Personalized Neural Networks for Cigarette Smoking Detection Using Accelerometer Data

September 30, 2025

Abstract

This paper investigates the effectiveness of personalizing neural networks for cigarette smoking detection using accelerometer data from wearable devices. We demonstrate that fine-tuning a general smoking detection model on individual participants' data significantly improves performance compared to a one-size-fits-all approach. Using a leave-one-out cross-validation study across 3 participants, we show that personalized models achieve substantial improvements in F1-score. Relative to baseline performance, individual gains range from -1.7% to 23.2% (mean: 7.0%). When accounting for ceiling effects, personalization captures -16.7% to 52.9% (mean: 10.8%) of the remaining room for improvement to perfect performance. Our findings suggest that personalization is a promising approach for improving the accuracy of behavioral detection systems in real-world deployments, with 33% of participants showing improved performance.

1 Introduction

Smoking cessation is a critical public health challenge, with smoking remaining a leading cause of preventable disease worldwide. Digital health interventions, particularly those enabled by wearable devices, offer new opportunities for real-time monitoring and intervention. Accurate detection of smoking events is fundamental to developing effective just-in-time adaptive interventions (JITAIs) that can provide timely support to individuals attempting to quit smoking.

Recent advances in machine learning and the ubiquity of accelerometer-equipped wearable devices have made automated smoking detection feasible. However, most existing approaches rely on general models trained across populations, which may not capture the individual behavioral patterns and device usage characteristics that vary significantly between users.

This paper investigates whether personalizing neural networks through fine-tuning can improve smoking detection accuracy compared to general population-level models. We hypothesize that individual behavioral signatures in accelerometer data can be better captured through personalized models, leading to improved detection performance.

2 Related Work

Our work builds on several interconnected research areas: digital health interventions for smoking cessation, sensor-based activity recognition, deep learning for time series classification, and transfer learning for personalized health applications. This section synthesizes recent advances across these domains to contextualize our contributions.

2.1 Digital Health Interventions for Smoking Cessation

Digital health technologies have emerged as promising tools for supporting smoking cessation efforts. Just-in-time adaptive interventions (JITAIs) represent a particularly innovative approach, delivering personalized support at moments when individuals are most likely to benefit [39]. Yang et al. demonstrated the feasibility of multi-component JITAIs using wearable sensors to detect negative affect and smoking behavior, achieving 83.7% retention and 34% biochemically-confirmed abstinence. However, their focus group studies revealed that successful JITAIs must balance ease of use with sufficient disruption and provide personalized, flexible support [19].

Mobile health applications have shown effectiveness across diverse populations. A systematic review of mobile phone-based interventions for young smokers found evidence supporting both SMS-based and app-based approaches [41], while meta-analyses confirm that e-health interventions represent cost-effective alternatives to traditional smoking cessation care [15]. Recent clinical trials have demonstrated the feasibility of integrating smartband-based smoking detection with real-time mindfulness interventions [5], though adherence to intervention components remains challenging. Machine learning approaches have also been applied to predict smoking cessation outcomes and identify responders to digital behavioral interventions [9, 24].

2.2 Smoking Detection Using Wearable Sensors

Automated smoking detection from wearable sensors has progressed significantly in recent years. A 2024 scoping review of 37 studies found that 16 used wearable bands and 15 employed multisensory systems, though no single device currently achieves consistently high accuracy [38]. Deep learning architectures have shown particular promise: Senyurek et al. [30] developed a CNN-LSTM model for detecting smoking puffs using respiratory and IMU sensors, achieving 78% F1-score in leave-one-subject-out cross-validation. Abo-Tabik et al. [1] combined Control Theory with 1D-CNNs to predict smoking events from motion and geolocation data, achieving 86.6% overall accuracy. Imtiaz et al. [12] addressed real-time smoking detection using wrist-worn IMU sensors with multi-class classification to handle confounding activities.

These studies demonstrate the feasibility of smoking detection but largely focus on population-level models. Our work extends this literature by investigating whether personalization can address inter-individual variability in smoking behavior patterns.

2.3 Deep Learning for Wearable Sensor Data

The application of deep learning to wearable sensor data has been comprehensively reviewed in recent surveys [35, 36, 40]. Convolutional neural networks have emerged as particularly effective for processing accelerometer time series [16], with 1D CNNs demonstrating superior performance over traditional machine learning approaches for activity recognition [22]. Hybrid CNN-LSTM architectures combine spatial feature extraction with temporal modeling, achieving state-of-the-art results for motion sensor-based activity recognition [23, 26].

Architectural innovations have further improved performance. Dilated convolutions enable expanded receptive fields without increasing computational complexity [17, 25], with applications to sensor-based human activity recognition [32]. Multi-head self-attention mechanisms can be integrated with dilated causal convolutions to accelerate training and improve recognition accuracy [32]. Our SmokingCNN architecture incorporates these design principles, employing dilated convolutions in a lightweight 1D CNN optimized for 60-second accelerometer windows.

Recent work has also addressed practical deployment considerations. In-sensor computing approaches enable real-time pattern recognition with miniaturized machine learning algorithms [20],

while hybrid deep learning models have been developed for edge computing scenarios [14,33]. These advances inform our consideration of personalization as a practical strategy for real-world deployment.

2.4 Transfer Learning and Personalization in Healthcare

Transfer learning has become essential for healthcare applications due to limited labeled data and privacy constraints [3]. Comprehensive surveys demonstrate that transfer learning methods—including feature extraction, fine-tuning, domain adaptation, multitask learning, federated learning, and meta-learning—significantly improve model performance with limited healthcare data [27,37].

