Deep Generative Domain Adaptation with Temporal Relation Knowledge for Cross-User Activity Recognition

Xiaozhou Ye[1] , **Kevin I-Kai Wang**[2]

[1] [2]Department of Electrical, Computer, and Software Engineering, The University of Auckland,
Auckland, New Zealand
xye685@aucklanduni.ac.nz, kevin.wang@auckland.ac.nz

Abstract

In human activity recognition (HAR), the assumption that training and testing data are independent and identically distributed (i.i.d.) often fails, particularly in cross-user scenarios where data distributions vary significantly. This discrepancy highlights the limitations of conventional domain adaptation methods in HAR, which typically overlook the inherent temporal relations in time-series data. To bridge this gap, our study introduces a Conditional Variational Autoencoder with Universal Sequence Mapping (CVAE-USM) approach, which addresses the unique challenges of time-series domain adaptation in HAR by relaxing the i.i.d. assumption and leveraging temporal relations to align data distributions effectively across different users. This method combines the strengths of Variational Autoencoder (VAE) and Universal Sequence Mapping (USM) to capture and utilize common temporal patterns between users for improved activity recognition. Our results, evaluated on two public HAR datasets (OPPT and PAMAP2), demonstrate that CVAE-USM outperforms existing state-of-the-art methods, offering a more accurate and generalizable solution for cross-user activity recognition.

1 Introduction

Human Activity Recognition (HAR) is an essential field within Human-Computer Interaction, ubiquitous computing [Ma *et al.*, 2019], and the Internet of Things [Abdallah *et al.*, 2018]. It involves identifying human activities using sensor data and contextual information [Qian *et al.*, 2019]. HAR finds applications in diverse areas like medical treatment, assisted living, fitness, security, and home automation. Current HAR methods, especially those processing time-series sensor data, operate under the assumption that training and testing data are drawn from the same distribution, meaning they are independent and identically distributed (i.i.d.) [Wilson and Cook, 2020]. This approach assumes that a model trained on source data will perform similarly on target data, as long as both sets of data are from the same domain with consistent features and distribution characteristics. However, this is often not the case in real-world scenarios, where training and testing datasets may have different distributions due to data heterogeneity, also known as the out-of-distribution (o.o.d.) problem. This discrepancy can lead to reduced model performance when applied to new, unseen data [Patel *et al.*, 2015].

Transfer learning, specifically domain adaptation [Pan and Yang, 2009], is a method that addresses this data heterogeneity issue. Its core principle is to identify and leverage common knowledge to minimize differences in data distribution between the source and target domains. Current research in domain adaptation primarily focuses on static data, like images, where each sample is considered i.i.d. within its domain [Patel *et al.*, 2015]. This assumption is extended to time-series data in HAR, treating each data segment as i.i.d. in its domain [Soleimani and Nazerfard, 2021]. However, this approach overlooks the inherent temporal relations in time-series data, where consecutive data segments are often interdependent. For example, in activity recognition, activities like walking consist of interconnected sub-activities (e.g., lifting the leg, moving it forward, placing the foot down), each with temporal dependencies. The prevalent oversight in current research methods is the insufficient attention to these temporal dependencies, which are crucial for understanding and accurately recognizing human activities. This lack of focus results in models that fail to capture the common sequential dynamics of activity data, leading to suboptimal performance on domain adaptation, especially when applied across diverse users.

This paper introduces a novel approach called Conditional Variational Autoencoder with Universal Sequence Mapping (CVAE-USM) for time series domain adaptation in HAR. This method focuses on capturing temporal relations in activity data by USM encoding, enabling better adaptation between different users' data. By preserving these temporal relations, our approach effectively regularizes the adaptation process, enhancing the alignment of sub-activity distributions across users. This results in improved performance in domain adaptation tasks. Moreover, We propose a new generative model of variational autoencoder (VAE) based architecture that captures temporal relation knowledge for cross-user HAR, which enhances the generalized capability of VAE generative model. Our extensive testing on two public HAR datasets shows that this method outperforms existing approaches in time series domain adaptation, particularly in cross-user scenarios.

