SN Computer Science (2026) 7:191
https://doi.org/10.1007/s42979-026-04776-1

ORIGINAL RESEARCH

A Clustering-Based Method for Reducing the Search Space for Mobile Stroke Unit Allocation

Muhammad Adil Abid[1] · **Johan Holmgren**[1] · **Fabian Lorig**[1] · **Jesper Petersson**[2,3]

Received: 7 August 2025 / Accepted: 20 January 2026

Abstract

Mobile Stroke Units (MSUs), which are specialised ambulances equipped with brain imaging devices and trained medical personnel, offer the potential for rapid on-site diagnosis and treatment, improving patient outcomes in prehospital stroke care. To fully realise their benefits, it is essential to strategically allocate. However, identifying optimal locations within a region for MSU deployment is typically computationally challenging due to the vast number of possible combinations of ambulance stations. Existing methods suffer from computational inefficiency, as they consider the whole search space when looking for the optimal solution to the MSU allocation problem. In the current paper, we propose a framework, *Quality Clustering for Reducing the Search Space (QCRSS)*, which uses clustering as a preprocessing step to significantly reduce the number of candidate MSU locations while maintaining high solution quality for the MSU allocation problem. In a real-world use case study, we evaluate our proposed framework in Sweden's southern healthcare region. Extensive experiments across multiple MSU settings demonstrate that QCRSS achieves the optimal solution for two, three, and four MSUs, and a highly satisfactory solution even for the larger and more complex case of five MSUs. The proposed framework reduces the search space by 5x, 11x, 26x, and 67x for two, three, four, and five MSUs, respectively. We illustrate the performance of QCRSS through both qualitative and quantitative analyses.

Keywords Decision support system · Ambulance allocation · Optimisation · Healthcare · Mobile stroke unit · Clustering · Reducing search space · Prehospital stroke care

Introduction

Stroke is a serious neurological condition caused by either a blockage in a blood vessel (ischemic stroke) or a rupture of a blood vessel (hemorrhagic stroke), both of which disrupt the blood supply to the brain [1]. Without timely medical intervention, a stroke can result in irreversible brain damage, long-term disability, or death. Globally, stroke remains a major public health concern, affecting one in six people over their lifetime, with an estimated 15 million cases and 5.8 million deaths annually [2]. In Sweden, more than 21,000 people suffer a stroke each year, of which approximately 3900 cases occur in the Southern Healthcare Region (SHR) [3], which is the focus of this study. The SHR includes a mix of densely populated urban centers and remote rural areas, presenting significant challenges for timely and equitable prehospital stroke care.

Beyond the immediate health risks, stroke also imposes a substantial long-term burden, leading to disability, reduced quality of life, and financial strain for patients, caregivers,

✉ Muhammad Adil Abid
muhammad.adil-abid@mau.se

Johan Holmgren
johan.holmgren@mau.se

Fabian Lorig
fabian.lorig@mau.se

Jesper Petersson
jesper.petersson@skane.se

[1] Department of Computer Science and Media Technology, Malmö University, 21119 Malmö, Sweden

[2] Department of Neurology, Lund University, 22185 Lund, Sweden

[3] Department of Health Care Management, Region Skåne, 21428 Malmö, Sweden

and the healthcare system [4, 5]. Timely treatment is critical, as early intervention significantly increases the likelihood of recovery [6]. However, delivering prompt and accurate treatment remains difficult due to logistical constraints and the clinical challenge of distinguishing between ischemic and hemorrhagic strokes at the patient site [7]. While ischemic strokes require clot-dissolving treatments (thrombolysis or thrombectomy), hemorrhagic strokes demand rapid blood pressure reduction to control the bleeding. Given the similarity in early symptoms, an accurate and immediate diagnosis is crucial to avoid administering incorrect or even harmful treatments.

Mobile Stroke Units (MSUs) have emerged as a transformative innovation for improving prehospital stroke care [8, 9]. These specialised ambulances are equipped with CT scanners and telemedicine capabilities, enabling on-site stroke diagnosis and treatment initiation, including thrombolysis, under the guidance of remote stroke experts [10]. This approach can significantly reduce the time to treatment, often eliminating the delay associated with hospital transport. Despite their clinical promise, MSUs are resource-intensive and costly to operate [11], limiting the number of units a healthcare region can afford. Therefore, strategic placement of MSUs is essential to maximise their impact across the region [12, 13]. This leads to the MSU allocation problem, an optimisation problem that aims to identify the optimal locations for a fixed number of MSUs at existing ambulance station locations within a geographic area. In this paper, we focus exclusively on the efficiency perspective, which seeks to maximise patient coverage by ensuring that as many patients as possible receive treatment within the shortest possible time window.

Previous studies have proposed various methods for solving the MSU allocation problem. For instance, Amouzad Mahdiraji et al. [14] apply exhaustive search, while their subsequent work [15] introduces a Mixed-Integer Linear Programming (MILP) model. However, both methods struggle to scale effectively due to the combinatorial explosion in larger regions [16]. More recent approaches, such as those by Abid et al. [16, 17], utilise genetic algorithms (GAs) to tackle the MSU allocation problem more efficiently. Still, these approaches typically evaluate solutions across the full search space, which increases the computational burden. In the current paper, we hypothesise that reducing the search space, by strategically filtering ambulance stations before optimisation, can significantly improve the optimisation process while maintaining high solution quality. This raises a key research question: How can we reduce the search space without significantly compromising the quality of the solution when solving the MSU allocation problem?

In the current paper, we propose the Quality Clustering for Reducing the Search Space (QCRSS) framework, which focuses on reducing the solution space in a heuristic preprocessing step to reduce the search space by exploiting the spatial distribution of ambulance stations for solving the MSU allocation problem. The ultimate aim is to allow the optimisation algorithm to explore a smaller, yet representative subset of possible solutions. In the QCRSS framework, we first perform a preprocessing step using clustering to group the ambulance locations (or stations). Thereafter, a single representative station is selected from each cluster. Finally, the MSU allocation problem is solved using the selected subset of ambulance locations. The central idea behind clustering is that geographically close ambulance stations are likely to have similar response times to emergency calls, making it sufficient to evaluate only representative locations from each cluster. This strategy preserves the spatial diversity of candidate locations while significantly reducing the number of combinations to evaluate, thereby improving computational efficiency without compromising solution quality. This paper extends our previous article [18] by introducing an enhanced and more generalised methodological framework for the proposed QCRSS. In contrast to our previous article, which primarily introduced clustering as a preprocessing step for reducing the search space in the MSU allocation problem, the current article provides a complete and systematic formulation of the QCRSS framework. We formalise its three key steps, explicitly showing how different clustering and optimisation methods can be flexibly integrated within each step in the proposed framework. Specifically, we introduce a generalised representative-selection strategy to select representatives when creating clusters with both actual and synthetic data points during the selection of representative ambulance stations (see Section "Step 2: Selection of Representatives Ambulance Stations"). We further extend the post-processing stage, that is, solving the MSU problem using the chosen representatives (see Section "Step 3: Solving the MSU Problem Using the ChosenRepresentatives"), in which we choose the number of clusters k, and show that algorithms can be applied to solve the MSU allocation. Furthermore, we demonstrate the generalisability and flexibility of the framework by incorporating an additional clustering method (i.e., K-means), thereby illustrating its adaptability. Moreover, to show more insights of the proposed framework, the current article provides an in-depth comparative analysis in terms of both qualitative and quantitative results across different clustering approaches, a systematic discussion of the hyperparameter k (for determining the suitable number of clusters) in search-space reduction, and expanded experimental evaluations across multiple MSU allocation settings.

