

Data-Driven Approaches in Physiotherapy Clinics: Improving Clinical Outcomes and Operational Performance

AFM Tanvir Anjum^{1*}, Priyanka Patel², Azhar Uddin³, Dr Mahzabin Mahmud⁴,
Shaiful Mahmud⁵ and Israt Jahan⁶

¹MBA in Business Analytics, Gannon University, 109 University Square, Erie, USA

²Department of Public Administration, Gannon University, 109 University Square, Erie, USA

³MBA in Business Administration, Gannon University, 109 University Square, Erie, USA

⁴Consultant Sonologist, Lab4 General Hospital, Rasel Bhaban, Dhaka, Bangladesh

⁵Doctor of Business Administration, Potomac University, 1401 H Street, Washington, D.C.

⁶Department of Information Technology, Washington University of Science and Technology,
Alexandria, USA

*Corresponding author e-mail: anjum002@gannon.edu

Abstract

Missed rehabilitation appointments disrupt treatment continuity, waste limited clinical capacity, and may reduce access for patients requiring repeated therapy. Machine learning has been proposed to identify appointments for targeted reminders or flexible scheduling, yet many no-show studies rely on random resampling, outcome-proximal variables, or accuracy measures that obscure temporal drift and poor calibration. This study assessed whether routinely available scheduling data could support a deployable no-show model in a rehabilitation service including physiotherapy. We analysed 49,593 appointments recorded at the Centro de Especialização em Reabilitação, southern Brazil, from June 2016 to February 2022. A deployment-safe feature set included service, sex/gender, disability category, municipality, appointment timing, calendar variables, age, proxy lead time, and companion requirement. Post-outcome, unavailable, incomplete, or redundant variables were excluded. Penalised logistic regression, random forest, and histogram gradient boosting were trained on appointments through December 2021 and evaluated on a January-February 2022 temporal holdout using discrimination, average precision, calibration, Brier score, and decision-curve net benefit. Overall non-attendance was 9.7% (4,832/49,593), increasing from 4.0% in 2016 to 17.2% in the 2022 holdout. Physiotherapy accounted for 9,074 appointments with a 10.4% no-show rate. Penalised logistic regression performed best, but discrimination was near chance (ROC-AUC 0.506; average precision 0.163; Brier score 0.146). At a 0.20 threshold, 81.9% of appointments were flagged while capturing 86.4% of no-shows, producing negative net benefit. Therefore, the available administrative variables do not justify automated patient-level targeting. Clinics should improve data capture, consider universal attendance supports, and prospectively test risk-guided workflows before claiming clinical or financial benefit.

Keywords: Physiotherapy; rehabilitation; missed appointments; no-show; machine learning; temporal validation; operational performance; calibration; decision-curve analysis.

1. Introduction

The operational impact of no-shows is especially high in physiotherapy services. Rehabilitation is often provided in a series of "time-dependent" interactions, instead of as a one-shot consultation [1, 2]. An empty time slot thus has two potential impacts: it will take up clinician time during the scheduling process and it may break a treatment sequence for an absent individual. The second effect is clinically plausible but should not be overstated. Not attending an appointment does not necessarily constitute a measured deterioration, nor does attendance serve as a proxy for recovery. The problem, however, is real: there is limited treatment capacity in a clinic and a limited number of individuals with disabilities or complex rehabilitation needs [3].

In the present paper, this problem is analysed from a data source of an outpatient rehabilitation service (Centro de Especialização em Reabilitação, CER) associated with the Unified Health System (SUS) of the Brazilian health care system. This data includes a total of 49,593 appointments for physiotherapy, speech therapy, psychotherapy, occupational therapy, and other rehabilitation between 2016 and 2022. Therefore, it would be more appropriate to describe this as a rehabilitation-service study with a physiotherapy-relevant subgroup rather than as a physiotherapy study. 9,074 appointments are covered by physiotherapy (18.3%). This is particularly important when considering appointment behaviour, referral pathways and appointment length, as this can differ between services and a pooled model might not be easily transferable to a stand-alone physiotherapy service [4]. The appeal of machine learning is simple. If there is a small group that can be identified, the service may provide a short-notice wait-list contact, transport check, tele-rehabilitation alternative, easy cancellation, or respectful reminder. These interventions should provide choices, not recrimination. Harmful reactions, on the other hand, are automatic cancellation, double-booking without patient permission, or prioritisation of people's social factors as risk. A risk score is only useful if it enhances a clear, fair process or procedure following the evaluation of the cost of a false alert. The no-show literature provides an alarming, but also encouraging, example. Decision-tree methods and prior attendance variables have been reported to be often influential in systematic reviews, and models have been created in heterogeneous clinics with varying outcome definitions and base rates [5, 6]. In more recent literature, there is a call for increased attention to imbalance and to repeated validation. However, a model can be performed well in one calendar period and fail in the next due to a change in scheduling, patient population, services offered or local conditions. Random split shuffles early and late appointments, enabling the model to borrow future distribution that it will never encounter during deployment. It is NOT an acceptable substitute for a forward temporal test if the service will utilise the model in the future.

Another challenge is the timing of information. There are some no-show reasons in the CER file and weather data for the appointment day. Both can create the illusion of more potency when predicting outcomes in retrospect, but neither should be assumed to be available at the time of booking. A written absence reason must be post-outcome in nature. For a future appointment, real weather data is not known at the time of the appointment, although it could be a forecast in a separately specified pre-appointment model. If such variables are included in a booking-time classifier, this creates leakage: the classifier gets information causally

downstream of the event to which it is asked to predict. The resulting performance would be overly optimistic and operationally misleading in terms of operations [7-10].

The research question here is deliberately limited in scope: Do booking and scheduling time characteristics of this administrative dataset predict future non-attendance sufficiently to guide targeted outreach efforts in a rehabilitation service? A related, but distinct, question is how attendance relates to the clinical outcome. The FYSIOPRIM supplementary analysis and the associated Norwegian cohort report show that it is possible to examine treatment contacts in conjunction with prognosis, perceived improvement, future service costs and work disability. Patient-reported outcome measures, pain, function, discharge status, costs, and a patient identifier are not included in the CER. It cannot determine if an attendance intervention works to aid recovery. This absence is not seen as a minor setback, but rather the main point of the paper's interpretation. The aims were: to describe non-attendance patterns of the CER cohort, with a particular focus on the physiotherapy cohort; to design three intentionally limited prediction models without using data collected outside of deployment; to assess temporal generalisation in 2022 using discrimination, calibration and a net benefit approach rather than accuracy; and to arrive at a cautious clinic-data strategy. The contribution does not discuss the deployment of machine learning. It is an empirical test of whether the available data are sufficient for the claim [4, 11, 12].

2. Literature review and conceptual rationale

In recent literature, missed appointments have been discussed as an operational support tool by means of machine learning and not as an end in itself. In physiotherapy and other outpatient settings, where predictive performance is apparent, there are four interconnected questions to be considered: Do there exist predictors at the decision point? Do estimates carry over to another setting? Is the score calibrated and useful? Does the score facilitate equitable supportive care? The review thus juxtaposes the reported promise with methodological and ethical limitations and motivates the present temporal-validation study.

Huang, Li [13], [14], Nagib, Pervez [15] suggest that if physiotherapy is provided as a series of visits, missed appointments may have material consequences and cost implications. The outpatient setting where the figures were reported, however, is from Nigeria and is significantly higher than the 10.4% (physiotherapy) rate in CER. This difference does not constitute evidence that one service is better managed: it can be a difference in payment system, referral patterns, location, definitions of appointments, and amount of time observed. The counterargument is that non-attendance is context-dependent. It can be a reflection of transport barriers, changing health, communication issues, carer availability or the mismatch between the planned services and the person's priorities. A model can explain patterns in these circumstances but not mechanisms nor justify a sanction against a person deemed "high risk". The no-show literature often shows that prior attendance and decision-tree approaches tend to be influential, according to Collins, Reitsma [16] and Ahmed, Zhinuk [17]. Predicting from administrative data systems is useful in some settings, as illustrated in their systematic review. But at the same time, the same review suggests a strong case against broad generalisation: studies vary with respect to target definition, forecast horizon, predictor timing, resampling,

and performance reporting. A published "best model" is thus a model validated within one information system, rather than a portable clinical fact. In a physiotherapy clinic, a base rate different from that at the clinic location may result in a different number of false alarms and workload given the same probability threshold [6, 18].

Dantas, Fleck [19] and Rahaman, Ferdousy [20] argue that decision analytic assessment is superior to sensitivity reporting, as it renders the cost of false-positive contact more explicit. This is a significant step forward from technical studies which are often understood as an end to discrimination. Decision analysis, however, is not a panacea for models that are not documented as temporally transportable and/or well-calibrated. A high-sensitivity classifier may send out nearly every appointment reminder and thus mimic a universal-reminder approach but without meaningfully targeting avoidable non-attendance. The question is not whether an algorithm generates a score, but whether it alters a workflow (which is already constrained) enough to warrant the administrative and patient burden. While machine-learning techniques do not generally outperform logistic regression after methodological quality is accounted for, they do perform better when the latter is applied to a dataset that has been aggregated from several independent studies, as with the WHO model [21-25]. Random forest and gradient boosting can capture nonlinearities and interactions, but these are not necessarily beneficial when predictors are sparse, unstable, or mistimed. The other side of an algorithm-first argument is that the quality of information and decision-time alignment can be more important than the choice of a classifier. In the CER data, there is no stable patient identifier to create a prior attendance, repeated cancellation, treatment sequence, or clinician-continuity feature; this limitation is especially important in the CER data [12].

In a previous, smaller evaluation that was also in the context of CER, machine learning had promising results, as reported by Davis and Goadrich [26]. However, that result is of local interest only and is not enough to answer the current deployment question. A model can be good for learning signals that are only present during service or overlap with random splits, or variables that are available only at the scheduling decision. The counterargument is not just statistical but practical: a clinic that pursues an optimistic retrospective model could be changing staff focus to low-yield contacts while neglecting the people who truly have avoidable barriers to care.