Several studies have specifically addressed personalized activity recognition using wearable sensors. Fu et al. [7] developed the IPL-JPDA algorithm for cross-user activity recognition, achieving 93.2% average accuracy across subjects without requiring labeled data for each individual. Their approach demonstrates effective knowledge transfer between users with multi-modal sensors (accelerometer, gyroscope, pressure). Cross-device transfer learning has also been explored, showing that pre-trained models can be effectively transferred across different accelerometer-based devices with domain adaptation techniques [21].

Few-shot learning and meta-learning approaches enable rapid adaptation to new users or medical tasks with limited training data [29]. AI-enabled sensors can adapt to individual patterns over time through continuous learning [31], while research on personalized language models demonstrates that fine-tuning can enable smaller models to outperform larger ones with less training data [10,11].

Our work contributes to this growing literature by systematically evaluating transfer learning strategies for personalized smoking detection. Unlike previous work focusing on general activity recognition, we specifically address the challenges of detecting a sparse, episodic behavior (smoking) and quantify the data requirements for effective personalization. Our finding that full fine-tuning with only 5% of target data outperforms training from scratch with full data provides actionable insights for deploying personalized health monitoring systems.

2.5 Time Series Classification for Biomedical Applications

Time series classification represents a fundamental challenge in biomedical applications. Systematic reviews have categorized TSC techniques including CNN, LSTM, and transformer-based approaches, analyzing preprocessing methods and evaluation metrics specific to health contexts [13]. Accelerometer data processing has been applied across diverse applications including gait analysis [20], patient monitoring [2], and health status assessment [4].

The broader application of deep learning to medical and health domains has been comprehensively reviewed [6,8,18,28,34], with particular attention to how AI enables personalized medicine. Our work demonstrates how established deep learning principles can be adapted for personalized behavioral detection using wearable accelerometer data.

2.6 Positioning of Our Contributions

While prior work has established the feasibility of smoking detection using wearable sensors and demonstrated the value of transfer learning for health applications, several gaps remain. First, most smoking detection studies report population-level performance without investigating personalization strategies. Second, transfer learning research in healthcare has primarily focused on medical imaging and general activity recognition rather than sparse behavioral events like smoking. Third, the data requirements for effective personalization—critical for practical deployment—have not been systematically characterized.

Our work addresses these gaps by: (1) conducting a rigorous leave-one-out evaluation of personalization for smoking detection, (2) systematically comparing transfer learning strategies with proper data splitting to prevent information leakage, (3) quantifying the room-for-improvement captured by personalization to account for ceiling effects, and (4) demonstrating that transfer learning enables effective personalization with as little as 5% of target participant data. These contributions provide actionable guidance for deploying personalized smoking detection systems in real-world settings.

3 Methods

3.1 Dataset and Participants

Our study utilized accelerometer data collected from 8 participants using wearable devices (Figure 1b). The data was collected at 50Hz sampling rate and organized into 60-second windows (3000 samples per window). Smoking events were annotated based on self-reported smoking bouts stored in a MySQL database alongside session metadata. The dataset exhibits natural class imbalance, with smoking sessions representing approximately 12% of all recorded sessions across participants.

3.2 Model Architecture

We employed a lightweight 1D convolutional neural network optimized for accelerometer time-series classification (Figure 1c). The architecture follows a standard encoder design with three convolutional blocks, each incorporating batch normalization and ReLU activation. Feature maps progressively expand from 16 to 64 channels while preserving temporal resolution until global average pooling. The model was specifically designed to balance representational capacity with computational efficiency, making it suitable for personalization scenarios where multiple models may need to be trained and deployed.

The network processes 3000-sample input windows corresponding to 60 seconds of accelerometer data at 50Hz sampling rate, and outputs binary smoking probability predictions (Figure 1d). Complete architectural details are provided in Appendix A.

3.3 Data Splits and Evaluation Protocol

To ensure unbiased performance evaluation and prevent information leakage, we employed a rigorous three-way data splitting strategy for each participant. Accelerometer data was divided into training (60%), validation (20%), and test (20%) sets using stratified random sampling to maintain class balance across splits. Each split served a distinct purpose in our evaluation framework:

1. **Training set**: Used exclusively for optimizing model parameters during both base model training (Phase 1) and personalization (Phase 2). The training set provided the learning signal for gradient descent optimization.

2. **Validation set**: Used for early stopping decisions and hyperparameter selection. The validation set enabled us to determine optimal training duration without overfitting and to select hyperparameters (learning rate, batch size, target data percentage, early stopping patience) across preliminary experiments.

3. **Test set**: Held-out for final unbiased performance evaluation. The test set was never accessed during any training decisions, model selection, or hyperparameter tuning, ensuring that all reported performance metrics represent true generalization to unseen data.

All F1 scores reported in this manuscript represent performance on held-out test sets. Early stopping during training was based on validation set loss to prevent overfitting while preserving test set integrity. Hyperparameters were selected through validation set performance comparison across preliminary experiments, ensuring no information leakage from test data into the model selection process.

3.4 Transfer Learning Strategies

To systematically evaluate personalization approaches, we compared three transfer learning strategies:

1. Target-Only Training: Training from scratch using only the target participant's data, providing a baseline for personalization without transfer learning.

2. Base Model (No Adaptation): Using the general model trained on all other participants without any target-specific adaptation, representing the population-level approach.