The paper is organized as follows: Section 2 reviews related work in HAR, transfer learning, and domain adaptation. Section 3 details our proposed CVAE-USM method, focusing on capturing and aligning temporal relations. Section 4 outlines our experimental setup and compares our method against existing approaches. Finally, Section 5 concludes the paper and suggests directions for future research.

2 Related work

2.1 Cross-user human activity recognition

Human Activity Recognition (HAR) is a fundamental aspect of ubiquitous computing, playing a vital role in enhancing daily human activities. Its primary goal is to recognize and analyze human behaviors by interpreting high-level knowledge derived from multi-modal sensor data and contextual information. Based on sensor modalities, HAR can be categorized into five distinct types: Smartphones/wearable sensors-based HAR, ambient sensors-based HAR, device-free sensors-based HAR, vision-based sensors-based HAR, and other modality sensors-based HAR [Chen *et al.*, 2021][Lentzas and Vrakas, 2020]. This paper specifically delves into HAR using wearable sensors.

Within machine learning research, sensor-based HAR is often approached as a time series classification problem [Andreas *et al.*, 2014]. Various classification models, including ensemble learning [Sekiguchi *et al.*, 2020], SVM [Bulling *et al.*, 2012], and HMM [Amft *et al.*, 2005], have been proposed to solve HAR challenges. With advancements in deep learning, numerous state-of-the-art results have been achieved for a wide range of tasks. Deep learning-based HAR techniques excel in learning high-level features and autonomously extracting features from extensive data sets [Qian *et al.*, 2019]. However, these approaches predominantly rely on the assumption that the training and testing data are drawn from the same distribution, meaning the data is considered independent and identically distributed (i.i.d.) [Wilson and Cook, 2020]. This assumption often fails in real-world applications, where the collected training and testing datasets are out-of-distribution (o.o.d.). Our focus in this paper is on addressing the challenges associated with sensor-based HAR in the context of o.o.d. data.

There are several categories of sensor-based HAR o.o.d. challenges: First, differences in data may result from the use of various sensor types, platforms, manufacturers, and modalities, leading to diverse data formats and distributions [Xing *et al.*, 2018]. Second, the data pattern might change over time, a phenomenon known as concept drift [Lu *et al.*, 2018]. For example, a person's walking pattern could be influenced by changes in health status. Third, behavioral differences between individuals can be significant [Ye and Wang, 2023]; for instance, walking pace may vary from person to person. Fourth, the positioning of physical sensors on the body [Rokni and Ghasemzadeh, 2018] or the layout of environmental sensors in smart homes [Sukhija and Krishnan, 2019] can also lead to different data distributions. Our research specifically targets sensor-based HAR o.o.d. problems stemming from behavioral differences among individuals.

2.2 Transfer learning and domain adaptation

Transfer learning enables models to learn from one or more source domains and then apply this knowledge to related target domains lacking labelled data. This process is crucial to address the out-of-distribution (o.o.d.) problem by minimizing distribution differences between the source and target domains. Domain adaptation, a subset of transfer learning, specifically tackles the o.o.d. issue while maintaining the same task across source and target domains. This approach typically involves leveraging labelled data from the source domain and unlabeled data from the target domain [Wilson and Cook, 2020].

Domain adaptation has experienced significant growth, particularly in the area of feature-based transfer learning [Pan and Yang, 2009]. Techniques like Subspace Alignment (SA) [Fernando *et al.*, 2013] focus on finding similarities between the feature subspaces of the source and target domains, using principal component analysis and linear transformations. Optimal Transport for Domain Adaptation (OTDA) [Flamary *et al.*, 2016], is grounded in optimal transport theory, aiming to establish a cost-effective correspondence between different domains. Deep domain adaptation has also seen advancements with approaches like the Domain Adversarial Neural Network (DANN) [Ganin *et al.*, 2016], which employs adversarial learning to train a model so that its features cannot be used to distinguish between source and target domains. This promotes the generation of domain-invariant features. FedMAT [Shen *et al.*, 2022] works by treating each individual's data as a distinct task within a federated learning system. It uniquely combines a central shared feature representation network with individual-specific networks equipped with attention modules in decentralized nodes, enabling the learning of both shared and individual-specific features from multi-modal sensor data.

