

A fuzzy-based prediction approach for blood delivery using machine learning and genetic algorithm

Marouane El Midaoui, Mohammed Qbadou, Khalifa Mansouri

Signals, Distributed Systems and Artificial Intelligence ENSET Mohammedia, University Hassan II of Casablanca, Casablanca, Morocco

Article Info

Article history:

Received Feb 22, 2021

Revised Jul 20, 2021

Accepted Aug 5, 2021

Keywords:

Blood demand prediction

Blood supply chain

Fuzzy logic

Genetic algorithm

MDVRP

Transfer learning

VRP

ABSTRACT

Multiple diseases require a blood transfusion on daily basis. The process of a blood transfusion is successful when the type and amount of blood is available and when the blood is transported at the right time from the blood bank to the operating room. Blood distribution has a large portion of the cost in hospital logistics. The blood bank can serve various hospitals; however, amount of blood is limited due to donor shortage. The transportation must handle several requirements such as timely delivery, vibration avoidance, temperature maintenance, to keep the blood usable. In this paper, we discuss in first section the issues with blood delivery and constraint. The second section present routing and scheduling system based on artificial intelligence to deliver blood from the blood-banks to hospitals based on single blood bank and multiple blood banks with respect of the vehicle capacity used to deliver the blood and creating the shortest path. The third section consist on solution for predicting the blood needs for each hospital based on transfusion history using machine learning and fuzzy logic. The last section we compare the results of well-known solution with our solution in several cases such as shortage and sudden changes.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Marouane El Midaoui

Signals, Distributed Systems and Artificial Intelligence ENSET Mohammedia, University Hassan II of Casablanca

Casablanca, Morocco

Email: marouane.elmidaoui@gmail.com

1. INTRODUCTION

These days, various diseases require a blood transfusion. Many patients, whatever the urgency of their situations, need a transfusion of red blood cells and/or platelets. The process of a blood transfusion is successful when the type and amount of blood needed is available and when the blood is transported at the right time. By having a blood unit from donors through blood banks and also those units should cover the need for quantity and type of blood. In the last decade, several countries attempted to create a social sensibility and requested for blood donation from public, because blood cannot be produced synthetically, and there is a positive outcome from the citizens. However, and unfortunately in many cases, the patients succumb to their injuries or illnesses and lose life due to delay in delivery process, or the blood not usable due to quality issues while transporting. Blood shortage remains and the blood demand over a time horizon is uncertain and also perishability. During each surgical operation, blood units are required. Transporting blood in an emergency on time and under the right conditions can, therefore, save many lives. Transportation needs to be more delicate since blood products have a limited lifespan hence, transportation-represents a critical stage in the transfusion chain which must be properly controlled. Transport must guarantee the integrity of blood while allowing rapid delivery to ensure the safety of the patient.

The idea is to determine the approach to be implemented to guarantee the quality of the goods, in our case blood units, transported both in terms of thermal and physical integrity within a well-defined period and in compliance with the hygiene and safety conditions of these organic products. Therefore, a good and efficient delivery system is needed, especially between blood-banks to the operation room, the system needs to provide a pulling approach from the blood bank and serve a set of hospitals while respecting the vehicle capacity and delivery time constraints. This approach is done using the vehicle routing problem (VRP) or traveling salesman problem (TSP) in case of only one route. VRP can be defined as the solution of creating optimal delivery from one or more depots to a certain number of dispersed customers. We can project this approach to our needs as blood-bank and hospital instead of depot and customer. However, in reality, many blood-banks can be located near to a set of hospitals, for this matter, the adaptation of new approaches such as the multiple depot vehicle routing problem (MDVRP) is required. There are three phases for optimization using MDVRP, grouping, routing, and scheduling. In the last phase, genetic algorithms can be used to attain more than one solution by fitness value. This transportation system needs to predict the amount of demand for each hospital, based on the history of demand.

This paper is organized as follows: Section 1 is about motivation and research method, section 2 presents the fuzzification of historical data to simplify the learning stage, section 3 discusses the prediction of hospital needs by using machine learning based on fuzzified data. Section 4 is about the research method between VRP and MDVRP for blood delivery using genetic algorithms after the defuzzification of machine learning data and followed with the result in section 5.

2. LITERATURE REVIEW

Blood storage and delivery is a well-known challenge that persists in the healthcare sector for decades, Martinez and Fedda [1] discuss those challenges and the requirement to store blood with different components like, red blood cells (RBCs) can be stored up to 42 days, platelets to 5 days, and plasma to one year. Jian *et al.* [2] author presents a decision cloud-based system for the risk assessment of heart disease by leveraging the techniques of a fuzzy expert system. Viegas *et al.* [3] use fuzzy logic to solve patients readmitted to care units, also Morsi and Gawad [4] use fuzzy logic in health-related diagnosis for the diagnosis of heart and blood pressure measurements. Viegas *et al.* [3] proposed a model to predict readmissions of care units using feature selection and fuzzy logic approaches. Alkahtani and Jilani [5] use machine learning to predict the return on donor blood donation with data mining. A 2016 study at New York City Blood Center [6] for blood supply demand forecast, to determine the optimal method of prediction, based on a comparison using moving average (MA), exponential smoothing (ES), autoregressive integrated moving average (ARIMA) and vector autoregressive integrated moving average (VARMA) models where the results showed that the accuracy of ARIMA models and their simplicity compared to VARMA made them the best models to predict blood demand, when compared the ARIMA model with artificial neural networks (ANNs) [7], and shows that ANN overcomes ARIMA in demand forecasting on monthly blood demand.

For the routing part, VRP natural computing algorithms show a suitable result. Yu *et al.* [8] proposed an improved ant colony algorithm to solve VRP which gives better results in the comparison between other heuristic methods, yet another approach proposed by the Gong *et al.* [9] based on particle swarm optimization (PSO) for VRP with time windows, the simulation result shows the efficiency of Solomon's benchmark testing algorithm. Multi-depot vehicle routing problem is considered one of the well-known solutions in the research field currently for solving the delivery of goods in multi-depot instead of one depot as VRP, and the goal is to optimize cost and exchange between multiple sources and destinations. Zhang *et al.* [10] proposed a model MDGVRP (MD green VRP) based on an ant colony with a solution to route the optimal path while taking into account the capacity of the vehicle and also fuel capacity of such vehicle. Vehicle routing problems cannot be resolved with a high precision, caused by the height computing power. To get better precision, the genetic algorithm (GA) can be used in this case, because of the stochastic properties and efficiency of GA, MDVRP can be resolved by using those techniques. Osaba *et al.* [11] describe several applied metaheuristic approaches. Laporte *et al.* [12] and Lenstra dan Kan [13] tried to solve the problem with exact methods, however in [14] demonstrated that both techniques are not effective. Aras *et al.* [15] solved for Priced MDVRP, which the vehicle visits a customer only if it is profitable, where in they propose two methods: the heuristic method to solve the MDVRP, and mixed-integer linear programming (MILP), and concluded that heuristics give better results.

