Published in final edited form as:
J Biomed Inform. 2022 March ; 127: 104015. doi:10.1016/j.jbi.2022.104015.

Predicting Physician Burnout using Clinical Activity Logs: Model Performance and Lessons Learned

Sunny S. Lou, MD, PhD[1], **Hanyang Liu, MEng.**[2], **Benjamin C. Warner**[2], **Derek Harford**[1], **Chenyang Lu, PhD**[2], **Thomas Kannampallil, PhD**[1,3]

[1]Department of Anesthesiology, School of Medicine, Washington University in St Louis, St Louis, Missouri

[2]Department of Computer Science, McKelvey School of Engineering, Washington University in St Louis, St Louis, Missouri

[3]Institute for Informatics, School of Medicine, Washington University in St Louis, St Louis, Missouri

Abstract

Background: Burnout is a significant public health concern affecting more than half of the healthcare workforce; however, passive screening tools to detect burnout are lacking. We investigated the ability of machine learning (ML) techniques to identify burnout using passively collected electronic health record (EHR)-based audit log data.

Method: Physician trainees participated in a longitudinal study where they completed monthly burnout surveys and provided access to their EHR-based audit logs. Using the monthly burnout scores as the target outcome, we trained ML models using combinations of features derived from audit log data—aggregate measures of clinical workload, time series-based temporal measures of EHR use, and the baseline burnout score. Five ML models were constructed to predict burnout as a continuous score: penalized linear regression, support vector machine, neural network, random forest, and gradient boosting machine.

Results: 88 trainee physicians participated and completed 416 surveys; >10 million audit log actions were collected (*Mean* [Standard Deviation]=25,691 [14,331] actions per month, per physician). The workload feature set predicted burnout score with a mean absolute error (MAE) of 0.602 (95% Confidence Interval (CI), 0.412–0.826), and was able to predict burnout status with an average AUROC of 0.595 (95% CI 0.355–0.808) and average accuracy 0.567 (95% CI 0.393–0.742). The temporal feature set had a similar performance, with MAE 0.596 (95% CI 0.391–0.826), and AUROC 0.581 (95% CI 0.343–0.790). The addition of the baseline burnout score to the workload features improved the model performance to a mean AUROC of 0.829 (95% CI 0.607–0.996) and mean accuracy of 0.781 (95% CI 0.587–0.936); however, this performance was not meaningfully different than using the baseline burnout score alone.

Conclusions: Current findings illustrate the complexities of predicting burnout exclusively based on clinical work activities as captured in the EHR, highlighting its multi-factorial and

Corresponding Author: Thomas Kannampallil, PhD, 660 S. Euclid Avenue, Campus Box 8054, Washington University School of Medicine, St Louis, Missouri 63110, thomas.k@wustl.edu.

individualized nature. Future prediction studies of burnout should account for individual factors (e.g., resilience, physiological measurements such as sleep) and associated system-level factors (e.g., leadership).

Introduction

Clinical settings place considerable psychological, emotional, and physical demands on practitioners, contributing to stress and burnout [1–3]. Physician burnout is widespread, with national surveys suggesting that it impacts nearly half of both physicians-in-training [4–7] and practicing physicians [8–12]—a prevalence twice that of the general population of the United States [11, 12]. Burnout has serious consequences: at a personal level, it is associated with poor physical and emotional health and suicidal ideation [5, 13, 14], alcohol abuse [15], and all-cause mortality [16]. For patients, it has been associated with lower quality of care [17], lesser patient satisfaction [18, 19], and increased risk for medical error [7, 18, 20, 21]; at an organizational level, it has been associated with increased provider turnover [22, 23] and decreased clinical productivity, costing the healthcare system over four billion dollars a year [24]. As such, burnout is a considerable public health concern, and tools for monitoring and managing clinician burnout are desperately needed [2]. However, current approaches to burnout measurement rely exclusively on surveys [25], with no known frameworks to unobtrusively monitor or predict clinician burnout.

Burnout is primarily a work-related phenomenon [3], and workload has been strongly correlated with burnout [26]. Measurement of workload has traditionally relied on self-reports, journals and time and motion studies [27–29]. With the widespread use of electronic health records (EHR), nearly 50% of a physician’s time is spent on EHR interactions [30–32]. Advances in clinical informatics have allowed researchers to leverage EHR-based audit logs—trails of activities performed on an EHR—to unobtrusively track clinical work activities and associated workload with considerable success [33, 34]. Previous studies have successfully conducted measurements of various contributors of clinical workload using audit logs, including administrative burden [35], cognitive load [36], interruptions [37], task switching [38], interface navigation [39], out-of-office work [30], and clinical documentation [40]. However, studies assessing the direct relationships between workload and burnout are very limited and have been focused on specific clinical tasks [41–44]. For example, Adler-Milstein et al. found that after-hours time spent on the EHR and inbox message volume among primary care physicians was associated increased emotional exhaustion, one of the three constructs associated with burnout [44]. Similarly, Tai-Seale and colleagues found associations between clinical inbox workload and burnout [43].

The ability of EHR audit logs to effectively capture workload measurements and its potential associations with burnout offers opportunities to develop a digital phenotype for burnout, similar to those that have been reported in the mental health literature [45, 46]. For example, machine learning (ML) techniques have used unobtrusively-collected mobile and wearable data [47–50] to predict depression, anxiety, and schizophrenia relapse in a variety of populations. Although burnout is not considered a mental health disorder [51], we hypothesized that burnout may be associated with changes in workload and work habits, which can potentially be captured from EHR audit logs.

The application of ML approaches for identifying burnout in clinicians has received limited attention. Towards this end, the primary objective of this study was to develop a predictive model for physician burnout using workload and work activity measures captured from EHR audit logs. Such a model can potentially be used as a screening tool to monitor and identify physicians at risk for burnout and to provide timely assistance and resources to those in need.

Method

Participants and Study Design

Intern physicians in Internal Medicine, Anesthesiology and Pediatrics training programs (*N*=104) and resident physicians in Internal Medicine (*N*=106) at the Washington University School of Medicine, BJC HealthCare and St Louis Children's Hospital were invited to participate through presentations and email solicitations. Participating physician trainees worked across both inpatient and outpatient settings during the study period. Participants were consented and enrolled between September and November 2020, and data were collected through April 2021.

