ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Lightweight Personalisation for MEMS-Based Wearables: A Padel Stroke Recognition Case Study

Alberto Gascon[1], Fatemeh Akbarian[2], Amir Aminifar[2], Alvaro Marco[1], Roberto Casas[1] *

[1]Aragon Institute of Engineering Research, University of Zaragoza, Spain

[2]Department of Electrical and Information Technology, Lund University, Sweden

Abstract.

This work investigates lightweight personalisation of micro-electro-mechanical systems (MEMS)-based wearables using a public padel database as a case study. We compare a centralised CNN model, single-user models and two fine-tuning schemes (full and last-layer) on wrist-worn IMU data from 23 players and 13 stroke classes. Personalised models with data augmentation achieve weighted F1-scores above 90%, closing most of the gap to an optimistic single-user upper bound while reducing inter-subject variability. FLOP and memory analyses show that last-layer fine-tuning offers a favourable trade-off between accuracy and efficiency for on-device deployment in MEMS-based wearables.

1 Introduction

Wearable devices based on MEMS sensors have become widely integrated into the field of sports, enabling the capture of both kinematic and physiological variables. These technologies have well-established applications in performance monitoring, injury prevention, and rehabilitation [1, 2, 3].

The raw signals produced by these devices are rarely interpretable in their native form. In most practical applications, neither end users nor practitioners inspect these signals directly. Instead, they rely on machine learning models to infer relevant events, behaviours or performance indicators.

However, training such models requires large and carefully labelled datasets, often synchronised with video recordings or expert annotations [4]. While this strategy improves scalability, it also introduces several limitations. Inter-individual differences in movement patterns, body morphology, equipment or even device placement can significantly degrade performance when a single model is applied to all users. In principle, individual richly annotated datasets could mitigate these effects; in practice, this is rarely feasible given the time, cost and expertise required.

This has led to the growing use of neural network personalisation, where generic base models are adapted to the final user [5, 6, 7] using a limited amount

*This work was supported by the Spanish Agencia Estatal de Investigación (PID2020-116011RB-C22, MCIN / AEI / 10.13039/501100011033), the Aragon Regional Government (T27_23R), Unizar, Fundación Bancaria Ibercaja and Fundación CAI (IT 1/25), and the Wallenberg AI, Autonomous Systems and Software Program (WASP).

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

of user-specific data. The aim of this approach is to retain the general knowledge encoded in the model while tailoring its behaviour to that particular user. However, most existing approaches assume either powerful off-device computation or relatively large user-specific datasets, which is often unrealistic in wearable scenarios [5, 6, 7].

In this work, we present a lightweight personalisation framework for MEMS-based wearables, in which a base model is adapted to a specific user using a limited amount of additional data and modest computational resources. The goal is to deploy compact neural networks that can reliably detect the specific actions performed by a given user, while remaining small enough to be trained, adapted and executed directly on the device. We demonstrate this approach in a sports context, using a public padel dataset [8] as a case study.

2 Methodology

This work proposes a lightweight user-personalisation framework for MEMS-based wearables, designed and evaluated through a padel stroke recognition task using inertial time-series data collected from a wrist-worn device. The aim is to personalise the model to the end user, maximising individual accuracy while still leveraging information from the broader player population.

Training with data from multiple players enhances model robustness and generalisation capabilities. However, in real-world use, the primary goal is to achieve high-precision stroke detection for the individual user. To reconcile these objectives, a transfer learning approach is applied, starting from a model pre-trained on data from the full cohort, followed by user-specific fine-tuning using examples from the target user. This setup mimics a realistic deployment scenario