Davis, Lasko [27] and Pervez, Nova [28] found that clear reporting is needed for intended use, outcome timing, predictor assessment, missing data, validation design, and implementation context for TRIPOD+AI. Essery, Geraghty [29] expand this analysis to include risk of bias and applicability in their PROBAST+AI. Such frameworks challenge the view that the best clinical model is the highest AUC. The counterargument is that a model could perform well in terms of ranking appointments without doing a good job of absolute probabilities or different error types or without being any better than contacting everyone. This is why transparent prediction reporting is required but not enough evidence to implement. Hasvold and Wootton [30] state that calibration is a central theme because probability estimates are employed to make decisions, rather than just to rank observations. Similarly, Huang et al. [13] explain that errors that are uncovered by the calibration intercepts, slopes, and plots are not detectable by AUC alone. The downside to relying on accuracy is particularly compelling under imbalance Sozib,

Islam [31]. In the 2022 CER holdout, correctly predicting attendance for all appointments was 82.8% accurate, while finding no no-shows. This implies that the average precision, calibration, Brier score, and threshold-specific workload are more informative than a headline accuracy score. Huang and Hanauer [32] affirms that decision-curve analysis is the comparison of the net benefit of a model with other simple alternatives like contacting everyone or nobody [33]. However, net benefit is not an intervention effect; it is based on a threshold trade-off between beneficial and non-beneficial contacts but does not demonstrate that a reminder, transport check, or rescheduling actually results in any change in attendance. The objection is not only of a methodological nature but also of an ethical nature. A model can be used to determine that there is a social pattern of risk, but it does not demonstrate that the action that is being recommended is acceptable, equitable, affordable, and useful to the people being recommended. The conceptually circumspect position of this paper is thus deliberately limited. The booking-time demographics and scheduling fields are treated as proxies for operational frictions rather than as patient characteristics that cannot be changed. A score can only provide staff support to a human-reviewed offer of assistance when staff capacity, patient safeguards and local evaluation support such action. It cannot justify automatic cancellation, overbooking, reprioritization or claim that a recorded characteristic will change to improve attendance [25, 34].

Kelly, Karthikesalingam [35] and Pervez, Nova [28] found that the greater the physiotherapists' contact frequency, the higher the likelihood of the patients' perceived improvement, and the greater their likelihood of improvement, the lower their future cost and disability pension odds were among patients with poor prognoses. This helps to highlight the importance of making linkages between operational and clinical information. Results from the authors' observation, however, do not rule out confounding by indication for treatment, access to treatment, engagement, and severity of the missed appointment in predicting recovery. There are no PROMs, patient identifiers, treatment exposures, discharge statuses, or cost outcomes in the CER analysis, and the counterargument is decisive. Operational efficiency cannot substitute for better health. As a combined result, the literature provides a conservative research contribution: evaluate the potential for a patient-supportive, prospective-like workflow with the available data before assuming that machine learning should be used. The real issue is not which model is the most technologically advanced; it is whether a locally validated model generates actionable, calibrated, equitable information that is better than a transparent non-algorithmic alternative. If a study in a future physiotherapy clinic is to be undertaken, the designed response is to connect appointment information with a longitudinal attendance record, PROMs, treatment exposure, discharge, and cost outcomes; to detail the time of prediction before modelling; and to compare supportive actions with universal care pathways prospectively. Without these elements, associations should not be treated as causal, and model scores should not be considered as evidence of clinical benefit but as decision-supporting evidence [33].

3. Materials and methods

3.1 Design, setting, and data provenance

This was an appointment-level retrospective prediction study and was reported in the spirit of TRIPOD+AI. The data file was provided by the researcher and is the Medical Appointments No-Show dataset made available by Salazar [36] (DOI: 10.17632/wm6w2fvkfj.1). The repository provides descriptions of the de-identified records from CER at the University of Vale do Itajai, southern Brazil, in a public outpatient rehabilitation environment that serves individuals with physical and intellectual disabilities. There are 49,593 appointments in a file, from 20 June 2016 to 25 February 2022. The entire file was retained, including all rows containing a valid appointment date and an observed no-show indicator [37].

The unit of analysis was an individual scheduled appointment, rather than a unique patient or episode of treatment. This decision is based on the operational question of whether the upcoming slot will be unused or not and has implications. There is no patient identifier in the file, so repeat appointments cannot be linked. The nominal sample size may then be inflated as a result of the within-person correlation of observations, and observations within individuals may not be independent. More significantly, there are many predictors that are useful in many patients but cannot be built, for example, a patient's prior attendance pattern, cumulative treatment exposure, prior cancellations, or time since last visit. This is a major data limitation and NOT a pre-processing detail. The main outcome was the binary field noshow, yes/no. According to the data documentation, it is the patient's attendance at the appointment. There is no distinction between cancelled and rescheduled appointments, clinician-cancelled appointments, late appointments, partially attended appointments or an audit of outcome coding. The analysis therefore uses the repository definition but does not claim a clinically richer adherence measure. The prevalence observed across the entire cohort was 9.7%, 9.6% for development data, and 17.2% for the 2022 holdout data [38].

Table 1. Dataset provenance, data quality, and pre-specified analytical decisions

Data element	Observed value	Analytical decision
Appointments	49,593	All rows retained; appointment-level unit of analysis
Outcome	4,832 no-shows (9.7%)	Binary scheduled no-show indicator
Holdout period	942 appointments; 17.2% no-show	Strict temporal test: Jan-Feb 2022
Development period	48,651 appointments; 9.6% no-show	June 2016-Dec 2021
Candidate predictors	14 fields engineered from 7 categorical and 7 numeric inputs	Booking/scheduling information only

Excluded leakage variables	No-show reason; observed-day weather; ICD; duplicate age/date fields	Unavailable, post-outcome, sparse, or redundant at intended decision time
High-missing candidate variables	ICD 78.4%; date of birth/age 20.9%; specialty 15.0%	Imputation only for retained predictors; no complete-case analysis
Patient identifier	Not supplied	Repeated appointments cannot be clustered; uncertainty may be understated

Source: Authors' analysis of the supplied CER Medical Appointments No-Show dataset [23, 36].

The table shows the reason this is considered an appointment-level as opposed to a patient-level prediction task. Clustered inference and features that require a patient identifier to find out if the patient had been previously attended to are not possible without it, and the 2022 higher outcome rates require a temporal holdout forward, not backwards [39].

3.2 Predictor timing, feature engineering, and leakage control

The predictors were chosen based on their likely availability at the intended decision time – the appointment booking/scheduling stage. Speciality, gender, disability, city, appointment shift, weekday and appointment month categorical fields were retained. Selected numeric fields were age, a proxy for scheduling lead time (appointment date minus service entry date), appointment hour, appointment week, day of month, calendar year index, and companion requirement. The date of the appointment was parsed as a day-first date. The scheduled date/time was used to calculate weekday, month, hour, day of month and week of year. The lead times of less than 0 days and more than 730 days were set to missing due to being incompatible with a simple booking interval and may have been date-quality issues or events not captured by the first entry date [40].

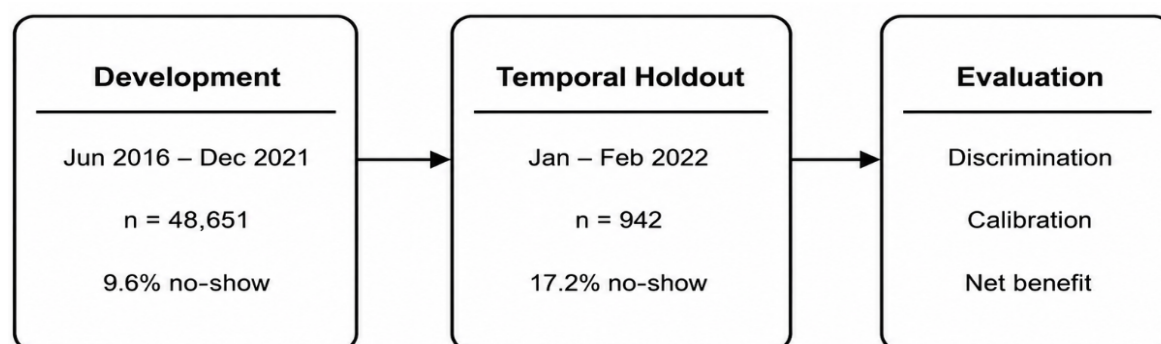


Figure 1. Temporal validation protocol for the appointment-level prediction study. The later period remains isolated until final evaluation.

The flow diagram is not intended to imply that the resulting model is deployable, but rather as a warning against being overly optimistic about it. Using appointments until December 2021 and keeping the data from the January-February 2022 time for a forward test ensures that

information from the future time does not affect feature preparation, tuning, or model selection. Each eligible appointment, however, does not necessarily give rise to independent observations because of the lack of a patient identifier; repeated visits may be correlated, and because of the nominal sample size, it may appear to give more independent predictive information than it does [4, 41]. The difference in the prevalence of outcomes between the development and holdout groups (9.6% vs. 17.2%) does not imply that there was any change in patient behaviour that caused the outcomes to shift. The holdout period also only includes two months and 162 no-shows, making it difficult to determine if the same trend would continue with other seasons or with other conditions. The diagram is therefore a reasonable test for the future, not an endorsement of the model or that a future outreach policy would be effective.

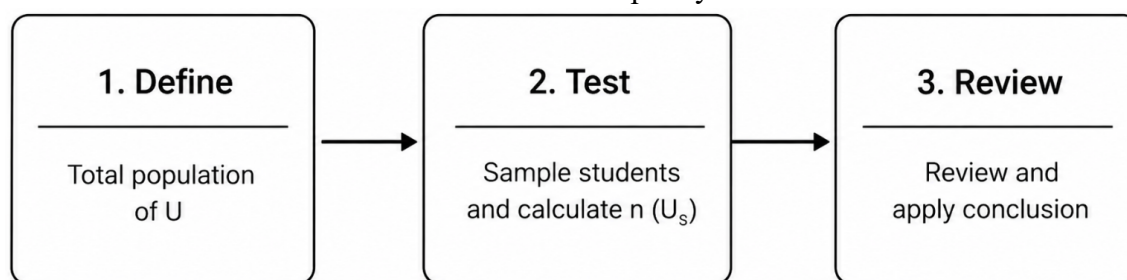


Figure 2. Responsible machine-learning decision architecture. The score may trigger supportive human review, not automatic action.

The pathway is explicit, separating prediction from action only information that is plausibly present before the booking decision goes into the model is relevant, and a score can at most trigger an offer of additional support. Obviously, one of the reasons for the restriction is that a no-show reason is a downstream variable or is not recorded when the appointment is made and adding them would result in a model that is operationally invalid and may be predictive but is operationally unavailable. The pathway is a governance plan and not a causal model for intervention. It does not indicate that attendance will change because of a reminder, transport check, or rescheduling offer and does not specify which patients will benefit. The use of the score should be checked by staff, limited by support available and monitored for any unequal distribution of false positive burden. It should never be a way to automatically cancel or overbook patients or to deprioritise them, where access barriers are embedded in administrative data [42-45].