This was a prospective longitudinal cohort study. During the study period of 6 months, participants completed monthly burnout surveys and consented to provide their EHR-based audit logs and related activities. Participants were compensated $10 for the first survey completion and $5 for each ensuing survey completions.

This study was approved by the institutional review board of Washington University (IRB# 202004260) and was part of a larger study investigating the relationship between clinical workload and trainee wellness [52]. This study is reported according to TRIPOD guidelines [53].

Data Collection

All consented participants completed a baseline survey that included socio-demographic characteristics (age, race, sex, marital status, number of dependents), clinical domain and experience. Participants completed a monthly survey using the Stanford Professional Fulfillment Index (PFI) at baseline and consecutively for the following 6 months. PFI is a 16-item survey that combines burnout and professional fulfillment [21]; the burnout component of PFI is based on workload exhaustion and interpersonal disengagement (depersonalization) and correlates with the commonly used Maslach Burnout Inventory (MBI) on the emotional exhaustion and depersonalization scales [21, 54].

Given the longitudinal, monthly data collection, PFI has the advantage of being able to capture burnout over shorter time intervals, with the questions assessing burnout over the "past two weeks." The surveys were electronically delivered via a secure REDCap link to each participant, aligned with their monthly rotation schedules to capture the burnout associated with the previous month's schedule.

Synchronized with the monthly rotation schedules and burnout surveys, we also collected the raw EHR-based audit logs for each participant for the corresponding month. As previously described, audit log files are maintained in most modern EHRs to track and monitor access to protected health information as mandated by the Health Insurance Portability and Accountability Act [33]. Using the audit logs of consented participants, for each month and participant, we captured all EHR access events (e.g., opening a note, viewing a lab result, signing an order), including information on the time stamp, patient, and type of action performed.

Primary Outcome

The primary outcome was the monthly burnout, as assessed by the PFI scale; with a range of values from 0–4 [21]. Burnout score was used a continuous variable in the presented analysis, and hence, regressor models were employed for analyses.

Feature Engineering from Audit Log Files

EHR-based activities for all recruited participants were retrieved from institutional databases for the study period (Epic Systems, Verona WI). We chose to use raw audit log files as opposed to vendor-derived metrics such as Epic Signal, which can change over time with system upgrades [55], and are currently not available for inpatient workflows. Using the audit log data, several features associated with trainee workload and work patterns were generated, without knowledge of the outcome (i.e., burnout). We formulated two types of features—those based on aggregated “workload” measures and those that incorporated the raw “temporal” measures associated with clinical work activities. An overview of the feature extraction process from the audit log files is shown in Supplemental Figure S1.

The “workload” measures feature set was chosen based on prior research that showed that clinical workload was associated with burnout [41–44], and hence might be used to predict burnout. The primary work responsibilities of trainees typically include review of patient data, note writing, order placement, and review of clinical inbox messages. Therefore, similar to methods that have been previously developed [31], we identified all of the audit log actions relating to patient data review, notes, orders, or reviewing inbox messages; time stamps for the audit log actions were used to compute a heuristic approximation of the time spent on each action, as the difference in time stamps between each consecutive action. Additionally, similar to prior literature, durations of no activity extending for greater than 5 minutes were categorized as “inactive” time periods [31] (see Supplemental Figure S1). This method of measuring time spent on EHR activities has previously been shown to be correlated with time-motion measurements [30, 56].

We computed the following prespecified measures that summarized EHR-based clinical workload: total time spent using the EHR; time spent using the EHR after-hours, measured as the time spent between 6pm and 6am (similar to [57]), patient load, measured as the number of patients seen per day; time spent on the clinical inbox; number of ordering sessions per patient per day; time spent writing notes per patient per day; and time spent on chart review, i.e., reviewing patient notes, flowsheets, or laboratory test results, per patient, per day (Table 1). Each measure was aggregated over the month preceding each survey completion. These measures have been previously described as potential indicators of EHR-based work activities [58, 59]. The “workload” feature set included all of the abovementioned pre-specified clinical workload measures as continuous variables, as well as one-hot encoded categorical variables for participant gender and specialty (Anesthesiology, Pediatrics and Internal Medicine).

For the “temporal” feature set, our assumption was that burnout might be associated with the temporal pattern of work activities on the EHR, potentially relating to increased cognitive load and decreased efficiency, and hence could be used to predict burnout. To summarize the temporal pattern of work activities, we measured the difference in time stamps between consecutive activities in the raw audit log and computed summary statistics for this vector including mean, minimum, maximum, skewness, kurtosis, entropy, total energy, autocorrelation, and slope [60].

Similarly, prior research has highlighted the increased cognitive load associated in the pattern of switches between tasks [61]. Therefore, we grouped actions in the raw audit log by patient and computed a vector representing the time spent consecutively on each patient’s chart before either inactivity (time gap > 5 min) or switching to a second patient’s chart. We summarized the temporal pattern of patient switches by computing the statistical features of this vector using similar summary statistics, as described above. The “temporal” feature set included the time series features summarizing the pattern of individual audit log actions and patient chart switches.

As burnout measurements for the same individual may be correlated over time [52], we also created a “first survey score” feature set that consisted of a single feature: a participant’s first recorded burnout score. This feature mimics the use of our model prospectively once a baseline survey was provided to help “calibrate” the model to the inter-individual variability. When the “first survey score” feature set was used, burnout surveys and associated EHR features from the first month of the study were excluded from the analysis to prevent label leakage.

The three feature sets (workload, temporal, and first survey score) were used alone and in combination to assess their relative contributions to burnout predictive performance (see Table 2 for a summary). To facilitate model training, all values of the extracted features were normalized and scaled to values between 0 and 1 using only the distribution of the training data.

Study participants also occasionally worked at patient care locations where audit log data was not routinely available (e.g., at the Veterans Affairs). Therefore, we excluded all time periods (for a participant’s survey month) with fewer than 3000 recorded audit log actions. Otherwise, there was no missing data.

Model Development and Training

The primary focus of our analysis was to develop models that could be translated for predicting burnout outcomes for “unseen” participants. Towards this end, participant identifiers were not included in the model, and each monthly burnout survey, associated workload and temporal features were treated as independent measurements. For an overview of model training and evaluation, see Figure 1.