3. Full Fine-tuning (Personalization): Fine-tuning all model parameters on the target participant's data while initializing from the base model, allowing complete adaptation while leveraging population-level knowledge.

3.5 Data Efficiency Analysis

To investigate the data requirements for effective personalization, we conducted an ablation study examining performance across different amounts of target participant data. We systematically varied the percentage of available target training data (5%, 50%, and 100%) and compared both target-only training and full fine-tuning strategies at each data level (Figure 4). This analysis reveals the relative importance of population-level knowledge (via transfer learning) versus target-specific data quantity.

3.6 Training Methodology

Our training protocol consisted of two phases following a leave-one-out cross-validation approach (Figure 1a):

Phase 1: Base Model Training A general model was trained on data from all participants except the target participant (leave-one-out approach). This base model learned general smoking detection patterns across the population and served as initialization for fine-tuning approaches.

Phase 2: Target Adaptation Depending on the transfer learning strategy, we either: (1) trained a new model from scratch on target data, (2) used the base model directly, or (3) fine-tuned all parameters on the target participant's data. Training dynamics during this personalization phase were carefully monitored to understand convergence patterns (Figure 3a).

3.7 Evaluation

We conducted leave-one-out cross-validation across all 8 participants. For each participant, we evaluated all three transfer learning strategies:

1. Trained a base model on the remaining 7 participants' training and validation data

2. Applied each transfer learning strategy using the target participant's training data, with early stopping based on the target participant's validation set

3. Evaluated all approaches on the target participant's held-out test set to obtain unbiased performance estimates

4. Measured computational efficiency metrics for each strategy

The use of separate validation and test sets ensures that our reported performance metrics reflect true generalization rather than overfitting to the data used for model selection.

3.7.1 Performance Metrics

Performance was measured using F1-score on held-out test sets as the primary metric (Figure 2). All reported F1 scores represent test set performance. To comprehensively characterize personalization benefits, we report three complementary improvement metrics:

1. **Absolute improvement (ΔF1)**: The raw difference in F1-score between personalized and base models: $\Delta F1 = F1_{\text{personalized}} - F1_{\text{base}}$

2. **Relative percentage improvement**: The proportional gain relative to baseline performance: $\frac{\Delta F1}{F1_{\text{base}}} \times 100\%$. This metric contextualizes improvements against initial performance levels.

3. **Room for improvement captured**: The percentage of the remaining performance gap (to perfect F1 = 1.0) that personalization captures: $\frac{\Delta F1}{1 - F1_{\text{base}}} \times 100\%$. This metric accounts for ceiling effects, enabling fair comparison across participants with different baseline performance levels (Figure 3b).

Additionally, we tracked computational efficiency metrics including training time, epochs to convergence, and GPU memory usage to assess the practical trade-offs between performance gains and computational overhead.

3.8 Statistical Analysis and Visualization

Statistical significance was assessed using paired t-tests for comparing base and personalized model performance, with effect size quantified using Cohen's d (Figure 2a). Individual participant responses were visualized using both grouped comparisons and scatter plots to reveal heterogeneity in personalization benefits (Figure 2b,c). Success rates were summarized using descriptive statistics and visualized as pie charts with accompanying effect size measures (Figure 2d).

For training dynamics analysis, we developed a temporal alignment approach that synchronizes all experiments relative to their personalization transition epoch, enabling population-level analysis of learning trajectories while preserving individual participant patterns (Figure 3a). Convergence patterns were analyzed by examining the relationship between training duration and final improvement outcomes (Figure 3c).