3. RESEARCH METHOD

There is uncertainty in the demand and distribution of blood in the healthcare industry, which drives it necessary to study and understand the history of demands for blood in order to be able to predict future

needs for a wide blood type. Blood unit bag distribution has a large share of the cost in hospital logistics. It causes high-cost transportation, but also risks to the quality and requirement of transportation, such as vibration avoidance, temperature maintenance, and timely delivery, especially in emergency situations. When the distribution is not optimal, blood quality is greatly influenced by transportation. Related to this matter, the distribution process needed to be optimized and predictable. The first step in our proposition is to predict the number of units that can be requested from each hospital. This prediction is based on historical data of demand of the hospitals. After the donor signs a consent form, the collection of blood from donors can be with two methods, manually or automated apheresis devices. This blood then will be processed into three components, red blood cells, Platelets, and Plasma, each one has specific conditions for storing and quality control based on the duration of storage, where the red blood cells (RBCs) can be stored up to 42 days, the platelets up to 5 days, and plasma up to one year [1]. Therefore, each hospital has a specific demand for these components. However, usually, the demand for blood units in the historical data is set as a numeric value, consequently, the decision about blood unit quota to be provided for each hospital based on current blood bank storage is unclear and hard to predict. Thus, a fuzzy system to perform a clustering can help to determine the demands based as a fuzzification of the numerical value inside the historical data demand for each blood component and for each hospital. The data after fuzzification will be introduced to the machine learning stage to perform a prediction based on historical data while respecting the priority of the newest data by being aware of its past experiences as a human. A defuzzification process is needed to numerically predict each hospital's demand based on current blood bank stock to set a fair distribution. Finally, perform routing and scheduling using GA to create the shortest path for blood unit's delivery, Figure 1 details this process.

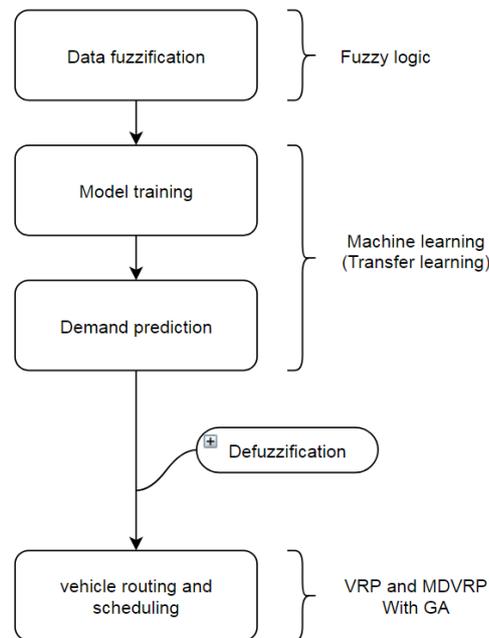


Figure 1. General schema of the proposition

3.1. Data fuzzification

Fuzzy logic is a generalization of standard logic, in which a concept can possess a degree of truth by using many-valued logic rather than binary or numerical logic, fuzzy logic can provide human experience and human decision-making behavior. In recent decades, fuzzy logic is interesting in industry and academic fields. It has been applied to many research areas, such as artificial intelligence and machine learning. This approach can deal with incomplete, imprecise, inconsistent, and uncertain data. It has been used to solve several nonlinear controls, such as modeling the number of patients readmitted to care unit [3]. The author proposes the use of demographic information, vital signs, and statistic from a laboratory, to accurately predict the number of patients to be readmitted to the care unit.

Fuzzy logic consists of four-stage, fuzzifier, a defuzzifier, an inference engine, and a rule base fuzzification is needed to convert the input data into an appropriate set of linguistic value which is required for the inference engine. At the rule base stage, a set of fuzzy rules are defined, that characterize the behavior

of the system. The inference engine stage is used to create inferences and provide conclusions based on the previous stage. Defuzzification is the stage of transforming back the output (fuzzy action) into a set of numerical or binary actions based on the output of the inference of the engine. Figure 2 presents the fuzzy logic system and Figure 3 present the pseudo code as implemented in our paper.

In our case, the blood bank data is a set of five values: The week (week date), hospital, the number of ordered red cells bag units, platelets bag units, and plasma bag units. During the fuzzification stage, numeric inputs are linked to linguistic membership Functions to get values that describe the respective input as presented in Figure 4. With this approach, the output data will be a linguistic data (very high, high, normal, low, very low) which describes the amount of blood unit order for each blood component (red cells, platelets, plasma) by a hospital in a specific week.

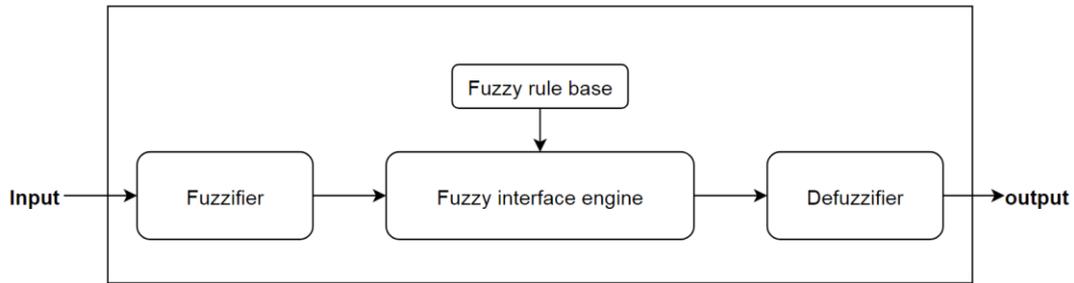


Figure 2. Fuzzy logic system diagram

```

Fuzzy logic algorithm
1. Collect data from the source to input variable.
2. Define linguistic taxonomy and terms
3. Construct the membership function
4. Create a rule base
5. Convert input data to fuzzy data
6. compare rules in the rule base
7. group the result of each rule
8. Convert output data to a numeric type
    
```

Figure 3. Fuzzy logic pseudo code

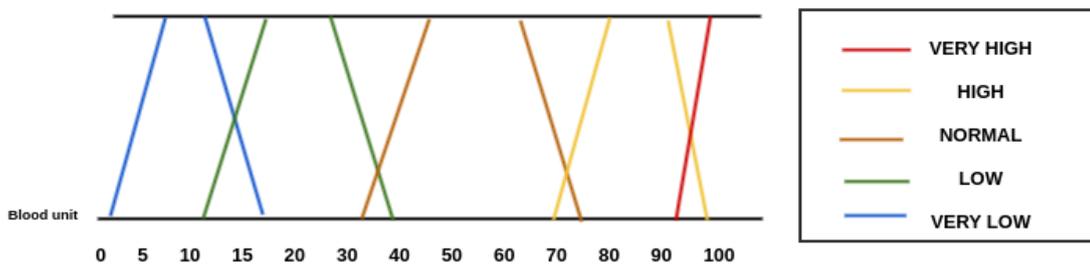


Figure 4. Membership functions for unit demand for each hospital (Very high, high, normal, low, very low).