3.3 Prediction target, candidate models, and mathematical specification

The binary variable Y_i equals 1 if appointment i is reported as a no-show and 0 if the appointment is not a no-show. The risk π_i for each model is estimated, given the booking-time feature vector x_i . The random forest model and the histogram gradient boosting model are used to test the hypothesis that nonlinear patterns and interactions are useful for better generalisation, while the logistic model is the clear base model. The hyperparameters were intentionally limited rather than exhaustively explored: logistic regression L2 regularisation ($C = 0.3$); random forest with 350 trees, maximum depth 16, and minimum leaf size 12; and histogram gradient boosting with 300 iterations, learning rate 0.055, 15 leaves, minimum leaf

size 35, and L2 penalty 1.0. It was an attempt to make a stable comparison and not optimise against the holdout [46].

The penalised logistic regression can be expressed as a logit link function with the penalisation term being the magnitude of the coefficients. This is a good reference model because it is easy to interpret and less prone to capturing transitory peculiarities in administrative data. Random forest takes the average of predictions from decorrelated decision trees; gradient boosting fits a series of shallow decision trees to decrease the residual loss. Both algorithms are compared without assuming superiority, unless the researcher knows otherwise. They need to be evaluated on upcoming appointments, rather than be assumed from training fits. The major evaluation was a holdout test wherein all appointments for June 2016-December 2021 ($n = 48,651$) were designated development data, and all appointments in January-February 2022 ($n = 942$) were designated as untouched test data. This test is more difficult than a random split, as prevalence rose rapidly, and 2022 is not included in the development. This design strengthens deployment relevance but reduces precision – just 162 no-shows were seen in the holdout. The uncertainty of ROC-AUC and average precision is quantified using 95% confidence intervals (500 resamples). No class resampling was used. Synthetic oversampling can enhance the sensitivity of the development process but may lead to a distortion of the probability calibration, especially if the desired action is based on absolute risk. The study instead provides measures that are relevant to imbalance and does not perform post-hoc optimisation of thresholds. Operational thresholds described in the operational table (0.05-0.30) do not represent recommended deployment thresholds.

3.4 Performance evaluation, operational utility, and critical safeguards

ROC-AUC and average precision (AP) were used as measures of discrimination. AP summarises precision over recall values and is sensitive to the low event proportion, and ROC-AUC is the probability that a randomly selected no-show gets a higher score than a randomly selected attendance. The prevalence of no-shows in the holdout is 0.172, so the resulting AP has limited meaning if below or close to this value. The following metrics were not considered primary measures, as they can mask a low-utility model with imbalanced data: accuracy, F1 score and confusion matrix. Decision-curve analysis was used to evaluate clinical utility. At threshold t a false positive carries a weight of $t/(1-t)$, reflecting the relative harm/cost to the organisation of a non-beneficial outreach contact versus a successfully targeted no-show. Two policies are compared to the model: treat all appointments and treat none. 'Treat' does not mean medical treatment, nor does it mean a punitive action, but it does mean supportive attendance outreach. Prior to assigning implementation resources to a model, a clinic needs to exceed both policies across a reasonable range of thresholds.

3.5 Ethics, transparency, and reproducibility

The data were a de-identified, publicly deposited dataset and were licensed CC BY 4.0 by the data repository, thus serving as a secondary analysis. It did not involve contact with participants or re-identification. However, it is important to note that local institutional requirements should be checked prior to submission, as wording and analysis of health-service records may require

an exemption statement. Analysis is reproducible from the provided CSV file, and the workflow (using Python) to create features, fit the three pipelines, and evaluate the temporal holdout is provided and is auditable. No patient-level raw data are reproduced in this manuscript [47].

4. Results

4.1 Cohort and descriptive operational pattern

This included a total of 49,593 appointments, with 4,832 of those being "no-shows". Physiotherapy represented 9,074 appointments (18.3%) with 941 no-shows (10.4%). The highest-volume services were speech therapy (11,877 appointments) and psychotherapy (11,675 appointments). The 'sem especialidade' category contained only 57 appointments, with a 22.8% no-show rate, which was insufficient for an operational conclusion. The percentage of appointments with missing specialty data (15.0%) and the 12.1% no-show rate for missing specialties highlight that the completeness of data can be useful operationally but is not an inherent attribute of patients.

The gender field had seven records coded 'I', which had a no-show rate of 0.0%, with intellectual and motor disabilities having no-show rates of 8.9% and 9.6%, respectively – this may be due to documentation rather than patient-specific factors. The observed no-show rate was lower for appointments with a companion (9.0%) as compared to those without a companion (10.7%). All such differences are descriptive and unadjusted and susceptible to the confounding of the service composition and of temporal differences. They should not be interpreted as a patient-level presumption. Time was a striking pattern. Annual no-show prevalence rose from 4.0% in 2016 to 8.3% in 2017, 8.0% in 2018, 12.3% in 2019, 13.9% in 2020, 15.1% in 2021, and 17.2% in the January-February 2022 holdout. The number of appointments also varied significantly during these periods. The increase does not imply causation but rather implies a distribution shift – a model that performed well for the earlier (narrower prevalence) service environment must generalise to a materially different outcome frequency [48].

Table 2. Appointment-level descriptive characteristics and unadjusted no-show prevalence

Characteristic	Category	Appointments	No-shows	Rate
Specialty	assist	194	13	6.7%
Specialty	enf	712	43	6.0%
Specialty	occupational therapy	6486	555	8.6%
Specialty	pedagogo	2064	154	7.5%
Specialty	physiotherapy	9074	941	10.4%

Specialty	psychotherapy	11675	1091	9.3%
Specialty	sem especialidade	57	13	22.8%
Specialty	speech therapy	11877	1121	9.4%
Specialty	Missing	7454	901	12.1%
Gender	F	12003	1240	10.3%
Gender	I	7	0	0.0%
Gender	M	37583	3592	9.6%
Disability		9	4	44.4%
Disability	intellectual	30882	2742	8.9%
Disability	motor	13565	1308	9.6%
Disability	Missing	5137	778	15.1%
Appointment shift	afternoon	27865	2762	9.9%
Appointment shift	morning	21728	2070	9.5%
Age group	0-11	21244	1870	8.8%
Age group	12-17	10206	851	8.3%
Age group	18-29	1480	136	9.2%
Age group	30-44	994	95	9.6%
Age group	45-59	1759	186	10.6%
Age group	60+	3560	356	10.0%
Age group	Missing	10350	1338	12.9%
Companion required	0	22603	2416	10.7%
Companion required	1	26990	2416	9.0%

Source: Authors' analysis of the supplied CER Medical Appointments No-Show dataset [36, 49].

The table does not provide a cause-and-effect relationship. It reveals that apparent differences between specialties, gender coding, disability recording, shifts, age and companion

requirements exist alongside gaps and shifts in data over the service period; none of these should be considered a measure of patient propensity or a ground for punitive scheduling.

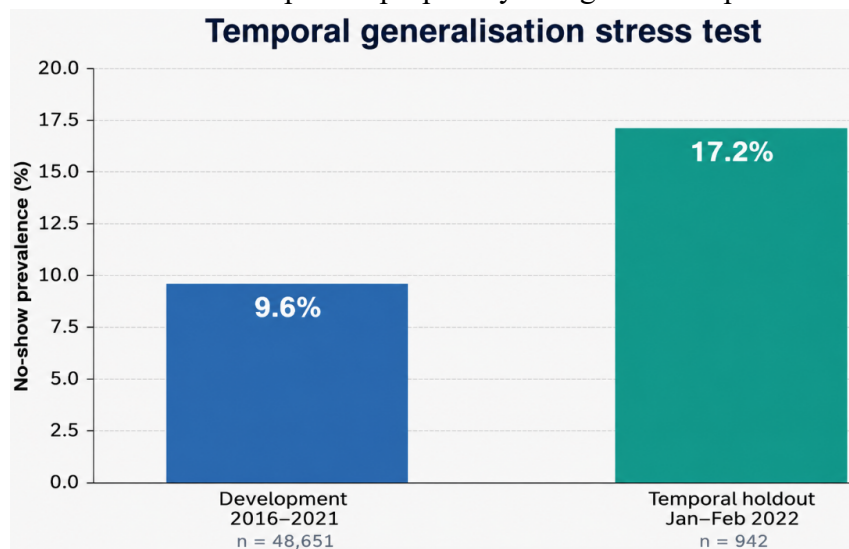


Figure 3. Temporal generalisation stress test. The holdout has a materially higher non-attendance prevalence than the development window.

The stacked bars depict two related analytical issues: no-shows are a small proportion of cases, and their share increases rapidly in the future-period information. In this situation, a classifier used to predict attendance for all appointments would appear accurate while missing every no-show. The goal of accuracy would then benefit the dominant class and not be an effective example of selective outreach. The difference in prevalence is also a warning about transportability. It could be due to changes in delivery, referrals, case mix, documentation or broader social factors; the visual does not differentiate between these factors. These later records would be spread across both samples if split at random, thus hiding the operational problem encountered during deployment. The result supports temporal validation but cannot provide any conclusive evidence of a drift mechanism.

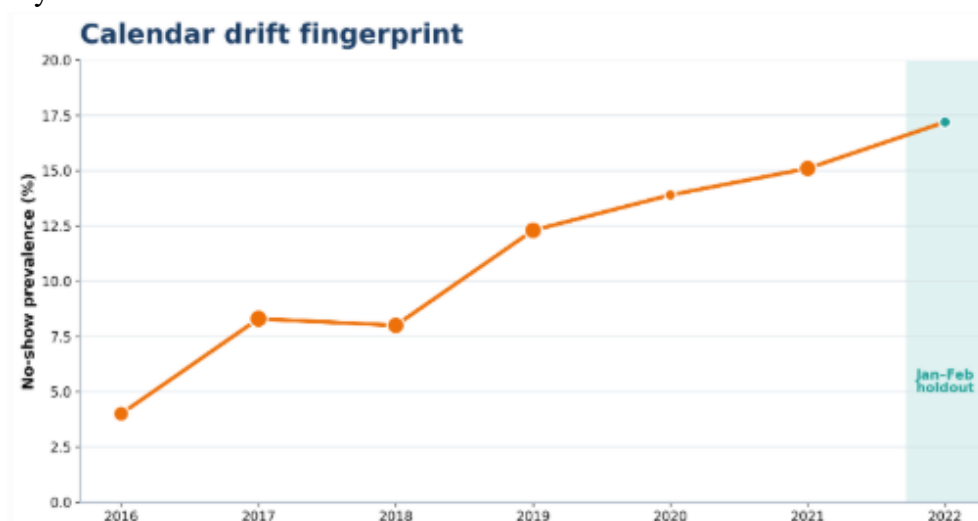


Figure 4. Calendar drift fingerprint. Annual no-show prevalence is plotted over time; marker area reflects appointment volume.