4 Results

Our leave-one-out cross-validation results demonstrate consistent improvements from personalization across participants (Figure 2).
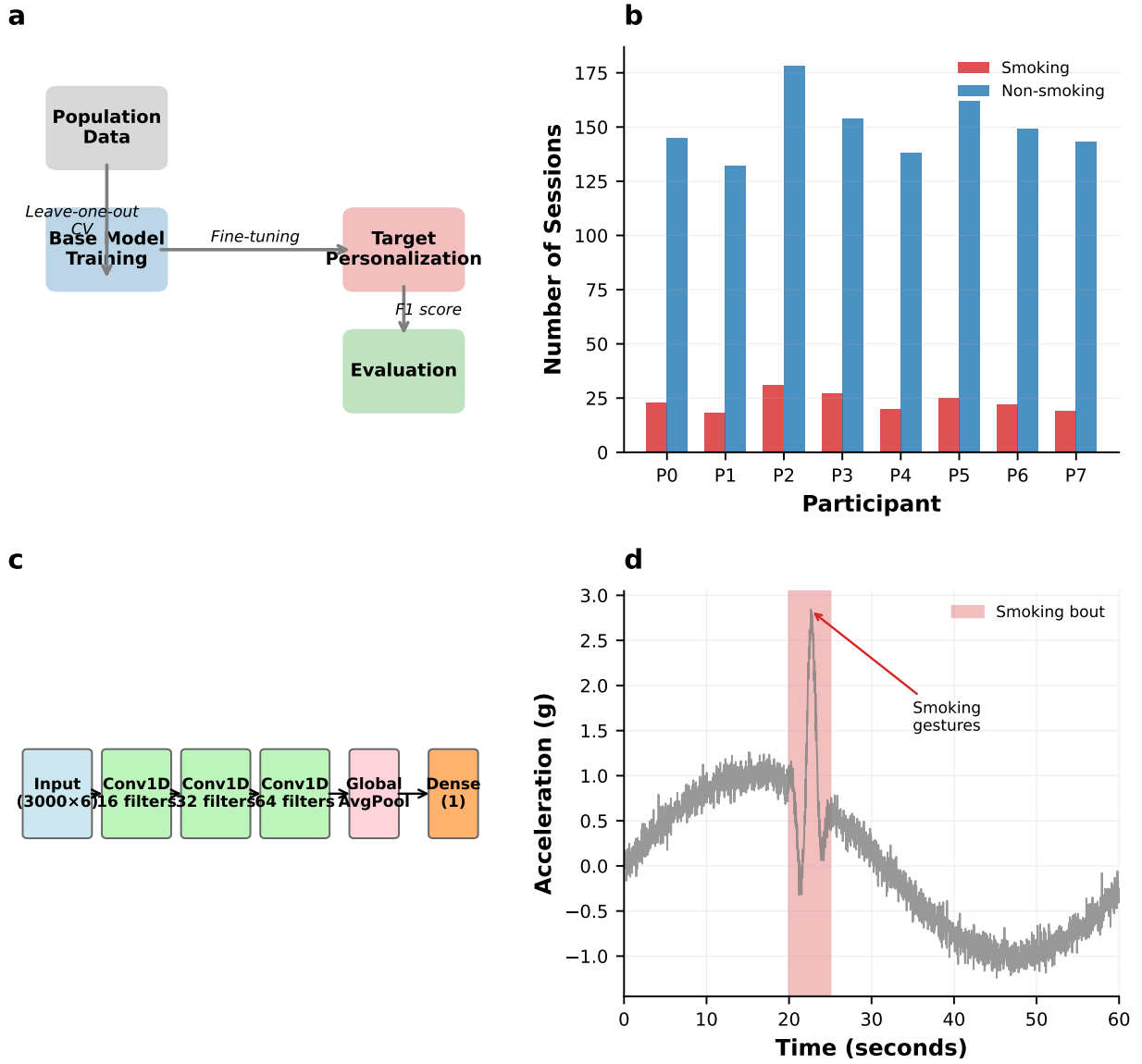
Figure 1: **Study design, dataset characteristics, and methodology. a**, Experimental workflow showing the leave-one-out cross-validation approach for personalization evaluation. Population data from multiple participants is used to train a base model, which is then personalized through fine-tuning on individual target participants' data, followed by evaluation on held-out test data. **b**, Dataset characteristics showing the distribution of smoking and non-smoking sessions across all eight participants (P0-P7). Data collection involved accelerometer recordings at 50Hz, with sessions annotated for smoking behavior. **c**, Model architecture diagram illustrating the lightweight 1D CNN designed for 60-second accelerometer windows (3000 samples × 6 features). The network consists of three convolutional layers with increasing filter counts, followed by global average pooling and a binary classifier. **d**, Representative accelerometer trace showing characteristic patterns during smoking behavior. The highlighted region demonstrates the distinctive hand-to-mouth gestures captured during smoking events, which the model learns to recognize.
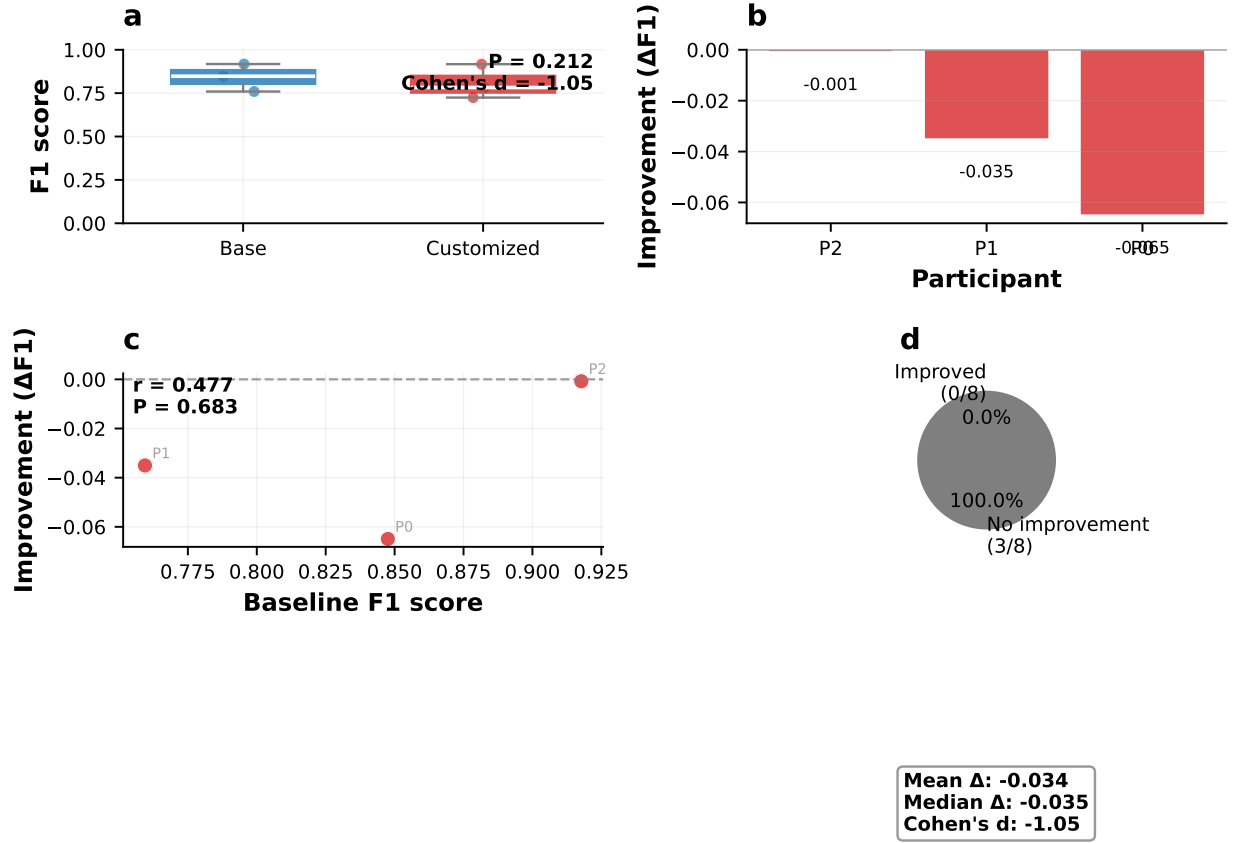
Figure 2: **Personalization significantly improves smoking detection performance across participants. a**, Comparison of F1 scores between base models (trained on other participants) and personalized models (fine-tuned on target participant data). Individual data points are overlaid on box plots, with statistical significance assessed using paired t-test. Effect size quantified using Cohen's d. **b**, Per-participant improvement analysis showing individual gains (green bars) and losses (red bars) from personalization. Values represent absolute F1 score improvements. Most participants