ORIGINAL RESEARCH **Open access**

Sequence Learning Model for Predicting Specialist Consultation Completion Delays Using Consultation Type, Patient Location, Specialty Workload, Ordering Service, Communication Logs, and Escalation History

Maria Silva[1*], Joao Pereira[1]

Abstract

Specialist consultation delays are a pervasive source of prolonged inpatient stays and disrupted throughput. They remain difficult to anticipate because delay risk emerges across ordering, communication, workload, and completion steps. A predictive model could support earlier recognition of consults likely to exceed expected completion windows. Existing consultation monitoring often depends on retrospective reports, manual tracking, or informal escalation. These approaches miss the opportunity to intervene while the consultation is still unfolding. A real-time model could convert consult workflow events into actionable delay forecasts. This article proposes a sequence learning model that predicts the probability of specialist consultation completion delay at the time of order entry. The model would refine this probability after each subsequent event, including messages, assignment, escalation, note drafting, and completion. The objective is conceptual model development rather than experimental evaluation. The proposed approach uses an LSTM-, GRU-, or Transformer-based architecture to ingest consultation milestones and static context. Inputs include consultation type, patient location, ordering service, specialty workload, communication logs, and escalation history. The output is a dynamically updated delay probability intended for consult workflow management. Conceptually, the model would identify high-risk consults early, such as a complex weekend consultation for a critically ill patient with no timely response from an overloaded service. It would be expected to support targeted escalation, workload redistribution, and proactive communication. No empirical performance claims are made. A sequence learning model could help hospitals move from passive consultation tracking to proactive delay management. By combining temporal workflow events with operational context, the model could support more timely specialist input and reduce avoidable length-of-stay pressure. Future evaluation should focus on safety, fairness, usability, and workflow impact.

Keywords Electronic health records, Clinical workflow operations, Inpatient consultation, Sequence learning, Consultation delay, Workload prediction

*Correspondence:
Maria Silva
maria.silva@gmail.com
[1] Department of Healthcare Informatics and AI Systems, Faculty of Medicine, University of Coimbra, Coimbra, Portugal

Introduction

Delayed specialist consultation completion can influence inpatient length of stay, emergency department boarding, patient throughput, and the timing of clinical decision-making. Variation in inpatient consultation practices suggests that delays are not merely administrative artifacts but reflect differences in service norms, patient complexity, and operational load [1, 2]. Emergency department studies also show that consultation and diagnostic decision delays can contribute to prolonged care episodes before disposition [3, 4]. In this context, consultation completion becomes both a clinical coordination problem and a hospital operations problem.

Current consultation management is often passive, with the requesting team placing an order and waiting for acknowledgement, evaluation, or note completion without a continuously updated estimate of completion risk. Studies of inpatient consultation practice emphasize that communication expectations, responsibility boundaries, and specialty-specific norms shape whether consultation unfolds efficiently or stalls [2, 5]. Hospitalist workload and admission-day busyness further illustrate how competing demands can alter clinical responsiveness and resource use [6, 7]. A delay prediction model would therefore need to represent consultation as a dynamic process rather than a single order timestamp.

Digitisation of consult orders, secure messages, assignment events, and note-signing timestamps creates an event stream suitable for sequence learning. Recurrent neural networks and Transformer models have been used conceptually and empirically for clinical event prediction from longitudinal electronic health record data, showing that timestamped patient histories can be encoded as evolving states [8-14]. Secure messaging research also indicates that communication content and timing can be structured into analyzable signals rather than treated as invisible workflow background [15, 16]. These developments create a foundation for modeling consultation delay as a time-varying prediction task.

The central thesis is that a sequence learning model could predict specialist consultation completion delay at order entry and update the risk estimate as the consultation unfolds. Such a model would combine static features, such as consultation type and patient location, with dynamic features, such as workload, response latency, and escalation events. Deep survival and time-to-event modeling provide a natural framework for censored or incomplete consult trajectories, where completion may not yet have occurred at prediction time. The intended use is not autonomous decision-making but proactive support for consultation tracking, escalation, and workload balancing.