As previously described, our primary model formulation was as a regression problem, where we first estimated the continuous burnout scores and then used the same burnout threshold to dichotomize the estimated burnout.

Considering the relatively small sample size (i.e., total number of surveys), ML models may not yield robust performance given different splits of training and test datasets. To get as close an estimation as possible to the true out-of-sample performance of each model on unseen individuals, we used a repeated, nested cross-validation approach, synthesizing the process of hyperparameter optimization and model evaluation (Figure 1); in this nested process, the inner k-fold cross-validation procedure for model hyperparameter optimization was nested inside the outer k-fold cross-validation procedure for model evaluation. We chose 10-folds for both outer and inner cross-validation. For each fold of the outer tenfold cross-validation, the whole dataset was split into 90% training set and 10% testing set, grouped by participant such that data for any individual participant only appeared in either the training or testing set. This participant-level grouping was performed because measurements from the same participant were correlated over time. Then, a machine learning model was trained on the training set, using the optimal set of hyperparameters selected by randomly searching from a predefined hyperparameter space to maximize the mean of area under the receiver operating characteristic (AUROC) over the ten folds of the inner cross-validation performed on the training set. This best-performing trained model was then evaluated on the corresponding testing set for this fold. We repeated the above process for each of the ten folds of the outer cross-validation. Finally, the whole nested cross-validation was repeated 20 times using different random seeds for the k-fold splitting (see Figure 1).

The following regressor models were constructed: penalized linear regression, support vector machine, multi-layer perceptron neural network, random forest, and gradient boosting machine. The following hyperparameters were tuned within nested cross-validation: for penalized linear regression, penalty and L1/L2 ratio; for support vector machine, the kernel and regularization parameter; for neural network, the activation function, solver, and regularization parameter; for random forest, the number of features considered for each split and evaluation criterion; for gradient boosting, the learning rate, maximum depth, child weight, regularization, and the number of boosting rounds. All models were implemented using Python 3.9.5 and sci-kit learn 0.24.2, with the exception of the gradient boosting machine, which was implemented using XGBoost 1.4.0 [62].

Model Evaluation

Overall model performance in predicting burnout score was evaluated using mean absolute error (MAE). Each models' ability to estimate the dichotomous burnout status was assessed across the range of possible predicted burnout score thresholds using the area under the receiver operating characteristic curve (AUROC). Accuracy was also evaluated by dichotomizing predicted and observed burnout scores using the previously validated burnout score threshold of 1.33 [21].

Results

General Characteristics

75 of 104 intern physicians and 13 of 106 resident physicians participated and provided a total of 528 surveys, of which 416 were completed (see Figure 1). A median of 6 (IQR 4–6) surveys were completed per participant. EHR audit log data associated with 25 surveys were excluded due to insufficient audit log events (<3000 actions). A total of 10,045,218 audit log actions were recorded across the study, with a mean (S.D.) of 25,691 (14,331) actions recorded per month per participant.

The demographics and summarized monthly workload measures associated with each completed survey are shown in Table 3 (also see Supplementary Material Table S1, for monthly temporal measures). 54% of the surveys were from Internal Medicine, 24% from Pediatrics, and 22% from Anesthesiology. 43% of surveys were from males. Participants spent a median of 88.3 (Interquartile Range (IQR) 53.0–124.1) hours each month using the EHR, with a median of 10.3 (IQR 4.4–29.8) hours occurring after-hours (i.e., between 6pm and 6am). Median patient load was 6.0 (IQR 4.8–7.1). Per patient per day, participants spent a median of 0.51 (IQR 0.37–0.72) hours writing notes and 0.49 (IQR 0.39–0.60) hours reviewing patient data. The median burnout score across the collected surveys was 1.8 (IQR 1.6–2.2). Burnout (PFI score ≥ 1.33) was identified in 164 surveys (42%).

Model Performance

Five machine learning models were constructed to predict burnout as a continuous score: penalized linear regression, support vector machine, multi-layer perceptron neural network, random forest, and gradient boosting machine (Table 4).

The “workload” feature set predicted burnout score with an average MAE of 0.602 (95% confidence interval CI, 0.412–0.826), and was able to predict burnout status with an average AUROC of 0.595 (95% CI 0.355–0.808) and average accuracy 0.567 (95% CI 0.393–0.742). The “temporal” feature set, had similar performance, with MAE 0.596 (95% CI 0.391–0.826), and AUROC 0.581 (95% CI 0.343–0.790). There was no increase in performance by combining the “workload” and “temporal” feature sets.

Due to the poor discriminative performance of the regression models, we conducted several secondary analyses. First, we evaluated model performance to directly classify burnout as a binary outcome, dichotomized using the previously validated 1.33 score threshold for burnout [21]. Discriminative performance was similar to the regression analysis (Supplementary Material Table S2).

Next, we also evaluated if the addition of the first reported burnout score (i.e., baseline score, at the start of the study) to the model (with workload features) could improve performance. With the addition of this baseline burnout score, model performance improved to a mean AUROC of 0.829 (95% CI 0.607–0.996) and mean accuracy of 0.781 (95% CI 0.587–0.936) (Table 4). However, knowledge of the baseline burnout alone had similar discriminative performance [AUROC=0.819 (95% CI 0.551–0.999), accuracy=0.765 (95% CI 0.547–0.952)].

Discussion

Using a longitudinal study of physician trainees, our goal was to develop a digital phenotype for burnout based on previously established associations between EHR-derived workload and burnout [11, 63, 64]. Our primary hypothesis, based on considerable prior literature, was that the EHR-based clinical activity logs would provide a proxy for developing meaningful EHR-based workload phenotypes associated with burnout. To the best of our knowledge, this is the first study attempting to develop a screening tool for burnout using passively collected data.

Using raw EHR-based audit log data, we engineered a panel of features that captured the primary work responsibilities of physician trainees, including all of the features previously shown to be associated with burnout in statistical analysis [41, 43, 44, 65], as well as time-series features that captured users' EHR activity patterns. We found that the workload feature set predicted burnout with modest discriminative ability (average AUROC of 0.595 and average accuracy 0.567). The temporal feature set had similar performance (average AUROC 0.581 and average accuracy 0.556), suggesting that the pattern of a clinician's EHR usage does provide some insight into their wellness status.

