

ORIGINAL RESEARCH

Open access

Deep Neural Network for Detecting Physician Documentation Burden Using Note Length, Time in Electronic Health Record, After-Hours Charting Activity, Inbox Volume, and Order Entry Patterns

Carlos Ramirez^{1*}, Elena Torres², Pablo Ortega¹, Sofia Mendes³

Abstract

Documentation burden is a leading contributor to physician burnout, yet detection often depends on periodic self-report surveys. Electronic health record audit logs provide an objective and continuous record of clinical work patterns that may reveal burden before physicians formally report distress. Moving from reactive survey assessment to proactive detection requires transforming complex, high-dimensional audit log signals into meaningful burden classifications. These signals must be modeled in a way that reflects workload intensity, temporal accumulation, and the interaction of documentation, inbox, and order-entry demands. This article proposes a conceptual deep neural network model for detecting physicians with high documentation burden. The model uses note length, time in the electronic health record, after-hours charting activity, inbox volume, and order entry patterns as core input domains. The proposed model uses a multi-input neural architecture that fuses aggregated and temporally aware features derived from raw electronic health record audit logs. The model would generate a burden probability score for each physician over a defined weekly period without requiring direct linkage to individual patient content. Conceptually, the model could identify physicians with rising documentation burden earlier than survey-based approaches. It would also be expected to reveal the dominant burden component, such as excessive inbox work, prolonged after-hours charting, unusually long notes, or high-complexity order entry. A deep learning model for documentation burden detection could help health systems move from burnout treatment to prevention. By connecting objective workload signals to targeted operational interventions, such a model could support physician well-being while preserving privacy and professional trust.

Keywords Deep learning, Physician burnout, Documentation burden, Electronic health record audit logs, After-hours charting, Inbox workload

*Correspondence:

Carlos Ramirez

carlos.ramirez@gmail.com

¹ Department of Health Informatics and AI Applications, Faculty of Medicine, University of Barcelona, Barcelona, Spain

² Department of Digital Clinical Analytics, Faculty of Engineering, University of Lisbon, Lisbon, Portugal

³ Department of Smart Health Systems, Faculty of Medicine, University of Porto, Porto, Portugal

Introduction

Physician burnout has become a central challenge for health systems because it is closely linked to emotional exhaustion, reduced professional satisfaction, and potentially adverse consequences for care delivery. Documentation burden is one of the most visible operational drivers of this problem, with time-motion evidence showing that physicians can remain tethered to electronic health record work beyond the clinical encounter [1]. EHR usability has also been

associated with professional burnout, suggesting that documentation systems are not merely administrative tools but part of the working environment that shapes physician well-being [2]. Periodic survey instruments remain valuable for measuring subjective distress, but they are retrospective and may miss rapidly worsening workload conditions that accumulate between survey cycles [3].

Electronic health record audit logs provide a granular behavioral data stream that can describe when physicians open notes, sign encounters, enter orders, use the inbox, and continue work after formal clinic hours. Standardized EHR activity metrics have been proposed to make these logs interpretable for physician workload assessment [4], while scoping work on clinical documentation burden has emphasized the value of measuring burden through routinely collected EHR data rather than relying only on self-report [5]. Audit logs can be structured at daily or weekly intervals, enabling detection of trajectories in documentation work rather than isolated impressions. Because these logs capture real-time interaction patterns, they offer a foundation for proactive model-based detection of physicians whose work patterns suggest rising burden [6].

Simple workload indicators, such as average note length or total after-hours time, are useful but insufficient when used in isolation. A physician with long notes but low inbox volume may experience a different burden profile from a physician whose notes are short but whose inbox work extends into evenings, and national studies show that after-hours EHR time varies substantially across ambulatory specialties [7]. Note composition patterns differ across specialty groups [8], and gender differences in documentation and EHR time indicate that burden measurement must account for contextual variation rather than apply universal thresholds [9]. A deep neural network is therefore justified because it can represent interactions among note length, EHR time, after-hours work, inbox demand, and ordering activity.

This article proposes a conceptual deep neural network for detecting physician documentation burden from multi-domain EHR audit log features. Clinical activity logs have already been used to support burnout prediction models, and temporal associations between EHR-derived workload and burnout-related outcomes suggest that longitudinal log patterns can carry meaningful well-being signals. The model proposed here is designed for private feedback, early intervention, and health-system learning rather than punitive productivity surveillance. By combining audit log analytics, machine learning, and human-centered deployment principles, the approach could enable individualized and privacy-preserving burden monitoring at scale while remaining aligned with organizational support for clinician well-being.