The annual trajectory shows that the rate of non-attendance was not constant over the time observed. This undermines the assumption that associations learned in early years can be transferred to later schedules without modification. The trend is particularly significant, as a probability threshold has different workload and false-alert impacts if the event rate changes. This line should not be regarded as a causal trend. All these factors can have changed at the same time, and the latest period consists of only January and February. Local structural change may then be captured by calendar year instead of as a stable component of a patient's risk. Any model should be monitored and adjusted over time, not because it is a “historical year effect” that has clinical significance [50].

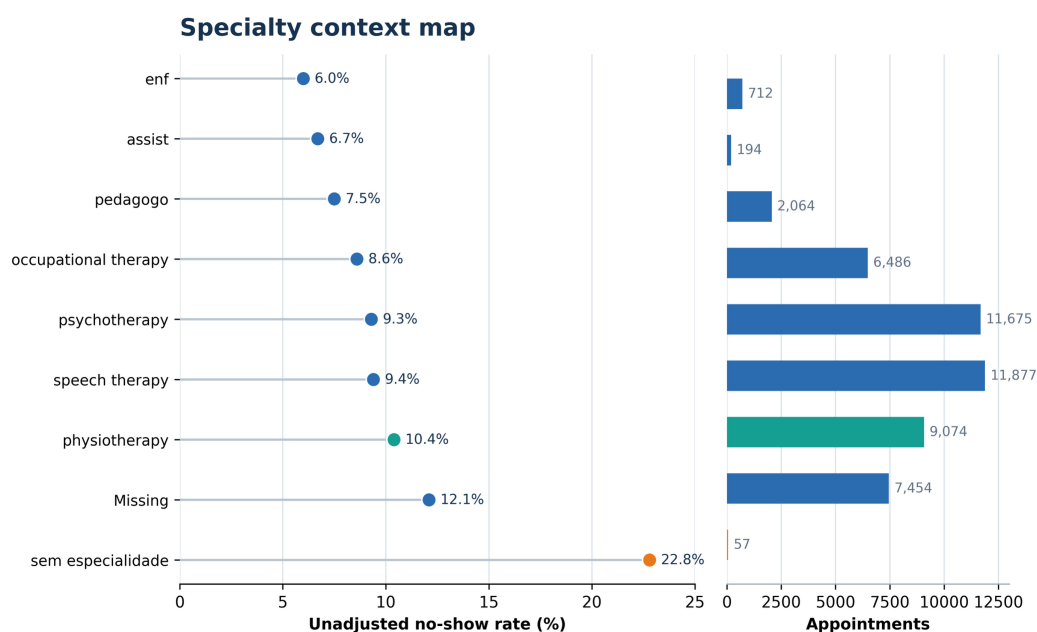


Figure 5. Specialty context map. The rate is interpreted jointly with appointment volume; physiotherapy is highlighted.

The plot refers to the heterogeneity between rehabilitation specialties, but it is not a behavioural mechanism for the specialty. The highest rate category only has 57 appointments and would be very sensitive to the number of events. The physiotherapy estimate is more accurate because the denominator is larger but is still expressed as an appointment-level proportion and not as an individual patient risk estimate [51, 52]. Service labels may also represent variations in appointment length, referral streams, availability of providers, disability mix and administration recording. There is an operational context, and the sizeable missing-speciality group indicates that data completeness is part of the operational context. If such rates were considered inherent features of the service or ways of scheduling resources without modifying them, the risk would be that they would be confused with documentation and composition effects as well as with causes of non-attendance that are modifiable [53].

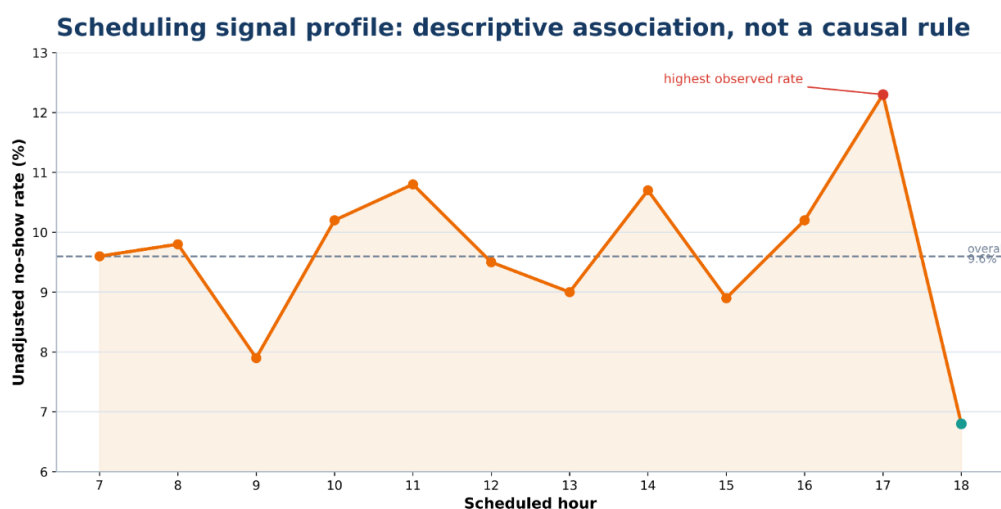


Figure 6. Scheduling signal profile. The observed hour pattern is descriptive and is not a causal scheduling rule

The hour-by-hour difference is relatively small and is presented as descriptive scheduling information and not as evidence that the timing will result in fewer absences. An appointment hour is probably bound up with speciality, transport arrangements, clinician availability, patient routine and calendar period. These sources of confounding cannot be resolved by the plotted proportions [54]. If this is the only pattern returned by the service, it may end up shifting the burden to patients who have even fewer choices. The clinic would have to have a prospective operational evaluation that assesses attendance, cancellation lead time, patient acceptance, and impacts by relevant access groups before rearranging timetables. The visual can be used to create an audit question but not necessarily for a causal scheduling recommendation.

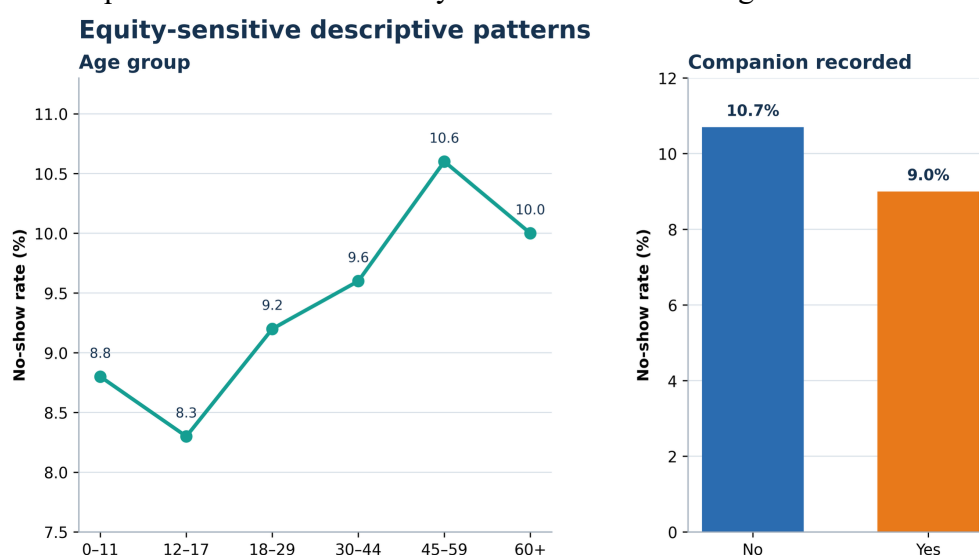


Figure 7. Equity-sensitive descriptive panel for age and companion requirement. Observed associations require supportive, not punitive, interpretation.

A lower no-show rate was observed when the panel recorded a companion as required for the appointment. This is not a device to make judgements on motivation, reliability or personal

propensity to attend. The companion requirement may be a result of accessibility issues, documentation, disability-related needs, referral patterns or staff expectations of disability-related issues, and not necessarily a direct impact on attendance. The age pattern is also constrained by the categorization and exclusion of age-unknown records from the plot. Grouped rates may mask within-group variation and are likely to be misleading if there is a difference in service mix or calendar period. These variables could be used for fairness monitoring or for providing supportive contact but should not be used for rationing care or as a modifiable causal explanation without prospective evidence [55].

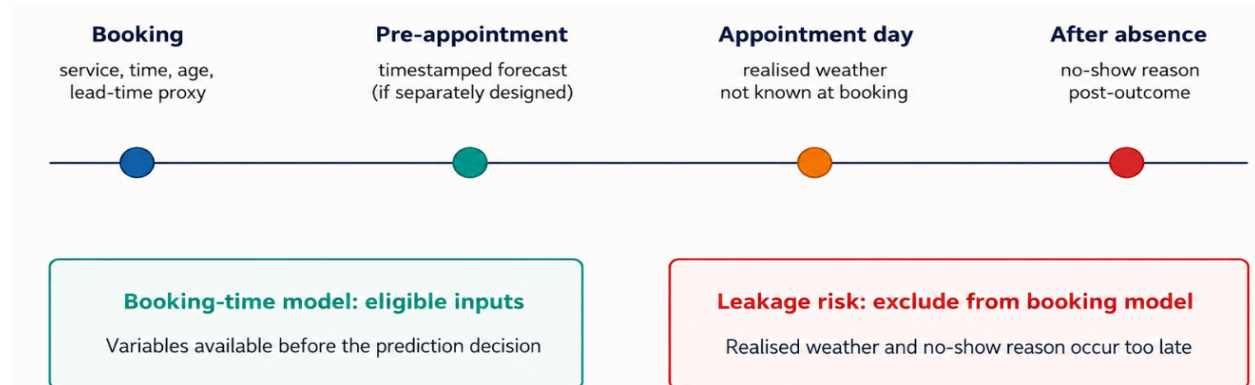


Figure 8. Predictor-timing and leakage audit. Realised weather and no-show reason are excluded from the booking-time model.

The observed-weather pattern is deliberately descriptive. Weather data did not exist at the time the appointment was booked but was recorded on appointment day. If it is used in the booking-time model, it would be hard to distinguish the prediction time from the actual scheduling context and would introduce information unavailable to a real scheduler [56].