Background

Inpatient specialty consultation workflows

Inpatient specialty consultation commonly begins with an electronic order, followed by notification, chart review, consultant assignment, communication between teams, bedside or remote assessment, and eventual documentation. Each step can introduce delay, particularly when responsibilities are unclear, the patient's location changes, or communication expectations differ between requesting and consulting teams [2, 5]. Pediatric and adult hospital medicine studies suggest that consultation patterns vary by physician, patient, admission context, and specialty, reinforcing the need to model workflow heterogeneity rather than assume a uniform process [1, 2]. A sequence model could represent these steps as ordered milestones and learn how different milestone patterns may indicate elevated delay risk.

Factors influencing consultation timeliness

Consultation timeliness is shaped by consultation type, patient acuity, unit location, specialty workload, ordering service, and the culture of communication between services. Hospitalist busyness and resource-use variation suggest that workload conditions can influence clinical throughput and the timing of downstream care actions [6, 7]. Emergency department consultation delays show how diagnostic testing, specialist input, and disposition decisions can interact to prolong patient flow [3, 4]. These factors imply that a consultation delay model should include both patient-context variables and operational variables reflecting the state of the hospital at the time of the request.

Communication logs and escalation pathways

Communication logs, including secure messages, pages, nurse triage notes, and phone-call documentation, can function as real-time markers of coordination quality. Secure messaging studies show that clinical communication can be characterized by timing, content, sender-recipient patterns, and response behavior, all of which could contribute to delay prediction [15, 16]. Escalation events, such as repeated messages, attending-to-attending calls, or critical laboratory triggers, may indicate that routine consultation flow has failed or that urgency has increased. A model that treats these events as part of the consultation sequence could update risk when communication becomes unusually sparse, delayed, or escalatory.

Sequence learning for clinical event prediction

Sequence learning architectures are well suited to clinical event prediction because they encode temporally ordered observations rather than isolated variables. Recurrent neural networks, multitask clinical time-series models, Transformer-based EHR representations, and deep patient-event embeddings demonstrate how timestamped health data can be mapped into evolving clinical states [8-14]. These methods are relevant for consultation delay because the meaning of an event depends on its timing, prior milestones, and surrounding clinical context. Survival-oriented deep learning extends this logic by allowing the model to estimate time-to-event risk while accounting for incomplete observation windows [17-22].

Prior work in consultation and workflow delay prediction

Prior work on consultation practice and hospital workflow has clarified the organizational sources of delay but has not fully transformed consult management into a dynamic prediction problem. Studies of consultation variability, hospitalist workload, emergency department boarding, and secure messaging provide the operational substrate for such modeling [1-7, 15, 16]. Deep learning studies in clinical event prediction and deterioration forecasting show that temporal EHR data can be used to generate continuously updated risk assessments, although consultation completion delay remains a distinct workflow target [23-25]. The proposed model would fill this gap by linking consultation-specific workflow events to a real-time probability of delayed completion.

Model Development Overview

High-level predictive pipeline

The proposed pipeline would listen to real-time EHR events, beginning with consult order placement and continuing through messages, consultant assignment, note drafting, note signing, cancellation, or escalation. After each new event, a sequence model would update a consultation delay risk score and send the current prediction to a consult tracking dashboard. This design follows the broader logic of continuous clinical prediction systems, in which risk is revised as new events accumulate rather than fixed at baseline [23-25]. For consultation workflows, this dynamic structure would be expected to make predictions more operationally useful than retrospective delay reports.

Figure 1 illustrates the proposed left-to-right sequence learning pipeline for dynamically predicting specialist consultation completion delay from order context, workflow events, operational signals, temporal encoding, risk estimation, and consult management outputs.

