Solving the Cold-Start Problem for the Edge: Clustering and Adaptive Deep Learning for Emotion Detection

1st Junjiao Sun
Centro de Electrónica Industrial
Universidad Politécnica de Madrid
Madrid, Spain
0000-0001-6940-3164
junjiao.s@upm.es

2nd Laura Gutiérrez Martín
Departamento de Tecnología Electrónica
Universidad Carlos III de Madrid
Madrid, Spain
0000-0002-0679-7300
lagutier@ing.uc3m.es

4nd Jose Miranda
Embedded Systems Laboratory
EPFL
Lausanne, Switzerland
0000-0002-7275-4616
jose.mirandacalero@epfl.ch

3rd Celia López Ongil
Departamento de Tecnología Electrónica
Universidad Carlos III de Madrid
Madrid, Spain
0000-0001-9451-6611
celia@ing.uc3m.es

5st Jorge Portilla
Centro de Electrónica Industrial
Universidad Politécnica de Madrid
Madrid, Spain
0000-0003-4896-6229
jorge.portilla@upm.es

6st Andrés Otero
Centro de Electrónica Industrial
Universidad Politécnica de Madrid
Madrid, Spain
0000-0003-4995-7009
joseandres.otero@upm.es

***Abstract*—Designing AI-based applications personalized to each user's behavior presents significant challenges due to the cold start problem and the impracticality of extensive individual data labeling. These challenges are further compounded when deploying such applications at the edge, where limited computing resources constrain the design space. This paper introduces a novel approach to AI-driven personalized solutions in biosensing applications by combining deep learning with clustering-based separation techniques. The proposed *Clustering and Learning for Emotion Adaptive Recognition (CLEAR)* methodology strikes a balance between population-wide models and fully personalized systems by leveraging data-driven clustering. CLEAR demonstrates its effectiveness in emotion recognition tasks, and its integration with fine-tuning enables efficient deployment on edge devices, ensuring data privacy and real-time detection when new users are introduced to the system. We conducted experiments for model personalization on two edge computing platforms: the Coral Edge TPU Dev Board and the Raspberry Pi with an Intel Movidius Neural Compute Stick 2. The results show that initial cluster assignment for new users can be achieved without labeled data, directly addressing the cold-start problem. Compared to baseline validation without clustering, this proposal improves accuracy metric from 75% to 81.9%. Furthermore, fine-tuning with minimal labeled data significantly improves accuracy, achieving up to 86.34% for the fear detection task in the WEMAC dataset while remaining suitable for deployment on resource-constrained edge devices.**

***Index Terms*—Emotion Recognition, Edge Devices, Clustering, Adaptive Deep Learning, Cold Start.**

I. Introduction

THE rapid development of edge computing technologies, mainly driven by the integration and deployment of artificial intelligence (AI) into such constraint systems, is enabling personalized experiences through immediate extraction and processing of relevant data tailored to each user or context. Typically, personalization in AI models follows two approaches: deploying general models trained on aggregated population data or utilizing personalized models customized for individual users or specific environments [1]. These are also known as subject-independent and subject-dependent, respectively. Although general models can be applied to new users who enter the system, they tend to exhibit lower accuracy due to the lack of adaptation to specific user behaviors. In contrast, personalized AI models, customized to each user's physiological responses, have been shown in the literature to deliver superior accuracy [2]. However, the development of such personalized models has largely been confined to academic research utilizing preexisting datasets rich in labeled data. This scenario is not easily replicated in real-world applications where the labeled data are scarce or non-existent.

Existing methods to mitigate the lack of labeled data, also known as the cold-start problem, often rely on either collecting extensive labeled data from new users [3] or employing transfer learning techniques that adapt pre-trained models to new users' data [4]. However, these approaches are of limited application in edge computing scenarios due to several constraints. Collecting large amounts of labeled data from users is impractical and raises privacy concerns [5]. At the same time, transfer learning methods typically require substantial computational resources and time, which are limited on edge devices. Furthermore, the heterogeneity of user behaviors and responses further complicates the development of a one-size-fits-all solution [6]. Therefore, there is a critical need for a system that can initially solve the cold-start problem while being suitable for deployment on edge devices, with the limitations inherent in resource-constrained environments, to further personalize it

978-3-9826741-0-0/DATE25/©

with minimal labeled data and computational overhead.

