International Journal of Advanced Technology and Engineering Exploration, Vol 9(92)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2021.875567

Research Article

An optimized deployment plan of ambulances for trauma patients

Zaheeruddin and Hina Gupta*
Department of Electrical Engineering, Jamia Millia Islamia University, New Delhi, India

Received: 01-March-2022; Revised: 19-July-2022; Accepted: 20-July-2022

Abstract
Emergency medical service (EMS) control centres should focus on strategically deploying ambulances to enable trauma patients to receive better care. The work proposed here aims to find an optimal deployment plan of ambulances for the existing base stations using the genetic algorithm (GA) based optimization component. The GA has been modified by incorporating a new proportion-based population seeding method for initializing the population. Considering a set of assumptions, the authors have applied the new strategy for allocating an optimal count of ambulances to 28 base stations in Delhi. The working environment of EMS that includes stochastic requests, travel time, and dynamic traffic conditions has been taken into account, and the optimization strategy has been implemented in a MATLAB environment. With the proposed work, the authors have been able to reduce the average response time (ART) by 6.7%. The simulation result has also demonstrated a comparison between GA and particle swarm optimization (PSO). In addition, some numerical experiments are performed to conclude the impact of different attributes on the value of ART.

Keywords
Ambulance allocation, Ambulance deployment, Emergency medical service, Trauma victims, Accident victims.

1.Introduction
Emergency medical service (EMS) control centres are essential components of modern healthcare systems. They are the pre-hospital component of the health system, which includes medical care and transport activities performed from the arrival of an emergency call with the release of a patient or its transfer to a hospital [1]. EMSs thus, play an important role in responding to emergency calls and significantly impact a patient's health and recovery [2, 3]. However, the major challenge for EMS is to strategize and work towards providing a quick service to society. Thus, EMS has to efficiently handle some issues like deciding on the locations of ambulance stations (base stations), allocation of ambulances to base stations, policies regarding the dispatch of the ambulance to handle service requests, and relocation of ambulances if needed [4, 5]. In recent years, various support tools have been developed using different optimization techniques. These tools help EMS organizations in decision-making and policy framing.

The effectiveness of any service can be measured in terms of quality. The measurement of the quality of EMSs can be done by evaluating response time, the type of care provided by EMS staff, types of equipment used by staff, the number of requests handled, and so on. Among all the attributes, response time is believed to be strongly correlated with a patient's survival rate. The response time in EMS is the time interval from patients calling for service until being reached [6]. Increasing call volumes and worsening traffic conditions in the metropolis are creating a challenge for the EMS control centres to achieve good performance. To shorten the time in which the emergency help reaches the service requesters, the EMS control centres need to decide the locations at which the ambulances should be stationed, the number of ambulances that should be allocated at these base stations, and dispatching policies. However, irregular population density across the city makes the allocation of ambulances at the base stations difficult. A dynamic and stochastic environment is associated with the process of decision-making by EMS. Some predictable and unpredictable attributes associated with EMSs are locations from where requests are raised, traffic conditions on roads, the arrival rate of requests, time used to serve the requests, travel time

*Author for correspondence

to and from different locations, and the service time at the request calls' scenes and hospitals. All the stated attributes associated with EMS cause obstruction in making decisions about ambulance deployment [7].

In recent years, the counts of road accidents in India have increased. Delhi, the capital of India, accounts for the maximum count of accidents in the country [8]. Road accidents in India are expected to cause 2,500 deaths in a year by 2025 [9]. The leading cause of death in accident cases is a road traffic crash (RTC). However, since not much focus has been placed on this topic, it has been termed "the neglected disease of modern society" [10]. To provide a good EMS to the trauma victims, a special organization named, centralized accident and trauma services (CATS) has been established in Delhi by the central government. CATS is an autonomous body that provides 24X7 free ambulance services to the victims of accidents and trauma. Currently, it comprises 151 ambulances deployed at 28 base stations serving the city with an average response time (ART) of approximately 13 minutes [11]. A deployment plan refers to the allocation and relocation of ambulances among the base stations. The proposed work intends to find a new optimal deployment plan for ambulances that can serve the city with an ART value of less than 13 minutes. To assess the operational performance of the new plan, the framework used here considers all the uncertainties associated with the working model of EMS. It aims to investigate and analyze the performance of ambulance deployment from different perspectives. The work also focuses on finding the maximum count of ambulances that can be added to the existing base stations beyond which no further improvements can be achieved in ART until more base stations are added.

The remaining paper is structured as follows: Section 2 offers a brief overview of the related works carried in the same domain. Section 3 focuses on the problem, proposed framework, methodology and the mathematical formulation of the work. The experiments, results and discussions are covered in Section 4. Finally, Section 5 contains the conclusion and future scope related to the work.

2.Literature review

A wide range of literature is available on EMSs focusing on strategies for improving the quality of medical service. However, we have confined the literature review to the work that is closely related to our topic. To provide a satisfactory level of service to the population, the ambulances used by EMS are strategically placed at different locations in the region. This static ambulance location problem deals with the selection of base stations or standby sites and the count of ambulances that should be placed at these locations. The very first explicit study on the location of ambulances was based on minimizing the count of ambulances that are supposed to cover all the demand points. This is also called location set covering model (LSCP) [12]. A major drawback of this model is that it considers equal demands from all the nodes along with only one facility site per demand location. This drawback was overcome by another model proposed by Church and ReVelle [13] to solve the maximal covering location problem (MCLP), which allowed the specification of demand at each node. The model worked on covering maximum demands with a given (fixed) number of ambulances. Recently, Zonouzi and Kargari [14] used the results of [13] and a method of data mining for allocating ambulances and rescue vehicles to handle the trauma victims. These are static models that do not consider the fact that ambulances will become unavailable throughout the day and certain demand points might not be covered anymore. Several approaches have been developed in order to handle this problem of uncertainty. For example, the probability model for unavailable ambulance developed by Daskin, presented a solution to the maximum expected covering location problem (MEXCLP) [15]. In this, a probability p is associated with the system indicating the non-availability of ambulance to serve a request. In the models proposed earlier, a value indicating the probability of non-availability of ambulances was considered to be constant. However, in the models proposed later, iterative methods were used to achieve a rational value indicating the probability of non-availability of ambulances [16]. Another possibility includes multiple coverages, i.e. demand points are supposed to be covered by more than one vehicle. Such a model called the double standard model (DSM) was introduced by Gendreau et al. [17]. DSM was further enhanced to work in a dynamic context, allowing it to take advantage of the time frame between calls by predicting fleet deployment decisions in the future [18]. DSM that initially worked for a single period was further modified by Schmid & Doerner to work for multiple periods [19]. Shariat-Mohaymany et al. proposed two reliability-based linear models for the optimal location of ambulances [20]. Similarly, a framework was proposed to help in decision-making strategies for locating and assigning the emergency vehicle (firefighter vehicle) for improving response time [21]. However, most optimization models do not effectively handle the dynamic situations, stochastic environment, randomness, and uncertainty related to locations, travel time, traffic situations, and service time [8]. To address such problems, optimal or heuristic search solutions have been developed using genetic algorithm (GA) and particle swarm optimization (PSO).

GA is search and optimization algorithm that imitates the processes involved in biological evolution, such as mutation and crossover [7]. A mathematical model was proposed to solve the ambulance coverage problem by Benabdouallah et al. [22]. In this work a hybrid combination of GA and guided local search was used to find an optimal solution. It resulted into a plan that distributed ambulances in each potential waiting site, minimizing the total latency of emergency intervention. McCormack et al. used GA with an integrated simulation model to optimize the EMS fleet allocation and location of base station [18]. The optimization and simulation work was done using the real call data from the London Ambulance service. The work optimized the existing resource plan showing significant improvement in the survival probability. In another work put forward by the authors in [4] three different case studies are analyzed using optimization techniques to support in the decision making policies for medical services. Apart from handling the deployment of ambulances, some authors have improved the emergency services by optimizing the management of human resources for handling emergencies in hospitals [23]. Another work incorporating the use of GA is in scheduling surgeries in a Mexican Public Hospital [24]. To improve the results of the GA; a technique of population seeding is used to initialize the population in GA. This method improves GA in terms of problem search space exploration, convergence speed, and the final optimal solution obtained. Random initialization, nearest neighbor, selective initialization, gene bank, knowledge based, and regression based initialization are some of the well-known population seeding methods used by different authors [25−28] to attain optimal results.

The field of ambulance allocation has also been extensively exploited by various researchers using another optimization algorithm of PSO. In the work by Kolomvatsos et al. [29], a scenario of disaster was considered to propose a methodology for allocating resources for emergency response. They used the PSO algorithm for the work. The authors distributed the area concerned into various cells and calculated the cell weight based on the spatial data for all the cells. Later, different counts of ambulances were allocated to each cell considering the run time and coverage of the area. In another work, adaptive PSO was used by the authors to determine the number of facility locations required along with the allocation process at the locations [30]. A work was proposed by the authors in [31] where a solution for optimally allocating the ambulances was proposed using Jumping PSO. To ensure that medical help is reached to the patient in the shortest time possible, the work of [32] deployed ambulances using the data of requests raised and the coverage to be achieved.