Background

Documentation burden and physician burnout

Documentation burden links EHR design, regulatory expectations, billing requirements, and clinical communication into a cumulative workload that can extend beyond scheduled clinical time. Primary care studies using EHR event log data show how much physician work can occur inside the record itself [1], and research on the timing of EHR use indicates that work outside standard hours is an important marker of documentation pressure [10]. Burnout is not caused by documentation alone, but inefficient documentation workflows can intensify emotional exhaustion by reducing perceived control over work and increasing administrative load. Studies connecting EHR usability, after-hours charting, and organizational support to burnout reinforce the need for detection systems that identify burden before it develops into sustained distress [2, 11, 12].

EHR audit log data and its potential

EHR audit logs capture timestamped user interactions with the system, including chart review, note editing, order entry, inbox use, and related event categories that can be mapped into clinical work patterns. The development of shared metrics for physician activity using EHR log data has made it possible to compare time-based workload measures across settings [4], and audit log approaches have also been used to assess broader changes in ambulatory EHR work during system shocks such as the COVID-19 pandemic [6]. Because audit logs include user identifiers, event types, timestamps,

and session information, they can support physician-level reconstruction of work intensity while avoiding unnecessary exposure of note content. However, interoperability limitations and vendor-specific logging practices can affect data completeness, making careful preprocessing and validation essential [13].

Existing metrics for documentation workload

Existing workload metrics include note length, note composition time, total EHR time, after-hours charting, inbox volume, and order-entry frequency. Note length and redundancy are important because outpatient progress notes have grown longer and more repetitive over time [14], while note bloat can affect downstream clinical NLP models and obscure meaningful content density [15]. Inbox metrics add a distinct dimension because message characteristics have been associated with physician burnout [16], and primary care physicians vary in inbox message volume in ways that may reflect panel, workflow, and organizational factors [17]. Alert-opening behavior and time-sensitive inbox activity further show that electronic communication work can create cognitive demand that is not captured by note metrics alone [18].

Machine learning for burnout and workload prediction

Machine learning offers a framework for detecting complex workload patterns that would be difficult to summarize using single thresholds. Clinical activity logs have been used to predict physician burnout [19], and machine learning has also been applied to classify clinical work settings from EHR audit logs [20]. These studies suggest that audit log features can contain predictive signals, but most existing approaches do not fully integrate note length, EHR time, after-hours activity, inbox load, and order-entry behavior into a unified burden model. A multi-input deep learning approach could address this gap by learning nonlinear interactions among workload domains while still producing interpretable summaries for physicians and operational leaders.

Deep learning for sequential human-computer interaction data

Deep learning is well suited to sequential human-computer interaction data because it can represent event order, temporal dependency, and repeated behavioral motifs across time. In the EHR context, physician stress during inbox work has been studied using in situ sensing methods [21], reinforcing the idea that interaction patterns can reveal cognitive and emotional load. A physician's burden profile may emerge from recurring sequences such as morning inbox triage, daytime order entry, evening note completion, and weekend chart closure. Recurrent networks, temporal convolutional networks, and attention-based encoders could therefore be adapted to weekly physician feature sequences to detect both persistent overload and accelerating workload trajectories [20, 22].

Model Development Overview

Surveillance philosophy and burden definition

The proposed model conceptualizes documentation burden as a latent construct that is not directly observed but is expressed through elevated, persistent, and interacting work patterns across multiple EHR domains. Rather than defining burden as equivalent to burnout, the model treats burden as an upstream operational state that may increase the risk of emotional exhaustion or reduced professional satisfaction if it remains unaddressed. EHR-based audit log data have been linked to burnout and clinical practice process measures [3], while after-hours charting has been examined as one organizationally relevant expression of EHR burden [11]. The five core signal domains are note length, time in the EHR, after-hours charting activity, inbox volume, and order-entry patterns, each representing a different mechanism through which documentation work becomes excessive. The operational definition would classify a physician as having high documentation burden when a composite pattern exceeds specialty-adjusted and physician-specific expectations for a sustained period. This sustained-period definition avoids overreacting to isolated acute spikes that may reflect temporary coverage problems, seasonal demand, or unusual patient complexity. The model should therefore be interpreted as an early-warning system for documentation strain, not as a diagnostic instrument for burnout itself.

High-level detection pipeline

The detection pipeline would begin with ingestion of raw EHR audit log events and extraction of timestamped activities relevant to documentation, messaging, and order entry. These events would be filtered to physician-role users and transformed into weekly physician-level representations that preserve both aggregate workload and temporal structure, consistent with prior metric-development work using EHR log data [4]. Feature extraction would summarize note length, EHR duration, after-hours activity, inbox load, and order-entry complexity before passing each feature domain into a dedicated neural input branch. The multi-input neural network would then fuse domain-specific embeddings and generate a physician-week burden probability score with an associated confidence estimate. The score could be categorized into low, moderate, or high burden for private feedback and operational triage, while continuous probabilities would remain available for trend monitoring. The pipeline should run on a weekly cadence because weekly aggregation balances operational timeliness with enough data density to reduce false alarms from isolated workdays. To preserve privacy, the pipeline can operate without exposing individual patient data, using event metadata and derived features rather than clinical note content whenever possible. This design follows the logic that audit log analytics can support physician activity measurement while limiting unnecessary access to sensitive clinical information [5, 6].