These feature-based methods primarily target static data like images, often applying the same framework to time series data [Lu *et al.*, 2021][Li *et al.*, 2020]. However, this approach may be less effective for time series data, where temporal relationships are crucial. Current domain adaptation strategies often overlook these temporal relations, leading to models that may not perform reliably, especially in cross-user HAR applications where models trained on data from one user are applied to others. Incorporating temporal relation knowledge could enhance the identification of commonalities between users, improving domain adaptation in sensor-based HAR. This paper focuses on exploring and integrating temporal relation knowledge to enhance domain adaptation effectiveness in cross-user HAR.

3 Method

3.1 Problem formulation

In a cross-user HAR problem, a labelled source user $S^{Source} = \left\{ \left(x_i^{Source}, y_i^{Source} \right) \right\}_{i=1}^{n^{Source}}$ drawn from a joint probability distribution P^{Source} and a target user $S^{Target} = \left\{ \left(x_i^{Target}, y_i^{Target} \right) \right\}_{i=1}^{n^{Target}}$ drawn from a joint probability distribution P^{Target}, where n^{Source} and n^{Target} are the

number of source and target samples respectively. S^{Source} and S^{Target} have the same feature spaces (i.e. the set of features that describes the data from sensor readings) and label spaces (i.e. the set of activity classes). The source and target users have different distributions, i.e., $P^{Source} \neq P^{Target}$, which means that even for the same activity, the sensor readings look different between the two users. Given the labelled source user data and unlabelled target user data, the goal is to obtain the labels $\left\{ \left(y_i^{Target} \right) \right\}_{i=1}^{n^{Target}}$ for the target user activities.

3.2 Conditional Variational Autoencoder with Universal Sequence Mapping

In this study, we present the Conditional Variational Autoencoder with Universal Sequence Mapping (CVAE-USM), a novel approach for cross-user HAR. The foundation of this method is the understanding that human physical movements are influenced by preceding actions, and these temporal dependencies are generally consistent across different individuals. By capturing these shared temporal patterns in human activity time series data, our model aims to enhance its performance and generalizability for cross-user HAR tasks.
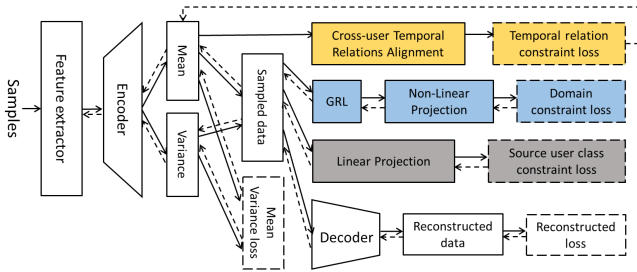

Figure 1: CVAE-USM Learning Process.

Generative Model Architecture

Our approach utilizes a generative model framework, as depicted in the CVAE-USM Learning Process shown in Figure 1, to achieve effective model generalization. Generative models are advantageous in that they can produce new data samples that are similar to the training data, aiding in data augmentation and improving generalization capabilities. Even in scenarios where some class labels are missing during training, generative models focus on modelling the data distribution itself, which is more effective in uncovering the underlying data patterns and structures [Ruthotto and Haber, 2021]. For cross-user HAR, where labels for target users are not available (or only partially available), generative models are particularly useful as they are capable of modelling user-invariant activity data distributions. They achieve this by encoding data into a compact latent space, capturing the underlying factors or patterns in the data [Van Kasteren *et al.*, 2010].

In implementing our method, we employ the Variational Autoencoder (VAE) architecture [Kingma and Welling, 2014] as the foundational generative model. The VAE aims to learn a structured latent space where similar data samples are closely aligned, such as patterns found in time series data, typically following a Gaussian distribution. In the VAE structure, input data (like sensor features in HAR) is compressed into a lower-dimensional latent space via an encoder. This latent space is represented probabilistically with mean and variance parameters, encapsulating intrinsic characteristics of human activities. The VAE employs the reparameterization trick for differentiability during training, sampling from the latent space based on Gaussian distributions and then using a decoder to reconstruct the original input. This method enables the capturing of complex temporal activity patterns, making the VAE an ideal model for cross-user HAR tasks.

