

ORIGINAL RESEARCH

Open access

Predictive Analytics Model for Estimating Outpatient Clinic Overbooking Risk Using Appointment History, Provider Schedule Density, Visit Complexity, No-Show Probability, Seasonal Trends, and Patient Communication Records

Jinwoo Park^{1*}, Minji Kim¹, Seung Lee²

Abstract

Outpatient clinics routinely overbook to compensate for patient no-shows, but poorly calibrated overbooking can create provider overtime, patient wait time, and staff burnout. The operational challenge is to preserve access without overwhelming clinical capacity. Current overbooking rules often rely on static session-level averages. These rules ignore the dynamic risk profile of the specific patient being added to the schedule. This article develops a predictive analytics model that estimates overbooking risk for each proposed additional appointment. The model uses patient-specific features together with provider, schedule, seasonal, and communication context. The proposed model would use gradient-boosted classification or regression trained on historical appointment data. Its output would be a risk score reflecting the likelihood of excessive wait time, overtime, or queue formation for the session. Conceptually, the model would flag situations in which adding a particular patient to a dense session creates high operational risk. It would also identify lower-risk overbooking opportunities when the schedule has sufficient flexibility. The model could enable precision overbooking in outpatient clinics. It would support access and efficiency while reducing the negative consequences of both no-shows and excessive overbooking.

Keywords Predictive analytics, Outpatient scheduling, Overbooking risk, No-show prediction, Appointment history, Provider utilisation

*Correspondence:

Jinwoo Park
jinwoo.park@outlook.com

¹ Department of Digital Health Engineering, College of Medicine, Seoul National University, Seoul, South Korea

² Department of Intelligent Health Systems, Faculty of Engineering, KAIST, Daejeon, South Korea

Introduction

Patient no-shows remain a persistent operational problem in outpatient clinics because unused appointment slots reduce access, waste provider capacity, and disrupt planned session flow. Overbooking is a common response because it attempts to offset expected non-attendance by adding appointments beyond nominal capacity. However, overbooking becomes effective only when the clinic can distinguish likely idle capacity from genuine congestion risk [1-3].

Aggressive overbooking can shift the problem from underutilisation to overcrowding. When more patients arrive than expected, clinics may experience prolonged waits, rushed consultations, provider overtime, and patient dissatisfaction.

These consequences are especially important in settings where diagnostic imaging, specialty consultations, and complex visits already create variable service times [4-7].

Traditional overbooking rules are blunt instruments because they often use historical averages rather than patient-level and session-level predictors. They may not distinguish a new patient with high visit complexity from a brief follow-up visit, or a lightly booked provider session from one already near capacity. Predictive scheduling research suggests that appointment history, lead time, provider patterns, and clinic context can be integrated into richer decision support than static heuristics [8-11].

This article proposes a predictive analytics model for estimating the risk associated with adding a specific appointment to an already scheduled outpatient session. The model would combine appointment history, provider schedule density, visit complexity, no-show probability, seasonal trends, and patient communication records into a single overbooking risk score. Such a score could support scheduler decisions by estimating whether a proposed overbooked appointment is likely to create operational disruption or preserve access without excessive congestion.

Background

Overbooking in outpatient clinics: rationale and consequences

Overbooking is grounded in the operational logic that some scheduled patients will not attend, creating otherwise unused capacity. Yet the same strategy can produce long waits and overtime when too many patients arrive, making the decision highly sensitive to uncertainty in attendance and service duration. Prior work on individualized no-show prediction, appointment overbooking, and access improvement shows that the value of overbooking depends on balancing utilisation against congestion and fairness concerns [1, 2, 7].

No-show prediction models

No-show prediction models commonly use lead time, prior attendance, demographics, insurance status, distance, appointment type, and contextual variables to estimate whether a patient will attend. Machine learning approaches have been applied across pediatric, neurology, radiology, dental, psychiatric, and general outpatient settings, showing that patient behavior can be partially predicted but remains uncertain at the individual level. This uncertainty is why overbooking risk should be modeled as a session-level operational consequence rather than as a simple patient-level no-show probability [12-24].

Visit complexity and provider schedule density

Visit complexity is a central determinant of outpatient clinic flow because it directly influences service time variability and downstream congestion. New-patient visits typically require extended history-taking, diagnostic clarification, and care planning, while procedure-based encounters and imaging visits introduce setup time, equipment constraints, and coordination delays that are absent in routine follow-ups. High-comorbidity patients further increase uncertainty because their visits are more likely to involve unexpected clinical findings, additional documentation, or coordination with other services. As a result, even a small number of complex visits within a session can create cascading delays that propagate across subsequent appointments, amplifying the operational impact of overbooking decisions [4, 6, 8, 9].