Figure 1 illustrates the hierarchical multi-input deep neural network pipeline for transforming EHR audit log signals into physician-level documentation burden predictions.

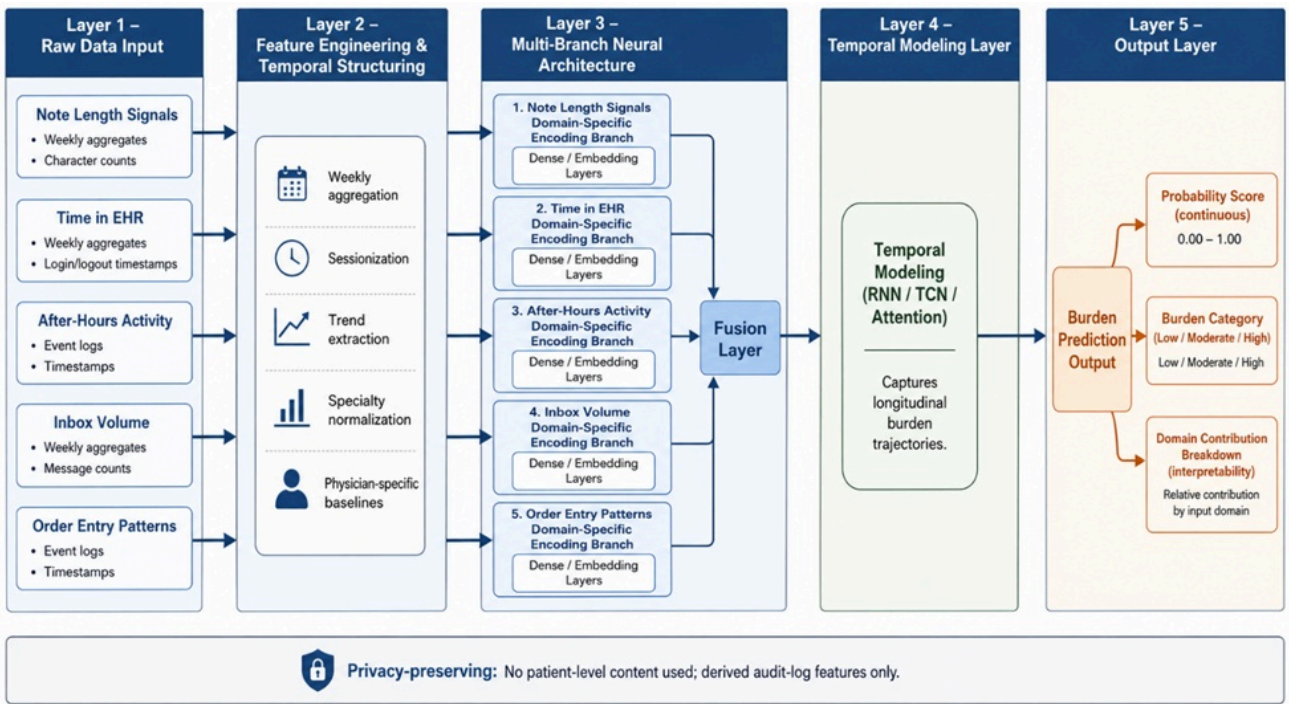


Figure 1. Hierarchical Multi-Input Deep Neural Network Pipeline for Early Detection of Physician Documentation Burden from EHR Audit Logs

Core input signals and their rationale

Note length is included because unusually long, redundant, or rapidly expanding notes may reflect note bloat, templating, copy-forward behavior, or excessive documentation requirements. Evidence that progress notes can become longer and more redundant supports the inclusion of note length and redundancy-sensitive features [14], while work on note bloat and NLP suggests that non-substantive note expansion can affect clinical information processing [15]. Time in the EHR captures the global burden of interacting with the system, including documentation, review, order entry, and message management. After-hours charting is a distinct signal because national ambulatory data show that EHR work outside

standard hours varies meaningfully across specialties [7]. Inbox volume is included because message characteristics have been associated with physician burnout [16], and primary care inbox volume can vary in ways that may reflect workload distribution and practice organization [17]. Order-entry patterns are included because high order volume, frequent order modification, and manual entry rather than streamlined order sets may indicate cognitive and interaction complexity. The rationale for integrating these signals is that each represents a partially independent burden pathway, and their interaction may be more informative than any single metric. For example, high inbox volume combined with rising after-hours EHR time may suggest a different intervention than long notes combined with high template fraction.

Table 1 provides an analytical decomposition of documentation burden signals into distinct operational mechanisms and corresponding intervention pathways.