However, despite our best efforts, including both classification and regression approaches using multiple ML modeling strategies with various feature sets, none of the EHR-derived features were able to meaningfully identify clinicians suffering from burnout; confidence intervals for all EHR-derived feature performance spanned below AUROC 0.5, indicating substantial risk for model performance worse than random on new (i.e., unseen) individuals. Although the disappointing performance of these models can be attributed to several factors, our study raises several questions regarding the implicit relationships regarding EHR-based workload measures and burnout, and the viability of using EHR-derived measures alone in such modeling efforts [25].

First, while burnout is primarily a work-related phenomenon, individual responses to workload are likely highly individualized [11, 66]. In a mixed effects statistical analysis of the same cohort (not reported here), we found statistically significant associations between the same proximal clinical workload elements described here (total time spent on the EHR, number of patients, time spent on chart review) and burnout [52, 67] (based on a linear mixed effects regression; results not reported here). However, the measured effect size was modest in comparison to the variation in burnout observed. In addition, we found that the intra-class correlation of the individual level effects was 0.65, suggesting that measurements within individuals were highly correlated, and that more variance in burnout score could be attributed to individual effects than all of the workload measures combined. In the current analysis, we replicated that finding; the model using an individual's first burnout survey response to predict all subsequent responses performed better than any of the models that used EHR measures alone (Table 4). In addition, inter-individual variability explains the wide confidence intervals for model performance we observed, as our cross-validation strategy was grouped by participant.

The addition of workload measures to the first survey response only marginally improved predictive performance (Table 4). We also found that our workload measures were no better at predicting the work exhaustion subscale of the burnout score, nor the degree of self-reported physical exhaustion at work (see summary in Supplemental Table S3). Taken together, these results suggest that although workload does contribute to the burnout phenotype, resulting in modest predictive ability (AUROC=0.595) of our workload-based model and the previously described statistical associations [41, 43, 44, 65], it may be that an individual's interpretation or experience of their workload that is more important. In other words, two individuals exposed to the exact same workload may have very different burnout experiences. Future efforts to identify burnout likely need to account for these intrinsic individual-level variables in order to be successful.

Burnout is also a multi-factorial problem influenced by personal, contextual, and organizational factors in addition to workload itself [3]. In the current analysis, the model formulation primarily relied on workload-related factors and basic socio-demographic characteristics of participants. As such, the models did not account for the contextual aspects of the clinician's personal behaviors (e.g., sleep, physical activity, coping skills, resilience, expectations, mental health) or that of the clinical work environment (e.g., interactions with patients, collaborative work with other clinicians). Additionally, as highlighted by recent research, system-level factors such as program leadership and other organizational characteristics also play a significant role in burnout [68–70], but were not incorporated in the current analysis. As described in a recent National Academy of Medicine report [2], burnout is likely more driven by the system-level factors such as work environment and its associated demands than individual-level factors.

Recently, there has been considerable research regarding the use of unobtrusive personal sensing frameworks using mobile and wearable platforms on a number of psychological constructs, feelings, traits and human behaviors [46]. These studies have relied on a number of measurements including that of physical activity, speech, behavioral indicators, location, sleep and related measures, and have had reasonably successful predictive performance (ranging from 0.65 to 0.8) [46, 71].

Perhaps the addition of more contextual and individualized measurement of physical and physiological activities—exercise, sleep, momentary assessments of wellness—using such platforms can provide situated assessments of individual characteristics that may improve burnout prediction. However, a point of concern is the fact that, as opposed to mood-related disorders that have associated symptomatology, some of the burnout symptoms (e.g., lack of empathy or lack of a feeling of personal accomplishment) are harder to capture with direct, unobtrusive measurement. In addition, recent research has highlighted that the symptoms and manifestations of burnout are distinct from that of depression or anxiety [72]. Nevertheless, future research should consider the use of mobile and wearable technologies and ecological momentary assessments for micro-measurement for further contextualizing the individual intangibles associated with burnout. Additionally, measures of resilience, coping abilities, and other mental health and wellness can provide additional context for burnout prediction.

Finally, our modeling efforts may also have been affected by work patterns of the considered study sample of residents. Residents are unique as they perform the role of learners and healthcare providers, often with limited autonomy. Their role as learners involves rotations in various clinical settings based on their clinical specialty. Such transitions between settings often result in varying clinical workload; for example, a clinical rotation in an intensive care unit versus one in an outpatient setting involve different work activities, patient characteristics, and time spent in the unit. This leads to constant fluctuations in workload and burnout, often on a month-to-month basis for the same trainee, contributing to the challenges in model-based prediction efforts. In addition, trainee physicians may have evolving EHR work habits over time as they gain experience. Modeling efforts may have more success among physicians in practice who have a more consistent work schedules (and workload) over time.

This study has several limitations. This was a single academic medical center study with a small group of trainee physicians and as such the findings may not translate to other settings or physician groups. The participation rate in this study was high, but all participants did not complete all their surveys, potentially leading to response bias. Survey completion was incentivized to facilitate participation; however, we could not assess the impact of the incentives on participation or survey responses. The number of burnout measurement cases (N=391 surveys) was moderate; a higher number of measurements, with a larger participant population may improve predictive performance by reducing overfitting or allowing the use of deep learning techniques to summarize raw audit log data and associated temporal features rather than relying on manual feature engineering.

Conclusions

Burnout is a significant public health concern, with more than half of healthcare workers being affected; the problem has been exacerbated during the COVID-19 pandemic highlighting the considerable challenges ahead for this unique workforce. Given that much of the clinical care activities are documented on the EHR, and that prior research has shown associations between clinical workload and activities and burnout, our focus was on using ML approaches to develop a EHR workload-based digital phenotype associated with burnout. However, workload metrics alone provided poor discriminative performance, highlighting the complexities of the burnout phenotype, especially its multi-factorial, and highly-individualized nature. Future work, especially in the prediction of burnout, should account for individual and behavioral characteristics that may provide novel discriminating features for prediction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Dyrbye LN, Shanafelt TD, Sinsky CA, Cipriano PF, Bhatt J, Ommaya A, West CP, and Meyers D, Burnout among health care professionals: a call to explore and address this underrecognized threat to safe, high-quality care, in NAM perspectives. 2017.

2. National Academies of Sciences, E., and Medicine, Taking Action Against Clinician Burnout: A Systems Approach to ProfessionalWell-Being. 2019, National Academies of Sciences, Engineering, and Medicine: Washington, DC.

