Original Paper

Constraint Programming and Neural-Inspired Optimization for Hospital Diagnostic Examination Scheduling: Comparative Development and Evaluation Study

João Lopes, PhD; Afonso Lobo, MSc; Fábio Rocha, BSc; Tiago Guimarães, PhD; Júlio Duarte, PhD; Manuel Filipe Santos, PhD

ALGORITMI/LASI Research Centre, University of Minho, Guimarães, Portugal

Corresponding Author:

João Lopes, PhD
ALGORITMI/LASI Research Centre
University of Minho
R. da Universidade
Guimarães 4804-533
Portugal
Phone: 351 934373667
Email: jlopes@dsi.uminho.pt

Abstract

Background: Diagnostic examination scheduling is a recurrent operational challenge in health care because limited daily capacity must be allocated across heterogeneous requests with different urgency and follow-up timing requirements. In this context, optimization methods can support more timely and efficient scheduling decisions.

Objective: In collaboration with Unidade Local de Saúde de Santo António, an optimization-based approach was proposed to improve the management of outpatient examination scheduling.

Methods: Anonymous operational data from Unidade Local de Saúde de Santo António, corresponding to 2025, were analyzed to model the scheduling process over a 4-week (28-d) horizon comprising 2460 examination requests, including 156 examinations with follow-up requirements. Two optimization approaches, Constraint Programming Satisfiability (CP-SAT) and a neural combinatorial optimization (NCO)–inspired constructive approach, were evaluated under the same daily capacity structure, prioritization logic, and penalty-based framework, confirmed across 10 independent runs under 4 solver configurations, with operational outcomes remaining unchanged. The best-performing approach was then compared with the hospital scheduling solution for the same period.

Results: CP-SAT achieved the best overall performance among the 2 optimization approaches. It scheduled 1400 examinations vs 1391 for the NCO-inspired approach, reduced free time from 42 to 3 minutes, and achieved a substantially lower total penalty (49,790.78 vs 83,652.10). Both methods achieved 0 unscheduled due examinations, 0 delayed scheduled examinations, and 0 delayed unscheduled examinations. The NCO-inspired approach required less runtime than CP-SAT (17.2 vs 214.6 s). When compared with the hospital scheduling solution, CP-SAT scheduled more examinations (1400 vs 1147) and eliminated all unscheduled due examinations (0 vs 80), delayed scheduled examinations (0 vs 23), and delayed unscheduled examinations (0 vs 40).

Conclusions: Both optimization approaches generated feasible schedules under the same evaluation framework, but CP-SAT produced the best overall results in the tested setting. The findings suggest that optimization-based scheduling can improve the timely allocation of diagnostic examinations and reduce follow-up–related delays when compared with routine hospital scheduling practice.

JMIR AI 2026;5:e97738; doi: 10.2196/97738

Keywords: health care scheduling; outpatient exam scheduling; diagnostic exam scheduling; Constraint Programming Satisfiability; CP-SAT; neural combinatorial optimization; optimization models; decision support; AI

Introduction

Background

The effective delivery of health care services depends on the efficient execution of multiple operational processes and is becoming increasingly reliant on computer-based systems capable of supporting complex decision environments [1-3]. Given the objectives that health care organizations must meet and the need to provide timely and appropriate patient care, systems capable of improving, evaluating, and anticipating operational situations are increasingly important [2,3]. At the same time, the health care sector has become progressively richer in data, creating new opportunities to support decision-making through digital and analytical methods [4,5].

This growing volume of clinical and organizational data has reinforced the relevance of decision-support, predictive, and prescriptive approaches in health care management. In practice, health care services must continuously make decisions under limited capacity, variable demand, and heterogeneous clinical requirements, which makes the quality of operational planning particularly important. In such settings, data-driven approaches can support more structured and timely decisions, especially in processes where scheduling directly affects access to care and the use of hospital resources [4-6].

Health care scheduling is not a single problem but a broad family of distinct operational problems, each with its own constraints, objectives, and decision structures. Well-established subdomains include the patient admission scheduling problem, which assigns inpatients to beds and wards over time; the operating theater scheduling problem, which plans surgical cases across rooms and downstream resources; and outpatient and diagnostic examination scheduling, which allocates limited daily service capacity across incoming requests. Although these problems share a common concern with the efficient use of constrained resources, they differ substantially in their temporal granularity, resource types, and clinical priorities, and solution approaches developed for one subdomain do not transfer directly to the others. This study focuses specifically on the examination scheduling problem, in which requests drawn from a waiting list must be assigned to limited daily room–shift capacity. This setting is distinguished from other scheduling subdomains by 3 characteristics that jointly shape the problem addressed here: capacity is defined by discrete room-shift slots with fixed daily availability; requests are heterogeneous in duration and in clinical urgency; and prioritization is driven by follow-up timing requirements, so that the cost of not scheduling an examination depends on how overdue its clinically recommended follow-up is. Framing the problem in this way clarifies both the scope of the contribution and the operational context against which the proposed approaches are evaluated.

Within this context, scheduling remains one of the most persistent operational problems in health care because it directly affects waiting time, resource utilization, and service responsiveness [7-10]. This is especially relevant in diagnostic services, where hospitals must allocate finite daily capacity across heterogeneous examination requests while also considering urgency and follow-up timing. At the same time, the growing availability of operational data and decision-support capabilities has reinforced interest in new algorithmic approaches, including AI, machine learning, and other data-driven optimization methods, to improve scheduling processes and support more timely care delivery [11,12].

Related Work

Health care scheduling has been widely studied through optimization methods because these problems typically involve competing constraints on time, capacity, and priority, often under uncertainty and limited resource availability [13,14]. In this context, exact methods provide structured formulations for constrained allocation problems, whereas heuristic and metaheuristic approaches have been extensively explored to address large-scale or computationally difficult scheduling settings [15,16]. This broader literature is relevant because it establishes the main optimization paradigms commonly used in health care operations.

A closely related stream of research has addressed hospital scheduling problems in surgery-related settings, where scheduling decisions must account for limited capacity, uncertainty, and downstream operational effects. Previous studies have shown that stochastic and optimization-based approaches can support surgery scheduling under constrained hospital resources, particularly when time-sensitive allocation decisions interact with capacity limitations and service variability [17]. A stochastic hierarchical approach has also been proposed for master surgical scheduling, considering multiple stakeholders and resources, including surgeons, machines, operating rooms, and recovery ward beds, while pursuing waiting time and utilization objectives [18]. More recent hospital-based work has also shown that optimization can improve scheduling performance when compared with routine hospital practice, reinforcing the practical value of algorithmic support in real health care environments [6].

More specifically, optimization has also been applied in diagnostic and examination scheduling contexts. Simulation optimization has been used for ultrasound appointment scheduling and examination room assignment, showing that scheduling decisions can improve waiting-related and utilization-related outcomes [8]. More recent work has highlighted that diagnostic scheduling may also depend on equipment diversity and physician expertise, which increases the operational complexity of these systems [9]. In positron emission tomography workflows, dynamic scheduling and control mechanisms have likewise been proposed to improve time efficiency and reduce operational burden [10].