Table 1. Conceptual Decomposition of Documentation Burden Signals into Distinct Operational Mechanisms and Intervention Targets

Signal Domain	Underlying Operational Mechanism	Latent Burden Pathway	Distinguishing Feature Patterns	Primary Risk Interpretation	Targeted Intervention Levers
Note Length	Documentation expansion, templating, redundancy	Cognitive overload from excessive narrative construction	Long-tail note distributions, high variance, rapid growth trends	Inefficient documentation workflows or regulatory burden	Template redesign, documentation coaching, NLP-assisted summarization
Time in EHR	Total system interaction load	Global workload accumulation	High total duration, persistent elevation across weeks	Sustained operational overload independent of specific task	Workflow optimization, schedule redesign, team documentation support
After-Hours Activity	Work outside scheduled clinical time	Temporal spillover burden	High after-hours proportion, late-night clustering, weekend activity	Loss of work-life boundaries, delayed task completion	Inbox redistribution, protected time policies, staffing adjustments
Inbox Volume	Asynchronous communication demand	Fragmented cognitive workload	High incoming volume, rapid response cycles, message backlog	Interrupt-driven workload and attention fragmentation	Team-based inbox triage, message routing protocols, automation
Order Entry Patterns	Clinical decision and interaction complexity	Task-switching and cognitive burden	Frequent modifications, high diversity of orders, manual entry patterns	Workflow inefficiency or high clinical complexity	Order set optimization, decision support redesign, delegation models

Design principles

The model should be physician-centered, non-punitive, and explicitly framed as a tool for workload support rather than productivity surveillance. Its purpose would be to identify modifiable documentation pressures and trigger supportive interventions, not to judge individual performance. Computational efficiency is important because health systems may need to run the model weekly across thousands of physicians, multiple specialties, and several practice settings. Interpretability must be built into the design so physicians can understand whether their burden score is driven primarily

by note length, after-hours work, inbox load, or order-entry complexity. Specialty-specific normalization is essential because system-level studies show that EHR time varies across organizational and clinical contexts [23], and physician note composition patterns also differ across specialty types [8]. The model should also be resilient to missing data, part-time schedules, leave, and fluctuations caused by rotation changes. Governance safeguards should define who can see individual scores, how aggregate dashboards are used, and how physicians can contest or contextualize signals. These design principles respond to evidence that EHR workload and burnout are shaped not only by individual behavior but also by usability, organizational support, and practice setting [2, 12].

Expected output and actionable insight

For each physician, the model would output a weekly burden probability score, a categorical burden level, and a decomposition of the major contributing signal domains. A physician-facing dashboard could show whether the current week's burden appears to be driven mainly by longer notes, more after-hours charting, elevated inbox messages, increased order-entry complexity, or a combination of these factors. A leadership-facing dashboard would use de-identified and aggregated outputs to identify clinics, specialties, or service lines with rising burden trends while avoiding exposure of individual scores to managers. This separation between private individual feedback and aggregate organizational insight is central to maintaining trust, especially because physicians may reasonably worry that EHR-based burden measurement could become surveillance rather than support. Evidence on organizational EHR support and after-hours charting suggests that interventions must address the surrounding work system, not only individual behavior [11]. The model would be most useful when its outputs are linked to concrete operational responses, such as inbox triage, documentation redesign, team-based order support, or targeted EHR training. Scoping work on electronic medical record-related burnout interventions further supports connecting measurement to practical burden-reduction pathways [24]. In this way, the model would function as both a detection system and a guide for selecting the most relevant operational response.

Data Sources and Feature Engineering

Raw audit log extraction and structuring

Raw data extraction would begin with EHR audit log tables containing timestamped user actions, event categories, session identifiers, and contextual metadata about the clinical module being used. The extraction process would filter for attending physicians or other physician-role users while excluding non-physician staff unless team-based attribution is explicitly modeled. Relevant event types would include note creation, note editing, note signing, chart review, order placement, order modification, inbox opening, inbox response, result review, refill processing, and alert interaction. Each event would be standardized into a common schema with physician identifier, timestamp, event type, clinical module, encounter context where allowed, and session boundary markers. Prior work using EHR event log data demonstrates that such raw interactions can be converted into physician workload measures [1], and later metric-development studies provide a structured basis for defining comparable activity categories [4]. The resulting data could be transformed into a physician-day-event-type tensor that preserves temporal order while allowing weekly aggregation. Sessionization would be required to distinguish active EHR use from idle time, using inactivity gaps and event density to infer meaningful work intervals. Privacy protection would be maintained by deriving workload features from metadata and event counts rather than retaining note text or patient-specific clinical details in the modeling table.