3. West CP, Dyrbye LN, and Shanafelt TD, Physician burnout: contributors, consequences and solutions. Journal of internal medicine, 2018. 283(6): p. 516–529. [PubMed: 29505159]

4. Dyrbye LN, West CP, Satele D, Boone S, Tan L, Sloan J, and Shanafelt TD, Burnout among US medical students, residents, and early career physicians relative to the general US population. Academic medicine, 2014. 89(3): p. 443–451. [PubMed: 24448053]

5. Hu Y-Y, Ellis RJ, Hewitt DB, Yang AD, Cheung EO, Moskowitz JT, Potts III JR, Buyske J, Hoyt DB, and Nasca TR, Discrimination, Abuse, Harassment, and Burnout in Surgical Residency Training. New England Journal of Medicine, 2019.

6. West CP, Shanafelt TD, and Kolars JC, Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. Jama, 2011. 306(9): p. 952–960. [PubMed: 21900135]

7. West CP, Tan AD, Habermann TM, Sloan JA, and Shanafelt TD, Association of resident fatigue and distress with perceived medical errors. JAMA, 2009. 302(12): p. 1294–1300. [PubMed: 19773564]

8. Allegra CJ, Hall R, and Yothers G, Prevalence of burnout in the US oncology community: results of a 2003 survey. Journal of Oncology Practice, 2005. 1(4): p. 140–147. [PubMed: 20871697]

9. Embriaco N, Azoulay E, Barrau K, Kentish N, Pochard F, Loundou A, and Papazian L, High level of burnout in intensivists: prevalence and associated factors. American journal of respiratory critical care medicine, 2007. 175(7): p. 686–692. [PubMed: 17234905]

10. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, and Mata DA, Prevalence of burnout among physicians: a systematic review. Jama, 2018. 320(11): p. 1131–1150. [PubMed: 30326495]

11. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, West CP, Sloan J, and Oreskovich MR, Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Archives of internal medicine, 2012. 172(18): p. 1377–1385. [PubMed: 22911330]

12. Shanafelt TD, Hasan O, Dyrbye LN, Sinsky C, Satele D, Sloan J, and West CP. Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. in Mayo Clinic Proceedings. 2015. Elsevier.

13. Shanafelt TD, Balch CM, Dyrbye L, Bechamps G, Russell T, Satele D, Rummans T, Swartz K, Novotny PJ, and Sloan J, Special report: suicidal ideation among American surgeons. Archives of surgery, 2011. 146(1): p. 54–62. [PubMed: 21242446]

14. Van der Heijden F, Dillingh G, Bakker A, and Prins J, Suicidal thoughts among medical residents with burnout. Archives of suicide research, 2008. 12(4): p. 344–346. [PubMed: 18828037]

15. Oreskovich MR, Kaups KL, Balch CM, Hanks JB, Satele D, Sloan J, Meredith C, Buhl A, Dyrbye LN, and Shanafelt TD, Prevalence of alcohol use disorders among American surgeons. Archives of surgery, 2012. 147(2): p. 168–174. [PubMed: 22351913]

16. Ahola K, Väänänen A, Koskinen A, Kouvonen A, and Shirom A, Burnout as a predictor of all-cause mortality among industrial employees: a 10-year prospective register-linkage study. Journal of psychosomatic research, 2010. 69(1): p. 51–57. [PubMed: 20630263]

17. Klein J, Grosse Frie K, Blum K, and von dem Knesebeck O, Burnout and perceived quality of care among German clinicians in surgery. International Journal for Quality in Health Care, 2010. 22(6): p. 525–530. [PubMed: 20935011]

18. Dewa CS, Loong D, Bonato S, and Trojanowski L, The relationship between physician burnout and quality of healthcare in terms of safety and acceptability: a systematic review. BMJ Open, 2017. 7(6): p. e015141.

19. Panagioti M, Geraghty K, Johnson J, Zhou A, Panagopoulou E, Chew-Graham C, Peters D, Hodkinson A, Riley R, and Esmail A, Association between physician burnout and patient safety, professionalism, and patient satisfaction: a systematic review and meta-analysis. JAMA internal medicine, 2018. 178(10): p. 1317–1331. [PubMed: 30193239]

20. Shanafelt TD, Balch CM, Bechamps G, Russell T, Dyrbye L, Satele D, Collicott P, Novotny PJ, Sloan J, and Freischlag J, Burnout and medical errors among American surgeons. Annals of Surgery, 2010. 251(6): p. 995–1000. [PubMed: 19934755]

21. Trockel M, Bohman B, Lesure E, Hamidi MS, Welle D, Roberts L, and Shanafelt T, A brief instrument to assess both burnout and professional fulfillment in physicians: reliability and validity, including correlation with self-reported medical errors, in a sample of resident and practicing physicians. Academic Psychiatry, 2018. 42(1): p. 11–24. [PubMed: 29196982]

22. Del Carmen M, Herman J, Rao S, Hidrue MK, Ting D, Lehrhoff SR, Lenz S, Heffernan J, and Ferris TG, Trends and factors associated with physician burnout at a multispecialty academic faculty practice organization. JAMA network open, 2019. 2(3): p. e190554–e190554. [PubMed: 30874776]

23. Shanafelt TD, Raymond M, Kosty M, Satele D, Horn L, Pippen J, Chu Q, Chew H, Clark WB, and Hanley AE, Satisfaction with work-life balance and the career and retirement plans of US oncologists. Journal of Clinical Oncology, 2014. 32(11): p. 1127. [PubMed: 24616305]

24. Han S, Shanafelt TD, Sinsky CA, Awad KM, Dyrbye LN, Fiscus LC, Trockel M, and Goh J, Estimating the attributable cost of physician burnout in the United States. Annals of internal medicine, 2019. 170(11): p. 784–790. [PubMed: 31132791]

25. Kannampallil T, Abraham J, Lou SS, and Payne PR, Conceptual considerations for using EHR-based activity logs to measure clinician burnout and its effects. Journal of the American Medical Informatics Association, 2021. 28(5): p. 1032–1037. [PubMed: 33355360]

26. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, and West CP. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. in Mayo Clinic Proceedings. 2016. Elsevier.