Provider schedule density interacts with visit complexity in a nonlinear manner, meaning that the same additional appointment may have vastly different consequences depending on the existing composition of the session. A session populated with short, predictable follow-ups may absorb an extra patient with minimal disruption, whereas a session already filled with new patients or procedures may be operating near its effective capacity even if nominal time slots remain. This distinction highlights that schedule density should not be measured solely as the number of booked slots, but rather as a weighted construct reflecting both volume and complexity of appointments. Incorporating such weighted

density measures aligns with predictive scheduling frameworks that emphasize the joint influence of appointment mix and provider throughput on wait times and overtime risk [6, 25-27].

Provider-specific characteristics further complicate the relationship between density and operational performance. Clinicians differ in pacing, documentation style, support staff availability, and tolerance for compressed schedules, leading to systematic variation in how densely appointments can be safely scheduled. Template design also plays a role, as some providers build buffers or flexible slots into their schedules while others rely on tightly packed appointment grids. These variations suggest that overbooking risk cannot be generalized across providers but must instead be calibrated to individual scheduling environments, reinforcing the need for provider-level features and clustering in predictive models [8, 9, 26].

Importantly, visit complexity and schedule density jointly determine not only average performance but also variability, which is critical for risk estimation. A session with moderate average utilisation but high variability in visit duration may pose greater overbooking risk than a uniformly dense but predictable session. This implies that the model should capture distributional properties such as variance in expected service times, sequencing of complex visits, and clustering of high-risk appointments within specific time windows. By integrating these factors, the model can better approximate the real operational dynamics that drive queue formation and overtime rather than relying on simplified representations of capacity [4, 6, 25].

Seasonal and temporal patterns in attendance

Temporal patterns play a significant role in outpatient clinic operations because both demand and attendance behavior fluctuate systematically over time. Day-of-week effects are commonly observed, with certain weekdays exhibiting higher no-show rates or increased demand due to work schedules, referral patterns, or clinic-specific practices. Monthly and seasonal variations further influence attendance, as factors such as weather conditions, school calendars, and public holidays alter patient availability and willingness to attend appointments. These recurring patterns suggest that time should be treated as a structured predictor rather than as random noise in scheduling models [3, 14, 17].

Seasonal illness trends, such as influenza peaks or allergy seasons, can simultaneously increase visit demand and alter visit complexity. During high-demand periods, clinics may experience surges in acute visits that require longer or less predictable consultation times, thereby increasing the operational sensitivity to overbooking. Conversely, holiday periods may reduce demand but increase no-show rates due to travel and competing commitments. These dynamics indicate that temporal features influence both the likelihood of attendance and the downstream consequences of attendance, reinforcing their importance in overbooking risk estimation [10, 28].

Temporal proximity to the appointment date also affects patient behavior, particularly through lead time. Appointments scheduled far in advance are more susceptible to cancellation or no-show, while shorter lead times may increase attendance but reduce scheduling flexibility. The interaction between lead time and seasonal context can further complicate predictions, as patients may behave differently depending on whether an appointment falls during a high-demand or low-demand period. Capturing these interactions allows the model to adjust risk estimates dynamically as the appointment date approaches [14, 17].

Incorporating temporal features into predictive models enables more realistic representation of outpatient operations by aligning predictions with known patterns of variability. Rather than treating sessions as independent and identically distributed, the model can learn periodic structures and temporal dependencies that influence both attendance and service demand. This approach supports more accurate estimation of overbooking risk by accounting for predictable fluctuations in patient behavior and clinic workload over time [3, 10, 28].

Patient communication and reminder effects

Patient communication is a critical operational lever because it directly influences attendance behavior in the period leading up to an appointment. Reminder systems, including SMS messages, phone calls, and patient portal notifications, are widely used to reduce no-show rates by prompting patients to confirm, reschedule, or cancel their appointments. The effectiveness of these interventions varies depending on timing, modality, and patient characteristics, but they consistently demonstrate that communication can modify attendance probabilities in a measurable way. As such, communication records provide valuable, time-sensitive signals that can refine predictive models beyond static patient attributes [29-31].

Confirmation status is particularly informative because it reflects an explicit patient response to a reminder. Patients who confirm their appointments are generally more likely to attend, whereas lack of response or failed delivery may indicate higher risk of no-show. Additionally, patterns of engagement with communication channels, such as portal usage or responsiveness to text messages, can serve as proxies for patient reliability and engagement with care. These behavioral indicators complement traditional predictors such as prior attendance history and demographics [32, 33].

The timing of communication relative to the appointment also matters, as reminders delivered closer to the appointment date may have stronger effects on attendance. Multiple reminders or multimodal communication strategies may further influence behavior, although their impact may depend on clinic context and patient population. Incorporating these temporal and modality-specific features allows the model to update risk estimates as new information becomes available, effectively treating communication as a dynamic input rather than a static attribute [29-33].

