Predicting Physician Burnout Using EHR Activity Logs

Hanyang Liu, PhD Student

Advisor: Chenyang Lu

Oct 12, 2023

Washington University in St. Louis
James McKelvey School of Engineering

CPSL
Cyber-Physical Systems Laboratory

Washington University in St. Louis
School of Medicine

Background



Physician Burnout

Burnout

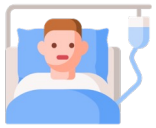
-- “mental exhaustion caused by one’s professional life”

Consequence



Substance Abuse

Suicidal Ideation



Medical error

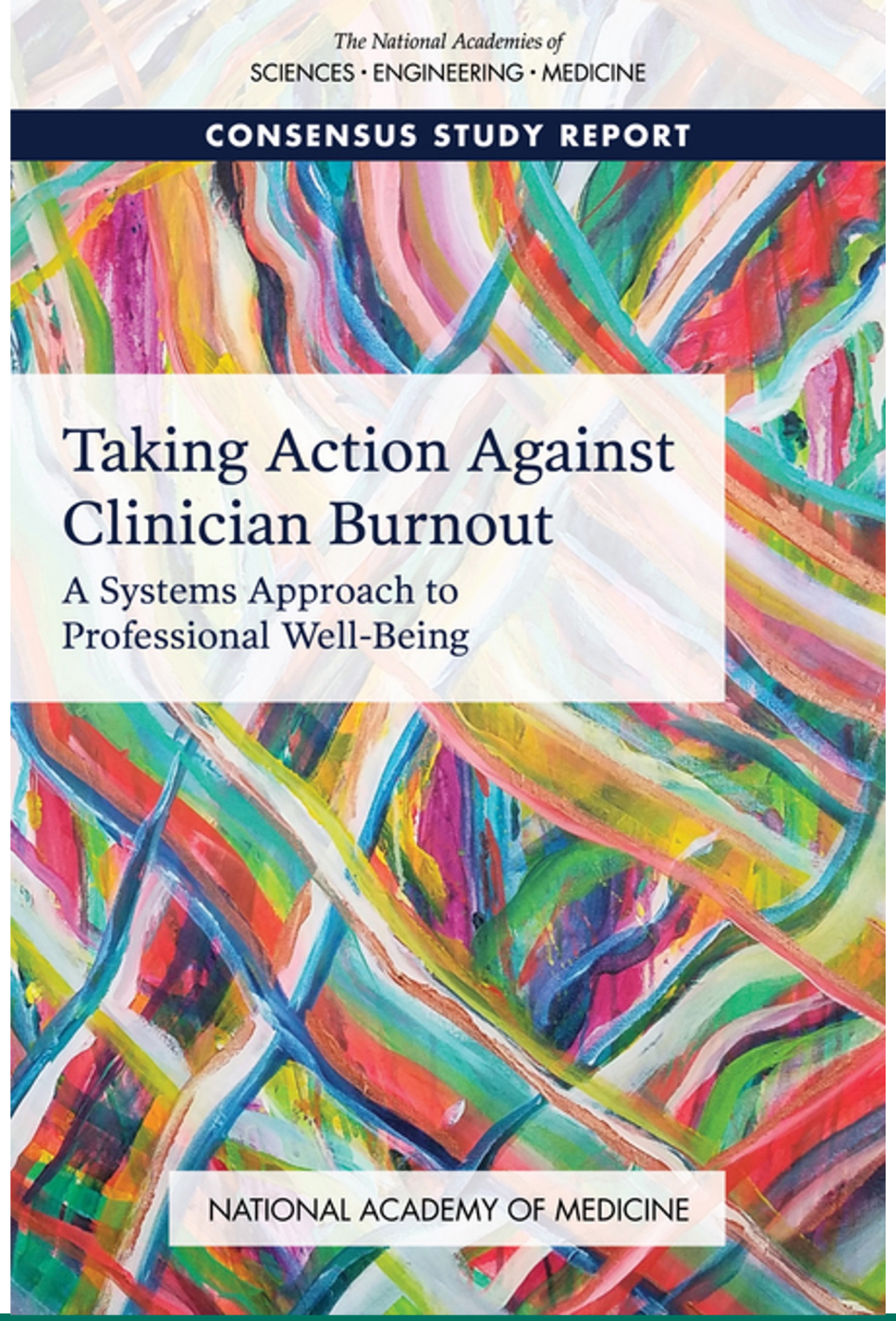
Poor care quality



Increased turnover

Decreased work hours

> $4 billion /year



CPSL
Cyber-Physical Systems Laboratory

50% of physicians

70% of trainee physicians (residents, fellows)

COVID-19 pandemic further fueled the negative impact

Measurement and Intervention

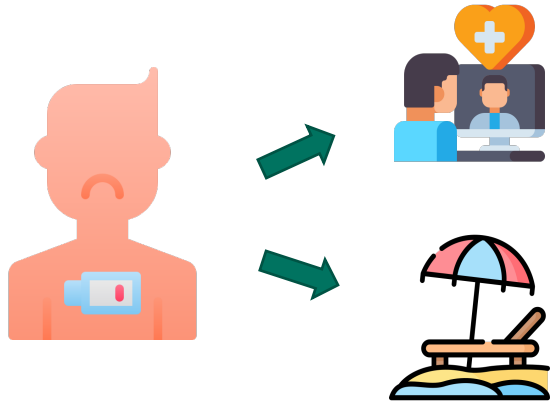
CPSL
Cyber-Physical Systems Laboratory

Traditional Burnout Assessment



Obtrusive, Slow and **Expensive**

Personalized Intervention



Requires monitoring & prediction system

No such system yet!

Approach I:
Feature Engineering + Standard ML



What are EHR Activity Logs?

CPSL
Cyber-Physical Systems Laboratory

Time	USER_ID	PAT_ID	Clinical Action
12/21/19 7:21	M58530		Inpatient Patient Lists list loaded
12/21/19 7:21	M58530		Inpatient system list accessed
12/21/19 7:21	M58530	A	Storyboard viewed
12/21/19 7:21	M58530	A	SmartSets activity selected
12/21/19 7:21	M58530	A	Visit Navigator template loaded
12/21/19 7:21	M58530	A	Problem List accessed
12/21/19 7:21	M58530	A	Chart Review Encounters tab selected
12/21/19 7:21	M58530	A	Chart Review Notes tab selected
12/21/19 7:21	M58530	A	Chart Review Note report viewed
12/21/19 7:22	M58530	A	Report with patient data viewed

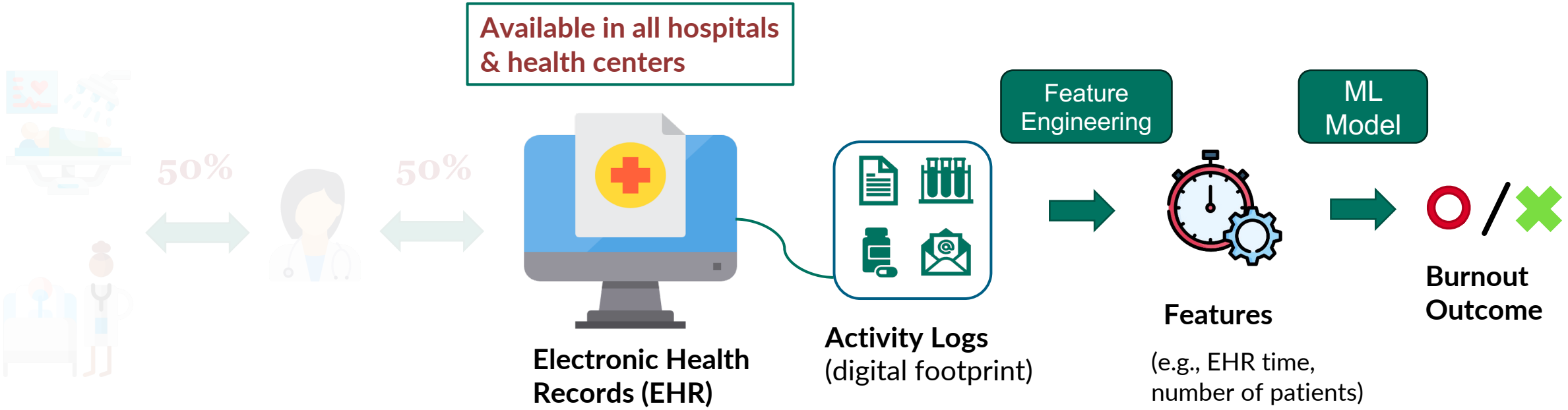


Tracks **who** performed **what** at **what time**.