4.2 Temporal model performance

The performance of the 942-appointment temporal holdout is reported in Table 3. Penalised logistic regression was the least poor model, with ROC-AUC 0.506 (95% CI 0.464-0.549), AP 0.163 (95% CI 0.137-0.193), and Brier score 0.146. The AP was slightly lower than the holdout prevalence (0.172) and had a confidence interval that included the holdout prevalence. Random forest had ROC-AUC 0.464 and AP 0.159. Histogram gradient boosting was even worse, with ROC-AUC 0.395 and AP 0.150. The flexible models were not able to find any meaningful pattern in the chosen predictors. The estimates are not to be taken to imply clinical acceptability of logistic regression. The apparent differences in ranking ability is within uncertainty, since the ROC-AUC is close to 0.50. This is more informative than a high in-sample score, as it is describing the performance of appointments that were made after the development period. It suggests that the previous years' association structure was not robust enough to rank 2022 no-shows. The ROC curves are grouped together close to the diagonal in figure 9. Precision is near the base rate of 17.2% across most of the recall range, as shown in figure 10. The first set of high-precision values are based on very few observations and should not be regarded as a viable high-risk group. An operational clinic requires a reasonably large number of enriched appointments; a single outlier score from a few appointments is no justification for a workflow for clinic staff [57].

Table 3. Temporal-holdout discrimination and overall accuracy of predicted probabilities

Model	ROC-AUC (95% CI)	Average precision (95% CI)	Brier score
Penalized logistic regression	0.506 (0.464-0.549)	0.163 (0.137-0.193)	0.146
Random forest	0.464 (0.419-0.511)	0.159 (0.134-0.195)	0.152
Histogram gradient boosting	0.395 (0.349-0.442)	0.150 (0.121-0.188)	0.150

Source: Authors' analysis of the supplied CER Medical Appointments No-Show dataset [36, 58].

The table provides the primary negative finding. Penalised logistic regression was the best of the three models, but its ROC-AUC of 0.506 and AP of 0.163 were essentially no better than chance ranking and the 0.172 event prevalence; the more flexible models performed worse.

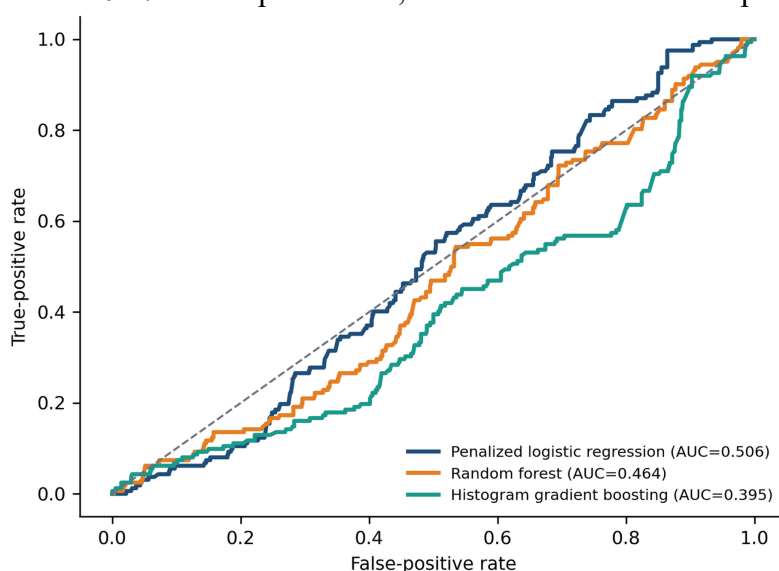


Figure 9. ROC curves on the January-February 2022 temporal holdout. A diagonal reference indicates chance ranking.

The curves lie close to, or below, the chance-level diagonal in the forward holdout. A ROC-AUC near 0.50 means that a randomly selected future no-show is not consistently ranked above a future attendance. The nominal difference between the models is therefore not a clinically meaningful basis for choosing a complex algorithm. ROC curves do not show whether predicted probabilities are accurate, whether a usable number of appointments can be targeted, or whether the score would improve a workflow. The small temporal holdout also means that apparent wiggles in the curves may be sampling variation. The visual should therefore be interpreted alongside calibration, precision-recall performance, and decision analysis; none of these models should be promoted on ranking performance alone.

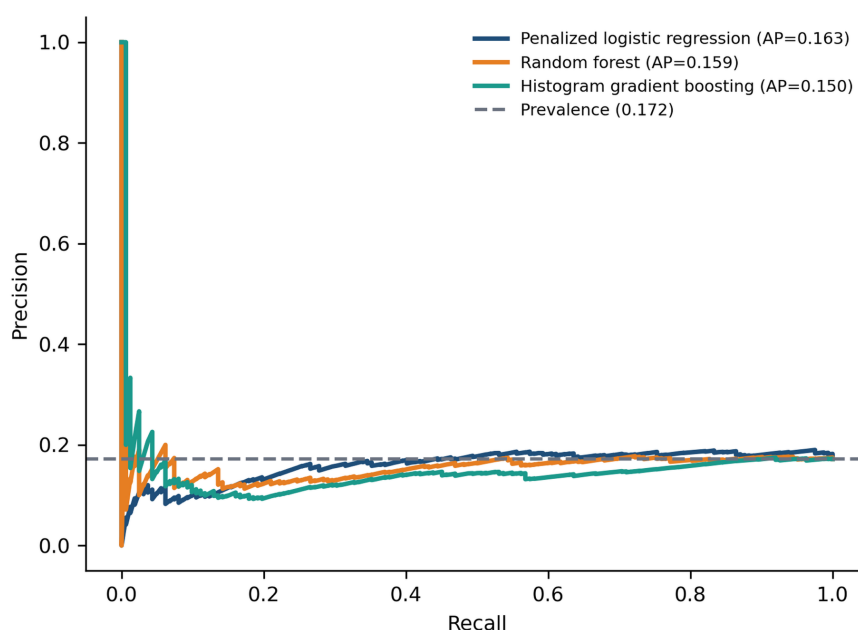


Figure 10. Precision-recall curves on the temporal holdout. The dashed line is no-show prevalence, the relevant null reference under outcome imbalance.

Precision remains close to the 17.2% holdout prevalence through much of the recall range, indicating little enrichment over selecting appointments without a model. The apparently high precision at the extreme left of the curves concerns very few observations, so it does not supply a stable or adequately sized outreach list for a clinic. This is the more relevant failure under outcome imbalance: a model may rank a small number of appointments unusually well yet still fail to identify enough avoidable no-shows to justify staff effort. Precision-recall curves do not quantify the burden of contacting false positives, nor do they establish that a contact changes behaviour. Threshold selection must therefore be tied to capacity, patient experience, and prospective intervention evidence rather than to an isolated high-precision point [59-61].

4.3 Calibration, importance, and decision utility

Calibration was poor for all three models. Logistic regression showed a calibration intercept of -0.998, a slope of 0.468 and an ECE of 0.063. Random forest and histogram gradient boosting had negative slopes, which were consistent with unstable or reversed relationships in the holdout. Figure 11 shows the spread of the observed frequencies from the expected probabilities. This does not change with the use of the modestly less accurate Brier score for logistic regression: a predicted probability very close to 0.20 cannot be relied upon as an operational risk estimate with similar observed probabilities varying considerably. The permutation importance of the logistic model was weak and unstable. Only small mean decreases in AP were found for appointment shift and lead-time proxy, while some of the calendar and demographic variables had negative importance. The negative permutation importance indicates that, for this test set, breaking the relationship between a variable and the outcome slightly decreased average precision. This is a warning that the apparent signals may

represent noise and/or unstable historical associations. The results from feature-importance rankings are not causal explanations and should not be used to change the scheduling priority. The operational consequence is made visible through the threshold analysis. All appointments were flagged at thresholds 0.05 and 0.10, and the model just mimicked a treat-all outreach policy. At 0.15, 923 of 942 appointments (98.0%) were flagged, while all 162 no-shows were captured. The results show that increasing the threshold to 0.20 resulted in a lower level of outreach (772 appointments, 81.9%) and a lower PPV (18.1%), while it did not increase the number of no-shows collected (140 no-shows) nor generate a net benefit. The sensitivity at 0.25 was reduced to 19.8% with only 241 appointments flagged. No feasible threshold was able to create a coherent and consistent target group for targeted intervention [62]. This is confirmed by the decision curve in Fig. 13. The logistic model is very similar to the treat-all strategy at very low thresholds, since almost all records will be classified as high risk. Where a clinic may want to engage in selective outreach, the model net benefit is zero or negative. The same failure is shown here as a capacity-yield trade-off: If staff were to reach out to almost the entire schedule to capture a large percentage of no-shows, staff time would be heavily consumed. Rather than searching for a more flattering metric, the appropriate conclusion is that we should not deploy any such metric [46, 63].

Table 4. Calibration assessment on the January-February 2022 temporal holdout

Model	Brier score	Calibration intercept	Calibration slope	ECE (10 bins)
Penalized logistic regression	0.146	-0.998	0.468	0.063
Random forest	0.152	-1.954	-0.178	0.095
Histogram gradient boosting	0.150	-3.110	-0.788	0.089

Source: Authors' analysis of the supplied CER Medical Appointments No-Show dataset [36]. None of the models provide stable absolute estimates of risk in 2022, as indicated in the table. The logistic model's slope of 0.468, combined with its intercept of -0.998, and the negative slopes of the ensemble models, demonstrates that there is no relationship between a stated risk threshold and the likelihood of non-attendance in later years.

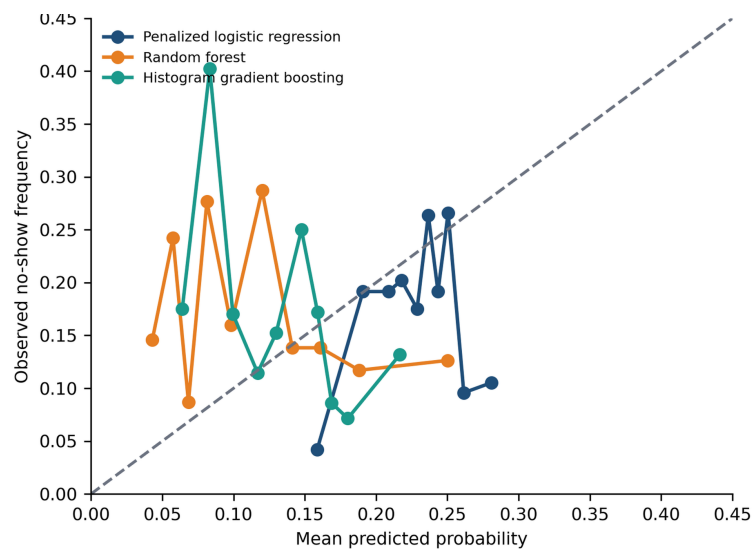


Figure 11. Calibration curves on the temporal holdout. Perfect calibration lies on the dashed diagonal; no model provides stable risk probabilities.