The key contributions of the current paper are summarised as follows:

1. An optimisation framework, *Quality Clustering for Reducing the Search Space (QCRSS)*, which consists of a preprocessing step, a selection of representatives step and a problem-solving step to solve the MSU allocation problem. The core innovation of the framework lies in its heuristic, clustering-based, preprocessing step, which significantly reduces the search space for the MSU allocation problem.
2. Applying the QCRSS framework in a real-world case study in Sweden's Southern Healthcare Region (SHR), a geographically diverse area comprising both densely populated cities and sparsely populated rural areas, which poses a significant challenge for prehospital stroke care.
3. An illustration of how the QCRSS framework reduces the computational burden across various MSU deployment scenarios. The effectiveness of the approach is validated through comprehensive quantitative and qualitative analyses.

The remainder of this paper is structured as follows: Section "Related Work" reviews related work. Section "The Mobile Stroke Unit Allocation Problem" defines the MSU allocation problem. Section "Quality Clustering for Reducing the Search Space" presents the QCRSS methodology. Section "Computational Study" details our computational study. Finally, Section "Conclusions" concludes the paper and Section "Future Work" outlines some directions for future research.

Related Work

Emergency Medical Service (EMS) allocation is a critical challenge in healthcare logistics, directly influencing both patient outcomes and resource utilization. The strategic placement of limited medical resources to maximize service coverage and minimize response times has been a central focus in EMS research for a long time and continues to attract significant attention [19, 20]. Over the years, various approaches have been developed to address this problem, ranging from exact mathematical programming formulations to heuristic and simulation-based methods [21–23]. These approaches typically aim to balance competing objectives such as spatial coverage, population demand, operational constraints, and computational tractability.

Within the specific context of stroke care, the introduction of Mobile Stroke Units (MSUs) has added new complexity to the ambulance allocation problem. MSUs are specialized, high-cost ambulances equipped with imaging and telemedicine capabilities that enable on-site diagnosis and treatment of stroke patients [24–27]. The time-critical nature of stroke management, where treatment delays of even a few minutes can lead to severe neurological deterioration, further emphasizes the need for optimally positioned MSUs [28]. Consequently, developing efficient and scalable optimization strategies for MSU allocation has become a central research problem in the EMS field.

Early studies on MSU allocation primarily relied on exact optimization methods. Amouzad Mahdiraji et al. [14] utilised an exhaustive search (ES) method to identify optimal MSU placements for scenarios involving one to three MSUs across 39 potential ambulance locations in Sweden's Southern Healthcare Region (SHR). Their study evaluated each possible combination to determine its suitability and quality based on an objective function. While this brute-force approach guarantees optimality, its scalability is limited. As the number of MSUs increases, the number of possible combinations grows exponentially, resulting in a rapidly expanding search space. Consequently, the computational time becomes prohibitively high, rendering exhaustive search impractical for larger problem instances. In a follow-up study, Amouzad Mahdiraji et al. [15] proposed a mathematical optimisation model based on Mixed Integer Linear Programming (MILP) to solve the MSU allocation problem. The model was also applied to the SHR. Despite its theoretical rigour, solving the MILP formulation using the Gurobi solver [29] proved computationally challenging. Due to the model's complexity, the approach was limited to only two counties within the SHR, indicating limitations in scalability.

To improve scalability and computational efficiency, Abid et al. [16] introduced a genetic algorithm (GA) tailored for solving the MSU allocation problem. This method demonstrated increased scalability and broader regional applicability compared to previous models. However, one limitation of the traditional GA is its reliance on randomly generated initial populations. Such randomness can lead to poor initial diversity, with ambulance stations selected in close proximity to one another. This results in slow convergence, as the algorithm requires many generations to evolve better solutions.

To address this issue, Abid et al. [17] later proposed a clustering-based GA variant that strategically constructs the initial population. By selecting MSU placements from geographically dispersed clusters, the method ensures greater spatial diversity in the initial solutions, leading to improved coverage and enhanced convergence.

The main limitation of recent and existing methods for solving the MSU allocation problem is their reliance on considering the whole search space, that is, all possible combinations of ambulance station locations. The search space often contains numerous solutions with similar performance or quality. As a result, when considering the whole search space, optimisation algorithms waste computational resources evaluating less relevant ambulance station locations or suboptimal solutions. To overcome this challenge, the current article aims to develop a more efficient approach that strategically reduces candidate locations or search space to a smaller, high-quality subset without compromising solution quality.

The Mobile Stroke Unit Allocation Problem

In this section, we present a mathematical model for the mobile stroke unit (MSU) allocation problem, which aims to identify the optimal locations for a fixed number of MSUs at existing ambulance stations within a geographic area, covering the efficiency perspective. Efficiency refers to covering as many patients as possible to receive treatment in a shorter window of time.

Let I represent the set of existing ambulance stations in the considered region and N the total number of MSUs to allocate. Each ambulance station is assumed to always have at least one regular ambulance available. The geographic region R is subdivided into smaller areas r, where all patients located within subregion $r \in R$ are assumed to be at the centroid of r.

Let t_{ir}^{RA} be the expected time to treatment for a patient in subregion r when served by a regular ambulance located in $i \in I$, and let t_{ir}^{MSU} be the expected time to treatment if the patients are served by an MSU stationed at i. The expected time to treatment for a patient in subregion $r \in R$ when served by a regular ambulance is given by $t_r^{RA} = \min_{i \in I}\{t_{ir}^{RA}\}$. The values of t_r^{RA} ($r \in R$) and t_{ir}^{MSU} ($r \in R$, $i \in I$) can be precomputed and are parameters in the optimisation model.