The servicing capability of an EMS system is also highly dependent on the dispatching and redeployment strategy of the ambulances. As per the related work carried out in [33, 34] there is not much scope for improvement in terms of servicing capability of EMS due to complex dispatching policies. Generally, the nearest ambulance is dispatched to the demand location to serve the request which does not always provide a good result. Therefore, ambulances should be redeployed so that the service provided could be improved. A two-stage stochastic optimization model was presented in [35] to solve the ambulance redeployment problem to minimize the number of relocations. The authors in [36] considered the workload of EMS personnel while planning for the redeployment of the ambulances. A real-time ambulance redeployment approach was proposed by Ji et al. [37]. It used real-time data to find a new optimal location for the free ambulance to optimize transporting capability. Yavari et al. considered the problem of ambulance dispatching and relocation to avoid overcrowding of emergency departments [38].In the maximum of the models and works stated above, the researchers have considered response time as the prime attribute while assessing any deployment configuration. The background of many articles has validated response time as a major factor for gaining better insight into the operational performance of EMS. The work of Wilde has also clearly demonstrated that mortality rate and recovery rate of patient is highly influenced by the response time [39]. The literature discussed gave some strong motivations to find a cost-effective solution that improves service performance of EMS. In most of the previous works, researchers have worked on finding new locations for base stations. However, finding new locations for ambulances may demand construction or setting up of new base stations. Therefore, the authors chose to utilize the existing locations by allocating ambulances to the existing base stations. The problem undertaken by the authors is novel in its form as both the allocation and relocation activities of ambulances will be done amongst the existing base stations. The proposed work takes into consideration the intricate and random evolution of the CATS EMS system over time, the uncertainty of request arrivals, real-time traffic conditions, patterns in road accidents, and accident-prone areas as stated in the report of road accidents by Delhi traffic police [40]. Although many heuristic approaches exist these days to solve the optimization problem, the authors chose GA and PSO due to their easiness in handling potential solutions and capability of combining optimal features from the population of solutions. The algorithms are capable to explore huge search space in less time, which otherwise is a time-consuming process. One more advantage of GA is that it has an inherent property of parallelism. The parallelism nature helps to avoid it being trapped in local optima and provides a global optimized result.

3.Methods

3.1Problem background

The working environment of EMS comprises of base stations, ambulances, hospitals, demand locations, and patients. The base station is a location where the ambulance is in standby mode and activates for movement whenever a request call arrives. Ambulances are vehicles that help in transporting patients to hospitals. Demand location is a location from where the request is raised. Patients are people who are in need of medical aid. As the medical service requests (demand sites) in a territory are irregularly distributed, the paper deals with the optimal allocation and relocation of ambulances. The aim is to optimize the operational efficiency of the EMS so that the response time is reduced [40, 41]. As shown in *Figure 1*, when a request call is initiated, ambulance is selected and dispatched to the demand site. The decision for dispatching the ambulance is taken as per the decision rules set by the EMS authority. The general rule used is to select and dispatch the ambulance that is nearest to the requested location [4, 33, 42, 43]. When the ambulance reaches the request location it may provide first aid to the patient or resuscitation. It then takes the patient to the hospital if needed or returns to the base station and waits until dispatched to serve new request. The counts of accidents in Delhi have increased in the last few years. In some cases, the victims of these accidents die as they are not provided a timely assistance. Therefore, there is a need to optimally allocate ambulances among the base stations so that ambulance reaches the requested site in the shortest time possible. Studies also reveal that the occurrence of accidents is more at night time in comparison to day. Limited brightness, range of illumination, and fatigue are some of the major reasons behind the high frequency of accidents at night [44, 45]. To deal with such time dependent scenarios and different request rate, there is a need to relocate ambulances [7, 46]. Relocation is concerned with changing the location i.e. redeploying the ambulances amongst the base station between different time frames of a day to meet the change in the arrival rate of request calls. Attaining two different allocation plans for ambulances for day and night, we need to find how to relocate ambulances so that the total relocation (distance) cost in minimized. Relocation cost is the cost incurred while moving ambulance from one base station to other in the city. The relocation plan should consider the following three conditions: (a) select a time at which the ambulances need to be relocated; (b) it is not mandatory to relocate all the ambulances; (c) the movements of the ambulances to go to the new base stations should minimize the cost incurred for the ambulances.

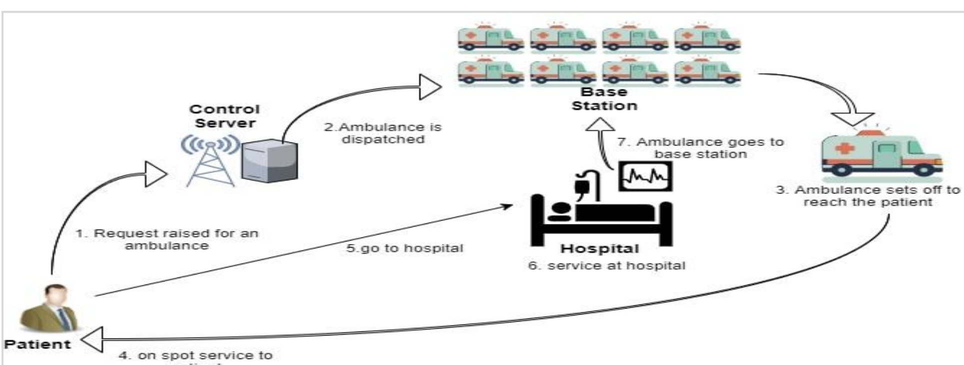


Figure 1 Working of EMS

3.2Framework

A simulation optimization (SO) framework is used in this work. The block diagram of SO framework used is illustrated in *Figure 2*. It consists of ambulance assignment component (AAC) and an optimization component (OC). AAC provides a simulation environment and provides a potential solution to OC which then optimizes the result. The figure depicts the flow of data from one entity to the other. The detailed working of AAC and OC are explained in the subsequent sections.

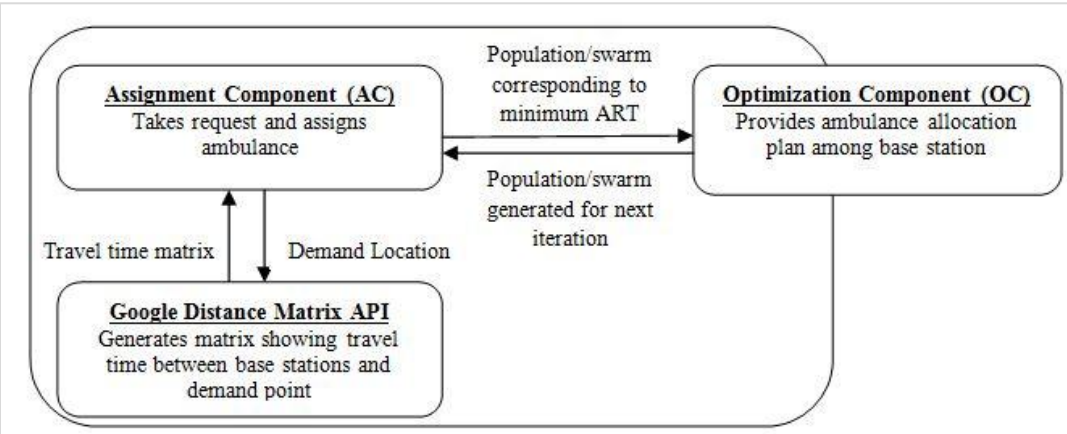


Figure 2 Block diagram showing the simulation optimization framework

3.2.1 Ambulance assignment component

An ambulance assignment component is a simulation program that captures the workflow of EMS. It is a program that assigns the ambulances to fulfil the random requests raised from the various locations of the city. In order to assign ambulance to serve a demand, Google distance matrix application programming interface (API) is used. This API generates a matrix of travel time between the demand point and all the base stations. The ambulance with least travel time from base station to the demand point is assigned to provide the required service. The algorithm for detailed working of AAC is shown in *Figure 3*. The selection of the ambulance, travelling to the demand point, and fulfilment of requests are carried out by AAC. This captures the operational level behaviour of EMS to measure the service performance by evaluating the ART of all the requests raised.

Algorithm 1: Ambulance Assignment Component (X)
Input: Chromosome (X)
Output: Objective value OV_x of chromosome X
Function: Ambulance Assignment Component (X)

Generate 300 instances randomly// every instance consists of randomly generated requests (r) as per the PDFs of request arrival and distribution and PDFs of service time of requests
ist = 1 // ’ist’ stands for index of instance
whileist<= 300 **do**
for all generated requests(r) in an instance **do**
  Call Google Distance Matrix API to find t// t is the travel time from all the base stations to the request location
  Sort the base station as per the travel time obtained above in increasing order
  Select the ambulance from the base station with the shortest travel time, and assign it to the request r
  Check the status of the ambulance
  if status is ’available’ **then**
    Update the status of ambulance as ’assigned’ The ambulance sets out from the base station and reaches request r’s location.
    Decrement the count of ambulance from the particular base station by 1
  else

Select the base station next in the increasing order obtained
End
Response time (RT) is recorded between the request(r) arrival and arrival of ambulance at the location of request(r)
Ambulance provides service to the patient as needed and returns back to its allotted location
Update the status of ambulance to 'available'
End
Evaluate the Average Response Time(ART) for all the requests
T(ist) = Average(RT)
ist++
end
For all the instances calculate the mean value of T(ist)
OVx $=\frac{1}{300}\sum_{ist=1}^{300} T(ist)$**return**
OVx

Figure 3 Algorithm of AAC

3.2.2 Optimization component

The OC consists of meta-heuristic algorithms GA and PSO to work on the results generated by AAC to find an optimized allocation plan for ambulances.

a) ***Genetic Algorithm***

GA imitates the natural evolution process of inheritance, mutation, selection, and crossover to find an optimal solution [47, 48]. Here, the current configuration of deployment of ambulances is reconfigured by the optimizer by finding an alternate solution for integer decision variable '*X*'. '*X*' comprises of integer values x_i that signifies the count of ambulances deployed at '*N*' base stations and is denoted by $X = \{x_1, x_2, \ldots, x_N\}$. Thus, as per Equation1.

$$\sum_{i=1}^{N} x_i = X \tag{1}$$

The algorithm for working of GA is illustrated in *Figure 4.*

In this research, the working of GA has not been modified with respect to the evolutionary operators. Here, the authors have incorporated a new method of population seeding to enhance the capability of GA. The population is a set of chromosomes or a subset of solutions generated in the different generations of GA. The population that gives the best result for the specified objective function is acknowledged as the outcome in GA. To obtain a global optimum solution, good quality and diversified population should be generated at every iteration. Since the population of every generation is dependent on the population of the previous generation, high importance is placed on the step of initial population seeding [49]. In the initial population seeding phase, the population of feasible solutions is randomly or heuristically generated to give it as input for the GA. The random initialization technique is used when complete knowledge about the problem is not available [50]. However, this may generate a poor fitness solution and reduce the likelihood of finding an optimal solution. On the other hand, having prior knowledge of the problem helps in generating the initial population heuristically. The results of the previous studies indicate that heuristic generation of the initial population enhances the capability of GA to provide optimum or near to optimum solutions. However, the generation of the entire population should not be done heuristically as it may result in a population having identical solutions with very little diversity [51]. Therefore, the best way is to use a mixed approach of heuristically seeding some populations with good solutions and allowing the random generation of remaining solutions. In this paper, the authors have proposed a novel method of proportion-based initial population seeding method. This method is used to attain a solution that can be used in the initial population seeding phase to define one population heuristically and the remaining population can then be randomly generated by GA. A detailed explanation of the steps involved in the proportion-based population seeding is explained below.