Machine Learning for Burnout Prediction

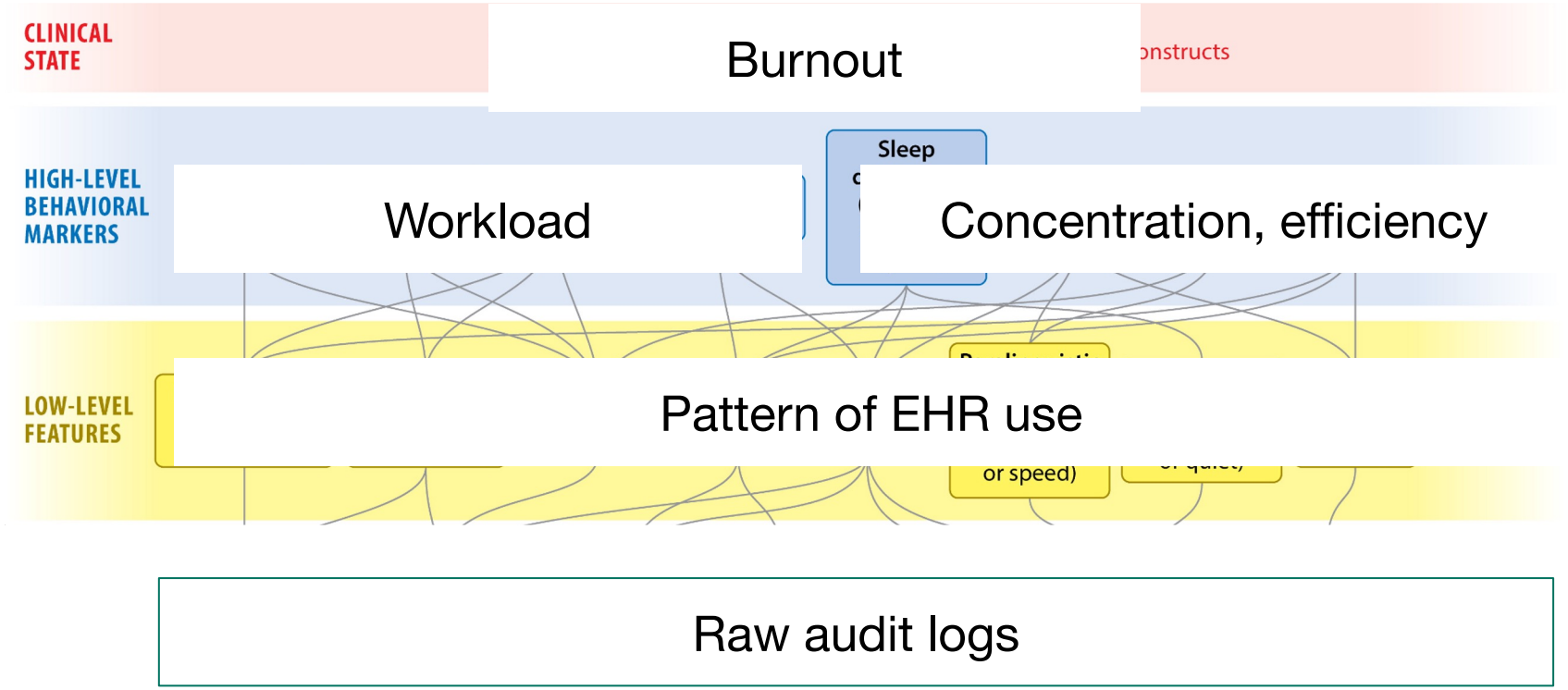
CPSL
Cyber-Physical Systems Laboratory

Goal: A system that can assess the risk of burnout



Digital Phenotyping for Burnout

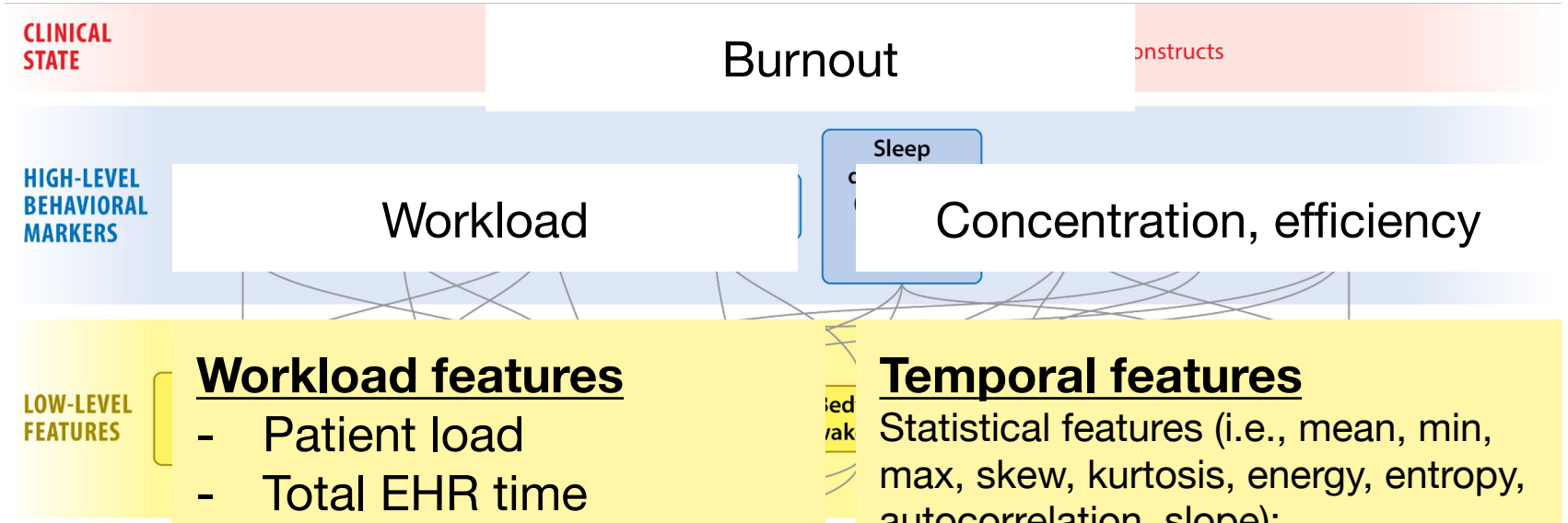
CPSL
Cyber-Physical Systems Laboratory



Mohr et al (2017) *Ann. Rev. Clin. Psych.*

Digital Phenotyping for Burnout

CPSL
Cyber-Physical Systems Laboratory



Workload features

- Patient load
- Total EHR time
- After-hours time
- Inbox time
- Chart review time
- Note time
- Order volume

Temporal features

Statistical features (i.e., mean, min, max, skew, kurtosis, energy, entropy, autocorrelation, slope):

- Temporal pattern of EHR click activities
- Temporal pattern of attention switching

Burnout Score

0.0 — 4.0

Low — High

EHR: Electronic Health Records

Association with Burnout

Statistical analysis shows:

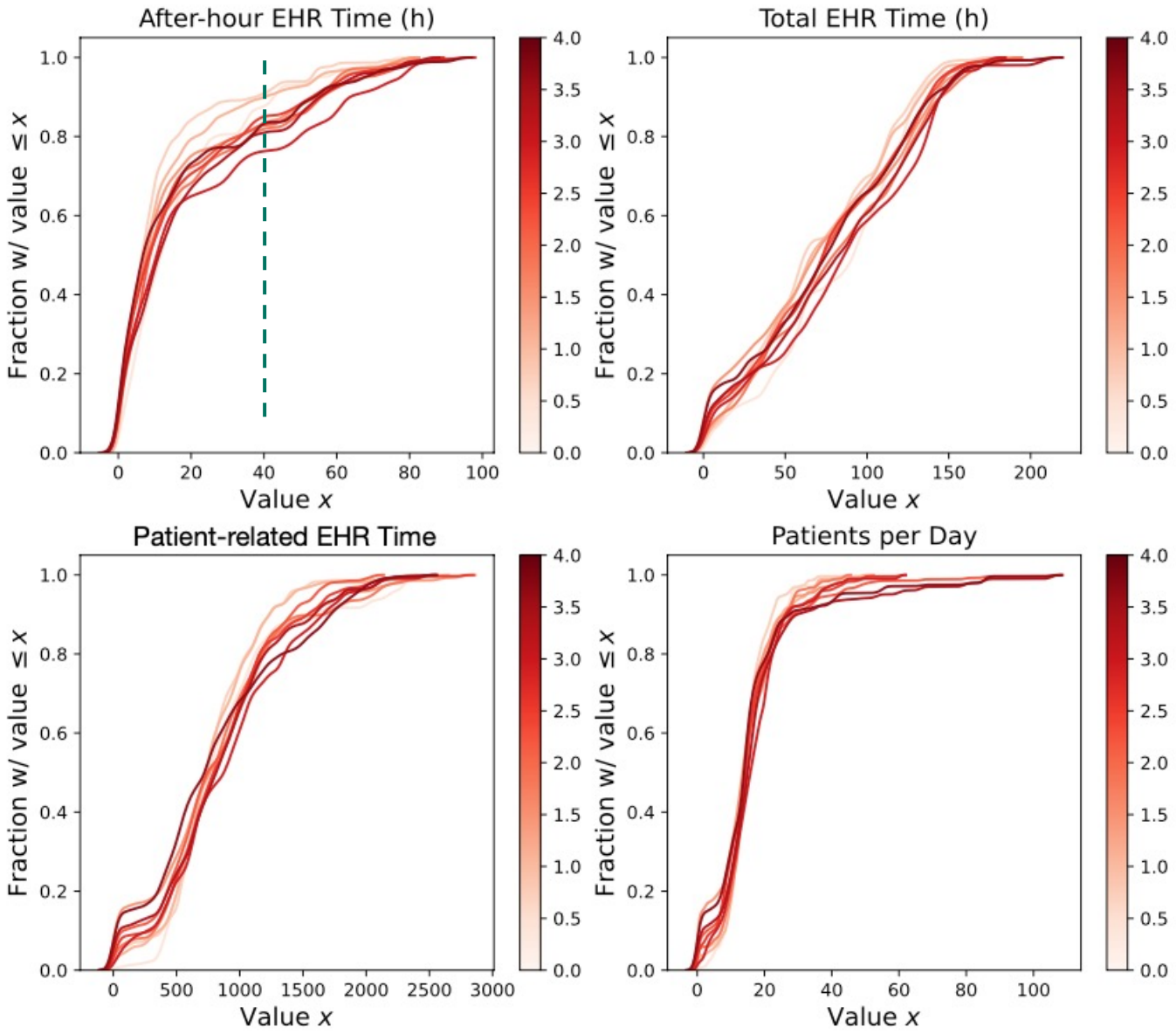
Higher workload links to higher risk of burnout.

e.g., 25% burned-out **VS** 10% healthy physicians have spent > 40H/month on "after-hour" working

There exists predictive information in EHR activity logs

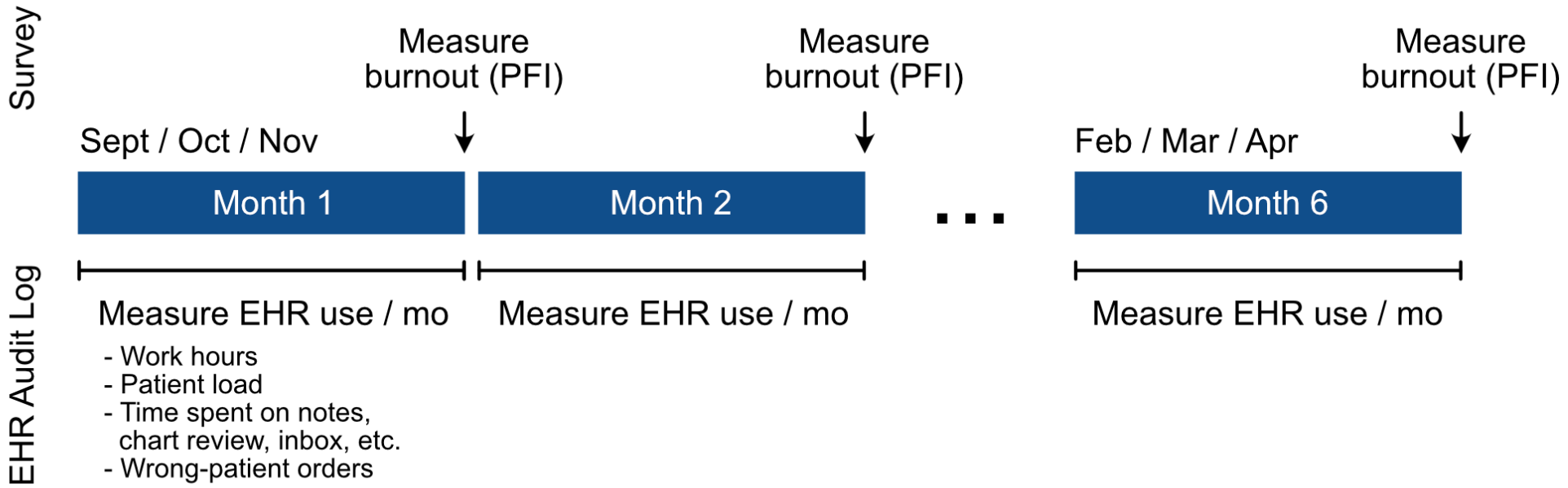
Workload & Burnout

Cumulative Distribution Functions (CDF)



