FOCUS FLOW: WORK MONITORING SYSTEM FOR REMOTE EMPLOYEES USING BEHAVIOURAL ANALYSIS

Group 1

September 4, 2025

1 Privacy-Preserving On-Screen Activity Tracking and Classification in E-Learning Using Federated Learning

Publication

Authors: D. Mistry, M. F. Mridha, M. Safran, S. Alfarhood, A. K. Saha, D. Che

Published In: IEEE Access, Volume 11 (2023), Pages 79315–79329

DOI: 10.1109/ACCESS.2023.3299331

1.1 Abstract

With the rapid rise of e-learning, students increasingly rely on digital devices for education. However, the same devices are also used for entertainment and distraction, making it difficult to ensure focused learning. This study addresses the challenge of monitoring student activity while preserving privacy by introducing a federated learning-based system that classifies on-screen behavior—productive or unproductive—without transferring raw visual data off the device.

The proposed framework uses FedInceptionV3, a privacy-preserving model that processes screenshots locally and trains collaboratively using federated learning. The system leverages pre-trained deep learning architectures such as VGG16, VGG19, ResNet50, and InceptionV3, comparing their performance across a dataset of over 4,000 labeled screenshots. FedInceptionV3 achieved a leading test accuracy of 99.75%, correctly classifying 798 out of 800 screenshots, outperforming all other variants. The study demonstrates how federated learning enables effective monitoring without compromising user data privacy and presents performance evaluations using metrics such as precision, recall, F1-score, loss functions, and confusion matrices. This approach offers a scalable, privacy-focused solution for maintaining engagement in e-learning environments.

1.2 Advantages

- Very high accuracy (99.75%) in classifying productive vs. unproductive screen activity.
- Local device processing protects user privacy—no screenshots leave the user's machine.

- Federated learning allows secure model updates without centralizing user data.

1.3 Limitations

- Relies on screenshot images, not behavioral signals like mouse or keystroke patterns.
- Deep learning models are resource-heavy, limiting performance on low-end devices.
- Not real-time—it works on periodic snapshots, not continuous input streams.

1.4 Relevance

- Aligns with the goal of privacy-preserving, on-device monitoring using federated learning.
- The federated training approach can be adapted for behavioral rather than visual input data.
- The performance evaluation methodology is suitable for validating intent-aware monitoring models.

2 A Machine Learning Approach to Leverage Individual Keyboard and Mouse Interaction Behavior From Multiple Users in Real-World Learning Scenarios

Publication

Authors: S. Salmeron-Majadas, R. S. Baker, O. C. Santos, J. G. Boticario

Published In: IEEE Access, Volume 6 (2018), Pages 39154–39179

DOI: 10.1109/ACCESS.2018.2854966

2.1 Abstract

This study proposes a non-intrusive affect detection approach using keyboard and mouse interaction data in real-world educational scenarios, specifically in English-as-a-Second-Language essay writing tasks. The authors introduce a baseline interaction model—captured via an initial "calibration task"—to normalize individual user differences in typing and mouse usage skills. Features are extracted from keystroke dynamics (both key-specific and key-independent n-graphs), mouse movement metrics (speed, acceleration, distance, click features), and task performance indicators (e.g., proportion of required words used). Affect labeling uses the Self-Assessment Manikin (SAM) scale for valence and arousal, discretized in multiple ways for modeling. Data preprocessing methods, including dimensionality reduction (PCA, forward selection) and class balancing (SMOTE, equal size sampling), are tested along with algorithms such as Random Forest, C4.5, SVM, and Naïve Bayes. Experiments with 41 participants generated 512 models, evaluated via 10-fold cross-validation. The user-normalized dataset consistently outperformed raw data, achieving higher accuracy and kappa values, with the best models yielding accuracy up to 87% in binary classification of affect states and notable improvement over baseline classifiers. The study demonstrates that incorporating per-user baselines and careful preprocessing can enhance affective state prediction in realistic, low-intrusion contexts.

2.2 Advantages

- The baseline calibration approach effectively compensates for individual differences, improving predictive accuracy in real-world settings.

- Integrates keystroke, mouse, and performance metrics for richer behavioral modeling without additional hardware.
- Evaluates multiple dimensionality reduction and class balancing techniques to address small, imbalanced datasets.

2.3 Limitations

- Only 41 participants, limiting generalizability and increasing risk of overfitting despite preprocessing.
- While accuracy is high in some cases, agreement beyond chance remains modest, suggesting potential label noise or model limitations.
- Designed around ESL essay writing; feature relevance and performance may not directly transfer to other domains or behaviors.