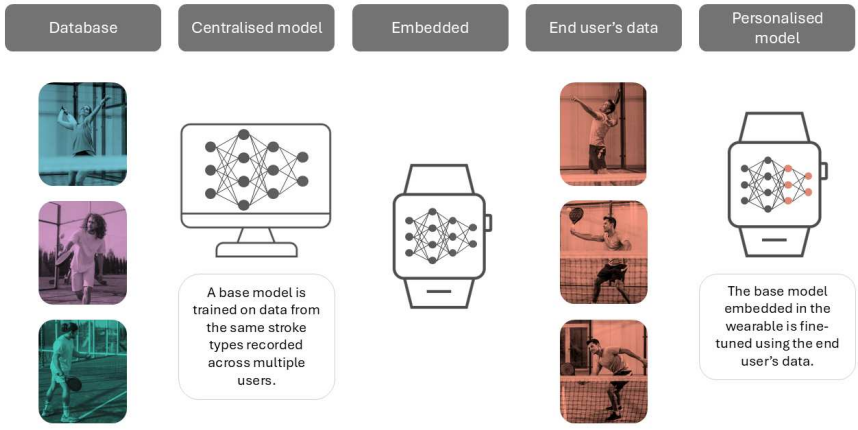

Fig. 1: Proposed framework for padel stroke recognition.Visual assets by Freepik.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

with an initial centralised training phase and a subsequent on-device adaptation stage. Figure 1 illustrates an overview of the proposed framework.

This study builds on the publicly available padel dataset released in [8]. The data are used exactly as provided, without additional preprocessing, filtering or relabelling. The original work focuses on a subset of 12 right-handed players and evaluates performance on seen subjects, which primarily captures within-subject behaviour. By contrast, we use the dataset as fully released, incorporating both right- and left-handed players to better reflect real-world variability and applicability.

On top of this dataset, we implement two fine-tuning strategies to personalise the model: (i) full fine-tuning (Full-FT), where all layers are updated using the target user's data; and (ii) last-layer fine-tuning (LL-FT), where the feature extractor is frozen and only the classification layer is retrained. Both strategies are benchmarked against the pretrained, non-personalised model to quantify the specific benefit of personalisation.

3 Evaluation

3.1 Preliminary Analysis

The dataset contains inertial recordings sampled at 20 Hz in 2 s windows from 23 players (skill levels 1-5) who differ in age, gender and anthropometric characteristics. Each player performed up to 13 different stroke types (e.g., forehand, backhand, volleys, serve). Three practical aspects are particularly relevant for our study: (i) seven of the 23 participants did not perform all stroke types; (ii) there is an unequal number of repetitions per stroke; and (iii) the number of samples per subject and class is relatively low (on average 13.55), which limits both model training and the robust estimation of performance.

3.2 Model Architecture

The architecture evaluated in this work is a 1D convolutional neural network with three convolutional blocks consisting of 128, 96 and 64 filters, respectively. Each convolutional layer is followed by batch normalisation and ReLU activation. A 1D global average pooling (GAP) layer aggregates the temporal features, followed by a dropout layer (rate $= 0.25$) to mitigate overfitting given the limited data volume.

We deliberately adopt a temporal CNN instead of recurrent architectures such as LSTMs or GRUs, as convolutions are typically more parameter- and compute-efficient and offer fixed-latency inference, which is advantageous for deployment on resource-constrained wearable devices.

A dense classification layer is placed on top of the aggregated representation and also serves as the fine-tuning head in the personalised last-layer training scenario. In this case, the convolutional layers are frozen and only the final dense layer is updated, minimising the number of trainable parameters during personalisation and further facilitating deployment on wearable devices.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

3.3 Experimental protocol

In all experiments, the input consisted of the six components of the inertial signals (triaxial accelerometer and gyroscope). From the training set, 20% of the data were reserved for validation using a class-stratified split. The model was trained from random initialisations, and each configuration was repeated five times with different random seeds to reduce the impact of stochastic variation.

The optimisation setup was kept fixed across all conditions, using a batch size of 16 and a learning rate of 10^{-4}. Each run was executed for up to 200 epochs with early stopping based on validation accuracy, using a patience of 10 epochs. The loss function was sparse categorical cross-entropy.