3.2. Prediction and machine learning

Machine learning is a new field of computer science which is growing fast in the last few years, and making a big difference in pattern recognition and learning in artificial intelligence (neural network). Machine learning use algorithms that allow learning from data to make a prediction or decision in a future event, by using statistics to solve many classifications and clustering problems. Machine learning techniques can be applied to all research and industry fields. Usually, machine learning (ML), uses data to train the system and make a prediction on unknown data by using one of three categories, unsupervised, supervised,

and semi-supervised, however in reality and more specifically in our case (blood transportation), and due to constraint of durability, quality, time conditions, specification can occur. Thus, the data need to be updated for escaping the old data training problem which is not accurate. To resolve that, we can use a reinforcement learning technique called transfer learning. This technique reduces the old data in case of new data inserted. Transfer learning consists of reuse the model created for a specific task as a starting point for training the second task. This method is a popular approach in deep learning when a model that has been trained before being used as a starting point to avoid consuming huge amount of time and resources required to train a new model and to use the old experience of the old model. As an example, the Yosinski *et al.* [16] present their results on how the lower layers act as extractors of conventional computer vision, such as edge detection feature, while the final layers work towards task-specific functionality. Kocçer and Arslan [17] developed a method for routing traveling paths based on the virtue of traffic density by using transfer learning. And such models can rapidly change the routing based on past experiences. Sharma *et al.* [18] use machine learning for products price prediction.

To deal with future events, a prediction system is needed, in our case, the need is to predict future hospital blood demand. Thus, machine learning is used to optimize performance using old or sample data. After receiving data from the fuzzification stage, we need to predict the hospital demand, based on historical data, and this data will be feeded as an input into a neural network, and start training the model. In this case, we can predict a fuzzification of the amount of blood units needed for each hospital in every blood component, however, we can also predict the demand based on a specific week-in case of epidemic disease which spreads in a specific time and needs a blood supply. Input data is the fuzzified week, hospital, the number of ordered red cells bag units, platelets bag units, and plasma bag units. Table 1 presents sample input data: (*NB: hospitals are presented as identifiers; those identifiers are linked geographical data.*)

Table 1. Sample input data for training models to make predictions

Week(W)	Hospital(H)	Red Cell(RC)	Platelets(PL)	Plasma(P)
1	H1	LOW	HIGH	VERY_LOW
1	H2	HIGH	HIGH	NORMAL
2	H1	LOW	VERY_HIGH	NORMAL

In our example, we predict the number of units needed by a specific hospital in a specific week of the year. The model can also predict the need based on geographical data or/and week. During this prediction, the trained model predicts the output for a given input (hospital and week) based on its learning. To evaluate the performance of the model, the test prediction concludes by predicting blood components for a hospital in a pre-stored week and comparing the values. As a result, 81% of prediction was correct, 9% was incorrect in one component, and 6% incorrect in two components and 4% was completely incorrect.

3.3. Evaluation metric

To gauge and evaluate the presented model among other models, a regression analyses using evaluation metric is implemented to measure the quality of the fit of the model, where (Root mean squared error) RMSE is widely used for this matter. It's the same metric used by [7] and [6], RMSE is a very common evaluation metric, measuring the quadratic mean of the differences between predicted values and the real ones, It can range between 0 and infinity, where it is based on penalizing large errors, which means values closer to 0 are better than higher values. Thus, this is a sensitive metric to outliers. Besides RMSE, mean absolute error (MAE) is also used for the same matter. MAE helps to measure error magnitude that consists in the results of average of the difference between the predicted values and the actual ones.

$$RMSE = \frac{\sqrt{\sum_{i=1}^N (x_i - x'_i)^2}}{N} \quad MAE = \frac{\sum_{i=1}^N |x_i - x'_i|}{N}$$

Where, N number of non-missing data points x_i actual observations time series and x'_i is the estimated series.

4. ROUTING AND SCHEDULING

Optimal distribution problem can be solved using the vehicle routing problem (VRP) model, however in some cases, several blood banks can serve a group of hospitals, in this case, the multi depot vehicle routing problem can be useful as a model. As we need, the main purpose of VRP is to find the shortest routing and shortest time. In the classic VRP case, only one blood bank can serve, we consider the

blood-bank as the depot, in this case, the depot will be the start and endpoint. The blood-bank has m vehicles for n hospitals. In case all hospitals had been visited, or the capacity of vehicles met, the vehicles need to return to the blood bank. Although, both vehicle routing problems are NP-hard.

4.1. VRP

The vehicle routing problem (VRP) is an extension of the traveling salesman problem, also known as the traveling salesman problem (TSP). The objective is to determine the routes of a vehicles fleet to customers or to perform tour of visits (medical, and commercial visits) or interventions (maintenance, repairs and checks). We will represent clients as hospital or a healthcare facility, and the deposit as blood-bank. VRP: as presented in Figure 5, a single depot (in our context a single blood bank) is serving several hospitals, thus all hospitals are served from one blood bank.

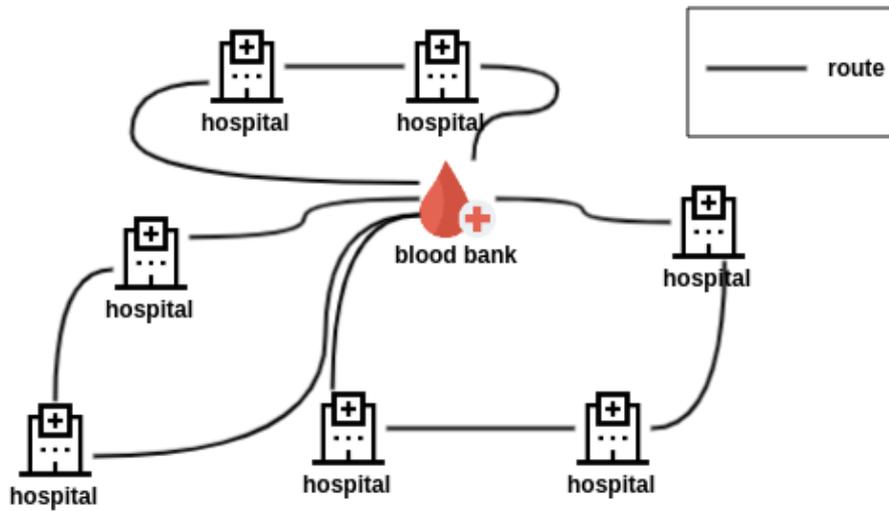


Figure 5. Blood delivery routing illustration based on the VRP approach

Based on transportation network, the notations used for the formulation of the VRP are based on a simple graph $G = (V, E)$ with:

- $V = \{V_0, V_1, \dots, V_n, V_{n+1}\}$ a set of vertices with $V_0 = V_{n+1}$ representing the deposit and $V' = V/\{V_0, V_1, \dots, V_n, V_{n+1}\}$ the set of clients.
- $E = \{(V_i, V_j) / V_i, V_j \in V', i \neq j\}$ is the set of edges forming the routes. Each tour starts at v_0 and ends at v_{n+1} .
- A matrix of costs or distances c_{ij} between customers V_i and V_j , with $(C_{0,n+1} = 0)$.
- d represents a vector of customer requests.
- R_i is the tour of vehicle i represented by a permutation of customers visited by the vehicle.
- The fleet of vehicles \mathfrak{S} is made up of all identical vehicles with capacity C and fixed cost F . A vehicle is assigned to each round.