From an overbooking perspective, communication features are valuable because they can reduce uncertainty about which patients are likely to attend. By integrating reminder delivery and response data into the model, schedulers can make more informed decisions about whether additional appointments can be safely added to a session. This dynamic updating of attendance probability aligns with the broader goal of precision overbooking, where decisions are tailored to the evolving state of both the schedule and the patient population [34].

Model Development Overview

High-level predictive pipeline

The predictive pipeline is designed to operate at the moment a scheduling decision is being made, ensuring that risk estimation reflects the most current information available. When a scheduler considers adding an appointment to a session, the model aggregates patient-specific features, appointment characteristics, provider context, and session-level state into a unified representation. This includes retrieving historical attendance patterns, estimating no-show probability, assessing visit complexity, and evaluating current schedule density. The resulting computation produces a session-specific overbooking risk score that reflects the expected operational impact of the proposed addition [1, 8, 9, 11, 25].

Figure 1 illustrates the hierarchical predictive pipeline that integrates patient, provider, schedule, temporal, and communication features into a unified overbooking risk score for real-time scheduling decisions.

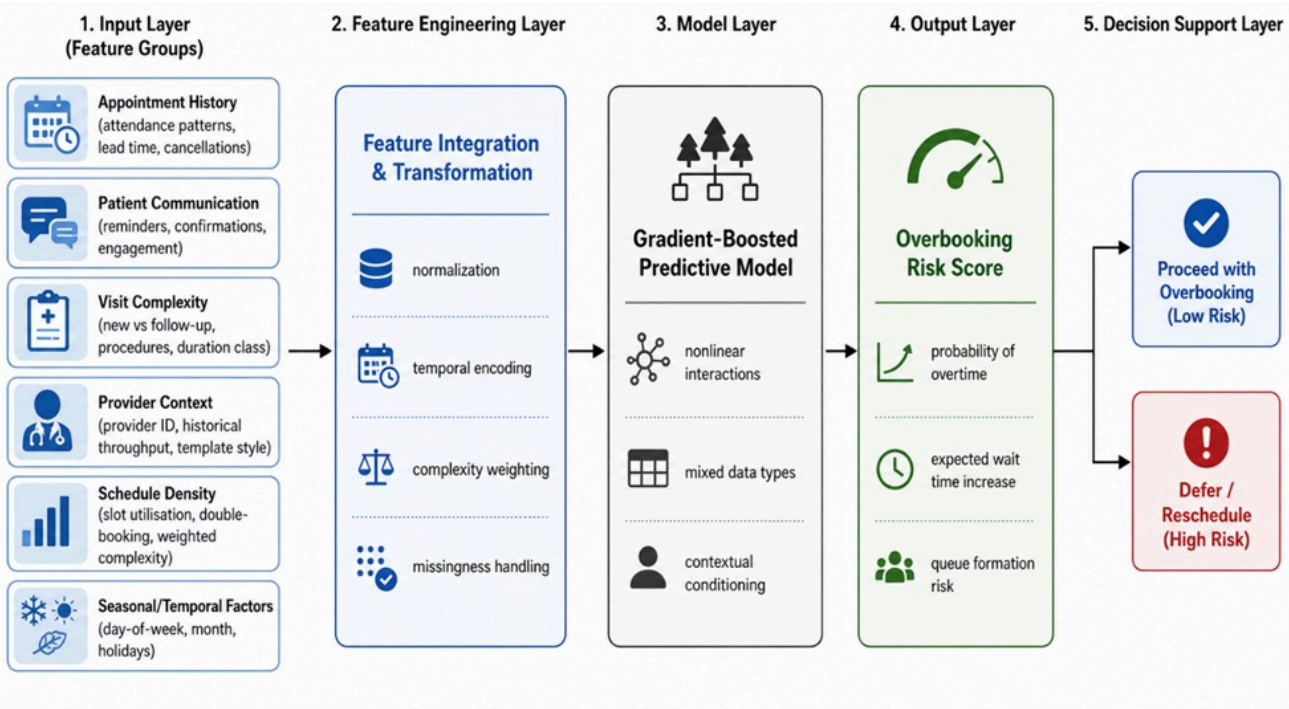


Figure 1. Hierarchical Predictive Pipeline for Patient-Specific Overbooking Risk Estimation in Outpatient Clinics

A key aspect of this pipeline is its conditional nature, as the risk estimate depends on both the candidate appointment and the existing composition of the session. This contrasts with traditional approaches that treat overbooking as a fixed adjustment applied uniformly across sessions. By conditioning on the current schedule, the model captures interactions between appointments, such as the cumulative effect of multiple complex visits or the buffering capacity of low-density periods. This design aligns with predictive scheduling research emphasizing context-aware decision-making [8, 9, 25].