Algorithm 2:	Optimization Component(GA)
Input	Data about ambulances and base stations
Output	Optimized deployment plan
Initialize N //	Total number of base stations
Initialize C //	Total number of ambulances
Initialize array	X []= [x_1 x_2 ...x_N] // The values of array X correspond to the number of ambulances at every base station and size of array is equal to the number of base station
Initialize	a population P(X) of twenty chromosomes// every chromosome is an integer array
Evaluate	the objective function value of each chromosome using Ambulance Assignment Component (X)
while	minimum average time has not been improved for 100 successive iterations **do**
	Sort the twenty chromosomes in ascending order of their objective value
	The best 10% chromosomes evolve to the next iteration without any modification.
	The moderate 80% chromosomes undergo mutation operation.
	The worst 10% chromosomes undergo crossover operation
	Evaluate the objective value for all the chromosomes of the new population obtained using Ambulance Assignment Component (X)
	For the current iteration **Record** the objective value of best chromosome
	Update the existing best solution obtained so far if necessary
end	

Figure 4 Algorithm of GA as OC

a. Proportion-based seeding method:

This new method is based on the concept of proportion that relates a part or share to a whole. To get a deeper insight into the working procedure, Let,

N = total base stations in the area concerned,

$X = \{x_1, x_2, \ldots x_N\}$ where 'x_i' signifies the count of ambulances deployed at 'N' base stations,

$T = \{t_1, t_2, t_3, \ldots, t_N\}$where '$t_i$' is the initial count of ambulances allocated at 'N ' base stations and as per Equation 2.

$$\sum_{i=1}^{N} t_i = T \quad (2)$$

R = maximum count of request calls generated in the area concerned. This data is retrieved using the historical data.

Since at the beginning of the simulation there is no certainty of from where and when the requests will be generated, the count of ambulances at each base station is set equal to the value 'R' so that for every request there is one ambulance always available at the base station. Thus, using Equation 3.

$$t_1 = t_2 = t_3 = \ldots = t_N = R \quad (3)$$

It is assumed that each request demands exactly one ambulance. Therefore, an integer variable 'C' is defined as $C = \{c_1, c_2, c_3, \ldots, c_N\}$where '$c_i$' is the count of requests fulfilled at each base station.

Each iteration 'k' is run for 'R' requests assuring that atleast one request is raised from every demand point and

$1 \le k \le K$,

where 'K' is the total number of iterations.

Let $F_i^{\,k}$ = the count of ambulances used in every iteration 'k' from each base station in fulfilling the requests. Thus, according to Equation 4.

$$F_i^k = [t_i - (t_i - c_i^k)], \text{where } i \epsilon N \quad (4)$$

Now, the total count of ambulances (A_i) used at every base station 'i' after the completion of 'K' iterations is given by Equation 5.

$$A_i = \sum_{k=1}^{K} F_i^k \quad (5)$$

The total average response time (RS) after the completion of 'K' iterations is given by Equation 6.

$$\mathrm{RS} = \sum_{k=1}^{K} S_k \quad (6)$$

where 'S_k' denotes the average response time of the ambulances achieved in each iteration 'k'.

Average count of ambulances (Avg_i) at every base station 'i' after the completion of 'K' iterations is calculated by Equation 7.

$$\mathrm{Avg_i} = {}^{\mathrm{A_i}}/_{\mathrm{K}} \quad (7)$$

After all the requests are fulfilled and the value of Avg_i is attained, we can calculate the proportion of ambulances 'P'$_i$ that are used at all the base stations. This can be evaluated by Equation 8.

$$P_i = \frac{Avg_i}{t_i} \times 100 \quad (8)$$

Using the value of P_i, the values of x_i in initial population can be defined as shown in Equation 9.

$$x_i = X * P_i \quad (9)$$

Mean of response time denoted as 'AT' taken by the ambulances after the completion of 'K' iterations is given by Equation 10.

$$AT = RS/K \quad (10)$$

The value of '*AT*' attained using this approach is the minimum ART achieved in an ideal situation where each base station has an ambulance available for every demand request at any point of time. Therefore, the value of '*AT*' is considered as the benchmark value for ART in the proposed work.

b. Particle swarm optimization

The technique of searching for food and the method of social interaction by the flock of birds was observed and modeled by Poli et al. in the year 2007 [52]. In this method popularly known as PSO, the swarm particles make movements to attain food (optimal objective). A group (swarm) of randomly generated particles is given as input to the PSO algorithm. Every particle in the swarm is a possible solution to the problem. Two major attributes linked with every particle are position and velocity. Considering the ideal optimum solution as zero, the position of the particle evaluates the distance of the particle from the optimum solution by finding the fitness value of the optimization function. The velocity of the particle establishes the movement of the particle in the solution space. Two other variables namely *pbest* and *gbest* are associated with the social behavior of every particle. The best position of the particle that it achieves in the iterations of the optimization is denoted by the value of *pbest*. On the other hand, gbest represents the best position attained by the whole swarm. The values of the variables are updated in the successive iterations only if the new achieved values are better than the last stored values of *pbest* and *gbest*. Thus, to achieve their objective, the swarm makes use of their personal (*pbest*) and swarm's (*gbest*) best experiences. The values of the current velocity of the particle, its best position, and the swarm's best position are used to update the position of each particle. These values are continuously updated until the optimal solution is attained. The algorithm of PSO is given in *Figure 5.*

Algorithm 3	Optimization Component(PSO)
Input	Data about ambulances and base stations
Output	Optimized deployment plan
Initialize N //	Total number of base stations
Initialize C //	Total number of ambulances
Initializeswarm	X []= [x_1 x_2 ...x_N] // The values of array X correspond to the number of ambulances at every base station and size of array is equal to the number of base station
Initialize	a population P(X) of particles , position and velocity
	for each particle find the fitness value using Ambulance Assignment Component (X)
	Update the best local solution P_{best}
	Update the best global solution g_{best}
	for population size **do**
	Compute $v_j^{k+1} = \mu^k v_j^k + c_1 r_1 (pbest_j - x_j^k) + c_2 r_2 (gbest_j - x_j^k)$
	Compute $x_j^{k+1} = x_j^k + v_j^{k+1}$
	end
	end **Update** the existing solution obtained

Figure 5 Algorithm of PSO as OC

3.3 Methodology

3.3.1 Area of concern

Delhi, the capital of India, has been taken as the area of concern for this work. It provides shelter to approximately 19.5 million people. Being one of the largest metropolitan areas of the world and having a population density of 11,297 persons per square kilometer, Delhi has area coverage of about 1484 square kilometers. Nowadays various private organizations are providing the ambulance facility in Delhi. However, we have confined our research to the CATS EMS organization that is operated and handled solely by the central government. In Delhi, 28 base stations and 151 ambulances are operated by CATS as shown in *Figure 6*. Base stations have been numbered from 1 to 28 as reference numbers used in this work and all the base stations are located to nearby government hospitals. Base stations are responsible for maintaining records of ambulances and handling other exigencies.

Being densely populated a huge amount of accidents are reported daily in Delhi. Out of these accidents, some are fatal and some are non-fatal. The authors have used the accident report of the year 2019 (January 1 to December 31) for the work [33]. The data of the report has mentioned many locations and zones that are accident-prone. As per the report, taking the different locations of accidents in an area, various clusters are formed. In all the clusters, a point is selected as cluster head that covered a distance 500 meters in diameter. Then, considering the frequency of accidents, the cluster head of the areas are categorized into accident black spot (ABS) and accident prone zone (APZ). If an area having a diameter of 500 meters has 3 or more fatal accidents in a day then its cluster head is classified as an ABS and if the total count of accidents (including fatal and non-fatal) within the same range is 10 or more in a day, its cluster head is classified as APZ. A total of 100 APZ and 20 ABS have been reported in Delhi as shown in *Figure 7* and *Figure 8* respectively. *Figure 9* shows all the accident-prone sites of Delhi. The work aims at finding out an optimal deployment plan with minimum ART for the accident-prone sites mentioned in this section.