This model contains noted decision variables x_{ij}^ℓ (defined $\square ij \square V, \square \ell \square \mathfrak{S}$) and take the value 1 if the vehicle ℓ goes from customer i to customer j , 0 otherwise. The VRP can therefore be formulated as (1)-(7):

$$\text{Min } Z = F \sum_{j \in V'} \sum_{\ell \in \mathfrak{S}} x_{0j}^\ell + \sum_{i \in V'} \sum_{j \in V'} c_{ij} \sum_{\ell \in \mathfrak{S}} x_{ij}^\ell \tag{1}$$

$$\sum_{i \in V'} \sum_{\ell \in \mathfrak{S}} x_{ij}^\ell = 1 \quad \forall j \in V' \tag{2}$$

$$\sum_{j \in V'} \sum_{\ell \in \mathfrak{S}} x_{ij}^\ell = 1 \quad \forall i \in V' \tag{3}$$

$$\sum_{j \in V'} x_{0j}^{\ell} = 1 \quad \forall \ell \in \mathfrak{S} \quad (4)$$

$$\sum_{i \in V} x_{ik}^{\ell} - \sum_{j \in V} x_{kj}^{\ell} = 0 \quad \forall k \in V', \forall \ell \in \mathfrak{S} \quad (5)$$

$$\sum_{i \in V'} x_{i,n+1}^{\ell} = 1 \quad \forall \ell \in \mathfrak{S} \quad (6)$$

$$\sum_{i \in V'} d_i \sum_{j \in V} x_{ij}^{\ell} \leq C \quad \forall \ell \in \mathfrak{S} \quad (7)$$

The objective function (1) indicates the total cost to be minimized and which includes the cost of vehicles and the cost of rounds. The constraints (2) and (3) indicate that each customer must be assigned to exactly one vehicle. The constraints (4), (5) and (6) model the flow constraints which guarantee that each vehicle leaves the depot, and then visits at least one customer to finally return to the depot. The constraint (7) indicates that the capacity of the vehicles must not be exceeded.

4.2. MDVRP

The multi-depot vehicle routing problem (MDVRP) is a well-known class of problems in distribution and logistics, it is a VRP with multiple depots from where vehicle with or without a fixed capacity can make a travel to serve the client or the destination. In our context, the hospital presents a client. If health-care or hospitals are placed near to blood-banks, then the distribution can be simplified as a normal set of VRPs as explained before. But if healthcare facilities and blood banks are not organized due to geographic shape or political requirements, or also urgent delivery, the issues will be a multi-depot vehicle routing. The vehicles in MDVRP serve a hospital in a separated region or group, each vehicle or set of vehicles is assigned to a single depot, which is usually both the origin and destination of the vehicle route. The Figure 6 presents the multiple blood bank deliveries for a set of hospitals. It shows that hospitals can be served from several blood banks and can alter the route length. The two blood banks in this figure is the minimum, it can be more based on the needs and infrastructure.

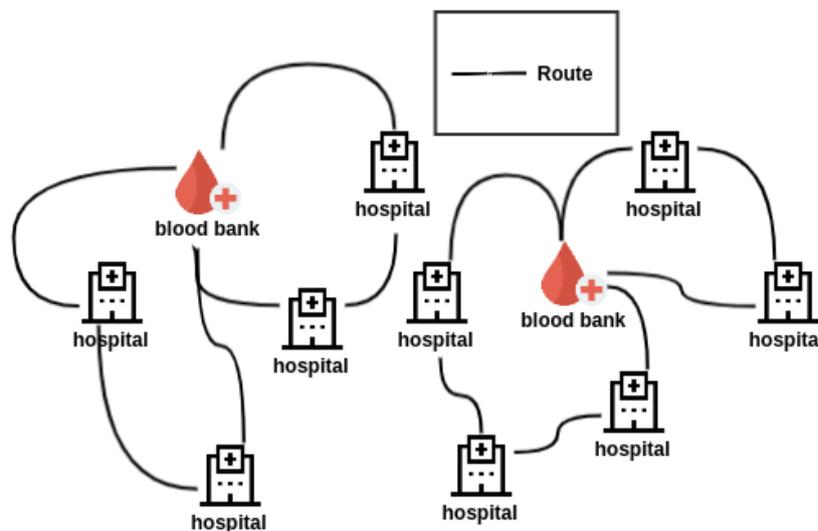


Figure 6. Blood delivery routing illustration in case of multiple blood banks based on the MDVRP approach

We can present our MDVRP mathematically as follows: $G = (V, E)$ is a graph where V is the set of vertices partitioned into two collections. $V_c = \{V_1, V_2, \dots, V_n\}$ Represent the list of the hospitals. $V_p = \{V_{n+1}, V_{n+2}, \dots, V_{n+m}\}$ represent the collection of blood banks (BB), E set of edges connecting two points. A

Cost matrix $D = (D_{ij})$ is the length of arc (v_i, v_j) (corresponding to distance). Expecting D symmetry, and the distance between sets of three points satisfies the triangle inequality. Each BB $V_{n+m} \in V_d$ has the same model. k_m (The number of vehicles at the BB) are identical in term of vehicle (model, capacity Q , shape).

