Staff Scheduling in a hospital context: The Case of Luz Saúde Hospital

Sara Maria Pereira Oliveira
sara.m.pereira.oliveira@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2020

Abstract

This work proposes a multi-objective genetic algorithm that aims to optimize the schedule of the staff in a hospital service. Health professionals are one of the indispensable resources in the quality of services provided and represent more than half of hospitals' operating costs, which requires an endless search for methods to improve the efficiency of operations. This work is motivated by Hospital da Luz de Lisboa, which identified some needs in the planning of these resources, since its management is time consuming, mostly hand-made and by trial and error, which often causes several inconsistencies. This can be reflected in a dissatisfied staff or situations of staff shortages, overloading the remaining staff. The Intensive Care Unit (ICU) served as case study and intends to incorporate personal preferences in a balanced way, ensuring a fair schedule. The algorithm was also applied in the Imagery service. Both have points in common, mainly the emerging need to find a solution for staff scheduling. The algorithm is able to effectively find good approximations of the Pareto-front in a timely manner. Additionally, it incorporates the possibility of choosing a single solution based on weights attributed by the decision-makers. The ICU is satisfied with the results, mainly by the fact that the algorithm is able to generate a balanced schedule with the possibility of using historical records from previous months, contributing to a more effective and fair scheduling in the long term.
Keywords: Staff Scheduling, Genetic Algorithm, Fairness, Preferences, Multi-objective

1. Introduction

One of the biggest challenges faced by the world is the provision of high-quality healthcare with affordable expenses. This requires a never-ending search for methods to improve the efficiency of the operations. Staff-related costs accounts for more than half of the operating costs in hospitals [1]. Furthermore, the health workforce is an indispensable and decisive resource in the quality of the healthcare delivery [2]. Consequently, it is crucial to find efficient and effective ways to plan and schedule the healthcare resources. This is a remarkably complex, time-consuming and error-prone work that should simultaneously consider multiple criteria from cost savings, preferences and employee and patient satisfaction to dependencies between human and material resources. Therefore, the number of conditions is high and most of the time they are variable and uncertain, thus complex to meet simultaneously.

Personnel scheduling problems have been widely discussed in literature [1]. Although, over time, the way it is approached has changed, there has been a growing concern regarding employee satisfaction and preferences [3]. There is an increased interest among hospitals to satisfy these demands, not only due to the scarcity of available health staff on the job market but also the difficulty in replacing, which often leads to understaffing [1, 4]. Hospitals cannot stop and they are under high uncertainty and fluctuating conditions and, these issues could lead to an increase of lengths of stay and waiting times, which leads to a loss of quality. Additionally, to compensate this scenario it is necessary to increase the workload of the team, creating a further unsatisfied and unmotivated staff, and this will also have a negative impact in effectiveness and can lead to an increase of absences [1]. The opposite scenario is also possible - overstaffing - in times of less demand. This can cause extra costs and staff dissatisfaction since they cannot use their time so usefully.

Currently, the most part of schedules are hand-made, which often raise unfairness and violations of labor regulation; for the most part it is planned by a team element - who wastes valuable time. Furthermore, this implies extra and needless costs.

The aim of this work is to develop an algorithm to support and optimize the staff scheduling, considering the available resources and satisfying the rules and targets that must be met, as well as the stakeholder's perspective. A decision support algorithm, based on Operations Research techniques, to automate and improve the hospital schedules is proposed. This follows the main steps of the Non-dominated Sorting Genetic Algorithm II (NSGA-II), a metaheuristic method based on the theory of evolution that advocates that the most able individuals perpetuate themselves in the species. This will be made analyzing the *Hospital da Luz de Lisboa* (HLL) case study. The preferences, restrictions and objectives must be identified in order to automate and improve scheduling activities.

The contributions of this study are four-fold: (i) reducing the time needed to obtain a feasible schedule; (ii) increasing the efficiency of the schedules produced; (iii) ensuring that a greater number of conditions are met; and (iv) incorporating preferences and fairness.

The remainder of this paper is organized as follows. Section 2 describes the case study description. Section 3 provides an overview of healthcare staff scheduling literature. Section 4 introduces the Genetic Algorithm (GA) developed. Section 5 presents the computational experiments and the discussion of the results. Finally, in Section 6 , conclusions, limitations and future work are stated.

2. Background

Most of hospitals continue to use manual scheduling, a time-consuming and error-prone task, which in most cases is conducted by a highly qualified staff, as happens in HLL, which additionally has been faced with shortages of healthcare personnel. Two cases studied are addressed in this work: Intensive Unit Care (ICU) and Imagery Service (IS).

HLL is a private hospital that belongs to Luz Saúde (LS) Group, which incorporates several Hospitals and Clinics.

2.1. Intensive Care Unit

The ICU carries patients with life-threatening illnesses and injuries, who require constant supervision. Furthermore, this implies highly trained staff and a higher staff-to-patient ratio comparing with other wards, and consequently this is an expensive resource. The ICU in HLL is composed by 15 physicians, where one is in charge with administrative tasks. Thus, he/she works less hours. There is the possibility of receiving external physicians in the less populated shifts, i.e. nights and weekend shifts. Currently, the scheduling task has been done in an *Excel* file, which is performed by trial-error method. The goal is to build an entire month schedule. There are five types of shifts: (i) Morning shift (from 8:30am to 4:30pm - 9 hours); (ii) Prolongation shift (from 8:30am to 9pm - 12.5 hours), coincident with the morning in part of the day; (iii) Night shift (from 3:30pm to 9:30am - 18 hours); (iv/v) 24-hour and weekend shifts (from 8:30am to 9:30am - 25 hours).