More recently, learning-based methods have emerged as a distinct line of research for scheduling problems, motivated by the growing availability of operational data and by advances in machine learning. In particular, neural combinatorial optimization (NCO) and deep reinforcement learning have been applied to construct scheduling solutions directly, rather than solving an explicit mathematical program at each decision point. In outpatient settings, a graph-based two-agent deep reinforcement learning approach has been

proposed to sequence patient activities and allocate service desks, achieving solution quality comparable to metaheuristics while offering advantages in computational efficiency [19]. Related work has combined optimization, machine learning, and overbooking strategies to improve resource utilization in medical appointment scheduling [20]. These approaches illustrate a shift toward data-driven and learning-based scheduling, but they are typically evaluated on their own terms, and comparisons between exact optimization and neural-inspired constructive methods under a common evaluation framework and on the same real-world data remain scarce.

Taken together, these studies show that optimization-based scheduling is relevant across different hospital contexts, from surgery to diagnostic services. They also suggest that diagnostic examination scheduling remains a suitable setting for testing alternative optimization paradigms, particularly when daily capacity decisions, operational timing, and follow-up-related constraints must be managed jointly.

Study Aim and Contribution

This study addresses the scheduling of diagnostic examinations from a waiting list under limited operational capacity. The aim is to support the generation of feasible daily schedules that balance clinical priority with the efficient use of available resources in a diagnostic setting.

Against this background, the article compares 2 distinct solution paradigms, namely a Constraint Programming Satisfiability (CP-SAT) approach implemented in Operations Research Tools (OR-Tools) and an NCO-inspired constructive approach. The paper contributes in 3 main ways: it formulates diagnostic examination scheduling as a daily capacity-constrained decision problem with follow-up-sensitive prioritization, it evaluates an exact constraint-based approach and a neural-inspired constructive approach under the same operational rules and objective framework, and it provides practical evidence on their comparative behavior against hospital scheduling practice.

Methods

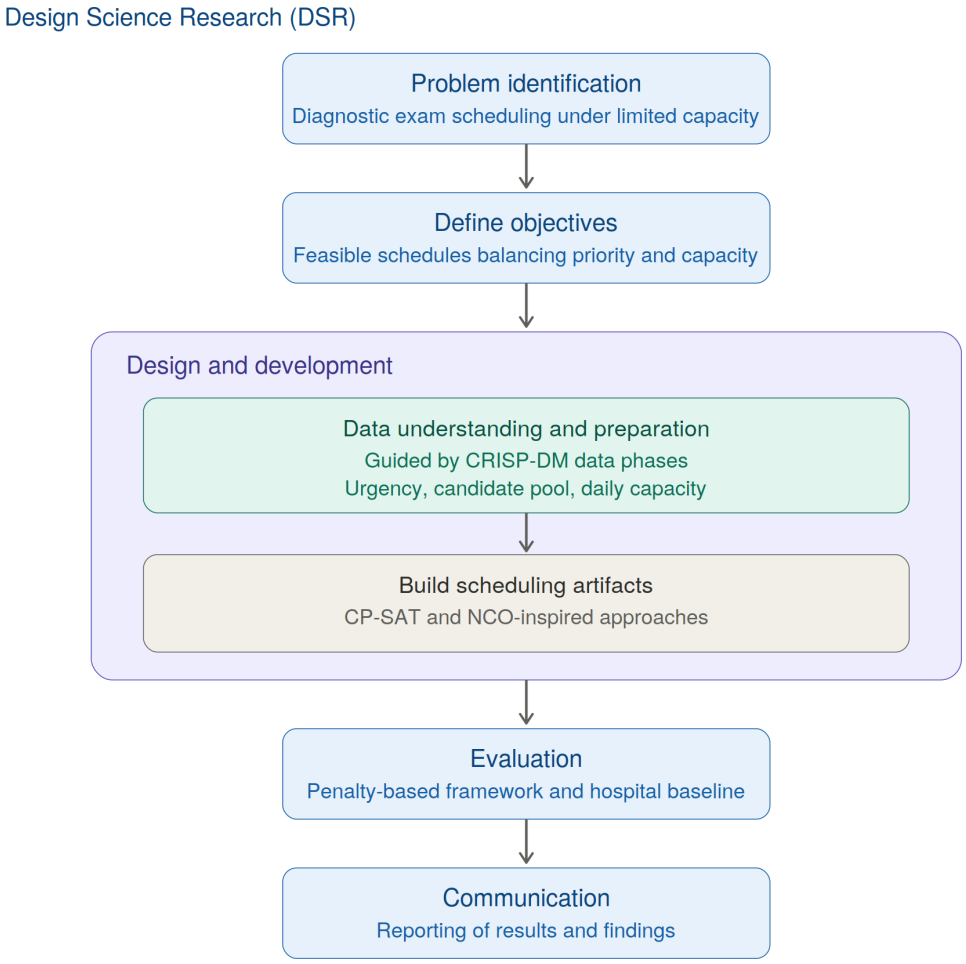
Ethical Considerations

This study was based exclusively on anonymous data provided by the organization involved in this research and did not involve sensitive personal information. Therefore, ethics board review in accordance with the General Data Protection Regulation was not required.

Study Design

This study followed design science research (DSR) as its overarching research methodology. DSR provides a framework for the construction and evaluation of artifacts intended to address real-world problems [21], and it is well suited to research whose central contribution is the design and assessment of algorithmic solution approaches. In this study, the artifacts correspond to the 2 proposed scheduling approaches, which are designed, implemented, and evaluated against real operational data and against current hospital scheduling practice. Because these approaches depend on operational data that must be interpreted and structured before they can be used, the data-oriented stages of the work (namely the understanding and preparation of the hospital's operational records) were organized following the data understanding and data preparation phases of the Cross-Industry Standard Process for Data Mining (CRISP-DM) [22]. CRISP-DM is used here only as a practical guide for these data-handling stages and not as the methodological framework of the study, which remains grounded in DSR. The subsequent formulation, solution, and evaluation of the scheduling approaches follow the design and evaluation logic of DSR. An overview of this methodology, including the role of the CRISP-DM–guided data preparation stage within the DSR process, is shown in Figure 1.

Figure 1. DSR methodology with a Cross-Industry Standard Process for Data Mining (CRISP-DM)–guided stage. CP-SAT: Constraint Programming Satisfiability; NCO: neural combinatorial optimization.



Data Understanding and Preparation

This stage focused on preparing a common operational dataset for both scheduling approaches. The objective was to ensure that the CP-SAT and NCO-inspired approaches used the same examination records, the same capacity structure, and the same prioritization logic before schedule generation.

Data Source and Operational Records

The operational examination data were derived from Unidade Local de Saúde de Santo António (ULSSA), corresponding to the year 2025. This period was selected to provide a recent operational snapshot of the diagnostic scheduling context under analysis. The available records included, among other fields, the examination request date, follow-up information when available, and the timestamps associated with examination execution. These records were then used as the common input basis for both scheduling approaches.

In addition to the examination data, structured operational capacity files describing the available room-shift slots were used. These files specify, for each slot, the operational day, shift name, room identifier, and corresponding time window. Rather than being defined arbitrarily, this slot structure was reconstructed from the historical operational records in order to approximate the actual temporal organization of the service, including the practical start and end times of shifts and the separation between operational periods. Together, the examination records and the slot definitions formed the common data basis required to support the day-by-day scheduling process.