We let Q_r ($r \in R$) denote the share of the stroke cases in R that is expected to take place in subregion r. The decision variables $x_i \in \{0, 1\}$ ($i \in I$) are defined as

$$x_i = \begin{cases} 1 & \text{if an MSU is placed at location } i \\ 0 & \text{otherwise.} \end{cases}$$

Using the decision variables x_i, the minimum expected time to treatment for a patient in subregion r when served by an MSU can be calculated as follows:

$$t_r^{MSU} = \min_{i \in I} \left\{ t_{ir}^{MSU} + (1 - x_i) \cdot M \right\}, \tag{1}$$

where M is a sufficiently large constant, such as the maximum expected time to treatment for any subregion r from any ambulance station i. This equation ensures that stations without an MSU are assigned such a long time to treatment that they will be excluded from consideration.

The objective function is the weighted average time to treatment across all subregions $r \in R$, which is formulated as follows:

$$\min z = \sum_{r \in R} Q_r \cdot \min \left\{ t_r^{RA}, t_r^{MSU} \right\}, \tag{2}$$

where the decision variables $(x_i, i \in I)$ are implicitly included in the calculation of t_r^{MSU} (see Eq. 1). The MSU allocation, represented by the x_i values, is constrained by

$$\sum_{i \in I} x_i = N, \tag{3}$$

which ensures that exactly N MSUs are allocated within the region.

Quality Clustering for Reducing the Search Space

The MSU allocation problem is a complex problem due to the need to search through a large combinatorial space of possible placements. To address this challenge, we introduce *Quality Clustering for Reducing the Search Space (QCRSS)*, which is designed to effectively reduce the computational burden associated with solving the MSU allocation problem. The core idea of the QCRSS is to utilise clustering as a preprocessing step to reduce the number of candidate locations under consideration without compromising solution quality too much. By exploiting the geographical distribution of ambulance stations, this method focuses on achieving better geographical spread and a structurally representative subset of locations, thus improving computational efficiency. This reduction in the search space allows the optimisation process to focus on a smaller, more relevant sub-search space. The QCRSS framework is composed of three main steps: (1) preprocessing using clustering, (2) selecting representatives for each cluster, and (3) solving the MSU allocation problem using the chosen representatives. A high-level overview of the QCRSS framework is illustrated in Fig. 1, and each step is detailed below. The proposed framework allows to apply different methods within each of the steps, which are discussed in more detail in the following sections.

Step 1: Preprocessing using Clustering

In the clustering-based preprocessing step, the central idea is that geographically close ambulance stations are likely

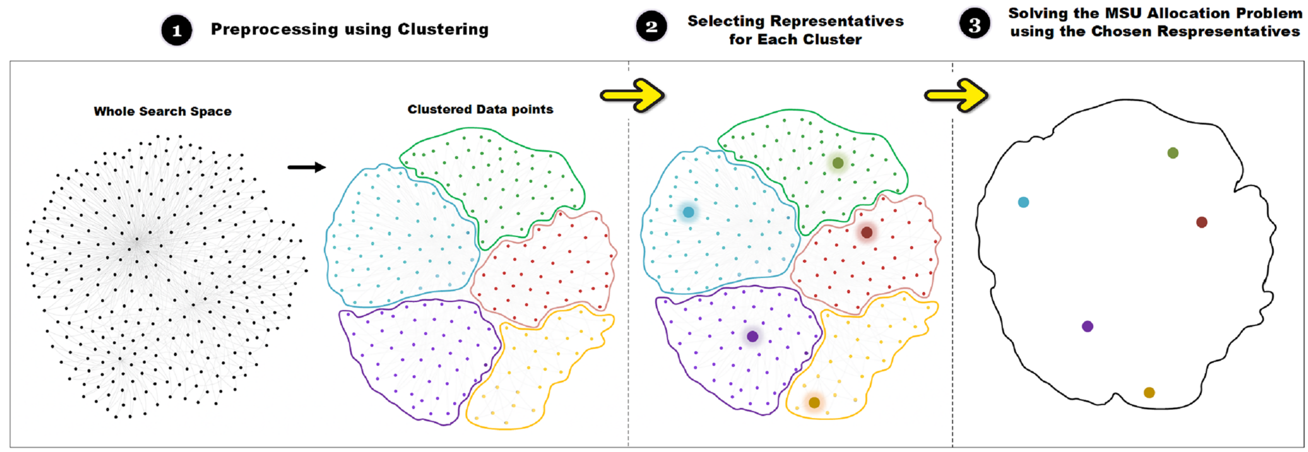


Fig. 1 Overview of the Quality Clustering for Reducing the Search Space (QCRSS) framework. The process begins with clustering as a preprocessing step (Step 1), followed by selecting representative locations from each cluster (Step 2). Finally, the optimisation problem (*i.e.,* MSU allocation) is solved using only the selected representatives (Step 3)

to have similar response times to emergency calls within their vicinity. By grouping these stations through clustering, we focus on evaluating a smaller set of representative locations by considering representatives from each cluster.[1] There is a large pool of clustering methods that can be used to group ambulance stations. In the current study, we explore three commonly used methods as use cases. These are (1) K-means clustering, (2) K-medoid clustering, and (3) Agglomerative hierarchical clustering (AHC). These methods are widely recognised for their simplicity, ease of implementation and effectiveness across various domains. Each method provides a different strategy for partitioning the data and selecting cluster representatives, which we explain in the next step (see Section "Step 2: Selection of Representatives Ambulance Stations"). Note that clustering can be performed using various features, such as geographical coordinates, ambulance travel times, or other relevant spatial attributes. We apply clustering based on the geographical coordinates of ambulance stations.

K-Means Clustering

The K-means clustering method is applied to a set of ambulance station locations I, where each location $i \in I$ has specific coordinates (x_i, y_i). The objective of K-means clustering is to partition I into K clusters, where each cluster is represented by a central point called a "centroid," which is the mean of all points in the cluster. The centroids minimise the total squared Euclidean distance between the ambulance stations and their respective centroids. See below for details about selecting the value of K.

The K-means clustering approach achieves minimal total squared distance through two main iterative steps:

1. Assignment Step: Each ambulance location i is assigned to the nearest centroid c_j, effectively minimising the distance between each location and its assigned centroid:

$$\mathcal{C}_j = \left\{ i \in I \mid j = \arg\min_{c_j \in C} \left\|(x_i, y_i) - c_j\right\|^2 \right\}, \tag{4}$$

   where $\mathcal{C}_j$ denotes the set of locations in cluster j, c_j is the centroid of cluster $\mathcal{C}_j$, and $\|(x_i, y_i) - c_j\|$ is the Euclidean distance between location i and centroid c_j.
2. Update Step: Each centroid c_j is updated to be the mean of all locations in the cluster:

$$c_j = \left(\frac{1}{|\mathcal{C}_j|} \sum_{i \in \mathcal{C}_j} x_i, \frac{1}{|\mathcal{C}_j|} \sum_{i \in \mathcal{C}_j} y_i \right), \tag{5}$$

   where $|\mathcal{C}_j|$ represents the number of locations in cluster $\mathcal{C}_j$. This step ensures that each centroid minimises the sum of squared Euclidean distances to all points in its cluster.