Note length feature engineering

Note length features would aggregate the number of characters or words across notes authored, edited, or signed by a physician within each week. The feature set would include weekly mean note length, median note length, upper-tail note length, and total documentation volume to distinguish routine documentation from unusually long or cumulative note production. Because outpatient progress notes can increase in length and redundancy over time [14], feature engineering should capture not only average length but also the distributional tail of very long notes. Rate-of-change features over

prior weeks would capture whether note burden is increasing gradually, remaining stable, or spiking abruptly. When text access is permitted under governance rules, NLP-derived indicators could estimate template fraction, copied-forward content, duplicated phrases, and meaningful content density. These derived indicators would help distinguish clinically necessary long notes from non-substantive note expansion caused by templating, auto-population, or redundant documentation habits. If note text cannot be used, proxy features such as note length distributions, edit duration, signature delay, and repeated note-type patterns could still provide evidence of documentation burden. Work showing that note bloat can affect deep learning-based NLP models reinforces the importance of separating meaningful documentation from redundant text expansion [15].

Time-in-EHR and after-hours activity metrics

Time-in-EHR features would be derived from sessionized audit log activity windows that estimate active physician interaction with the EHR during each week. Total active EHR time would capture overall system workload, while domain-specific time features could separate documentation time, inbox time, order-entry time, and chart-review time where event categories allow. After-hours activity would be defined as actions occurring outside a locally specified workday window, such as evenings, nights, weekends, or organizationally defined non-clinic periods. Evidence that total and after-hours EHR time differ across ambulatory specialties supports the need to treat after-hours activity as a context-sensitive signal rather than a universal threshold [7]. The model would use both absolute after-hours time and after-hours proportion because a physician with high total EHR time may differ from a physician whose smaller amount of work occurs mostly outside scheduled hours. Rolling averages would smooth expected variation, while trend features would identify burden creep, such as progressively later note completion over several weeks. Gender differences in EHR workload and documentation time further suggest that time-based features should be evaluated carefully for fairness and contextual validity [9, 25, 26]. Acute spikes would be retained as separate features because they may indicate temporary workload shocks that should be interpreted differently from persistent overload.

Inbox volume and order entry pattern features

Inbox features would quantify weekly incoming and processed messages, including patient messages, result notifications, refill requests, staff messages, and other actionable communications. The model would also include sent-message counts, unread carryover, time-to-first-response, message reopening, and the proportion of inbox work occurring after hours. Inbox message characteristics have been associated with physician burnout [16], and in situ work on physician stress during inbox activity suggests that electronic message management may be an important source of cognitive and emotional strain [21]. These variables are important because inbox burden can create fragmented cognitive work that is not captured by encounter counts or documentation time alone. Order-entry features would quantify total orders, unique order types, order modifications, cancellation or replacement patterns, use of order sets, and manual entry patterns that may indicate workflow complexity. Time-sensitive alerts sent to electronic inboxes provide an example of how message and action demands can combine into operational workload [18]. Order-entry features should be interpreted in clinical context because high order volume may reflect specialty mix, patient acuity, or team structure rather than inefficient workflow by itself. Interactions between inbox and order features may be particularly informative, such as result messages triggering additional orders or refill requests increasing after-hours work.

Feature alignment and transformation pipeline

All engineered features would be aligned into weekly vectors for each physician, with each row representing a physician-week and each column representing an aggregated, normalized, or trend-based workload feature. Specialty-specific normalization would reduce the risk that physicians in inherently documentation-heavy specialties are systematically classified as burdened simply because their normal work differs from other groups. Physician-specific historical baselines would allow the model to detect deviation from an individual's own prior pattern, which is essential for identifying change rather than only comparing physicians to population averages. System-level analyses of EHR time show that organizational context can shape physician workload [23], so alignment procedures should preserve clinic, specialty, and schedule context as covariates or stratification factors. Partial weeks caused by vacation, leave, onboarding, reduced full-

time equivalent status, or rotation changes would be flagged and normalized rather than discarded automatically. Outlier capping could reduce the influence of rare logging artifacts, while missing values could be imputed using physician-specific history, specialty medians, and missingness indicators that allow the model to learn when absence of data is itself informative. The final transformation pipeline would standardize continuous features, encode categorical indicators, and construct sequential feature windows for temporal modeling. Quality-control checks would examine impossible timestamps, duplicate events, unusually long sessions, and changes in audit log definitions after EHR upgrades, which is particularly important because audit log structures and interoperability constraints may vary across systems [13].

Deep Neural Network Architecture

Input branches and multi-modal fusion

The model would use separate branches for note length, EHR time, after-hours work, inbox activity, and order-entry patterns. Each branch would learn a domain-specific representation before fusion into a shared layer. This structure preserves distinct workload signals while allowing the model to learn interactions across documentation, messaging, ordering, and time-based burden [4, 16, 19, 27].

Temporal context through recurrent or attention layers

Weekly feature vectors could pass through recurrent, temporal convolutional, or attention layers to capture burden trajectories. This would help distinguish stable high workload from rapidly worsening burden. Temporal modeling is important because EHR workload and burnout-related signals may evolve gradually over time [10, 11, 22].