2.4 Relevance

- Baseline normalization offers a transferable strategy for calibrating individual typing/mouse behavior before monitoring intent or focus in remote work.
- Feature engineering methods (n-graphs, mouse movement metrics) are directly applicable for intent-aware monitoring without invasive sensing.
- On-device feasibility: Demonstrates lightweight, interpretable models that can operate with minimal resource demands, aligning with privacy-preserving desktop analytics.

3 Real-Time Detection of Insider Threats Using Behavioral Analytics and Deep Evidential Clustering

Publication

Authors: Anas Ali, Mubashar Husain, Peter Hans

Published In: arXiv, May 2025, arXiv:2505.15383 [cs.CR]

DOI: 10.48550/arXiv.2505.15383

3.1 Abstract

This paper tackles the growing challenge of insider threats—malicious actions by authorized users—by introducing a real-time behavioral analytics framework enhanced with deep evidential clustering (DEC). The system captures rich user activity signals (e.g., file access, process execution, command usage), encodes these into temporal embeddings via RNNs, and applies DEC with Dirichlet distribution outputs to estimate both cluster assignments and the model's epistemic uncertainty. This dual insight allows the system to autonomously classify highly confident threats while flagging ambiguous cases for manual review—important in high-risk environments where trust and interpretability are essential. The model also incorporates online learning to adapt to behavioral drift, making it robust to evolving user patterns. Evaluated on the CERT and TWOS insider threat datasets, this method achieves an impressive average detection accuracy of 94.7%, while reducing false-positive rates by 38% compared to traditional clustering techniques such as k-means or Isolation Forests . These results showcase the value of modeling uncertainty and concept drift in real-time detection systems, offering a practical and reliable pipeline that can serve as an intelligent pre-filter for security operations environments.

3.2 Advantages

- The DEC approach quantifies confidence via Dirichlet outputs, enabling more nuanced decision-making (autonomous vs. human-reviewed alerts).
- Uses online learning and temporal embeddings to stay effective amidst evolving user behavior.
- Demonstrates strong results—94.7% accuracy and 38% fewer false positives—on industry-standard insider threat datasets.

3.3 Limitations

- Evaluated only on CERT and TWOS datasets, which may not fully reflect other enterprise contexts or real-world complexities.
- RNN-based embedding plus evidential clustering could be resource-intensive for large-scale or real-time deployment.
- While uncertainty is quantifiable, the feature-level interpretability (e.g., which behaviors trigger alerts) isn't fully explored.

3.4 Relevance

- The temporal encoding of user actions mirrors our intent-detection approach using keyboard/mouse dynamics—both rely on sequential behavior modeling.
- DEC's ability to flag low-confidence cases maps well to our project's goal of providing focused feedback while avoiding misclassification.
- Incorporating real-time drift adaptation aligns with our aim of maintaining model relevance in dynamic work-from-home environments.

4 User Authentication Method Based on Keystroke Dynamics and Mouse Movement (SIURUA)

Publication

Authors: Xiujuan Wang; Yutong Shi; Kangfeng Zheng; Yuyang Zhang; Weijie Hong; Siwei Cao

Published In: Sensors (MDPI), Volume 22 (2022), Article number 6627

DOI: 10.3390/s22176627

4.1 Abstract

This paper introduces SIURUA, a system designed to identify users based on how they type and use a mouse during a variety of activities, including text typing, online shopping, social media browsing, and gaming. The method works by separating two types of information from the captured data: scene-independent features, which remain consistent regardless of the activity being performed, and user-related features, which reflect the unique behavior patterns of each individual. These features are then combined using a Multiple Kernel Learning approach and classified with an SVM to improve recognition accuracy. The system was tested on data from 41 participants across four activity types, achieving an accuracy of 84.0%, precision of 84.1%, true positive rate of 85.0%, false positive rate of

16.9%, false reject rate of 15.0%, and an F1 score of 83.9%. Compared to other existing methods, SIURUA performed better across all measured metrics, demonstrating its ability to reliably recognize users even when their activities vary, making it suitable for use in diverse and changing real-world environments.

4.2 Advantages

- effectively handles variability across multiple user activity scenes by isolating scene-independent features and fusing them with user-specific features.
- achieves high authentication metrics (e.g., aAcc 84%, aTPR 85%) in real-world hybrid settings.
- AMI-based selection clarifies whether features emphasize identity traits or scene artifacts, aiding model transparency.