Urgency Construction

Because examination urgency was not available as a standardized field in the original records, a synthetic urgency indicator was constructed to support scheduling prioritization. The variable was designed to combine two temporal dimensions of relevance: the age of the examination request and, when available, the proximity of the associated follow-up date. This construction allowed the prioritization process to distinguish between older requests without follow-up information and examinations whose clinical relevance was reinforced by an upcoming or already delayed follow-up event.

The first component of urgency was defined from request age. For each examination i, the number of days elapsed since the request was computed relative to the most recent request in the dataset, and then normalized to the interval [0,1]:

$$AgeScore_i = \frac{max(RequestDate) - RequestDate_i}{max(RequestDate) - min(RequestDate)} \quad (1)$$

This component was available for all examinations.

For examinations without valid follow-up information, urgency was not assigned directly from request age alone. Instead, request age was evaluated within the subgroup of examinations without follow-up through a percentile-based filter. Let $AgeRank_i^{NoFU}$ denote the percentile rank of the examination within this subgroup. Only examinations above the 90th percentile were assigned a relevant urgency, using a scaled excess function:

$$AgeExcess_i = max\left(0, \frac{AgeRank_i^{NoFU} - 0.90}{1 - 0.90}\right) \quad (2)$$

$$Urgency_i = 0.75 \times AgeExcess_i \quad (3)$$

Under this rule, examinations without follow-up and with a relatively low age receive urgency values close to 0, whereas the oldest examinations in this subgroup can reach a maximum urgency of 0.75. For examinations with valid follow-up information, urgency was constructed by combining the request age with a follow-up proximity score. The number of days between the planning start date and the follow-up date was first computed, clipped to the interval [0,180], and transformed through a logarithmic function in order to assign stronger priority to examinations with nearer follow-up dates:

$$DaysToFollowUp_i = FollowUpDate_i - PlanningStartDate \quad (4)$$

$$DaysClipped_i = min(max(DaysToFollowUp_i, 0), 180) \quad (5)$$

$$FollowUpBase_i = 1 - \frac{log(1 + DaysClipped_i/7)}{log(1 + 180/7)} \quad (6)$$

This score was then weighted according to follow-up type, with a weight of 1.0 for a set follow-up event and 0.6 for a probable follow-up event:

$$FollowUpScore_i = FollowUpBase_i \times TypeWeight_i \quad (7)$$

A preliminary urgency score was then defined by the following equation:

$$Urgency_i = 0.2 \times AgeScore_i + 0.8 \times FollowUpScore \quad (8)$$

Finally, a priority override was applied to examinations with follow-up dates either falling within the planning horizon (*today, horizon_end*) or already overdue (*FollowUpDate<today*). These examinations were reordered in 2 consecutive blocks: first, examinations with follow-up dates inside the planning horizon, sorted by follow-up date in ascending order, and second, overdue examinations, also sorted by follow-up date in ascending order. Urgency values were then reassigned linearly in the interval (0.90, 1.00) according to rank position in this final ordered list, with the first examination receiving urgency 1.00 and the last receiving urgency 0.90. This override ensures that examinations with time-sensitive follow-up requirements dominate the upper part of the scheduling priority scale.

Daily Capacity and Candidate Pool Preparation

Daily operational capacity was defined from the set of room-shift slots available on each day. For each slot, the corresponding duration was derived from the recorded start and end timestamps. The nominal daily capacity was then computed as the sum of the duration of all slots available on that day.

Based on this daily capacity structure, a candidate pool was generated for each decision day from the remaining waiting list. The remaining examinations were ordered according to the urgency indicator, and candidates were incrementally added while the cumulative duration remained within a predefined allowance based on the nominal daily capacity. This mechanism allowed the daily search space to be expanded beyond the nominal capacity through an overbook factor, thereby increasing the set of candidate examinations considered by the scheduling approaches while preserving a consistent day-level screening logic for both methods.

Problem Formulation and Evaluation Framework

Overview

The scheduling problem is formulated over a finite planning horizon composed of operational days. For each day $d \in D$, a set of room-shift slots S_d defines the available scheduling capacity. In the current implementation, the operational template includes 2 shifts per working day, whereas Sundays contain no operational slots. The exact end times of the slots are determined by the reconstructed operational template derived from the hospitals' historical records.

Let cap_{srd} denote the nominal capacity, in minutes, of slot (s, r) on day d, where s represents the shift and r the room. The nominal daily capacity is then given by the following equation:

$$C_d = \sum_{(s,r) \in S_d} cap_{srd} \quad (9)$$

where C_d is the total nominal capacity available on day d.

For each decision day, a daily candidate pool K_d is generated from the remaining waiting list. Examinations are ordered according to the urgency indicator, and candidates are incrementally included until the cumulative examination duration reaches a predefined allowance based on the daily nominal capacity. Let dur_i denote the duration of examination i, and let α denote the overbook factor. The daily candidate pool is therefore constrained by the following equation:

$$\sum_{i \in K_d} dur_i \leq \alpha C_d \quad (10)$$

In the main experimental setting, the reference value used for the overbook factor was α=1.3. This mechanism enlarges the daily search space beyond nominal capacity while preserving

a controlled and identical screening rule for both scheduling approaches. The shared daily evaluation framework is defined through a penalty-based objective function that aggregates unscheduled examinations, accumulated waiting time, follow-up-related delay, unused nominal capacity, and overtime. The total daily cost for day d is defined by the following equation:

$$J_d = \sum_{i \in K_d} \left(w_1 urg_i g_i u_{i,d} + w_2 wait_{i,d} u_{i,d} + w_3 \delta^{set}_{i,d} delay_{i,d} u_{i,d} + w_4 \delta^{prob}_{i,d} delay_{i,d} u_{i,d} + w_5 \delta^{today}_{i,d} u_{i,d} \right) + \sum_{(s,r) \in S_d} \left(w_6 free_{s,r,d} + w_7 over_{s,r,d} \right) \quad (11)$$

where J_d is the total cost of the schedule generated for day d, K_d is the candidate pool considered on day d, S_d is the set of room-shift slots available on day d, $u_{i,d}$ equals 1 if examination i remains unscheduled on day d, and 0 otherwise, urg_i is the urgency score of examination i, g_i is the normalized duration factor of examination i, bounded in the interval (0.30, 1.00), $wait_{i,d}$ is the accumulated waiting time, in days, of examination i by day d, $delay_{i,d}$ is the number of delayed days associated with examination i on day d, $\delta^{set}_{i,d}$ equals 1 when examination i has an overdue *set follow-up appointment* on day d, and 0 otherwise, $\delta^{prob}_{i,d}$ equals 1 when examination i has an overdue *probable follow-up appointment* on day d, and 0 otherwise, $\delta^{today}_{i,d}$ equals 1 when the follow-up date of examination i is exactly day d, and 0 otherwise, $free_{s,r,d}$ is the number of unused minutes in slot (s, r) on day d, $over_{s,r,d}$ is the number of overtime minutes in slot (s, r) on day d, and $w1$,...,$w7$ are the penalty weights associated with each component of the objective function.