Output layer and burden classification

The output layer could generate a burden probability or classify physicians into low, moderate, and high burden categories. Thresholds should support early detection without treating the score as a burnout diagnosis. Confidence estimates and contribution summaries are needed because EHR workload varies by specialty, gender, organization, and practice pattern [2, 8, 9, 12, 23, 25, 26].

Handling Temporal Dynamics and Longitudinal Patterns

Weekly time-window construction

The model would use rolling weekly windows to summarize recent note length, EHR time, after-hours work, inbox activity, and order-entry patterns. Only data available before the prediction period should be used. Weekly windows support prospective detection because EHR workload changes over time and may signal emerging burden [6, 22].

Physician-specific baselines and change detection

Physician-specific baselines would help detect deviations from each clinician's usual workload rather than relying only on system-wide averages. This matters because EHR use varies by specialty, gender, and organizational setting [8, 9, 23, 25, 26]. Change-detection features would compare current workload with personal history, specialty norms, and clinic-level expectations.

Table 2 establishes a temporal-analytical framework for distinguishing transient workload fluctuations from clinically meaningful documentation burden trajectories.

Table 2. Analytical Framework for Distinguishing Transient Workload Spikes from Persistent Documentation Burden Trajectories

Analytical Dimension	Transient Workload Spike	Emerging Burden Trajectory	Persistent High Burden State	Model Feature Signals	Interpretation Implication
Temporal Duration	Short-term (1 week)	Multi-week increasing trend	Sustained elevation (≥4–6 weeks)	Rolling averages, slope features	Distinguishes noise from meaningful trend
Pattern Stability	Irregular, isolated	Gradually stabilizing increase	Consistent high-level plateau	Variance and autocorrelation metrics	Identifies stabilization of overload
Domain Consistency	Single-domain spike	Expanding across domains	Multi-domain co-elevation	Cross-domain feature interactions	Signals systemic vs localized burden
After-Hours Shift	Temporary extension	Increasing evening spillover	Persistent off-hours work pattern	After-hours proportion trends	Indicates erosion of schedule boundaries
Baseline Deviation	Within normal variation	Exceeds personal baseline	Far exceeds both personal and specialty norms	Physician-specific normalization	Enables individualized detection
Recovery Behavior	Rapid return to baseline	Partial recovery between weeks	No recovery periods observed	Lagged temporal features	Detects resilience vs sustained strain
Intervention Sensitivity	Self-resolving	Responsive to early intervention	Resistant without structural change	Change-point detection features	Guides timing and intensity of intervention

Handling FTE changes and leave periods

The model should account for FTE changes, leave, reduced clinical schedules, and rotation-based work. Partial weeks or return-from-leave inbox accumulation could otherwise be misclassified as true workload change. Schedule indicators and adjusted denominators would improve fairness across physicians with different clinical commitments [13, 23].

Model Interpretability for Stakeholders

Explaining burden scores to physicians

Physician-facing interpretability should explain the burden score in terms of recognizable work domains rather than abstract model weights. For example, a private dashboard could show whether a weekly score was mainly influenced by after-hours charting, unusually long notes, inbox accumulation, or order-entry complexity, using contribution methods such as SHAP-style feature attribution. This approach is important because EHR usability and perceived lack of control are associated with professional distress [2, 12], and an opaque model could intensify rather than relieve mistrust. The explanation should therefore be framed as supportive feedback that helps physicians understand workload patterns and request assistance. It should also allow physicians to add context, such as unusual coverage responsibilities or temporary staffing disruptions, so that model output remains clinically and organizationally interpretable.

Aggregate insights for health system leadership

Leadership-facing interpretability should emphasize de-identified trends across clinics, specialties, and operational units rather than individual physician ranking. Aggregate reports could identify whether rising burden in a department is driven mainly by inbox expansion, increased after-hours work, documentation redundancy, or order-entry complexity. This design aligns with evidence that organizational EHR support and after-hours charting are connected to physician burden

[11], while reviews of EMR-related burnout interventions emphasize the importance of system-level remediation rather than individual blame [24]. De-identified SHAP-based summaries could help leaders select appropriate interventions, such as inbox team redesign when messaging is the dominant driver or note template reform when documentation length is the dominant driver. The goal would be to translate model output into organizational learning while protecting individual confidentiality.

Operational Deployment and Intervention Integration

Privacy-respecting model deployment

The model should run inside the institution using derived audit-log features, not patient-level content. Individual scores should remain private to physicians or wellness-support teams, while managers receive only aggregate trends. Because EHR metrics are sensitive [5, 21], deployment requires access controls and clear limits on performance use.

Linking detection to intervention pathways

High burden scores should trigger support, not punishment. Options could include inbox triage, scribe support, documentation coaching, order-set redesign, or note-template review. The model's value depends on whether detection leads to practical workflow improvement [24, 28, 29].

Evaluation Strategy

Burden label derivation and validation

Burden labels should come from validated well-being surveys or expert-defined audit-log thresholds. Because audit-log measures have been linked to burnout and workload prediction [3, 19], validation should test alignment with physician-reported burden.