The pipeline also supports iterative updates as the schedule evolves, allowing risk estimates to be recalculated whenever new information becomes available. For example, confirmation of a reminder or cancellation of an existing appointment can immediately alter the risk profile of the session. This dynamic updating ensures that scheduling decisions remain responsive to real-time changes rather than relying on outdated assumptions [1, 11, 25].

Core input features

The feature set integrates multiple dimensions of outpatient scheduling to capture the multifactorial nature of overbooking risk. Patient-level features include prior no-shows, cancellations, rescheduling behavior, arrival punctuality, and lead time, all of which have been shown to influence attendance probability. Appointment-level features include visit type, duration category, and new versus established patient status, which collectively define visit complexity and expected service time [4, 12-15].

Provider-level and session-level features capture schedule density, template utilisation, double-booking frequency, and historical throughput patterns. These features reflect the operational environment in which the appointment will occur, allowing the model to distinguish between sessions with different capacity constraints. Seasonal and temporal features add context by encoding time-related patterns, while communication features provide real-time signals about patient engagement and likelihood of attendance [29-33].

The integration of these feature groups enables the model to capture interactions that would be missed by simpler approaches. For example, the risk associated with a complex visit may be mitigated if the patient has a high probability of

no-show, or amplified if the session is already densely packed with similar visits. By combining patient, provider, and contextual information, the model provides a holistic representation of overbooking risk [4, 12-15, 29-33].

Table 1 presents a structural decomposition of the multidimensional determinants of overbooking risk, highlighting how feature domains interact to influence operational outcomes beyond isolated effects.

Table 1. Multidimensional Determinants of Overbooking Risk: Structural Roles, Interactions, and Operational Implications

Feature Domain	Primary Role in Model	Key Variables	Interaction Effects	Operational Interpretation
Appointment History	Behavioral prediction	Prior no-shows, cancellations, lead time, punctuality	Modulates reliability of attendance estimates when combined with communication signals	Identifies patient-level uncertainty and reliability
Patient Communication	Dynamic attendance updating	Reminder delivery, confirmation status, modality, timing	Reduces uncertainty when combined with appointment history; modifies real-time risk	Enables adaptive risk recalibration close to appointment
Visit Complexity	Service time estimation	New vs follow-up, procedures, imaging, duration class	Amplifies risk when combined with high schedule density	Drives variability and potential bottlenecks in session flow
Provider Context	Capacity heterogeneity	Provider throughput, template design, staffing support	Interacts with density and complexity to define effective capacity	Explains variability in tolerance to overbooking
Schedule Density	Load measurement	Slot utilisation, double-booking, complexity-weighted load	Nonlinear interaction with complexity and provider characteristics	Determines proximity to congestion threshold
Seasonal/Temporal Factors	Systematic variability	Day-of-week, month, holidays, demand cycles	Interacts with no-show probability and visit mix	Captures predictable fluctuations in demand and attendance

Design principles

The model is designed to meet practical requirements for deployment in outpatient scheduling environments. Real-time computability is essential because risk estimates must be available within the scheduling workflow without introducing delays. Interpretability is equally important, as schedulers need to understand and trust the model's recommendations, particularly when they conflict with existing heuristics or clinical priorities. Providing clear explanations for risk scores supports adoption and appropriate use [2, 34].

Flexibility is another key principle, as clinics differ in their tolerance for wait times, overtime, and access constraints. The model should allow adjustment of risk thresholds to align with local operational goals, enabling different strategies for high-demand versus low-demand periods. This adaptability ensures that the model remains relevant across diverse clinical settings and evolving organisational priorities [27, 35].

Finally, the model should include mechanisms for monitoring fairness and unintended consequences. Because predictive models are trained on historical data, they may inadvertently reinforce existing disparities in access or scheduling

practices. Incorporating fairness checks and ongoing evaluation helps ensure that the model supports equitable care delivery while improving operational efficiency [2, 27, 35, 34].

Data Sources and Feature Engineering

Appointment history and attendance labelling

Appointment history forms the foundation of the predictive model, as it provides direct evidence of patient behavior over time. Each historical encounter can be labeled according to attendance outcome, including attended, no-show, cancelled in advance, rescheduled, or late arrival. These labels enable construction of longitudinal patient profiles that capture patterns such as frequent cancellations, consistent punctuality, or sporadic attendance. Such patterns have been widely used in no-show prediction models as key indicators of future behavior [3, 12, 14, 16].

Feature engineering from appointment history involves aggregating these behaviors into meaningful predictors, such as the proportion of missed appointments, time since last no-show, and variability in attendance patterns. Lead time between booking and appointment date is also critical, as longer lead times are associated with higher uncertainty. Additionally, rescheduling frequency may indicate patient engagement or instability in availability, both of which can influence attendance probability [18, 20, 21].

