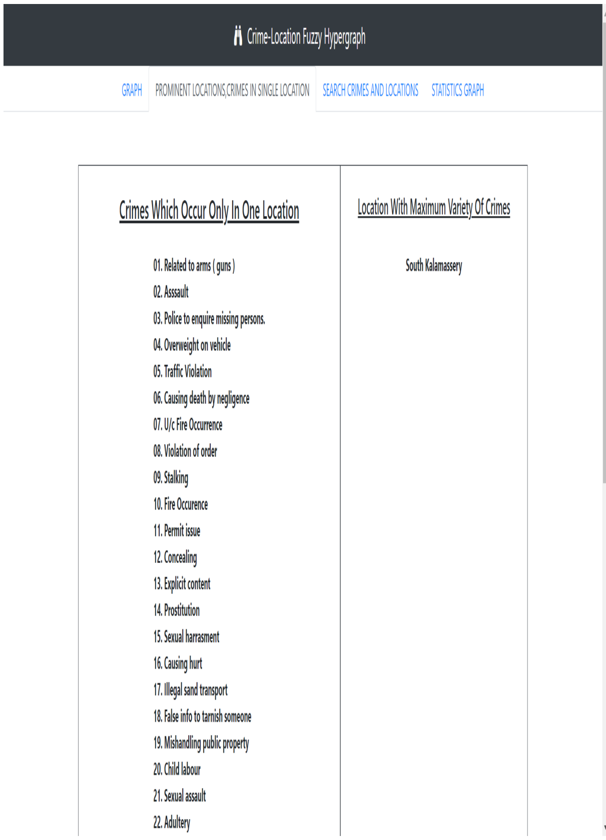


Figure 20. Rare crimes

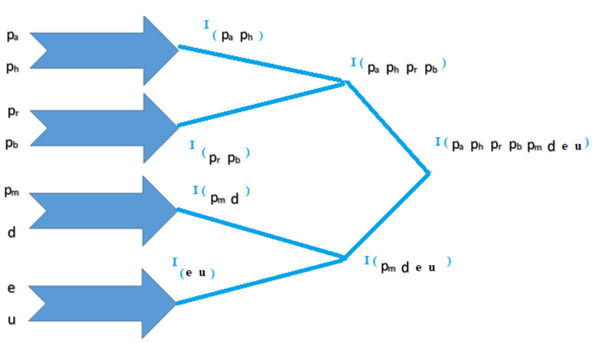


Figure 21. Lukasiewicz

$((x-x_0)(x-x_2)....)/(x_1-x_0)(x_1-x_2).....)y_1.......$
Where x is a future year and y is the crime rate expected in that year.

8. RESULT ANALYSIS

The data set obtained from the Police department is divided in to two parts where the set 1 consists of 75% of the data and set 2 consists of 25% of the data. Table.1. shows Lukasiewicz implication of $f_{\rightarrow}(p_a, p_h)$, where p_a stands for the fuzzy value of a location for its proximity to ATM and p_h shows its fuzzy value for its proximity to highways.

Algorithm 1 Crime prediction of a location(Training phase)

procedure FuzzyPredict($S1$) ▷ S1 is the data set(2016,2017)
  Threshold $\theta = 0.25$;
  $L_{chigh} \in S1 \leftarrow crime\_rate > \theta * average\_crime\_rate$
  $L_{clow} = S1 - L_{chigh}$