Predictive performance and calibration

Future evaluation should assess discrimination, calibration, and subgroup fairness without claiming results in advance. Because EHR workload varies by specialty, system context, and physician group [7-9, 23, 25, 26], calibration should be checked across relevant subgroups.

Prospective silent-mode and impact evaluation

The model should first run silently without triggering interventions or exposing individual scores. Silent-mode testing could compare scores with surveys, physician feedback, and operational indicators. Later studies could assess whether model-triggered support reduces documentation burden [22].

Limitations

Data availability and audit log fidelity

Audit logs may miss cognitive work outside the EHR, such as phone calls, teaching, and informal care coordination. Vendor differences and configuration changes may affect event definitions [13]. Ongoing data-quality monitoring and recalibration would be required [4, 5].

Cultural and organisational sensitivity

Physicians may view burden monitoring as surveillance if governance is unclear. Because EHR usability and documentation pressure already contribute to distress [2, 12], deployment must emphasize transparency, physician participation, and separation from disciplinary use.

Conclusion

The proposed deep neural network offers a conceptual framework for detecting physician documentation burden from electronic health record audit log data. By integrating note length, time in the EHR, after-hours charting, inbox volume, and order-entry patterns, the model would represent documentation burden as a multi-domain and temporally evolving workload state.

A central strength of the approach is its capacity to combine heterogeneous signals into a single burden probability while still decomposing the major contributors for interpretation. Its privacy-preserving design would support individual feedback and aggregate organizational learning without requiring routine exposure of patient-level content or individual scores to managers.

Important challenges remain, including label fidelity, audit log quality, specialty variation, and cultural acceptance. The model would need careful validation, transparent governance, and integration into supportive clinical operations rather than performance surveillance.

Future work should involve collaborative pilot studies among physician wellness programs, clinical informatics teams, frontline clinicians, and EHR vendors. Such collaboration could help establish standardized, ethical, and actionable burden surveillance tools that support early intervention and healthier clinical work systems.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 26 Feb 2023

Revised: 17 Mar 2023

Accepted: 12 Apr 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

- Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan WJ, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med*. 2017;15(5):419-26.
<https://doi.org/10.1370/afm.2121>.
- Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clin Proc*. 2020;95(3):476-87.
<https://doi.org/10.1016/j.mayocp.2019.09.024>.
- Dyrbye LN, Gordon J, O'Horo J, Belford SM, Wright M, Satele DV, et al. Relationships between EHR-based audit log data and physician burnout and clinical practice process measures. *Mayo Clin Proc*. 2023;98(3):398-409.
<https://doi.org/10.1016/j.mayocp.2022.11.020>.
- Sinsky CA, Rule A, Cohen G, Arndt BG, Shanafelt TD, Sharp CD, et al. Metrics for assessing physician activity using electronic health record log data. *J Am Med Inform Assoc*. 2020;27(4):639-43.
- Moy AJ, Schwartz JM, Chen R, Sadri S, Lucas E, Cato KD, et al. Measurement of clinical documentation burden among physicians and nurses using electronic health records: a scoping review. *J Am Med Inform Assoc*. 2021;28(5):998-1008.
- Holmgren AJ, Downing NL, Tang M, Sharp C, Longhurst C, Huckman RS. Assessing the impact of the COVID-19 pandemic on clinician ambulatory electronic health record use. *J Am Med Inform Assoc*. 2022;29(3):453-60.
- Rotenstein LS, Holmgren AJ, Downing NL, Bates DW. Differences in total and after-hours electronic health record time across ambulatory specialties. *JAMA Intern Med*. 2021;181(6):863-5.
<https://doi.org/10.1001/jamainternmed.2021.0341>.
- Rotenstein LS, Apathy N, Holmgren AJ, Bates DW. Physician note composition patterns and time on the EHR across specialty types: a national, cross-sectional study. *J Gen Intern Med*. 2023;38(5):1119-26.
<https://doi.org/10.1007/s11606-022-07993-1>.
- Rotenstein LS, Fong AS, Jeffery MM, Sinsky CA, Goldstein R, Williams B, et al. Gender differences in time spent on documentation and the electronic health record in a large ambulatory network. *JAMA Netw Open*. 2022;5(3):e223935.
<https://doi.org/10.1001/jamanetworkopen.2022.3935>.
- Micek MA, Arndt B, Tuan WJ, Trowbridge E, Dean SM, Lochner J, et al. Physician burnout and timing of electronic health record use. *ACI Open*. 2020;4(1):e1-e8.
<https://doi.org/10.1055/s-0040-1702207>.
- Eschenroeder HC Jr, Manzione LC, Adler-Milstein J, Bice C, Cash R, Duda C, et al. Associations of physician burnout with organizational electronic health record support and after-hours charting. *J Am Med Inform Assoc*. 2021;28(5):960-6.
- Tajirian T, Stergiopoulos V, Strudwick G, Sequeira L, Sanches M, Kemp J, et al. The influence of electronic health record use on physician burnout: cross-sectional survey. *J Med Internet Res*. 2020;22(7):e19274.
<https://doi.org/10.2196/19274>.