By structuring these features at the patient level while preserving temporal ordering, the model can learn both stable behavioral tendencies and recent changes. This dual perspective allows the model to adapt to evolving patient behavior while maintaining sensitivity to long-term patterns. As a result, appointment history becomes a rich and dynamic source of predictive information for overbooking risk estimation [3, 12, 21].

Provider schedule density and visit complexity metrics

Provider schedule density is operationalised through a combination of quantitative and structural features that reflect how fully a session is booked. Metrics such as the ratio of scheduled appointments to nominal capacity, frequency of double-booked slots, and distribution of appointments across the session timeline provide a baseline measure of density. However, these measures must be augmented with complexity-weighted adjustments to account for variation in service time across different visit types [8, 9, 25].

Visit complexity metrics are derived from appointment attributes such as new-patient status, procedure indicators, imaging requirements, and expected duration categories. When available, additional markers such as comorbidity burden or need for multidisciplinary coordination can further refine complexity estimates. Combining these features allows the model to estimate effective workload rather than relying solely on appointment counts, which may underestimate the burden of complex sessions [6, 16, 26].

The interaction between density and complexity is particularly important for identifying tipping points where additional appointments transition from manageable to disruptive. By capturing both dimensions, the model can differentiate between sessions that appear similar in size but differ substantially in operational risk. This nuanced representation supports more accurate estimation of overbooking consequences in real-world clinical settings [6, 25].

Seasonal and communication features

Seasonal features are engineered to capture recurring temporal patterns that influence both attendance and demand. These include calendar-based variables such as month, week of year, and day of week, as well as indicators for holidays, school breaks, and known high-demand periods. Encoding these features allows the model to learn periodic trends and adjust predictions accordingly, improving alignment with observed clinic operations [10, 28].

Communication features are derived from patient engagement with reminder systems, including modality, timing, delivery success, and response behavior. For example, a confirmed SMS reminder may reduce uncertainty about attendance, while a failed phone call may signal increased risk of no-show. These features are inherently time-dependent, as their predictive value changes as the appointment date approaches and additional communication events occur [29-33].

Combining seasonal and communication features enables the model to incorporate both macro-level patterns and micro-level signals into risk estimation. Seasonal trends provide a baseline expectation of attendance and demand, while communication records refine these expectations for individual patients. This integration supports dynamic updating of risk scores and enhances the model's ability to reflect real-time operational conditions [10, 28-34].

Predictive Analytics Architecture

Model choice and rationale

Gradient-boosted decision trees would be appropriate because they can handle nonlinear relationships, mixed data types, missingness patterns, and interactions among patient, provider, and schedule features. For example, the operational risk of adding a patient may depend jointly on prior attendance, provider density, visit complexity, and reminder confirmation rather than on any single predictor alone. Quantile regression or probabilistic extensions could also be considered to represent uncertainty in wait time, overtime, or queue formation [8, 9, 11, 16, 17].

Input feature vector and preprocessing

The input vector would combine patient-level, appointment-level, provider-level, session-level, seasonal, and communication features into a single representation evaluated at the moment of scheduling. Density measures could be normalised within clinic and provider, categorical variables could be encoded with safeguards against leakage, and missing communication records could be represented explicitly rather than assumed to indicate no reminder. These preprocessing choices would support transportability across clinics while preserving the operational meaning of local scheduling practices [14, 19-24].

Output: overbooking risk score

The model output would be a continuous overbooking risk score representing the probability that adding the appointment would breach an operational threshold such as excessive waiting, overtime, or queue length. Alternatively, the output could be framed as expected incremental operational burden, allowing clinics to rank candidate overbook slots by relative risk. The score would be intended for decision support, not automatic scheduling, so that staff can weigh access needs, clinical urgency, and operational constraints together [1, 2, 5, 7, 34, 25].

Handling Temporal Dynamics and Confounding

Session-level context and temporal ordering

Overbooking risk should be conditional on the current state of the session rather than estimated from the candidate patient alone. The same appointment may be acceptable early in a lightly booked session but risky once several complex visits, procedures, or historically punctual patients have already been scheduled. Temporal ordering also helps prevent leakage by ensuring that the model uses only information available at the time the scheduling decision is made [1, 8, 9, 25].

Clustering by provider and clinic

Provider and clinic clustering should be represented because session flow depends on local template design, staffing, room availability, specialty mix, and provider-specific throughput. Random effects, provider identifiers, clinic-level

encodings, or hierarchical calibration could help account for systematic variation without assuming that all clinics tolerate overbooking in the same way. This is especially important when operational consequences differ across primary care, specialty care, imaging, and behavioral health settings [4-6, 26, 27, 35].

Evolving patient behaviour and model decay

Patient attendance behavior and reminder responsiveness may change as clinic policies, communication channels, and population characteristics evolve. The model should therefore be monitored for concept drift, calibration decay, and changing relationships between reminders and attendance. Periodic retraining or recalibration would be expected to maintain reliability as scheduling practices and patient behavior shift over time [29-34].