After following the process it was possible to identify some inconsistencies, such as the same physician assigned to the morning or afternoon shift on Monday being assigned to the Sunday before, violating a hard constrain. There were also writing mistakes which affected the counts; in addition, sometimes the same physician is assigned in a morning and prolongation shift, which could not be possible since they are simultaneous shifts.

2.2. Imagery Service

A hospital IS is equipped with image technology able to give support and find the correct diagnostic based in images of organs. The machines are operated - and exams are provided - by technicians. As aforementioned, the LS Group incorporates several Hospitals and Clinics; for this reason,IS technicians can be scheduled in HLL and in Amadora and in Oeiras Clinics.

In total, there are 46 technicians and they are divided into two main groups: Central (composed by 34 technicians) and Emergency Group (subdivided into two groups of 6 technicians: Group E1 and Group E2). The Central Group is in charge of the elective exams that come mainly from appointments, while the emergency group is in charge of unpredictable situations that appear in the Emergency Room. Groups differ in shift types: on the one hand the E1 group works until 1 a.m. at the most, while E2 are in charge of night shifts. Here there is a greater offer of shifts, and they are divided in morning, afternoon, all-day and night. There are morning and afternoon shifts of 6 and 8 hours, while the all-day shifts have a duration of 10 hours.

The technicians scheduling task takes almost two days to complete and it is performed by two technicians, who lose valuable time, which represents an extra-cost for the hospital and a decrease of effectiveness. The technicians' preferences are reported by e-mail and by phone, often during the process, which implies excessive and spread information, potentiating inaccuracies. The service schedule is made in a *Excel* file and the three units are in the same sheet, thus it is necessary to be always scrolling it when building the clinics schedule. The file counts the number of hours and the shifts assigned, in order to compare them and perceive if the needs are achieved.

3. Literature Review

The literature on staff scheduling in healthcare has increased in the last years, due to the impact this can have from the operating costs in hospitals and the quality of services provided to the workers life [1, 5]. Over time, the approach has changed, in particular, the importance of staff satisfaction, preferences and flexible work hours has grown [3].

Due to the uncertainties, demand may face unbalanced distributions over time. This is usually an input parameter that can be deterministic or stochastic, depending on whether it is a known or an uncertain value, respectively. In the literature, the majority approaches demand in a deterministic way [1]. The flexible modelling of shifts can be used to overcome this obstacle, i.e. the start time and the duration of each shift are flexible as long as work regulations are satisfied [1, 6]. [7] relied on flexible shifts in order to better provide demand coverage and minimize the cost of hiring physicians. On the other hand, several studies deal with different level skills. A cross-trained workforce, despite having specific skills, is allowed to perform tasks that usually are performed by other workers. This increases flexibility and allows to cover unexpected demand peaks, avoiding the hiring of new expensive workers or even layoffs. [8] concluded that the flexibility that results from this type of workers is better than having perfect information regarding demand.

Regarding objectives, they are mainly divided into financial or non-financial. In order to achieve their financial goals, [9] used cost as their measure, since it depends on the day of the week – in off-days (e.g. weekends) costs are higher – and highly-skilled personnel demands higher spending. [7, 10] had as a main objective to minimize staffing costs related to hiring physicians with different experience levels, that are needed to provide the demanded coverage. Alternatively, [11, 12]'s goal was to minimize the overtime working hours that are paid out. Although financial goals are important, they do not appear regularly in the literature. The non-financial objectives mainly focus on individual aspects from both patients and personnel. [13] concentrated on aspects with respect to well-being and patients' quality of life, in particular, on minimizing the number of hand-offs. On the other hand, [14] chose to minimize the average total patient waiting times in an Emergency Department. Waiting times are strongly influenced by the number of staff present in the unit.

Concerning individual preferences, they can vary greatly from person to person – *e.g.* there may be employees willing to work on weekend or night shifts, in order to receive extra money, while others prefer evenings or nights off [15]. Therefore, these preferences may include the staff requests, such as preferred days-off, duties, daytimes and so on [1].

The fairness aspects are often addressed in scheduling problems and its definition has varied greatly. [16] interpreted fairness as the equal distribution of preference fulfilment among schedule personnel. Their model ensured that the preferred solution is one in which several nurses have a small number of violated preferences, rather than a single nurse suffering a considerable amount of violations. Often the greatest concern is to maximize overall quality of the schedule, however this can lead to unfair individual schedules. In order to overcome this problem [17] minimizes the difference between maximum and minimum individual penalties from soft constraint violations. On the other hand, [5] tackled a re-scheduling problem where the main objective was to minimize the overall penalty costs, which derived from fairness violations. [15] proposed a model using a satisfaction-based preference weight, where each physician has a satisfaction indicator, which measures a physician's satisfaction according to the preferences fulfilled in the roster. They also used previous planning horizons to track fairness measures. Both [5] and [15] concluded that, using previous periods, the fairness level of schedules improves, and unfair fulfilment of preferences accumulated over time is avoided. Nevertheless, fairness does not consist only in maximizing or minimizing a particular goal, but rather about trying to find a balance for all parties, considering not only the balanced distribution of the workload, but also taking into account the individual preferences as well as particularities for each day (e.g. working on a Sunday is not the same as working on a Monday). Fairness aspects are significantly improved by automated scheduling, as demonstrated by [18].