Each vehicle starts from a blood bank (BB), serving a set of hospitals, at the end of the route or when Q is not enough to serve the next Hospital, the vehicle must return to the starting BB to refill. Every hospital is visited only once by one vehicle, we consider in this example that a route can be done only by one vehicle. The total demand for each route does not exceed the vehicle capacity Q . The vehicle cannot exceed the vehicle capacity Q , and the number provided by each BB is limited. Blood bank cannot serve each other's (BB cannot serve another BB). The main objective for MDVRP, in this case, is to link hospitals into the proper BB with minimization of distance travelled by vehicle across the network. n hospitals are grouped to form m clusters. Clusters consists of n_1, n_2, \dots, n_m number of hospitals. The number of vehicles based at a BB is k_m . k_j is the group of hospital linked to a vehicle. MDVRP formulation for finding x which is the minimizes can be described as following:

Decision variables X and Y :

$$x_{ijkm} = \{\text{hospital } j \text{ else } 0\}$$

$$y_{ikm} = \{\text{BB } m \text{ else } 0\}$$

$$\text{MIN} = \sum_{p=1}^m \sum_{q=1}^{k_p} \sum_{i=1}^n \sum_{j=1}^n d_{ij} x_{ijqp} \quad (8)$$

$$\sum_{i=1}^n q_i y_{iqp} \leq Q \quad (9)$$

$$0 \leq n_{jq} \leq n_j \quad (10)$$

$$\sum_{q=1}^{k_j} n_{jq} = n_j \forall j = 1 \dots m \quad (11)$$

$$\sum_{j=1}^m n_j = n \quad (12)$$

$$\sum_{p=1}^m \sum_{q=1}^{k_p} y_{iqp} = 1 \quad (13)$$

$$x_{ijqp} = 1 \vee 0 \quad (14)$$

$$y_{iqp} = 1 \vee 0 \quad (15)$$

The objective function according to (8) minimizes the total delivery distance and cost of each vehicle within a blood bank. Equation (9) limits capacity Q of the vehicle, (10) indicates the hospitals served by each vehicle must not overtake the number of the hospital a blood-bank can deliver to. The (11) shows the sum of the hospitals served by the whole route should be the sum of the hospitals served by the blood bank. The (12) shows every hospital served by a blood bank. The (13) Guarantee that every hospital must be visited only once by a vehicle (14), (15) is the value bound of the decision variables.

4.3. Genetic algorithm

The genetic algorithm (GA) is a stochastic optimization technique based on a parallel search mechanism which makes it more efficient than other classical optimization techniques such as branch and bound, Tabu search method and simulated annealing. The GA implementation is used to avoid local optimum trap using genetic operations like crossover and mutation. A population of possible randomly-generated solutions represent as chromosomes that can evolve through iterations. The chromosome is ranked by fitness in each generation. The score of the fitness of chromosome determine if it will be selected for the next generation or not by using the genetic operations.

GA was used by many authors to simulate an evolutionary system based. It is a class of heuristics methods used in several field including the VRP and MDVRP, Surekha and Sumathi [19] present a comparative study of different genetic algorithm variant applied to VRP and MDVRP, Ho *et al.* [20] propose two solutions for the MDVRP. The first one is a simple implementation of GA for solving MDVRP. The second one is a with Clarke and Wright saving based method with nearest neighbour approach, and he provide a comparative study between both solution and the author conclude that the second one is more suitable.

I initial population: the first chromosome is generated based on Euclidian distance. Blood bank and hospitals will be grouped according to the nearest group with consideration of the grouping, routing, and scheduling. For example, an MDVRP instance with 5 health care facilities and 0 is the blood-bank (BB). If the path representation for this instance is (0 2 4 1 0 3 5 0), then two routes are required to serve all the facilities. The first route depart from the BB at 0 and travels to facilities (2 4 1). Once the vehicle has finished the first route, the BB receive the vehicle for next load to serve for second route (if there is any hospital left). Second route start from the BB 0, serves hospitals and facilities 3, 5, and returns to the blood-bank.

Evaluation: chromosome can be evaluated by calculating the fit score or fitness value, the ideal situation in MDVRP is to reduce the time spending on delivery and reduce cost in n blood bank and health-care facilities. The process starts in each blood bank at the same time that explains why some vehicles are able to the delivery simulation before other vehicles. In our context we consider fitness score calculated based on a total of cost route CR_{mk} of each vehicle (k) in every blood-bank (m). The cost formula is presented as the following:

$$cost = \sum_{m=1}^m \sum_{k=1}^k CR_{mk}$$

The cost for one vehicle of the i BB is calculated as the following:

$$CR_{i1} = D(d_i, c_1) + \sum D(c_j, d_i) + D(c_n, d_i) \text{ for all } j \in n_{jq}$$

Selection: the best chromosome (based on fitness score) is selected for next generation. Ho *et al.* [20] uses a roulette-wheel based selection to generate the new chromosome for next generation, while others use tournament selection [19]. The chromosome selection strategy is based in almost all situation on fitness score evaluation, the fittest solution is to be selected for the next iteration. It can be also a selection for several chromosomes to perform the genetic operation.

Genetic operation: both mutation and crossover affect the selected chromosomes with genetic operations rate, several methods and approach can be implemented for picking the right case, but overall, both operation impact the results in term of time and iteration. Crossover: The combination of informations from two or several genomes to generate new generation, it can be done with picking the random (classical) case or by implementing a more suitable solution to make the process faster or more efficient. Berman and Hanshar [21] present an approach for crossover, the approach is based on best cost route crossover (BCRC) projected to vehicle routing problem with time windows (VRPTW). The classical order crossover was implemented by [20] for an MDVRP application. Mutation: exchanging or swapping two distinct customers or depots randomly to generate more heterogenic population. The mutation operation assure the diversity in the each generation to avoid being trapped in a local optimum. There are two types of mutation method, heuristic mutation and inversion mutation [22]. Flipping a chromosome case to build a new offspring. The creation of the population and more specifically the initial population will be presented as set of cases, hospitals presented as numbers while 0_a , 0_b are BB. Figure 7 presents a chromosome as a set of healthcare facility with number from 1 to 9 presents hospitals, and 0 presents blood banks.

Grouping: We implement weighted-k-means to be our clustering method for not only group by distance, but also by hospital bed capacity (HBC), but the data in the real case was not available especially HBC. So we used Euclidean distance instead [23]. He *et al.* [24] present a new approach using the Tabu search algorithm. However, some geographical distribution need to use haversine [25] especially if the country is large and the hospital along with blood bank are way to separate from each other. Figure 8 presents grouping sample, *BB* presents blood banks and numbers presents hospitals.

1	2	3	4	5	6	7	8	9	10	0_a	0_b
---	---	---	---	---	---	---	---	---	----	-------	-------

Figure 7. Chromosome presentation of hospitals and blood banks

BB_A 1, 2, 3, 4, 5	BB_B 6, 7, 8, 9, 10.
----------------------	------------------------

Figure 8. Grouping samples between blood banks and hospitals

Figure 9 present the whole process as a flowchart from the initial parameters to the routing and scheduling implementation using genetic algorithm operation (Mutation, Crossover, Selection), and looping after preserving the population that has a better fitness score.