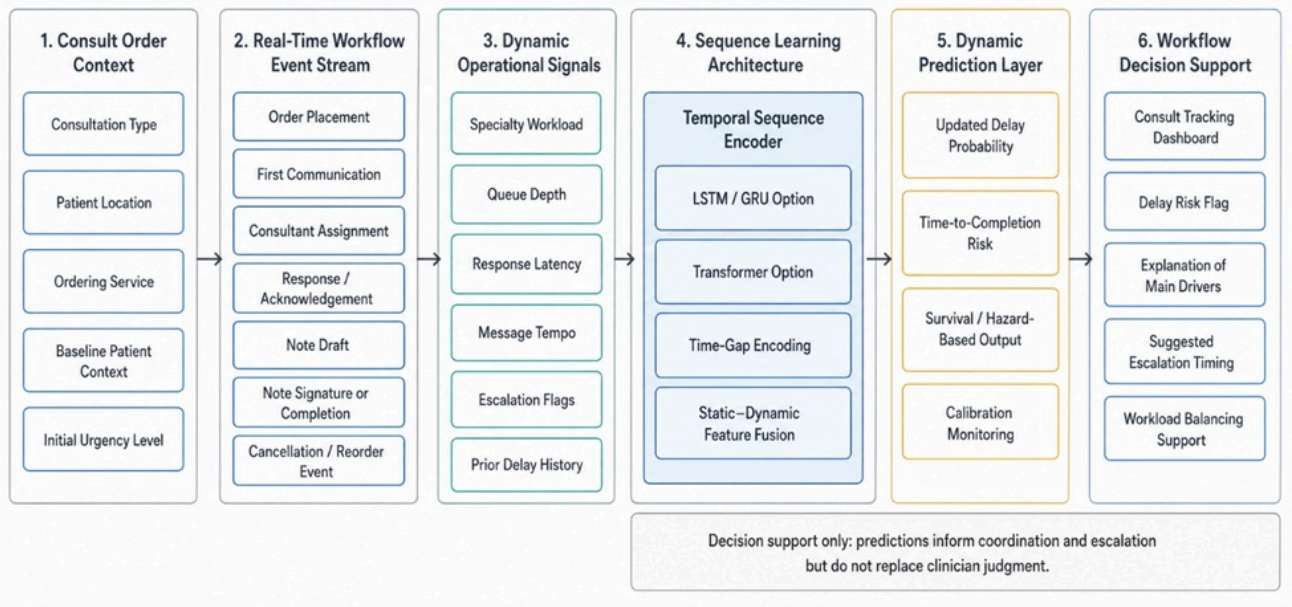


Figure 1. Dynamic Sequence Learning Pipeline for Predicting Specialist Consultation Completion Delay

Core input features

Core input features would include consultation type, patient location, specialty workload, ordering service identifier, communication log attributes, and escalation history. Consultation type and patient location would capture baseline complexity and urgency, while ordering service and workload would reflect operational context and service-specific patterns [1-7]. Communication features, including message count, response latency, and acknowledgement timing, would represent whether the consult is progressing or becoming stagnant [15, 16]. Escalation history would provide a structured signal that the current request may already be deviating from routine workflow expectations.

Design principles

The model should be real-time, dynamically updating, interpretable, and embedded within existing consult management workflows without requiring additional manual data entry. This design principle is consistent with prior clinical machine learning work emphasizing that predictive models must fit the timing and context of clinical action rather than merely generate retrospective labels [23-26]. Interpretability is particularly important because hospitalists and consultants would need to understand whether risk is driven by workload, lack of response, patient location, or repeated escalation. The model should therefore support shared situational awareness rather than replace clinical judgment.

Data Sources and Feature Engineering

Extracting the consultation event sequence

The consultation event sequence would be extracted from timestamped EHR data, including order time, consultation type, patient location, first communication, consultant assignment, note draft, note signature, cancellation, and escalation events. This representation builds on the idea that clinical histories can be modeled as ordered event streams, where each timestamped observation changes the estimated state of the patient or workflow [8-14]. In consultation workflows, the same logic applies to operational state: a message without response, an assignment without documentation, or a note draft without signature may signal different delay trajectories. The event sequence should therefore preserve both order and elapsed time between milestones.

Static and dynamic features

Static features would include consultation type, patient unit, ordering service, and baseline patient context available at order entry. Dynamic features would include rolling specialty workload, queue depth, communication count, message tempo, response latency, and escalation flags that change as the consultation unfolds [6, 7, 15, 16]. Clinical time-series modeling supports this distinction between baseline covariates and evolving observations, because each new event can revise the representation of the current state [12-14, 27, 28]. For consult delay prediction, this separation would allow the model to produce an initial estimate and then refine it as workflow evidence accumulates.