The VAE's sampling process offers several benefits for generalization and domain adaptation. The randomness introduced through sampling acts as a regularization form, preventing overfitting and enhancing generalization. By encoding data into a distribution instead of a deterministic decision boundry and encouraging the latent variables to approximate a Gaussian distribution (via the Kullback–Leibler divergence), the latent space becomes more structured and continuous [Rybkin *et al.*, 2021]. In such a structured latent space, alignment techniques like Gradient Reversal Layer (GRL) adversarial learning, which we will discuss later, can be more effectively applied for domain adaptation in HAR.

Moreover, we implement an adversarial learning strategy, specifically utilizing the Gradient Reversal Layer (GRL) technique [Ganin *et al.*, 2016] as the blue parts in Figure 1. This strategy is integral to modifying the model's loss functions related to domain constraints (i.e. confusing the data distributions between source and target users). The GRL adversarial learning operates by reversing the gradients during the training phase. The primary goal of this technique is to ensure that the features learned by the model are not solely tied to the specific user but also lean towards a generalized model. Through this approach, our model can generate user-invariant representations. Therefore, this strategy is integrated as a constraint condition to the original CVAE for model learning.

Furthermore, after acquiring the sampled features, they are input into a process focusing on the source user's activity class constraint as the grey parts in Figure 1. This involves using a linear transformation to shape how the data distribution is generated. Specifically, this step takes into consideration the activity class labels associated with the source user. This means that the transformation is guided by the types of activities the source user performed, helping to tailor the model's understanding of the data based on these specific activities. Therefore, this source user class constraint is integrated as another constraint condition to the original CVAE for model learning.

Cross-user Temporal Relations Alignment

The key element of the CVAE-USM method is the cross-user temporal relations alignment. This CVAE constraint condition is shown in the yellow parts in Figure 1. In this component, Gaussian Mixture Model (GMM), Universal Sequence Mapping (USM), and Wasserstein Distance techniques are included. Each component plays a pivotal role in capturing and adapting temporal relation knowledge for domain adaptation. The overall process of cross-user temporal relations

alignment is described in Algorithm 1.

Algorithm 1 Cross-user Temporal Relations Alignment

Input: Mean values from CVAE encoder for source (μ_s) and target (μ_t) users, number of components (K) in GMM.
Output: Aligned temporal relation features, Wasserstein distance (W) temporal relation loss.
Begin Algorithm
1. Sub-Activity Identification with GMM:
- Apply GMM to sensor data for source and target users: $GMM_{init}(\{\mu_s, \mu_t\}, K)$.
- This step is to identify and decompose complex activities into simpler sub-activities.

2. Temporal Relations Encoding with USM:
- Transform sub-activities into feature space using USM: $F_s = \mathcal{U}(Seq_s)$, $F_t = \mathcal{U}(Seq_t)$.
- This step aims to capture and encode temporal relation for activities. Here, $\mathcal{U}$ denotes the USM function, and Seq_s, Seq_t are sequences of encoded sub-activities for source and target, respectively, with each value in the sequences encoded into the set $\{0, K{-}1\}$.

3. User-Specific Temporal Feature Analysis:
- Perform separate GMM clustering on the USM-encoded features for source (F_s) and target (F_t) users: $GMM_s(F_s, K)$ and $GMM_t(F_t, K)$.
- This step is to analyze the variations and similarities between the sub-activities' data distributions across users. Let P_s and P_t denote the resulting distributions from GMM_s and GMM_t respectively, capturing distinct temporal patterns of each user group.

4. Data Distribution Alignment with Wasserstein Distance:
- Minimize the Wasserstein distance between source and target distributions: $W(P_s, P_t) = \inf_{\gamma \in \Pi(P_s, P_t)} \mathbb{E}(x, y) \sim \gamma[\|x - y\|]$.
- This step aims to adjust model parameters to align Ps and P_t, focusing on common temporal relations.

End Algorithm

The first step is the Gaussian Mixture Model (GMM) for sub-activity identification. GMM [Zhang *et al.*, 2021] is a probabilistic model for representing normally distributed sub-populations within an overall population. It is defined as:

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{1}$$

Where x is the data point. K is the number of Gaussian distributions in the mixture. π_k are the mixture weights, with $\sum_{k=1}^{K} \pi_k = 1$ and $\pi_k \geq 0$. $\mathcal{N}(x|\mu_k, \Sigma_k)$ denotes the Gaussian distributions with mean μ_k and covariance Σ_k.