These two steps are repeated until the cluster assignments and centroid locations stabilise, indicating convergence. At convergence, the minimum total squared distance across all clusters can be expressed as:

$$\min \sum_{j=1}^{K} \sum_{i \in \mathcal{C}_j} \left\|(x_i, y_i) - c_j\right\|^2, \tag{6}$$

---

[1] The representative from each cluster can be one or more.

which represents the total sum of squared Euclidean distances from each location i to its centroid c_j across all clusters.

K-Medoid Clustering

The K-medoids clustering method is applied to a set of ambulance station locations I, where each location $i \in I$ has specific coordinates (x_i, y_i). The objective of the K-medoids clustering is to partition I into K clusters, where each cluster is represented by a central point called a “medoid," selected from actual ambulance station locations. Let M denote the set of these medoids, where $M \subseteq I$. The medoids minimise the total dissimilarity (sum of distances) between the ambulance stations and their respective medoid locations. See below for details about selecting the value of K.

The K-medoids clustering approach achieves minimal dissimilarity through two main iterative steps:

1. Assignment Step: Each ambulance location i is assigned to the nearest medoid m_j, effectively minimising the distance between each location and its assigned medoid:

$$\mathcal{C}_j = \left\{ i \in I \mid j = \arg\min_{m_j \in M} d\left(i, m_j\right) \right\}, \tag{7}$$

   where $\mathcal{C}_j$ denotes the set of locations in cluster j, m_j is the medoid of cluster $\mathcal{C}_j$, and $d(i, m_j)$ is the dissimilarity measure (distance) between location i and medoid m_j.
2. Update Step: Each medoid m_j is updated to be the location within $\mathcal{C}_j$ that minimises the sum of distances to all other locations in the cluster:

$$m_j = \arg\min_{p \in \mathcal{C}_j} \sum_{q \in \mathcal{C}_j} d(p, q), \tag{8}$$

   where p represents a candidate medoid, and q ranges over all locations in cluster $\mathcal{C}_j$. This step ensures that each medoid is the location that reduces intra-cluster dissimilarity.

These two steps are repeated until the cluster assignments and medoid locations stabilise, indicating convergence. At convergence, the minimum total dissimilarity across all clusters can be expressed as:

$$\min \sum_{j=1}^{K} \sum_{i \in \mathcal{C}_j} d\left(i, m_j\right), \tag{9}$$

which represents the total sum of distances from each location i to its medoid m_j across all clusters.

Agglomerative Hierarchical Clustering

Consider the set of ambulance locations $I = \{i_1, i_2, \ldots, i_{|I|}\}$, where each location $i \in I$ has coordinates (x_i, y_i), which represent the geographical coordinates (latitude and longitude) of ambulance station i.

Initially, each ambulance station location is treated as its own cluster, resulting in $k = |I|$ clusters (one for each station). We then compute the Euclidean distance $d(i, j)$ between every pair of ambulance locations,

$$d(i, j) = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}, \tag{10}$$

to compute the geographical proximity between them.

The next step is to iteratively merge the closest clusters (ambulance stations) using Ward’s linkage criterion [30], which minimises the increase in the sum of squared deviations within clusters (*i.e.,* variance). The centroid of each cluster C_j is defined as:

$$\mu_j = \frac{1}{|C_j|} \sum_{x \in C_j} x, \tag{11}$$

where $|C_j|$ is the size of cluster C_j, and μ_j is its centroid.

The increase in the sum of squared deviations within clusters when merging two clusters C_i and C_j is given by:

$$\Delta E\left(C_i, C_j\right) = \frac{|C_i|\,|C_j|}{|C_i| + |C_j|} \cdot d\left(\mu_i, \mu_j\right)^2, \tag{12}$$

where $d(\mu_i, \mu_j)$ is the Euclidean distance between the two centroids μ_i and μ_j. Upon merging clusters C_i and C_j, the new centroid μ' of the merged cluster $C' = C_i \cup C_j$ is computed as:

$$\mu' = \frac{|C_i|\,\mu_i + |C_j|\,\mu_j}{|C_i| + |C_j|}. \tag{13}$$

Step 2: Selection of Representatives Ambulance Stations

Once clusters have been formed, the second step involves selecting representative ambulance stations from each cluster. The chosen locations are used as representatives for the entire clusters in the optimisation (step 3). The objective is to preserve the structural integrity of the spatial distribution while significantly reducing the number of candidate locations.

There are two general scenarios when selecting representatives, depending on which clustering method is applied in step 1: (1) clustering methods that produce actual data points as centres, and (2) clustering methods that produce synthetic centres that do not correspond to real locations and therefore need to be approximated by actual data points.

Case 1: Clustering with Actual Data Points. Some clustering algorithms, such as *K-medoids* and *Agglomerative Hierarchical Clustering (AHC)*, produce cluster centres that are real data points (*i.e.,* actual ambulance station locations). These methods are well-suited for our use case, where MSUs can only be deployed at real stations.

K-medoids Clustering: This method identifies medoids (central data points within each cluster) that minimise the total dissimilarity to other points in the cluster. The algorithm directly returns medoids as representative locations.

Agglomerative Hierarchical Clustering (AHC): This bottom-up approach merges the most similar pairs of data points based on a linkage criterion (*i.e.,* Ward's method) until the desired number of clusters is formed. Since AHC does not select representatives by default, we adopt a deterministic strategy of choosing the first element from each cluster, which often tends to be centrally located due to the nature of hierarchical merging.

Case 2: Clustering with Synthetic Data Points. Methods such as *K-means clustering* calculate centroids as the arithmetic mean of data points in a cluster. These centroids typically do not correspond to real-world ambulance stations and thus cannot be directly used for MSU deployment. In such cases, a post-processing needs to be applied. For each centroid, we identify and select the nearest actual ambulance station within the same cluster. This allows us to map synthetic centres to real candidate sites.

By applying these strategies, our proposed method preserves the key spatial properties of the cluster and reduces the search space for the MSU allocation problem, reducing the number of candidate stations from $|I|$ (the total number of stations) to k (the number of clusters) as we now focus on the representative ambulance locations $\{m_1, m_2, \ldots, m_k\}$ as candidate locations for MSU placement.

Step 3: Solving the MSU Problem Using the Chosen Representatives

In the final step, we solve the MSU allocation problem using only the selected k representatives from the clusters. In other words, the reduced search space, composed of the k representative stations, is used to solve the MSU allocation problem. Note that the k is the hyperparameter of our method as it directly influences the size of the reduced search space. The smaller value of k (i.e., fewer clusters) leads to a smaller search space, which helps improve computational efficiency. However, too few clusters may risk omitting strategically important locations. Therefore, the wiser way is to identify a balanced number of clusters (the value of k) that effectively reduces the search space by exploiting the spatial distribution of ambulance stations while still preserving the diversity and quality of candidate locations necessary to find optimal MSU placements.