This paper introduces *Clustering and Learning for Adaptive Emotion Recognition (CLEAR)*, a novel approach that integrates deep learning (DL) with clustering techniques for emotion recognition. Instead of training a single monolithic DL model for the entire population, CLEAR involves clustering users based on their initial responses to stimuli and identifying groups of individuals with similar physiological responses. This enables the creation of semi-personalized models that strike a balance between fully personalized and generalized models. When new users enter the system, they are assigned to a cluster based on their initial and unlabeled data input, effectively addressing the cold-start problem from an unsupervised perspective. Additionally, after the cluster assignment, our methodology allows for fine-tuning the model with only a few labeled samples from the new user.

Original contributions to this paper are the following.

- Introduction of CLEAR, a novel methodology that effectively addresses the cold-start problem from an unsupervised perspective. By assigning new users to appropriate models based solely on their initial, unlabeled data input, we eliminate the need for prior data or extensive labeling. This makes this approach highly practical for real-world applications on edge devices.
- Integration and validation of the proposed approach in the context of emotion recognition. Using wearable physiological signals as input, we demonstrate the practical viability and effectiveness of our methodology in real-world edge computing environments.
- The experimental results show an average accuracy of the targeted emotion of up to 86.34% after fine-tuning. To the best of our knowledge, CLEAR outperforms any other state-of-the-art result for emotion recognition under cold-start and fine-tuning considerations.

II. Related Work

Clustering-based approaches, adaptive DL, and cold-start problems are all active areas of research. This section reviews key related work in these areas, highlighting the contributions and limitations of existing methodologies in the context of edge computing and biosensing applications.

A. Clustering-Based Approaches

Several studies have demonstrated the effectiveness of clustering techniques in reducing model complexity and improving accuracy in biosensing applications. For example, the authors in [7] proposed a cluster-based analysis for personalized stress evaluation using physiological signals. Their method uses k-means clustering to group individuals based on their stress responses and employs a cluster-specific model for more accurate stress evaluation. Similarly, in [8], a community-based federated learning algorithm using clustering to improve the prediction of patient mortality was introduced. By clustering patients into communities based on similar clinical characteristics, the system reduced data complexity and improved model performance in distributed environments. More recently, in [9], a globally generalized emotion recognition system was explored. Specifically, it applied clustering to enhance the robustness of emotion detection across different individuals and physiological sensors. These studies highlight the potential of clustering to address inter-subject variability in biosensing applications. In contrast to these approaches, our CLEAR methodology integrates clustering with DL to create semi-personalized models that are both accurate and computationally efficient, addressing the cold-start problem, enabling fully unsupervised assignment of new users to existing clusters.

B. Adaptive Deep-Learning and Cold-Start Problem

DL has shown exceptional performance across different biosensing applications [10]. However, traditional DL models often struggle when they are applied to highly dynamic environments. In these scenarios, static models, which are trained once on a fixed dataset, tend to fall short because they cannot adapt to the variability and heterogeneity of user-specific data. This limitation has led to the rise of adaptive DL, where models are designed to adjust their parameters based on new data or changing user conditions.

One notable example of adaptive DL is presented in [11], where domain localization techniques are used to dynamically adapt the models in robotic applications. While this approach allows for specialized adaptation, it requires significant computational resources for retraining and fine-tuning. Within the context of this work, this is a critical disadvantage as it limits real-time adaptability, especially for applications with strict latency and power consumption requirements. Furthermore, the method focuses primarily on domain shifts rather than user-specific adaptation. Note that this fact is essential in applications where inter-user variability is significant, such as emotion recognition. Particularly focusing on the latter type of applications, which are strongly driven by the need for personalisation, different works have also successfully applied adaptive DL schemes. For instance, the authors in [12] proposed a user-adaptive model that leverages deep transfer learning and data augmentation to personalize models with minimal user data. This method offers the advantage of efficient user adaptation without the need for extensive labelled data, making it ideal for cold-start scenarios. However, they still rely on a single pre-trained general model for the entire set of users. This approach may not generalize well across all the different users, especially when the underlying data distributions differ significantly. In fact, that can lead to sub-optimal performance for new incoming individuals whose data deviates from the average patterns captured by the general model. In [13], the authors developed a DL model with adaptive regularization for bio-potential based emotion recognition. In this case, they integrated adaptive regularization to eliminate biases in the model caused by data from different users. While adaptive regularization improves the robustness of the model, it requires careful tuning of hyperparameters, which can be a high resource-intensive process. Moreover, the method primarily addresses variability at the model level but does not tackle the challenge of real-time personalisation, a key requirement for edge computing in emotion recognition tasks.