GMM identifies sub-activities hidden in the sensor data, effectively decomposing complex activities into simpler sub-activities which represent common knowledge across different users. This model identifies inherent sub-structures within the dataset, irrespective of their originating domain (source or target). This decomposition is vital for two reasons: firstly, it simplifies the complex temporal patterns into more manageable segments, and secondly, the common sub-activities are used for subsequent temporal relation capture and analysis between the source and target domains.

Following the identification of sub-activities, the next step is using USM for temporal feature encoding. This step is crucial for transforming the identified sub-activities into a feature space that encapsulates both their spatial structural and temporal characteristics. USM [Almeida and Vinga, 2002] is a technique for encoding sequences to capture their inherent structure and temporal relations. It can be mathematically represented as:

$$F = \mathcal{U}(Seq) \tag{2}$$

Where F represents the feature space encoding the sequence. $\mathcal{U}$ is the USM function. Seq is the sequence of sub-activities.

USM plays a pivotal role in maintaining the integrity of temporal relations within the activity sequences. This encoding not only represents the sequence of sub-activities but also preserves the temporal dynamics associated with each activity – that is, the temporal order and sequence of sub-activities, particularly when analyzing complex human behaviours, thus offering a more precise representation of human activities. This encoding respects the chronological order of events, ensuring that the temporal dynamics—such as the duration, sequence, and intervals between sub-activities—are accurately captured and reflected in the feature vectors. This is particularly important for understanding how sub-activities unfold over time.

By effectively encoding the temporal order, USM enables the identification of recurrent temporal patterns and relationships within the data. This can include regular sequences of activities, the typical duration of certain activities, or the common intervals between sub-activities. Moreover, USM-encoded features facilitate the comparison of temporal sequences across different users. By providing a uniform method of encoding, USM allows for the direct comparison of temporal features, supporting the data distribution alignment of temporal relations across users. This step paves the way for the following temporal relation alignment across users.

The next step is user-specific feature analysis. The USM-encoded features, now rich with temporal relational information, undergo separate GMM clustering for the source and target domains. This bifurcation is key to understanding how temporal patterns manifest uniquely in each domain. By analyzing these user-specific GMMs, the approach gains insights into the variations and similarities in activity patterns across users. This comparative analysis is essential for identifying user-specific differences in temporal activity structures.

Finally, Wasserstein Distance is applied for data distribution alignment across users with the above-learned temporal relation knowledge. The Wasserstein Distance [Wu *et al.*, 2020], also known as the Earth Mover's Distance, measures the distance between two probability distributions. It is defined as:

$$W(P_s, P_t) = \inf_{\gamma \in \Pi(P_s, P_t)} \mathbb{E}_{(x,y) \sim \gamma}[\|x - y\|] \tag{3}$$

Where $W(P_s, P_t)$ is the Wasserstein Distance between two distributions. P_s and P_t are the probability distributions of source and target users being compared. γ represents a joint distribution with marginals P_s and P_t. $\Pi(P_s, P_t)$ denotes the set of all possible joint distributions $\gamma(x, y)$ with marginals P_s and P_t. $\mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|]$ is the expected value of the Euclidean distance between sampled points x and y from γ.

Wasserstein distance is computed using the means of the GMMs from the source and target users. This distance metric quantifies the disparity in the temporal feature distributions between the two domains. Minimizing the Wasserstein distance is used to align the source and target user distributions in the temporal relation view based on the idea that source and target users should have common temporal relation sub-activities. This alignment is central to the model's ability to generalize and accurately recognize activities across users.

4 Experiments

4.1 Datasets and Experimental setup

For validating our cross-user domain adaptation method in sensor-based HAR, we employed two widely recognized public datasets as outlined in Table 1. Our focus was on the practical application scenario involving a smartwatch. To this end, we only utilized sensor data from accelerometers and gyroscopes positioned on the right lower arm. Below, we provide a brief overview of each dataset, highlighting their key aspects.