Over the full planning horizon D, the total evaluation cost was defined by the following equation:

$$J = \sum_{d \in D} J_d \quad (12)$$

The numerical values assigned to each penalty weight are reported in Table 1. These values were set to reflect the relative operational severity of each penalized component, with the highest penalties assigned to missing a follow-up due on the current day and to incurring overtime, and the lowest to incremental waiting time. Reporting these weights explicitly supports the reproducibility of the reported penalty values and the interpretation of the optimization results.

Table 1. Penalty weights used in the objective function.

Weight	Penalized component	Value, n
$w1$	Unscheduled examination, scaled by urgency and duration	200
$w2$	Accumulated waiting time	1
$w3$	Delay of a set follow-up	50
$w4$	Delay of a probable follow-up	20
$w5$	Follow-up due on the current day	5000
$w6$	Unused nominal capacity (free time)	500
$w7$	Overtime	1500

Hospital Baseline and Comparison Protocol

To assess the practical value of the proposed approaches, the scheduling outcomes were compared against the hospital's actual scheduling practice over the same 4-week horizon. The hospital baseline was reconstructed from the operational records of ULSSA, using the examination appointments effectively made by the institution during the study period. For each examination, the recorded scheduling date and its follow-up requirements were extracted, allowing the same timing–related and follow-up–related indicators computed for the optimization approaches to be derived for the hospital practice. Because certain optimization-specific indicators depend on the internal capacity model of the solver and have no direct equivalent in the hospital records, the comparison was restricted to metrics that are meaningful across both settings. Accordingly, the baseline comparison focused on the number of scheduled examinations, unscheduled due examinations, and delayed scheduled and unscheduled examinations, which reflect the operational ability of each scheduling process to support timely examination allocation and follow-up compliance. The daily evolution of unscheduled examinations, for both the full set of examinations and the subset of urgent examinations, was additionally tracked across the day horizon to characterize performance over time.

Proposed Scheduling Approaches

Overview

This study evaluates two scheduling approaches with distinct search paradigms under the same daily candidate-pool logic and the same penalty-based evaluation framework. The first corresponds to a CP-SAT–based optimization approach implemented in OR-Tools. The second corresponds to an NCO-inspired constructive approach in which a lightweight neural scoring model guides the generation of feasible daily schedules. Because both approaches operate on the same prepared data and are evaluated under the same objective structure, their comparison reflects differences in search logic rather than differences in input preparation or scoring design.

CP-SAT

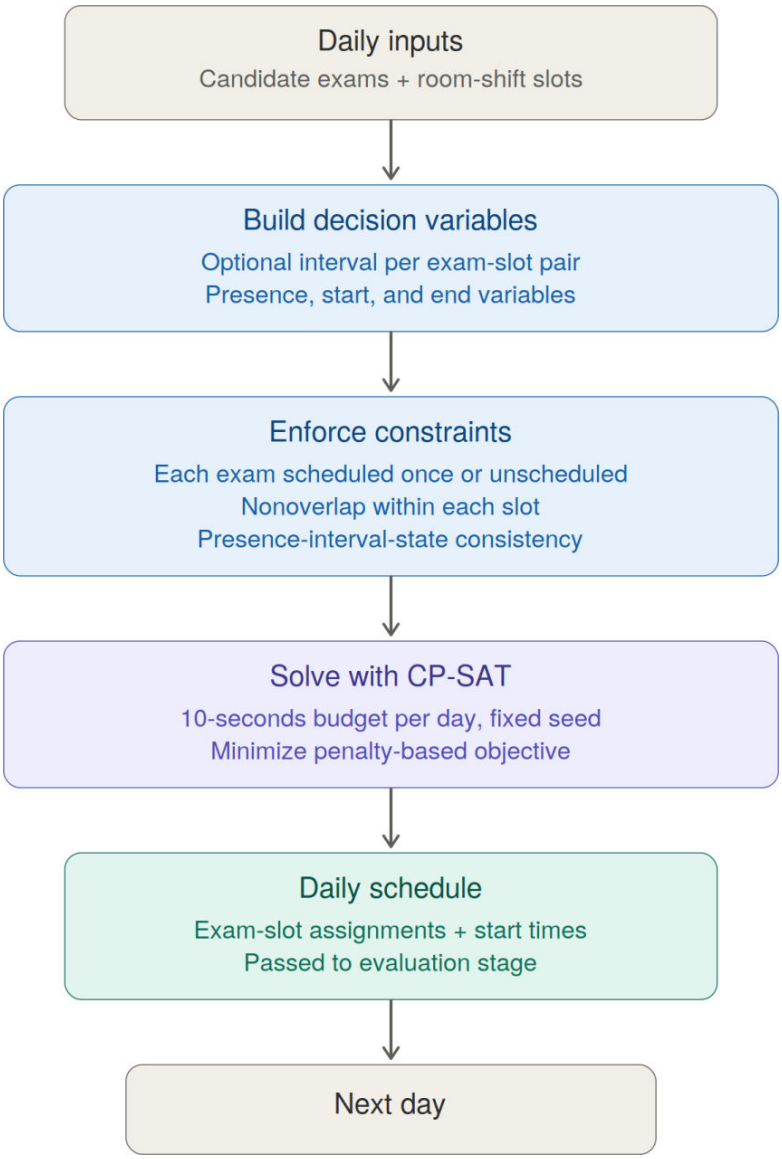
The CP-SAT approach relies on a hybrid optimization framework that combines constraint programming with Boolean satisfiability–based search. This makes it particularly suitable for scheduling problems in which logical decisions, temporal constraints, and shared resources must be handled simultaneously [23,24]. An overview of the CP-SAT scheduling approach, from the daily inputs to the resulting schedule, is presented in Figure 2.

In the present study, it is used to model the daily scheduling task as a constrained allocation and sequencing problem over the candidate examinations and the available room-shift slots. For each candidate examination, the model decides whether it is scheduled on that day or left unscheduled. For each feasible examination-slot pair, an optional interval is created, together with an associated presence variable and start and end time variables. This representation allows the solver to reason jointly about assignment, timing, and slot occupation within the daily operational horizon.

The model enforces that each examination is either scheduled once or remains unscheduled. Temporal feasibility within each slot is ensured through nonoverlap constraints, which prevent simultaneous occupation of the same operational slot by multiple examinations. Additional consistency constraints link the presence variables, the interval timing variables, and the scheduled state of each examination.

Daily optimization is performed with the OR-Tools CP-SAT solver under a fixed time budget of 10 seconds per day and a fixed random seed. Under this setting, the solver may return either a feasible solution or a proven optimal solution, depending on the complexity of the daily instance and the time required to complete the search. The output of the model includes the examinations assigned to each room-shift slot and the corresponding start times obtained from the solver, which are then used in the subsequent evaluation stage.

Figure 2. Constraint Programming Satisfiability (CP-SAT) scheduling approach.