These studies fully demonstrate that adaptation makes models more flexible, efficient, and capable of generalizing across diverse scenarios. However, several critical disadvantages remain, particularly in the context of real-time and resource-constrained edge environments. Issues such as high computational cost, the reliance on only one pre-trained model for the whole population, and limited real-time adaptability restrict their practical applications in edge computing scenarios.

C. Edge Computing for Emotion Detection

Building on the need for real-time adaptability and personalized models in emotion recognition, edge computing has emerged as a crucial enabler for these applications. However, the deployment of DL models on resource-constrained edge devices presents significant challenges, particularly in terms of balancing accuracy, computational efficiency, and personalisation.

Within this context, several recent studies have explored the potential of edge-based emotion recognition using lightweight machine learning techniques. For instance, the authors in [14] developed an adaptive emotion recognition system using smartwatches. By leveraging machine learning classifiers, they achieved an accuracy of 74.30% across 30 volunteers. While the study demonstrates the feasibility of real-time emotion detection on wearables, the reliance on traditional machine learning techniques limits the system's ability to generalize across a wider range of emotional states and users. Following this approach, different studies can be found facing the same limitations with respect to the lack of deep feature extraction capabilities [15].

Looking instead at DL workloads deployment, several recent works have addressed the challenge of deploying different models in environments requiring user-specific adaptation or personalisation [16], [17]. These approaches highlight the ongoing efforts to improve model adaptability and personalisation in DL. However, they do not provide any mechanism to handle the cold-start problem in edge environments. In this context and to the best of our knowledge, [3] is the only study that addresses the cold-start problem in real-life emotion recognition with wearables. Their approach introduces per-group personalisation, in which models are personalized for groups of users rather than individuals, to mitigate the cold-start issue. The study shows that a small amount of labelled data collected over a week can significantly improve model performance. However, the reliance on labelled data, even if minimal, to tackle the cold-start problem and ideally a further personalisation, introduces a bottleneck, as obtaining labelled data from new users can be time-consuming and impractical in many real-world scenarios. Additionally, while their group-based approach offers a solution to the cold-start problem, it lacks a fully unsupervised mechanism for initial model assignment, which could limit scalability and responsiveness in real-time applications.

The CLEAR methodology proposed in this work bridges this gap by leveraging clustering and adaptive DL techniques to create semi-personalized models that can be fine-tuned with minimal data. Unlike existing methods, CLEAR uses an unsupervised clustering approach to assign new users to clusters based on their physiological responses.

III. Proposed Methodology

The proposed method assumes the availability of an initial set of users, likely during a pre-deployment phase, with sufficient labelled data for these users. While this labelled data is crucial for developing the baseline DL models, it is impossible to guarantee that labelled data will also exist for new incoming users to the system once it is deployed. The second stage, involving cold-start handling and fine-tuning, has been incorporated to address this latter challenge. Figure 1 presents a simplified representation of the architecture of the CLEAR methodology. The two stages of the methodology are described next.

A. Initial Clustering and Learning on the Cloud (CL)

The initial training of the DL models is conducted in the cloud as part of the offline design process. However, rather than generating monolithic models personalized for each user or generalized for the entire population, a clustering process is first applied to identify commonalities among users in the initial dataset. Models are then trained for each identified cluster. Instead of using raw signals to train the models, a preliminary step is performed to extract feature maps, which are subsequently used for clustering and model training.

1) Feature map generation: The raw data from the collected dataset is processed to generate two-dimensional (2D) feature maps. This method draws on the strategy described in [18] and extracts 123 features from raw physiological signals, including the time domain, frequency domain and non-linear feature. Let $X \in \mathbb{R}^{T \times S}$ represent the raw data, where T is the number of time samples and S represents the different sensor modalities. A feature extraction process is applied to X, yielding a set of F features per time window to capture meaningful patterns. These features are then organized into a 2D matrix, $M \in \mathbb{R}^{F \times W}$, where W represents the number of time windows. This matrix serves as the feature map, which will be used in the subsequent stages of the methodology. Note that the feature map reduces the complexity of the raw data by capturing relevant temporal and statistical characteristics while discarding redundant information. This step ensures that the downstream clustering methods and DL models can effectively learn from the processed data, improving the training efficiency and model robustness.