Table 1: Two sensor-based HAR datasets information.

Dataset	Subjects	#Activities	Common Activities
OPPT	S1, S2, S3	4	1 standing, 2 walking, 3 sitting, 4 lying
PAMAP2	1, 5, 6	11	1 lying, 2 sitting, 3 standing, 4 walking, 5 running, 6 cycling, 7 Nordic walking, 8 ascending stairs, 9 descending stairs, 10 vacuum cleaning, 11 ironing

The OPPORTUNITY (OPPT) dataset [Chavarriaga *et al.*, 2013] contains recordings of people doing morning activities in a natural way, without strict guidance, recorded at 30 Hz. The Physical Activity Monitoring (PAMAP2) dataset [Reiss and Stricker, 2012] features over 10 hours of data recorded at 100 Hz, with subjects performing specific, pre-defined activities.

In the experimental setup for cross-user HAR, the sliding window technique is employed for data segmentation. This widely-used approach in sensor-based HAR involves creating fixed time intervals of 3 seconds with a 50% overlap between windows, in line with standard practices in HAR tasks [Wang *et al.*, 2018]. The study compares four methods, divided into traditional domain adaptation and deep domain adaptation categories. Notably, the TrC method requires a few target domain labels for fine-tuning, unlike other methods which do not require any labelling in the target domain.

Traditional Domain Adaptation Methods:

CORAL (Covariance Alignment) [Sun *et al.*, 2016]: Aims to align the covariance of feature layers across domains, fostering domain-agnostic features.

SOT (Substructure Optimal Transport) [Lu *et al.*, 2021]: Focuses on underlying domain structures to facilitate detailed substructure mapping, balancing broad and specific mapping.

Deep Domain Adaptation Methods:

DANN (Domain-Adversarial Neural Network) [Ganin *et al.*, 2016]: A deep learning approach using adversarial training to produce features indistinguishable between domains by a discriminator.

TrC (Temporal Regularized CNNs) [Rokni *et al.*, 2018]: Employs CNNs for discerning temporal features, trained on source domain data and fine-tuned on the target domain.

Hyper-parameters tuning is implemented on these methods with the methodology suggested by Flamary et al. [Flamary *et al.*, 2016], specifically to prevent overfitting during testing. We divide the target user's data into two parts: a validation subset and a test subset. The validation subset is utilized to fine-tune the hyper-parameters to achieve the highest accuracy. After optimizing these hyper-parameters, we evaluate the performance of the model on the test subset. The key metric we use for evaluation is the classification accuracy specific to the target user.

4.2 Cross-user HAR performance results

We evaluate the performance of various methods on the OPPT and PAMAP2 datasets. Each method is tested for its ability to transition between individual users. Within each dataset, three users are selected randomly, and a one-to-one cross-user HAR task is conducted for every possible user pair.
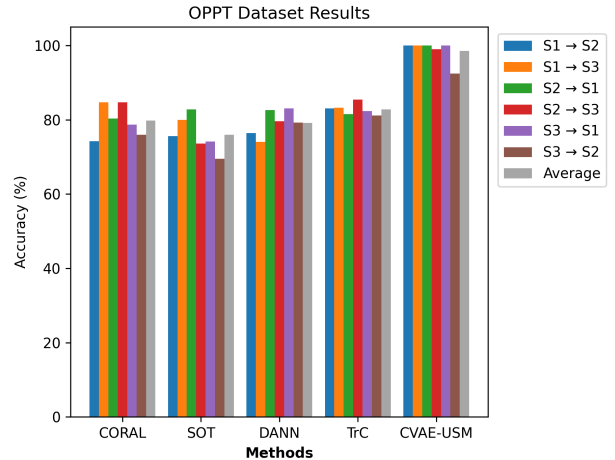

Figure 2: OPPT dataset classification results.

In the results of OPPT dataset as shown in Figure 2, CVAE-USM stands out by achieving almost 100% accuracy in all testing scenarios. This exceptional performance underscores CVAE-USM's adaptability and effectiveness in various conditions specific to the OPPT dataset. TrC also achieves a reasonably good performance, consistently reaching accuracy levels in the low to mid-80% range. This indicates its competence in the OPPT dataset, though it doesn't match the superior performance of CVAE-USM. Another method DANN demonstrates moderate performance, with accuracy generally fluctuating between the mid-70% and mid-80% range.