  while $r \neq 0$ **do** ▷ r is the number of records in L_{chigh}
  $f_{\rightarrow}(p_a, p_h) = \min\{1, 1 - p_a + p_h\}$
  $f_{\rightarrow}(p_r, p_b) = \min\{1, 1 - p_r + p_b\}$
  $f_{\rightarrow}(p_m, d) = \min\{1, 1 - p_m + d\}$
  $f_{\rightarrow}(e, u) = \min\{1, 1 - e + u\}$.
  $f_{\rightarrow}(p_a, p_h, p_r, p_b) = \min\{1, 1 - \{f_{\rightarrow}(p_a, p_h)\} + \{f_{\rightarrow}(p_r, p_b)\}\}$
  $f_{\rightarrow}(p_m, d, e, u) = \min\{1, 1 - \{f_{\rightarrow}(p_m, d)\} + \{f_{\rightarrow}(e, u)\}\}$
  $L_v = f_{\rightarrow}(p_a, p_h, p_r, p_b, p_m, d, e, u) = \min\{1, 1 - \{f_{\rightarrow}(p_a, p_h, p_r, p_b)\} + \{f_{\rightarrow}(p_m, d, e, u)\}\}$
  return L_v ▷ L_v is the fuzzy value of a location

Algorithm 2 Crime prediction of a location(Testing phase)

procedure FuzzyPredict($S2$) ▷ S2 is the data set(2018)
  L_{cr} = crime prone area;
  L_{ncr} = non crime prone area;

  while $s \neq 0$ **do** ▷ s is the number of records in $S2$
  $L_v = f_{\rightarrow}(p_a, p_h, p_r, p_b, p_m, d, e, u) = \min\{1, 1 - \{f_{\rightarrow}(p_a, p_h, p_r, p_b)\} + \{f_{\rightarrow}(p_m, d, e, u)\}\}$
  k=5;
  Compare L_v with all records in L_{chigh} and L_{clow}.
  neighbours(L_v) = Find k values nearer to L_v

  if neighbours(L_v) $\in L_{cr}$ **then**
  label(L_v) = L_{cr}

  else
  label(L_v) = L_{ncr}
  return *label*(L_v) ▷ L_v is the fuzzy value of a location

According to the crime analysis of various locations, it is found that crimes are frequent in areas near to ATMs and in areas near to highways. As given in Table I., if a location is with an ATM within its 200m proximity, its fuzzy value is high(0.9) and if the location is with ATMs more than 1km proximity, then its fuzzy value is low(0.6). Eq. (2) is also considered for assigning these fuzzy values. Also if a location is within 100m of highways its fuzzy value is high (0.7) and if the location is more than 1 km away from highway, its fuzzy value is low (0.2). So $f_{\rightarrow}(p_a, p_h)$ of a location which is within 200m proximity to ATM and within 100m proximity to highway is 0.8, which is a high value. All locations with $f_{\rightarrow}(p_a, p_h) >= 0.7$ are considered to be unsafe locations. As per this method several locations

TABLE I. lukasiewicz implication of $f_{\rightarrow}(p_a, p_h)$

Proximity to highways and ATMs	**ATM within**	<= 200m	> 200 m<= 500m	> 500m <= 1km	> 1 km
	$p_a \rightarrow$	0.9	0.8	0.7	0.6
Highway	$p_h \downarrow$				
<= 100m	0.7	0.8	0.9	1	1
> 100m <= 200m	0.6	0.7	0.8	0.9	1
> 200m <= 1km	0.3	0.4	0.5	0.6	0.7
> 1km	0.2	0.3	0.4	0.5	0.6

TABLE III. lukasiewicz implication of $f_{\rightarrow}(p_m, p_d)$

Proximity to Mall, density	**Mall**	<= 50m	> 50 m<= 200m	> 200m <= 500m	> 500m
	$p_r \rightarrow$	0.9	0.8	0.5	0.4
Density/sq km	$p_b \downarrow$				
> 1500	0.9	0.8	1	1	1
> 1000 <= 1500	0.7	0.8	0.8	1	1
> 500 <= 1000	0.2	0.3	0.4	0.7	0.8
< 500	0.1	0.2	0.3	0.6	0.7