2) Global Clustering (GC): After generating the feature maps, the next step is to separate users into clusters based on the similarity of their physiological responses. This will reduce data complexity and improve model performance. Thus, for the initial clustering separation or GC, the clustering algorithm is applied to the initial extracted features, where a matrix $D \in \mathbb{R}^{F \times N}$ is formed, with N representing the number of users. The goal is to determine the optimal number of clusters K using standard techniques or algorithms. For each cluster $k \in \{1, \ldots, K\}$, centroids $C_k \in \mathbb{R}^F$ are computed, representing the cluster's central point in feature space. An iterative process is then used to refine these clusters. Training subsets of data are repeatedly sampled, and the centroids are recalculated. Users are reassigned if their current cluster is no longer the closest based on the updated centroids. This ensures

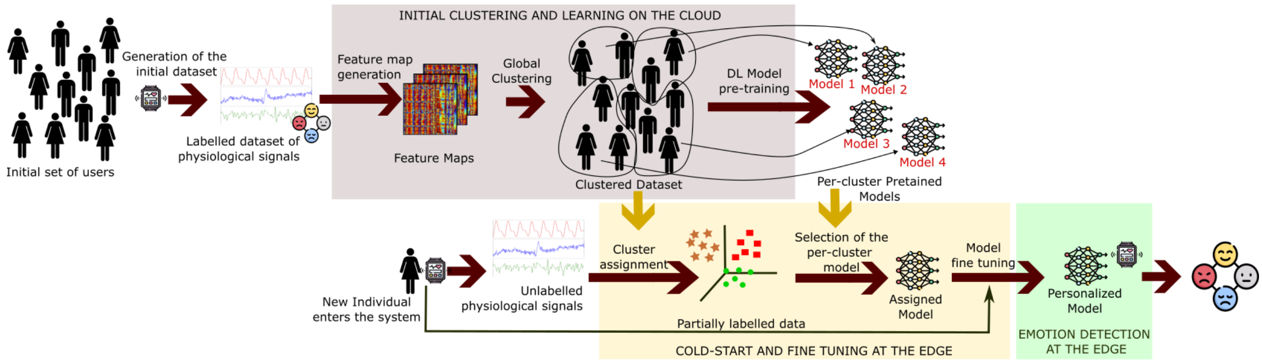

Fig. 1. Overview of the architecture of the CLEAR methodology, including first initial clustering and learning carried out on the cloud, and the cold-start and fine-tuning solutions at the edge.

robust and optimized clustering across the population, capturing the main patterns in the user data. Note that this process is based on the proposal in [19].

3) Initial System Pre-training: Once users are grouped into clusters, the next step is to pre-train a DL model for each cluster. The 2D feature maps generated in the earlier step allow for the task to be handled as an image classification problem, enabling the use of convolutional neural networks (CNNs), which have been proven highly effective. We introduce a simple DL architecture combining two convolutional layers with a long short-term memory (LSTM) layer, as shown in Figure 2. LSTMs have proven strong capabilities for learning contextual sequential information. Emotions can be considered a continuous physiological response, and the physiological signals used to record emotions also contain rich sequential features [20]. Therefore, we chose the CNN-LSTM architecture, which can effectively integrate feature maps' global and sequential information, ultimately enhancing classification accuracy. Following this process, we obtain as many pre-trained DL models as clusters in the system.

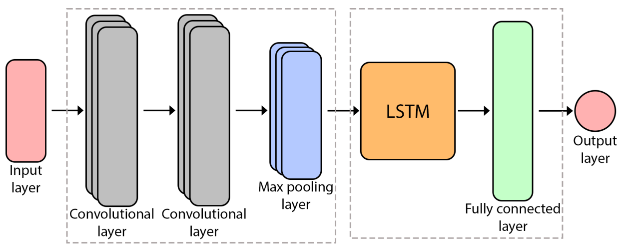

Fig. 2. The CNN-LSTM architecture proposed for emotion recognition from 2D feature maps.

B. Cold-start and Fine-tuning at the Edge

After the cloud-based clustering and pre-training process, the cluster-specific models are ready for deployment on edge devices. However, when a new user adopts the DL-based system, it faces a cold-start problem due to the absence of prior data for that user. To address this, we propose a cold-start and fine-tuning strategy that leverages the clustering approach from the initial training phase. This process is lightweight enough to be performed at the edge, minimizing or eliminating the need for new users to share their personal physiological data. It consists of two subsequent steps, as described next.