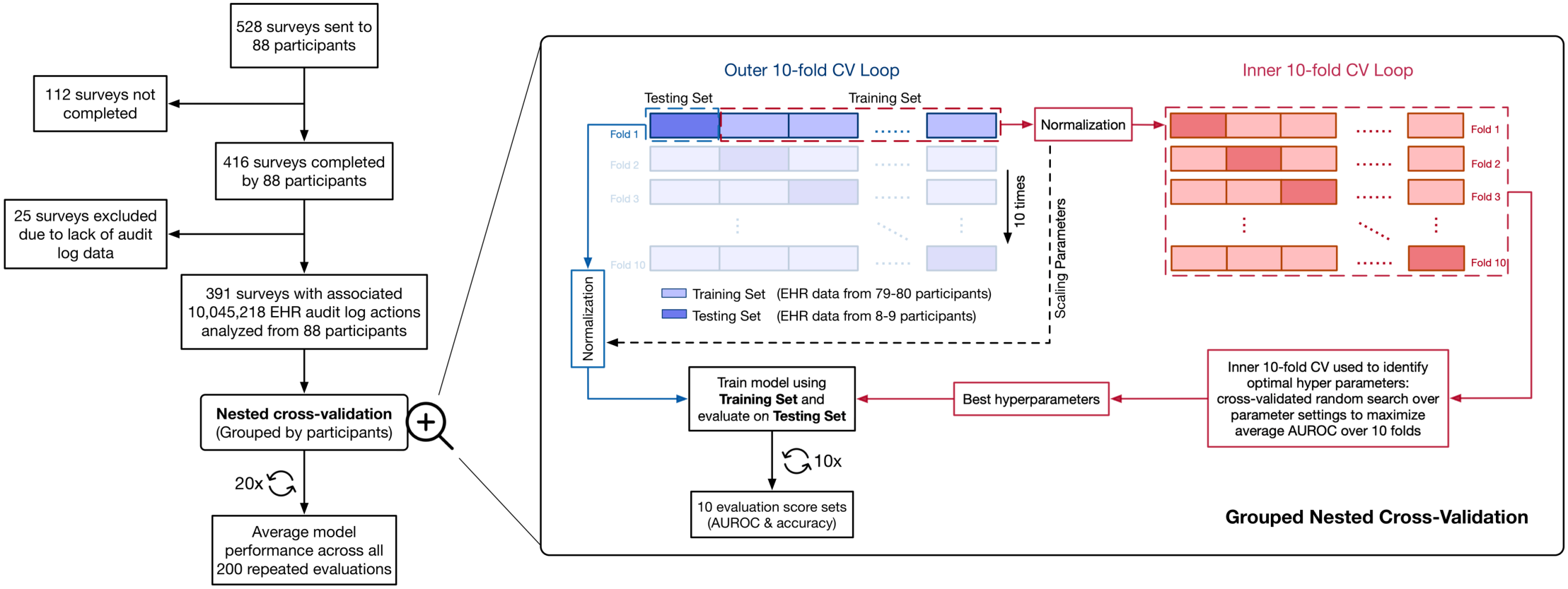
Study Design and Data Collection

CPSL
Cyber-Physical Systems Laboratory



Machine Learning Evaluation Pipeline

CPSL
Cyber-Physical Systems Laboratory



Lou et al. (2022) *Journal of Biomedical Informatics*.

Workload and temporal features weakly predict burnout

CPSL
Cyber-Physical Systems Laboratory

Feature Set	Best Model	Mean Absolute Error	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)

Median PFI burnout score: 1.2 (IQR 0.7-1.7)

PFI score > 1.33 used to indicate burnout

Lou et al. (2022) *Journal of Biomedical Informatics*.

Baseline burnout is highly predictive of future burnout

CPSL
Cyber-Physical Systems Laboratory

Feature Set	Best Model	Mean Absolute Error	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 + Workload	Neural Network	0.423 (0.293, 0.567)	0.829 (0.607, 0.996)	0.781 (0.587, 0.936)
First Survey Score	Neural Network	0.432 (0.304, 0.570)	0.819 (0.551, 0.999)	0.765 (0.547, 0.952)

Lou et al. (2022) *Journal of Biomedical Informatics*.

Limitations

CPSL
Cyber-Physical Systems Laboratory

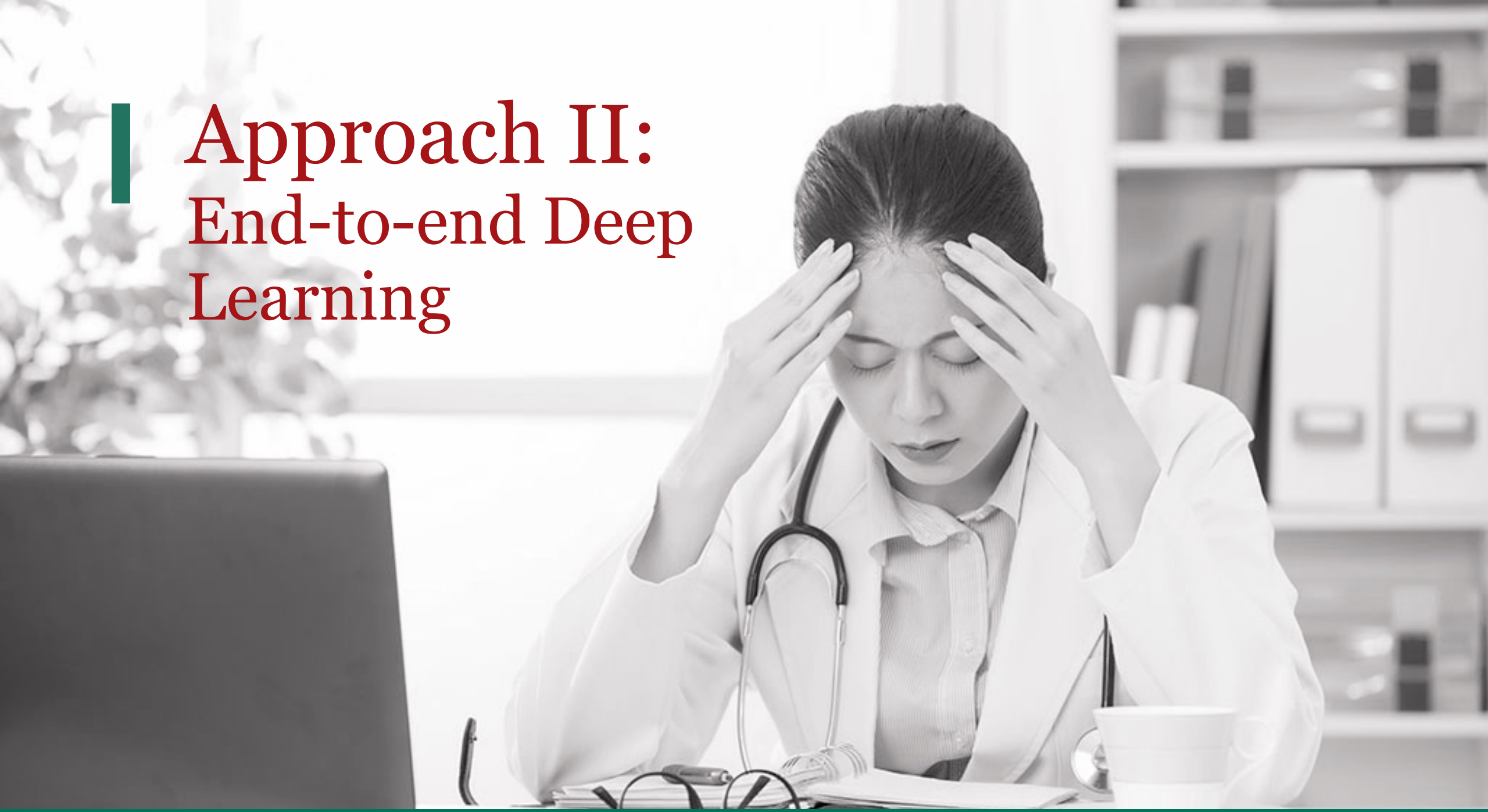
Study Design

- Small participant set (88 physicians x 6 months)
- Resident physicians may not be the right population to develop a stable model – not “steady state”
- Inter-individual variability in response to workload & workflow

Modeling

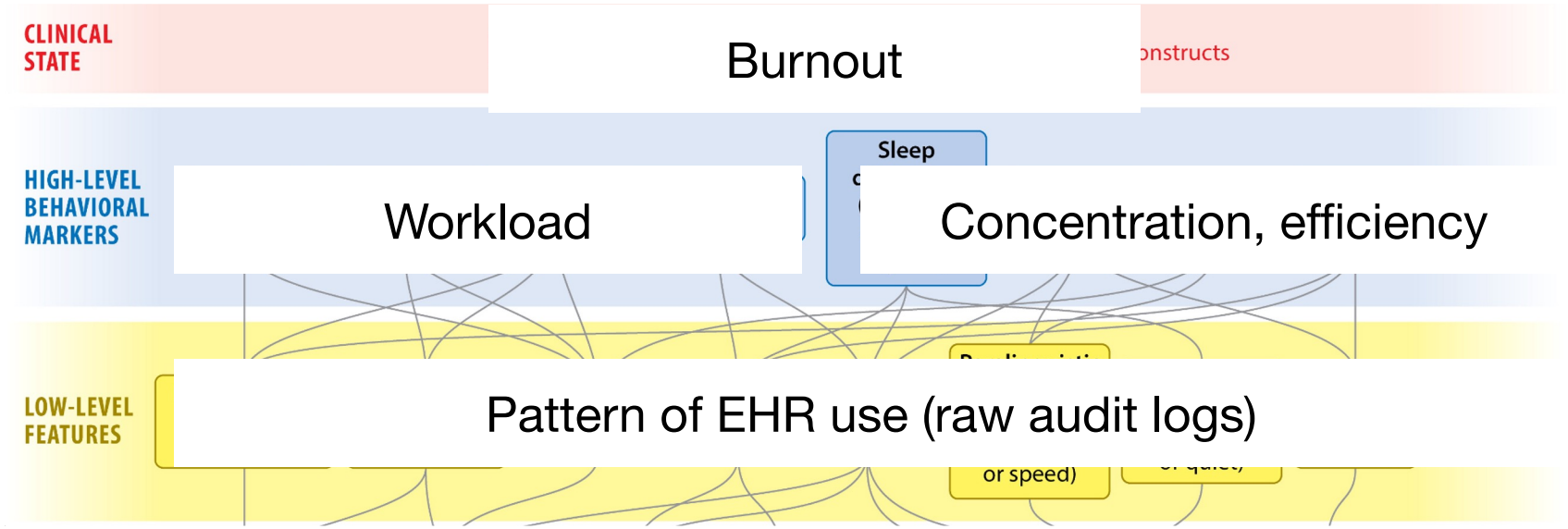
- Inability to capture non-EHR work, personal, environmental factors that also contribute to burnout
- Heavily require domain knowledge for feature engineering