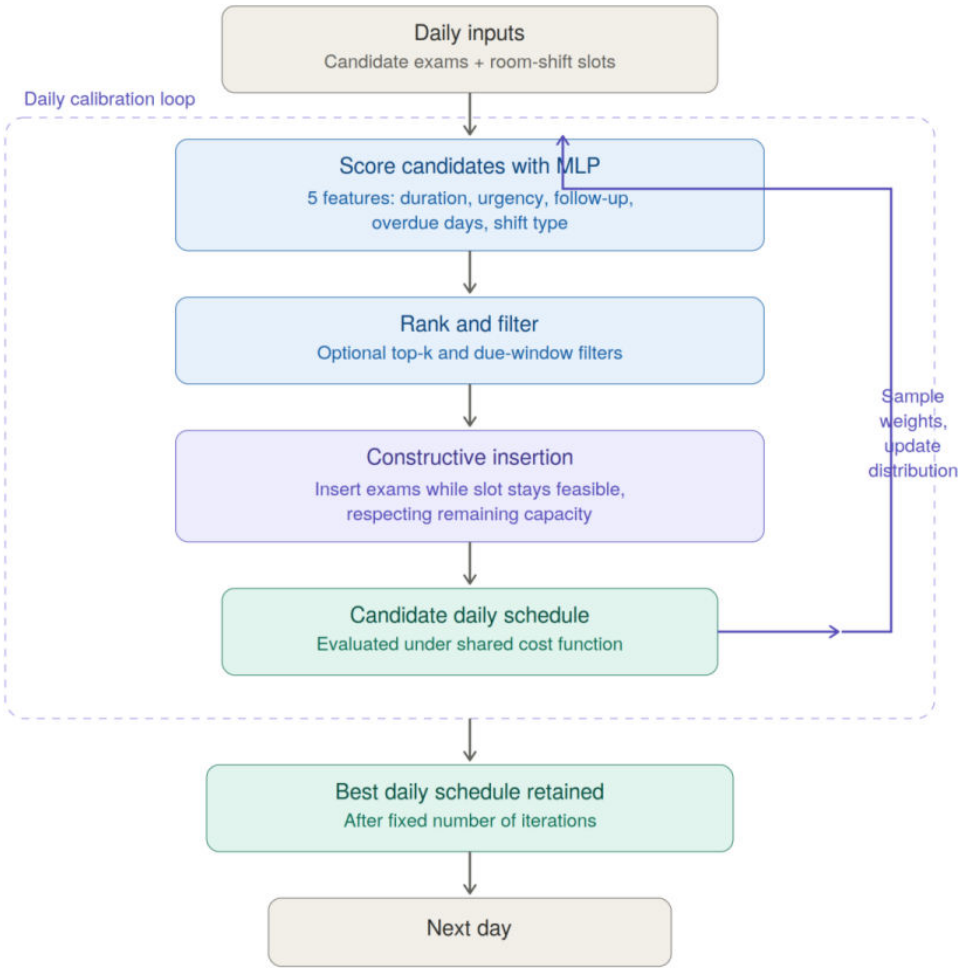
NCO-Inspired Approach

The NCO-inspired approach constructs feasible daily schedules through a score-guided constructive decoding process. It is described as NCO-inspired because it borrows the core idea of neural combinatorial optimization, using a neural scoring model to guide the constructive generation of a solution, while differing from standard NCO approaches. Conventional neural combinatorial optimization relies on a neural network that is trained offline on a large set of instances and then applied, with fixed parameters, to unseen

instances. In contrast, the present method does not use offline pretraining: a lightweight multilayer perceptron scores candidate examination-slot alternatives, and its parameters are calibrated per instance, so that the scoring model is adapted to each daily scheduling problem rather than learned once and reused. This distinction is revisited when the parameter calibration procedure is described below. The scoring model uses 5 input features: examination duration normalized in hours, urgency, days to follow-up, overdue days, and a shift-type indicator. When no follow-up date is available, a default sentinel value is used for the corresponding follow-up feature. The network architecture consists of a single hidden layer with hyperbolic tangent activation and a linear output neuron, producing a scalar score for each candidate alternative.

For each day, candidate examinations are scored and ranked for the available slots. Candidate ranking is optionally restricted by top-k and due-window filters before insertion. A constructive decoder then inserts examinations sequentially whenever slot feasibility is preserved, respecting the remaining capacity of each slot and the operational conditions defined for the day. In this way, the daily schedule is generated through a ranking-and-insertion procedure rather than through exact combinatorial search. The parameters of the scoring model are calibrated through an iterative population-based search procedure similar to the cross-entropy method [25]. At each iteration, multiple weight vectors are sampled from a normal distribution, decoded into feasible daily schedules, and evaluated under the shared daily cost function. The elite solutions are then used to update the parameter distribution, and the process is repeated for a fixed number of iterations. Parameter calibration is performed daily for each instance, not via offline pretraining. The best schedule found during this calibration loop is retained as the daily solution produced by the NCO-inspired approach. Although the calibration procedure is stochastic, in that weight vectors are randomly sampled at each iteration, all reported experiments used a fixed random seed. The NCO-inspired approach is therefore deterministic and fully reproducible for the reported runs: repeated executions under the same configuration and seed yield identical schedules. For this reason, and unlike CP-SAT, the NCO-inspired approach was not subjected to a multirun stability analysis, as repeated runs would reproduce the same results. Figure 3 summarizes the NCO-inspired approach, including the scoring of candidate examinations, the constructive insertion procedure, and the daily calibration loop through which the scoring model parameters are updated.

Figure 3. Neural combinatorial optimization (NCO)–inspired constructive scheduling approach. MLP: multilayer perceptron.



Results

For the initial head-to-head comparison, a fixed reference configuration was adopted for each approach. CP-SAT used 8 parallel search workers (num_search_workers=8) and a search time limit of 10 seconds per daily instance (max_time_in_seconds=10). The NCO-inspired approach was configured with a hidden dimension of 14 (hidden_dim=14), 30 calibration iterations (n_iter=30), a population size of 12 (pop_size=12), and an elite fraction of 0.2 (elite_frac=0.2), resulting in the retention of the 2 best-performing parameter vectors per iteration. Each method retained its respective parameter configuration throughout the 28-day horizon, while the common candidate-pool and evaluation settings remained identical across both approaches. Table 2 reports the results obtained under these reference configurations over the 4-week scheduling horizon, under the common evaluation framework developed in this study.

Table 2. General analysis of optimization algorithms performance.

Metrics	CP-SAT[a]	NCO[b]
Scheduled examinations (2460)	1400	1391
Free time (min)	3	42
Overtime (min)	1	0
Scheduled examinations mean urgency	0.7687	0.7711
Unscheduled due examinations (n=156)	0	0
Delayed scheduled examinations	0	0
Delayed unscheduled examinations	0	0
Penalty	49,790.78	83,652.10
Run time (s)	214.6	17.2

[a]CP-SAT: Constraint Programming Satisfiability.
[b]NCO: neural combinatorial optimization.

Based on the results shown in Table 2, CP-SAT achieved the best overall performance and was therefore selected for subsequent comparison with the hospital's current scheduling solution, using the same 4-week period.

Since CP-SAT achieved the best overall performance, it was subjected to a more detailed analysis to assess the stability of its results and their sensitivity to the main solver parameters. As CP-SAT incorporates randomized search strategies, a single execution may not fully characterize its behavior. It was therefore executed 10 independent times under each of four configurations, obtained by varying the per-instance search time limit (10 vs 5 s) and the number of parallel search workers (8 vs 16), while all other problem, candidate pool, and evaluation settings were held constant over the full 28-day horizon.