Table 1 provides an analytical mapping between consultation delay signals, their sequence-modeling function, and their operational interpretation within inpatient consult management.

Table 1. Analytical Mapping of Consultation Delay Signals to Model Functions and Workflow Interpretation

Signal domain	Example features	Model function	Workflow interpretation	Actionable implication
Consult order context	Consultation type, urgency, ordering service, patient unit	Establishes baseline delay risk at order entry	Some consults begin with structurally higher completion complexity	Early triage before delay becomes visible
Patient location	ICU, ED boarding area, ward, procedural unit	Captures location-specific urgency and access constraints	Location may alter consultant prioritization and communication pathways	Location-aware escalation thresholds
Specialty workload	Active consult census, queue depth, recent completion tempo	Updates risk as operational pressure changes	Delay may reflect service congestion rather than individual inattention	Queue redistribution or attending-level review
Communication tempo	Message count, response latency, acknowledgement timing	Detects whether consult progression is active or stagnant	Sparse or delayed response suggests emerging coordination failure	Proactive outreach before missed completion window
Escalation history	Repeated pages, urgent re-consultation, attending-to-attending contact	Identifies deviation from routine workflow	Escalation may indicate clinical urgency, workflow friction, or both	Structured escalation pathway rather than ad hoc follow-up
Documentation milestones	Note draft, note signature, verbal recommendation proxy	Distinguishes clinical completion from administrative closure	Signed note may lag behind clinical recommendation	Avoid mislabeling documentation delay as clinical delay
Temporal spacing	Time since last event, time of day, weekend/after-hours status	Preserves irregular timing between consult milestones	The same event may carry different meaning depending on timing	Time-sensitive dashboard prioritization

Label construction: defining a completion delay

Completion delay should be defined using clinically informed thresholds that distinguish routine, urgent, and emergent consultation expectations. A survival-oriented framework would allow the model to handle consultations that remain

incomplete at the time of prediction, are cancelled, or are reordered as separate outcomes [17-22]. The label should also recognize that note signature may not always equal clinical completion, especially when verbal recommendations precede documentation [2, 5]. For this reason, institutions should evaluate multiple operational definitions of completion while avoiding labels that reward delayed documentation or penalize clinically appropriate sequencing.

Sequence Learning Architecture

Input representation

Each consultation would be represented as a sequence of event vectors, with each vector combining event type, time since last event, time of day, service context, and relevant communication attributes. Static context, including consultation type, patient location, and ordering service, would be encoded separately and fused with the event sequence so that early predictions can be made before many milestones occur [8-14]. This structure would allow the model to distinguish, for example, an expected overnight waiting period from an unexpected lack of response during a high-priority daytime consult. The representation should preserve temporal irregularity because consultation events occur at uneven intervals rather than fixed measurement times.

Sequence encoder

The sequence encoder could use an LSTM, GRU, temporal convolutional structure, or Transformer to process the evolving consultation event stream. Recurrent models are conceptually attractive for ordered clinical histories, while Transformer-based encoders can represent longer dependencies and use positional information to capture temporal spacing [8, 12, 13]. Multichannel and contrastive approaches to medical event prediction also suggest that heterogeneous event types can be integrated into a shared clinical or workflow representation [27, 28]. In this application, the encoder would output the current latent state of the consult, reflecting what has happened so far and what delay risk may be emerging.

Table 2 compares candidate sequence learning and survival-oriented architectures according to their suitability for dynamic consultation delay prediction.