Many studies do not pursue a single goal, but rather try to consider multi-objective functions. Although the main goal of [12] was to minimize the operating costs, which incurred due to overtime, the authors also include employee preferences, fairness aspects and consistent workloads. On the other hand, [19] did not only want to maximize total revenue, but also to minimize all scheduling preferences discrepancies, in order to maximize the overall workload fairness. [20] had a bi-objective function: maximize the number of assigned shifts while minimizing the number of assigned inconvenient shifts, in order to provide fairness. [21] created a multi-objective model for a nurse scheduling problem by highlighting human factors, such as skills, preferences and compatibility between nurses; besides the goal of minimizing the total cost related to staff, it also intended to minimize incompatibilities and maximize overall satisfaction. Lastly, [18] used a multi-criteria objective function which maximizes the number of assignments according to labor regulations and internal department rules, and minimizes costs for violating fairness goals.

The current work accounts for a multi-objective approach for healthcare staff scheduling that meets the needs of the hospital, considering different skills, satisfying preferences, and incorporating fairness aspects. In addition, it will be applied to different health professionals, namely physicians and technicians, however not simultaneously. This work uses real data from the hospital.

4. Methodology

A NSGA-II – a metaheuristic method –, based on [22], is adapted to each case study. This is a GA applied to multi-objective problems. It uses the usual genetic operators, namely the chromosome, a two-point crossover, mutation and tournament selection, as well as multi-objective genetic operators such as non-dominated sorting and crowding distance. This algorithm is based on three important features: (i) it considers *elitism* property, this is, it preserves the best solutions of a generation and takes them to the next one; (ii) through crowding distance is possible to preserve diversity; and (iii) it highlights non-dominated solutions.

The chromosome represents the set of variables that define a proposed solution. In this case it is a $[i \times j]$ matrix, where i denotes the total number of staff, and j the days that are intended to be scheduled. Each entry of the matrix indicates the shift assigned to a worker during a day (see Figure 1).

The two-point crossover is applied in the proposed algorithm, as represented in Figure 1a. Two points are chosen at random, between 1 and $j-1$. Afterwards, the first columns from the parent P_0 to the first cutting point will pass to the child C_0, the portion between the cutting points will be inherited from the parent P_1, and the remainder from the second point to the end (j) will come from the first parent.

The mutation operator inserts small changes in the chromosome. In this case, for each column j, two rows are chosen at random, i_1 and i_2. If the entries corresponding to these choices ($i_1 j$ and $i_2 j$) are different, they are exchanged. Otherwise, a new choice is made. This process is repeated for all columns of the chromosome (see Figure 1b).

Tournament selection is used, where sets of T individual are randomly chosen. The selection takes place as often as necessary until the desired number of individuals (r) is reached and the value of T can vary.

The Algorithm 1 provides a pseudo-code of the main steps of the algorithm. As defined in line 6, there are two stopping criteria: (i) maximum number of generations desired; and (ii) based on how long the algorithm should run, stopping after the defined time limit.

Algorithm 1 NSGA-II algorithm.

```
1: Initialize random Population
2: Evaluate Individual Fitness (f_m)
3: Non-dominated sorting
4: Calculate Crowding Distance
5: Tournament Selection to select P_0
6: while generation number < maximum generation or timer < defined time do
7:     Generate Children Population (Q_t)
8:     for i < Parents size do
9:         Crossover
10:        Mutation
11:    end for
12:    R_t = P_t ∪ Q_t
13:    F = Non-Dominated sort (R_t)
14:    Crowding Distance
15:    Select N individuals
16: end while
```

4.1. Algorithm Overview

In both services, the algorithm receives several common data as input, such as the size of the staff to be scheduled, the year, the month, the employees on vacation or sick leave during the month, and other restrictions. Furthermore, the algorithm receives the needs for the scheduled period, which may vary for each day of the week. There is also the possibility of receiving previous schedules or information from these. This is important to know the shifts worked by each employee so hard constraints are not violated. For instance, assigning a morning shift on the first day of the month to whom worked one night on the last day of the previous month, it should be avoided. In addition, GA specific parameters are also required, such as the population size, mutation probability, the number of chromosomes in the Tournament Selection T, the number of parents desired r and the number of generations.

The algorithm aims to start the population with feasible solutions, for this is required to satisfy the hard constraints. It is able to distinguish weekends from weekdays in every month under construction, and the Portuguese national holidays, both fixed and those that vary every year (like Easter Sunday), are also included. This is important to adjust the needs correctly, since these days are considered differently from the general working days.

The flowchart of the algorithm is similar for both cases under study. Initially a random population is generated, then each individual is evaluated and distributed across the different Pareto fronts, where the non-dominated solutions are on the first front. Afterwards a tournament selection is performed to choose the parents that will cross over and originate offspring, which may or may not mutate. The parents and children are combined in a large pool from which N individuals are chosen, its number being equal to the population size defined as input, according to the front they belong to. Crowding