27. Ballermann MA, Shaw NT, Mayes DC, Gibney RN, and Westbrook JI, Validation of the Work Observation Method By Activity Timing (WOMBAT) method of conducting time-motion observations in critical care settings: an observational study. BMC medical informatics and decision making, 2011. 11(1): p. 32. [PubMed: 21586166]

28. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, Westbrook J, Tutty M, and Blike G, Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Annals of internal medicine, 2016. 165(11): p. 753–760. [PubMed: 27595430]

29. DesRoches C, Donelan K, Buerhaus P, and Zhonghe L, Registered nurses' use of electronic health records: findings from a national survey. The Medscape Journal of Medicine, 2008. 10(7): p. 164. [PubMed: 18769691]

30. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W-J, Sinsky CA, and Gilchrist VJ, Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. The Annals of Family Medicine, 2017. 15(5): p. 419–426. [PubMed: 28893811]

31. Ouyang D, Chen JH, Hom J, and Chi J, Internal medicine resident computer usage: an electronic audit of an inpatient service. JAMA internal medicine, 2016. 176(2): p. 252–254. [PubMed: 26642261]

32. Tai-Seale M, Olson CW, Li J, Chan AS, Morikawa C, Durbin M, Wang W, and Luft HS, Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. Health Affairs, 2017. 36(4): p. 655–662. [PubMed: 28373331]

33. Adler-Milstein J, Adelman JS, Tai-Seale M, Patel VL, and Dymek C, EHR audit logs: a new goldmine for health services research? Journal of biomedical informatics, 2020. 101: p. 103343. [PubMed: 31821887]

34. Rule A, Chiang MF, and Hribar MR, Using electronic health record audit logs to study clinical activity: a systematic review of aims, measures, and methods. Journal of the American Medical Informatics Association, 2019.

35. DiAngi YT, Stevens LA, Halpern–Felsher B, Pageler NM, and Lee TC, Electronic health record (EHR) training program identifies a new tool to quantify the EHR time burden and improves providers' perceived control over their workload in the EHR. JAMIA Open, 2019.

36. Ratanawongsa N, Matta GY, Bohsali FB, and Chisolm MS, Reducing misses and near misses related to multitasking on the electronic health record: observational study and qualitative analysis. JMIR human factors, 2018. 5(1): p. e4. [PubMed: 29410388]

37. Gardner RL, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, and Linzer M, Physician stress and burnout: the impact of health information technology. Journal of the American Medical Informatics Association, 2018. 26(2): p. 106–114.

38. Lieu TA, Warton EM, East JA, Moeller MF, Prausnitz S, Ballesca M, Mark G, Akbar F, Awsare S, and Chen Y-FI, Evaluation of Attention Switching and Duration of Electronic Inbox Work Among Primary Care Physicians. JAMA network open, 2021. 4(1): p. e2031856–e2031856. [PubMed: 33475754]

39. Kroth PJ, Morioka-Douglas N, Veres S, Pollock K, Babbott S, Poplau S, Corrigan K, and Linzer M, The electronic elephant in the room: Physicians and the electronic health record. JAMIA open, 2018. 1(1): p. 49–56. [PubMed: 31093606]

40. Baumann LA, Baker J, and Elshaug AG, The impact of electronic health record systems on clinical documentation times: A systematic review. Health Policy, 2018. 122(8): p. 827–836. [PubMed: 29895467]

41. McPeek-Hinz E, Boazak M, Sexton JB, Adair KC, West V, Goldstein BA, Alphin RS, Idris S, Hammond WE, and Hwang SE, Clinician burnout associated with sex, clinician type, work culture, and use of electronic health records. JAMA network open, 2021. 4(4): p. e215686–e215686. [PubMed: 33877310]

42. Tajirian T, Stergiopoulos V, Strudwick G, Sequeira L, Sanches M, Kemp J, Ramamoorthi K, Zhang T, and Jankowicz D, The influence of electronic health record use on physician burnout: cross-sectional survey. Journal of medical Internet research, 2020. 22(7): p. e19274. [PubMed: 32673234]

43. Tai-Seale M, Dillon EC, Yang Y, Nordgren R, Steinberg RL, Nauenberg T, Lee TC, Meehan A, Li J, and Chan AS, Physicians' well-being linked to in-basket messages generated by algorithms in electronic health records. Health Affairs, 2019. 38(7): p. 1073–1078. [PubMed: 31260371]

44. Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, and Grumbach K, Electronic health records and burnout: Time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. Journal of the American Medical Informatics Association, 2020. 27(4): p. 531–538. [PubMed: 32016375]

45. Insel TR, Digital phenotyping: technology for a new science of behavior. Jama, 2017. 318(13): p. 1215–1216. [PubMed: 28973224]

46. Mohr DC, Zhang M, and Schueller SM, Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annual review of clinical psychology, 2017. 13: p. 23–47.

47. Huckins JF, DaSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, Wu J, Obuchi M, Murphy EI, and Meyer ML, Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. Journal of medical Internet research, 2020. 22(6): p. e20185. [PubMed: 32519963]

48. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, Zhou X, Ben-Zeev D, and Campbell AT. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. in Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing. 2014.

49. Wang R, Wang W, daSilva A, Huckins JF, Kelley WM, Heatherton TF, and Campbell AT, Tracking depression dynamics in college students using mobile phone and wearable sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018. 2(1): p. 43.

50. Cao B, Zheng L, Zhang C, Yu PS, Piscitello A, Zulueta J, Ajilore O, Ryan K, and Leow AD. Deepmood: modeling mobile phone typing dynamics for mood detection. in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017.

51. Menon NK, Shanafelt TD, Sinsky CA, Linzer M, Carlasare L, Brady KJ, Stillman MJ, and Trockel MT, Association of physician burnout with suicidal ideation and medical errors. JAMA network open, 2020. 3(12): p. e2028780–e2028780. [PubMed: 33295977]

52. Lou SS, Lew D, Harford D, Lu C, Evanoff BA, Duncan JG, and Kannampallil TG, Temporal associations between EHR-derived workload, burnout, and errors: a prospective cohort study, Journal of General Internal Medicine (Accepted).

53. Collins GS, Reitsma JB, Altman DG, and Moons KG, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Journal of British Surgery, 2015. 102(3): p. 148–158.

54. Maslach C, Jackson SE, Leiter MP, Schaufeli WB, and Schwab RL, Maslach burnout inventory. Vol. 21. 1986: Consulting psychologists press Palo Alto, CA.