Table 2. Conceptual Comparison of Candidate Sequence Learning Architectures for Consultation Delay Prediction

Architecture	Strength for consult-delay modeling	Limitation	Best-fit use case	Interpretability consideration
LSTM	Captures ordered consult milestones and evolving workflow state	May struggle with long-range dependencies and highly irregular gaps	Moderate-length consult sequences with clear milestone progression	Hidden-state explanations require feature attribution or milestone summaries
GRU	Efficient recurrent alternative with fewer parameters	Less expressive than larger temporal architectures	Real-time deployment where computational simplicity matters	Easier to operationalize but still requires explanation layer
Temporal convolutional model	Handles local temporal patterns and short event windows	Less naturally suited to variable-length clinical narratives	Detecting short bursts of delayed communication or repeated escalation	Can highlight influential temporal windows
Transformer encoder	Represents long-range dependencies and heterogeneous events	Requires larger datasets and careful calibration	Complex consult trajectories with multiple messages, delays, and escalations	Attention summaries may help but should not be treated as complete explanation

Deep survival model	Directly models time-to-completion and censoring	Requires careful definition of event endpoint and censoring rules	Consults unresolved at prediction time, cancelled, or still active	Hazard-based outputs must be translated into clinically meaningful risk windows
Hybrid sequence-survival model	Combines temporal event encoding with time-to-event prediction	More complex to validate and explain	Dynamic prediction that updates after each workflow milestone	Requires both statistical calibration and action-oriented explanations

Output layer and dynamic risk score

The output layer would translate the current encoded consult state into a dynamically updated delay probability. A survival-oriented head, such as a Cox-style or discrete-time hazard formulation, would be appropriate because consultation completion is a time-to-event outcome and some requests may be censored or unresolved at prediction time [17-22]. The risk score would update whenever a new milestone occurs, such as consultant assignment, response message, escalation, note draft, or signature. This design would allow the model to support ongoing consultation management rather than producing a single static prediction at order entry.

Incorporating Workload, Communication, and Escalation Signals

Specialty workload and its temporal influence

Specialty workload would be represented as a time-varying feature reflecting active consult census, queue depth, recent completion tempo, and competing clinical demand. Hospitalist busyness and resource-use variation suggest that operational load can shape patient throughput and care timing, making workload a central signal for delay prediction [6, 7]. Emergency department consultation delays further indicate that service congestion and decision bottlenecks can propagate across the hospital [3, 4]. A dynamically updated workload feature would allow the model to adjust risk when a consulting service becomes overloaded during the life of an active consult.

Communication log features as real-time signals

Communication log features would include message count, sender role, acknowledgement timing, time since last response, repeated pages, and escalation language when available. Secure messaging studies show that clinical communication patterns can be characterized from electronic records, supporting their use as structured workflow signals [15, 16]. A lack of acknowledgement after the initial consult message could increase predicted delay risk, while a rapid consultant response could reduce it. These features would help the model distinguish a consult that is simply waiting from one that is actively progressing.

Escalation history and repeat consults

Escalation history would capture whether the patient has had repeated recent consults, prior delayed consults, urgent re-consultation, or documented attending-level involvement. Variation in consultation practice and expectations suggests that escalation can reflect both clinical complexity and coordination difficulty [1, 2, 5]. In sequence form, an escalation event would not be treated as a static label but as a temporal signal that changes the current consult state. The model could therefore learn that certain escalation patterns are compatible with high urgency, workflow friction, or both.

Model Interpretability for Clinical Stakeholders

Explaining delay predictions to care teams

Interpretability should help care teams understand why a consult is being flagged rather than merely displaying a risk score. Attention-based summaries, feature-attribution methods, or milestone-level explanations could identify whether risk is driven by workload, lack of response, patient location, ordering service, or repeated escalation [13, 26]. Prior clinical prediction work underscores that deployment requires explanations aligned with action, because clinicians need to know what can be changed after a warning appears [23-25]. For consultation management, useful explanations would translate model output into operational next steps.

Consultant and hospitalist views

A consult-tracking dashboard would present different but aligned views for hospitalists, consultants, and operational leaders. Hospitalists could see which consults are at risk of delay and why, while consulting services could see queue pressure and backlog in clinical context [2, 6, 7]. This shared display would be expected to reduce ambiguity about whether delay risk reflects workload, missing communication, or unresolved escalation. The model's purpose would be to support coordination between teams, not to assign blame to individual clinicians.