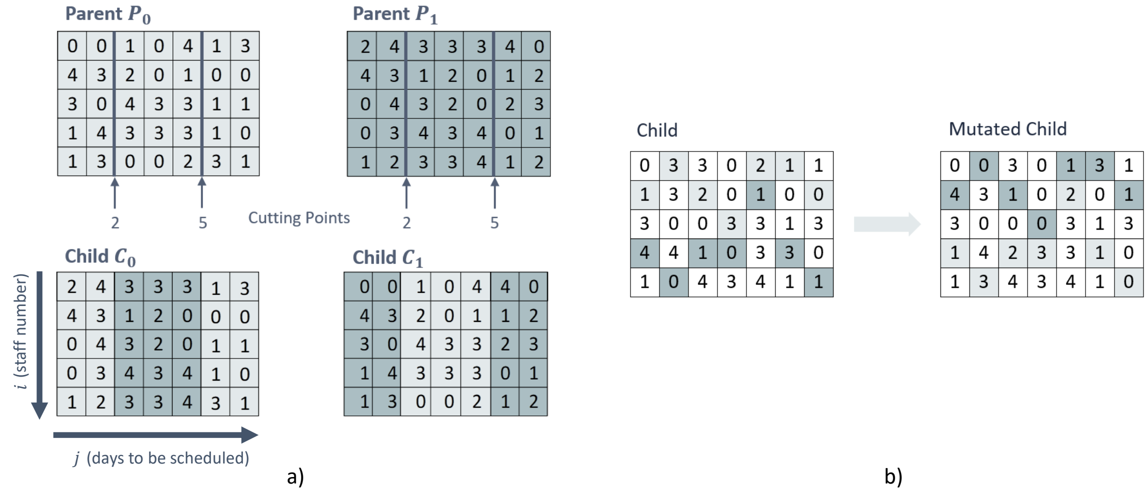

Figure 1: Chromosome representation: $[i \times j]$ matrix; (a) Crossover operator and (b) Mutation operator.

distance is used if it is necessary to choose individuals from the same front. This process is repeated as many times as needed until the defined stop criterion is reached.

The last step of the algorithm consists in determining the final list of non-dominated solutions, as a subset of the first Pareto-front (F_1) of all generations. In each generation, the list with chromosomes belonging to the front of non-dominated solutions is kept. In the end, they all come together and non-dominant sorting is applied as they may no longer be all non-dominated. Then, the duplicates are removed and the best solution can be obtained from the different alternatives. For this purpose, the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method, a multi-criteria decision method, is applied. Here, it is necessary to highlight and distinguish two terms: alternatives and criteria. Alternatives are the different options which need to be evaluated in order to select the best one or, in other words, alternatives are the different chromosomes that result in schedules. Meanwhile the criteria, also known as attributes, will impact the selection of alternatives. Three characteristics must be ensured: completeness - to guarantee that all the important criteria are included -; exclusiveness - to avoid redundant criteria - and operationality - each alternative should include all the criteria, for the alternatives to be confronted with each other according to each criterion [23].

4.2. ICU

ICU's planning is done for an entire month. For each shift in the day, an ideal number of physicians to be allocated is defined. However, in this case, it is not always possible to meet those needs, because the lack of physicians is a recognized problem. Thus, an ideal limit is defined – that is, the ideal number of physicians that must be assigned in each shift –, and a minimum limit that needs to be reached.

There are some conditions that should not be violated – i.e. hard constraints –, namely if a physician was assigned to a night shift, he/she cannot be assigned to the morning, prolongation or night shift the next day, and the minimum coverage limit considered for the number of physicians in each shift must be satisfied. Although the initial population is created randomly, these conditions are repaired in the end if they are violated, to start with feasible solutions. On the other hand, it is possible to assign the same physician to more than one shift in a day, as long as they are not simultaneous. For instance, someone assigned to a morning shift can be assigned to the night shift in the same day, which is equivalent to a 24-hour shift. Nonetheless, this is a condition less desired by the majority, so this occurrence should be penalized. The same happens if two consecutive prolongation shifts are assigned. Another soft constraint is that shifts may not be filled in an ideal way due to the lack of physicians.

In this case, the Decision Maker (DM) is one of the physicians, who knows the needs of the physicians and has experience in the task of decision making. One of the objectives is to minimize the penalties resulting from fairness deviations between physicians. In this case, a cumulative scoring system is created. These scores provide an overview of the physicians most affected over time, through the evaluation of the total number of hours, number of nights, weekends and 24-hour shifts worked. These scores are assigned to each physician as follows: first they are organized in descending order in relation to each of the parameters. A score equal to the position of the physician is given. The procedure is repeated for each of the parameters, however, the penalty is different for each of the shifts (night shift:

score × 1.50; weekend shift: score × 2; and 24-hour shift: score × 1.75)

Thus, to evaluate the solutions a first fitness function will evaluate the following factors: (i) the difference in absolute value between the hours worked by each physician and the average hours worked, where the penalty cost is equal to the sum of differences; (ii) the same is applied for number of 24-hour shifts, weekends, and nights, only with a different weight; (iii) the previously described cumulative score, where the penalty score is equal to the sum of differences; and (iv) maximum number of hours worked: in order not to overload physicians, the penalty cost is equal to the difference of number of hours worked and 180 hours times 50.

The other objective tries to meet the individual preferences of each physician as much as possible. To evaluate this objective the next factors are evaluated by a second fitness function: (i) the number of consecutive prolongation shifts, for which the penalty is the proper value; (ii) count the number of shifts ideally filled, for which the penalty cost is that value unless the limit coverage is not fulfilled, meaning a hard constraint would be violated, so the penalty cost is the number of shifts times 1000; (iii) the team affinity (the physicians with more affinity should be together), the penalty cost being equal to the number of shifts where there are no team affinity; (iv) the individual preferences satisfaction, the penalty cost being equal to the number of preferences not satisfied; and (v) the average number of days-off after a 24-hour shift – in this case, the lower the average number of days-off the higher is the penalty cost.