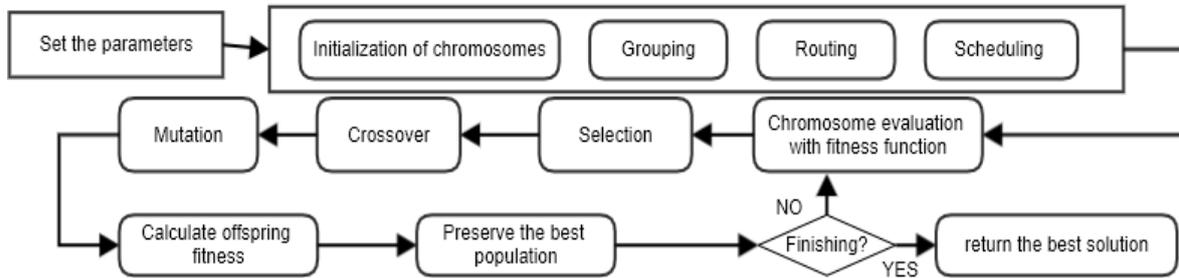


Figure 9. Flowchart of GA implementation

5. RESULTS AND DISCUSSION

As we mentioned in previous chapters, the fuzzification of the data set was translated into 5 values (Very high, high, normal, low, very low), those data and beside the hospital information (an identifier, latitude-longitude) and week (is of date format), was used to train our machine learning model with up to 800.000 records, and the prediction was well acceptable with 81% of a correct prediction for all three components (Red cells, plasma, platelets). Our model is more likely to be compared with both models ARIMA [6] and ANN [7] as mentioned previously, the present approach is based on transfer learning (TL) and fuzzy logic instead of ANN, to include sudden changes in blood demand. Also, the model is using fuzzy logic to determine the amount of blood bag/unit the delivery in case of shortage. Additionally, we specified in the previous chapter that the prediction is on a weekly basis due to the nature of some blood types which can be usable only for less than 5 days, thus the prediction on a monthly basis will be not suitable for all blood types. Using our generated data, the simulation of the three models in two different cases shows the results presented in Table 2.

Table 2. MAE and RMSE records of ARMA, ANN and TL-FUZZY in shortage and sudden change cases

	RMSE			MAE		
	NORMAL	SHORTAGE	SUDDEN CHANGE	NORMAL	SHORTAGE	SUDDEN CHANGE
ARIMA	103.32	204.56	280.02	134.22	234.32	310.08
ANN	87.11	102.45	192.56	55.32	144.23	207.53
TL-FUZZY	101.34	80.32	102.86	56.43	81.67	99.45

The Figure 10 is a graphical presentation (chart) of the Table 2 with the upper section using RMSE and the bellow section using mean absolute error (MAE). Using MAE and RMSE metrics, Figure 11 shows performance of the three models. This figure displays a confrontation between the models in usually normal situations, under shortage of blood resources and when sudden changes occur. In fact, as detailed in Table 2, Root mean squared error metric enlighten the right performance of the present TL-FUZZY model compared with others, especially in shortage and sudden change situations (204.56 for ARIMA, 102.45 for ANN but only 80.32 for TL-FUZZY in shortage cases). Also, MAE metric accentuate the performance of the present model in both unusual cases (310.08 for ARIMA, 207.53 for ANN and 99.45For TL-FUZZY). Finally, in usual situations, TL-FUZZY nearly rivals ARIMA model but still far from ANN feat. However under shortage circumstances or when a sudden change happens, where TL-FUZZY performed better when compared to both ANN and ARIMA models. For the routing and scheduling, our simulation shows a regression in fitness score from generation to the next one from 7011 fitness score to 5254, and stabilized at generation 75 and start to rise after, which means the generation 75 is more suitable. In terms of time as presented in Figure 12, the total consumption time from generation to the next is considerably changing. The first generation starts with 120 min to complete all visits, the 5th generation raised to 130 min, but just after start decrease to 71 min at generation 45 and start raising again to 85 min in generation 65, at last generation (70) the time was 70 min which is the “better result” for the simulation, which means an escape from the optimal minimum.