The calibration curves are not close to the diagonal, and the no-show rates were not similar in the latter period for appointments with similar predicted probabilities. This means that a numerical risk score is not an appropriate trigger for selective outreach because a predicted 20% does not necessarily represent a stable future risk of absence. The calibration plots based on a small holdout and the grouped bins are just as uncertain in the same unfavourable direction as the companion intercept and slope and Brier results. Recalibration would require an additional sample that is sufficiently large and relevant to the target setting. Thresholds should not determine resource allocations until they are shown to be accurate in a prospective manner.

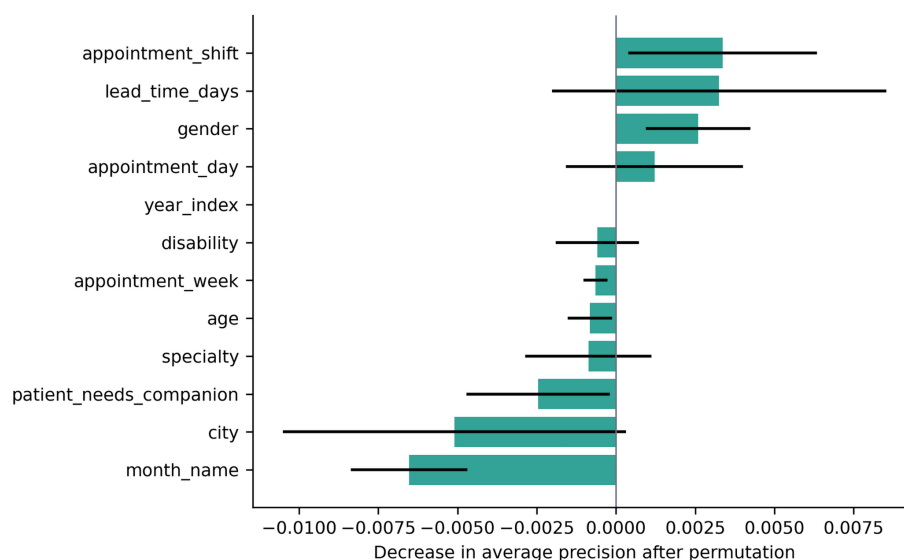


Figure 12. Permutation importance of booking-time variables for the best-performing model by average precision. Small and negative values indicate unstable information rather than robust drivers.

The relatively small and negative permutation changes suggest that the variables for booking time do not provide a consistent and repeatable contribution to average precision in the holdout. If a negative value is obtained when permuting a variable, it means that the model performed better after permutation in this sample, which is more suggestive of noise or variable instability over time, as opposed to a stable operational signal. Permutation importance is not a causal estimate and is sensitive to correlated predictors, category coding, and temporal shift. For instance, a calendar or service variable may overlap with information represented in other fields, and the significance of the calendar in the display shouldn't be taken to mean that the change will affect attendance. The plot thus contradicts the idea of individualized scheduling judgements, rather than the argument for a hierarchy of patient characteristics [64].

Table 5. Capacity-oriented threshold scenarios for penalized logistic regression on the temporal holdout

Thresh old	Appointmen ts flagged	Flagged (%)	No-shows captured	Sensitiv ity	PPV	Specif icity	Net benefit
0.050	942	100.0%	162	100.0%	17.2 %	0.0%	0.128
0.100	942	100.0%	162	100.0%	17.2 %	0.0%	0.080
0.150	923	98.0%	162	100.0%	17.6 %	2.4%	0.029
0.200	772	82.0%	140	86.4%	18.1 %	19.0%	-0.019
0.250	241	25.6%	32	19.8%	13.3 %	73.2%	-0.040
0.300	4	0.4%	0	0.0%	0.0%	99.5%	-0.002

Source: Authors' analysis of the supplied CER Medical Appointments No-Show dataset [36].

The table translates model scores into workload. Capturing all 162 no-shows at threshold 0.15 would require outreach to 923 of 942 appointments (98.0%); at threshold 0.20, 81.9% would still be contacted and net benefit is negative. These are not viable selective-targeting scenarios.

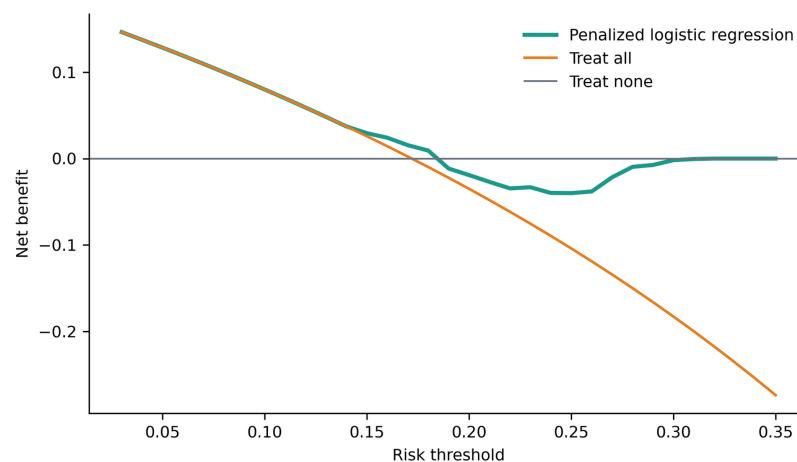


Figure 13. Decision-curve analysis. The model does not provide a useful advantage over contacting everyone at thresholds that would make outreach selective.

The decision curve shows that at low thresholds the model adds little to a policy of offering outreach to everyone, while at more selective thresholds its net benefit is zero or negative. This means that the score does not concentrate enough true no-shows to compensate for the burden of contacting many appointments that would have been attended [65]. Net benefit formalises a threshold-specific trade-off, but it does not prove that the proposed outreach is effective, affordable, or harmless. The threshold implicitly encodes the clinic's tolerance for unnecessary contact and must be justified with patient and staff input. With no advantage over simple strategies in the holdout, investment in model-guided outreach would be premature; universal low-burden support or better data collection are more defensible comparators [47, 66].

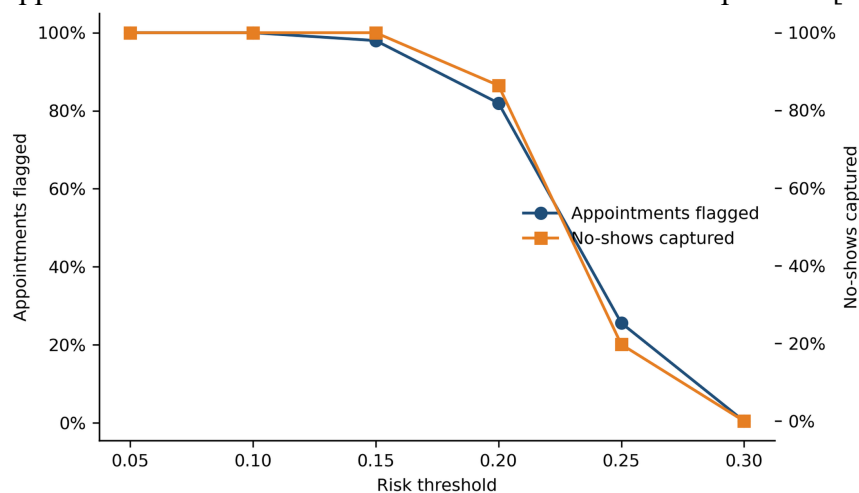


Figure 14. Threshold capacity-yield trade-off. A high capture rate requires contacting nearly the entire appointment list.

The capacity-yield trade-off translates poor discrimination into a service consequence: capturing a large share of no-shows would require contacting nearly the entire appointment list. Raising the threshold reduces workload only by losing a substantial proportion of the appointments that the intervention intends to reach. This is not a selective targeting solution.

The graph does not include intervention uptake, staff time per contact, successful rescheduling, or patient-reported burden, so it should not be converted into a cost-saving estimate. Even a high capture rate has limited operational value if the action merely substitutes one administrative task for another or disadvantages people with communication and access barriers. A future implementation study should compare this workload with a transparent universal reminder pathway before asserting efficiency gains.

5. Discussion

This study examined the feasibility of using routinely collected data in booking time fields to predict future non-attendance in a rehabilitation service, for the purpose of informing targeted outreach efforts. In the available data, the answer is no. The optimal model was a penalised logistic regression that had neither decision-curve advantage at selective thresholds nor reliable calibration. Random forest and gradient boosting performed worse. This does not mean that machine learning is not appropriate for physiotherapy or rehabilitation. It is evidence that algorithmic complexity is not sufficient to overcome non-attendance in the longitudinal dimension, irregularity in field timing, missingness, and temporal service change [67].

The most important finding is the disparity between historical and forward performance. The no-show prevalence rate nearly doubled from the development cohort to the 2022 holdout cohort. The temporal evaluation exposed this challenge and not smooth it out by splitting it randomly. A random split would spread out later, high-prevalence appointments to both development and test sets, giving the illusion to the model that it is learning some of the patterns that are time-specific. A clinic that implements a model always forecasts future scheduled appointments. The question asked in the test used here is therefore more realistic but with a small final-period sample and large confidence intervals [68].

The discovery also serves as a reminder of why the accuracy might have been misleading. In the holdout, predicting attendance for every appointment gives a high accuracy of 82.8% despite detecting no non-attendance. A paper that reported accuracy alone could mislead one to think that a usable classifier was provided. Similarly, even if AUC were slightly greater, it would not have been enough: outreach may only be done with well-calibrated absolute probability and with an acceptable workload. Using AP, Brier score, calibration intercept and slope and net benefit, the analysis shows that there is no safe operational threshold. This assessment is based on the principle of existing prediction model guidance, not optimizing a single headline metric [69].

The outcomes support a straightforward interpretation of the data. There is historical variation associated with some appointment characteristics (such as the timing of an appointment, patient age, the category of disability, and the city) and some of the services to which the appointment is related (such as the service type). These are rough proxies for what drives attendance. Prior non-attendance, cancellation pathway, booking channel, reminder delivery, preferred communication, travel difficulty, wait time, appointment duration, clinician continuity, carer availability, perceived value of the visit, and patient-reported barriers were not included in the dataset. Many of these may vary over time. Many no-show studies attribute the most predictive variables to factors linked to their history of no-shows and the scheduling process, information

which can be recovered only with a patient identifier. The physiotherapy subgroup requires an independent interpretation. It had a 10.4% no-show rate, which was slightly higher than the full cohort but less than one fifth of all records. Other rehabilitation disciplines can provide a sample size for a pooled model but may conceal speciality-specific processes. The current file contains insufficient information to confirm whether associations with appointment hour, gender, companion requirement or disability hold within physiotherapy. Future research should obtain sufficient longitudinal data from physiotherapy only or model explicitly the heterogeneity of services with prespecified validation by speciality. It should not be a mere description of a multi-service model as a physiotherapy model [70].