1) Cluster assignment: When a new user enters the system, its feature map $M_{\text{new}} \in \mathbb{R}^{F \times W}$ is generated from its physiological data, and it must be assigned to one of the pre-existing clusters. It must be noticed that this Cluster Assignment (CA) must handle new, unseen observations from volunteers who were not part of the initial clustering process. To achieve this, the system computes the distance between the new user's feature map and the centroids of the existing clusters C_k. For each cluster k, internal centroids $C_{k,i} \in \mathbb{R}^P$ for sub-clusters $i \in \{1, \dots, I_k\}$ are used to refine this process. The distance from the new user's data to the centroids of the main clusters and their internal subdivisions is calculated. The new user is assigned to the cluster that minimizes the overall summation of distances to these internal centroids.

This approach allows for a more granular and accurate initial assignment of new users by leveraging the hierarchical structure within each cluster. This first, unsupervised assignment provides a robust starting point for the system to personalize the model in the last step. Note that, as for the global clustering, this process is also based on [19].

2) Final personalisation: While the cluster-specific model provides a solid starting point, further personalisation is needed to tailor the model to the unique physiological patterns of the new user. The system achieves this by fine-tuning the selected model with a small set of labelled data from the new user. In practice, the user is asked to label a minimal portion of their physiological data, which is then used to update the model. Due to the pre-trained models' small size and efficient architecture, this fine-tuning process can be executed on the edge device itself, allowing for real-time personalisation without sacrificing computational efficiency and privacy.

IV. Experimental Results

This section details the experiments conducted to validate the effectiveness and stability of the proposed methodology. It involves the setup of our experimental environment, how the clustering strategy proceeds, and the description of used data sets. Then, validation metrics for CLEAR are presented to demonstrate our method's advantages against several comparison tests and prove the method's robustness. Finally, several edge devices are used for the cloud-edge system with fine-tuned training to test whether it can produce better results for a specific user, aiming to validate our method's usability in real world.

A. Experimental setup

For our experiments, we utilized the WEMAC dataset [21], a comprehensive multi-modal repository that includes physiological signals such as BVP, GSR, and SKT. The dataset is annotated with ten emotional labels: *fear*, *joy*, *hope*, *surprise*, *anger*, *tedium*, *tenderness*, *calm*, *disgust*, and *sadness*, with the *fear* label representing 44.4% of the data. To address the class imbalance and streamline the analysis, we restructured the dataset into a binary classification task by consolidating the labels into fear and non-fear categories. Specifically, we focused on the first release of the dataset, which comprises data from $N = 44$ volunteers.

For all experiments and following the method for generating 2D feature maps outlined in Section III, we generated approximately 800 feature maps from the WEMAC dataset. Specifically, 123 features were extracted, including 34 for GSR, 84 for BVP, and five for SKT. Moreover, with respect to the total number of clusters to be deployed in the first phase of CLEAR, we selected $K = 4$ clusters. This division was based on preliminary analysis, which indicated it offered the best balance between intra-cluster similarity and inter-cluster separation. With such cluster division, on average, regardless of the specific validation or iteration performed, we obtained 17, 13, 7, and 7 volunteers belonging to clusters 1, 2, 3, and 4, respectively.

The Leave-One-Subject-Out (LOSO) cross-validation technique was employed to evaluate the proposed CLEAR methodology. This approach is widely used in scenarios involving multiple subjects, where the model is trained and tested across different individuals. In each iteration, one subject is excluded from the dataset, and the remaining $N - 1$ subjects are used to train and validate. This process is repeated until every subject has been left out once. This process is repeated for each subject, ensuring unbiased clustering and training by excluding the test subject's data. This approach validates the entire dataset.

B. CLEAR Validation

To the best of our knowledge, only two previous studies, [22] and [18], have been conducted using the WEMAC dataset. Although they did not present any clustering or adaptive DL methods, they will be used as the reference for accuracy in this work. Based on the method proposed in Section III, this experiment was divided into two validation scenarios: *Clustering and Learning validation (CL validation)* and complete *CLEAR validation*. Meanwhile, robustness tests (RT) were also carried out to verify whether the proposed method can effectively improve the performance of other alternatives. Mainly, the two envisaged validations are as follows:

- CL validation: This first phase of CLEAR (involving only the clustering and pre-training on the cloud) is validated using GC to separate clusters for the complete dataset of N users, as explained in Section III-A. Then, we perform intra-cluster LOSO for each cluster to ensure the training process is unbiased. For the RT in CL validation, we also use the volunteers from the other clusters as the test set to evaluate the performance of the trained model on data outside the current cluster.
- CLEAR validation: This is the validation of the whole CLEAR pipeline. In this case, LOSO is applied from the beginning by selecting one volunteer, denoted as V_x, to be the one later considered new volunteer. V_x is left out of the initial clustering and model pre-training. The remaining volunteers are clustered. The four resulting clusters are trained separately, and the best-performing training checkpoints for each cluster are saved. Then, V_x is assigned to its respective cluster through fully unsupervised CA using just 10% of the data. This process is repeated until each volunteer has served as V_x (CLEAR w/o FT). RT in this validation is applied when V_x is assigned to its respective cluster so that such volunteers are also tested using the models belonging to the other clusters. Moreover, the fine-tuning strategy mentioned in Section III-A was also applied in CLEAR validation to determine whether it can improve performance (CLEAR w FT). Specifically, 20% of labelled data from V_x was used as a new training set, which was input into the trained checkpoint corresponding to its cluster for re-training. The remaining 80% was used as the test set.

Finally, a General model was developed for comparison. Specifically, x volunteers were randomly selected without clustering, and the same DL model was used for training. In this case, the training process of the general model is the same as CL validation. As mentioned earlier, this experiment divided all volunteers into four clusters. Therefore, x was identified as 11, for a fair comparison with an average cluster size. This comparison aims to analyze whether the combination of clustering with DL could enhance performance versus classifying data of the same size without clustering using DL.

Table I shows the validation results, highlighting that CL and CLEAR models outperform previous methods, proving the effectiveness of combining clustering with DL in enhancing classification performance over population-wide models. The comparison between CL validation and the General model highlights how clustering users with similar behaviors helps classification algorithms learn patterns more effectively, improving accuracy.

The results from RT CL and CL validation confirm that our GC approach effectively separates volunteers into optimal groups for DL training. Similarly, the outcomes for RT CLEAR and CLEAR without fine-tuning (w/o FT) demonstrate that our CA method successfully assigns new users to the correct clusters. Lastly, the results for CLEAR with fine-tuning (w/ FT) show that incorporating a small amount of labelled data from new users through retraining significantly enhances model precision.

TABLE I
COMPARISON OF THE PROPOSED METHOD WITH THE EXISTING REFERENCES FOR WEMAC *(fear and non-fear)*.

Validation func	Accuracy	STD (Acc)	F1-score	STD (F1)
Previous works in the state-of-the art				
Bindi [22]	64.63	16.56	66.67	17.31
Sun et al. [18]	79.90	4.16	78.13	6.52
Without Clustering				
General Model	75.00	2.76	72.57	3.12
Clustering and Learning (CL) validation				
RT CL	64.33	1.80	62.42	1.57
CL validation	**81.90**	3.44	**80.41**	3.58
CLEAR Validation				
RT CLEAR	72.68	5.10	70.98	4.26
CLEAR w/o FT	80.63	4.22	79.97	4.74
CLEAR w FT	**86.34**	4.04	**86.03**	5.04

C. Edge validation

As described in Section III, a Cloud-Edge system has been designed to develop an optimization scheme based on CLEAR that is more suitable for hardware platforms. This experiment utilizes two hardware platforms: the Coral Edge TPU Dev Board (TPU) [23], and Raspberry Pi combined with Intel Movidius Neural Compute Stick 2 (NCS2) [24]. Coral Edge TPU Dev Board has a hardware accelerator designed to speed up machine learning inference, which could also help accelerate the DL process. NCS2 is a plug-and-play USB device which could combine with a development board such as Raspberry Pi, which allows it to prototype and deploy AI applications quickly.

The results are presented in Table II upper part. Based on the validation CLEAR, the best checkpoints have been obtained and deployed into those edge devices. After that, new user V_x data would be loaded into the devices and assigned to a cluster based on CA. Then fine-tune the models on the edge to get the validation result. Meanwhile, this experiment also uses RT CLEAR for the comparison. The training results on the GPU are considered the baseline. The performance of TPU is lower than baseline due to it only support for only 8-bit data.

Furthermore, the fine-tuning process is conducted on each edge device. As shown in Table II bottom part, the results after fine-tuning are significantly better than those in Table II upper part, demonstrating that the cloud-edge system proposed in this paper offers optimal performance for real-world applications tailored to individual users.