Operational Integration and Proactive Delay Management

Real-time consult dashboard and alerts

The model would populate a live consultation dashboard with dynamically updated delay risk, recent milestones, communication status, and suggested escalation timing. Continuous prediction systems in clinical care demonstrate the conceptual value of updating risk as new data arrive, especially when alerts can be linked to timely action [23-25]. In a consultation workflow, an alert could be triggered when risk exceeds a locally defined threshold and when the explanation suggests a modifiable bottleneck. Alert design should avoid excessive interruption by focusing on consults where proactive outreach is clinically meaningful.

Proactive workload balancing and escalation protocols

High-risk consults on overloaded services could be reviewed for prioritization, attending-level escalation, redistribution of queue responsibilities, or alternative specialty routing when clinically appropriate. Evidence on hospitalist workload, consultation variability, and emergency department delays suggests that operational context can influence timeliness and downstream throughput [1, 3, 4, 6, 7]. The model would support proactive workflow management by identifying consults likely to become delayed before the delay becomes clinically entrenched. Escalation protocols should remain governed by clinical judgment and institutional policy.

Evaluation Strategy

Predictive accuracy metrics

The model should be evaluated using time-to-event and dynamic prediction metrics appropriate for consultation completion delay. Candidate metrics include time-dependent discrimination, concordance-based survival assessment, and calibration of predicted delay probability at clinically meaningful time points [17-22]. Evaluation should focus on whether predictions remain reliable as additional consultation events accumulate. No single metric would be sufficient, because operational usefulness depends on both statistical validity and clinical interpretability.

Temporal validation and prospective testing

Temporal validation should train on earlier historical periods and evaluate on later periods to reflect real-world drift in staffing, messaging practices, service structure, and consultation culture. Silent prospective testing would allow

comparison of predicted delay risk with actual consult progression without initially altering clinical behavior [23-26]. This phase should examine whether predictions remain stable across specialties, patient locations, ordering services, and workload conditions. The model should be evaluated for calibration drift before any active alerting is introduced.

Clinical and operational impact

Clinical and operational evaluation should examine whether model-informed workflows could reduce avoidable consultation delay, improve throughput, and support hospitalist-consultant coordination. Outcomes should include consult completion timeliness, length-of-stay pressure, escalation appropriateness, and user trust in the dashboard [1-7]. Qualitative assessment would also be important because consultation delays often arise from service expectations, communication norms, and informal coordination patterns [2, 5]. The goal should be to determine whether the model changes workflow in a safe, acceptable, and equitable manner.

Table 3 presents a governance and evaluation framework for determining whether dynamic consultation delay prediction can be implemented safely, fairly, and usefully in clinical workflow operations.

Table 3. Governance and Evaluation Framework for Safe Deployment of Dynamic Consultation Delay Prediction

Evaluation dimension	Core question	Recommended assessment	Risk if neglected	Governance implication
Predictive validity	Does the model accurately estimate delay risk over time?	Time-dependent discrimination, calibration, survival metrics	Misleading risk scores may trigger unnecessary escalation	Require temporal validation before deployment
Calibration drift	Do predictions remain reliable as staffing and workflows change?	Periodic recalibration by specialty, unit, and time period	Model may become unreliable after workflow redesign	Establish scheduled monitoring and retraining criteria
Fairness	Are some services, units, or patient groups over-flagged?	Subgroup calibration and error analysis	Alerts may reinforce operational inequities	Require fairness review before active alerting
Workflow usability	Do clinicians understand and use the risk output appropriately?	Silent testing, usability interviews, dashboard observation	Alerts may be ignored or misinterpreted	Co-design dashboard with hospitalists and consultants
Explanation quality	Are risk drivers actionable and clinically meaningful?	Review of milestone-level and feature-level explanations	Clinicians may distrust unexplained warnings	Display risk drivers linked to possible next steps
Alert burden	Does the system increase interruption or escalation fatigue?	Alert volume, override rates, user feedback	Excessive alerts may worsen workflow burden	Use threshold governance and escalation rules
Clinical impact	Does prediction improve coordination or reduce avoidable delay?	Prospective impact study, consult completion time, LOS pressure	Statistically accurate model may not improve care	Deploy only if operational benefit is demonstrated
Accountability	Who acts on flagged consults?	Role-based workflow protocol	Alerts may create blame or ambiguity	Define responsibilities before implementation

Limitations

Data granularity and missing events

The model would be limited by incomplete capture of real clinical communication, especially phone calls, bedside discussions, informal hallway conversations, and verbal recommendations not reflected in electronic logs. Secure messaging data can characterize some communication patterns, but they may not fully represent the consultation relationship or the actual timing of clinical decision-making [15, 16]. Note signature may also occur after clinical completion, creating potential misclassification if documentation is treated as the only completion endpoint [2, 5]. These limitations require careful local validation of event definitions before operational use.