55. Hron JD and Lourie E, Have you got the time? Challenges using vendor electronic health record metrics of provider efficiency. Journal of the American Medical Informatics Association, 2020. 27(4): p. 644–646. [PubMed: 32016394]

56. Sinha A, Stevens LA, Su F, Pageler NM, and Tawfik DS, Measuring electronic health record use in the pediatric ICU using audit-logs and screen recordings. Applied Clinical Informatics, 2021. 12(04): p. 737–744. [PubMed: 34380167]

57. Overhage JM and McCallie D Jr, Physician time spent using the electronic health record during outpatient encounters: a descriptive study. Annals of internal medicine, 2020. 172(3): p. 169–174. [PubMed: 31931523]

58. Sinsky CA, Rule A, Cohen G, Arndt BG, Shanafelt TD, Sharp CD, Baxter SL, Tai-Seale M, Yan S, and Chen Y, Metrics for assessing physician activity using electronic health record log data. Journal of the American Medical Informatics Association, 2020. 27(4): p. 639–643. [PubMed: 32027360]

59. Baxter SL, Apathy NC, Cross DA, Sinsky C, and Hribar MR, Measures of electronic health record use in outpatient settings across vendors. Journal of the American Medical Informatics Association, 2021. 28(5): p. 955–959. [PubMed: 33211862]

60. Barandas M, Folgado D, Fernandes L, Santos S, Abreu M, Bota P, Liu H, Schultz T, and Gamboa H, TSFEL: Time series feature extraction library. SoftwareX, 2020. 11: p. 100456.

61. Monsell S, Task switching. Trends in cognitive sciences, 2003. 7(3): p. 134–140. [PubMed: 12639695]

62. Chen T and Guestrin C. Xgboost: A scalable tree boosting system. in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.

63. Maslach C, Schaufeli WB, and Leiter MP, Job burnout. Annual review of psychology, 2001. 52(1): p. 397–422.

64. West CP, Dyrbye LN, Erwin PJ, and Shanafelt TD, Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. The Lancet, 2016. 388(10057): p. 2272–2281.

65. Tran B, Lenhart A, Ross R, and Dorr DA, Burnout and EHR use among academic primary care physicians with varied clinical workloads. AMIA Summits on Translational Science Proceedings, 2019. 2019: p. 136.

66. Balch CM, Shanafelt TD, Dyrbye L, Sloan JA, Russell TR, Bechamps GJ, and Freischlag JA, Surgeon distress as calibrated by hours worked and nights on call. Journal of the American College of Surgeons, 2010. 211(5): p. 609–619. [PubMed: 20851643]

67. Lou SS, Lew D, Harford D, Lu C, Evanoff BA, Duncan JG, and Kannampallil TG, A Longitudinal Study of Burnout and Clinical Workload Measured With Electronic Health Record Audit Logs, in American Medical Informatics Association Annual Symposium. 2021: San Diego, CA.

68. Shanafelt TD, Gorringe G, Menaker R, Storz KA, Reeves D, Buskirk SJ, Sloan JA, and Swensen SJ. Impact of organizational leadership on physician burnout and satisfaction. in Mayo Clinic Proceedings. 2015. Elsevier.

69. Shanafelt TD, Makowski MS, Wang H, Bohman B, Leonard M, Harrington RA, Minor L, and Trockel M, Association of burnout, professional fulfillment, and self-care practices of physician leaders with their independently Rated leadership effectiveness. JAMA network open, 2020. 3(6): p. e207961–e207961. [PubMed: 32543700]

70. Shanafelt T and Swensen S, Leadership and physician burnout: using the annual review to reduce burnout and promote engagement. American Journal of Medical Quality, 2017. 32(5): p. 563–565. [PubMed: 28651438]

71. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, Marx BP, Feast J, Deckersbach T, and Nierenberg A, Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. Journal of medical Internet research, 2017. 19(3): p. e75. [PubMed: 28302595]

72. Fischer R, Mattos P, Teixeira C, Ganzerla DS, Rosa RG, and Bozza FA, Association of Burnout With Depression and Anxiety in Critical Care Clinicians in Brazil. JAMA network open, 2020. 3(12): p. e2030898–e2030898. [PubMed: 33355676]

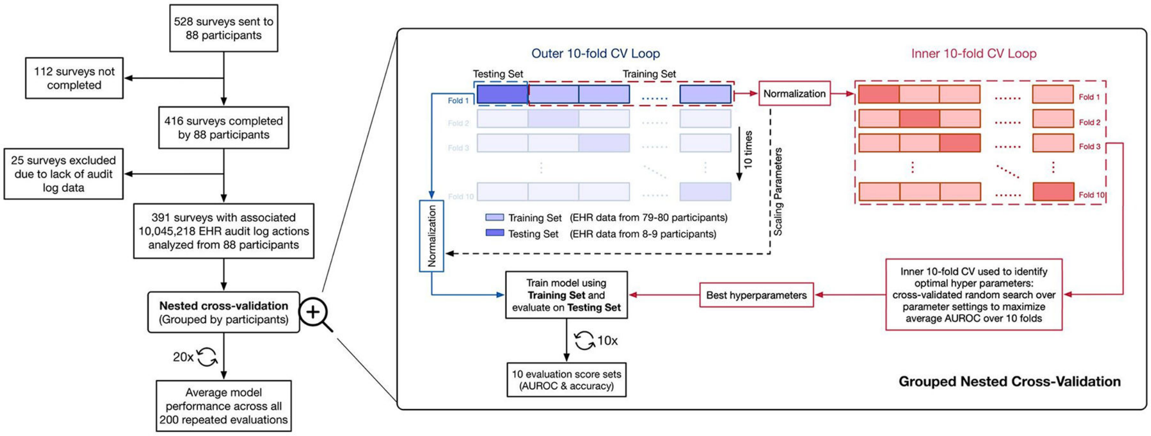
528 surveys sent to 88 participants
112 surveys not completed
416 surveys completed by 88 participants
25 surveys excluded due to lack of audit log data
391 surveys with associated 10,045,218 EHR audit log actions analyzed from 88 participants
Nested cross-validation (Grouped by participants)
20x
Average model performance across all 200 repeated evaluations
Outer 10-fold CV Loop
Testing Set
Training Set
Fold 1
Fold 2
Fold 3
Fold 10
10 times
Normalization
Scaling Parameters
Training Set (EHR data from 79-80 participants)
Testing Set (EHR data from 8-9 participants)
Normalization
Inner 10-fold CV Loop
Fold 1
Fold 2
Fold 3
Fold 10
Inner 10-fold CV used to identify optimal hyper parameters: cross-validated random search over parameter settings to maximize average AUROC over 10 folds
Best hyperparameters
Train model using Training Set and evaluate on Testing Set
10x
10 evaluation score sets (AUROC & accuracy)
Grouped Nested Cross-Validation

Figure 1.

