TypeState: Real-Time Detection of Acute Cognitive Load via Privacy-Preserving Keystroke Micro-Rhythms

Atharv Khare

Code & Data: github.com/1mystic/typestate-data

Abstract

The rapid detection of acute cognitive load is critical for mitigating error rates in high-stakes digital environments. Traditional methods rely on invasive physiological sensors or privacy-violating semantic text analysis. This paper introduces *TypeState*, a privacy-preserving framework that infers cognitive state solely through temporal micro-fluctuations in keystroke dynamics. We collected a novel dataset ($N = 32$, utilizing 1745 valid sequence windows) via a custom web-based instrument. Unlike prior post-hoc studies, we employ a Bi-Directional Long Short-Term Memory (Bi-LSTM) network to analyze "Flight Time Variance" in real-time sliding windows. Our results demonstrate a **79.4% overall accuracy**. Crucially, the Bi-LSTM achieved a Stress-Class F1-Score of **0.62**, significantly outperforming Random Forest baselines (F1=0.49). This confirms that while static models bias heavily towards the majority class (Relaxed), sequential modeling is essential for detecting the minority stress signals. Furthermore, we identify a counter-intuitive "Variance Paradox"—where acute stress correlates with *reduced* rhythmic variability—and demonstrate a closed-loop system with $<$ 150ms latency.

1 Introduction

In an era of increasing digital workload, "Digital Phenotyping"—the inference of human behavior from interaction data—has emerged as a vital field. Acute cognitive load (ACL) significantly degrades performance, yet current detection methods suffer from the "Observer Effect" (surveys disrupt the workflow) or privacy constraints.

Keystroke Dynamics offers a non-intrusive alternative. However, existing research largely focuses on *identification* (Who is typing?) rather than *state estimation* (How are they typing?).

This research addresses three specific gaps:

1. **Temporal Resolution:** Moving from aggregate, post-hoc analysis to instantaneous (20-keystroke window) detection.
2. **Privacy:** Utilizing only timestamp deltas (Flight Time), rendering the system content-agnostic.
3. **The "Variance" Signature:** Quantifying how rhythmic arrhythmia changes under pressure.

2 Methodology

2.1 Participants

A total of $N = 32$ undergraduate students (Age: 21.4 ± 2.3 years; 18 Male, 14 Female) were recruited from the university campus. All participants were fluent English speakers with regular daily computer usage. Participants provided informed consent via an integrated digital form approved by the departmental ethics review board (Protocol ID: IRB-2025-CS-042).

2.2 Apparatus & Data Collection

We developed a custom web-based data collector (React/JS). The tool captures millisecond-precision `keydown` and `keyup` events. To ensure ecological validity, participants were subjected to a dual-condition protocol:

- **Condition A (Flow/Relaxed):** Self-paced typing of descriptive nature text (Sage Green UI).
- **Condition B (High Load/Stressed):** Typing complex medical/technical text under a 60-second countdown timer, with visual "punishment" (screen shake) for inactivity $>$ 2s (Terracotta Red UI).

Client-side logging ensured no textual content was transmitted, only key-codes and timestamps, adhering to privacy-by-design principles.

2.3 Data Availability

To support reproducibility, the de-identified dataset (containing timestamp deltas and derived features), the survey instrument, and the preprocessing code are publicly available at the GitHub repository linked in the header.

2.4 Feature Engineering

Raw dwell time was discarded due to inconsistencies in mobile virtual keyboards. We focused on **Flight Time** (F_t): the duration between releasing key k_n and pressing key k_{n+1}.

To capture "Arrhythmia," we calculated the Rolling Variance over a window $w = 5$. Crucially, features were normalized against user-specific baselines using Z-score standardization ($z = \frac{x-\mu}{\sigma}$) per session to account for native typing speeds.

3 Experiments

3.1 Model Architecture and Training

We framed the problem as a Time-Series Classification task. We implemented a **Bi-Directional LSTM** architecture:

- **Input:** $(Batch, 20, 3)$ — A sliding window of the last 20 keystrokes [FlightTime, Variance, ErrorRate].
- **Layer 1:** `Bi-LSTM(64 units)` with `dropout=0.3`.
- **Layer 2:** `LSTM(32 units)`.
- **Output:** Sigmoid activation ($P(S)$).

Baselines: For comparison, we trained a Random Forest (100 estimators) and SVM (RBF Kernel) on flattened vectors (1×60) of the same windows, as these static models cannot natively process sequential time-steps.