For both fitness functions, the objective is to minimize the values resulting from the penalty costs assigned. Through the objective functions, it is now possible to calculate non-dominated solutions. The next step is to find the best solution using the TOPSIS method. For this, it is necessary to define criteria and weights, as is described followingly.

Survey Results

A study was carried out through a survey. This had as main goal to understand which criteria gives rise to a good solution and the degree of importance for each one of the physicians, in order to assign them the right weights or, at least, to rank them accordingly. Two alternatives were given (plus the indifferent option) so that the preferred option was chosen. Through the choices it is possible to evaluate the criteria and understand the degree of importance of each one concerning the others. For this purpose, the Condorcet method is used to adequately assess this pairwise comparison.

Figure 2 represents the preferences matrix regarding pairwise confrontation. This shows an election with 8 voters. There are three numbers in each cell according to the vote options (For, Against, and Neutral), and these are always related to the candidate on the left column when confronted with the candidate in the top row. To illustrate it better, four physicians preferred C_1 over C_2, three voters preferred C_2 over C_1, and one physician had no preference between the two. Note that C_x represents the different criterion: C_1 is the absolute value difference in the number of hours worked between physicians; C_2 is the absolute value difference of weekends number; C_3 is the absolute value difference of cumulative scores; C_4 is related to preferences and C_5 to the number of shifts ideally filled. From the very beginning it is possible to identify the Condorcet winner and loser, identified in Figure 2. Criterion C_3 was always chosen over others, so it is the preferred among all confrontations, having won them all. On the other hand, criterion C_5 is never chosen in relation to others, having lost all confrontations. Therefore, the first and last places in the ranking are already known. There is a tie between criteria C_1 and C_4. In this way, the rank, which satisfied the greatest number of physicians, obtained through the survey is: 1^{st} balance between physicians (C_3), 2^{nd} difference in absolute value between hours worked and preferences (C_1 and C_4), 3^{rd} difference in absolute value between weekends worked (C_2) and lastly the filling of shifts ideally (C_5).

	C_1	C_2	C_3	C_4	C_5	
C_1	-	4,3,1	0,8,0	3,3,2	6,2,0	2 wins, 1 loss
C_2	3,4,1	-	2,4,2	2,4,2	3,2,3	1 win, 3 losses ⇒ Condorcet Winner
C_3	8,0,0	4,2,2	-	5,1,2	7,1,0	4 wins, 0 losses
C_4	3,3,2	4,2,2	1,5,2	-	4,3,1	2 wins, 1 loss
C_5	2,6,0	2,3,3	1,7,0	3,4,1	-	0 wins, 4 losses ⇒ Condorcet Loser

Legend: For, Against, Neutral

Figure 2: Preference matrix with 8 voters, highlighting the Condorcet Winner and Condorcet Loser.

Once the rank has been determined, it is possible to assign the weights according to the relative importance of each criterion. In this case, the approach applied is one of the *ranking methods*, in particular the rank sum. In this approach weights are computed from the individual ranks. The normalized weight can be expressed as:

$$w_j = \frac{n - p_j + 1}{\sum_{k=1}^{n} n - p_k + 1} \qquad (1)$$

where n is the number of criteria ($n = 5$) and p_j is the rank of the $j - th$ criterion [24]. Table 1 provides the weights obtained according to the ranking defined through the Condorcet method.

Table 1: Computing the weights using the Ranking method.

	Rank	Weight (n-p_j+1)	Normalized Weight
C_1	3	3	0.21
C_2	4	2	0.15
C_3	1	5	0.36
C_4	3	3	0.21
C_5	5	1	0.07
	Sum	14	1

Once the criteria are ranked and computed, one can choose the best solution. For this purpose, the TOPSIS method is used.

4.3. IS

The IS's scheduling horizon spans over an entire calendar month, starting on the first day and ending on the last one. On each day and for each shift there are specific needs that should, whenever possible, be met. Usually, week needs are the same for the whole month, with slight differences in cases where equipment is under maintenance. Weekends have fewer needs than weekdays.

In this case the hard constraints are: (i) a technician cannot be assigned to more than one shift within a day; (ii) if a technician is in breastfeeding period she must take shifts of 6 hours; (iii) the hospital needs should be satisfied; (iv) according to legislation, a technician cannot work more than 6 consecutive days and (v) if a technician was assigned to a night shift, he/she cannot be assigned to a morning or an afternoon shift in the next day. Concerning soft constraints, a technician should not work more than 10 consecutive hours; on their birthday they are entitled to a day-off; the most part of technicians have a certain functional area or skills, so they should be assigned to those; and ideally, the technician should do just a day in a weekend during the entire month. They intend to maintain the preset rotations. In the case of the central group, they work mornings for one week and afternoon shifts for the next.

As in the case of the ICU, there is also a DM here that supports the definition of the fitness functions. In this case, the DM is one of the technicians responsible for building the schedule every month. Unfortunately, there was no opportunity to gather more opinions, as happened at the ICU. After some conversations with the DM, it was realized that the needs and preferences of the hospital and the individual preferences of each technician are the main objectives.

To evaluate the fitness function concerning hospital's preferences, penalty costs are attributed to following factors: (i) to meet the needs in relation to each technician functional area; (ii) the hospital needs regarding the number of technicians needed in each shift; (iii) the number of hours worked by each technician, which should not be less than 40 hours a week, except in special cases, such as breastfeeding period or weeks with holidays.