Given the limited number of samples per class and subject, Data Augmentation (DA) was applied using a sliding window over each original segment. Segments of 40 samples per channel were transformed into sub-windows of length 35, generating six overlapping windows per segment with a stride of one. This DA strategy is also straightforward to implement on-device by buffering overlapping signal windows.

Evaluation followed a per-subject protocol. In each iteration, all instances from one player were held out as the test set, while the remaining data were used for training and validation as described above. To ensure fair comparisons, only players with samples for all stroke classes were used as test subjects (16), whereas the remaining participants were included only in the training set. All experiments were run on an Intel(R) Core(TM) i5-9400F CPU @ 2.90 GHz with 32 GB RAM and no GPU, using TensorFlow (Python) developed in PyCharm.

3.4 Results

We then evaluated the effect of per-user personalisation. Table 1 reports the mean and standard deviation (SD) of the weighted F1-score across the 16 subjects for the centralised model baseline, the single-user models, and the personalised models with and without DA. For comparability, we always used 35-sample windows, discarding the last 5 samples in the non-augmented case.

Table 1: Weighted F1-scores (mean ± SD) across the 16 subjects for centralised, single-user and personalised models, with and without data augmentation.

Method	DA	F1-weighted (mean ± SD)
Central (no personalisation)	No	56.7% ± 14.0%
Central (no personalisation)	Yes	56.2% ± 13.7%
Single-user	No	65.7% ± 22.0%
Single-user	Yes	93.0% ± 6.3%
Full-FT	No	82.6% ± 14.6%
LL-FT	No	75.3% ± 15.2%
Full-FT	Yes	93.1% ± 6.1%
LL-FT	Yes	91.3% ± 6.8%

As a reference, we first trained a centralised model using data from all avail-

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

able players, following the subject-wise evaluation protocol. Using the 35 sample window representation with and without DA, the non-personalised central models attain a mean weighted F1-score of approximately 56.5% in both cases. This provides an empirical justification to adopt the 35-sample configuration as our main baseline for the subsequent comparisons.

To characterise the potential performance of a fully personalised model, we trained subject-specific models using only data from each individual player (single-user models). It is worth noting that, when these single-user models are trained without DA, their performance varies greatly across subjects (mean weighted F1-score 65.7%, SD 22.0%), indicating strong inter-subject variability. Applying DA both increases the mean weighted F1-score to 93.0% and markedly reduces inter-subject variability (SD 6.3%), leading to more consistent results across players.

While this single-user configuration provides an optimistic upper bound on achievable performance, it requires training and storing a complete model per player and running a comparatively expensive optimisation process on all its parameters. However, starting from the central model and applying per-user personalisation yields a more favourable trade-off between accuracy and practicality. Without DA, Full-FT and LL-FT achieve 82.6% (SD 14.6%) and 75.3% (SD 15.2%), respectively. When DA is enabled and all user data are used, Full-FT reaches 93.1% (SD 6.1%) and LL-FT 91.3% (SD 6.8%), thus recovering most of the gap to the optimistic single-user configuration while maintaining substantially lower variability across subjects.

From a resource perspective, the model comprises 68,473 parameters, of which 8,320 belong to the final classification layer (12.15%). A single forward pass of the whole model requires 4,051,712 FLOPs ($\approx$4.05 M), while the last layer accounts for 16,384 FLOPs. Assuming a standard backpropagation scheme, one training step of Full-FT (forward and backward through all layers) costs approximately 3 times the forward FLOPs (12.2 M FLOPs per step). In contrast, LL-FT still requires a full forward pass (4.05 M FLOPs), but the backward pass is restricted to the final layer (0.033 M FLOPs), resulting in a total of approximately 4.08 M FLOPs per training step, i.e. about one third of the cost of Full-FT.