Table 3 summarizes the results of experiments carried out under 4 distinct configurations, with multiple independent runs. The operational outcomes were identical to those obtained under the reference configuration, with only the total penalty and the run time varying between runs. The penalty remained stable across all configurations, with only minor differences in its mean and dispersion, while the run time was determined almost entirely by the search time limit and was not significantly affected by the number of workers. This indicates that the quality of the CP-SAT schedules is robust to the randomized components of the search, and that reducing the per-instance time limit roughly halves the computational cost without significantly degrading solution quality. The complete results of the 10 independent runs for each configuration are provided in Multimedia Appendix 1.

Table 3. Constraint Programming Satisfiability (CP-SAT) penalty and run time across 10 independent runs under each of 4 solver configurations, reported as mean (SD).

Configuration	Penalty, mean (SD)	Run time, mean (SD)
10 s / 8 workers	49,782.99 (60.66)	212.7 (0.2)
10 s / 16 workers	49,879.76 (130.86)	214.9 (0.4)
5 s / 8 workers	49,879.76 (130.86)	107.5 (0.2)
5 s / 16 workers	50,172.14 (409.54)	109.9 (1.6)

Having established that the CP-SAT approach produces stable, high-quality schedules and confirmed its robustness across independent runs and parameter configurations, its best configuration was next compared against the hospital's actual scheduling practice, to assess the practical value of the proposed approach under real operating conditions. Table 4 presents the comparison between the best CP-SAT approach and the hospital's scheduling practice over the same 4-week period.

Table 4. Comparison between the Constraint Programming Satisfiability (CP-SAT) approach and the hospital scheduling solution.

Metrics	ULSSA[a]	CP-SAT
Scheduled examinations (n=2460)	1147	1400
Unscheduled due examinations (n=156)	80	0
Delayed scheduled examinations	23	0
Delayed unscheduled examinations	40	0

[a]ULSSA: Unidade Local de Saúde de Santo António.

To complement the aggregate comparison presented in Table 4, Figures 4 and 5 show the evolution of scheduling performance over the 28-day planning horizon. Figure 4 shows the evolution of the total number of unscheduled examinations, whereas Figure 5 focuses on urgent examinations only. In this visualization, urgent examinations were defined as those whose follow-up date was already overdue at the start of the planning horizon or occurred within the planning horizon itself. These visualizations allow the progression of both the implemented approaches and the hospital scheduling solution to be observed day by day across the scheduling period.

Figure 4. Daily comparison of total unscheduled examinations over the planning horizon for Constraint Programming Satisfiability (CP-SAT), the neural combinatorial optimization (NCO)–inspired approach, and the hospital scheduling solution. ULSSA: Unidade Local de Saúde de Santo António.

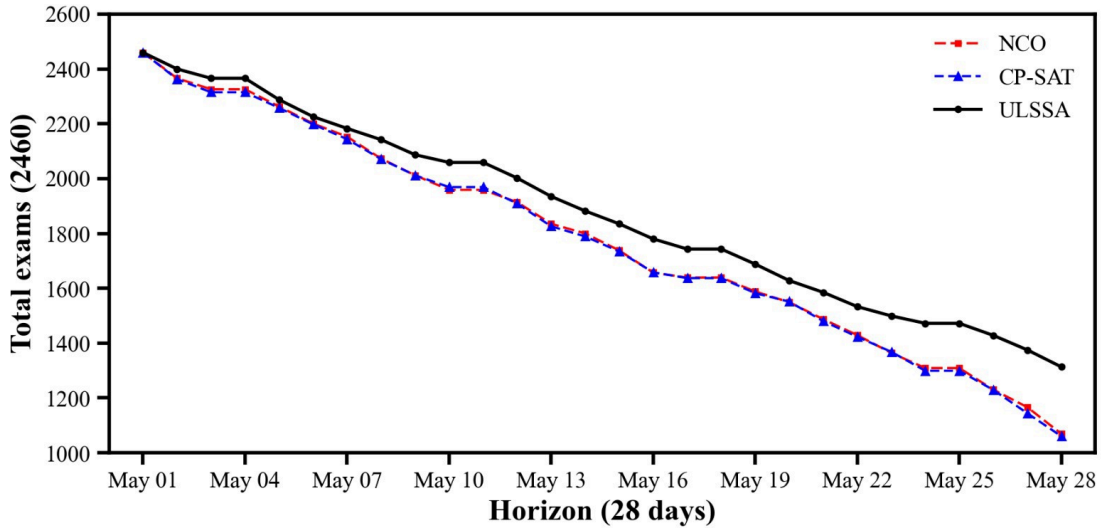
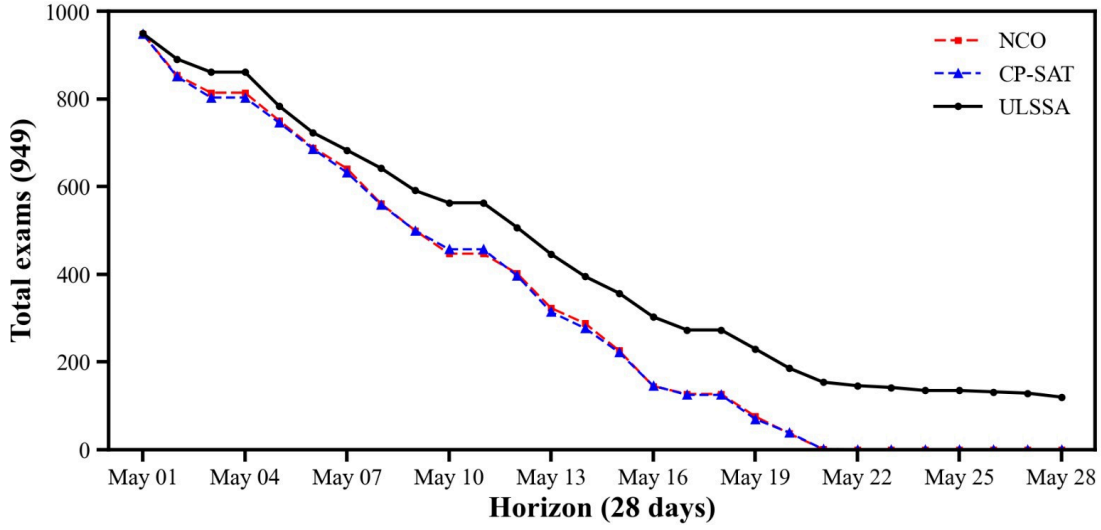


Figure 5. Daily comparison of total unscheduled urgent examinations over the planning horizon for Constraint Programming Satisfiability (CP-SAT), the neural combinatorial optimization (NCO)–inspired approach, and the hospital scheduling solution. ULSSA: Unidade Local de Saúde de Santo António.



Discussion

Principal Results

This study evaluated two optimization approaches for diagnostic examination scheduling and compared the best-performing one with the hospital's current practice. The main finding is that both approaches produced feasible schedules that eliminated follow-up-related delays, and that CP-SAT achieved the best overall performance, scheduling 1400 examinations with the lowest total penalty. When compared with the hospital scheduling solution over the same 4-week period, CP-SAT scheduled more examinations (1400 vs 1147) and removed all unscheduled due examinations (0 vs 80), delayed scheduled examinations (0 vs 23), and delayed unscheduled examinations (0 vs 40), indicating that optimization-based scheduling can meaningfully improve timely examination allocation and follow-up compliance. The robustness of these results was further supported by repeated executions under different solver configurations, in which the operational outcomes remained stable.