In addition to accuracy, we recorded time and power consumption during the experiments, as these are critical factors in edge computing. Table II bottom part presents the average computation time and power consumption during the re-training and testing phases. "Re-training" refers to the period from the start of fine-tuning until model Convergence. In contrast, "Test" refers to inputting a feature map into the model and obtaining the output. "Baseline" represents the state when no programs run on the two edge devices. Meanwhile, the unit of each measurements are indicated on the right.

The results indicate that the re-training and testing times on the Pi+NCS2 are significantly longer than those on the TPU, likely due to the TPU's machine learning accelerator and its ability to process 8-bit data, which may converge faster during training. These findings on time and power consumption further support the practicality and efficiency of the proposed method for real-world edge deployments.

TABLE II
THE VALIDATION RESULTS OF CLOUD-EDGE SYSTEM IN DIFFERENT PLATFORMS FOR CLEAR. MTC AND MPC REPRESENT MEAN TIME CONSUMPTION AND MEAN POWER CONSUMPTION RESPECTIVELY.

Platform	Accuracy	STD (Acc)	F1-score	STD (F1)
GPU (baseline)	**80.63**	4.22	**79.97**	4.74
Coral TPU	74.17	3.84	73.57	4.44
RT CLEAR	65.32	5.42	64.79	4.82
Pi + NCS2	79.03	4.10	78.48	4.76
RT CLEAR	68.47	3.25	69.02	4.14

	GPU	TPU	Pi + NCS2	unit
Accuracy	86.34	79.40	84.49	-
Accuracy std	4.04	4.51	4.82	-
F1-score	86.03	79.14	84.07	-
F1 std	5.04	4.66	5.16	-
MTC Re-training	-	32.48	78.52	s
MPC Re-training	-	1.82	3.78	W
MTC Test	-	47.31	239.70	ms
MPC Test	-	1.64	3.43	W
MPC Baseline	-	1.28	2.76	W

V. CONCLUSION AND FUTURE WORKS

In this paper, we proposed CLEAR, a novel approach designed to address the cold-start problem in edge computing. The methodology combines DL with clustering techniques to enable semi-personalized models that balance fully personalized and general models. By clustering users based on initial, unlabeled data and assigning new users to appropriate clusters, CLEAR effectively tackles the cold-start problem from an unsupervised perspective. We demonstrated the effectiveness of this approach by validating it on edge platforms such as the Coral Edge TPU and Raspberry Pi combined with Intel Movidius Neural Compute Stick 2, where minimal labelled data was required to fine-tune cluster-specific models. The proposed CNN-LSTM architecture for classification tasks was optimized to balance performance and deployability, ensuring that the models are lightweight enough for edge devices without sacrificing accuracy. The results showed that CLEAR significantly improves emotion recognition, particularly in the context of new users, while adhering to the requirements of edge computing, such as low latency and privacy preservation.

Future work will explore expanding the methodology to other physiological signals and domains and further optimizing the clustering and model personalisation processes to reduce the need for labelled data. Additionally, testing CLEAR in more datasets will help further validate its robustness and practical applications in wearable technologies and personalized healthcare. The methodology will be also improved to assure low power devices to further enhance real-world usability.

ACKNOWLEDGMENT

This work has been supported by the National Research Project Bindi-Tattoo PID2022-142710OB-I00 (Agencia Estatal de Investigación, AEI, Spain), as well as by the Spanish MCIN/AEI/10.13039/501100011033 and the European Commission "NextGenerationEU"/PRTR under the project COGNITION (TED2021-132475B-I00).

REFERENCES

[1] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "Deep learning and model personalization in sensor-based human activity recognition," *Journal of Reliable Intelligent Environments*, vol. 9, no. 1, pp. 27–39, 2023.

[2] D. Nath, Anubhav, M. Singh, D. Sethia, D. Kalra, and S. Indu, "A comparative study of subject-dependent and subject-independent strategies for eeg-based emotion recognition using lstm network," in *Proceedings of the 2020 4th International Conference on Compute and Data Analysis*, ser. ICCDA '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 142–147. [Online]. Available: https://doi.org/10.1145/3388142.3388167

[3] S. Saganowski, D. Kunc, B. Perz, J. Komoszyńska, M. Behnke, and P. Kazienko, "The cold start problem and per-group personalization in real-life emotion recognition with wearables," in *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 2022, pp. 812–817.

[4] S. Omidvar and T. Tran, "Tackling cold-start with deep personalized transfer of user preferences for cross-domain recommendation," *International Journal of Data Science and Analytics*, pp. 1–10, 2023.