Li E, Clarke J, Ashrafian H, Darzi A, Neves AL. The impact of electronic health record interoperability on safety and quality of care in high-income countries: systematic review. *J Med Internet Res.* 2022;24(9):e38144.
<https://doi.org/10.2196/38144>.

Rule A, Bedrick S, Chiang MF, Hribar MR. Length and redundancy of outpatient progress notes across a decade at an academic medical center. *JAMA Netw Open.* 2021;4(7):e2115334.
<https://doi.org/10.1001/jamanetworkopen.2021.15334>.

Liu J, Capurro D, Nguyen A, Verspoor K. "Note Bloat" impacts deep learning-based NLP models for clinical prediction tasks. *J Biomed Inform.* 2022;133:104149.
<https://doi.org/10.1016/j.jbi.2022.104149>.

Baxter SL, Saseendrakumar BR, Cheung M, Savides TJ, Longhurst CA, Sinsky CA, et al. Association of electronic health record inbasket message characteristics with physician burnout. *JAMA Netw Open.* 2022;5(11):e2244363.
<https://doi.org/10.1001/jamanetworkopen.2022.44363>.

Margolius D, Siff J, Teng K, Einstadter D, Gunzler D, Bolen S. Primary care physician factors associated with inbox message volume. *J Am Board Fam Med.* 2020;33(3):460-2.
<https://doi.org/10.3122/jabfm.2020.03.190324>.

Cutrona SL, Fouayzi H, Burns L, Sadasivam RS, Mazor KM, Gurwitz JH, et al. Primary care providers' opening of time-sensitive alerts sent to commercial electronic health record InBaskets. *J Gen Intern Med.* 2017;32(11):1210-9.
<https://doi.org/10.1007/s11606-017-4131-8>.

Lou SS, Liu H, Warner BC, Harford D, Lu C, Kannampallil T. Predicting physician burnout using clinical activity logs: model performance and lessons learned. *J Biomed Inform.* 2022;127:104015.
<https://doi.org/10.1016/j.jbi.2022.104015>.

Kim S, Lou SS, Baratta LR, Kannampallil T. Classifying clinical work settings using EHR audit logs: a machine learning approach. *Am J Manag Care.* 2023;29(1):e1-e7.
<https://doi.org/10.37765/ajmc.2023.89353>.

Akbar F, Mark G, Prausnitz S, Warton EM, East JA, Moeller MF, et al. Physician stress during electronic health record inbox work: in situ measurement with wearable sensors. *JMIR Med Inform.* 2021;9(4):e24014.
<https://doi.org/10.2196/24014>.

Lou SS, Lew D, Harford DR, Lu C, Evanoff BA, Duncan JG, et al. Temporal associations between EHR-derived workload, burnout, and errors: a prospective cohort study. *J Gen Intern Med.* 2022;37(9):2165-72.
<https://doi.org/10.1007/s11606-021-07333-7>.

Rotenstein LS, Holmgren AJ, Horn DM, Lipsitz S, Phillips R, Gitomer R, et al. System-level factors and time spent on electronic health records by primary care physicians. *JAMA Netw Open.* 2023;6(11):e2344713.
<https://doi.org/10.1001/jamanetworkopen.2023.44713>.

Li C, Parpia C, Sriharan A, Keefe DT. Electronic medical record-related burnout in healthcare providers: a scoping review of outcomes and interventions. *BMJ Open.* 2022;12(8):e060865.
<https://doi.org/10.1136/bmjopen-2022-060865>.

Rittenberg E, Liebman JB, Rexrode KM. Primary care physician gender and electronic health record workload. *J Gen Intern Med.* 2022;37(13):3295-301.
<https://doi.org/10.1007/s11606-022-07422-3>.

Rule A, Shafer CM, Micek MA, Baltus JJ, Sinsky CA, Arndt BG. Gender differences in primary care physicians' electronic health record use over time: an observational study. *J Gen Intern Med.* 2023;38(6):1570-2.
<https://doi.org/10.1007/s11606-022-07950-y>.

Apathy NC, Rotenstein L, Bates DW, Holmgren AJ. Documentation dynamics: note composition, burden, and physician efficiency. *Health Serv Res.* 2023;58(3):674-85.
<https://doi.org/10.1111/1475-6773.14136>.

Johnson KB, Neuss MJ, Detmer DE. Electronic health records and clinician burnout: a story of three eras. *J Am Med Inform Assoc.* 2021;28(5):967-73.

Budd J. Burnout related to electronic health record use in primary care. *J Prim Care Community Health.* 2023;14:21501319231166921.
<https://doi.org/10.1177/21501319231166921>.