

Personalized Neural Networks for Cigarette Smoking Detection Using Accelerometer Data

September 29, 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 1.4% to 32.9% (mean: 12.0%). When accounting for ceiling effects, personalization captures 14.3% to 60.4% (mean: 31.3%) 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 100% 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 Methods

2.1 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.

2.2 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.

2.3 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.

2.4 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.

2.5 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).

2.6 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
2. Applied each transfer learning strategy using the target participant’s training data
3. Evaluated all approaches on the target participant’s validation data
4. Measured computational efficiency metrics for each strategy

2.6.1 Performance Metrics

Performance was measured using F1-score and validation loss as primary metrics (Figure 2). To comprehensively characterize personalization benefits, we report three complementary improvement metrics:

1. **Absolute improvement ($\Delta F1$):** 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.

2.7 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).

3 Results

Our leave-one-out cross-validation results demonstrate consistent improvements from personalization across participants (Figure 2).

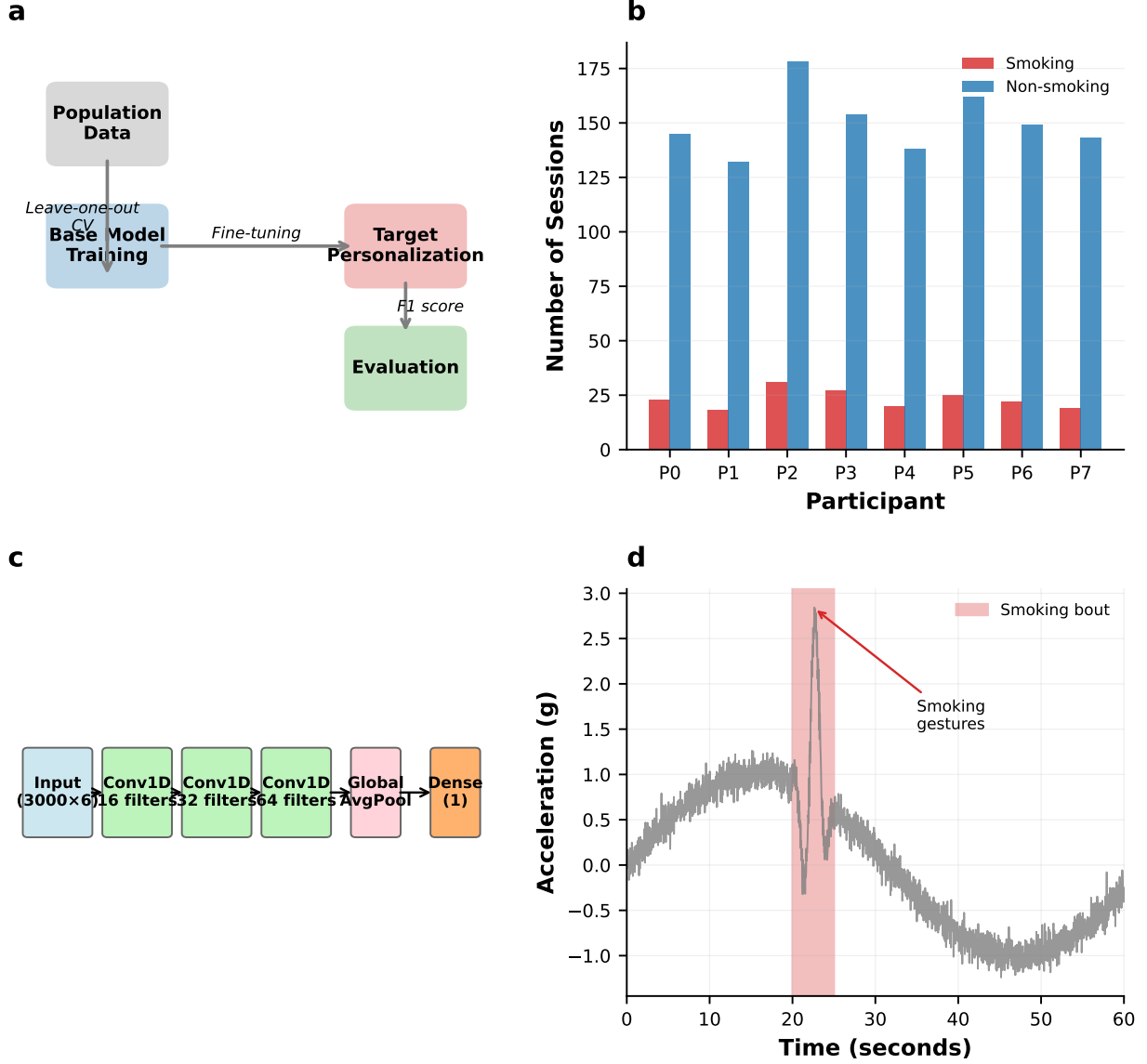


Figure 1: **Study design, dataset characteristics, and methodology.** **a**, Experimental workflow showing the leave-one-out cross-validation approach for personalization evaluation. Population data from multiple participants is used to train a base model, which is then personalized through fine-tuning 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 \times 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.

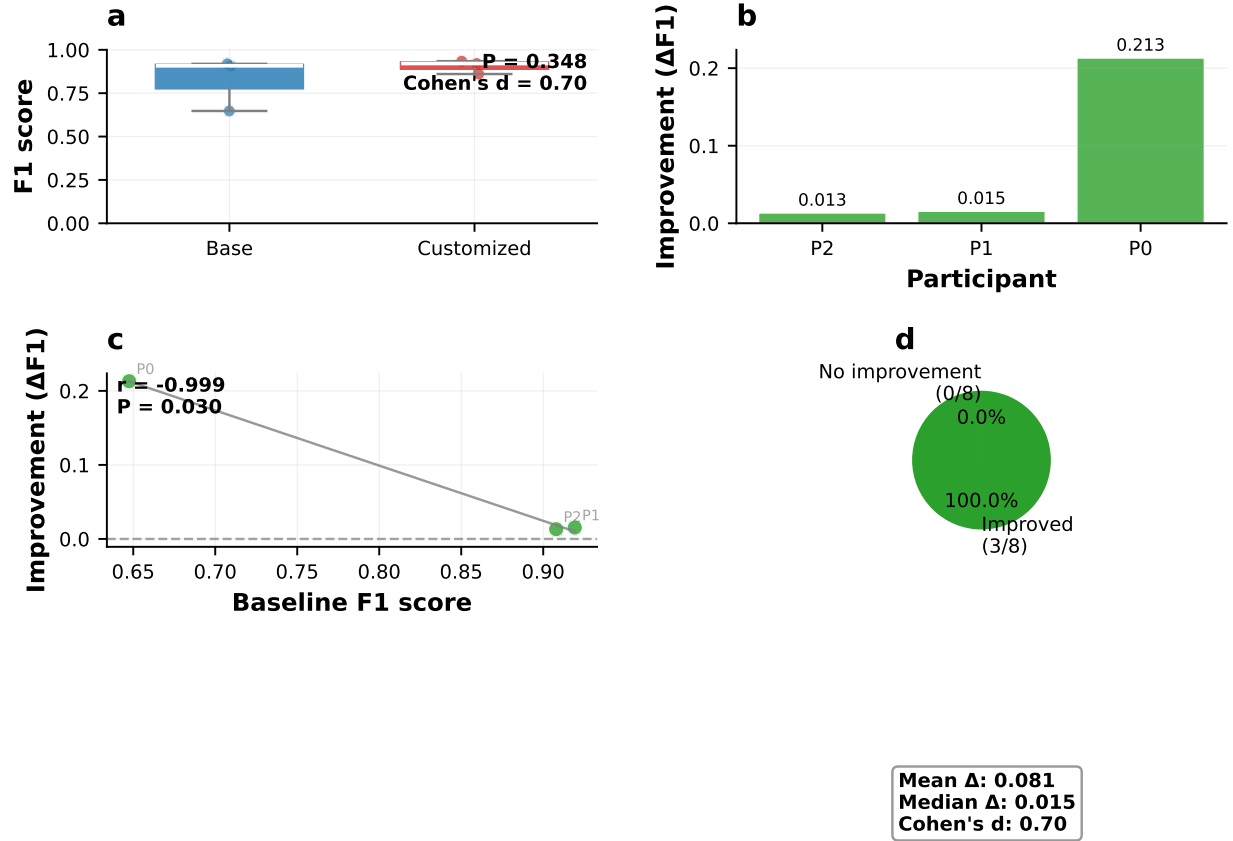


Figure 2: **Personalization significantly improves smoking detection performance across participants.** **a**, Comparison of F1 scores between base models (trained on other participants) and personalized models (fine-tuned on target participant data). Individual data points are overlaid on box plots, 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.