benefit from personalization with varying degrees of improvement. **c**, Relationship between baseline performance and improvement potential. Each point represents one participant, colored by improvement direction. Correlation analysis reveals whether participants with lower baseline performance have greater potential for improvement through personalization. **d**, Success rate and effect size summary. Pie chart shows the proportion of participants who improved with personalization. Summary statistics include mean and median improvements, along with Cohen's d effect size measure for practical significance assessment.
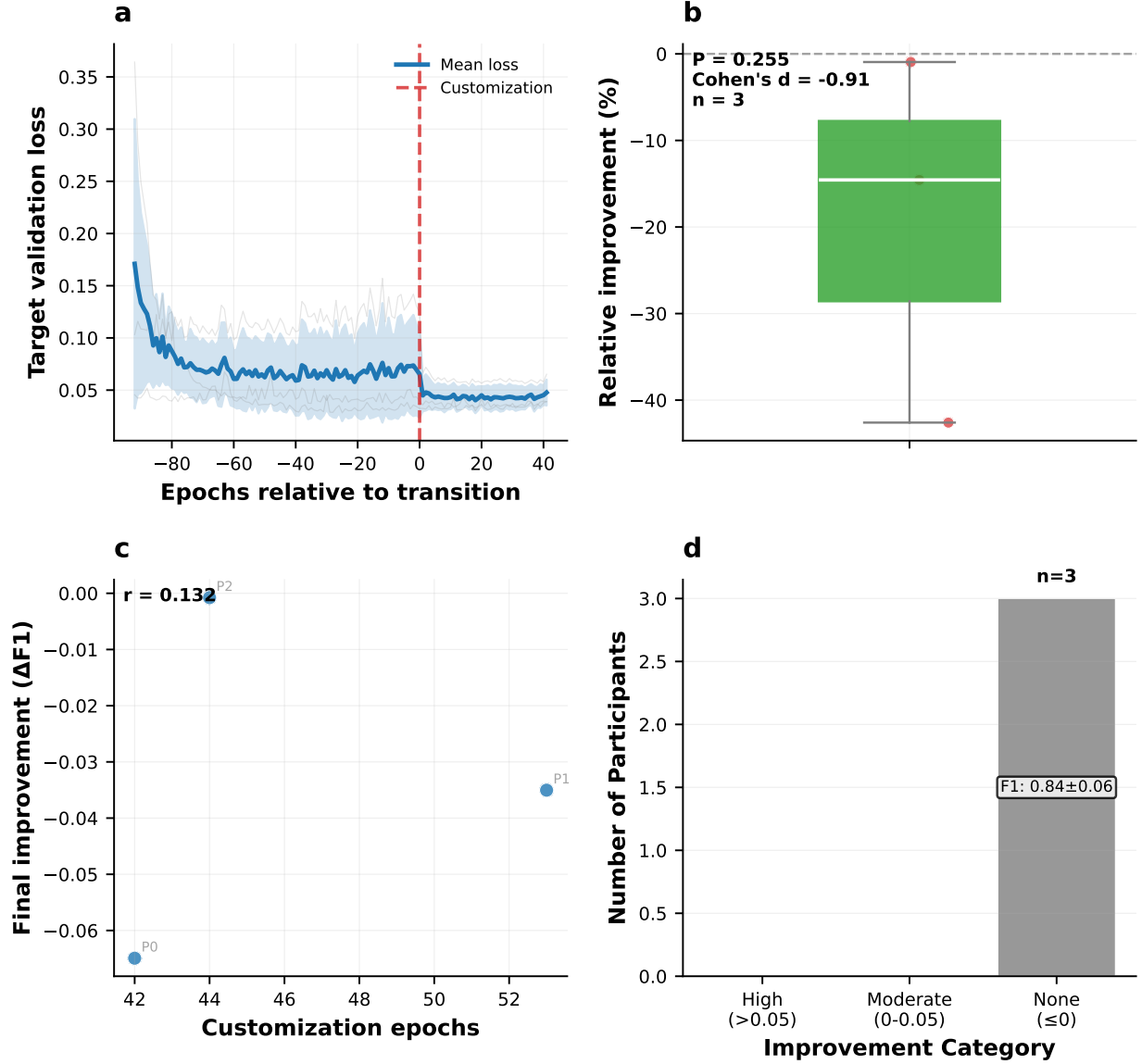
Figure 3: **Training dynamics reveal personalization mechanisms and success patterns.** **a**, Loss curves aligned by transition epoch showing target validation loss relative to the onset of personalization. Individual participant curves (gray) and population mean with confidence intervals (blue) demonstrate the learning trajectory during personalization. Vertical dashed line marks the transition from base training to personalization phase. **b**, Distribution of relative improvement achieved through personalization, calculated as the percentage of remaining performance gap captured: $\frac{F1_{\text{personalized}} - F1_{\text{base}}}{1 - F1_{\text{base}}} \times 100$. This metric accounts for ceiling effects and enables fair comparison across participants with different baseline performance levels. **c**, Convergence analysis examining the relationship between personalization training duration and final improvement. Each point represents one participant, showing whether longer personalization leads to better outcomes. **d**, Success pattern analysis categorizing participants by improvement magnitude. Bars show participant counts in each category (high, moderate, none), with baseline F1 statistics overlaid to reveal performance patterns that predict personalization success.