Service-specific workflows and generalizability

Consultation culture, staffing models, escalation norms, and documentation practices vary substantially across hospitals and specialties. Prior work on consultation variability and hospitalist resource use suggests that a model trained in one setting may not transfer cleanly to another without recalibration or retraining [1, 2, 6, 7]. Deep learning models using EHR sequences can be sensitive to institutional coding, timestamp conventions, and workflow design [10, 11, 26]. Generalizability should therefore be treated as an empirical question rather than assumed from model architecture alone.

Conclusion

A sequence learning model for specialist consultation completion delay could transform consultation management from retrospective tracking into dynamic prediction. By representing the consult as a sequence of workflow events, the model could update risk as new milestones occur. This approach is especially suited to a process in which delay risk emerges gradually through workload, communication, and escalation patterns.

The key strength of the proposed model is its integration of static context with evolving operational signals. Consultation type, patient location, ordering service, specialty workload, communication tempo, and escalation history could jointly support a more complete picture of delay risk. Interpretable predictions would help clinicians understand why a consult is being flagged and what action might be appropriate.

Important challenges remain before such a model could be safely implemented. Data completeness, undocumented communication, service-specific norms, and shifting hospital operations could all affect reliability. The model would also need governance to ensure that alerts support collaboration rather than increase friction between requesting and consulting teams.

Future work should pursue multi-site validation, silent prospective testing, usability assessment, and integration into hospitalist workflow tools. The central test should be whether dynamic consultation delay prediction leads to earlier action, better coordination, and reduced avoidable length-of-stay pressure. A carefully designed model could become a practical component of clinical workflow operations.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 22 Oct 2023 Revised: 23 Nov 2023 Accepted: 16 Dec 2023
Published online: 25 February 2024

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

Kern-Goldberger AS, Dalton EM, Rasooly IR, Congdon M, Gunturi D, Wu L, et al. Factors associated with inpatient subspecialty consultation patterns among pediatric hospitalists. JAMA Netw Open. 2023;6(3):e232648. https://doi.org/10.1001/jamanetworkopen.2023.2648.

Kachman M, Carter K, Arora VM, Flores A, Meltzer DO, Martin SK. Describing variability of inpatient consultation practices: physician, patient, and admission factors. J Hosp Med. 2020;15(3):164-8. https://doi.org/10.12788/jhm.3344.

Ataman MG, Sariyer G, Saglam C, Karagoz A, Unluer EE. Factors relating to decision delay in the emergency department: effects of diagnostic tests and consultations. Open Access Emerg Med. 2023;15:119-31. https://doi.org/10.2147/OAEM.S398225.

Pearlmutter MD, Dwyer KH, Burke LG, Rathlev N, Maranda L, Volturo G. Analysis of emergency department length of stay for mental health patients at ten Massachusetts emergency departments. Ann Emerg Med. 2017;70(2):193-202.e16. https://doi.org/10.1016/j.annemergmed.2016.10.002.

Pavitt S, Bogetz A, Blankenburg R. What makes the “perfect” inpatient consultation? A qualitative analysis of resident and fellow perspectives. Acad Med. 2020;95(1):104-10. https://doi.org/10.1097/ACM.0000000000002944.

Stevens JP, Hatfield LA, Nyweide DJ, Landon B. Comparison of health outcomes among patients admitted on busy vs less busy days for hospitalists. JAMA Netw Open. 2022;5(1):e2144261. https://doi.org/10.1001/jamanetworkopen.2021.44261.