TABLE II. lukasiewicz implication of $f_{\rightarrow}(p_r, p_b)$

Proximity to railways, Bus stands	**Railway within**	<= 100 m	> 100 m<= 500m	> 500m <= 1km	> 1 km
	$p_r \rightarrow$	0.9	0.8	0.7	0.5
Bus stand	$p_b \downarrow$				
<= 100m	0.7	0.8	0.9	1	1
> 100m <= 200m	0.6	0.7	0.8	1	1
> 200m <= 1km	0.4	0.5	0.6	0.7	0.9
> 1km	0.3	0.4	0.5	0.6	0.8

TABLE IV. lukasiewicz implication of $f_{\rightarrow}(e, u)$

literacy, urban factors	**literacy**	>= 90%	< 90% >= 70%	< 70% $\geq$ 50%	< 50%
	$p_r \rightarrow$	0.9	0.7	0.5	0.4
urban factor	$p_b \downarrow$				
city	0.7	0.8	0.8	1	1
town	0.5	0.6	0.8	1	1
village	0.2	0.5	0.5	0.7	0.8
remote area	0.1	0.3	0.4	0.6	0.7

are considered and the results shows a 94% accuracy in the method.

Table II. shows the Lukasiewicz implication of $f_{\rightarrow}(p_r, p_b)$, where pr stands for the fuzzy value of a location for its proximity to railway station and p_b shows its fuzzy value for its proximity to bus stations. According to the crime analysis of various locations, it is found that crimes are frequent in areas near to railway stations and in areas near to bus stations. As given in Table .2, if a location is with 100m proximity of a railway station, its fuzzy value is high(0.9) and if the location is with more than 1km proximity of railway station, then its fuzzy value is low(0.5). Also a location within 50m proximity to a bus station is with a fuzzy value 0.7(high) and the one with more than 1km proximity is with a fuzzy value 0.3(low). Applying $f_{\rightarrow}(p_r, p_b)$, we have taken all areas with value $\geq$ 0.7 as unsafe (crime prone) areas.

Similar calculation are done for proximity to malls and for density of population and values are shown in Table III. Crimes are very few in locations with less population density(fuzzy value 0.1) and crimes are seen to be high in locations with high population density(fuzzy value 0.9). As per Table .3, all $f \rightarrow (p_m, d)$ which are $\geq$ 0.8 are crime prone areas.

Table IV. shows $f_{\rightarrow}(e, u)$ where e is the literacy rate and u is the urban factor. According to crime analysis, crime is more in urban areas than in rural areas. Also new types of crimes are there when literacy rate is high. Crimes using modern technology is seen when literacy rate is high. The effect of these two factors is plotted in Table .4. Crime prone locations are those where value is more than or equal to 0.8. Finally we take the combined effect of all the factors as $f_{\rightarrow}(p_a, p_h, p_r, p_b, p_m, d, e, u)$ = min $\{1, 1 - \{f_{\rightarrow}(p_a, p_h, p_r, p_b)\}+ \{f_{\rightarrow}(p_m, d, e, u)\}\}$.

After this classification of locations into crime prone areas and safe areas, we calculate the following:-

- True positives t_p = Number of locations originally crime prone and classified as crime prone.
- False positives(Type I error) f_p = Number of locations originally safe but classified as crime prone.
- False negatives(Type II error) f_n = Number of locations originally crime prone, but classified as safe.
- True negatives t_n = Number of locations originally safe and classified as safe.

With these four values we calculate True Positive rate (recall), False Positive rate, True Negative rate, False Negative Rate, Positive Predictive Value (Precision), False Omission Rate, Positive Likelihood ration(LR+), Negative Likelihood Ratio(LR-), Negative Predictive value, F1 score and are indicated in Table V. The ROC curves for all these Lukasiewicz implications are plotted and shown in the Figs.