Approach II:
End-to-end Deep Learning



Directly Using Raw Logs

CPSL
Cyber-Physical Systems Laboratory



Raw audit logs

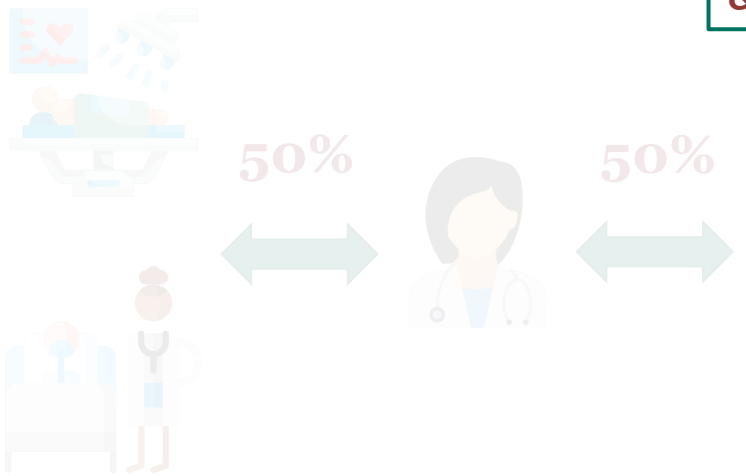
Mohr et al (2017) *Ann. Rev. Clin. Psych.*

Proposal: End-to-end Burnout Prediction

CPSL
Cyber-Physical Systems Laboratory

Goal: A system that can unobtrusively monitor and assess the risk of burnout in real time

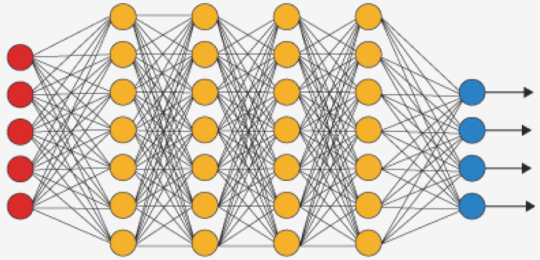
Available in all hospitals & health centers





Electronic Health Records (EHR)

Deep Neural Networks





Activity Logs
(digital footprint)

Workload
(e.g., EHR time, number of patients)



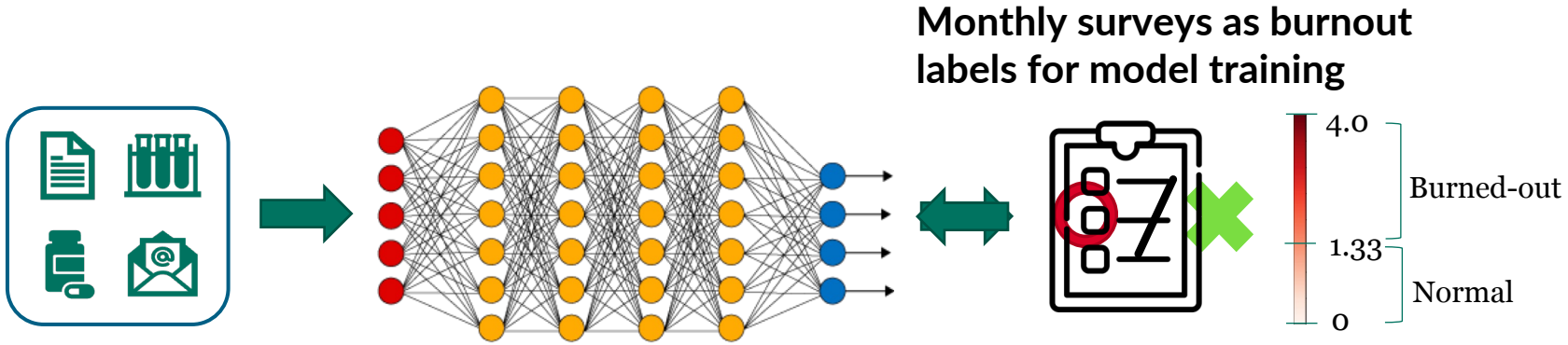
Burnout Outcome

Running in Background (Unobtrusive)

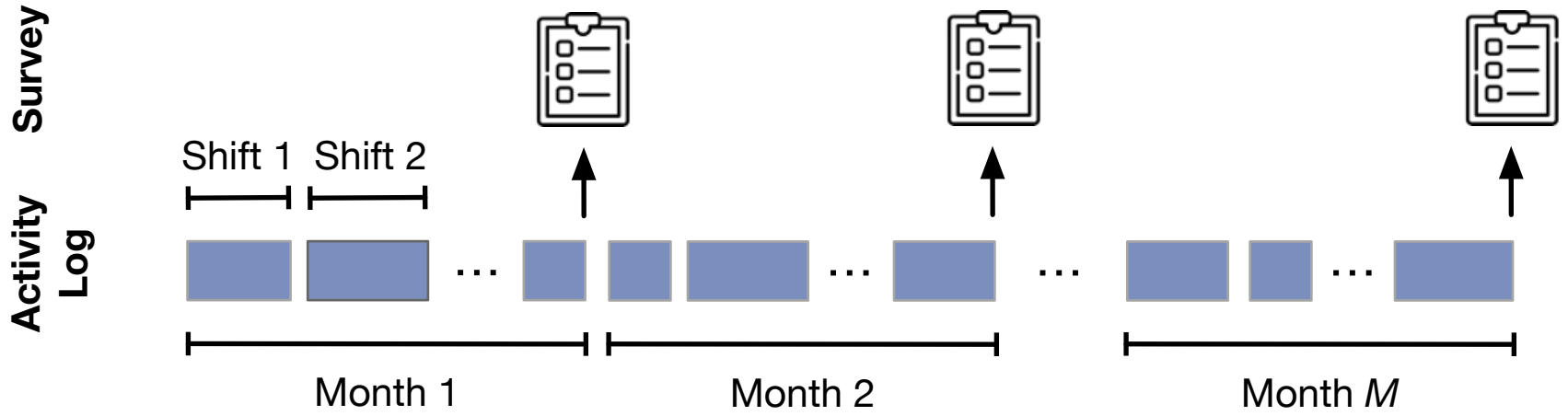
Data Collection

CPSL
Cyber-Physical Systems Laboratory

Model Training



Data Collection



(till now, 88 intern physicians across 6 months)

CPSL
Cyber-Physical Systems Laboratory

Challenges & Proposed Model

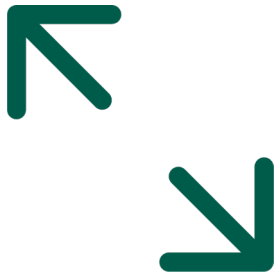
Challenges to End-to-end Modeling

CPSL
Cyber-Physical Systems Laboratory



Challenge I

Content Understanding (Activity2Vec)



Challenge II

Large-scale Model Input



Challenge III

Limited Labels

I –Understanding Raw Activity Logs

CPSL
Cyber-Physical Systems Laboratory

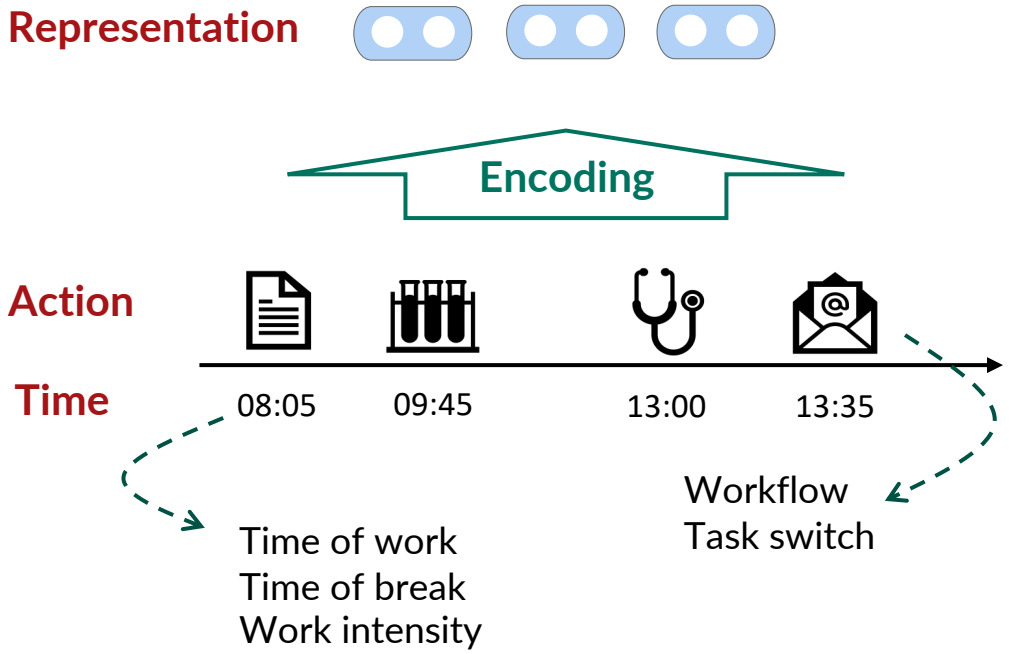
Challenge I

Raw Content Understanding

Extract useful (data-driven) knowledge from raw activity logs.