Figure 6 Base stations of CATS

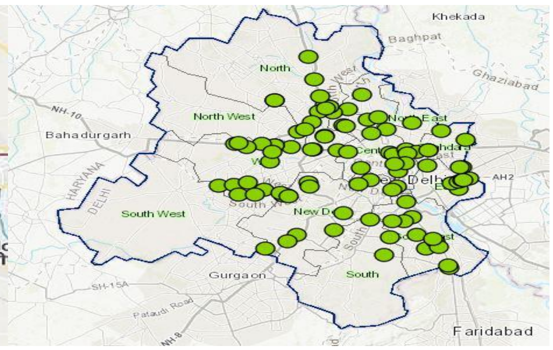


Figure 7 Accident prone zones (APZ)

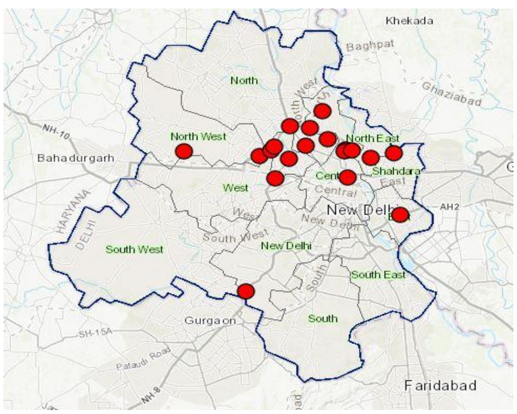


Figure 8 Accident black spots (ABS)

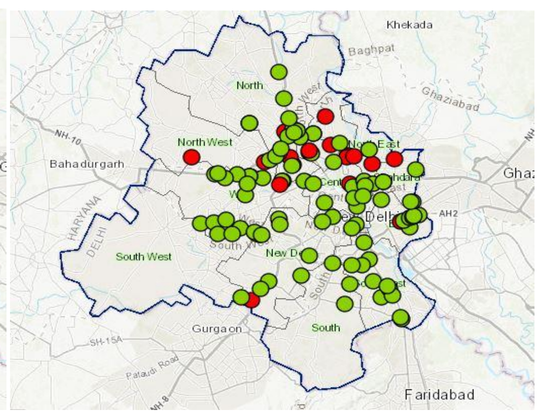


Figure 9 Accident-prone Sites

3.3.2 Data

Two types of data are usedin the work: *static data* and *real-time data*.

(i) Static data**:**This is the information that is known before running the framework. Static data comprises of(a) thedata about the number of base stations, (b) geographical location of base stations, and (c) accident-prone sites as stated in the report. The coordinates of the sites are obtained using Google Map API. This helped in the segregation of zone with a higher count of accidents. As per the data of the report, the frequency of accidents has been used as a key attribute to divide the day into two-time frames: *peak hours* and *lean hours*. *Peak hours* include the hours (7 p.m.-7 a.m.). Since the frequency of accidents in this frame is more due to poor illumination, fatigue of driver, reflection of light, improper judgment etc., the arrival rate of request calls at the base stations is increased [31]. On the contrary, *lean hours* include the hours (7 a.m.-7 p.m.), when the accidents are less and therefore call arrival rate decreases. To ascertain a good performance of CATS EMS throughout the day, the authors have proposed a relocation plan for relocating the ambulances once between peak hours and lean hours.

(ii) The real-time data**:** This comprises of the travel time from one location to another. Google Distance Matrix API provides this data in form of a matrix. The data is integrated into AAC and OC framework as explained in Section 3.2.

3.3.3 Assumptions

The authors have taken two assumptions for the work. First, if two ambulances are required at a single demand location, then the request will be split into two requests from the same location. Second, the dispatch decision will only consider the vehicles available at the base stations and not the ambulances that are busy serving the patient.

3.3.4 Ambulance allocation

For the research, a map enclosing the coordinate plane values for the city is defined using Arcgis software ArcPro 2.8. The coordinate plane shows various locations like base stations and accident-prone sites as stated by the Delhi traffic police report. Initial number (initial population) of ambulances at the base stations is defined using the proportion based population seeding method. Random requests are generated from different locations of the city that have been marked as accident prone areas making certain that atleast one request is generated from every location. In this work, 300 random requests are generated for every generation of OC. The emergence of a request triggers the search for an appropriate ambulance from a set of available ambulances. As soon as a request is raised, the coordinates of the location of the request is identified. Google Distance Matrix API then generates a matrix indicating the travelling time from every base station to the demand location. The traveling time from the base station to the demand site is used in handling the decision as to which ambulance has to be sent to serve the request. The deployment plan (population) with minimum ART is then sent to OC to generate more population by crossover and mutation. The OC and AAC continue to run iteratively using the outcome of each other until the value of ART becomes constant for 100 successive generations. Several runs of the framework are performed to attain statistically valid results.

3.3.5 Ambulance relocation

For relocating the ambulances we need to take the following key points into consideration [4, 6]:

K1: The count of ambulances available at every base station at the time of relocation

K2: For every base station, the volume of request calls should be noted both for current and future situations. This value is predicted using the data from the base stations.

K3: The geographical position of each base station should be considered. If the station is in a remote location, the cost of relocating any ambulance will be more. It is possible that there is a high density of calls at present or in near future. In order to handle a large volume of calls, it will be good if the base station has the right count of ambulances available to serve the requests.

K4: The traveling time taken by ambulance to move from the station *a* to *b* denoted as t_{ab} and traveling cost for moving from the station *a* to *b* denoted as c_{ab} for the relocating ambulance should be as minimum as possible. Taking these into consideration, the relocation plan for ambulances should be designed so that the ambulance takes less time to reach the new base station.

K5: Since the proposed work is not focused on the real-time relocation of ambulances, it is assumed that all the ambulances are available while the relocation plan is projected.

In order to redeploy the ambulances appropriately, all the key points K1-K5 mentioned above have been taken into consideration in order to obtain a new deployment plan using the relocation model as shown in *Figure 10.*

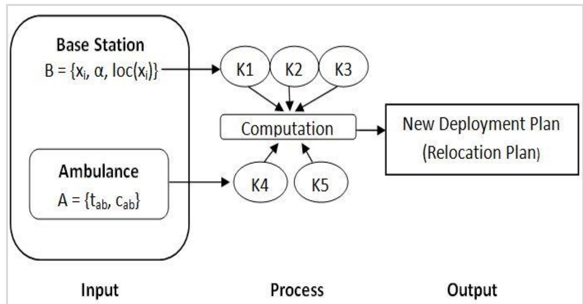


Figure 10 Relocation model

The steps in computing the new relocation plan is as follows:

Step 1: Input the allocation plan for peak hour.

Step 2: Input the allocation plan for lean hour.

Step 3: Evaluate the requirement matrix by calculating the difference between the peak and lean hour allocation plans stating abundance or deficiency of ambulances at each base station.

Step 4: The travel distance (cost) matrix to and from all the base stations is obtained from Google Distance Matrix API.

Step 5: Using the data of distance matrix, relocation of ambulances is done by moving the ambulances between base stations having with minimum travel distance.

Step 6: Update the data in all the matrices and continue the process until the required ambulances are relocated.

The stated steps help in evaluating the relocation plan for changing the location of ambulances between peak hour and lean hour.

3.4 Mathematical formulation

The mathematical formulation for obtaining the deployment plan with minimum response time is explained below. It specifies the input data, decision variables, auxiliary variables, and objective equation.

3.4.1 Input data

The ambulance deployment problem for a geographical location deals with assigning a certain count of ambulances (*X*) to the base stations (*N*). The area consists of several demand points or sites (*D*), from where the request for an ambulance is initiated. The solution for the ambulance deployment problem is represented by an integer variable x_i where $i \in N$, specifying the number of ambulances distributed among the base stations *N*.

3.4.2 Decision and auxiliary variable

We assume that *'a'* ambulances are available out of a total of *'X'* ambulances to serve the requests at an instant. The number of ambulances available at an instant for each station i can be written as $x_i(a)$. A binary value $y_i(a)$ will be associated with each station to specify the presence or absence of the ambulance at a station at the arrival of request. Therefore, the value of y_i will be zero if no ambulance is available at station 'i'. When there are 'a' ambulances available in the system, y_i will be 1. In the proposed work, response time has been taken as the prime attribute to measure the quality of service of EMS. Hence, we will pay more emphasis on the reducing the value of response time.

3.4.3 Objective function for minimizing response time

The objective function, which aims to minimize the response time, can be formulated as shown in Equation 11.

$$min\ RT = \sum_{i \in D} v_i * T_{ij} \tag{11}$$

subject to the constraints

$$\sum_{i \in D} x_i(a) = a \tag{12}$$

$$\sum x_i = X \tag{13}$$

$$y_i(a) \le x_i(a) \tag{14}$$

$$x_i(a) \ge 0 \tag{15}$$

$$y_i(a) \ge 0 \tag{16}$$

$$y_i(a) \in (0,1) \tag{17}$$

where T_{ij} is the travel time from node *i* (base station) to node *j* (demand location), given that '*a*' ambulances are available out of *X* and v_i denotes the arrival rate of calls per hour from node *i*. Constraint (12) checks that the total ambulances available at each base station are equal to the total number of ambulances present in the system at the same instant. Constraint (13) limits the fleet size of the ambulance to *X*. The fulfillment of requests is constrained by the presence of a certain number of ambulances at the base station by Constraints (14)-(16). Constraint (17) restricts the values of the variable.

3.4.4 Relocation cost

Since time frame of the day for handling the trauma victims has been divided into two frames termed as peak hours and lean hours, hence deployment plan will be different for two time frames. Therefore, a mathematical model has been proposed to obtain a relocation plan for the ambulances in this section. Suppose the amount of ambulances at every base station i ϵ N in the peak hours is denoted as x_i^p and in the lean hours is denoted as x_i^l, where *N* is the set of all the base stations. The number of ambulances that need to be relocated can be evaluated as the difference between x_i^p and x_i^l in the following manner as shown in Equation 18

$$z = x_i^p - x_i^l \tag{18}$$

A positive value of *z* will indicate that the base station '*i*' has surplus count of ambulance during peak hours whereas a negative value will indicate that that base station '*i*' has surplus count of ambulance during lean hours. The value of *z* will help in finding out the count of ambulances that will be required to be relocated. After finding out the count of ambulances that need to be relocated, a decision has to be made to to find out how many ambulances will be relocated to and from every base station.