Data collection, model training and cross-validation approach.

Table 1.

Description for how EHR workload measures were calculated. Computer code for reproducing these measurements is available at https://github.com/sslou/publications/tree/main/2021_burnout_lmm. The method is illustrated in Supplemental Figure S1.

Metric	**Description**
Total EHR time per month (hours)	Divide audit log actions in the 30-day window preceding survey completion into chunks separated by at least 5 minutes of inactivity. Sum the difference in timestamps between the first action and last action in each chunk.
After-hours EHR time (hours/month)	Same as above, except only considering audit log actions that occur between 6pm and 6am.
Patient load per day	Mean number of patients per day for which the user either placed an order or wrote a note.
Inbox time per month (hours)	Assign time to audit log actions as the time between each action and subsequent action, excluding when that difference is greater than 5 minutes. Sum the time spent on all audit log actions assigned to the Inbox category.
Ordering sessions per patient per day	Count of audit log actions where METRIC_NAME == “Order list changed”, normalized per patient per day (measured with “Patient load per day” above). “Order list changed” appears in the audit log whenever a clinician signs one or more orders, including those placed using order sets. Thus, this variable captures the number of ordering sessions where at least one order is placed, rather than the total number of individual orders placed.
Note time per patient per day (hours)	Sum of total time spent on each note, measured using Epic metadata recording the time spent with the note-editing pane open for each note, normalized per patient per day using the number of patients per day for which the user wrote a note.
Review time per patient per day (hours)	Same as Inbox time, except for all audit log actions categorized to Note Review, Chart Review, or Results Review, and normalized per patient per day (measured with “Patient load per day” above)

Table 2.

Various combinations of feature sets that were utilized for model development: workload-based, temporal, first survey score, and combination of first survey score and workload. Two combinations of these feature sets were also considered: workload+temporal and first survey score+workload.

Feature set	**Description**
Workload	Gender and clinical specialty (as categorical variables). All of the EHR-associated features detailed in Table 1 (as continuous variables): total EHR time, after-hours EHR time, patient load, inbox time, time spent on notes, chart review, and number of orders, per patient per day.
Temporal	Statistical features (mean, minimum, maximum, skewness, kurtosis, entropy, total energy, autocorrelation, and slope) that describe the temporal pattern of time gaps between consecutive access events recorded in the raw audit log, and of time gaps for switches between different patient charts.
First survey score	PFI-based burnout score in the first month of study participation (i.e., baseline burnout score).

Table 3.

Demographic characteristics of the survey responses and workload-associated EHR measures. Count and % presented for categorical variables. Median (IQR) presented for continuous variables.

Variable	Burned Out N = 164	Not Burned Out N = 227	All Survey Months N = 391
Specialty – Medicine	88 (54%)	124 (55%)	212 (54%)
Pediatrics	40 (24%)	76 (33%)	116 (30%)
Anesthesiology	36 (22%)	27 (12%)	63 (16%)
Gender (= Male)	70 (43%)	100 (44%)	170 (43%)
Avg total EHR time per month (hours)	94.5 (62.5–129.9)	81.4 (48.3–115.7)	88.3 (53.0–124.1)
Avg after-hours EHR time per month (hours)	13.0 (6.1–36.0)	8.7 (3.9–21.7)	10.3 (4.4–29.8)
Patient load	6.2 (5.3–7.3)	5.8 (4.6–7.0)	6.0 (4.8–7.1)
Number of orders/patient/day	3.4 (2.6–4.4)	3.2 (2.5–4.0)	3.2 (2.6–4.1)
Note time/patient/day (hours)	0.53 (0.40–0.72)	0.51 (0.35–0.72)	0.51 (0.37–0.72)
Review time/patient/day (hours)	0.50 (0.38–0.62)	0.49 (0.40–0.57)	0.49 (0.39–0.60)
Inbox time per month (hours)	0.98 (0.43–2.19)	1.21 (0.34–2.94)	1.07 (0.37–2.61)
1st survey-month PFI burnout score	1.8 (1.7–2.1)	1.0 (0.5–1.1)	1.3 (0.9–1.8)
Overall PFI burnout score	1.8 (1.6–2.2)	0.8 (0.3–1.1)	1.2 (0.7–1.7)

Table 4.

Predictive performance for each feature set. Five ML regressor models were fit for each feature set to predict burnout score on the PFI scale (which ranges 1–4): linear regression, support vector machine, random forest, gradient boosting tree, and multi-layer perceptron neural network. The results for the best-performing model for each feature group is shown, along with the resulting mean absolute error (MAE) in burnout score prediction. The predicted score was dichotomized throughout the range of possible thresholds to measure area under the receiver operating characteristic curve (AUROC) for burnout identification. Accuracy is reported using the score threshold of 1.33. 95% confidence intervals are shown for all metrics as determined by nested cross-validation grouped by participant.

Feature Set	Best Model	MAE	AUROC	Accuracy
Workload	Random Forest	0.602 (0.412, 0.826)	0.595 (0.355, 0.808)	0.567 (0.393, 0.742)
Temporal	Support Vector Machine	0.596 (0.391, 0.826)	0.581 (0.343, 0.790)	0.556 (0.318, 0.756)
Workload + Temporal	Gradient Boosting Machine	0.619 (0.438, 0.844)	0.583 (0.270, 0.831)	0.559 (0.386, 0.780)
First Survey Score	Neural Network	0.432 (0.304, 0.570)	0.819 (0.551, 0.999)	0.765 (0.547, 0.952)
First Survey Score + Workload	Neural Network	0.423 (0.293, 0.567)	0.829 (0.607, 0.996)	0.781 (0.587, 0.936)