- Over 1,900 unique kinds of actions
- Irregular intervals
- Need capture dynamics/temporality

No standard data encoding method for clinical activity logs.



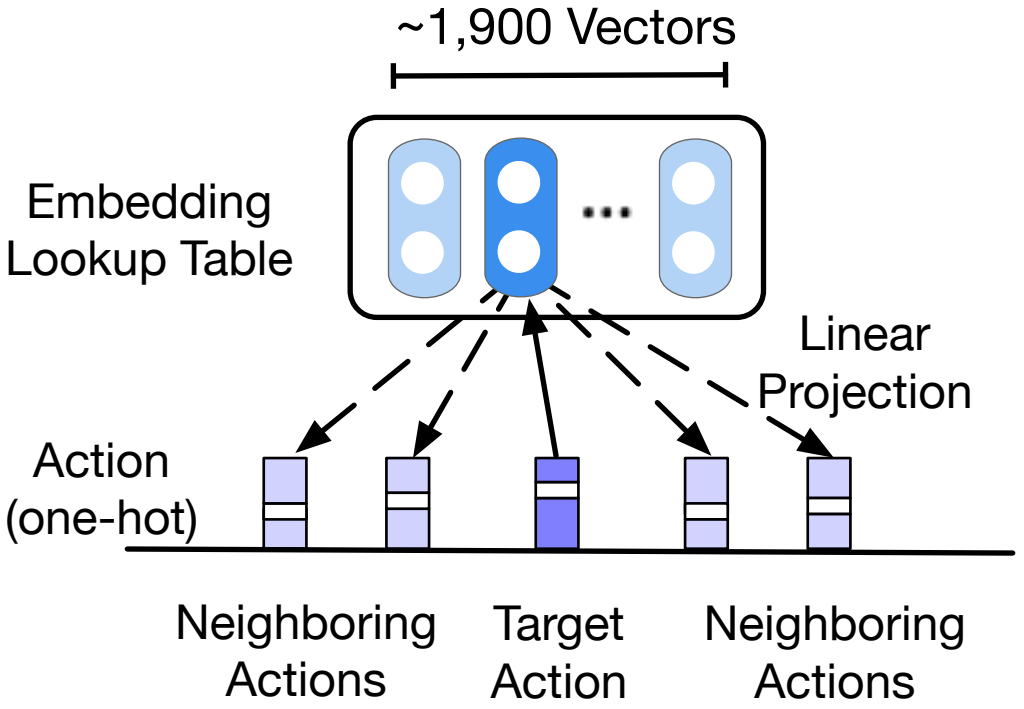
Learning Activity Context



A: Note Review
B: Report Review

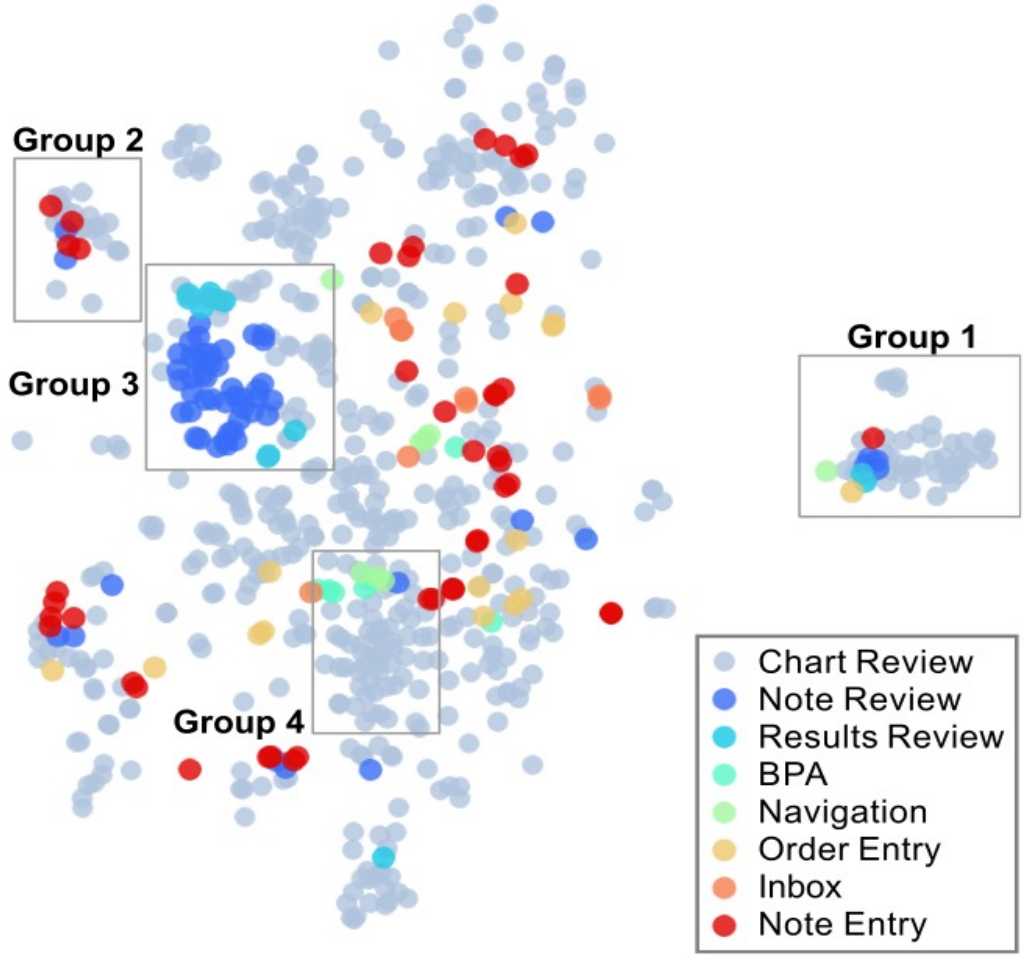
Semantically similar

Similar embedding (representation)



"Skip-Gram Algorithm"

Learned Action Embedding



Identified Groups

Group 1:
Mobile EHR client

Group 2:
Preoperative assessment module

Group 3:
Inpatient pre-rounding

Group 4:
Outpatient setting

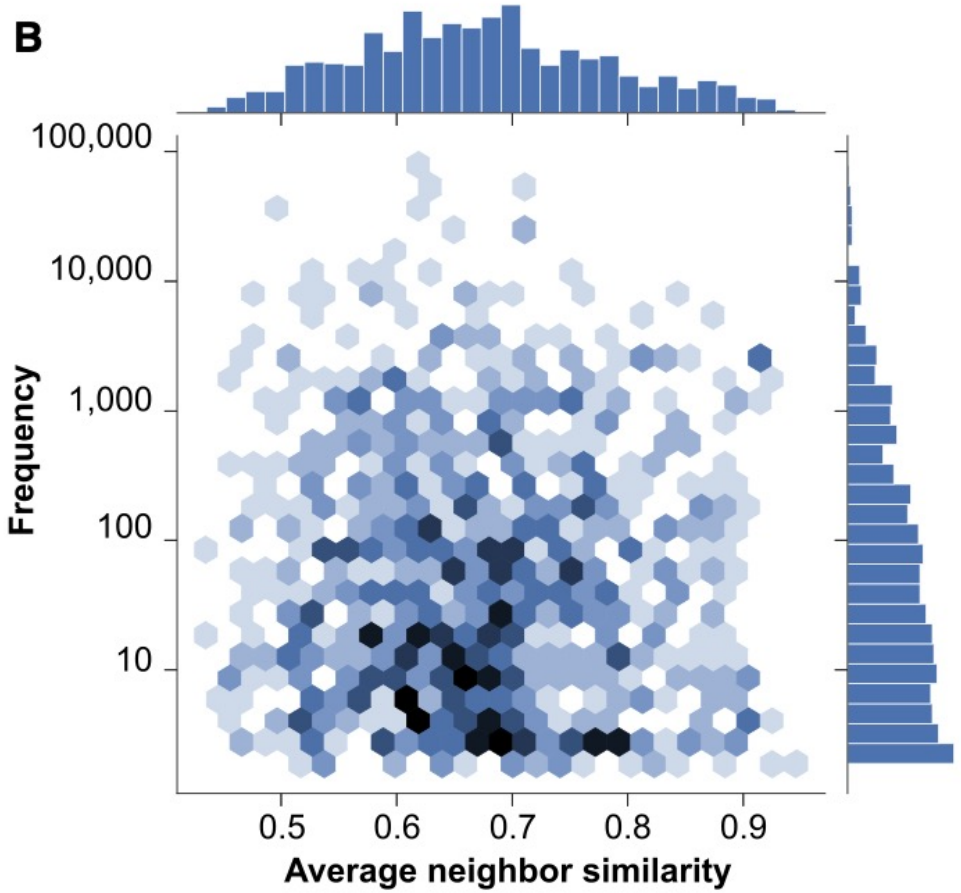
Learned Action Embedding

A

Most similar actions to 'Results Review accessed'

Action	Similarity
Results Review exited	0.779
Report viewed - Radiology Results	0.634
Report viewed - Result Detail	0.626
Report viewed - Lab Result Template	0.606
Report viewed - IP Vitals	0.582
Chart Review Encounters tab selected	0.557
Report viewed - CV Echo	0.551
Order list changed	0.550
Report viewed for an order	0.537
Report viewed - IP Flowsheet	0.529

B



Action-time Joint Embedding

Contextual **Action** Embedding

$f_A(\boldsymbol{e}_j)$

Time **Interval** Embedding

$f_I(t_{i+1} - t_i)$

Time **Periodicity** Embedding

$f_P(\sin(t_i))$

⇨

Action-time **Joint** Embedding

$\boldsymbol{g}_i = f(\boldsymbol{e}_j, t_i)$

$\boldsymbol{e}_j$: one-hot encoding of the j-th unique action

$\boldsymbol{f}$: Function implemented by a neural network

Action-time Joint Embedding

Action Embedding

$$\mathbf{a}_i^{(k)} = \mathbf{W}_a \mathbf{e}_i^{(k)}$$

Time **Interval** Embedding

$$\mathbf{b}_i^{(k)} = \tanh\left(\mathbf{W}_b \cdot \log\left(t_i^{(k)} - t_{i-1}^{(k)}\right) + \mathbf{d}_b\right)$$

Time **Periodicity** Embedding

$$\mathbf{c}_i^{(k)}[j] = \begin{cases} \omega_j t_i^{(k)} + \varphi_j, & \text{if } j = 1 \\ \sin(\omega_j t_i^{(k)} + \varphi_j), & \text{if } 1 < j \leq d \end{cases}$$

$$\mathbf{g}_i^{(k)} = \text{Concat}([\mathbf{a}_i^{(k)}; \mathbf{b}_i^{(k)}; \mathbf{c}_i^{(k)}])$$