Let R_{ij} be the decision variable stating the amount of ambulances that need to be relocated from base station *i* to other base stations *j*. Let t_{ij} denotes the travelling time which is defined as the time that an ambulance will take to travel from station *i* to station *j*. This value will be evaluated using Google Distance Matrix API.The value of c_{ij} denotes the transportation cost that will incur by ambulance in making a movement from station *i* to station *j*. A constant or a fixed value of cost will also be associated with every ambulance relocation and is denoted as f_{ij}. Another decision variable is denoted by p_{ij} which tells us whether the relocation has been made or not from station *i* to station *j* where $i, j \in N$ and $i \neq j$. The value of p_{ij} will be 1 if relocation has been made and 0 otherwise.These two values denoted as t_{ij} and c_{ij} where $i, j \in N$ and $i \neq j$ are associated with each relocation.

3.4.5 Objective function for minimizing relocation cost

Using the above parameters, the mathematical formulation for minimizing the relocation cost (RLC) can be written as shown in Equation 19.

$minRLC = \sum_{i\in N} \sum_{j\in N i\neq j} (R_{ij} * c_{ij} + f_{ij} * p_{ij})$ (19)
subject to:
$\sum_{i\in N} R_{ij} = (x_j^l - x_j^p) \quad j\forall N \quad , i \neq j$ (20)
$\sum_{j\in N} R_{ij} = (x_i^p - x_i^l) \quad i\forall N \quad , i \neq j$ (21)
$R_{ij} \leq p_{ij}.M \; i,j \in N \;,\; i \neq j$ (22)
$R_{ij} \geq 0 \quad i,j \in N \;,\; i \neq j$ (23)
$p_{ij} \in \{0,1\} \; i,j \in N \;,\; i \neq j$ (24)

Equation 19 minimizes the relocation transporting cost and fixed cost of the ambulances. The number of ambulances that are relocated is limited using constraints (20) and (21). M is considered as a very large value in Equation (22). The values of decision variables are defined using constraints (23) and (24).

4.Numerical experiments, results and discussions

4.1Numerical experiments

Numerical experiments are performed here to find an optimized allocation plan having minimum ART. The correlation and impact of different attributes (count of ambulances, the frequency of requests) on ART are also analyzed in this section. The experiments and their results are explained in detail below. The framework is executed 20 times for GA and PSO, each as OCs handling random requests generated from the areas covered by APZ and ABS. The population is initialized for GA using the proportion-based seeding method. During the execution of the seeding method for the stated scenario, a benchmark value *'AT'* of 12.06 minutes is attained. The evolution graph depicting the variation in the value of ART with respect to iterations for PSO and GA is shown in *Figure 11*. It is evident from the graph that the value of global fitness has changed from 13.533 to 12.47 minutes in 1000 iterations of PSO and 13.355 to 12.12 minutes in 1000 iterations of GA. The graph also states that the convergence rate of PSO is faster than GA; where PSO converges in 223 iterations and GA converges in 728 iterations. Although PSO shows a fast convergence rate it fails at attaining the objective of this work i.e. minimum value of ART.

The fitness values obtained in 20 different executions of the SO framework using GA and PSO have been used here to infer the constancy and repeatability of the algorithms for this work. The graphs in *Figure 12* and *Figure 13* depict that the changes in the fitness values are from 12.4 min to 12.12 minutes in the case of GA and from 13.0202 minutes to 12.175 minutes in the case of PSO. The consistency of any algorithm can be measured by the value of variance. An algorithm is said to be consistent if the value of variance is between 0 and 1. In the proposed work, the value of variance is 0.0681 for PSO and 0.0054 for GA signifying that both the algorithms are stable and consistent. However, the value is less for GA showing it to be more consistent than PSO with respect to this work.

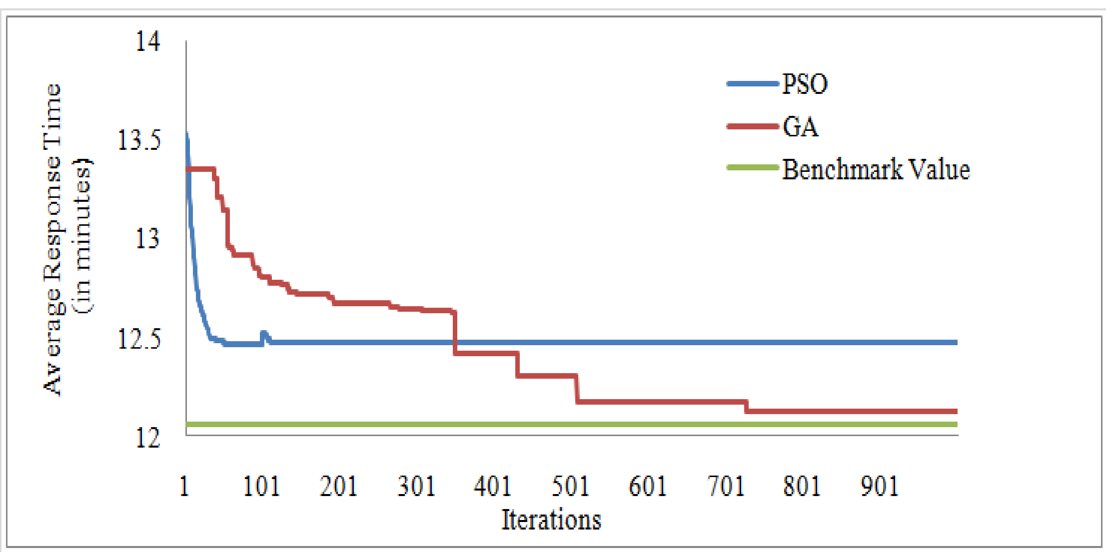


Figure 11 Evolution graph of GA and PSO

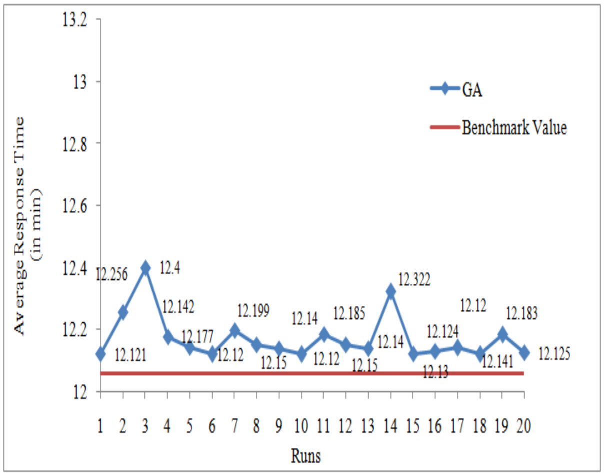


Figure 12 Consistency graph of GA

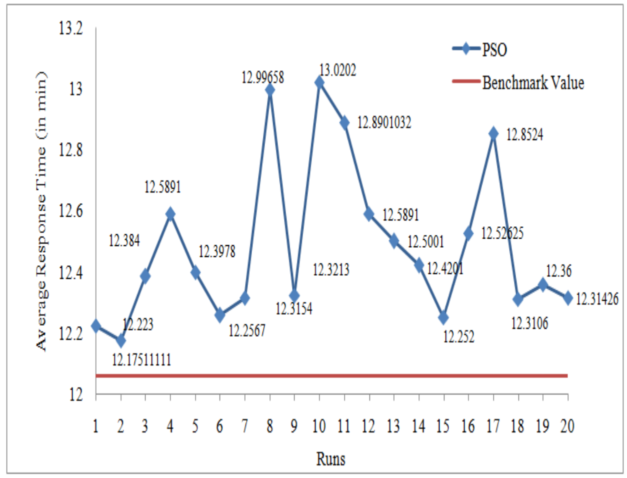


Figure 13 Consistency graph of PSO

Since GA is better than PSO in our work, further experiments are performed using only GA as OC. The different experiments that are performed are as follows:

i) Experiment 1**:** The count of ambulance in the fleet is changed to observe the impact on deployment plans.

ii) Experiment 2: The count of ambulance in the fleet is changed to observe the impact on the ART.

iii) Experiment 3: The frequency of requests is changed to observe the impact on deployment plans.

iv)Experiment 4: The frequency of requests is changed to observe the impact on the ART.

v)Experiment 5: The impact of relocation movements on the relocation cost.

4.2Results and discussion

The results of all the numerical experiments performed are shown and discussed in the following section.

a) Experiments 1 and 2 are conducted to observe the impact of the count of ambulances on deployment plans and ART. The deployment plan obtained for the different counts of ambulances in the fleet is shown in *Table 1*. The easiest way to deploy the ambulances is to place an equal number of ambulances at all the base stations. However, this would not handle the needs of the city as the demand rate varies at different base stations. Therefore, the values of coefficient of variance (CoV) and standard deviation (SD) are calculated to assess the uneven status of the ambulance count among the base stations. Moreover, it is a general perception that the response time of EMS can be reduced by deploying more ambulances in the fleet to serve the requests. To validate this assumption, the experiment is conducted to find the maximum count of ambulances that can be placed in the fleet at the existing base stations so that minimum ART (same or nearly same to the benchmark value) can be achieved. It can be concluded from *Table 1* and the graph in *Figure 14* that the value of ART is indirectly proportional to the number of ambulances after a threshold count. For a count of 130, 140, 150, and 160 ambulances, the average response time is 16.52 min, 14.24 min, 13 min, and 12.12 min respectively. However, for the count of ambulances 170, 180, and 190 the value of ART is the same as that of 160 i.e. 12.12 min.