For (i) if all needs are met (even if there are more than supposed), no penalty cost is assigned. Otherwise, it is assigned a penalty cost equal to the number of unfilled functional areas times 7.5. In (ii) if a need is not met, one of the hard constraints is violated so a high cost is assigned, in particular, one thousand times the number of needs not met. Also, it is evaluated the average difference in the absolute value of needs for each day. Finally, for (iii) a penalty cost is not assigned if all technicians work the minimum number of hours. Otherwise, the penalty cost is equal to the difference in absolute value between the minimum number of hours which are supposed to be worked and the number of hours worked times 15.

On the other hand, the fitness evaluation concerning technicians preferences is carried through the following factors: (i) pairs of days-off; (ii) the number of M6 shifts compared to the number of 10-hour shifts; (iii) comparison between number of days-off and weekend shifts; (iv) individual preferences and (v) rotation desired.

In (i) and (iii) if the number of pairs of days-off is equal to the number of weekends, no penalty cost is assigned. Otherwise, the penalty cost is equal to 30 times the difference in absolute value between the number of weekends and the number of pairs. In order to maintain fairness, the sum of differences between technicians regarding pairs is also evaluated. In (ii) if the number of M6 shifts is equal to the number of S/I10 shifts worked, no penalty is assigned. Otherwise, the penalty cost is the difference in absolute value between both numbers times 10. For (iv) will be evaluated if a technician works an extra shift and does not desire it, in this case, the penalty score is equal to the number of extra shifts times 50. Finally, for (v) if the rotation is not maintained the penalty cost is equal to the number of rotations not respected times 100.

In contrast to what happened in ICU case, in the IS there was no opportunity to collect information from different technicians. However the weights and criteria were defined with the DM, as follow: (i) 30% for the skills; (ii) 25% for the average number of pairs; (iii) 15% for the sum of differences concerning pairs; (iv) 25% for the minimum number of hours worked and (v) 5% for the individual preferences.

5. Computational Experiments

The algorithm was tested using real instances. The algorithm is coded in Python. The tests were performed on a computer with an Intel Core i7-8565U processor (4 cores/8 threads each with a base frequency of 1.80 GHz) and 8 GB of RAM.

Through some preliminary experiments it was found the best combination of parameters, which will be used throughout the computational experiments. These are the following: (i) population size, $N = 160$; (ii) number of individuals participating in the Tournament Selection, $T = 4$; (iii) number of the parents resulting from the Tournament, $r = 80$; and (iv) mutation probability, $p_m = 0.08$.

5.1. Fitness Function over Time

First, an analysis of the values of the fitness function over time was made. In this one the December conditions were considered as input, but without considering the historical data available. Under these conditions different times were used as stop criteria. It was noticed that the values of fitness 2 stabilize faster than those of fitness 1. Fitness 2 values remain more or less constant after 25 minutes, while fitness 1 values only stabilize after 60 minutes have passed. Nevertheless, the values of fitness 1 decrease much slowly after the second 30 minutes than in the first. This allows to conclude that using 30 minutes as stopping criterion is enough to obtain good solutions. Therefore, this was the stopping criterion used for the remaining experiments.

5.2. Evaluating the impact of Historical Records

The ICU service provided a historical record with information on all schedules, number of hours, absences and holidays for the year 2020. In this way, computational experiments were conducted to understand the influence of these records in the results. For this purpose, three scenarios were used: (scenario 0 + 12) where schedules for all months of the year were created through the algorithm; (scenario 6 + 6) where historical information provided by the ICU in the first half of the year was used, and the remaining months were generated by the algorithm; and (scenario 1 + 11) where historical records of the first eleven months are used and the last month of the year was created through the algorithm. The solutions used throughout the analysis are those that were selected by the TOPSIS method. In all three scenarios all other conditions used were the same. Here, the difference in cumulative scores (which represents the number of hours and weekend, night and 24-hour shifts), the average number of hours in each month and the number of shifts ideally completed were evaluated.

The results show that the variability between physicians tends to be lower when using the algorithm, which is synonymous with fairer results (see Figure 3). On the other hand, the results of scenario 0 + 12 concerning the average number of hours show the algorithm starts with higher average values but, in total, it presents a smaller variation and thus, a better balance among the physicians. The highest points are observed, for the three situations, in months that register the greatest number of physicians on vacation. The high number of hours is justified with the number of shifts ideally filled, since the algorithm was always closer to the ideal number of shifts than in reality. This shows that there is a tendency to satisfy this need of the hospital but that, despite this, there is not an exaggerated number of hours worked on average. This said, it is possible to state that the algorithm tries to find the best balance between the number of hours, the ideal filling of shifts and the fairness within the schedule.

Through all the described above, it is possible to notice that the algorithm always tends to stabilize, increasing the number of shifts ideally completed but without compromising the average number of hours. In addition, it is still possible to validate that the algorithm adapts to all months of the year, identifying holidays and weekends, adapting to a variable number of physicians, both internal and external.

Table 2: Weights' range for which the current rank does not change.