SOT and CORAL exhibit slightly lower performance levels compared to the others. One notable distinction contributing to the performance discrepancies is the utilization of temporal information. CVAE-USM effectively leverages temporal data, contributing significantly to its superior accuracy. In contrast, TrC considers the temporal features , however it does not exploit temporal information as effectively as CVAE-USM, which partially explains its lower performance metrics.
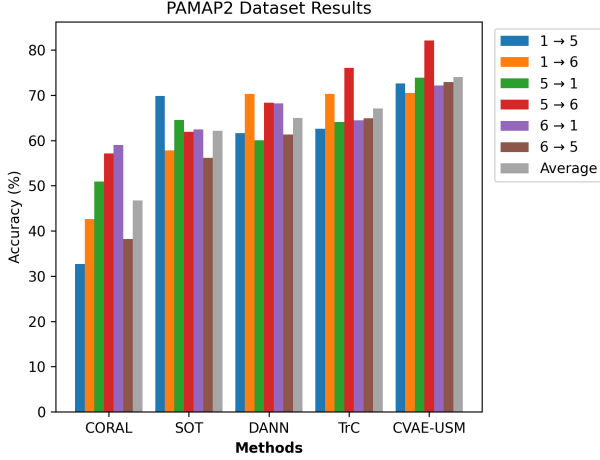

Figure 3: PAMAP2 dataset classification results.

In the results of PAMAP2 dataset as shown in Figure 3, CVAE-USM again dominates, with accuracy consistently above 70% and peaking at 82.12% in certain scenarios. This dominance indicates CVAE-USM's strong adaptability and learning capabilities across different conditions in the PAMAP2 dataset. TrC, SOT and DANN show respectable performances in specific scenarios, suggesting that they may be better adapted to situations with minimal variance between training and testing data or where data patterns are more consistent and less complex. In contrast, CORAL struggles significantly, with their accuracies mainly ranging between 30%-40%. This poor performance reflects their limited adaptability to the PAMAP2 dataset. A critical aspect contributing to these varied performances is the utilization of temporal relationships within the data. CVAE-USM's effective use of temporal information likely contributes to its outstanding accuracy, as it can capture and learn from the dynamics and temporal patterns inherent in the PAMAP2 dataset. In contrast, models like SOT and DANN do not leverage temporal data, which could explain their not outstanding results.

Overall, across both the OPPT and PAMAP2 datasets, CVAE-USM consistently achieves the highest accuracy, proving its adaptability, robustness, and effectiveness in handling cross-user HAR tasks. TrC shows a reasonable performance across both datasets, indicating its potential adaptability to different scenarios, albeit not at the level of CVAE-USM. CORAL, SOT, and DANN exhibit varied and mixed results, suggesting that their effectiveness and adaptability are

conditional based on specific dataset characteristics and scenarios. Some methods, particularly CORAL, show significant struggles in adapting to these datasets, highlighting the need for further development to enhance their effectiveness in time series data that involves temporal relation knowledge.

4.3 Effect of temporal information

In this section, we analyze the impact of integrating temporal relation knowledge into cross-user HAR, concentrating on the OPPT and PAMAP2 datasets. The comparison involves SOT, a traditional domain adaptation method; DANN, known for its deep domain adaptation approach; and our CVAE-USM, which is specifically designed for time series domain adaptation.

Figure 4 displays the confusion matrices for CVAE-USM, SOT, and DANN, reflecting their average performance across the OPPT and PAMAP2 datasets. The matrices use color gradients, with warmer colors (closer to yellow) indicating higher values. The x-axis represents predicted activities, while the y-axis corresponds to the actual activities. The diagonal line from the top left to the bottom right of these matrices signifies correct predictions, where the predicted activity aligns with the true activity.
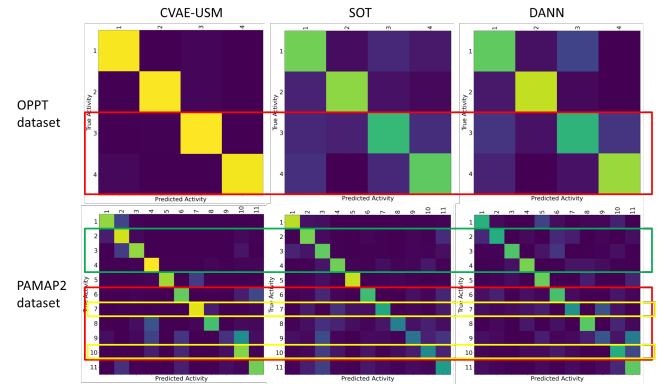