Model Interpretability and Clinical Trust

Explaining risk scores to schedulers

Schedulers need concise explanations for why a proposed overbooked appointment is classified as low, moderate, or high risk. Feature attribution methods could show whether the risk is driven by high existing schedule density, a complex visit type, prior late arrivals, lack of reminder confirmation, or low predicted no-show probability. Transparent explanations are also important for identifying bias, especially when historical scheduling data may encode inequitable access patterns [2, 11-13].

Integration with scheduling software

The risk score should appear inside the scheduling interface rather than in a separate analytics dashboard. A scheduler could view the score alongside available overbook slots, existing session density, and a brief explanation of the main risk drivers. Prior work on predictive scheduling and appointment management supports embedding model outputs directly into operational workflows so that decision support is timely and usable [8, 9, 25, 26].

Clinical Deployment and Scheduling Workflow Integration

Point-of-scheduling decision support

At the point of scheduling, the model would function as a decision-support service that receives patient, appointment, provider, session, seasonal, and communication features. It could return a risk score and suggest whether another session or time slot would create lower operational risk. This workflow aligns with predictive appointment-scheduling research that uses machine learning to reduce uncertainty before finalising outpatient appointments [1, 8, 9, 25].

Post-session learning and adjustment

After each clinic session, observed wait times, arrivals, cancellations, no-shows, and overtime could be used to recalibrate future risk estimates. Clinic leaders could adjust the risk threshold depending on whether their priority is faster access, lower overtime, shorter waits, or more stable provider workload. Such feedback would allow the model to remain operationally aligned rather than functioning as a static prediction tool [5-7, 34].

Evaluation Strategy

Predictive performance metrics

Evaluation should distinguish between predicting attendance and predicting operational disruption. Classification metrics such as discrimination and calibration would be relevant when the output is a threshold-breach risk, while regression

metrics and prediction-interval assessment would be relevant when the output is expected incremental wait time or overtime. Because no-show prediction alone does not fully determine overbooking safety, performance should be interpreted in relation to session-level operational outcomes [12, 13, 23, 24, 34].

Temporal validation and simulated deployment

Temporal validation should train the model on earlier appointments and evaluate it on later sessions to reflect real deployment conditions. Simulated deployment could compare hypothetical model-guided overbooking decisions with standard scheduling heuristics while preserving the chronological order of booking and communication events. This approach would help assess whether the model could support safer overbooking without claiming experimental results or prospective impact [1, 8, 9, 11].

Prospective impact assessment

A future prospective assessment could compare machine-guided overbooking with usual scheduling practice using pragmatic designs such as stepped-wedge implementation or controlled rollout. Outcomes should include clinic overtime, patient wait time, same-day access, provider utilisation, and no-show-adjusted capacity use. Such evaluation would be necessary before concluding that the model improves outpatient operations in routine practice [5-7, 25, 34].

Limitations

Data sparsity and cold-start problem

New patients may lack prior attendance history, cancellation patterns, punctuality records, and communication-response history. In these cases, the model would rely more heavily on population-level features such as visit type, lead time, clinic, provider, and seasonal context. This cold-start limitation should be explicitly communicated so that schedulers understand when individualised risk estimates are less certain [3, 14, 18, 22, 23].

Contextual factors not captured

Some disruptions are not easily captured in historical scheduling data, including sudden provider illness, severe weather, transportation failures, room shortages, or unusual patient preferences. Communication logs may also be incomplete, inconsistent, or unavailable across clinics, limiting the reliability of reminder-based features. These constraints mean that the model should support human judgment rather than replace operational oversight [10, 24, 28-33].

Conclusion

A predictive analytics model for outpatient clinic overbooking risk could help clinics move from static overbooking rules to patient-specific, session-aware decision support. By integrating appointment history, provider schedule density, visit complexity, no-show probability, seasonal patterns, and communication records, the model could estimate the operational risk of each proposed additional appointment.

The main strength of the proposed approach is that it treats overbooking as a contextual scheduling decision rather than a fixed percentage added to every session. Its output could be integrated into real-time scheduling workflows and presented in an interpretable format for front-line staff. This would allow clinics to balance access goals with the need to avoid excessive waiting, overtime, and staff burden.

Important challenges remain. New patients may have limited historical data, communication logs may be incomplete, and each clinic may require local calibration before the risk score becomes operationally meaningful. Fairness monitoring

would also be necessary to ensure that historical inequities are not reproduced through automated scheduling recommendations.