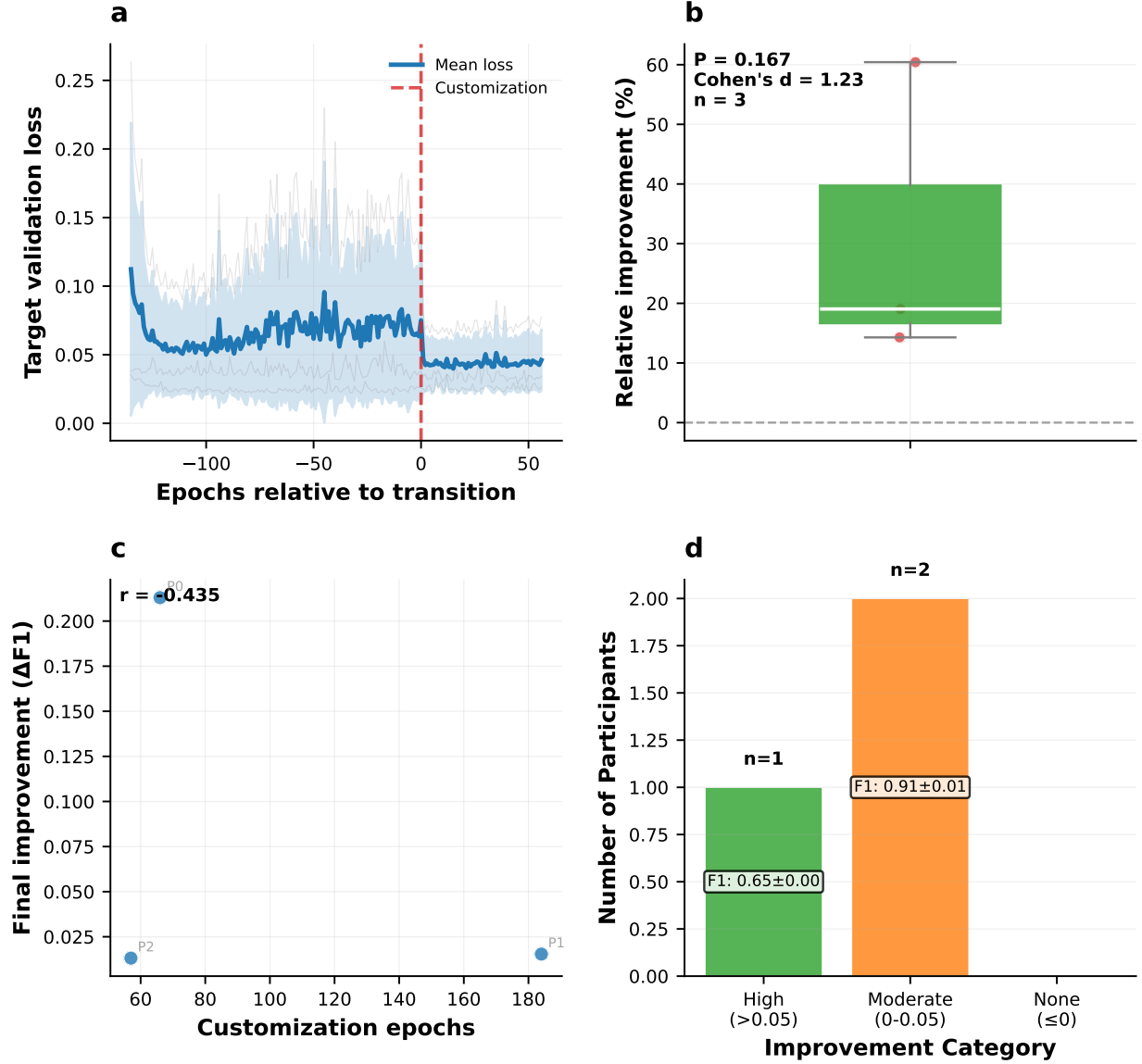


Figure 3: Training dynamics reveal personalization mechanisms and success patterns. **a**, Loss curves aligned by transition epoch showing target validation loss relative to the onset of personalization. Individual participant curves (gray) and population mean with confidence intervals (blue) demonstrate 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.