4.3 Limitations

- the dataset encompasses only four predetermined contexts; generalization to unseen or expanding usage scenarios is uncertain.
- initial keystroke feature space spans over 48,500 dimensions before reduction, requiring heavy processing.
- 5-minute time window might limit responsiveness in real-time monitoring contexts.

4.4 Relevance

- the selection and fusion of scene-independent and user-related features offers a blueprint to differentiate intent from context in desktop behavior analysis.
- the hybrid-scene framework parallels the variability in remote work environments, enhancing model resilience across task contexts.
- attaining 84% accuracy in identity inference based solely on keyboard and mouse dynamics sets realistic expectations for analogous intent detection tasks.

5 A Review of Emotion Recognition Methods from Keystroke, Mouse, and Touchscreen Dynamics

Publication

Authors: L. Yang, S.-F. Qin

Published In: IEEE Access, Vol. 9, 2021, pp. 162197–162213

DOI: 10.1109/ACCESS.2021.3132233

5.1 Abstract

Emotions play a critical role in both individual performance and team dynamics in digital work environments. This paper presents a comprehensive review of emotion recognition techniques based on keystroke, mouse, and touchscreen (KMT) dynamics, which are non-intrusive and unobtrusively collected during everyday computer interactions. Compared to facial expressions, voice, or physiological signals, KMT data offers privacy-friendly alternatives suitable for real-time monitoring.

The paper analyzes studies from the past two decades, answering six key research questions related to data elicitation methods, emotion types detectable from KMT signals, feature selection, classification techniques, practical applications, and deployment contexts. Emotional states such as stress, boredom, frustration, and happiness can be inferred through KMT features like typing speed, mouse movement patterns, and screen interaction rhythms. It also categorizes commonly used classification models (e.g., SVMs, decision trees, neural networks) and highlights performance metrics across studies. Challenges such as dataset inconsistency, limited generalization, and lack of cross-context deployment strategies are also discussed, along with a future roadmap for improving emotion recognition systems in behavioral monitoring contexts.

5.2 Advantages

- Focuses on KMT dynamics, which are non-intrusive and align well with privacy-preserving behavioral monitoring systems.
- Maps emotions to specific behavioral features, helping design feature extraction pipelines for real-time affect detection.
- Provides structured comparisons of classification models and features, useful for benchmarking and selecting model architectures.

5.3 Limitations

- Does not focus on specific environments like remote work or office settings, requiring additional adaptation.
- Heterogeneous methods across reviewed studies limit direct comparison or synthesis for implementation.
- Emotion categories differ from intent states like focus or distraction, requiring further contextual mapping for intent-aware systems.

5.4 Relevance

- Establishes the validity of KMT data as reliable input for recognizing internal states, supporting behavioral feature use in intent inference.
- Reinforces the privacy advantages of non-intrusive data collection, aligning with local-only processing and ethical monitoring goals.
- Offers guidance on feature selection and model choice, aiding implementation of real-time, behavioral-based intent recognition systems.

6 An Interpretable Machine Learning Approach to Multimodal Stress Detection in a Simulated Office Environment

Publication

Authors: Mara Naegelin, Raphael P. Weibel, Jasmine I. Kerr, Victor R. Schinazi, Roberto La Marca, Florian von Wangenheim, Christoph Hoelscher, Andrea Ferrario

Published In: Journal of Biomedical Informatics, Vol. 139, 2023, Article No. 104299

DOI: 10.1016/j.jbi.2023.104299

6.1 Abstract

This study investigates stress detection using multimodal machine learning models in a simulated office environment involving 90 participants. It combines behavioral data—keyboard and mouse dynamics—with physiological signals (heart rate variability, HRV) to predict emotional states such as perceived stress, arousal, and valence. Classifiers including support vector machines (SVM), random forests (RF), and gradient boosting models (GBM) were trained and evaluated using 10-fold cross-validation. The best-performing model (GBM) achieved an F1 score of 0.775 for valence prediction, indicating strong classification potential.

Importantly, SHAP (SHapley Additive exPlanations) analysis revealed that behavioral features (e.g., typing speed, mouse movement patterns) significantly influenced model predictions, sometimes outperforming physiological indicators like HRV. The authors highlight that while physiological fusion adds minimal value in this setting, behavioral-only approaches may be more scalable and privacy-friendly for real-world applications. The study also acknowledges the importance of baseline calibration and personalization, supporting the design of adaptive, user-specific stress monitoring systems suitable for office environments.

6.2 Advantages

- Uses real-world-like office simulations with 90 participants, improving ecological validity.
- Demonstrates strong performance using only keyboard and mouse data, validating non-invasive behavioral signals.
- Utilizes SHAP analysis to interpret model predictions, supporting transparent and explainable ML.