Action-time **Joint** Embedding

II – Modeling on Long Dependencies

CPSL
Cyber-Physical Systems Laboratory

Challenge II

Large-scale Model Input

- 20,000 ~ 90,000 actions per month*person
- 1,000 ~ 8,000 actions per shift*person
- Need Long dependencies

Table: Statistics of Dataset

Statistics	Number
# total data points (actions)	15,767,634
# types of actions	1,961
# participants	88
# total months of audit logs	754
# months with eligible surveys (labels)	391

e.g., single-level RNNs

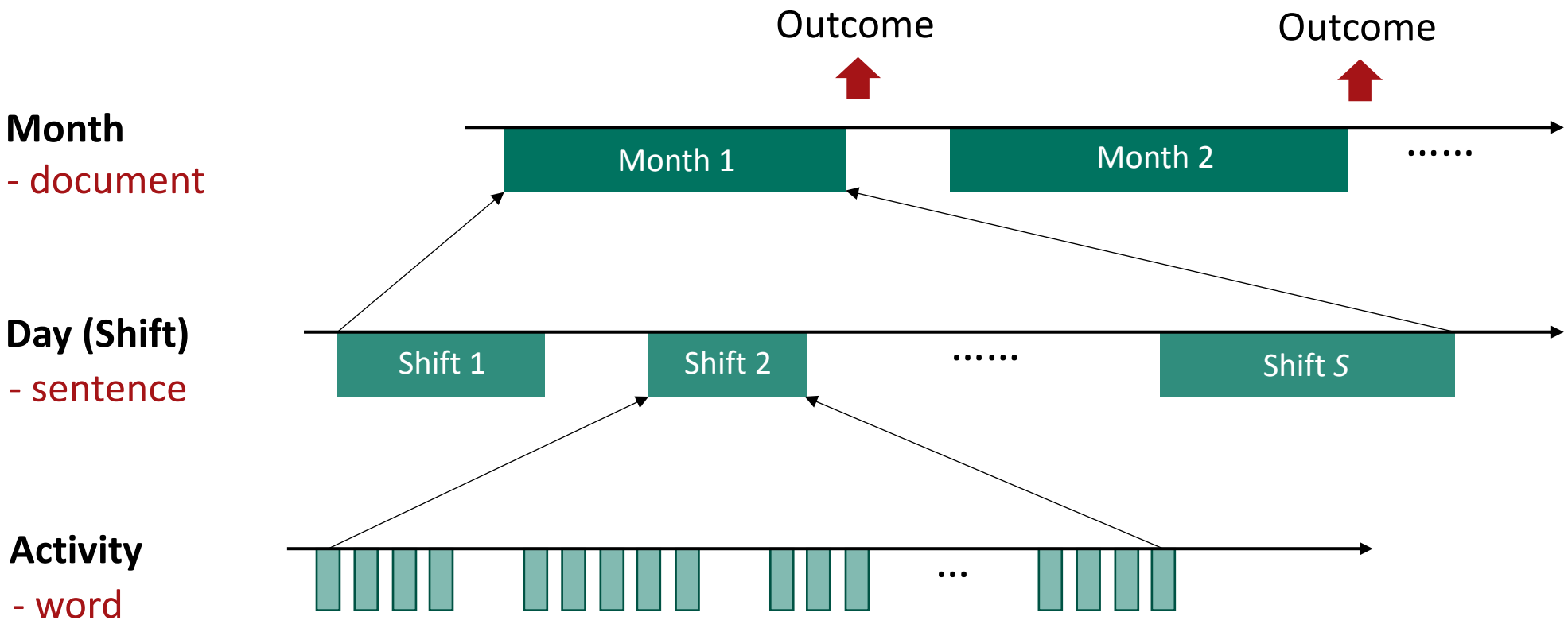
- "no-brainer" for sequential data
- Recurrent structure → slow to train
- Gradient issue → worse for long data sequence

Model	**Training Time** (per epoch)
RNNs	> 5 h

RNN: Recurrent Neural Networks

Hierarchical Sequence Modeling

CPSL
Cyber-Physical Systems Laboratory



Temporal clustering of activities may contain useful information associated with burnout.

Hierarchical Sequence Modeling

High-level Encoder:
Daily measurement → Monthly measurement

Hi
Er

Lo
En

Low-level Encoder:
Activity embedding → Daily measurement

f_{High}:
RNNs
(e.g., LSTM)

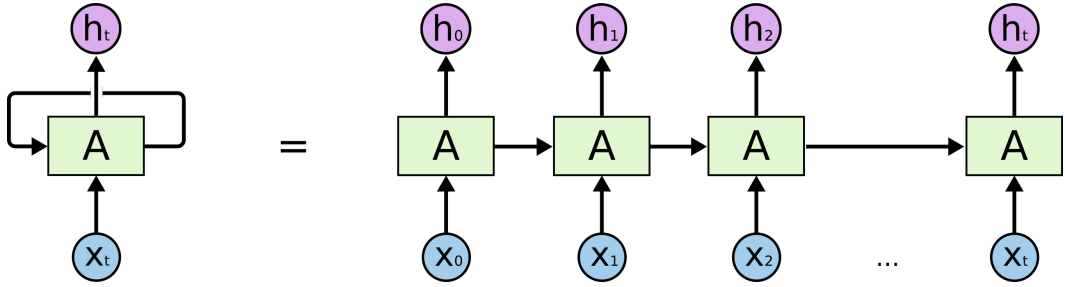
f_{Low} :
Convolutional
(e.g., TCN)

Raw data encoding:
Raw activity logs → Activity embedding (action & time)

Hierarchical burnout Prediction based on Activity Logs (HiPAL)

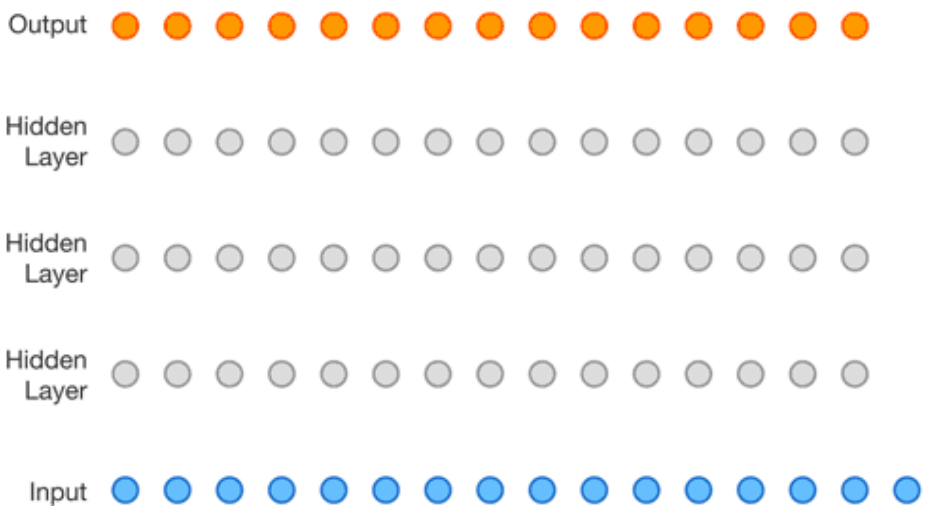
Design Choice: Sequence Models

Recurrent Neural Networks (RNN)
(e.g., LSTM, GRU)



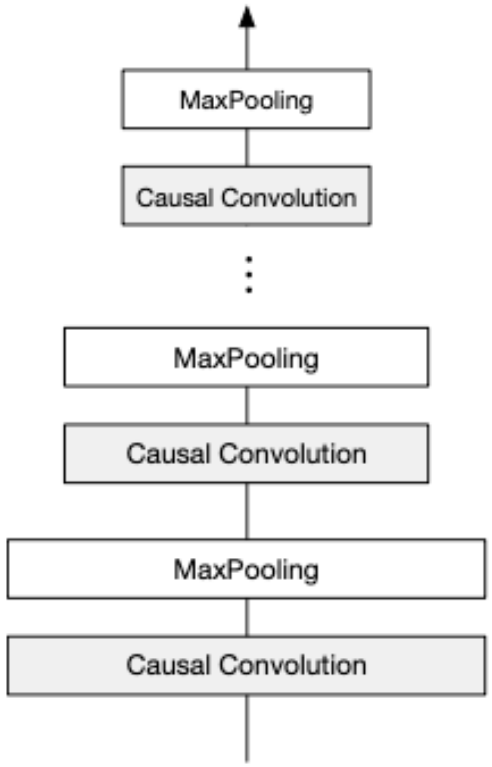
- Compute steps recurrently
- Time-consuming
- Gradient issue for long sequences

Temporal Convolutional Networks (TCN)

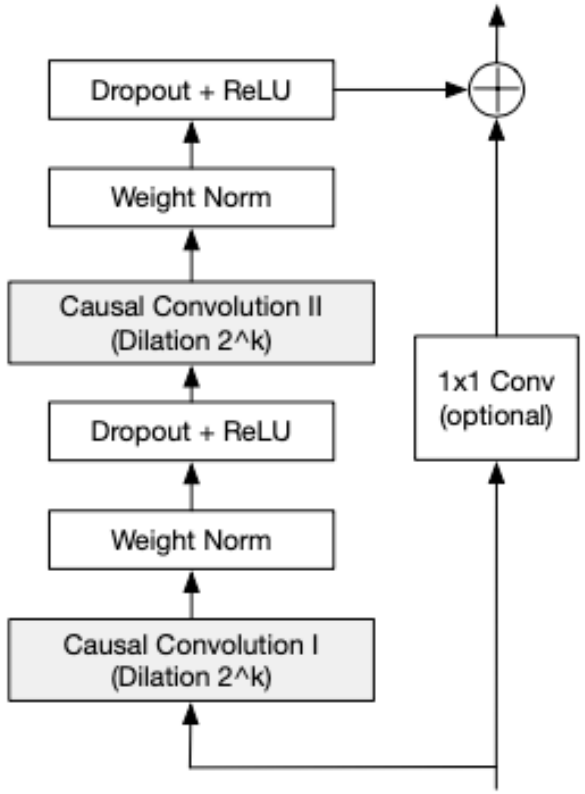


- Compute steps concurrently
- Efficient for long sequences
- Behave like RNN, compute like CNN

Variations of TCN



(A) CausalNet



(B) ResTCN

$$\mathbf{y} = \mathbf{W} * \mathbf{x} + \mathbf{d}$$
$$\mathbf{z} = \mathrm{BatchNorm}(\mathbf{y})$$
$$\mathbf{x}' = \mathrm{ReLU}(\mathbf{z})$$