Future work should focus on pragmatic implementation trials in diverse outpatient clinic settings. Shared benchmarks for overbooking optimisation would also help compare models across specialties, patient populations, and operational constraints. Ultimately, predictive overbooking should be evaluated not only by statistical accuracy but also by its ability to improve access, workflow stability, and patient experience.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 08 Apr 2023

Revised: 04 Jun 2023

Accepted: 12 Jul 2023

Published online: 20 July 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Li Y, Tang SY, Johnson J, Lubarsky DA. Individualized no-show predictions: effect on clinic overbooking and appointment reminders. *Prod Oper Manag.* 2019;28(8):2068-86.
<https://doi.org/10.1111/poms.13042>.
- Samorani M, Harris SL, Blount LG, Lu H, Santoro MA. Overbooked and overlooked: machine learning and racial bias in medical appointment scheduling. *Manuf Serv Oper Manag.* 2022;24(6):2825-42.
<https://doi.org/10.1287/msom.2021.1050>.
- Goffman RM, Harris SL, May JH, Milicevic AS, Monte RJ, Myaskovsky L, et al. Modeling patient no-show history and predicting future outpatient appointment behavior in the Veterans Health Administration. *Mil Med.* 2017;182(5-6):e1708-e1714.

<https://doi.org/10.7205/MILMED-D-16-00236>.

Chong LR, Tsai KT, Lee LL, Foo SG, Chang PC. Artificial intelligence predictive analytics in the management of outpatient MRI appointment no-shows. *AJR Am J Roentgenol.* 2020;215(5):1155-62.

<https://doi.org/10.2214/AJR.19.22443>.

Rothenberg S, Bame B, Herskovitz E. Prospective evaluation of a machine-learning prediction model for missed radiology appointments. *J Digit Imaging.* 2022;35(6):1690-3.

<https://doi.org/10.1007/s10278-022-00720-5>.

Rastpour A, McGregor C. Predicting patient wait times by using highly deidentified data in mental health care: enhanced machine learning approach. *JMIR Ment Health.* 2022;9(8):e38428.

<https://doi.org/10.2196/38428>.

Parente CA, Salvatore D, Gallo GM, Cipollini F. Using overbooking to manage no-shows in an Italian healthcare center. *BMC Health Serv Res.* 2018;18(1):185.

<https://doi.org/10.1186/s12913-018-2997-4>.

Srinivas S, Ravindran AR. Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: a prescriptive analytics framework. *Expert Syst Appl.* 2018;102:245-61.

<https://doi.org/10.1016/j.eswa.2018.02.022>.

Golmohammadi D, Zhao L, Dreyfus D. Using machine learning techniques to reduce uncertainty for outpatient appointment scheduling practices in outpatient clinics. *Omega.* 2023;120:102907.

<https://doi.org/10.1016/j.omega.2022.102907>.

Drewek R, Mirea L, Adelson PD. Lead time to appointment and no-show rates for new and follow-up patients in an ambulatory clinic. *Health Care Manag (Frederick).* 2017;36(1):4-9.

<https://doi.org/10.1097/HCM.000000000000141>.

Samorani M, Harris S. The impact of probabilistic classifiers on appointment scheduling with no-shows. In: *Proceedings of the Fortieth International Conference on Information Systems*. Munich; 2019.

Liu D, Shin WY, Sprecher E, Conroy K, Santiago O, Choi Y, et al. Machine learning approaches to predicting no-shows in pediatric medical appointment. *NPJ Digit Med.* 2022;5(1):50.

<https://doi.org/10.1038/s41746-022-00594-2>.

Dunstan J, Villena F, Hoyos JP, Riquelme V, Royer M, Román C, et al. Predicting no-show appointments in a pediatric hospital in Chile using machine learning. *Health Care Manag Sci.* 2023;26(2):313-29.

<https://doi.org/10.1007/s10729-022-09627-9>.

Daghistani T, AlGhamdi H, Alshammari R, AlHazme RH. Predictors of outpatients' no-show: big data analytics using Apache Spark. *J Big Data.* 2020;7(1):108.

<https://doi.org/10.1186/s40537-020-00370-3>.

Hamdan AF, Bakar AA. Machine learning predictions on outpatient no-show appointments in a Malaysia major tertiary hospital. *Malays J Med Sci.* 2023;30(5):169-84.

<https://doi.org/10.21315/mjms2023.30.5.15>.

Fan G, Deng Z, Ye Q, Wang B. Machine learning-based prediction models for patients no-show in online outpatient appointments. *Data Sci Manag.* 2021;2:45-52.

<https://doi.org/10.1016/j.dsm.2021.02.002>.

Dashtban M, Li W. Predicting non-attendance in hospital outpatient appointments using deep learning approach. *Health Syst.* 2022;11(3):189-210.

<https://doi.org/10.1080/20476965.2020.1833369>.

Peng J, Patel AD, Burch M, Rossiter S, Parker W, Rust S. Predicting patient no-shows in an academic pediatric neurology clinic. *J Child Neurol.* 2022;37(7):582-8.