Post-processing After Choosing k Clusters. Once the value of k is chosen, the MSU allocation problem is solved using the reduced set of k representative locations. By limiting the candidate locations to these representatives, the search space is reduced from $\binom{|I|}{M}$ possible combinations of MSU locations to $\binom{k}{M}$, where M is the number of MSUs to deploy. Here, we can use any algorithm to solve the MSU allocation problem. For example, in the current study, we used an exhaustive search to solve the problem. In the exhaustive search, one can pass each combination of ambulance locations to the objective function. In other words, the reduced search space (containing all ambulance locations) is passed to the objective function (i.e., Eq. 2) as outlined in Section "The Mobile Stroke Unit Allocation Problem". From this reduced search space, we select the solution that yields the minimum weighted time to treatment as the acceptable solution for placing the MSUs.

Computational Study

Scenario Description

To evaluate the effectiveness of the proposed QCRSS framework in solving the Mobile Stroke Unit (MSU) allocation problem, we applied it to the Southern Healthcare Region (SHR) of Sweden. The aim was to identify optimal locations for deploying M MSUs, ranging from two to five, with a focus on the efficiency perspective, represented using weighted time to treatment (WATT).

The SHR presents a challenge for pre-hospital care, as it encompasses both densely populated urban centers and sparsely populated rural areas. The region spans four counties, includes 49 municipalities, and covers approximately 24, 000 square kilometres. It has a population of around 1.9 million, 13 acute hospitals and 39 ambulance stations. Figure 2 provides an overview of Sweden's Southern Healthcare Region (SHR), where ambulance site locations are marked with green triangles and acute hospitals are indicated by purple circles.

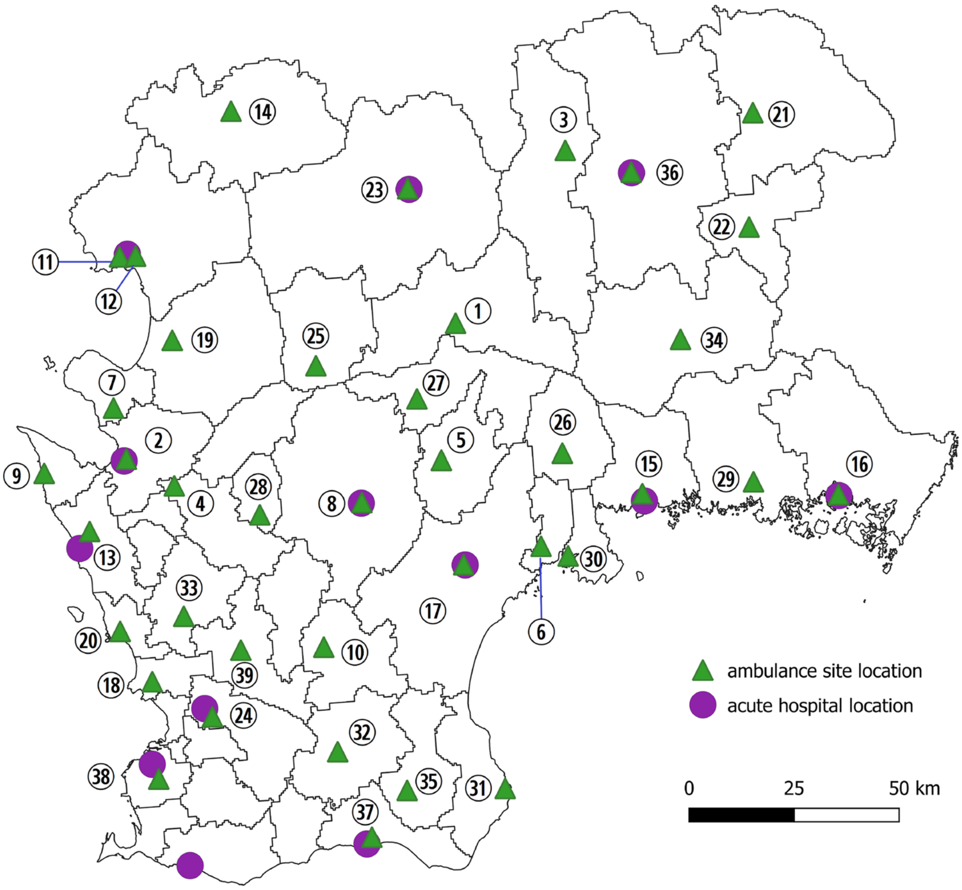


Fig. 2 Overview of Sweden's Southern Healthcare Region, with ambulance sites represented by green triangles and acute hospital locations by purple circles. The circled numbers indicate the corresponding ambulance site IDs

Evaluating Clustering Strategies for Reducing Search Space in the QCRSS Framework

In the preprocessing step of QCRSS (i.e., step 1), we explored three clustering techniques: K-means, K-medoids, and Agglomerative Hierarchical Clustering (AHC). Each method offers a distinct approach to partitioning the data and selecting representative cluster centres, affecting the efficiency and quality of the optimisation process.

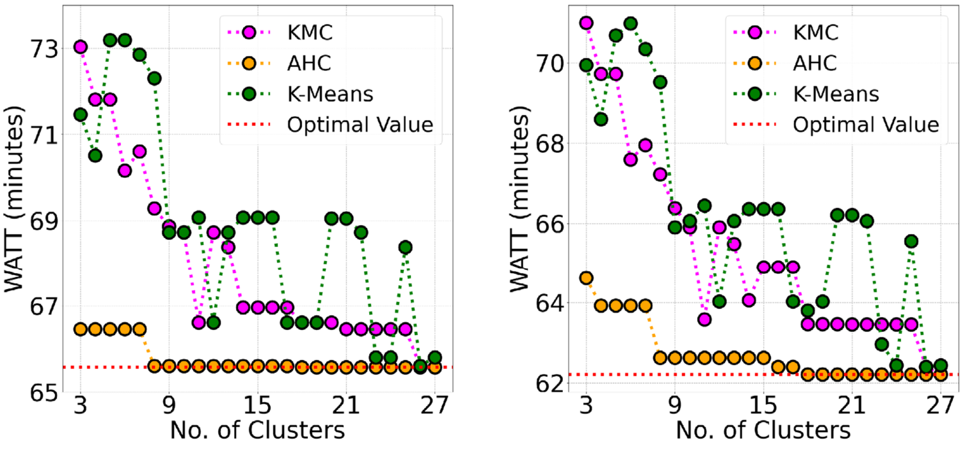
To assess the impact of these clustering methods, we conducted a comparative analysis based on the *Weighted Average Time to Treatment (WATT)*. The WATT is calculated using the objective function defined in Section "The Mobile Stroke Unit Allocation Problem" (Eq. 2). We varied the number of clusters from 3 to 27 and evaluated each method across four MSU allocation scenarios: deploying two, three, four, and five MSUs. The goal was to observe how the number and quality of clusters influence the ability to find high-quality solutions to the problem. The quantitative results are provided in Fig. 3 and the qualitative results are presented in Fig. 5.