TABLE V. Results of lukasiewicz implications

Measures	$f_{\rightarrow}(p_a, p_h)$	$f_{\rightarrow}(p_r, p_b)$	$f_{\rightarrow}(p_m, p_d)$	$f_{\rightarrow}(e, u)$
t_p	1	1	1	0.875
t_n	0.875	0.625	0.875	0.875
f_p	0.125	0.375	0.125	0.125
f_n	0	0	0	0.125
Σ pos	1	1	1	1
Σ neg	1	1	1	1
Recall (TPR)	1	1	1	0.875
Miss rate (FNR)	0	0	0	0.125
False +ve Rate(FPR)	0.125	0.375	0.125	0.125
True -ve Rate(TNR)	0.875	0.625	0.875	0.875
+ve Predictive value(PPV)	0.889	0.73	0.889	0.875
False Omission Rate(FOR)	0	0	0	0.125
+ve likelihood ratio(LR+)	8	1.66	8	7
-ve likelihood ratio(LR-)	0	0	0	0.143
Accuracy	0.938	0.813	0.938	0.875
False discovery rate(FDR)	0.111	0.273	0.111	0.125
-ve predictive value (NPV)	1	1	1	0.875
F_1 score	0.94	0.844	0.94	0.874

22 to 25.

9. Contributions of the work

This work is unique in the sense that hypergraphs used in this work are unique representations of data. Morphological operators are nonlinear operators which can be used in hypergraphs in an effective way and fuzzification of hypergraph give better mode such that the parameters involved are tuned with realistic observations. The work has taken into consideration all parameters like the current crime pattern in a locations, socio-economic feature of the location and its development aspects. When compared to other works in the same area, the fuzzy operator, Lukasiewicz gives a better analysis of dependency among the parameters used in the model.