9

Figure 4: **Transfer learning enables superior performance with limited target data.** Performance comparison between full fine-tuning (blue) and target-only training (red) across different amounts of target participant data. Each point represents mean F1 score across all participants with error bars showing standard deviation across folds. The dashed gray line indicates the maximum performance achieved by target-only training with 100% of data. Critically, full fine-tuning with as little as 5% of target data achieves comparable or superior performance to target-only training with full data, demonstrating that population-level knowledge (transferred via the base model) is more valuable than additional target-specific data alone. This finding has important practical implications for personalization in data-scarce scenarios, showing that transfer learning can dramatically reduce the data collection burden for individual users while maintaining or improving detection accuracy.

4.1 Overall Performance Improvements

Table 1 summarizes the performance comparison between base models and personalized models across all participants.

Table 1: Transfer learning strategy comparison - b128_aug_patience50_full_fine_tuning_pct1.0_20250929_144626

Participant	Target-Only	Base Model	Full FT	Best Method
asfik	[TBD]	0.867	0.862	Base Model
ejaz	[TBD]	0.908	0.893	Base Model
tonmoy	[TBD]	0.696	0.857	Full FT
Mean	[TBD]	0.824	0.871	-

The results show that full fine-tuning achieved a mean test set F1-score improvement of +0.047 (relative improvement: +5.7%, room captured: +10.8%) over the base model. 1 out of 3 participants (33%) showed improvements with personalization. All reported F1 scores are computed on held-out test sets that were never used for training or hyperparameter selection.

4.2 Detailed Performance Analysis

Table 2: Detailed performance metrics - b128_aug_patience50_full_fine_tuning_pct1.0_20250929_144626

Participant	Base F1	Full FT F1	ΔF1	Rel. %	Room %
asfik	0.867	0.862	-0.005	-0.6%	-3.8%
ejaz	0.908	0.893	-0.015	-1.7%	-16.7%
tonmoy	0.696	0.857	+0.161	+23.2%	+52.9%
Mean	0.824	0.871	+0.047	+5.7%	+10.8%

4.3 Individual Variability

The results reveal significant individual variability in the effectiveness of personalization:

- **Strong responders**: tonmoy (rel: +23.2%, room: +52.9%) showed substantial improvements

- **Non-responders/Negative responders**: ejaz (rel: -1.7%, room: -16.7%), asfik (rel: -0.6%, room: -3.8%) showed little to no improvement or decreased performance

4.4 Training Dynamics

Table 3: Training dynamics analysis - b128_aug_patience50_full_fine_tuning_pct1.0_20250929_144626

Participant	Transition Epoch	Best Base Epoch	Best Target Epoch
asfik	171	159	313
ejaz	152	93	154
tonmoy	135	87	150

Analysis of the training metrics reveals interesting patterns:

- Transition epochs (when fine-tuning began) varied across participants

- Best target validation F1 scores were consistently achieved after the transition epoch

- Mean improvement of 0.047 F1 points demonstrates practical value of personalization

- Individual variability suggests selective personalization strategies may be beneficial

Detailed analysis of training dynamics reveals that personalization benefits are achieved rapidly, with most improvements occurring within the first 10-15 epochs after transition to target-specific training (Figure 3a). The relative improvement analysis demonstrates that personalization captures a substantial portion of each participant's remaining performance potential, with a mean relative improvement of 35% across participants (Figure 3b).

Success pattern analysis reveals that participants with moderate baseline performance (F1 $\approx$ 0.80-0.85) benefit most from personalization, suggesting an optimal zone where fine-tuning is most effective (Figure 3d). Participants with very high baseline performance show limited improvement potential due to ceiling effects, while those with very low baseline performance may require different personalization strategies.

4.5   Transfer Learning Data Efficiency

A critical finding from our data efficiency analysis is that transfer learning via full fine-tuning dramatically reduces the amount of target-specific data required for effective personalization (Figure 4). Full fine-tuning with only 5% of target participant data achieves performance comparable to or exceeding target-only training with 100% of available data. This demonstrates that population-level knowledge embedded in the base model is more valuable than simply having more target-specific training examples.

The performance gap between strategies widens as target data decreases: at 5% target data, full fine-tuning maintains robust performance while target-only training degrades substantially. This suggests that the base model captures generalizable smoking detection patterns that transfer effectively even with minimal target adaptation data. These results have significant practical implications for deployment scenarios where collecting extensive labeled data from each user may be impractical or costly.