(C) Fully Convolutional Networks (FCN)

III – Utilizing Unlabeled Data

CPSL
Cyber-Physical Systems Laboratory



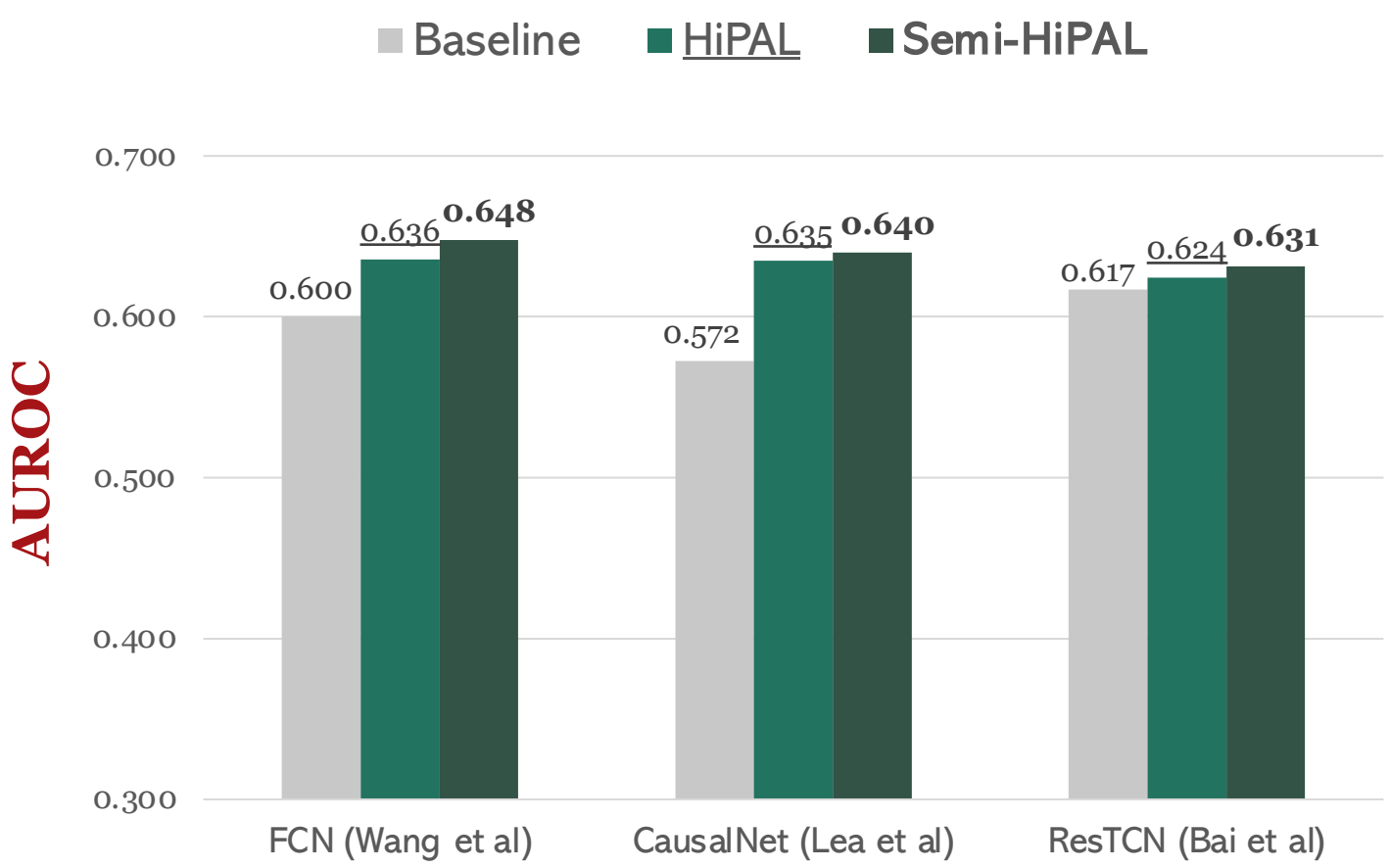
Challenge III
Limited Labels

CPSL
Cyber-Physical Systems Laboratory

Experiment

Prediction Performance

CPSL
Cyber-Physical Systems Laboratory



AUROC (higher → better):
Average true positive rate over all possible true negative rate

Training Efficiency

Method	Training Time (per epoch)
RNNs	> 5 h
HiPAL	< 1 min

Size of training set - 6 million activity logs
GPU - Nvidia RTX 3090

Practical Effectiveness

CPSL
Cyber-Physical Systems Laboratory

AUROC = 0.6358

% True Positive Rate ⟷ (Trade off) % True Negative Rate

	Intervention	Cost	Model Preference	% True Positive	% True Negative
I	 EHR training Online therapy	Low	Sensitive	80%	41.5%
II	 Days off Vacations	High	Specific	47.8%	80%

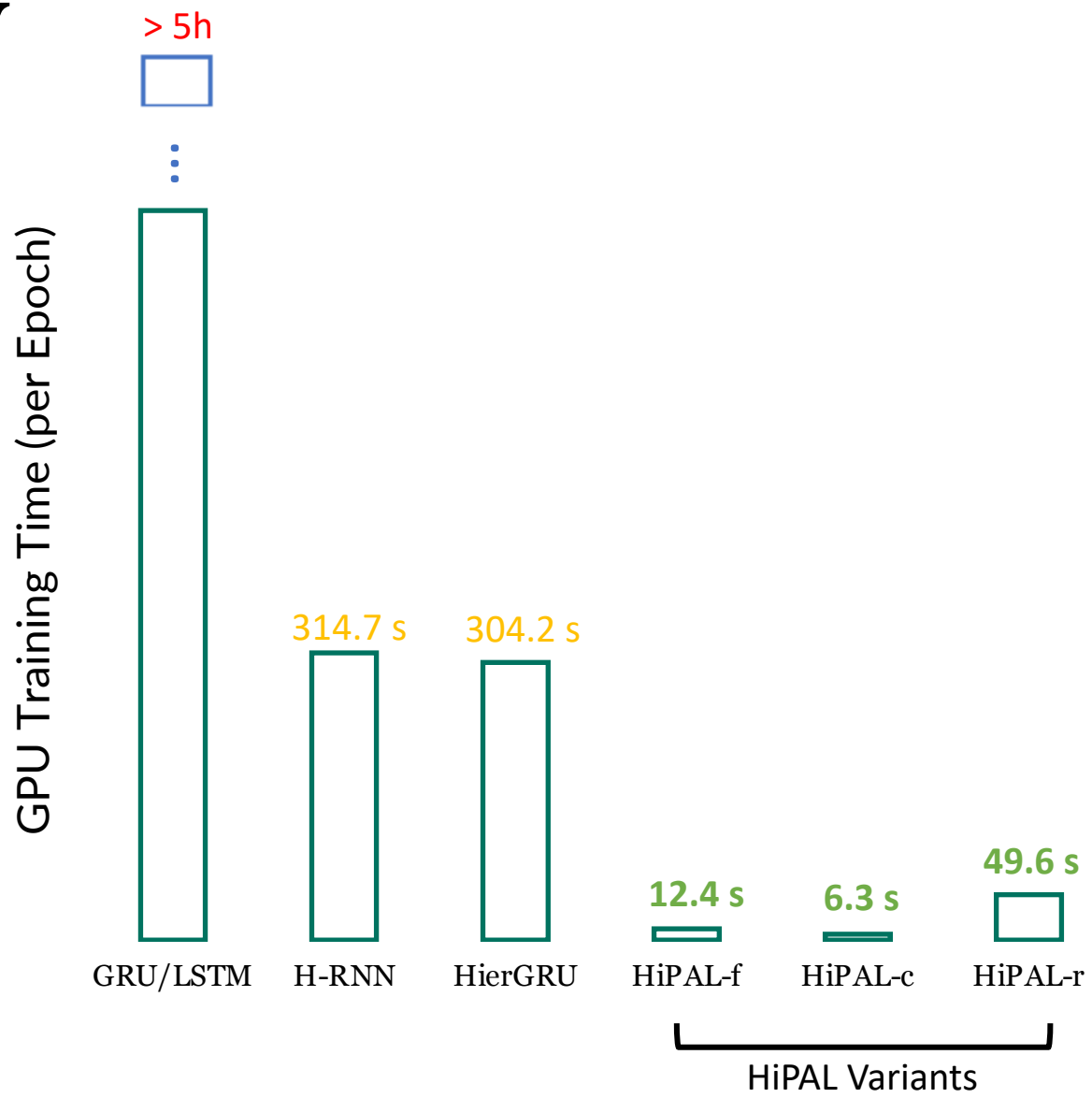
Training Efficiency

Single-level:
FCN, CausalNet and ResTCN

Hierarchical:
H-RNN: multi-level LSTM for video classification
HierGRU: replace f_{Low} as GRU

Ours:
HiPAL-f: f_{Low} = FCN
HiPAL-c: f_{Low} = CausalNet
HiPAL-r: f_{Low} = ResTCN
(f_{High} = LSTM)

RNN-based:
GRU and LSTM (recurrent, slow)



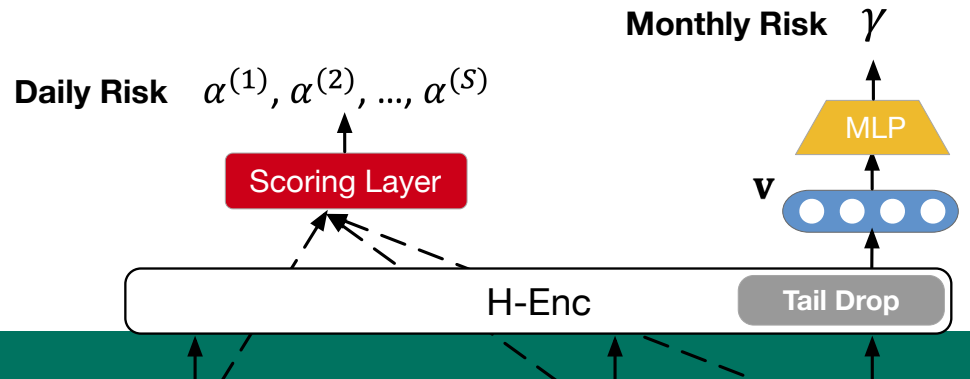
Temporal Consistency Regularization

Daily risk score: $\alpha^{(k)} = \text{Softmax}(\mathbf{W}_R \cdot \mathbf{r}^{(k)} + \mathbf{d}_R)$