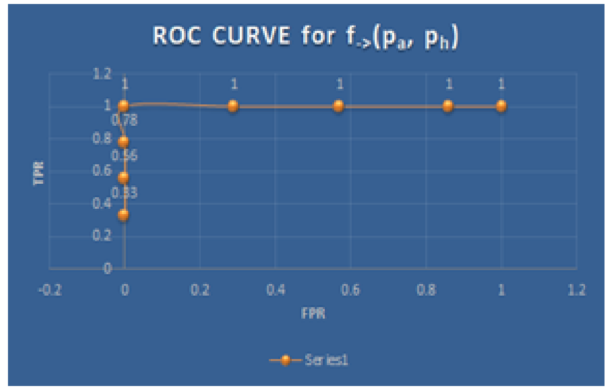


Figure 22. ROC for $f_{\rightarrow}(p_a, p_h)$

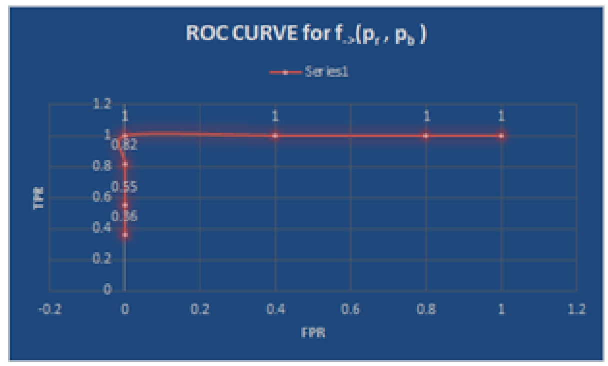


Figure 23. ROC for $f_{\rightarrow}(p_r, p_b)$

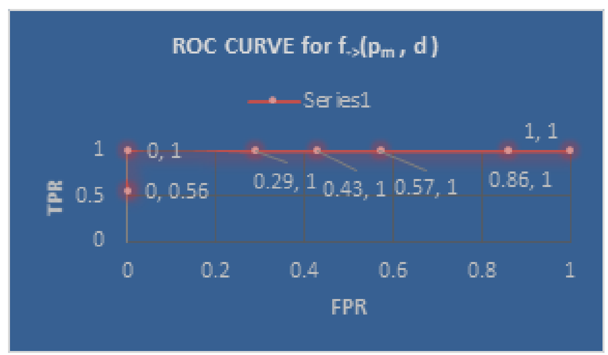


Figure 24. ROC for $f_{\rightarrow}(p_m, d)$

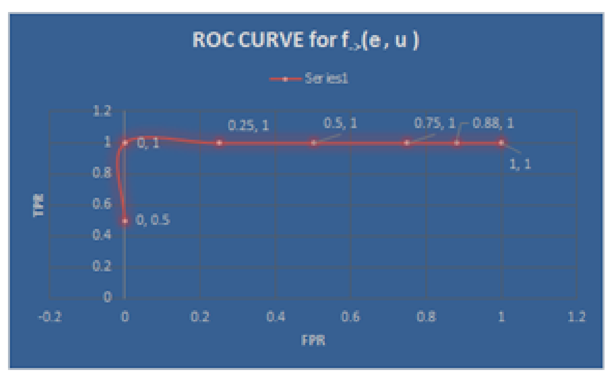


Figure 25. ROC for $f_{\rightarrow}(e, u)$

10. Conclusion

In this work, both crime mapping mobile application and the web version is created. The implemented features are alerts, hypergraph, the statistical report, crime mapping, reporting and prediction. The crime prediction which passes through the training and testing phase using the Lukasiewicz fuzzy method has shown good results. It addresses issues like the unavailability of the secure online sources to understand or analyze the crime data both for the public and the police. The interface can be accessed by both the public and police with a simple and responsive GUI for an enhanced experience. Also, many other features can be implemented further, if all the privileges and permissions are given from the police side. In the prototype, crime is being mapped to the region and while trying real-time fetching of latitude and longitude of regions within Kalamassery, we are unable to fetch the precise latitude and longitude. So if a precise system to fetch the latitude and longitude is integrated with the project, it will reduce the data storage as well as make the fetching process automated and faster. As mentioned in Table. 5, system has got an average precision of 0.85 and an average accuracy of 0.89. The system is an innovative one as it was modelled on a fuzzy hypergraph with Lukasiewicz implication applied on it for crime prediction. The prediction has an average F1 score of 89.95%

Acknowledgment

We acknowledge the financial support given by the Kerala Technological University- Center for Research and Development, Kerala, India. We also thank the Kalamassery, Kerala Police station for providing us the data set required for the work. We would like to make a special mention on the technical support given by Rajagiri School of Engineering and Technology, Kochi, Kerala, India for the successful completion of the work.

References

[1] R. Boba, "Introductory Guide to Crime Analysis and Mapping," Tech. Rep., 2001.

[2] A.Malathi and D. Baboo, "Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters," *Global Journal of Computer Science and Technology*, vol. 11, July 2011.

[3] D. M. S. Shiju Sathyadevan and S. S. Gangadharan, "Crime Analysis and Prediction Using Data Mining," in *Proceedings of the First International Conference on Networks and Soft Computing (ICNSC)*, 2014, pp. 406–412.

[4] K. K. Deepika and V.Smitha, "Crime analysis in India using data mining techniques," *International Journal of Engineering and Technology*, vol. 7, pp. 253–258, 2018.

[5] P. Yerpude and V. Gudur, "Preditive modelling of crime dataset using data mining," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. 7, no. 4, July 2017.

[6] R. M. Tahani Almanie and E. Lor, "Crime prediction based on crime types and using spatial and temporal criminal hotspots," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. 5, no. 4, July 2015.

[7] D. S. R. Alkesh Bharati, "Crime Prediction and Analysis Using Machine Learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, 2018.

[8] "Crimeview: Geographic mapping technologies corporation," Tech. Rep., 2017. [Online]. Available: http://gmtgis.com/products/crimeview/

[9] "RAIDS online: Municipality of Anchorage," Tech. Rep. [Online]. Available: http://www.muni.org/departments/police/pages/raidsonline/

[10] "Risk ahead crime map," Tech. Rep. [Online]. Available: http://www.riskahead.net.