The first main result is that CP-SAT achieved the best overall performance in the direct comparison with the NCO-inspired approach. Although both methods performed equally well in the most critical follow-up-related indicators, with 0 unscheduled due examinations, 0 delayed scheduled examinations, and 0 delayed unscheduled examinations, CP-SAT scheduled slightly more examinations overall and produced a substantially lower total penalty. The largest practical difference between the 2 methods emerged in the temporal use of capacity. CP-SAT left only 3 minutes of free time across the evaluated horizon, whereas the NCO-inspired approach left 42 minutes. This suggests that the CP-SAT formulation was more effective in tightening the allocation of examinations within the available slots and reducing avoidable idle capacity. The NCO-inspired approach achieved a marginally higher mean urgency among scheduled exams, which may indicate a slightly stronger focus on highly urgent cases, but this difference should be interpreted with caution because it occurred together with a lower total number of scheduled exams.

The second main result is that despite the superiority of CP-SAT in schedule quality, the NCO-inspired approach required substantially less runtime. This confirms that the 2 approaches offer different practical trade-offs. In the present setting, the CP-SAT formulation was more effective in minimizing the overall penalty and improving temporal slot occupation, whereas the NCO-inspired approach produced competitive schedules with much lower computational effort.

The third main result is that the comparison with the hospital scheduling solution showed a clear advantage for the optimization-based approach in the timing-related and follow-up-related metrics that could be directly compared. CP-SAT scheduled more examinations over the same monthly horizon and, at the same time, eliminated all cases of unscheduled due examinations, delayed scheduled examinations, and delayed unscheduled examinations. By contrast, the hospital solution still exhibited nonnegligible numbers in all 3 categories. From an operational perspective, these differences are highly relevant because they reflect whether clinically time-sensitive examinations are being completed within an acceptable window. In this sense, the results suggest that the proposed optimization framework can support a more timely allocation of diagnostic examinations than the current manual or operational scheduling process.

Limitations

This study has several limitations that should be considered when interpreting the results. First, the comparison with the hospital scheduling solution was necessarily restricted to the subset of indicators that were directly comparable. Optimization-specific metrics such as free time and overtime could not be reliably compared with the hospital baseline because the real minute-level operational schedule of the hospital was not fully observable from the available records.

Second, the slot structure used in the optimization process was reconstructed from historical data rather than being extracted from a complete operational planning system. As a result, the study assumed a predefined room-shift structure and evaluated schedules under a fully usable nominal capacity model. In practice, however, the real hospital environment may include additional restrictions that were not explicitly modeled, such as variations in staff availability, parallel room usage conditions, local interruptions, unfilled periods, and operational adjustments not visible in the historical records. This limitation is particularly relevant when interpreting the higher number of scheduled examinations obtained by CP-SAT relative to the hospital solution.

A further limitation concerns the scope and validation of the evaluation. The study relied on data from a single hospital and a 4-week planning horizon, and the room-shift capacity structure was reconstructed retrospectively from the same operational records used in the evaluation, rather than being obtained from an independent planning system. No separate validation period or external dataset was used. Consequently, the results characterize the behavior of the proposed approaches in this specific setting and should not be read as evidence of generalization to other institutions, departments, or longer horizons. Evaluating the approaches on longer periods and on data from additional hospitals or diagnostic services would be needed to assess the robustness and transferability of the findings and is left for future work.

The scalability of the proposed approaches also warrants further study. The evaluation considered daily instances of a size representative of the participating hospital, and the two approaches exhibited different computational profiles, with the CP-SAT approach requiring substantially more total run time than the NCO-inspired approach over the planning horizon. For larger institutions, higher daily volumes, or more complex capacity structures, the exact search performed by CP-SAT may become more computationally demanding, whereas the constructive nature of the NCO-inspired approach may offer more favorable scaling at the cost of solution quality. A systematic assessment of how both approaches behave on larger instances was outside the scope

of this study and represents a relevant direction for future work.

Finally, the urgency score used in the study was synthetically constructed from request timing and follow-up information. Its design was intended to reflect clinically meaningful operational priorities: older pending requests accumulate urgency over time, examinations with nearer or already overdue follow-up dates are prioritized, and set follow-up events are weighted more heavily than merely probable ones. Nevertheless, this construction was a pragmatic design choice made in the absence of a standardized urgency field in the source data, and the resulting score was not validated against clinical ground truth or expert judgment. Because the scheduling objective depends substantially on this score, its unvalidated nature is a limitation, and validating or replacing it with a clinically endorsed prioritization scheme is an important direction for future work.

Comparison With Prior Work

The present study is aligned with previous research showing that health care scheduling problems can benefit from optimization-based decision support, particularly in settings where time, capacity, and priority must be jointly managed [6,7]. Much of this work has focused on surgery scheduling and has explored heuristic, stochastic, and metaheuristic approaches to improve the allocation of limited hospital resources [7,8,17,18].

The present work extends that line of research to the scheduling of diagnostic examinations and does so through a different methodological perspective. Rather than comparing only heuristic or metaheuristic strategies, this study contrasts an exact constraint-based approach, implemented through CP-SAT, with an NCO-inspired constructive approach under a shared evaluation framework. In this sense, the contribution of the article lies not only in addressing a different health care scheduling context, but also in testing more recent optimization paradigms for a day-by-day diagnostic allocation problem.

The results suggest that this extension is meaningful. As in prior hospital scheduling studies, the use of optimization improved the quality of the scheduling process when compared with the hospital solution. At the same time, the comparison between CP-SAT and the NCO-inspired approach shows that the choice of optimization paradigm can affect both schedule quality and computational behavior, even when both methods operate over the same candidate pool and the same penalty structure.

Conclusions

This study showed that both proposed scheduling approaches were capable of generating feasible daily schedules under a common penalty-based evaluation framework, but CP-SAT achieved the best overall performance in the tested setting. In particular, it produced lower total penalty, tighter temporal occupation of available capacity, and better results than the hospital scheduling solution in the timing-related and follow-up-related metrics considered in the comparison.

From a practical perspective, the results suggest that optimization-based scheduling can support more timely and operationally consistent diagnostic examination allocation in real hospital contexts. At the same time, the comparison between CP-SAT and the NCO-inspired approach highlights that schedule quality and computational effort may evolve differently across paradigms, making method choice dependent not only on penalty minimization but also on operational deployment constraints.

Future work should focus on extending the model with richer operational information, particularly real staff-related availability, parallel room usage conditions, and explicit examination-room compatibility rules. It would also be relevant to evaluate the proposed approaches over longer time horizons and under alternative overbook and capacity settings, in order to better understand their robustness and practical deployment potential.

Acknowledgments

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: (1) literature search and systematization and (2) valuation of the novelty of the research and identification of gaps.
The GenAI tool used was Claude Opus 5 (Anthropic), which was used to assist with the literature review. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.
Declaration submitted by JL and AL.