Table 1 Optimal deployment plan for different ambulance count

Count of ambulances	Optimal deployment plan (Xi)	SD	CoV	ART
130	{11,9,1,10,3,14,6,7,5,9,5,1,4,3,18,6,1,1,1,1,1,5,1,1,1,1,3,1}	4.4336	0.9549	16.52
140	{11,9,1,11,5,14,6,8,5,9,6,1,5,3,14,6,1,1,1,1,1,6,1,1,6,1,5,1}	4.0532	0.8106	14.24
150	{11,10,1,8,7,13,7,7,4,10,5,1,5,4,16,5,2,1,1,1,1,8,4,1,9,1,4,3}	4.0463	0.7553	13
160	{16,8,3,8,6,15,5,8,5,6,5,5,8,3,15,7,2,7,1,1,3,8,4,1,3,1,5,1}	4.0871	0.7152	12.12
170	{16,8,3,8,6,15,6,7,5,7,5,5,8,3,15,7,3,7,3,2,3,8,4,2,3,4,5,2}	3.7601	0.6193	12.12
180	{16,8,4,8,6,15,6,7,5,7,5,5,8,4,15,7,4,7,4,4,3,8,4,4,3,4,5,4}	3.4582	0.5380	12.12
190	{16,8,5,8,6,15,6,7,5,7,5,5,8,5,15,7,5,7,5,5,5,8,4,4,5,4,5,5}	3.2111	0.4732	12.12

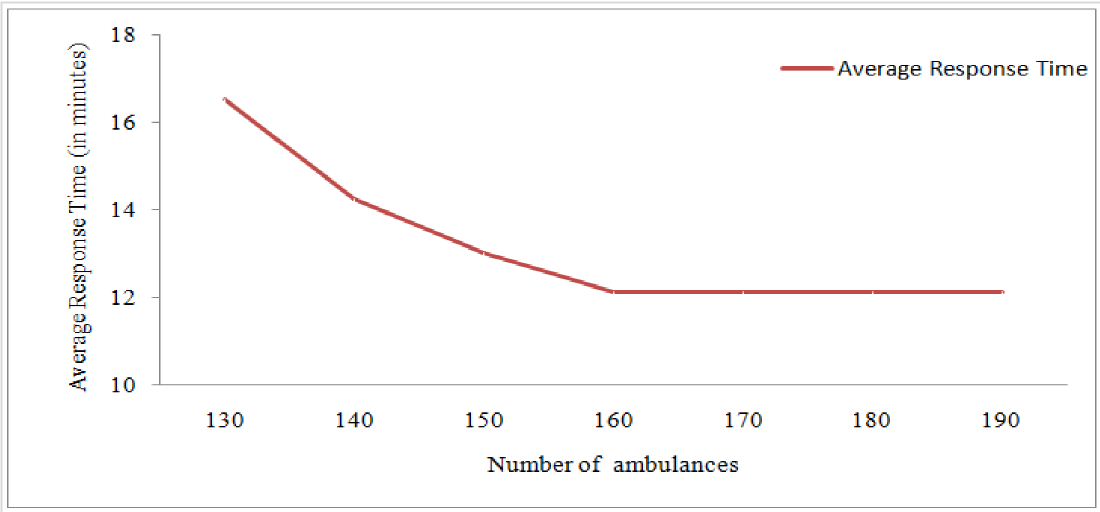


Figure 14 Change in ART WRT count of ambulances

Two inferences can be drawn from experiments 1 and 2. They are as follows:
(i) The count of ambulances in the fleet should be increased from 150 ambulances to 160 ambulances to reduce the value of ART from 13 minutes to 12.12 minutes i.e. by 6.77%.
(ii) Having more than 160 ambulances in the fleet will not further reduce the value of ART.

b) Experiments 3 and 4 are performed using the new fleet count of 160 ambulances to observe the impact of demand rate on the deployment plan and value of ART. *Table 2* shows the different deployment plans of 160 ambulances obtained by varying frequency of requests per minute. The values of SD and CoV have been calculated for each plan to monitor the variability in the deployment plan. The results obtained are used to plot a graph to derive the relationship between the arrival rate of requests and ART. It is observed from the graph shown in *Figure 15* that the value of ART increases evidently when the average time between the requests is less than 0.5 but does not decrease when the average time between the requests is more than 0.5.

Thus, it can be inferred from experiment 3 and 4 that a fleet count of 160 ambulances having the deployment plan {16, 8, 3, 8, 6, 15, 5, 8, 5, 6, 5, 5, 8, 3, 15, 7, 2, 7, 1, 1, 3, 8, 4, 1, 3, 1, 5, 1} as mentioned in *Table 1* and *Table 2* is sufficient to handle 2 requests per minute with ART equal to 12.12 minutes which is near to the benchmark value.

The new deployment plan (fleet 160 ambulances) with ART (12.12 min) and the existing deployment plan (fleet 151 ambulances) with ART (13 min) is shown in *Table 3*. The SO framework is executed in different time frames of a day to achieve deployment plans of 160 ambulances for peak hours and lean hours as shown in *Table 4*. This data is then used to relocate the ambulances. and observe the change in relocation cost and movement.

c) The experiments from 1 to 4 show that with the new deployment plan of 160 ambulances, requests can be served with better ART value. However, the deployment plan for peak and lean hours is different and requires the ambulances to be relocated. To eliminate the overhead of relocating the ambulances, this experiment is performed to observe if the requests during peak hours can be fulfilled with the same or nearly the same ART value using the deployment plan of the lean hour. As shown in *Table 5*, the value of ART for fulfilling the requests in peak hours was evaluated using
(i) deployment plan of the lean hour;
(ii) deployment plan of peak hours.

The values of ART for the two cases are 13.5113 minutes and 12.12502 minutes respectively. From the values obtained, it is noted that ART(i) > ART (ii). This makes it clear that the deployment plan of the lean hour and peak hours should be different to maintain the operational performance of EMS throughout the day. Therefore, some ambulances should be relocated amongst the base stations between peak hours and lean hours.

The count of ambulances that should be relocated is also shown in the table. From the table, it is observed that increasing the count of ambulances from 150 to 160 shows decrement in the count of ambulances that need to be relocated and the distance travelled for relocation.

Table 2 Optimal deployment plan for different request arrival rate

Average time between requests (in min)	Optimal deployment plan (X_i)	SD	CoV	ART(in min)
0.25	{17,11,2,6,1,18,5,8,5,3,6,11,4,3,17,4,2,7,1,1,5,11,2,1,4,1,3,1}	5.0133	0.8773	32.3
0.3	{17,9,1,6,1,15,7,4,5,7,6,4,6,3,20,6,2,10,1,1,5,5,2,1,4,6,5,1}	4.74234	0.8299	16.59
0.4	{16,8,2,9,5,15,6,8,5,7,5,5,7,3,15,7,2,8,1,1,3,9,3,1,2,1,5,1}	4.2161	0.7378	13.3
0.5	{16,8,3,8,6,15,5,8,5,6,5,5,8,3,15,7,2,7,1,1,3,8,4,1,3,1,5,1}	4.0871	0.7152	12.12
0.6	{16,8,3,7,2,14,5,5,5,6,7,5,3,3,16,5,3,7,1,5,3,9,2,5,4,5,4,2}	3.8346	0.6711	12.119
0.7	{15,9,5,6,3,13,4,3,10,8,5,2,5,3,15,7,3,5,3,3,5,7,1,2,4,8,4,2}	3.7211	0.6512	12.119
0.8	{13,7,3,6,5,11,5,5,6,5,6,5,6,5,11,6,3,6,5,2,2,7,4,6,5,5,4,6}	2.4328	0.4257	12.119

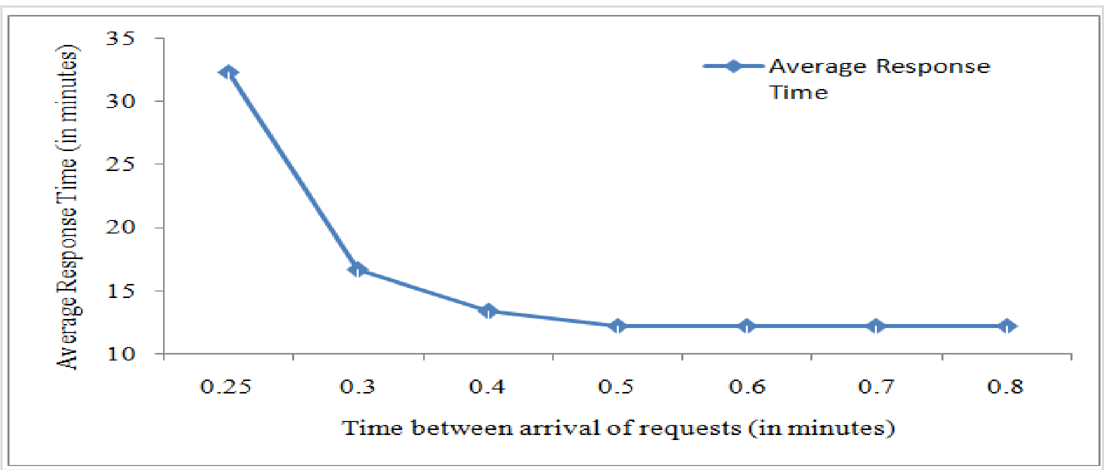


Figure 15 Change in ART wrt arrival time between requests

Table 3 Current and new deployment plan for 28 base stations

Deployment	Base Station Index									
Plan (Xi)	BS 1	BS 2	BS 3	BS 4	BS 5	BS 6	BS 7	BS 8	BS 9	BS 10
Current Plan of 151 ambulances 151 Ambulances	11	10	1	8	7	13	7	7	4	10
New Plan of 160 ambulances	16	8	3	8	6	15	5	8	5	6
Deployment	**Base Station Index**									
Plan (X_i)	BS 11	BS 12	BS 13	BS 14	BS 15	BS 16	BS 17	BS 18	BS 19	BS 20
Current Plan of 151 ambulances 151 Ambulances	5	1	5	4	16	6	2	1	1	1
New Plan of 160 ambulances	5	5	8	3	15	7	2	7	1	1
Deployment	**Base Station Index**									
Plan (X_i)	BS 21	BS 22	BS 23	BS 24	BS 25	BS 26	BS 27	BS 28		
Current Plan of 151 ambulances 151 Ambulances	1	8	4	1	9	1	4	3		
New Plan of 160 ambulances	3	8	4	1	3	1	5	1		