[11] A. Almaw and K. Kadam, "Survey paper on crime prediction using ensemble approach," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 8, pp. 133–139, 2018.

[12] G. M. Hitesh Kumar Reddy, Bhavna Sainia, "Crime prediction & monitoring framework based on spatial analysis," *Procedia Computer Science 132*, p. 696–705, 2018.

[13] H. B. F. David and A. Suruliandi, "Survey on crime analysis and prediction using data mining techniques," *ICTACT journal on soft computing*, vol. 07, April 2017.

[14] C.Pletikosa, "Mining large-scale human mobility data for long-term crime prediction," *EPJ Data Sci.*, vol. 7, no. 26, 2018.

[15] I. Bloch and A. Bretto, "Mathematical morphology on hypergraphs application to similarity and positive kernel," *Computer Vision and Image Understanding Elsevier*, 2013.

[16] I. Bloch, "Morphological links between formal concepts and hypergraphs," *Springer International Publishing AG*, 2017.

[17] J. H. Dengyong Zhou and B. Scholkopf, "Learning with hypergraphs clustering classification and embedding," *Advances in Neural Information Processing*, pp. 1601–1608, 2007.

[18] R.Dharmarajan and K.Kannan, "Hypergraph-based edge detection in gray images by suppression of interior pixels," *Global Journal of Science Frontier Research Mathematics & Decision Sciences*, vol. 12, 2012.

[19] C. BERGE, "Hypergraphs , combinatorics of finite sets," *North–Holland Mathematical Library Elsevier Science Publishers*, 1989.

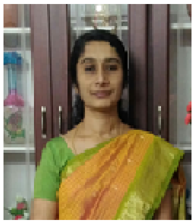
[20] D. P. M. S. A. J. M. R. P.B., "Algebra of morphological dilation on intuitionistic fuzzy hypergraphs," *IJSRSET*, pp. 300–308, 2018.

[21] B. S. A. U. K. Balakrishnan and P. B. Ramkumar, "Mathematical morphology on hypergraphs using vertex-hyperedge correspondence," *International Scholarly Research Notices*, vol. 2014.

[22] P. M. D. A. S. M. Jathavedan and P. B. Ramkumar, "On constructing morphological erosion of intuitionistic fuzzy hypergraphs," *Journal of Analysis Springer*, pp. 583–603, 2018.

[23] H. V. Lierde and T. W. S. Chow, "Learning with fuzzy hypergraphs: a topical approach to query-oriented text summarization," *Information Sciences*, vol. 496, pp. 212–224, 2019.

[24] K. B. B. S. U. A and R. P. B, "Morphological filtering on hyper graphs," *ISRN Discrete Mathematics*, 2014.



Dr. Dhanya P.M Dr. Dhanya P.M is working as Associate Professor and HoD in the Department of Computer Science and Engineering, Rajagiri School of Engineering and Technology, Kochi, India having a total 20 years of teaching experience. Her research fields are NLP, Fuzzy Hypergraphs, Data mining and Soft computing. She has 20 research publications and an International Patent.

Dr. Ramkumar P.B Dr. Ramkumar P.B is working as Associate Professor and HoD in the Department of Mathematics, Rajagiri School of Engineering and Technology, Kochi, India. His research field is Mathematical morphology and Discrete Mathematics. He has 14 publications and 2 patents.

Nirmaljith Cletus Nirmaljith Cletus is a B.Tech Computer Science & Engineering graduate from Rajagiri School of Engineering and Technology (APJ Abdul Kalam Technological University). He is currently working as a Software developer at Verizon.

Paul Joy Paul Joy is a B.Tech Computer Science and Engineering graduate from Rajagiri School of Engineering and Technology (APJ Abdul Kalam Technological University) and is currently working as a Software developer at Incture Technologies. His areas of interest are Web Application Development, Mobile Application Development.