<https://doi.org/10.1177/08830738221081521>.

Elkhider H, Sharma R, Sheng S, Thostenson J, Kapoor N, Kerr WT, et al. Predictors of no-show in neurology clinics. *Healthcare (Basel).* 2022;10(4):599.

<https://doi.org/10.3390/healthcare10040599>.

AlMuhaideb S, Alsweilem O, Alsubaie N, Ferwana I, Alnajem A. Prediction of hospital no-show appointments through artificial intelligence algorithms. *Ann Saudi Med.* 2019;39(6):373-81.

<https://doi.org/10.5144/0256-4947.2019.373>.

Chen J, Goldstein IH, Lin WC, Chiang MF, Hribar MR. Application of machine learning to predict patient no-shows in an academic pediatric ophthalmology clinic. *AMIA Annu Symp Proc.* 2020;2020:293-302.

Alqahtani HM, Alawaji YN. Exploring factors associated with missed dental appointments: a machine learning analysis of electronic dental records. *Cureus.* 2023;15(10):e47359.

<https://doi.org/10.7759/cureus.47359>.

Carreras-García D, Delgado-Gómez D, Llorente-Fernández F, Arribas-Gil A. Patient no-show prediction: a systematic literature review. *Entropy (Basel).* 2020;22(6):675.

<https://doi.org/10.3390/e22060675>.

Salazar LH, Parreira WD, Fernandes AM, Leithardt VR. No-show in medical appointments with machine learning techniques: a systematic literature review. *Information.* 2022;13(11):507.

<https://doi.org/10.3390/info13110507>.

Babayoff O, Shehory O, Geller S, Shitrit-Niselbaum C, Weiss-Meilik A, Sprecher E. Improving hospital outpatient clinics appointment schedules by prediction models. *J Med Syst.* 2023;47(1):5.

<https://doi.org/10.1007/s10916-022-01886-2>.

Azad N, Armstrong C, Depue C, Crimmins TJ, Tousey JC. An application of computable biomedical knowledge to transform patient centered scheduling. *Learn Health Syst.* 2023;7(4):e10393.

<https://doi.org/10.1002/lrh2.10393>.

Taheri-Shirazi M, Namdar K, Ling K, Karmali K, McCradden MD, Kelly J, et al. Exploring potential barriers in equitable access to pediatric diagnostic imaging using machine learning. *Front Public Health.* 2023;11:968319.

<https://doi.org/10.3389/fpubh.2023.968319>.

Yılmaz H, Kocyigit B. Factors associated with non-attendance at appointments in the gastroenterology endoscopy unit: a retrospective cohort study. *PeerJ.* 2022;10:e13518.

<https://doi.org/10.7717/peerj.13518>.

Liu C, Harvey HB, Jaworsky C, Shore MT, Guerrier CE, Panykh OS. Text message reminders reduce outpatient radiology no-shows but do not improve arrival punctuality. *J Am Coll Radiol.* 2017;14(8):1049-54.

<https://doi.org/10.1016/j.jacr.2017.03.019>.

Fischer HH, Moore SL, Johnson TL, Everhart RM, Batal H, Davidson AJ. Appointment reminders by text message in a safety net health care system: a pragmatic investigation. *eGEMS (Wash DC).* 2017;5(1):20.

<https://doi.org/10.5334/egems.214>.

Saeed S, Somani N, Sharif F, Kazi AM. Evaluating the effectiveness of text messaging and phone call reminders to minimize no show at pediatric outpatient clinics in Pakistan: protocol for a mixed-methods study. *JMIR Res Protoc.* 2018;7(4):e9294.

<https://doi.org/10.2196/resprot.9294>.

Moran L, O'Loughlin K, Kelly BD. The effect of SMS (text message) reminders on attendance at a community adult mental health service clinic: do SMS reminders really increase attendance? *Ir J Med Sci.* 2018;187(3):561-4.

<https://doi.org/10.1007/s11845-017-1695-2>.

Crutchfield TM, Kistler CE. Getting patients in the door: medical appointment reminder preferences. *Patient Prefer Adherence.* 2017;11:141-50.

<https://doi.org/10.2147/PPA.S117396>.

Oikonomidi T, Norman G, McGarrigle L, Stokes J, van der Veer SN, Dowding D. Predictive model-based interventions to reduce outpatient no-shows: a rapid systematic review. *J Am Med Inform Assoc.* 2023;30(3):559-69.

Clouse KM, Williams KA, Harmon JM. Improving the no-show rate of new patients in outpatient psychiatric practice: an advance practice nurse-initiated telephone engagement protocol quality improvement project. *Perspect Psychiatr Care.* 2017;53(2):127-34.

<https://doi.org/10.1111/ppc.12152>.