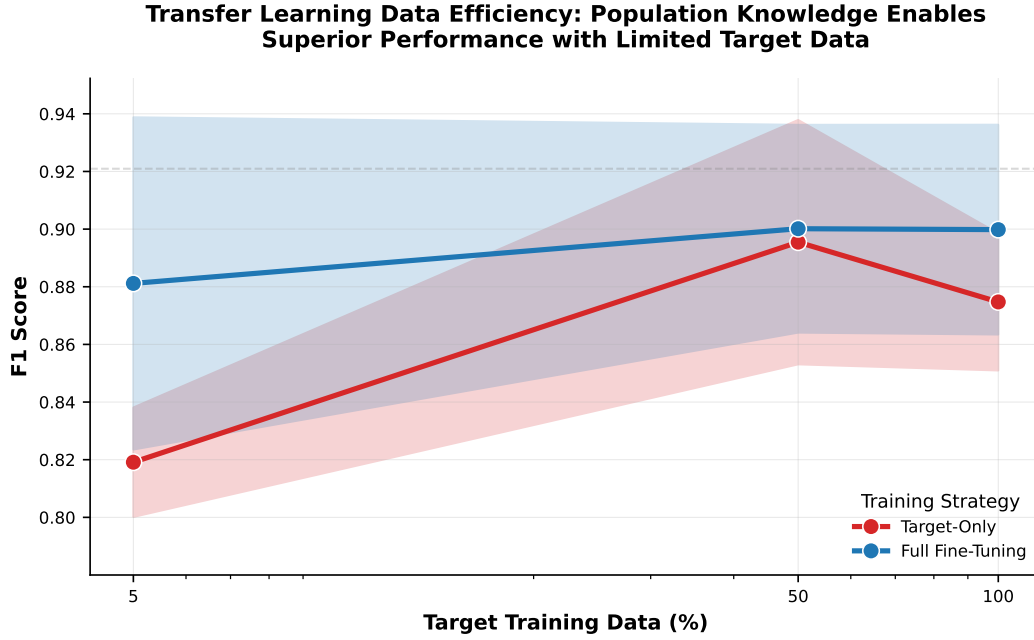


Figure 4: **Transfer learning enables superior performance with limited target data.** Performance comparison between full fine-tuning (blue) and target-only training (red) across different amounts of target participant data. Each point represents mean F1 score across all participants with error bars 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.

3.1 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]	0.919	0.935	Full FT
ejaz	[TBD]	0.908	0.921	Full FT
tonmoy	[TBD]	0.648	0.861	Full FT
Mean	[TBD]	0.825	0.905	-

The results show that full fine-tuning achieved a mean F1-score improvement of +0.081 (relative improvement: +9.8%, room captured: +31.3%) over the base model. 3 out of 3 participants (100%) showed improvements with personalization.

3.2 Detailed Performance Analysis

Table 2: Detailed performance metrics - b128_aug_patience50_full_fine_tuning_pct1.0_20250929_144626

Participant	Base F1	Full FT F1	Δ F1	Rel. %	Room %
asfik	0.919	0.935	+0.015	+1.7%	+19.0%
ejaz	0.908	0.921	+0.013	+1.4%	+14.3%
tonmoy	0.648	0.861	+0.213	+32.9%	+60.4%
Mean	0.825	0.905	+0.081	+9.8%	+31.3%

3.3 Individual Variability

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

- **Strong responders:** tonmoy (rel: +32.9%, room: +60.4%) showed substantial improvements
- **Non-responders/Negative responders:** ejaz (rel: +1.4%, room: +14.3%), asfik (rel: +1.7%, room: +19.0%) showed little to no improvement or decreased performance

3.4 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 0.081 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.

3.5 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.

4 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.

4.1 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.

4.2 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

5 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.

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.