[5] N. Gahlan and D. Sethia, "Federated learning inspired privacy sensitive emotion recognition based on multi-modal physiological sensors," *Cluster Computing*, vol. 27, no. 3, pp. 3179–3201, 2024.

[6] C.-H. Lu and X.-Z. Lin, "Toward direct edge-to-edge transfer learning for iot-enabled edge cameras," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4931–4943, 2020.

[7] Q. Xu, T. L. Nwe, and C. Guan, "Cluster-based analysis for personalized stress evaluation using physiological signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 275–281, 2015.

[8] L. Huang, A. L. Shea, H. Qian, A. Masurkar, H. Deng, and D. Liu, "Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records," *Journal of Biomedical Informatics*, vol. 99, p. 103291, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046419302102

[9] M. Ali, F. Al Machot, A. Haj Mosa, M. Jdeed, E. Al Machot, and K. Kyamakya, "A globally generalized emotion recognition system involving different physiological signals," *Sensors*, vol. 18, no. 6, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/6/1905

[10] T.-P. Jung, T. J. Sejnowski *et al.*, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 96–107, 2019.

[11] G. Angeletti, B. Caputo, and T. Tommasi, "Adaptive deep learning through visual domain localization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 7135–7142.

[12] E. Garcia-Ceja, M. Riegler, A. K. Kvernberg, and J. Torresen, "User-adaptive models for activity and emotion recognition using deep transfer learning and data augmentation," *User Modeling and User-Adapted Interaction*, vol. 30, no. 3, pp. 365–393, 2020.

[13] A. Samavat, E. Khalili, B. Ayati, and M. Ayati, "Deep learning model with adaptive regularization for eeg-based emotion recognition using temporal and frequency features," *IEEE Access*, vol. 10, pp. 24 520–24 527, 2022.

[14] Z. Wang, Z. Yu, B. Zhao, B. Guo, C. Chen, and Z. Yu, "Emotionsense: An adaptive emotion recognition system based on wearable smart devices," *ACM Transactions on Computing for Healthcare*, vol. 1, no. 4, pp. 1–17, 2020.

[15] L. Shu, Y. Yu, W. Chen, H. Hua, Q. Li, J. Jin, and X. Xu, "Wearable emotion recognition using heart rate data from a smart bracelet," *Sensors*, vol. 20, no. 3, p. 718, 2020.

[16] G. A.V., M. T., P. D., and U. E., "Multimodal emotion recognition with deep learning: Advancements, challenges, and future directions," *Information Fusion*, vol. 105, p. 102218, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253523005341

[17] A. Kaklauskas, A. Abraham, I. Ubarte, R. Kliukas, V. Luksaite, A. Binkyte-Veliene, I. Vetloviene, and L. Kaklauskiene, "A review of ai cloud and edge sensors, methods, and applications for the recognition of emotional, affective and physiological states," *Sensors*, vol. 22, no. 20, p. 7824, 2022.

[18] J. Sun, J. Portilla, and A. Otero, "A deep learning approach for fear recognition on the edge based on two-dimensional feature maps," *IEEE journal of biomedical and health informatics*, 2024.

[19] L. Gutiérrez-Martín, C. López-Ongil, J. M. Lanza-Gutiérrez, and J. A. Miranda Calero, "Personalized clustering for emotion recognition improvement," *Sensors*, vol. 24, 2024.

[20] J. Sun, J. Portilla, and A. Otero, "Negative emotion recognition based on physiological signals using a cnn-lstm model," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2024, pp. 3736–3741.

[21] J. A. Miranda, E. Rituerto-González, L. Gutiérrez-Martín, C. Luis-Mingueza, M. F. Canabal, A. R. Bárcenas, J. M. Lanza-Gutiérrez, C. Peláez-Moreno, and C. López-Ongil, "Wemac: Women and emotion multi-modal affective computing dataset," *arXiv preprint arXiv:2203.00456*, 2022.

[22] J. A. Miranda Calero, E. Rituerto-González, C. Luis-Mingueza, M. F. Canabal, A. R. Bárcenas, J. M. Lanza-Gutiérrez, C. Peláez-Moreno, and C. López-Ongil, "Bindi: Affective internet of things to combat gender-based violence," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21 174–21 193, 2022.

[23] "Google coral tpu dev board." [Online]. Available: https://coral.ai/products/dev-board

[24] "Intel movidius." [Online]. Available: https://www.developer.movidius.com/