Regarding memory consumption during training, the network has 68,473 parameters, which in 32-bit float format correspond to 273,892 bytes. Of these, 8,320 parameters belong to the final classification layer (33,280 bytes). With an optimiser such as Adam, storing two additional values per trainable parameter, the memory needed for weights, gradients and optimiser states is 1,095,568 bytes when all layers are updated in Full-FT and 373,732 bytes when only the last layer is updated in LL-FT. Since both strategies share the same forward activations, this roughly threefold reduction in training-related memory makes LL-FT particularly attractive for devices with tight RAM budgets.

Taken together, these results indicate that per-user personalisation can recover most of the performance gap between the centralised baseline and the optimistic single-user upper bound, while still relying on a compact model. At the same time, the FLOP and parameter analyses, based on operation counts, sug-

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

gest that the proposed architecture is compatible with the memory and compute budgets of contemporary MEMS-based wearables, and that LL-FT in particular offers a favourable compromise between accuracy and resource usage.

4 Conclusions

This work investigated lightweight per-user personalisation of a CNN-based stroke classifier for MEMS-based wearables, using a public padel dataset as a case study. Starting from a single central model trained on all players, we showed that per-user fine-tuning substantially improves performance over the non-personalised baseline and approaches an optimistic single-user upper bound, while relying on a compact architecture compatible with wearable constraints. In particular, last-layer fine-tuning achieves competitive weighted F1-scores with markedly lower computational and memory requirements, offering a favourable accuracy-efficiency trade-off for resource-constrained devices.

FLOP and memory analyses show that the proposed models have a small footprint and that LL-FT significantly reduces training-related RAM usage compared to Full-FT, reinforcing the feasibility of personalised models on embedded platforms. Key limitations are that our deployment analysis is based on operation and parameter counts rather than measurements on an actual wearable; we do not yet evaluate personalisation under more severely limited user-specific data; and we have not systematically explored the optimal window length. Future work will implement and profile the proposed strategies on representative hardware, explore quantisation, online adaptation, and assess their robustness in more data-constrained settings and across different window configurations.

References

[1] L. Liu, X. A. Zhang, Yan Liu, Alberto Martín-Pérez, Lei Liu, and Xuefeng Zhang. A focused review on the flexible wearable sensors for sports: From kinematics to physiologies. *Micromachines 2022, Vol. 13, Page 1356*, 13:1356, 8 2022.

[2] M. Rana and Vi. Mittal. Wearable sensors for real-time kinematics analysis in sports: A review. *IEEE Sensors Journal*, 21:1187–1207, 1 2021.

[3] L. Yang, O. Amin, and B. Shihada. Intelligent wearable systems: Opportunities and challenges in health and sports. *ACM Computing Surveys*, 56:1–42, 7 2024.

[4] Johannes Windischbauer and Jürgen Cito. Label generation for time series data.

[5] J. Schneider and M. Vlachos. Personalization of deep learning. In P. Haber, T. Lampoltshammer, M. Mayr, and K. Plankensteiner, editors, *Data Science – Analytics and Applications*, pages 89–96, Wiesbaden, 2021. Springer Fachmedien Wiesbaden.

[6] Melik Ozolcer and Sang Won Bae. Personalized neural modeling for daily injury risk assessment via wearable health data. In *2025 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 401–406, 2025.

[7] Ziyang He, Yuan Liu, Laurence T. Yang, Zhaoyang Ge, Ling Kuang, Cong Yang, Hangcheng Cao, and Nan Lin. Pfedal-ecg: A lightweight arrhythmia diagnostic system leveraging personalized federated active learning. *Information Fusion*, 127:103860, 2026.

[8] Guillermo Cartes Domínguez, Evelia Franco Álvarez, Alejandro Tapia Córdoba, and Daniel Gutiérrez Reina. A comparative study of machine learning and deep learning algorithms for padel tennis shot classification. *Soft Computing*, 27:12367–12385, 9 2023.