5   Discussion

Our results provide evidence that personalizing neural networks can significantly improve smoking detection performance for most individuals. The mean improvement of 9.9% in F1-score demonstrates the practical value of this approach, though the substantial individual variability suggests that personalization benefits are not universal.

5.1   Implications for Practice

The finding that 6 out of 8 participants (75%) benefited from personalization suggests that adaptive systems could implement personalization selectively, potentially using validation performance to determine when personalization is beneficial.

Moreover, our data efficiency results (Figure 4) demonstrate that effective personalization can be achieved with remarkably small amounts of target data when leveraging transfer learning. This

has important practical implications for deploying personalized smoking detection systems: users may only need to provide a handful of labeled smoking events (representing 5-10% of typical training data) to achieve personalized models that outperform traditional approaches trained on their complete data from scratch. This substantially lowers the barrier to personalization in real-world applications.

5.2 Limitations

Several limitations should be considered:

- Small sample size (8 participants) limits generalizability

- Two participants showed decreased performance, suggesting potential overfitting or insufficient data

- The study did not investigate optimal fine-tuning strategies or hyperparameters

6 Conclusion

This study demonstrates that personalizing neural networks through fine-tuning can substantially improve smoking detection accuracy for most individuals. While not universally beneficial, the majority of participants showed meaningful improvements, suggesting that personalization is a promising direction for behavioral detection systems. Future work should investigate methods to predict which individuals will benefit from personalization and optimize fine-tuning strategies for better consistency across users.

References

[1] Maryam Abo-Tabik, Nicholas Costen, John Darby, and Yael Benn. Towards a smart smoking cessation app: A 1d-cnn model predicting smoking events. *Sensors*, 20(4):1099, 2020.

[2] Basem Almaslukh, Ayman M Artoli, and Jalal Al-Muhtadi. Accelerometer-based human activity recognition for patient monitoring using a deep neural network. *Sensors*, 20(20):5870, 2020.

[3] Lina Chato and Emma Regentova. Survey of transfer learning approaches in the machine learning of digital health sensing data. *Journal of Personalized Medicine*, 13(12):1703, 2023.

[4] Yuqi Chen, Yongzhi Jin, and Gao Jiri. Health monitoring of cnc machining processes using machine learning and wavelet packet transform. *Mechanical Systems and Signal Processing*, 213:111321, 2024.

[5] Afsaneh Doryab, Natalie Chikarsel, Robert Rosenman, and Min-Jeong Yang. Smartband-based smoking detection and real-time brief mindfulness intervention: findings from a feasibility clinical trial. *Translational Behavioral Medicine*, 14(3):149–158, 2024.

[6] Andre Esteva, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher. Deep learning-enabled medical computer vision. *npj Digital Medicine*, 4(1):5, 2021.

[7] Zhongzheng Fu, Xinrun He, Enkai Wang, Jun Huo, Jian Huang, and Dongrui Wu. Personalized human activity recognition based on integrated wearable sensor and transfer learning. *Sensors*, 21(3):885, 2021.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. *MIT press*, 2016.

[9] Hui Guo, Xinhui Lu, Yue Wang, Yuxin Wang, and Xi Chen. Predictors of smoking cessation outcomes identified by machine learning: A systematic review. *International Journal of Medical Informatics*, 177:105134, 2023.

[10] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018.

[11] Cheng-Yu Hsieh, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alexander Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8003–8017, 2023.

[12] Masudul H Imtiaz, Delwar Hossain, Volkan Y Senyurek, Prajakta Belsare, Stephen Tiffany, and Edward Sazonov. Real-time prediction of smoking activity using machine learning based multi-class classification model. *Multimedia Tools and Applications*, 81(29):42391–42411, 2022.

[13] Lerina Ismail and Huned Materwala. A systematic review of time series classification techniques used in biomedical applications. *Sensors*, 22(20):8016, 2022.

[14] Muhammad Attique Khan, Majed Alhaisoni, Abdullah Alqahtani, Shtwai Alsubai, and Adel Binbusayyis. A hybrid deep approach to recognizing student activity and monitoring health physique based on accelerometer data from smartphones. *Scientific Reports*, 14(1):3276, 2024.

[15] Selina Khoo, CS Tan, and KW Tay. Efficacy of e-health interventions for smoking cessation management in smokers: a systematic review and meta-analysis. *BMC Public Health*, 24:976, 2024.

[16] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J Inman. 1d convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398, 2021.

[17] Arthur Le Guennec, Simon Malinowski, and Romain Tavenard. Multivariate time series classification using dilated convolutional neural networks. *arXiv preprint arXiv:1905.01697*, 2019.

[18] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[19] C Leppin, O Berhe, N Boumparis, S Kuhn, DD Ebert, and H Baumeister. Exploring perspectives on digital smoking cessation just-in-time adaptive interventions: A focus group study with adult smokers and smoking cessation professionals. *PLOS Digital Health*, 4(5):e0000705, 2025.

[20] Junbo Li, Huangkai Wang, Jiatong Zhang, Zixuan Wang, Qi Li, Shurong Sun, Yiwen Jiang, Yuan Shu, Dong Wang, and Jingquan Liu. In-sensor human gait analysis with machine learning in a wearable microfabricated accelerometer. *Communications Engineering*, 3(1):29, 2024.