Quantitative Results

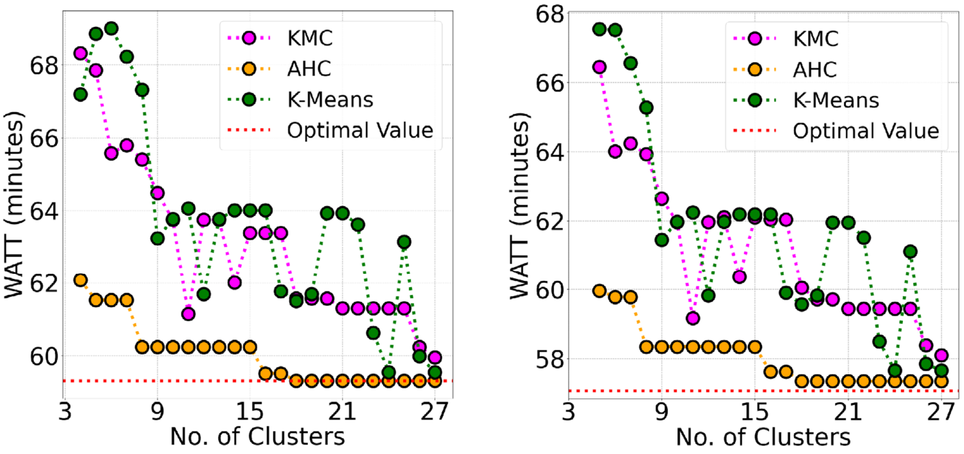
The results in Fig. 3 clearly show that AHC consistently outperforms the other considered clustering methods across all MSU deployment settings, that is, two, three, four, and five MSUs. For example, as shown in Fig. 3a

Note that across all scenarios, K-medoids clustering (KMC) and K-means clustering perform suboptimally for smaller numbers of clusters. As evident in the plots, the WATT values of these two clustering methods diverge considerably from the optimal values, especially in lower-cluster configurations. This indicates that KMC and K-means suffer from underfitting, as they fail to effectively capture the spatial structure of the data. Conversely, AHC achieves high-quality solutions with fewer clusters, reflecting its ability to better preserve the spatial integrity of ambulance station distributions.

Another notable observation is that the KMC and K-means results show irregular fluctuations. For example, for two MSUs, the performance improves at eight clusters but then worsens at nine. A similar dip and rise are observed as the number of clusters increases. This inconsistency highlights the instability of KMC and K-means in identifying robust cluster structures. In contrast, AHC remains



(a) Two MSUs (b) Three MSUs



(c) Four MSUs (d) Five MSUs

Fig. 3 Comparison of WATT across different numbers of clusters for each MSU setting. AHC consistently shows the best performance and fastest convergence. The results for K-medoids and AHC are reproduced from our previous study [18], while the results for K-means are newly obtained in the current, extended study

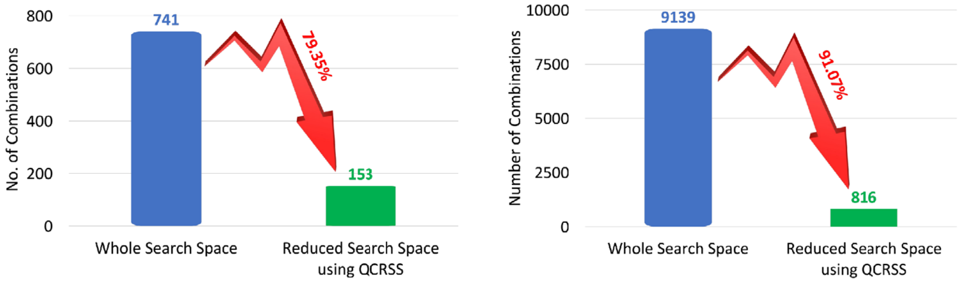
stable and reliable across different numbers of clusters, consistently yielding optimal or close to optimal results.

A key observation is that the performance of KMC and K-means deteriorates as the number of MSUs increases. While these two clustering methods reach the optimal solution for two MSUs at 28 clusters, they fail to do so for three to five MSUs. This suggests that KMC and K-means struggle to form high-quality clusters under large-scale MSU deployment scenarios. On the other hand, AHC achieves the optimal solution across all numbers of MSU deployments, demonstrating its effectiveness in finding the optimal solution across all scales of deployments, whether small or large.

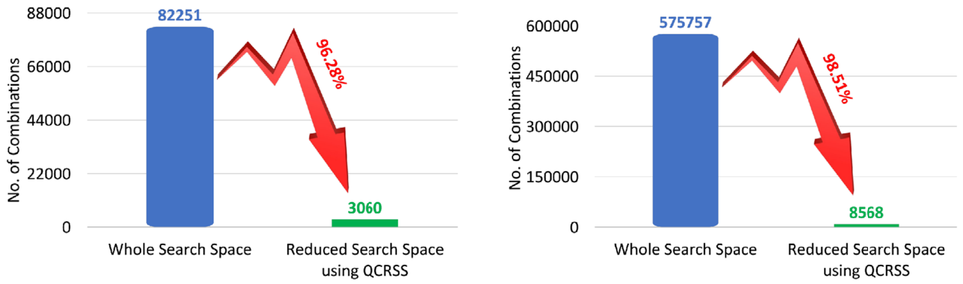
Search Space Reduction and Selecting Optimal Cluster Count in the QCRSS Framework. Selecting the optimal number of clusters k is a crucial step in the QCRSS framework, as it directly affects both the size of the search space and the quality of the final solution. A small k leads to a smaller search space, enabling faster convergence, but may risk overlooking key candidate locations. In contrast, a large k increases spatial granularity but comes with higher computational overhead.

To identify a suitable balance, we empirically evaluated values of k ranging from 3 to 27 across all MSU deployment scenarios. As illustrated in Fig. 3, clustering with $k = 18$ consistently delivered best results. Specifically, Agglomerative Hierarchical Clustering (AHC) was able to reach the optimal solution for two, three, and four MSUs with just 18 clusters. Even in the five MSU setting, where the complexity is significantly higher, AHC at $k = 18$ provided results very close to the optimal solution.