Funding

This work has been supported by Fundação para a Ciência e Tecnologia within the R&D Unit Project Scope 2024.07481.IACDC.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed per-run results of the Constraint Programming Satisfiability robustness analysis across four solver configurations. [DOCX File (Microsoft Word File), 22 KB-Multimedia Appendix 1]

References

1. Sun TQ, Medaglia R. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. Gov Inf Q. Apr 2019;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
2. Cutillo CM, Sharma KR, Foschini L, et al. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit Med. 2020;3(1):47. [doi: 10.1038/s41746-020-0254-2] [Medline: 32258429]
3. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. J Med Internet Res. Oct 5, 2022;24(10):e40238. [doi: 10.2196/40238] [Medline: 36197712]
4. Jayaratne M, Nallaperuma D, De Silva D, et al. A data integration platform for patient-centered e-healthcare and clinical decision support. Future Gener Comput Syst. Mar 2019;92:996-1008. [doi: 10.1016/j.future.2018.07.061]
5. Feldman K, Johnson RA, Chawla NV. The state of data in healthcare: path towards standardization. J Healthc Inform Res. May 2018;2(3):248-271. [doi: 10.1007/s41666-018-0019-8] [Medline: 35415409]
6. Lopes J, Guimarães T, Santos MF. Predictive and prescriptive analytics in healthcare: a survey. Procedia Comput Sci. 2020;170:1029-1034. [doi: 10.1016/j.procs.2020.03.078]
7. Fügener A, Hans EW, Kolisch R, Kortbeek N, Vanberkel PT. Master surgery scheduling with consideration of multiple downstream units. Eur J Oper Res. Nov 2014;239(1):227-236. [doi: 10.1016/j.ejor.2014.05.009]
8. Chen PS, Chen GYH, Liu LW, Zheng CP, Huang WT. Using simulation optimization to solve patient appointment scheduling and examination room assignment problems for patients undergoing ultrasound examination. Healthcare (Basel). Jan 15, 2022;10(1):164. [doi: 10.3390/healthcare10010164] [Medline: 35052327]
9. Wu Q, Ouyang Y, Li L, et al. Building an ultrasound appointment system: focusing on instrument diversity and the scope of physician expertise. BMC Med Inform Decis Mak. Aug 14, 2025;25(1):308. [doi: 10.1186/s12911-025-03150-0] [Medline: 40814005]
10. Chiou SY, Lin KJ, Dong YX. A real-time, automatic, and dynamic scheduling and control system for PET patients based on wearable sensors. Sensors (Basel). Feb 5, 2021;21(4):1104. [doi: 10.3390/s21041104] [Medline: 33562605]
11. Musa NS, Rafique SH, Mathew E. Hospital scheduling systems current challenges and emerging solutions. 2024 Global Digital Health Knowl Exch Empowerment Conf (gDigiHealthKEE). 2024:1-9. [doi: 10.1109/gDigiHealth.KEE62309.2024.10761758]
12. Knight DRT, Aakre CA, Anstine CV, et al. Artificial intelligence for patient scheduling in the real-world health care setting: a metanarrative review. Health Policy Technol. Dec 2023;12(4):100824. [doi: 10.1016/j.hlpt.2023.100824]
13. Ala A, Chen F. Appointment scheduling problem in complexity systems of the healthcare services: a comprehensive review. J Healthc Eng. 2022;2022:5819813. [doi: 10.1155/2022/5819813] [Medline: 35281532]
14. Niu T, Lei B, Guo L, et al. A review of optimization studies for system appointment scheduling. Axioms. 2024;13(1):16. [doi: 10.3390/axioms13010016]
15. Abdalkareem ZA, Amir A, Al-Betar MA, Ekhan P, Hammouri AI. Healthcare scheduling in optimization context: a review. Health Technol (Berl). 2021;11(3):445-469. [doi: 10.1007/s12553-021-00547-5] [Medline: 33868893]
16. Masmoudi M, Euchi J, Siarry P. Home healthcare routing and scheduling: operations research approaches and contemporary challenges. Ann Oper Res. 2024;343(2):701-751. [doi: 10.1007/s10479-024-06244-6]
17. Min D, Yih Y. Scheduling elective surgery under uncertainty and downstream capacity constraints. Eur J Oper Res. Nov 2010;206(3):642-652. [doi: 10.1016/j.ejor.2010.03.014]
18. Britt J, Baki MF, Azab A, Chaouch A, Li X. A stochastic hierarchical approach for the master surgical scheduling problem. Comput Ind Eng. Aug 2021;158. [doi: 10.1016/j.cie.2021.107385]
19. Zuo J, Jin Y, Liu W. Outpatient scheduling problem in smart hospital with two-agent deep reinforcement learning algorithm. Discov Comput. Nov 6, 2024;27(1):41. [doi: 10.1007/s10791-024-09474-1]
20. Valenzuela-Núñez C, Latorre-Núñez G, Troncoso-Espinosa F. Smart medical appointment scheduling: optimization, machine learning, and overbooking to enhance resource utilization. IEEE Access. 2024;12:7551-7562. [doi: 10.1109/ACCESS.2024.3349953]
21. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst. 2007;24(3):45-77. [doi: 10.2753/MIS0742-1222240302]
22. Schröer C, Kruse F, Gómez JM. A systematic literature review on applying CRISP-DM process model. Procedia Comput Sci. 2021;181:526-534. [doi: 10.1016/j.procs.2021.01.199]

23. Perron L, Didier F, Gay S. The CP-SAT-LP solver. In: Yap RHC, editor. 29th International Conference on Principles and Practice of Constraint Programming (CP 2023). Schloss Dagstuhl – Leibniz-Zentrum für Informatik; 2023. [doi: 10.4230/LIPIcs.CP.2023.3]
24. Perron L, Furnon V, Le Molgat C. How to cite OR-tools and its solvers. Google for Developers. 2024. URL: https://developers.google.com/optimization/support/cite [Accessed 2026-08-10]
25. de Boer PT, Kroese DP, Mannor S, Rubinstein RY. A tutorial on the cross-entropy method. Ann Oper Res. Feb 2005;134(1):19-67. [doi: 10.1007/s10479-005-5724-z]

Abbreviations

CP-SAT: Constraint Programming Satisfiability
CRISP-DM: Cross-Industry Standard Process for Data Mining
DSR: design science research
NCO: neural combinatorial optimization
OR-Tools: Operations Research Tools
ULSSA: Unidade Local de Saúde de Santo António

Edited by Gang Luo; peer-reviewed by Su Liu, Zeratul Izzah Mohd Yusoh; submitted 10.Apr.2026; final revised version received 02.Sep.2026; accepted 08.Sep.2026; published 20.Nov.2026

Please cite as:
Lopes J, Lobo A, Rocha F, Guimarães T, Duarte J, Santos MF
Constraint Programming and Neural-Inspired Optimization for Hospital Diagnostic Examination Scheduling: Comparative Development and Evaluation Study
JMIR AI 2026;5:e97738
URL: https://ai.jmir.org/2026/1/e97738
doi: 10.2196/97738

© João Lopes, Afonso Lobo, Fábio Rocha, Tiago Guimarães, Júlio Duarte, Manuel Filipe Santos. Originally published in JMIR AI (https://ai.jmir.org), 20.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on https://www.ai.jmir.org/, as well as this copyright and license information must be included.