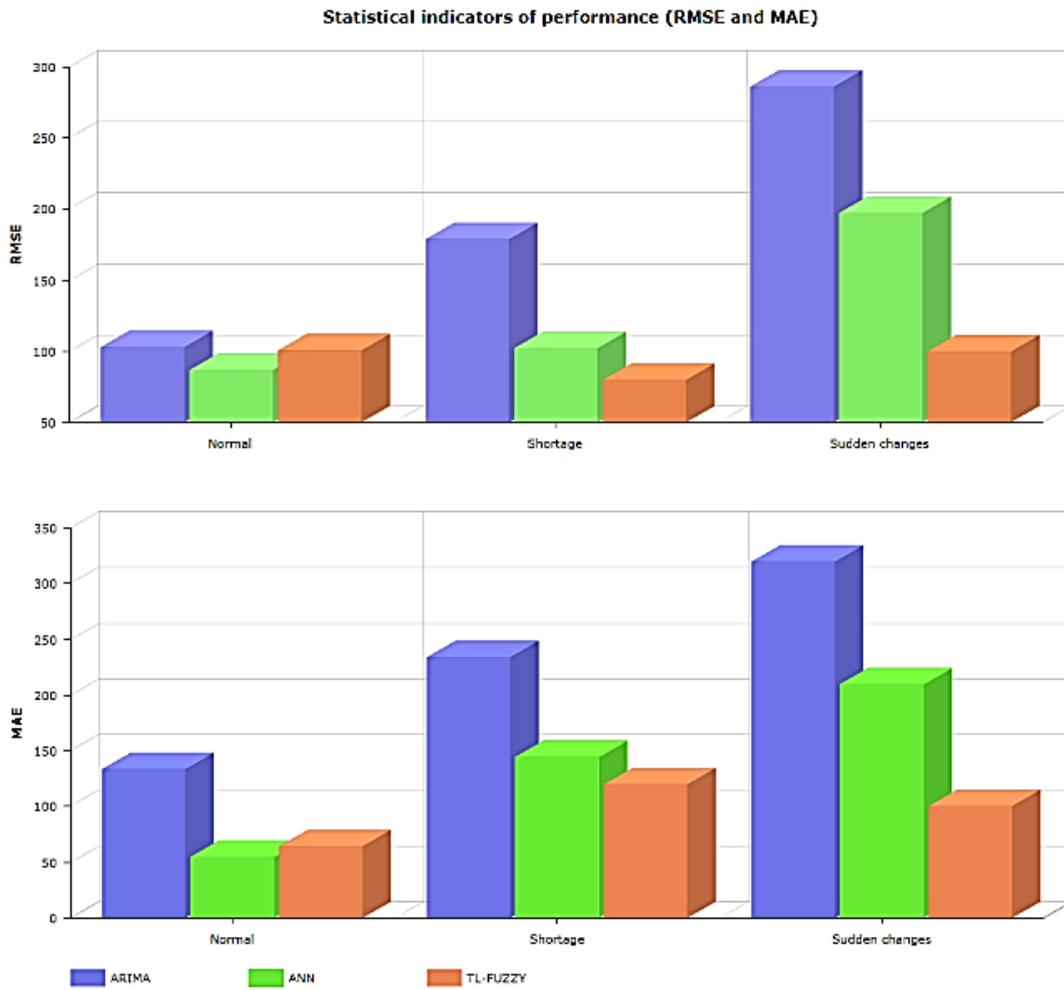


Figure 10. TL-FUZZY compared to ANN and ARIMA models with MAE and RMSE metrics

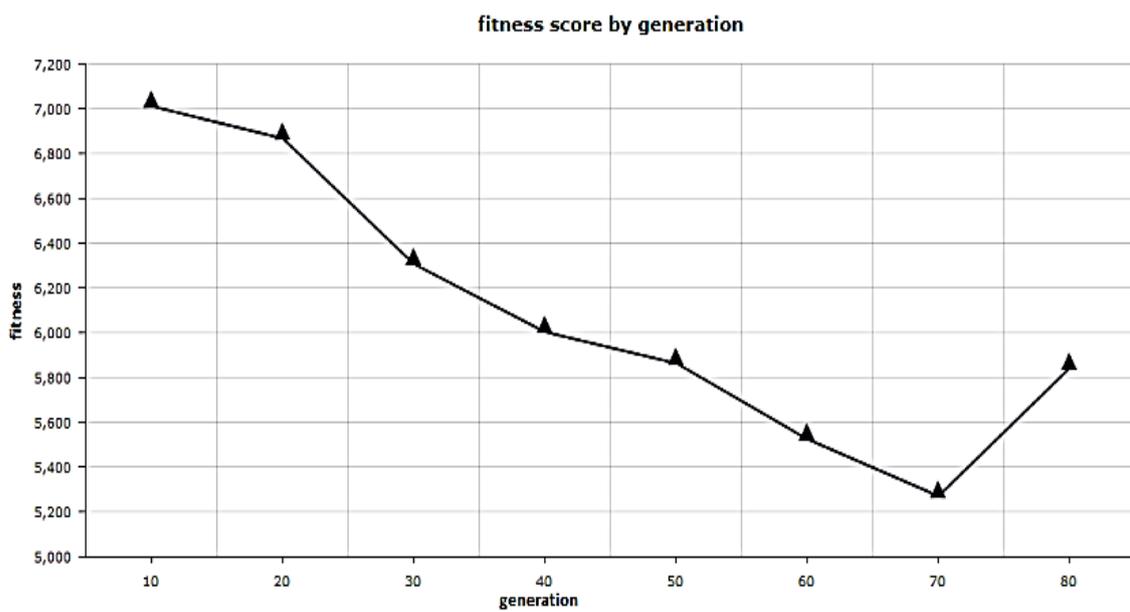


Figure 11. Fitness score (F) for each generation (G)

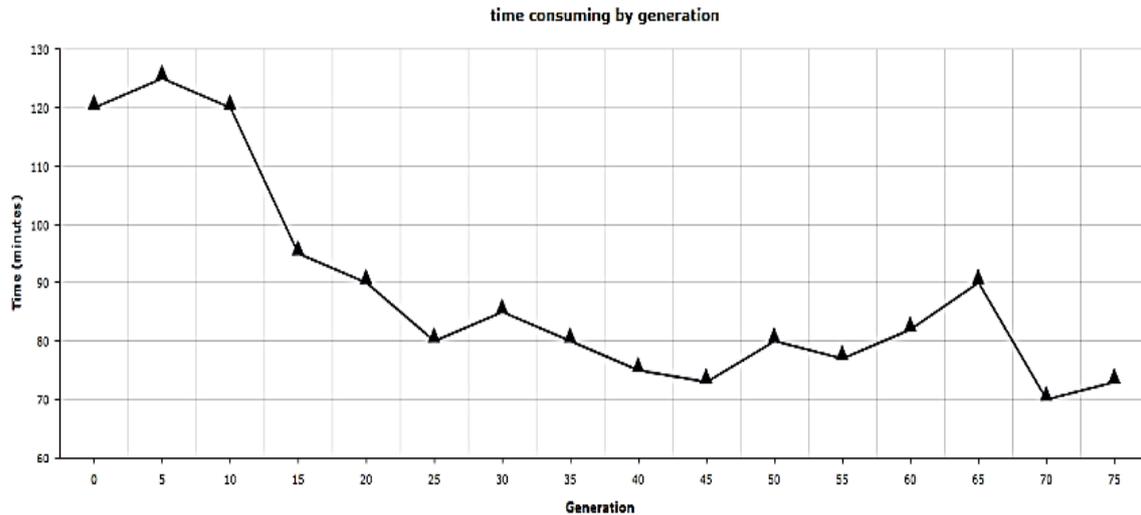


Figure 12. Time-consuming in minutes from generation to the next

6. CONCLUSION

Health-care and especially hospitals suffer from blood delivery delay and distribution due to the lack of infrastructure and vision of future demand. Our contribution consists of the adoption of fuzzification data to get a clear idea about demand with linguistic values, train a model by using transfer learning to predict a future need in blood components by using historical data. And using an algorithm based on genetic algorithms to set a routine and schedule plan for blood unit bag distribution. The combination of fuzzy logic and machine learning besides genetic algorithms is rarely used in the resolution of similar matters. Also, even if this solution appears to be specific to the health-care and blood delivery sector, this work can concern further fields like agricultural and food-processing. It should be noticed that the model can be improved by considering applying parallelization for GA and taking into consideration reverse logistics to recover the unused/expired blood.