	Current Rank	Current Weight	Weights' range for which the current rank does not change
C_1	3	0.210	[0.000, 0.299]
C_2	4	0.140	[0.100, 0.729]
C_3	1	0.360	[0.000, 0.909]
C_4	3	0.210	[0.111, 0.849]
C_5	5	0.070	[0.065, 0.159]

5.3. TOPSIS sensitivity Analysis

Any decision support model that depends on personal and qualitative judgments may be subject to uncertainty due to the inherent subjectivity. The goal is to evaluate how changes in the weight of criteria would interfere with the solution resulting from TOPSIS. The criteria and respective weights were described in Section 4. Once the best non-dominated individuals were selected, TOPSIS is applied to choose the best solution. In total, 25 non-dominated solutions were obtained, and one of them was considered the best according to the previously defined weights. Table 2 presents the ranges of criterion weights that do not change the best solution. For each criterion the weight is varied until a new solution was obtained. Based on the results presented in Table 2, one can notice that even small variations of the weights may give rise to different solutions. For example, for C_5 only a 0.005 decrease or a 0.09 increase in weight already results in a different solution. This shows that the defined ranking is not very stable. On the other hand, the criterion C_3 presents a large range where no changes are verified, a large variation is necessary to obtain a new

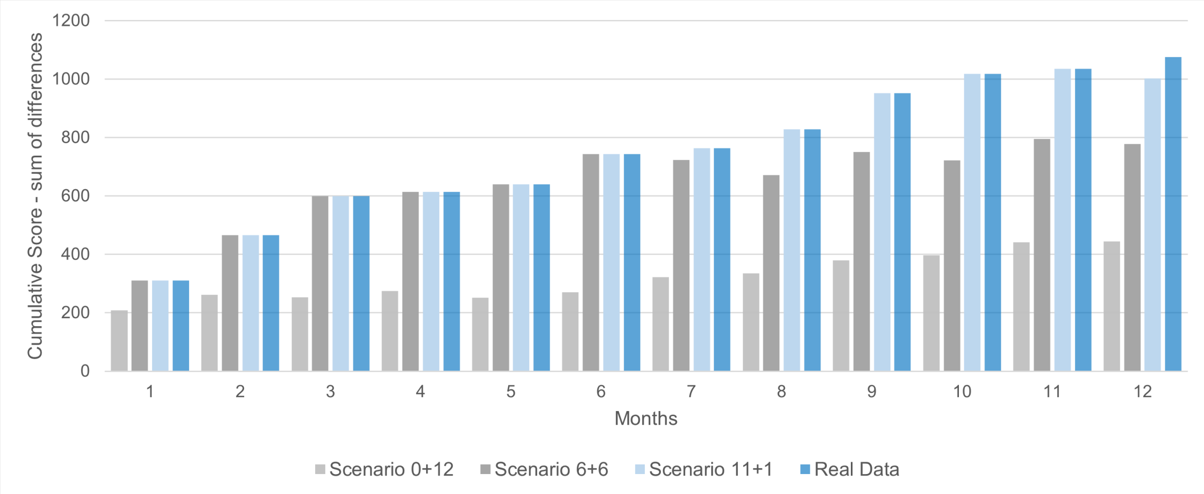

Figure 3: Result of the sum of the cumulative scores differences for the three scenarios and real case.

solution. Moreover, no matter how much the weight is reduced, it will not find a new best solution. Criterion C_5 is the most sensitive, because it is the one that has a smaller weight range for changes to occur, followed by C_1. The remaining criteria present the same solution for a longer interval, so they are less susceptible to changes.

5.4. Validation with HLL

A potential proposal for the December 2020 schedule was presented to HLL. It was possible to sustain that the algorithm is capable of generating admissible solutions respecting all hard constraints, and tries whenever possible to reduce penalties associated with the lack of achievement of soft constraints. Additionally, it did so in a very short time, contrary to what happens in reality. The algorithm presented worse values regarding the average number of hours worked, however this can be explained through the number of shifts ideally filled, which is about 20% higher than in the schedule built in the unit. Even so, the average number of hours is acceptable and within the limit defined of 180 hours, as mentioned above. Table 3 presents a comparison between real schedule and algorithm schedule results.

Overall, the DM is satisfied with the solutions presented and the feedback was positive. According to him/her, the proposed solution is already sufficiently balanced between elements. He/she highlights the fact that it takes into account not only the month under construction, but also the previous ones. He/she considers this is a very important aspect, as it contributes to long-term fairness and allows to compensate less benefited physicians in previous schedules. More importantly, in addition to these aspects, the algorithm can present good solutions within a reasonable time.

The main aspect to consider is the construction of a method that allows the schedule to be generated by a user other than the programmer.

6. Conclusions

This work addressed a scheduling problem in a hospital context, focused mainly on the ICU of HLL, although it has also been implemented for the IS. The main goal was to develop, propose and validate an algorithm able to optimize the staff scheduling in a timely manner, without inconsistencies, respecting all constraints and incorporating the main objectives of each service. In the case of the ICU these were to find a balance in fairness and satisfaction of preferences, while in the IS the biggest focus was the preference satisfaction, both individual and the hospital's. A decision support algorithm, based on NSGA-II, was developed. For this purpose, the criteria and needs of each service were identified. The algorithm was able to respect the needs and hard constraints and shows a tendency to decrease the violation of soft constraints. It is capable of presenting several non-dominated solutions and choosing one according to the defined ranking. In the ICU this was achieved through the results of a survey made to the physicians.

Overall, the solutions generated for the ICU by the algorithm show better results for most criteria, except for the average number of hours worked,

Table 3: Comparison between real schedule and algorithm schedule.

	Real Schedule	Algorithm Schedule
Ideal Shifts Filled	79.00%	90.16%
Team affinity	6	9
Average number of hours	138.40	160.93
Cumulative Score	1076.00	987.33

however this can be justified by the increase in the number of shifts ideally filled, and the values found are not higher than those considered admissible. In addition, it is still possible to validate that the algorithm adapts to all months of the year, identifying holidays and weekends, adjusting its results to a variable number of physicians, both internal and external. Through the results it was possible to conclude that regardless of using information from the past or not, the algorithm tends to find a balance between physicians concerning the previous months and the number of hours and types of shifts worked among them. This is a very relevant factor because it can contribute to increased satisfaction.