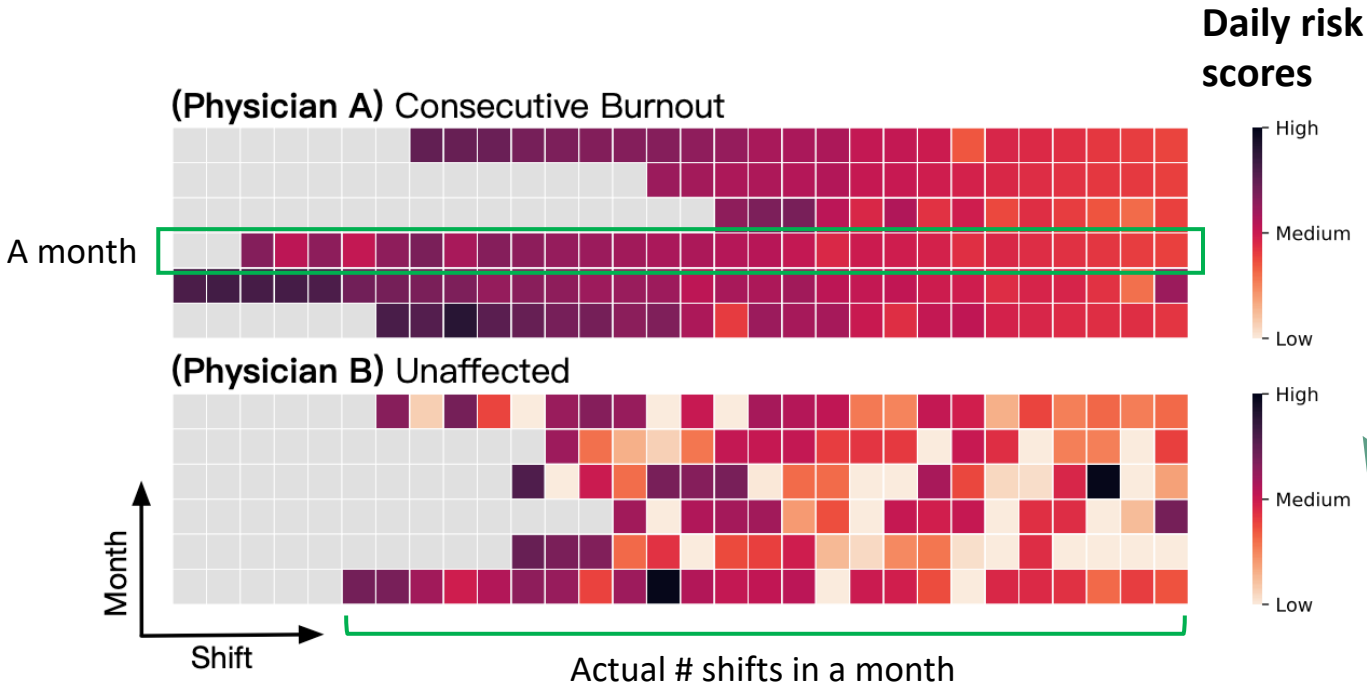
Daily loss: $\mathcal{L}_L = \frac{1}{T}\sum_{k=1}^{T} \text{CE}(\alpha^{(k)}, y)$

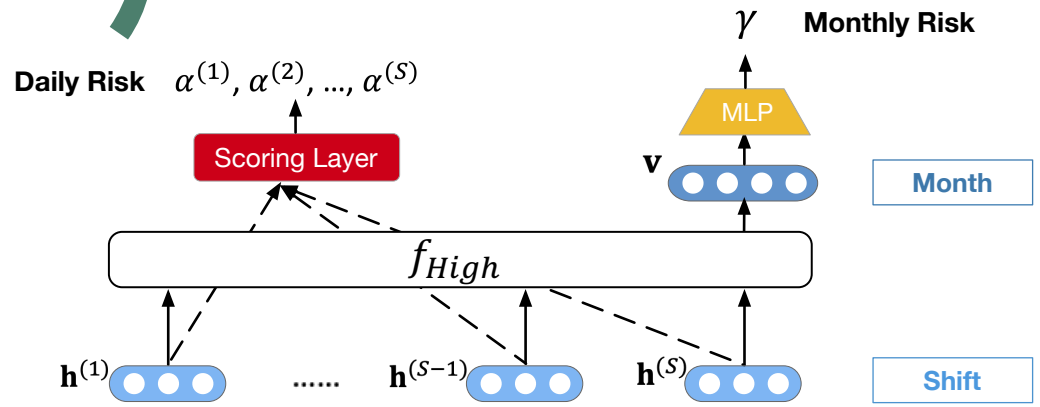
Monthly loss: $\mathcal{L}_H = \text{CE}(\gamma, y) = y \log \gamma + (1-y)\log(1-\gamma)$

Overall loss: $\mathcal{L} = \mathcal{L}_H + \lambda \mathcal{L}_L$



Interpretable Prediction





Can demonstrate how **daily risk** evolved over time and accumulated to final **monthly risk** of burnout.

Ablation Study

CPSL
Cyber-Physical Systems Laboratory

TC:
Temporal Consistency

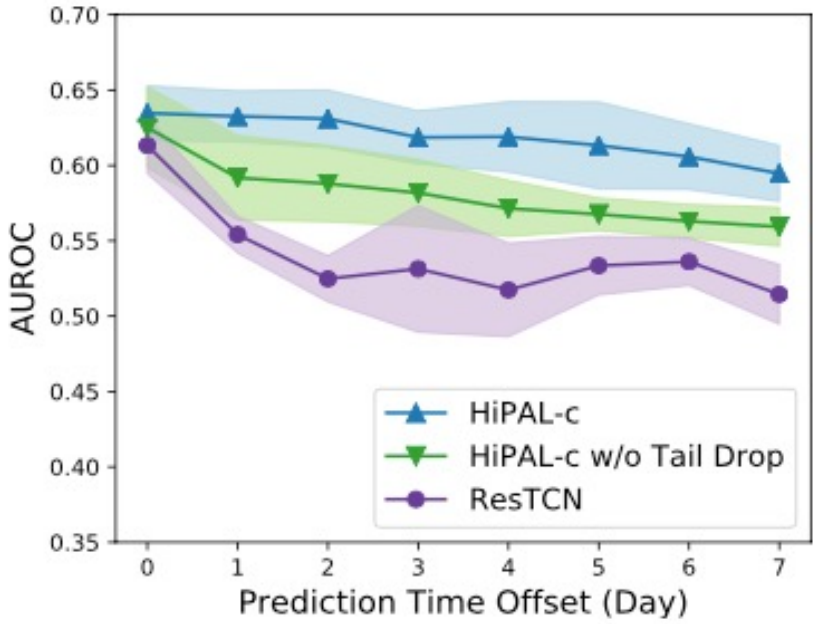
Pretrain:
Action embedding pretraining (unsupervised)

Concat -- > Add:
Replace the embedding aggregation function from concatenation to numerical addition

Variant	AUROC	AUPRC	Accuracy
w/o Regularization			
w/ Different Embedding Settings			
w/ All	**.6347 (.0181)**	**.5502 (.0191)**	**.6400 (.0218)**

Ablation Study - Taildrop

CPSL
Cyber-Physical Systems Laboratory



CPSL
Cyber-Physical Systems Laboratory

Conclusion

Conclusion

CPSL
Cyber-Physical Systems Laboratory

- End-to-end deep learning-based burnout prediction framework
- Learn contextual representations directly from raw activities
- Capture natural hierarchical structure of clinician activity logs
- Model long dependencies on large-scale activity logs with high efficiency
- Facilitate fast burnout phenotyping and individualized intervention

Limitations

CPSL
Cyber-Physical Systems Laboratory

Study Design

- Small participant set (88 physicians x 6 months)
- Resident physicians may not be the right population to develop a stable model – not “steady state”
- Inter-individual variability in response to workload & workflow

Modeling

- Did not consider dependencies across months
- Lack of lower-level explainability

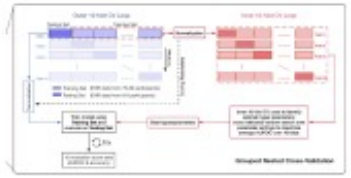
Evaluating the Burnout Project

CPSL
Cyber-Physical Systems Laboratory

1. Making the prediction actionable
   - Outcome: Burnout as defined by PFI survey
   - Clinical action: Interventions are unclear (scheduling, training, resources)
2. Potential biases in the outcome
   - Outcome is subjective, has reasonable validity but is imperfect
3. Training data quality
   - Small dataset (~80 people, ~300 surveys), maybe wrong population, not representative
   - Audit logs are reproducible and automatically generated
4. Practical predictors
   - Audit logs are required by law but are data intensive
5. Clinically meaningful measures of performance
   - AUROC because clinical action was unclear
   - Didn't measure calibration

Related Published Work

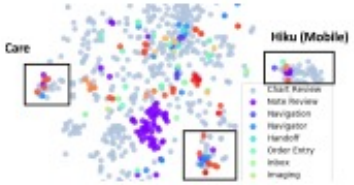
CPSL
Cyber-Physical Systems Laboratory



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

Sunny S. Lou, **Hanyang Liu**, Benjamin Warner, Derek Harford, Chenyang Lu, Thomas Kannampallil

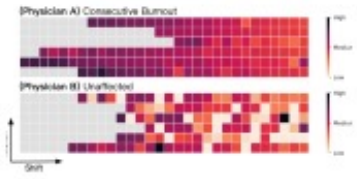
Journal of Biomedical Informatics, 2022. [pdf]



Characterizing the Microstructure of EHR Work Using Raw Audit Logs: An Unsupervised Action Embeddings Approach

Sunny S Lou, **Hanyang Liu**, Derek Harford, Chenyang Lu, Thomas Kannampallil

Journal of the American Medical Informatics Association (**JAMIA**), 2022. [pdf]



HiPAL: A Deep Framework for Physician Burnout Prediction Using Activity Logs in Electronic Health Records

Hanyang Liu, Sunny S. Lou, Ben Warner, Derek Harfort, Thomas Kannampallil, Chenyang Lu

ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2022. [pdf] [code][presentation]

Acknowledgement

CPSL
Cyber-Physical Systems Laboratory


Hanyang Liu


Sunny S. Lou


Ben Warner


Derek Harford


Thomas Kannampallil


Chenyang Lu

This study was funded by the Fullgraf Foundation, the Washington University/BJC HealthCare Big Ideas Innovation Award, and NIH 5T32GM108539-07.