REFERENCES

- [1] F. Martinez and F. Fedda, "Blood components," *Oncologic Critical Care*. Springer International Publishing, pp. 1171-1176, 2019, doi: 10.1007/978-3-319-74588-6_201.
- [2] S. J. Jian, R. K. Chiu, and S. A. Wang, "A cloud decision support system for the risk assessment of coronary heart disease," *2012 International Conference on Machine Learning and Cybernetics*, 2012, pp. 1435-1440, doi: 10.1109/ICMLC.2012.6359576.R.
- [3] R. Viegas, C. M. Salgado, S. Curto, J. P. Carvalho, S. M. Vieira, and S. N. Finkelstein, "Daily prediction of ICU readmissions using feature engineering and ensemble fuzzy modeling," *Expert Systems Applications*, vol. 79, pp. 244-253, 2017, doi: 10.1016/j.eswa.2017.02.036.
- [4] I. Morsi and Y. Z. A. El Gawad, "Fuzzy logic in heart rate and blood pressure measuring system," *2013 IEEE Sensors Applications Symposium, SAS 2013 - Proceedings*, 2013, pp. 113-117, doi: 10.1109/SAS.2013.6493568.
- [5] A. S. Alkahtani and M. Jilani, "Predicting return donor and analyzing blood donation time series using data mining techniques," *International Journal Advance Computer Science and Applications (IJACSA)*, vol. 10, no. 8, pp. 113-118, 2019, doi: 10.14569/ijacsa.2019.0100816.
- [6] S. M. Fortsch and E. A. Khapalova, "Reducing uncertainty in demand for blood," *Operations Research for Health Care*, vol. 9, pp. 16-28, 2016, doi: 10.1016/j.orhc.2016.02.002.
- [7] R. Khaldi, A. El Afia, R. Chiheb, and R. Faizi, "Artificial neural network-based approach for blood demand forecasting: Fez transfusion blood center case study," *Proceedings of the 2nd International Conference on Big Data, Cloud and Applications*, 2017, pp. 1-6, doi: 10.1145/3090354.3090415.
- [8] B. Yu, Z. Z. Yang, and B. Yao, "An improved ant colony optimization for vehicle routing problem," *European Journal Operational Research*, vol. 196, no. 1, pp. 171-176, 2009, doi: 10.1016/j.ejor.2008.02.028.
- [9] Y. J. Gong, J. Zhang, O. Liu, R. Z. Huang, H. S. H. Chung, and Y. H. Shi, "Optimizing the vehicle routing problem with time windows: A discrete particle swarm optimization approach," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 42, no. 2, pp. 254-267, 2012, doi: 10.1109/TSMCC.2011.2148712.
- [10] S. Zhang, W. Zhang, Y. Gajpal, and S. S. Appadoo, "Ant colony algorithm for routing alternate fuel vehicles in multi-depot vehicle routing problem," *Decision Science in Action*, pp. 251-260, 2019.
- [11] E. Osaba, X-S. Yang, and J. D. Ser, "Is the vehicle routing problem dead? An overview through bioinspired perspective and a prospect of opportunities," *Nature-Inspired Computation in Navigation and Routing Problems*, pp. 57-84, 2020, doi: 10.1007/978-981-15-1842-3_3.
- [12] G. Laporte, Y. Nobert, and S. Taillefer, "Solving a family of multi-depot vehicle routing and location-routing problems," *Transportation Science*, vol. 22, no. 3, pp. 161-172, 1988, doi: 10.1287/trsc.22.3.161.

- [13] J. K. Lenstra and A. H. G. R. Kan, "Some simple applications of the travelling salesman problem," *Operational Research Quarterly*, vol. 26, no. 4, pp. 717-733, 1975, doi: 10.2307/3008306.
- [14] B. Crevier, J.-F. Cordeau, and G. Laporte, "The multi-depot vehicle routing problem with inter-depot routes," *European Journal Operational Research*, vol. 176, no. 2, pp. 756-773, 2007, doi: 10.1016/j.ejor.2005.08.015.
- [15] N. Aras, D. Aksent, and M. T. Tekin, "Selective multi-depot vehicle routing problem with pricing," *Transportation Research Part C Emerging Technologies*, vol. 19, no. 5, pp. 866-884, 2011, doi: 10.1016/j.trc.2010.08.003.
- [16] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems*, pp. 3320-3328, 2014.
- [17] B. Kocçer and A. Arslan, "Transfer learning in vehicle routing problem for rapid adaptation," *International Journal Innovative Computing Information and Control*, vol. 8, no. 10 A, pp. 6799-6809, 2012.
- [18] A. Sharma, D. Bhuriya, and U. Singh, "Survey of stock market prediction using machine learning approach," in *Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017*, 2017, pp. 506-509, doi: 10.1109/ICECA.2017.8212715.
- [19] P. Surekha and S. Sumathi, "Solution to multi-depot vehicle routing problem using genetic algorithms," *World Applied Programming*, vol. 1, no. 3, pp. 118-131, 2011.
- [20] W. Ho, G. T. S. Ho, P. Ji, and H. C. W. Lau, "A hybrid genetic algorithm for the multi-depot vehicle routing problem," *Engineering Applications Artificial Intelligence*, vol. 21, no. 4, pp. 548-557, 2008, doi: 10.1016/j.engappai.2007.06.001.
- [21] B. O. Berman and F. T. Hanshar, "Using genetic algorithms for multi-depot vehicle routing," *Studies in Computational Intelligence*, vol. 161, pp. 77-99, 2009, doi: 10.1007/978-3-540-85152-3_4.
- [22] R. Cheng and M. Gen, "Genetic algorithms and engineering design (engineering design and automation) (1st ed.)," *Wiley-Interscience*, 1996, doi: 10.1002/9780470172254.
- [23] J. Carlsson, D. Ge, A. Subramaniam, A. Wu, and Y. Ye, "Solving min-max multi-depot vehicle routing problem," *Lectures on Global Optimization*, pp. 31-46, 2009, doi: 10.1090/fic/055/03.
- [24] Y. He, W. Miao, R. Xie and Y. Shi, "A tabu search algorithm with variable cluster grouping for multi-depot vehicle routing problem," *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2014, pp. 12-17, doi: 10.1109/CSCWD.2014.6846809.
- [25] K. Kolomvatsos, K. Panagidi, and S. Hadjiefthymiades, "A load balancing module for post emergency management," *Expert Systems with Applications*, vol. 42, no. 1, pp. 657-667, 2015, doi: 10.1016/j.eswa.2014.07.055.

BIOGRAPHIES OF AUTHORS



Marouane El Midaoui    is from Casablanca, Morocco, and he attended Hassan II University of Casablanca where he graduated in 2017 with a master degree in distributed information system at ENSET Mohammedia. He is also concerned about computer science, IA, programming and software design. For his Ph.D thesis, he is working on Intelligent Transport Systems and Logistics at SSDIA laboratory in the same university, His researches are focused on artificial intelligence, logistics, IoT and blockchain. He can be contacted at email: marouane.elmidaoui@gmail.com.



Mohamed Qbadou    has born in 1970 at Kalaa Sraghna, Morocco. Since 1997, he is a teacher of computer science and researcher at the University Hassan II Casablanca, ENSET Institute. His research is focused on semantic Web, Distributed computing system, Knowledge database system, educational information system and e-learning, Assistive robotics, Natural Language processing. Diploma ENSET Mohammedia in 1992, CEA in 1993 and Ph.D. in 1998 at the Mohammed V University of Rabat, Morocco. He can be contacted at email: qbmedn7@gmail.com.



Khalifa Mansouri    has born in 1968 at AZILAL, Morocco. He is now a teacher of computer science and researcher at the University Hassan II of Casablanca, ENSET Institute. His research is focused on Real-time modeling systems, the governance of information systems, e-Leraning systems and the calculation and optimization of structures. Diploma ENSET Mohammedia in 1991, CEA in 1992 and first Ph.D. in 1994 to Mohammed V University of Rabat, HDR in 2011 and second Ph.D to Hassan II University of Casablanca, Morocco. He can be contacted at email: khmansouri@hotmail.com.