6.3 Limitations

- Conducted in a simulated setting, not tested in real remote work environments.
- Physiological data showed limited contribution, but fusion potential in other contexts remains unexplored.
- Personalization is necessary, but full implementation of baseline calibration was not achieved.

6.4 Relevance

- Supports the use of behavioral signals like keystrokes and mouse dynamics for detecting stress-related states.
- Validates the design choice of excluding physiological sensors for privacy and simplicity in real-world deployments.
- Highlights the importance of user-specific baseline calibration, aligning with adaptive intent inference and feedback goals.

7 Robust Remote Detection of Depressive Tendency Based on Keystroke Dynamics and Behavioral Characteristics

Publication

Authors: Ruba Fadul, Aamna AlShehhi, Leontios J. Hadjileontiadis

Published In: Scientific Reports, Vol. 14, No. 1, 2024, Article No. 28025

DOI: 10.1038/s41598-024-78489-x

7.1 Abstract

This study introduces a remote, non-intrusive method for detecting depressive tendency (DT) using keystroke dynamics and behavioral characteristics. The approach processes flight time and hold time—timing-based metadata from keyboard interactions—to infer mental health status without accessing content. Using mutual information for feature selection and a gradient boosting classifier, the system achieved an impressive AUC of 0.98, indicating high diagnostic accuracy.

The research is grounded in real-world data collected from unsupervised typing behavior, validating its applicability in everyday settings. The study treats keystroke patterns as digital biomarkers, emphasizing the potential for scalable mental health assessment tools that respect privacy and require minimal user engagement. Although the focus is on depressive symptoms, the underlying methodology has broader relevance for real-time affective or intent classification systems in remote environments.

7.2 Advantages

- Achieves high accuracy (AUC = 0.98) using simple keystroke metadata, validating flight and hold time as meaningful behavioral indicators.
- Collected from in-the-wild usage, increasing real-world reliability for remote deployment scenarios.
- Applies mutual information for feature selection, enhancing both model interpretability and generalizability.

7.3 Limitations

- Focuses on depressive tendency, not directly on short-term states like distraction or frustration.
- Does not include other behavioral signals, such as mouse activity or multitasking patterns.
- Lacks integration with feedback systems, limiting its use as a dynamic or adaptive monitoring tool.

7.4 Relevance

- Confirms keystroke-based behavioral signals as effective for inferring internal states, supporting their use in intent-aware systems.
- The use of lightweight, interpretable ML models like gradient boosting aligns with efficient on-device processing goals.
- Emphasizes privacy-preserving data collection, reinforcing the local-only, metadata-based design of the proposed system.

8 Translating Keystroke and Mouse Dynamics Data to Classify Human Mood

Publication

Authors: B. Aldrich, Y. Liu, M. Almousa, M. Anwar

Published In: Proceedings of the 2023 Fifth International Conference on Transdisciplinary AI (TransAI), 2023, pp. 79–86

DOI: 10.1109/TransAI60598.2023.00023

8.1 Abstract

The study investigates unobtrusive mood detection through machine learning applied to keystroke and mouse interaction data. Over 16,000 labeled instances were collected, with participants reporting their mood using the Brief Mood Introspection Survey (BMIS). Models trained on keystroke and mouse dynamics were evaluated across six classification algorithms. Mouse-based models, particularly Random Forest and XGBoost, achieved superior performance. Mood classification accuracy was higher for "pleasant" states than "unpleasant." The results support the premise that mental states influence everyday computer interactions

8.2 Advantages

- Over 16,000 instances of keystroke and mouse interaction paired with mood labels, providing robust training and validation data.
- Describes a full software pipeline—from data capture (via custom application) to feature extraction to model evaluation—implemented in a real setting.
- Uses Random Forest and XGBoost—non-deep models known for interpretability and low computational cost; suitable for edge deployment.
- Shows mental state inference is feasible from interaction data alone, offering practical input signals for real-time intent detection systems.

8.3 Limitations

- Mood detection is limited to "pleasant" vs. "unpleasant," which may not encompass nuanced intent states such as frustration or distraction.
- Pleasant mood instances (9,330 mouse events vs. 3,848 unpleasant) may bias classifiers toward detecting positive mood more accurately.
- The system stops at mood classification and does not provide real-time feedback, recommendations, or smart nudging.
- Although interaction data is used, the paper does not discuss privacy protections, local processing, or anonymization strategies.

8.4 Relevance

- Demonstrates that keystroke and mouse dynamics can infer mental states reliably, supporting your project's intent detection component.