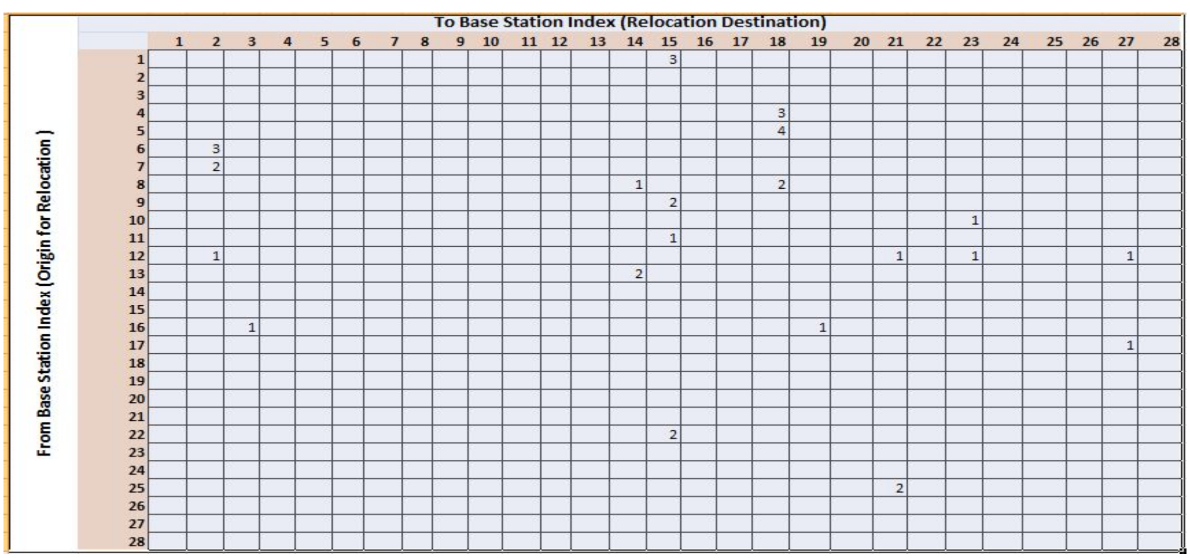
Table 4 Optimal deployment plan for peak hours and lean hours

Deployment Plan	Base Station Index									
	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8	BS9	BS10
Lean hours 151 Ambulances	16	8	3	8	6	15	5	8	5	6
Peak hours	13	14	4	5	2	12	3	5	3	5
Deployment Plan	**Base Station Index**									
	BS11	BS12	BS13	BS14	BS15	BS16	BS17	BS18	BS19	BS20
Lean hours 151 Ambulances	5	5	8	3	15	7	2	7	1	1
Peak hours	4	1	6	6	23	5	1	16	2	1
Deployment Plan	**Base Station Index**									
	BS21	BS22	BS23	BS24	BS25	BS26	BS27	BS28		
Lean hours 151 Ambulances	3	8	4	1	3	1	5	1		
Peak hours	6	6	6	1	1	1	7	1		

The table also highlights that for a count of 170 ambulances in the fleet, the number of ambulances needed to be relocated and the distance travelled for relocation is minimum. But increasing the fleet count from 150 to 160 is more cost-effective in all the aspects (results of experiments 1-4) than increasing the fleet count from 150 to 170. Therefore, it can be concluded that with a fleet count of 160 ambulances with different allocation plans for peak and lean hours, an ART value of 12.12 minutes can be achieved. The relocation movement of 35 ambulances during peak and lean hours is shown in *Figure 16.* The grid value in *Figure 16* indicates the count of ambulances that are relocated from one base station to another base station.

Table 5 Comparison of relocation movement for different scenarios

Number of ambulances	150	160	170
ART for peak hours with the old deployment plan (in min)	14.566	13.5113	13.3982
Number of ambulances relocated	42	35	34
ART for peak hours with the new deployment plan (in min)	12.78822	12.12502	12.48606
Total distance travelled by the ambulances during relocation (in km)	583.506	424.253	341.946


To Base Station Index (Relocation Destination)

From Base Station Index (Origin for Relocation)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1															3													
2																												
3																												
4																		3										
5																		4										
6		3																										
7		2																										
8														1				2										
9															2													
10																							1					
11															1													
12		1																			1		1				1	
13														2														
14																												
15																												
16			1																1									
17																											1	
18																												
19																												
20																												
21																												
22															2													
23																												
24																												
25																					2							
26																												
27																												
28																												

Figure 16 Relocation table

4.2.1 Key findings
The key findings of the experiments are stated below.
GA provides better-optimized results than PSO in this work as it provides a lesser value for ART.
Increasing the fleet count from 151 to 160 improves the value of ART by 6.7%.
Increasing the fleet count beyond 160 at the existing base stations does not give any significant change in the ART value.
The proposed deployment plan can handle 2 requests per minute with an ART of 12.12 minutes.
Relocation activity maintains the value of ART throughout the day.
In the case of CATS EMS, Delhi, a count of 160 ambulances allocated at the existing base stations can handle an average of 2 requests per minute with an ART of 12.12 minutes**.**

4.2.2 Managerial implications
Some managerial implications can be drawn from the numerical experiments conducted for this work. They can be summarized as follows:
The value of ART is not always positively influenced by the arrival rate of request; if the rate of request is below a certain threshold, then the value of ART shows no significant change.
The number of ambulances does not always show a positive influence on the value of ART. In other words, when the number of ambulances exceeds a certain count, the value of ART remains constant at a certain value rather than decreasing gradually.
A good deployment plan provides a good service level. Therefore, the ambulances should be deployed strategically and not in a random or balanced way to ease out the task of deployment.
The variation in the distribution of requests significantly impacts ambulance relocation. There will be a greater movement distance involved with ambulance relocation activities if the distribution of requests varies more during time intervals.

4.2.3 Limitations of the work
There are certain limitations associated with this research that will be handled by the authors in future research. The proposed work does not take into consideration heterogeneous ambulances categorized based on types of equipment, capabilities of the crew, and experience of the driver. In addition, the current work focuses on handling the requests raised by accident and trauma victims handled by CATS and can be extended to handle all sorts of request calls raised across the city. In future research, the authors will also work on handling the relocation of ambulances dynamically by finding a new location while it is returning after providing service. A complete list of abbreviations is shown in *Appendix I.*

5.Conclusion and future work
In this paper, we discussed on improving the ambulance-based medical service for trauma victims. Allocation and relocation plans for the ambulances have been proposed using a tightly coupled framework of AAC and OC. A new proportion-based seeding method of population initialization was used to get good population individuals in GA. The application of the work was carried out for the data of Delhi and CATS EMS organization to find an optimal deployment plan by minimizing the value of ART of all the demand requests. The work also aimed at finding the maximum count of ambulances that can be added to the existing CATS EMS structure to reduce the ART. Different rates of arriving requests, traffic conditions, and other spatial patterns were taken into consideration to map to real-time situations. The work has been supported by results obtained and inferences drawn after conducting some numerical experiments. The results of the work show that a maximum count of 9 ambulances can be added to the existing fleet to improve the average response time from 13 min to 12.12 minutes i.e by 6.77%. On adding more ambulances, the value of ART will not decrease until more base stations are constructed at new viable locations. The reduction in the response time will improve the efficiency of EMS. The results also state that the proposed deployment plan will help in servicing 2 demands per minute. Relocation activities of ambulances will balance the performance of EMS throughout the day.

In the future, the authors will focus on ambulance allocation considering heterogeneous ambulances categorized based on types of equipment, capabilities of the crew, and experience of the driver. The authors will also work towards relocating the ambulances dynamically by finding a new location for the ambulance while it is returning after providing service.

Acknowledgment
The CATS EMS organization and Jamia Millia Islamia University are acknowledged for supporting this work.

Conflicts of interest
The authors have no conflicts of interest to declare.

Author’s contribution statement
Zaheeruddin: Supervision, methodology, writing-reviewing and editing. **Hina Gupta:** Conceptualization, methodology, data collection, writing-original draft preparation.

Zaheeruddin and Hina Gupta.

References

[1] Matinrad N, Reuter-oppermann M. A review on initiatives for the management of daily medical emergencies prior to the arrival of emergency medical services. Central European Journal of Operations Research. 2022; 30(1):251-302.

[2] Bélanger V, Ruiz A, Soriano P. Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles. European Journal of Operational Research. 2019; 272(1):1-23.

[3] Prabhu SB, Ravithejaswi UR, Shetty S, Hegde SS, Prasad SM. NavIC driven dynamic ambulance allocation and tracking. In international students' conference on electrical, electronics and computer science 2022 (pp. 1-6). IEEE.

[4] Andersson H, Granberg TA, Christiansen M, Aartun ES, Leknes H. Using optimization to provide decision support for strategic emergency medical service planning–three case studies. International Journal of Medical Informatics. 2020.

[5] Tshokey T, Tshering U, Lhazeen K, Abrahamyan A, Timire C, Gurung B, et al. Performance of an emergency road ambulance service in Bhutan: response time, utilization, and outcomes. Tropical Medicine and Infectious Disease. 2022; 7(6):87.

[6] Zhang R, Zeng B. Ambulance deployment with relocation through robust optimization. IEEE Transactions on Automation Science and Engineering. 2018; 16(1):138-47.

[7] Zhen L, Wang K, Hu H, Chang D. A simulation optimization framework for ambulance deployment and relocation problems. Computers & Industrial Engineering. 2014; 72:12-23.

[8] Sharma N, Bairwa M, Gowthamghosh B, Gupta SD, Mangal DK. A bibliometric analysis of the published road traffic injuries research in India, post-1990. Health Research Policy and Systems. 2018; 16(1):1-11.

[9] Singh SK. Road traffic accidents in India: issues and challenges. Transportation Research Procedia. 2017; 25:4708-19.

[10] Blanchard IE, Doig CJ, Hagel BE, Anton AR, Zygun DA, Kortbeek JB, et al. Emergency medical services response time and mortality in an urban setting. Prehospital Emergency Care. 2012; 16(1):142-51.