This empirical study suggests that 18 clusters represent a robust and effective configuration that provides a favourable trade-off between computational efficiency and solution quality for the considered scenario. Accordingly, we used $k = 18$ as the reference point for evaluating search space reduction. Figure 4 demonstrates the dramatic impact of this choice. When reducing the original candidate set of 39 ambulance stations to 18 representatives, the number of



(a) Two MSUs (b) Three MSUs



(c) Four MSUs (d) Five MSUs

Fig. 4 Comparison of the total number of MSU location combinations in the full search space versus the reduced search space using the QCRSS framework with $k = 18$ clusters. These results are directly quoted from our previous study [18]

combinations drops significantly for all MSU settings. For example:

- For two MSUs, the number of combinations reduced from $\binom{39}{2} = 741$ to $\binom{18}{2} = 153$, a 79.35% reduction (around 5x smaller).
- For three MSUs, the reduction was from $9,139$ to 816 combinations (91.07%, over 11x smaller).
- For four MSUs, the number of solutions dropped from $82,251$ to $3,060$ combinations (96.28%, over 26x smaller).
- For five MSUs, it decreased significantly from $575,757$ to $8,568$ combinations (98.51%, more than 67x smaller).

This substantial reduction makes the problem far more tractable, especially for larger MSU configurations, without sacrificing solution quality. Once k is chosen, the MSU allocation problem is solved considering the reduced set of k representatives, and the best configuration is selected based on the minimum WATT value as defined in Eq. 2.

In summary, using 18 high-quality cluster representatives through AHC enables the QCRSS framework to maintain strong solution performance while significantly reducing the computational complexity of the MSU allocation problem.

Qualitative Results

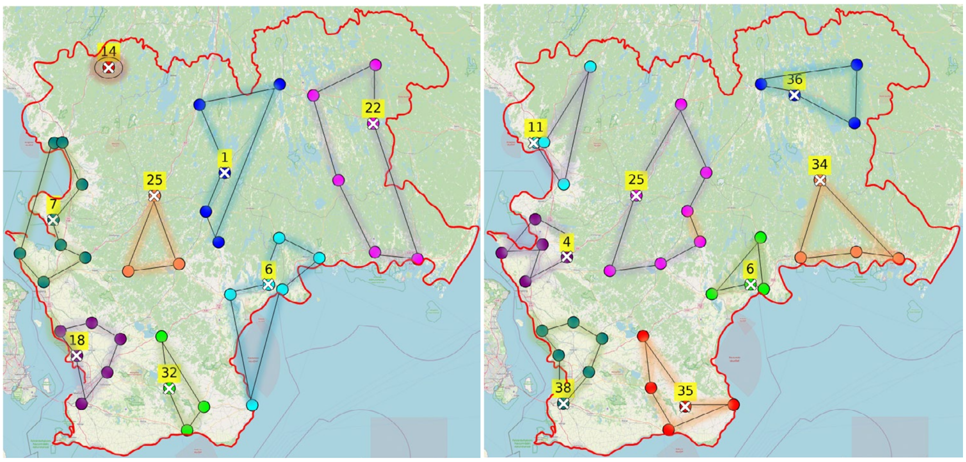
Following the quantitative analysis presented in Fig. 3, we further examine the qualitative aspects of the clustering outcomes in the QCRSS framework to understand how representative ambulance stations differ across the three clustering methods: Agglomerative Hierarchical Clustering (AHC), K-Medoids Clustering (KMC), and K-Means Clustering. The purpose of this visual analysis is to examine how effectively each method captures the spatial structure of the region and to evaluate the coherence and representativeness of the resulting clusters. Figure 5 visualises one example of forming eight clusters using each method for the Southern Healthcare Region (SHR). Each cluster is shown in a different colour, and the corresponding representative station, selected according to the QCRSS framework, is marked with a yellow label.

As shown in Fig. 5a

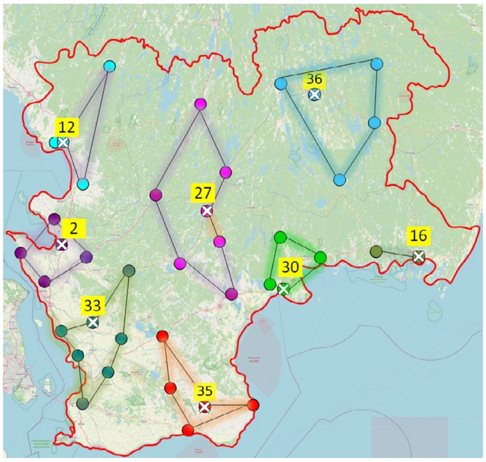
The K-means results (Fig. 5c) reflect similar issues to those for KMC. Although the clusters are generally tighter than those for KMC, the method suffers from the limitation that cluster centroids are not guaranteed to correspond to real ambulance locations. As a result, the nearest real station must be selected post hoc, which can skew representativeness. Additionally, K-means is sensitive to outliers and initial centroid positions, which can lead to potential variability across runs and inconsistency in results.

In contrast, AHC demonstrates superior spatial coherence, as seen in Fig. 5b

Overall, this qualitative comparison supports the quantitative findings: within the preprocessing clustering step of the QCRSS framework, AHC consistently outperforms other methods by producing spatially coherent clusters and selecting representative stations that closely reflect the true



(a) KMC Clustering (b) AHC Clustering



(c) K-means Clustering

Fig. 5 An example of eight clusters with their selected representative ambulance stations (highlighted with yellow labels) using the QCRSS framework for the SHR. The results for **a** K-medoids clustering and **b** AHC clustering are reproduced from our previous study [18]

geographic distribution of ambulance locations. Its deterministic merging strategy and strong spatial fidelity make it a reliable and effective approach for reducing the search space in the MSU allocation problem.

Conclusions

This paper introduces the *Quality Clustering for Reducing the Search Space (QCRSS)* framework as a solution-finding method for the Mobile Stroke Unit (MSU) allocation problem. QCRSS serves as a preprocessing step for optimisation algorithms, leveraging the spatial distribution of ambulance locations to reduce the size of the search space. By working within this reduced yet strategically constructed search space, QCRSS significantly improves computational efficiency while maintaining the possibility of identifying high-quality solutions. The framework consists of three main steps: 1) preprocessing using clustering, 2) selecting representatives for each cluster, and 3) solving the problem using only these representatives. In a computational study focusing on the Southern Healthcare Region of Sweden, we demonstrate that QCRSS can effectively filter out less promising ambulance locations without significantly compromising solution quality. The proposed method achieved the optimal solution while substantially accelerating convergence in two, three, and four MSU settings. Even for the larger and complex five MSUs setting, it achieved a highly satisfactory solution, illustrating both reliability and computational efficiency. We believe the QCRSS opens promising research directions for healthcare logistics optimisation and broader applications in prehospital care planning, including strategic MSU placement and resource allocation. It can serve as a foundation for future research on scalable, cluster-based optimisation in healthcare logistics.