- Mouse interactions appear to be more predictive than keystroke patterns, informing your feature engineering and modality prioritization.
- Emphasizes a clear data-to-inference pipeline (capture → feature extraction → classification), aligning with your system's modular design.
- Provides a template for gathering labeled interaction data linked to subjective states, which you can adapt for your own intent categories and feedback-aware data collection.

9 Workplace Surveillance and Worker Well-Being

Publication

Authors: Paul Glavin, Alex Bierman, Scott Schieman

Published In: Social Currents, Vol. 11, No. 4, 2024, pp. 327–345

DOI: 10.1177/23294965241228874

9.1 Abstract

This study investigates how workplace surveillance impacts employee well-being using data from a nationally representative sample of 3,508 Canadian workers. The findings reveal that workers' perceptions of being monitored are significantly linked to higher psychological distress and lower job satisfaction. These effects operate indirectly through increased job pressure, reduced autonomy, and perceived privacy violations—mechanisms associated with the stress-process framework. While surveillance shows a small direct positive link to job satisfaction in some contexts, the broader trend indicates that the psychological costs outweigh any perceived benefits. Using structural equation modeling, the study illustrates how surveillance leads to stress proliferation, raising concerns about the long-term impacts of electronic monitoring on workforce morale and mental health. These results challenge the assumption that heightened surveillance leads to better outcomes and instead point toward the need for less intrusive, more trust-centered systems in the workplace.

9.2 Advantages

- Based on a large, representative sample (N = 3,508), supporting strong generalizability.
- Demonstrates clear links between surveillance, stress, and job dissatisfaction using validated psychological models.
- Uses robust statistical modeling (structural equation modeling) to uncover both direct and indirect effects.

9.3 Limitations

- The findings are associational, not derived from experimental or time-based causal analysis.
- Results are based on a single national context, which may limit generalization across work cultures and industries.
- The study does not differentiate between types of monitoring technologies, which limits applicability for evaluating specific systems.

9.4 Relevance

- Provides strong evidence that electronic surveillance increases workplace stress and reduces satisfaction, which supports the move toward non-intrusive monitoring systems.
- Reinforces the importance of maintaining user autonomy and minimizing privacy intrusions, aligning with the project's behavioral and intent-aware design principles.
- Offers a social science foundation to justify design choices that prioritize ethical, privacy-preserving alternatives to traditional surveillance.

10 A Comprehensive Survey on Privacy-Preserving Techniques in Federated Recommendation Systems

Publication

Authors: Muhammad Asad, Saima Shaukat, Ehsan Javanmardi, Jin Nakazato, Manabu Tsukada

Published In: Applied Sciences, Vol. 13, No. 10, 2023

DOI: 10.3390/app13106201

10.1 Abstract

This survey presents an in-depth analysis of privacy-preserving techniques within federated recommendation systems (FRSs), focusing on how collaborative model training can occur without direct access to user data. It reviews various system architectures and cryptographic methods such as local differential privacy (LDP), homomorphic encryption (HE), and secure multiparty computation (SMC). The paper categorizes approaches based on privacy models and system design, and identifies critical issues including privacy–utility trade-offs, communication overhead, scalability, and trustworthiness of the system.

The survey also discusses recent threats such as model inversion and gradient leakage, while emphasizing the need for standard evaluation benchmarks and noise calibration techniques. Though centered around recommender systems, the principles and methods reviewed offer broad relevance to any privacy-sensitive distributed learning system, particularly those involving human-centric behavioral data. The paper concludes by outlining research gaps and future directions that aim to enhance privacy without compromising system performance or user experience.

10.2 Advantages

- Summarizes a wide range of privacy-preserving techniques (e.g., LDP, HE, SMC) applicable to federated learning systems.
- Classifies architectures by privacy guarantees and performance, supporting systematic design decisions.
- Identifies core challenges like scalability, data leakage, and utility loss, providing design insights for real-world deployments.

10.3 Limitations

- Focuses on recommender systems, not directly on behavioral intent recognition.

- Lacks empirical testing or real-world benchmarks, relying mainly on theoretical comparisons.
- Does not address real-time or edge-based performance, which are critical for low-latency behavioral analytics systems.

10.4 Relevance

- Provides technical foundations for privacy-preserving federated architectures, directly informing the local-only ML design in behavioral monitoring.
- Discusses trade-offs between privacy and accuracy, relevant to deploying models under resource and trust constraints.
- Highlights open challenges like adaptive privacy tuning and communication efficiency, useful for extending the system with secure aggregation or calibration strategies.