Additionally, the study is also quite conservative when it comes to clinical outcomes. While improved attendance can reasonably be expected to contribute to continuity of exercise progress, exercise monitoring and co-developed goals, these downstream outcomes are not measured by CER. A stronger linkage between the number of treatment contacts, prognosis, perceived improvement, costs and disability outcomes is illustrated in the FYSIOPRIM cohort. There, associations between improvement and frequency of contact do not necessarily imply causation. In the current data there is no patient identifier or PROM to evaluate for the impact of an attendance intervention on function, pain, quality of life, return to work, or discharge. A reduction in empty slots would not justify an inference of clinical improvement. Operational implication is not inaction. First, a clinic can consider low-burden (no model needed) interventions that are universal: reminders in preferred accessible formats, one-click cancellation, advance transport/accessibility checks, flexible rescheduling, and wait-list pathways. These should be assessed prospectively, using attendance and equity outcomes as endpoints. Second, it is important to enhance data architecture before trying another model. To model prior attendance and to use group-aware validation, a pseudonymised stable patient key is required. A causal and predictive study would be possible with records of appointment timestamps, the reason for cancellation, reminder channel and delivery, rescheduling, visit modality, and travel/accessibility needs volunteered by the patient, as well as clinician/service context. Third, clinical outcomes need to be connected by design: baseline PROMs, treatment dosage, discharge destination, and adequate health-economic measures [71, 72].

There are a few restrictions that warrant attention. Data is from one rehabilitation centre in southern Brazil and is over an irregular service period, so generalisability to other health systems is unknown. The 2022 holdout is only 942 appointments and two calendar months, which creates uncertainty around estimates and seasonality. The field definitions and data collection procedures were not independently audited. There is no clustering of patients within the appointment-level records, and missing information could represent documentation practices. Calendar year may reflect site-specific structural changes as opposed to a consistent risk mechanism. Short-term (less than one day) forecasts are not included in the analysis, as it could not be confirmed that the forecasts were available at the time of booking. Lastly, the paper was developed using a public, anonymised dataset, so it does not consider the impact of real-time integration, clinician uptake, patient acceptance, or unintended effects [73].

These restrictions are not evidence that a more flattering but less defensible model, should be presented. The significance of the study is that the credible stopping rule is illustrated: if

prospective-like validation, calibration, and decision analysis reveal there is no advantage, it is more appropriate to improve the data and workflow rather than deploy. This is particularly relevant in rehabilitation when risk labels may overlap with disability, communication requirements and access issues. Support and capacity should be enhanced, not automated decisions about appointment access from incomplete administrative records.

The negative result also reveals an issue in implementation. A risk model can be neutral on the surface but generate structural barriers that exist in scheduling and documentation. In a rehabilitation context, unnecessary contact, stigma or scrutiny of people with disability, transport or communication needs might occur because of false-positive labels, while preventable access barriers are not addressed because of false-negative labels. Avoiding punitive use alone does not ensure fairness. It entails the prospective measurement of those flagged, those supported, those who benefit and those burdened, and patient-clinician involvement in the design of safeguards.

Finally, the final issue is administrative data quality. The recorded no-show field does not differentiate between patient cancellation, clinic cancellation, rescheduling, patient late arrival, or partial attendance and does not indicate whether or not an unused slot could have been avoided. This composite label might consequently be a way for the model to learn documentation and workflow patterns instead of an actionable behavioural signal. There should be a data-quality/service-design response to the lack of predictive utility: Operational outcomes should be more precisely defined; longitudinal attendance history needs to be captured with clear time stamps; and testing should support interventions against clear non-algorithmic alternatives before making claims on clinical outcomes, efficiency or cost savings [67, 74].

6. Conclusion

The administrative variables routinely collected at booking did not provide a useful temporal prediction model for the non-attendance of appointments in a large public rehabilitation-service dataset with a physiotherapy subgroup. The ROC-AUC value for penalised logistic regression was 0.506, and it failed to provide an ‘outreach threshold’ that was useful for making decisions; random forest and histogram gradient boosting performed worse. The main adverse finding is the lack of reliable generalisation to the future, when the proportion of patients with no-shows is rising, while patient-history data are unavailable. The study thus does not validate automated no-show targeting, overbooking, cancellation or service deprioritisation based on the present CER file. It supports a different data-driven agenda: longitudinal attendance and scheduling variables with clear timestamps; future-period validation, calibration, and net benefit; no punitive use of the system against patients; and evaluation of operational changes along with PROMs and equity outcomes. The most defensible use of data in physiotherapy clinics is not to say an algorithm has solved attendance; it is to say what information and workflow evidence is needed for such a claim to be made.

References

1. Sabit Md Asad Sajib Debnath, S.B., Kamana Parvej Mishu, Mohammad Tahmid Ahmed, Md Ismail Jobiullah, *Advanced Cyber Attack Analysis with DNS: A Machine Learning Approach for Threat Detection and Classification*, in *9th International Conference on I-SMAC 2025 (IoT in social, Mobile, Analytics and Cloud)*. 2025/9, IEEE. p. 6, paper ID: I-SMAC-889.
2. Mir, M.N.H., et al., *Decentralized Tumor Classification with Federated Learning: A Privacy-Preserving Approach*, in *Federated Learning in Health Care Technology*. 2025, Springer. p. 19–35.
3. Aziz, A.S., et al. *A Comprehensive Framework Analysis of Cycle GAN-Based Modality Translation: Enhancing Brain Tumor Diagnostics from FLAIR to T2w*. in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*. 2025. IEEE.
4. Mishu, K.P., et al., *ADVANCING US CRITICAL INDUSTRY SUPPLY CHAIN THROUGH AI AND MACHINE LEARNING-DRIVEN LOGISTICS, RISK MITIGATION, AND OPERATIONAL EXCELLENCE*. *Cuestiones de Fisioterapia*, 2023. **52**(3): p. 971–977.
5. Kamana Parvej Mishu, I.U.R., Mohammad Tahmid Ahmed, Apurbaa Sarker, Afia Masuda Supti, *Transforming Supply Chain Operations through AI and Machine Learning: Real-Time Demand Forecasting, Inventory Optimization and Logistic Systems*. *Journal of Posthumanism*, 2026/3/11. **6**: p. 17.
6. Arif, M.H., et al., *Machine Learning Models for Cybersecurity Risk Analysis in US Critical Infrastructure Systems*. *Journal of Information Systems Engineering and Management*, 2026. **11**(1s).
7. Nagib, M.N., et al. *FedSAT-Detect: A Federated Learning Framework for Privacy-Preserving Cybersecurity Threat Detection in Cloud Systems*. in *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*. 2025. IEEE.
8. Sharif, K.S., et al. *DNA Sequence Classification: An Advanced Machine Learning Framework For Accurate Splice Junction Detection*. in *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)*. 2025. IEEE.
9. Asad, S.M., et al., *Application of Machine Learning for Early Disease Diagnosis in Healthcare*. *Cuestiones de Fisioterapia*, 2022. **51**(3): p. 332–355.
10. Alim, M.A., et al., *Enhancing fraud detection and security in banking and E-Commerce with AI-powered identity verification systems*. 2020.
11. Mishu, K.P., et al., *AI-driven supply chain management in the United States: Machine learning for predictive analytics and business decision-making*. *Cuestiones de Fisioterapia*, 2024. **53**(03): p. 5755–5768.
12. Babu, R.H., et al., *Cognitive Shielding: An Ai-Based Predictive Model for Real-Time Threat Detection in Financial Iot Networks*.

13. Huang, Y., et al., *A tutorial on calibration measurements and calibration models for clinical prediction models*. Journal of the American Medical Informatics Association, 2020. **27**(4): p. 621–633.
14. Islam, K., et al., *Enhancing Cybersecurity with Adversarial Defense: A Multi-Domain Machine Learning Perspective*. Well Testing Journal, 2025. **34**(S4): p. 236–260.
15. Nagib, M.N., et al., *Tusegnet: A transformer-based and attention-enhanced architecture for brain tumor segmentation*. IEEE Open Journal of the Computer Society, 2025. **6**: p. 750–761.
16. Collins, G.S., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement*. Annals of Internal Medicine, 2015. **162**(1): p. 55–63.
17. Ahmed, M.S., et al., *Ai-driven predictive operations management: a business science framework for dynamic hospital resource optimization and clinical workflow efficiency*. International Journal of Professional Business Review: Int. J. Prof. Bus. Rev., 2025. **10**(8): p. 5.
18. Md Abdul Alim, M.R.R., Md Shakhawat Hossen, Mamunur Rahman, Md Habibul Arif, Iftekhar Rasul, *Zero-Trust Security Models in Multi-Cloud Environments: Scalability, Challenges, and Implementation Strategies*. Journal of Advanced Research in Applied Sciences and Engineering Technology, 2025. **56**(282–29).
19. Dantas, L.F., et al., *No-shows in appointment scheduling – a systematic literature review*. Health Policy, 2018. **122**(4): p. 412–421.
20. Rahaman, M.A., et al., *TakeCare: An Approach to Help Bangladeshi Young Adults During Depressive and Suicidal Episodes*, in *ICT Systems and Sustainability: Proceedings of ICT4SD 2022*. 2022, Springer. p. 189–197.
21. Hasan, M.R., et al., *AI-Powered Inventory Optimization Models: A Strategic A Framework for Improving Stock Management In Our Supply Chains*.
22. Khatun, M. and M.S. Oyshi, *Advanced machine learning techniques for cybersecurity: Enhancing threat detection in us firms*. Journal of Computer Science and Technology Studies, 2025. **7**(2): p. 305–315.
23. Nova, A.A., et al., *CustXaiNet: A multi-modal deep learning framework for predicting customer behavior with explainable AI*. IEEE Open Journal of the Computer Society, 2025.
24. Pervez, R., M. Usmani, and T. Mahboob, *Augmented reality marker-based technology for augmenting newspaper advertisement*. Journal of Multimedia Tools and Applications, 2021. **80**: p. 12345–12367.
25. Md Habibul Arif, I.R., Md Abdul Alim, Md Reduanur Rahman, Md Shakhawat Hossen, Mamunur Rahman, *AI-Powered DDoS Detection and Mitigation: Developing Adaptive Machine Learning Frameworks to Predict and Block Next-Generation Attacks*. Journal of Advanced Research in Applied Sciences and Engineering Technology, 2025. **56**(231–243).
26. Davis, J. and M. Goadrich. *The relationship between Precision-Recall and ROC curves*. in *23rd International Conference on Machine Learning (ICML '06)*. 2006. ACM.