Future Work

In this study, we selected a single representative from each cluster during the second step of the proposed framework. As part of our future work, we aim to extend this approach by investigating the use of two or more representatives from each cluster. Furthermore, we intend to evaluate the performance of the proposed framework across larger geographic regions to gain more comprehensive insights into the performance of QCRSS.

Author Contributions Conception and design of the study: all authors. Data analysis: MA. A. Drafting the manuscript: MA. A. Revision of manuscript: all authors. Final approval of the version to be published: all authors.

Funding Open access funding provided by Malmö University. The research was partially funded by the Kamprad and partially by direct governmental research grants.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Research Involving Human and /or Animals Not applicable.

Informed Consent Not applicable.

References

1. Patil S, Rossi R, Jabrah D, Doyle K. Detection, diagnosis and treatment of acute ischemic stroke: current and future perspectives. Front Med Technol. 2022;4:748949.
2. World Stroke Organization: Facts and Figures About Stroke. Accessed 30 June 2025. https://www.world-stroke.org/world-stroke-day-campaign/about-stroke
3. The Swedish Stroke Register: Stroke Registrations. Accessed 15 June 2025. https://www.riksstroke.org/annual-reports-and-graphs/
4. Luengo-Fernandez R, Violato M, Candio P, Leal J. Economic burden of stroke across Europe: a population-based cost analysis. Eur Stroke J. 2020;5(1):17–25.
5. Majersik JJ, Woo D. The enormous financial impact of stroke disability. AAN Enterprises; 2020.
6. Ashraf S, Masood S, Shahbaz A, Saboor QA. Factors responsible for worse outcomes in Stemi patients with early vs delayed treatment presenting in a tertiary care center. Pak Heart J. 2023;56(Supplement–2):11–11.
7. Blacker D, Hankey G. Pre-hospital stroke management: an Australian perspective. Intern Med J. 2014;44(12a):1151–3.
8. Bowry R, Grotta JC. Bringing emergency neurology to ambulances: mobile stroke unit. In: Seminars in respiratory and critical care medicine, vol. 38. Thieme Medical Publishers; 2017, p. 713–7. https://doi.org/10.1055/s-0037-1607994
9. Navi, B.B., Audebert, H.J., Alexandrov, A.W., Cadilhac, D.A., Grotta, J.C., Group, P.P.S.T.O.W. Mobile stroke units: evidence, gaps, and next steps. Stroke. 2022;53(6):2103–13.
10. Harris J. A review of mobile stroke units. J Neurol. 2021;268:3180–4.
11. Southerland AM, Brandler ES. The cost-efficiency of mobile stroke units: where the rubber meets the road. AAN Enterprises; 2017.
12. Amouzad Mahdiraji S, Dahllöf O, Hofwimmer F, Holmgren J, Mihailescu R-C, Petersson J. Mobile stroke units for acute stroke care in the south of Sweden. Cogent Eng. 2021;8(1):1874084.
13. Nour M, Vassar SD, Brown AF, Bosson NE, Chidester C, Liebeskind DS, et al. Geospatial modeling to optimize mobile stroke unit system deployment in a large metropolitan region. Stroke. 2022;53(Suppl-1):22–22.
14. Mahdiraji SA, Holmgren J, Mihailescu R-C, Petersson J. An optimization model for the tradeoff between efficiency and equity for mobile stroke unit placement. In: Innovation in medicine and healthcare: Proceedings of 9th KES-InMed 2021, Springer; 2021, pp. 183–193.
15. Amouzad Mahdiraji S, Abid MA, Holmgren J, Mihailescu R-C, Lorig F, Petersson J. An optimization model for the placement of mobile stroke units. In: International Conference on advanced research in technologies, information, innovation and sustainability, Springer; 2023, pp. 297–310.
16. Abid MA, Mahdiraji SA, Lorig F, Holmgren J, Mihailescu R-C, Petersson J. A genetic algorithm for optimizing mobile stroke unit deployment. Proc Comput Sci. 2023;225:3536–45.
17. Abid MA, Holmgren J, Lorig F, Petersson J. An enhanced genetic algorithm with clustering for optimizing mobile stroke unit deployment. In: 2024 IEEE 24th International Conference on bioinformatics and bioengineering (BIBE), IEEE; 2024, pp. 1–8.
18. Abid MA, Holmgren J, Lorig F, Petersson J. Quality clustering for reducing the search space for mobile stroke unit allocation. In: 15th International Joint Conference on biomedical engineering systems and technologies, February 20–22, 2025, Porto, Portugal; 2025, pp. 105–114. INSTICC.
19. Chen Y, Lai Z. A multi-objective optimization approach for emergency medical service facilities location-allocation in rural areas. Risk Manag Healthc Policy. 2022; 15, p . 473–90. https://doi.org/10.2147/RMHP.S332215
20. Reuter-Oppermann M, Berg PL, Vile JL. Logistics for emergency medical service systems. Health Syst. 2017;6(3):187–208.
21. Wang W, Wu S, Wang S, Zhen L, Qu X. Emergency facility location problems in logistics: status and perspectives. Transport Res Part E Log Transport Rev. 2021;154:102465.
22. Liu J, Li Y, Li Y, Zibo C, Lian X, Zhang Y. Location optimization of emergency medical facilities for public health emergencies in megacities based on genetic algorithm. Eng Constr Archit Manag. 2023;30(8):3330–56.
23. Yousefi M, Yousefi M, Fogliatto FS. Simulation-based optimization methods applied in hospital emergency departments: a systematic review. SIMULATION. 2020;96(10):791–806.
24. Markhorst B, Jagtenberg C, Hov MR, Mei R, Larsen K. Optimizing mobile stroke unit deployment: a strategic case study in the greater Oslo area. Eur Stroke J. 2025; 10(4):1438–1444. https://doi.org/10.1177/23969873251329862

25. Reimer AP, Zafar A, Hustey FM, Kralovic D, Russman AN, Uchino K, et al. Cost-consequence analysis of mobile stroke units vs. standard prehospital care and transport. Front Neurol. 2020;10:1422.
26. Navi, B.B., Audebert, H.J., Alexandrov, A.W., Cadilhac, D.A., Grotta, J.C., Group, P.P.S.T.O.W. Mobile stroke units: evidence, gaps, and next steps. Stroke. 2022;53(6):2103–13.
27. Ehntholt MS, Parasram M, Mir SA, Lerario MP. Mobile stroke units: bringing treatment to the patient. Curr Treat Options Neurol. 2020;22(2):5.
28. Chen J, Lin X, Cai Y, Huang R, Yang S, Zhang G. A systematic review of mobile stroke unit among acute stroke patients: time metrics, adverse events, functional result and cost-effectiveness. Front Neurol. 2022;13:803162.
29. Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual. Accessed 12 June 2025 https://www.gurobi.com
30. Developers, S. Scipy.cluster.hierarchy.ward. Accessed 17 June 2025. https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.ward.html