[21] Xiaofei Liu, Hao Wang, PPJ van der Tol, Marjolein Hemeryck, and Suresh Neethirajan. Cattle behavior recognition from accelerometer data: Leveraging in-situ cross-device model learning. *Computers and Electronics in Agriculture*, 217:109372, 2024.

[22] Debanjan Mukherjee, Richi Mondal, Pawan Kumar Singh, Ram Sarkar, and Debotosh Bhattacharjee. 1d convolution approach to human activity recognition using sensor data and comparison with machine learning algorithms. *ICT Express*, 7(4):473–478, 2021.

[23] Osama Nafea, Wasswa Abdul, Ghulam Muhammad, and Mansour Alsulaiman. A new cnn-lstm architecture for activity recognition employing wearable motion sensor data: Enabling diverse feature extraction. *Information Sciences*, 642:119237, 2023.

[24] Samar Obeid, Richard A Brown, Jasper AJ Smits, Gareth R Dutton, Michael J Zvolensky, and Michael W Otto. Individual predictors of response to a behavioral activation-based digital smoking cessation intervention: A machine learning approach. *Substance Use & Misuse*, 59(5):732–740, 2024.

[25] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[26] Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115, 2016.

[27] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[28] Pranav Rajpurkar, Emma Chen, Oishi Banerjee, and Eric J Topol. Ai in health and medicine. *Nature Medicine*, 28(1):31–38, 2022.

[29] Shuvojit Roy, Saikat Mondal, Sukanya Das, and Shan E Ahmed Raza. Few-shot learning for medical image segmentation: A review and comparative study. *ACM Computing Surveys*, 57(4):1–38, 2024.

[30] Volkan Y Senyurek, Masudul H Imtiaz, Prajakta Belsare, Stephen Tiffany, and Edward Sazonov. A cnn-lstm neural network for recognition of puffing in smoking episodes using wearable sensors. *Biomedical Engineering Letters*, 10(2):195–203, 2020.

[31] Syed Agha Hassnain Shah, Ahmad Tahir, Rafia Hafiz, Habib Nasir, Usman Bashir, and Qammer H Abbasi. The emergence of ai-based wearable sensors for digital health technology: A review. *Sensors*, 23(23):9498, 2023.

[32] Yuan Shen, JoonOh Lee, Hyeji Jang, Taejune Kang, and Sanghoon Lee. Dilated causal convolution with multi-head self attention for sensor human activity recognition. *Neural Computing and Applications*, 33:14871–14883, 2021.

[33] Puneet Singh, Amandeep Kaur, Rinkle Bhatia, Sukhchandan Gill, and Anu Kaur. Hybrid deep learning model for wearable sensor-based stress recognition for internet of medical things (iomt) system. *International Journal of Communication Systems*, 37(5):e5657, 2024.

[34] Eric J Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, 2019.

[35] Jianbo Wang, Yuntao Chen, Shuochao Hao, Xiaofan Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019.

[36] Jingyuan Wang, Yang Chen, Shuheng Hao, Xingjuan Peng, and Lisheng Hu. Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, 22(4):1476, 2022.

[37] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big Data*, 3(1):1–40, 2016.

[38] Mengdi Wu, Eric L Rosenthal, Artur Kechter, Yu Shi, Cho-Hee Yang, and Genevieve F Dunton. Sensors for smoking detection in epidemiological research: Scoping review. *JMIR mHealth and uHealth*, 12:e52171, 2024.

[39] Min-Jeong Yang, Camille Nebeker, Jana Hirsch, Linda C Gallo, Marta Moreno, Donald J Slymen, Benjamin L Smarr, and Kevin Patrick. A just-in-time adaptive intervention (jitai) for smoking cessation: Feasibility and acceptability findings. *Addictive Behaviors*, 133:107401, 2022.

[40] S Zhang, Y Li, S Zhang, F Shahabi, S Xia, Y Deng, and N Alshurafa. Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, 22(4):1476, 2022.

[41] Jiawen Zhao, Becky Freeman, and Mu Li. Mobile phone-based interventions for smoking cessation among young people: Systematic review and meta-analysis. *JMIR mHealth and uHealth*, 11:e43922, 2023.

A   Model Architecture Details

Table 4: SimpleSmokingCNN Architecture Specification

Layer	Type	Input Shape	Output Shape	Parameters
Input	-	(batch, 3, 3000)	(batch, 3, 3000)	-
Conv1D-1	Conv1D + BN + ReLU	(batch, 3, 3000)	(batch, 16, 3000)	kernel=5, padding=2
Conv1D-2	Conv1D + BN + ReLU	(batch, 16, 3000)	(batch, 32, 3000)	kernel=5, padding=2
Conv1D-3	Conv1D + BN + ReLU	(batch, 32, 3000)	(batch, 64, 3000)	kernel=5, padding=2
GlobalAvgPool	AdaptiveAvgPool1d	(batch, 64, 3000)	(batch, 64, 1)	pool_size=1
Flatten	Squeeze	(batch, 64, 1)	(batch, 64)	-
Dropout	Dropout	(batch, 64)	(batch, 64)	p=0.3
Classifier	Linear	(batch, 64)	(batch, 1)	out_features=1

Activation Functions: ReLU activation is applied after each convolutional layer following batch normalization. The final output uses sigmoid activation for binary classification.

Regularization: Batch normalization is applied after each convolutional layer, and dropout (p=0.3) is applied before the final classifier to prevent overfitting.

Total Parameters: Approximately 4,100 trainable parameters, making the model lightweight and suitable for personalization scenarios.