27. Davis, S.E., et al., *Calibration drift in regression and machine learning models for acute kidney injury*. Journal of the American Medical Informatics Association, 2017. **24**(6): p. 1052–1061.
28. Pervez, R., et al. *CryptoQNet: Leveraging Quantum Computing for Enhanced Cryptocurrency Market Analysis*. in *2025 International Conference on Emerging Smart Computing and Informatics (ESCI)*. 2025. IEEE.
29. Essery, R., et al., *Predictors of adherence to home-based physical therapies: a systematic review*. Disability and Rehabilitation, 2017. **39**(6): p. 519–534.
30. Hasvold, P.E. and R. Wootton, *Use of telephone and SMS reminders to improve attendance at hospital appointments: a systematic review*. Journal of Telemedicine and Telecare, 2011. **17**(7): p. 358–364.
31. Sozib, H.M., et al., *Cloud Computing in Business: Leveraging SaaS, IaaS, and PaaS for Growth*. Journal of Applied Research, 2025: p. 38.
32. Huang, Y. and D.A. Hanauer, *Patient no-show predictive model development using multiple data sources for an effective overbooking approach*. Applied Clinical Informatics, 2014. **5**(3): p. 836–860.
33. Nadia, N.Y., et al., *Security Analytics Using Machine Learning Methods for Federal Government Cloud Systems*. 2026.
34. Begum, S., et al., *AttenGene: A deep learning model for gene selection in PDAC classification using autoencoder and attention mechanism for precision oncology*. Well Testing Journal, 2025. **34**(S3): p. 705–726.
35. Kelly, C.J., et al., *Key challenges for delivering clinical impact with artificial intelligence*. BMC Medicine, 2019. **17**: p. 195.
36. Salazar, L.H.A., *Medical Appointments No-Show*, in *Mendeley Data*. 2023, Mendeley Data.
37. Nova, A.A., et al. *Riskawaretnet: A multi-task deep learning framework for stock price prediction and risk assessment*. in *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. 2025. IEEE.
38. Prince, N.U., et al. *Deep Transfer Learning Approach to Detect Dragon Tree Disease*. in *2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*. 2024. IEEE.
39. Mishu, K.P., et al., *AI-Driven Intelligent Compliance and Resilience Framework (AI-ICRF) for Critical US Aviation Spare Parts and Semiconductor Supply Chains*. Well Testing Journal, 2022. **31**(2): p. 240–271.
40. Sazzadul, I., et al., *Designing a Real-Time Human-AI decision support framework for US Healthcare providers using big data analytics*. Journal of Computer Science and Technology Studies, 2025. **7**(8): p. 716–723.
41. Ahmad, J., F. Tauseef, and Z. Akbar, *Predictive analytics for AI-assisted patient no-show management and clinic revenue optimization: a simulation-based research*. Migration Letters, 2024. **21**(S13): p. 1901–1924.
42. Ahona, M.H., et al., *Aidriven discovery of novel biomarkers for early cardiovascular disease detection*. Well Testing, 2026. **35**: p. 04.

43. Mahdia Amina, B.P., Khan Raqib Mahmud, *Depressive and Suicidal Episodes*. ICT Systems and Sustainability: Proceedings of ICT4SD 2022, 2022/10/31: p. 189.
44. Talukder, A.R., et al., *Underwater image enhancement and restoration with YOLO-based object detection and recognition*. Well Test J, 2025. **34**(S3): p. 727–48.
45. Abubakkar, M., et al., *PROGNOSTIC MODELING FOR HEPATIC DISORDERS: A PARADIGM OF EQUILIBRATED AND GENERALIZED MACHINE LEARNING METHODOLOGIES*. American Journal of Scholarly Research and Innovation, 2025. **4**(01): p. 352–362.
46. Jamal, A., et al., *Detecting Governance Risk In Generative AI Adoption: A Predictive Analysis Of Organizational Misalignment And AI Failure Signals*.
47. Jamal, A., et al., *Predicting Human-Genai Collaboration Effectiveness: A Machine Learning Investigation Of Skill Configurations, Trust, And Work Design*. Migration Letters, 2022. **19**(S8): p. 2303–2324.
48. Abubakkar, M., et al., *AI in Neurology: Speech-Based Detection of Parkinson's Disease using Machine Learning Models*.
49. Rabby, H.R., et al. *Comparative Analysis of Machine Learning Algorithms for Breast Cancer Prediction Using Fine Needle Aspirate Image Features*. in *2025 International Conference on Emerging Trends in Industry 4.0 Technologies (ICETI4T)*. 2025. IEEE.
50. Rabby, H.R., et al. *Coronavirus Disease Outbreak Prediction and Analysis Using Machine Learning and Classical Time Series Forecasting Models*. in *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA)*. 2024. IEEE.
51. Jamal, A., et al. *Explainable AI for Predicting Patient Readmission in Hospitals*. in *2026 9th International Conference on Inventive Computation Technologies (ICICT)*. 2026. IEEE.
52. Akter, S., M. Amina, and N. Mansoor. *Early diagnosis and comparative analysis of different machine learning algorithms for myocardial infarction prediction*. in *2021 IEEE 9th region 10 humanitarian technology conference (R10-HTC)*. 2021. IEEE.
53. Abubakkar, M., et al. *Explainable suicide risk prediction with DeepFusion: A hybrid intelligence approach*. in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*. 2025. IEEE.
54. Jamal, A., et al., *BUSINESS ANALYTICS FOR HEALTHCARE COST OPTIMIZATION: A DATA-DRIVEN STUDY ON IMPROVING FINANCIAL SUSTAINABILITY IN HEALTHCARE SYSTEMS*. Spectrum of Engineering Sciences, 2024: p. 723–736.
55. Niaz, S., et al., *AI for Inclusive Educational Governance and Digital Equity Examining the Impact of AI Adoption and Open Data on Community Trust and Policy Effectiveness*. Contemporary Journal of Social Science Review, 2024. **2**(04): p. 2557–2567.
56. Uddin, M.M., et al. *Interpretable Alzheimer's Disease Diagnosis Via CNNs and MRI: an Explainable AI Approach*. in *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*. 2025. IEEE.

57. Arif, M.H., et al., *AI-Driven Risk Assessment in National Security Projects: Investigating machine learning models to predict and mitigate risks in defense and critical infrastructure projects*. Journal of Computer Science and Technology Studies, 2025. 7(2): p. 71–85.
58. Hossen, M.S., et al., *Parallel Deep Learning for Cybersecurity-Oriented Multi-Platform Social Media Bot Detection*. Frontiers in Computer Science and Artificial Intelligence, 2026. 5(2): p. 13–20.
59. Jamshaid, M.M., et al., *IMPACT OF ARTIFICIAL INTELLIGENCE ON WORKFORCE DEVELOPMENT: ADAPTING SKILLS, TRAINING MODELS, AND EMPLOYEE WELL-BEING FOR THE FUTURE OF WORK*. Spectrum of Engineering Sciences, 2024.
60. Zamil, M.Z.H., et al. *Stroke Prediction on Healthcare Data Using SMOTE and Explainable Machine Learning*. in *2025 13th International Symposium on Digital Forensics and Security (ISDFS)*. 2025. IEEE.
61. Biswas, A.K., et al., *A Dual Output Temporal Convolutional Network With Attention Architecture for Stock Price Prediction and Risk Assessment*. IEEE Access, 2025.
62. TOHFA, N.A., et al., *Machine learning-enabled anomaly detection for environmental risk management in banking*. 2025.
63. Islam, M.T., et al., *LaplaceSalesNet: A Neural Laplace-Transformer Framework for Continuous-Time Sales Forecasting*. IEEE Open Journal of the Computer Society, 2025.
64. Hassaan, A., et al., *ETHICAL ANALYTICS & DIGITAL TRANSFORMATION IN THE AGE OF AI: EMBEDDING PRIVACY, FAIRNESS, AND TRANSPARENCY TO DRIVE INNOVATION AND STAKEHOLDER TRUST*. Contemporary Journal of Social Science Review, 2023. 1(04): p. 1–18.
65. Islam, M.T., et al. *FinGraphRisk: A Graph Neural Network-Based Framework for Sentiment-Driven Financial Risk Assessment*. in *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*. 2025. IEEE.
66. Amina, M., et al. *AI-Powered Personalized Therapy for Children with Autism: A Reinforcement Learning Approach*. in *2025 International Conference on Advances in Machine Intelligence, and Cybersecurity Technologies (AMICT)*. 2025. IEEE.
67. Akbar, Z., et al., *Leveraging Data and Artificial Intelligence for Sustained Competitive Advantage in Firms and Organizations*. Journal of Innovative Computing and Emerging Technologies, 2023. 3(1).
68. HASSAAN, A., et al., *AI-driven administrative automation: Enhancing operational efficiency and security*. TPM–Testing, Psychometrics, Methodology in Applied Psychology, 2025. 32(S7 (2025)): Posted 10 October): p. 2451–2460.
69. Mir, M.N.H., et al., *ABMF-Net: An Attentive Bayesian Multi-Stage Deep Learning Model for Robust Forecasting of Electricity Price and Demand*. IEEE Open Journal of the Computer Society, 2025.

70. Jamshaid, M., et al., *Artificial intelligence generated deepfakes as instruments of disinformation: Examining their influence on public opinion, digital trust, and governance*. Journal of Information Systems Engineering and Management, 2025. **10**.
71. Jamshaid, M.M., et al., *Preparing human oversight talent for agentic AI workplaces: A competency framework for education and workforce systems*. Contemporary Journal of Social Science Review, 2025. **3**(4): p. 1561–1574.
72. Shaker, A.S.M., et al., *TACS-Net: Temporal-Aware Customer Segmentation Network*. IEEE Open Journal of the Computer Society, 2025.
73. Javed Mehedi Shamrat, F., et al. *A model based on convolutional neural network (CNN) for vehicle classification*. in *Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 1*. 2022. Springer.
74. Mishu, K.P., et al., *Utilizing Blockchain Technology for the US Supply Chain Management*. Well Testing Journal, 2026. **35**(S1): p. 1–19.