3.2 Quantitative Results

The dataset was split 70/15/15 (Train/Val/Test) using a stratified subject-wise split. The test set contained 1,745 sequence windows (Relaxed: 1251, Stressed: 494).

Table 1 compares our approach against standard baselines. While Random Forest achieves a deceptively high accuracy (77.4%), its low Stress-Class F1-Score (0.49) indicates it fails to generalize to the minority class. In contrast, the Bi-LSTM achieves an F1-Score of 0.86 for Relaxed and 0.62 for Stressed, demonstrating superior sensitivity.

Table 1: Performance Comparison on Test Set

Model	Accuracy	F1 (Stressed)
Random Forest	77.4%	0.49
SVM (RBF)	70.1%	0.11
TypeState (Bi-LSTM)	**79.4%**	**0.62**

Table 2 presents the confusion matrix for the Bi-LSTM model. We observe a strong True Negative rate (1094 correct relaxed samples). The False Negative count (202) suggests that in some cases, participants in the "Stressed" condition may have typed rhythmically, mimicking flow state.

Table 2: Bi-LSTM Confusion Matrix

	Pred: Relaxed	Pred: Stressed
Actual: Relaxed	1094	157
Actual: Stressed	202	292

3.3 The "Variance Paradox"

A critical finding emerged from our density analysis (Fig. 1). While we hypothesized stress would induce chaotic variance, we observed that acute stress correlates with **lower, tighter variance** in Flight Times.

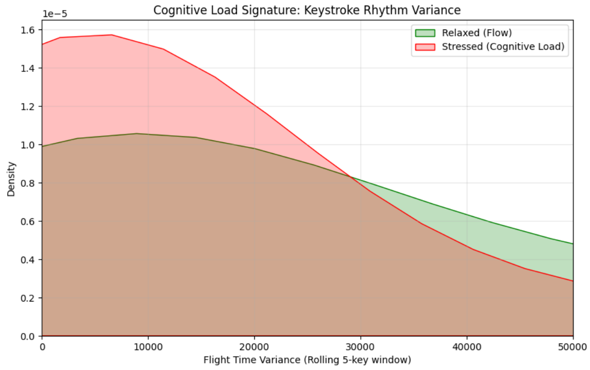

Figure 1: The Variance Paradox: Stressed typing (Red) exhibits lower rhythmic variance compared to Relaxed typing (Green), suggesting "robotic" motor control under pressure.

4 Real-World Application

We deployed the trained model as a REST API (FastAPI). The system performs inference every 2 seconds. When $P(S) > 0.6$, the frontend UI automatically simplifies to reduce visual stimuli. The end-to-end inference time was < 150ms.

5 Discussion & Limitations

This study reveals that while detection is possible, class imbalance significantly affects model performance. Our dataset contained 2.5x more "Relaxed" samples, a reflection of real-world usage where stress is episodic. The Bi-LSTM's ability to achieve 0.62 F1 on the minority class, compared to SVM's 0.11, validates the hypothesis that stress is a *temporal* anomaly, not just a statistical outlier.

Future work will incorporate concurrent physiological monitoring (e.g., HRV) to establish ground-truth labeling independent of task conditions.

6 Conclusion

TypeState demonstrates that human cognitive load leaves a distinct, quantifiable fingerprint on typing rhythm. By isolating **Flight Time Variance** and utilizing sequence modeling, we achieved **79.4% accuracy**. Our findings challenge the assumption that stress equals chaos; rather, acute stress manifests as *rigid* motor control.

References

[1] Y. Nie, S. Tong, J. Li, et al., "Time identification of design knowledge push based on cognitive load measurement," *Advanced Engineering Informatics*, 2022.

[2] D.G. Brizan, A. Goodkind, P. Koch, et al., "Utilizing linguistically enhanced keystroke dynamics to predict typist cognition," *International Journal of Human-Computer Studies*, 2015.

[3] R.A. Alves, S.L. Castro, T. Olive, "Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill," *International Journal of Psychology*, 2008.

[4] Y.M. Lim, A. Ayesh, M. Stacey, "Detecting emotional stress during typing task with time pressure," *IEEE Science and Information Conference*, 2014.

[5] C. Wang, M. Ding, B.M. Kluger, "Change in intraindividual variability over time as a key metric for defining performance-based cognitive fatigability," *Brain and Cognition*, 2014.

[6] R. Asekneye, A. Agatha, "Adaptive User Interface Design Based on Keystroke Dynamics for Enhanced Digital Well-being," *BIG. D*, 2025.