Stevens JP, Nyweide DJ, Maresh S, Hatfield LA, Howell MD, Landon BE. Comparison of hospital resource use and outcomes among hospitalists, primary care physicians, and other generalists. JAMA Intern Med. 2017;177(12):1781-7. https://doi.org/10.1001/jamainternmed.2017.5281.

Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc. 2017;24(2):361-70.

Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1(1):18.
https://doi.org/10.1038/s41746-018-0029-1.

Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform. 2018;22(5):1589-604.
https://doi.org/10.1109/JBHI.2017.2767063.

Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc. 2018;25(10):1419-28.

Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Sci Data. 2019;6(1):96.
https://doi.org/10.1038/s41597-019-0103-9.

Li Y, Rao S, Solares JR, Hassaine A, Ramakrishnan R, Canoy D, et al. BEHRT: transformer for electronic health records. Sci Rep. 2020;10(1):7155.
https://doi.org/10.1038/s41598-020-62922-y.

Si Y, Du J, Li Z, Jiang X, Miller T, Wang F, et al. Deep representation learning of patient data from electronic health records (EHR): a systematic review. J Biomed Inform. 2021;115:103671.
https://doi.org/10.1016/j.jbi.2021.103671.

Liu X, Sutton PR, McKenna R, Sinanan MN, Fellner BJ, Leu MG, et al. Evaluation of secure messaging applications for a health care system: a case study. Appl Clin Inform. 2019;10(1):140-50.
https://doi.org/10.1055/s-0038-1677016.

Huang M, Fan J, Prigge J, Shah ND, Costello BA, Yao L. Characterizing patient-clinician communication in secure medical messages: retrospective study. J Med Internet Res. 2022;24(1):e17273.
https://doi.org/10.2196/17273.

Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018;18(1):24.
https://doi.org/10.1186/s12874-018-0482-1.

Lee C, Zame W, Yoon J, Van Der Schaar M. Deephit: a deep learning approach to survival analysis with competing risks. Proc AAAI Conf Artif Intell. 2018;32(1):2314-21.
https://doi.org/10.1609/aaai.v32i1.11842.

Kvamme H, Borgan Ø, Scheel I. Time-to-event prediction with neural networks and Cox regression. J Mach Learn Res. 2019;20(129):1-30.

Nagpal C, Li X, Dubrawski A. Deep survival machines: fully parametric survival regression and representation learning for censored data with competing risks. IEEE J Biomed Health Inform. 2021;25(8):3163-75.
https://doi.org/10.1109/JBHI.2021.3052441.

Ren K, Qin J, Zheng L, Yang Z, Zhang W, Qiu L, et al. Deep recurrent survival analysis. Proc AAAI Conf Artif Intell. 2019;33(1):4798-805.
https://doi.org/10.1609/aaai.v33i01.33014798.

Wang P, Li Y, Reddy CK. Machine learning for survival analysis: a survey. ACM Comput Surv. 2019;51(6):1-36. https://doi.org/10.1145/3214306.

Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. Crit Care Med. 2019;47(11):1485-92. https://doi.org/10.1097/CCM.0000000000003891.

Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature. 2019;572(7767):116-9. https://doi.org/10.1038/s41586-019-1390-1.

Shamout FE, Zhu T, Sharma P, Watkinson PJ, Clifton DA. Deep interpretable early warning system for the detection of clinical deterioration. IEEE J Biomed Health Inform. 2020;24(2):437-46. https://doi.org/10.1109/JBHI.2019.2938387.

Pungitore S, Subbian V. Assessment of prediction tasks and time window selection in temporal modeling of electronic health record data: a systematic review. J Healthc Inform Res. 2023;7(3):313-31. https://doi.org/10.1007/s41666-023-00139-3.

Liu S, Wang X, Xiang Y, Xu H, Wang H, Tang B. Multi-channel fusion LSTM for medical event prediction using EHRs. J Biomed Inform. 2022;127:104011. https://doi.org/10.1016/j.jbi.2022.104011.

Zhang Z, Yan C, Zhang X, Nyemba SL, Malin BA. Forecasting the future clinical events of a patient through contrastive learning. J Am Med Inform Assoc. 2022;29(9):1584-92.