The feedback and validation from the HLL was positive, both strengths and weaknesses were highlighted. The most appreciated contribution is to reduce the time of a task that took two days to a few minutes and still improve some results, incorporating fairness and preferences. There are still aspects to improve, however they can be overcome.

As for future work, the implementation of a proper user interface will bring added value. At the moment the algorithm is able to produce good solutions but the programmer is required. In particular, the input concerning vacations, availability restrictions, etc., is done by hand, and a way to do this through the import of files would greatly improve the usability of the solution proposed.

References

[1] Melanie Erhard, Jan Schoenfelder, Andreas Fügener, and Jens O Brunner. State of the art in physician scheduling. *European Journal of Operational Research*, 265(1):1–18, 2018.

[2] Serpil Ozcan and Peter Hornby. Determining hospital workforce requirements: a case study. *Hum Resour Dev J*, 3(3):210–20, 1999.

[3] Jorne Van Den Bergh, Jeroen Beliën, Philippe De Bruecker, Erik Demeulemeester, and Liesje De Boeck. Personnel scheduling: A literature review. *European Journal of Operational Research*, 226(3):367–385, 2013.

[4] Jorge De Almeida Simões, Gonçalo Figueiredo Augusto, and Cristina Hernández-quevedo. Portugal - Health System Review. *Health Systems in Transition*, 19(2), 2017.

[5] Lena Wolbeck, Natalia Kliewer, and Inês Marques. Fair shift change penalization scheme for nurse rescheduling problems. *European Journal of Operational Research*, 2020.

[6] Jonas Volland, Andreas Fügener, and Jens O Brunner. A column generation approach for the integrated shift and task scheduling problem of logistics assistants in hospitals. *European Journal of Operational Research*, 260(1):316–334, 2017.

[7] Jens O Brunner and Günther M Edenharter. Long term staff scheduling of physicians with different experience levels in hospitals using column generation. *Health care management science*, 14(2):189–202, 2011.

[8] Gerard M Campbell. A two-stage stochastic program for scheduling and allocating cross-trained workers. *Journal of the Operational Research Society*, 62(6):1038–1047, 2011.

[9] Uwe Aickelin and Kathryn A Dowsland. An indirect genetic algorithm for a nurse-scheduling problem. *Computers & Operations Research*, 31(5):761–778, 2004.

[10] Subhamoy Ganguly, Stephen Lawrence, and Mark Prather. Emergency department staff planning to improve patient care and reduce costs. *Decision Sciences*, 45(1):115–145, 2014.

[11] Jens O Brunner, Jonathan F Bard, and Rainer Kolisch. Flexible shift scheduling of physicians. *Health care management science*, 12(3):285–305, 2009.

[12] Raik Stolletz and Jens O Brunner. Fair optimization of fortnightly physician schedules with flexible shifts. *European Journal of Operational Research*, 219(3):622–629, 2012.

[13] Pooyan Kazemian, Yue Dong, Thomas R Rohleder, Jonathan E Helm, and Mark P Van Oyen. An ip-based healthcare provider shift design approach to minimize patient handoffs. *Health care management science*, 17(1):1–14, 2014.

[14] Omar EL-Rifai, Thierry Garaix, Vincent Augusto, and Xiaolan Xie. A stochastic optimization model for shift scheduling in emergency departments. *Health Care Management Science*, 18(3):289–302, 2015.

[15] Christopher N Gross, Jens O Brunner, and Manfred Blobner. Hospital physicians can't get no long-term satisfaction–an indicator for fairness in preference fulfillment on duty schedules. *Health care management science*, 22(4):691–708, 2019.

[16] Jonathan F Bard and Hadi W Purnomo. Preference scheduling for nurses using column generation. *European Journal of Operational Research*, 164(2):510–534, 2005.

[17] Lorraine Trilling, Alain Guinet, and Dominiue Le Magny. Nurse scheduling using integer linear programming and constraint programming. *IFAC Proceedings Volumes*, 39(3):671–676, 2006.

[18] Andreas Fügener, Jens O Brunner, and Armin Podtschaske. Duty and workstation rostering considering preferences and fairness: a case study at a department of anaesthesiology. *International Journal of Production Research*, 53(24):7465–7487, 2015.

[19] Richard Baum, Dimitris Bertsimas, and Nathan Kallus. Scheduling, revenue management, and fairness in an academic-hospital radiology division. *Academic radiology*, 21(10):1322–1330, 2014.

[20] Renato Bruni and Paolo Detti. A flexible discrete optimization approach to the physician scheduling problem. *Operations Research for Health Care*, 3(4):191–199, 2014.

[21] Mahdi Hamid, Reza Tavakkoli-Moghaddam, Fereshte Golpaygani, and Behdin Vahedi-Nouri. A multi-objective model for a nurse scheduling problem by emphasizing human factors. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 234(2):179–199, 2020.

[22] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

[23] Hsu-Shih Shih, Huan-Jyh Shyur, and E Stanley Lee. An extension of topsis for group decision making. *Mathematical and computer modelling*, 45(7-8):801–813, 2007.

[24] GO Odu. Weighting methods for multi-criteria decision making technique. *Journal of Applied Sciences and Environmental Management*, 23(8):1449–1457, 2019.