[11] https://healthmarketinnovations.org/program/centralized-ambulance-trauma-services-/cats-delhi. Accessed 28 February 2022.

[12] Toregas C, Swain R, Revelle C, Bergman L. The location of emergency service facilities. Operations Research. 1971; 19(6):1363-73.

[13] Church R, Revelle C. The maximal covering location problem. In papers of the regional science association 1974 (pp. 101-18). Springer-Verlag.

[14] Zonouzi MN, Kargari M. Modeling uncertainties based on data mining approach in emergency service resource allocation. Computers & Industrial Engineering. 2020.

[15] Daskin MS. A maximum expected covering location model: formulation, properties and heuristic solution. Transportation Science. 1983; 17(1):48-70.

[16] Erkut E, Ingolfsson A, Erdoğan G. Ambulance location for maximum survival. Naval Research Logistics. 2008; 55(1):42-58.

[17] Gendreau M, Laporte G, Semet F. Solving an ambulance location model by tabu search. Location Science. 1997; 5(2):75-88.

[18] Gendreau M, Laporte G, Semet F. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. Parallel Computing. 2001; 27(12):1641-53.

[19] Schmid V, Doerner KF. Ambulance location and relocation problems with time-dependent travel times. European Journal of Operational Research. 2010; 207(3):1293-303.

[20] Shariat-mohaymany A, Babaei M, Moadi S, Amiripour SM. Linear upper-bound unavailability set covering models for locating ambulances: application to Tehran rural roads. European Journal of Operational Research. 2012; 221(1):263-72.

[21] Rodriguez SA, Rodrigo A, Aguayo MM. A simulation-optimization approach for the facility location and vehicle assignment problem for firefighters using a loosely coupled spatio-temporal arrival process. Computers & Industrial Engineering. 2021.

[22] Benabdouallah M, Yaakoubi OE, Bojji C. Genetic algorithm hybridised by a guided local search to solve the emergency coverage problem. International Journal of Mathematical Modelling and Numerical Optimisation. 2017; 8(1):23-41.

[23] Apornak A, Raissi S, Keramati A, Khalili-damghani K. Human resources optimization in hospital emergency using the genetic algorithm approach. International Journal of Healthcare Management. 2021; 14(4):1441-8.

[24] Rivera G, Cisneros L, Sánchez-solís P, Rangel-valdez N, Rodas-osollo J. Genetic algorithm for scheduling optimization considering heterogeneous containers: a real-world case study. Axioms. 2020; 9(1):1-16.

[25] Paul PV, Moganarangan N, Kumar SS, Raju R, Vengattaraman T, Dhavachelvan P. Performance analyses over population seeding techniques of the permutation-coded genetic algorithm: an empirical study based on traveling salesman problems. Applied Soft Computing. 2015; 32:383-402.

[26] Victer PP, Ganeshkumar C, Dhavachelvan P, Baskaran R. A novel ODV crossover operator-based genetic algorithms for traveling salesman problem. Soft Computing. 2020; 24(17):12855-85.

[27] Mundhenk T, Landajuela M, Glatt R, Santiago CP, Petersen BK. Symbolic regression via deep reinforcement learning enhanced genetic programming seeding. Advances in Neural Information Processing Systems. 2021; 34:24912-23.

[28] Mirshekarian S, Süer GA. Experimental study of seeding in genetic algorithms with non-binary genetic

representation. Journal of Intelligent Manufacturing. 2018; 29(7):1637-46.
[29] Kolomvatsos K, Panagidi K, Hadjiefthymiades S. Optimal spatial partitioning for resource allocation. In ISCRAM 2013 (pp.747-57).
[30] Hajipour V, Pasandideh SH. Proposing an adaptive particle swarm optimization for a novel bi-objective queuing facility location model. Economic Computation and Economic Cybernetics Studies and Research. 2012; 46(3):223-40.
[31] Tsai Y, Chang KW, Yiang GT, Lin HJ. Demand forecast and multi-objective ambulance allocation. International Journal of Pattern Recognition and Artificial Intelligence. 2018; 32(7).
[32] Swalehe M, Aktas SG. Dynamic ambulance deployment to reduce ambulance response times using geographic information systems: a case study of Odunpazari district of Eskisehir province, Turkey. Procedia Environmental Sciences. 2016; 36:199-206.
[33] Jagtenberg CJ, Bhulai S, Van DMRD. Dynamic ambulance dispatching: is the closest-idle policy always optimal? Health Care Management Science. 2017; 20(4):517-31.
[34] Schmid V. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. European Journal of Operational Research. 2012; 219(3):611-21.
[35] Naoum-sawaya J, Elhedhli S. A stochastic optimization model for real-time ambulance redeployment. Computers & Operations Research. 2013; 40(8):1972-8.
[36] Enayati S, Mayorga ME, Rajagopalan HK, Saydam C. Real-time ambulance redeployment approach to improve service coverage with fair and restricted workload for EMS providers. Omega. 2018; 79:67-80.
[37] Ji S, Zheng Y, Wang W, Li T. Real-time ambulance redeployment: a data-driven approach. IEEE Transactions on Knowledge and Data Engineering. 2019; 32(11):2213-26.
[38] Yavari M, Maihami R, Esmaeili M. Ambulance dispatching and relocation problem considering overcrowding of emergency departments. IISE Transactions on Healthcare Systems Engineering. 2022:1-2.
[39] Wilde ET. Do emergency medical system response times matter for health outcomes? Health Economics. 2013; 22(7):790-806.
[40] https://delhitrafficpolice.nic.in/sites/default/files/uploads/2020/Road-accident-in-delhi-2019. Accessed 23 June 2022.
[41] Bertsimas D, Ng Y. Robust and stochastic formulations for ambulance deployment and dispatch. European Journal of Operational Research. 2019; 279(2):557-71.
[42] Koenig KL. Emergency ambulance utilization in Harlem, New York (July 1985). Prehospital Emergency Care. 2022; 26(3): W1-7.
[43] El-masri S, Saddik B. An emergency system to improve ambulance dispatching, ambulance diversion and clinical handover communication—a proposed model. Journal of Medical Systems. 2012; 36(6):3917-23.
[44] Liu J, Li J, Wang K, Zhao J, Cong H, He P. Exploring factors affecting the severity of night-time vehicle accidents under low illumination conditions. Advances in Mechanical Engineering. 2019; 11(4).
[45] Pitchipoo P, Vincent DS, Rajakarunakaran S. Analysis of prime reasons for night time accidents in public transport corporations. In proceedings of international conference on advances in industrial engineering applications 2014.
[46] Carvalho AS, Captivo ME, Marques I. Integrating the ambulance dispatching and relocation problems to maximize system's preparedness. European Journal of Operational Research. 2020; 283(3):1064-80.
[47] Houck CR, Joines J, Kay MG. A genetic algorithm for function optimization: a Matlab implementation. Ncsu-ie tr. 1995; 95(9):1-10.
[48] Yong-jie M, Wen-xia Y. Research progress of genetic algorithm. Application Research of Computers. 2012; 4:1201-6.
[49] Li X, Xiao N, Claramunt C, Lin H. Initialization strategies to enhancing the performance of genetic algorithms for the p-median problem. Computers & Industrial Engineering. 2011; 61(4):1024-34.
[50] Bajer D, Martinović G, Brest J. A population initialization method for evolutionary algorithms based on clustering and Cauchy deviates. Expert Systems with Applications. 2016; 60:294-310.
[51] Xu P, Luo W, Xu J, Qiao Y, Zhang J, Gu N. An alternative way of evolutionary multimodal optimization: density-based population initialization strategy. Swarm and Evolutionary Computation. 2021.
[52] Poli R, Kennedy J, Blackwell T. Particle swarm optimization. Swarm Intelligence. 2007; 1(1):33-57.

Zaheeruddin is currently Professor in the Department of Electrical Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia (A Central University), New Delhi, India since 2003. He was Head of the Department of Electrical Engineering from 2011 to 2014. He joined the Department as a Lecturer from the very beginning of its establishment in 1986. He received the B. Sc. Engg. Degree in Electrical and M. Sc. Engg. Degree in Electronics and Communication from Aligarh Muslim University (AMU), Aligarh (UP) in 1982 and 1988 respectively, and a Ph. D. Degree in Computer Science and Technology from Jawaharlal Nehru University (JNU), New Delhi in 2002. He has published eighty (80) research papers in International Journals and Conferences in the areas of Soft Computing, Noise Pollution, Wireless Sensor Networks, Optimization Techniques, and Smart Grid.
Email: zaheeruddin@jmi.ac.in

Zaheeruddin and Hina Gupta.

Hina Gupta is a PhD scholar in the Department of Electrical Engineering, Jamia Millia Islamia, Central University, New Delhi, India. She has a teaching experience of 6 years. She completed her B.Tech in 2010 and M.Tech in 2016. Her area of interests are Networking, Cloud Computing, Internet of Things. Her current area of research is Optimization of the Healthcare Facilities. She has published seven (7) research papers in International Journals and Conferences.
Email: guptahina189@gmail.com

Appendix I

S. No.	Abbreviation	Description
1	AAC	Ambulance Assignment Component
2	ABS	Accident Black Spot
3	API	Application Programming Interface
4	APZ	Accident Prone Zone
5	ART	Average Response Time
6	BS	Base Station
7	CATS	Centralized Accident and Trauma Services
8	CoV	Coefficient of Variance
9	DSM	Double Standard Model
10	EMS	Emergency Medical Service
11	GA	Genetic Algorithm
12	LSCP	Location Set Covering Model
13	MCLP	Maximal Covering Location Problem
14	MEXCLP	Maximum Expected Covering Location Problem
15	OC	Optimization Component
16	PSO	Particle Swarm Optimization
17	RTC	Road Traffic Crash
18	SD	Standard Deviation
19	SO	Simulation Optimization