Figure 4: Confusion matrices of CVAE-USM, SOT, and DANN methods for the average performance of tasks in the OPPT and PAMAP2 datasets.

In these analyses, all methods exhibit pronounced values along their diagonals, indicating accurate recognition of most activities and underscoring their effectiveness in cross-user scenarios. Common activities such as 'lying', 'walking', 'standing' and 'cycling' are efficiently identified by all methods. The reason could be attributed to their distinct movement patterns, making them easily distinguishable from other activities. However, confusion arises in differentiating activities with inherent similarities, such as activities like 'ascending stairs' and 'descending stairs'.

CVAE-USM, in particular, excels in identifying activities with complex temporal relations like 'Nordic walking' and 'vacuum cleaning' as the yellow squares shown in Figure 4, with higher accuracy compared to SOT and DANN. Additionally, while SOT and DANN show dispersed classification outcomes, CVAE-USM exhibits a more focused and clear classification pattern due to its effective handling of tempo-

ral states and extraction of sub-activities as the red squares shown in Figure 4. CVAE-USM demonstrates a superior ability to distinguish activities like 'standing', 'sitting', and 'walking' as the green squares shown in Figure 4, possibly due to its nuanced capture of temporal relation context and subtle sensor data variations. Moreover, transitional movements within activities, often overlooked by SOT and DANN, are effectively captured by CVAE-USM. For example, repositioning a vacuum cleaner during 'vacuum cleaning' is better understood and classified by CVAE-USM, thanks to its comprehension of temporal relations.

In conclusion, the use of temporal relation knowledge in CVAE-USM effectively reduces the variance in activity patterns between different users, especially in dynamic activities with complex temporal relations. Extracting temporal relation knowledge from time series data enhances the robustness and accuracy of activity recognition. This approach is especially beneficial for activities with rhythmic or sequential traits, allowing for a more nuanced understanding of various activities and highlighting the importance of temporal knowledge in improving cross-user activity recognition capabilities.

5   Conclusion

In this study, we present a new approach called CVAE-USM for cross-user HAR tasks from time series data. This method is particularly effective in adapting to different users' data by focusing on the temporal relation knowledge of activities. CVAE-USM uses VAE generative model and adversarial learning techniques to align different users' data distributions, leading to more accurate activity recognition.

The validation is implemented on two public datasets show that CVAE-USM outperforms existing methods in recognizing activities across various users. This success demonstrates CVAE-USM's potential to analyse time series data in HAR. Overall, CVAE-USM represents its advantages in understanding temporal relations and adapting to the complex activity data from different users. In future, we aim to explore the application of CVAE-USM to more complex or less structured activities, to evaluate its adaptability and robustness.

References

[Abdallah *et al.*, 2018] Zahraa S Abdallah, Mohamed Medhat Gaber, Bala Srinivasan, and Shonali Krishnaswamy. Activity recognition with evolving data streams: A review. *ACM Computing Surveys (CSUR)*, 51(4):1–36, 2018.

[Almeida and Vinga, 2002] Jonas S Almeida and Susana Vinga. Universal sequence map (usm) of arbitrary discrete sequences. *BMC bioinformatics*, 3:1–11, 2002.

[Amft *et al.*, 2005] Oliver Amft, Holger Junker, and Gerhard Troster. Detection of eating and drinking arm gestures using inertial body-worn sensors. In *Ninth IEEE international symposium on wearable computers (ISWC'05)*, pages 160–163. IEEE, 2005.

[Andreas *et al.*, 2014] Bulling Andreas, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):33, 2014.

[Bulling *et al.*, 2012] Andreas Bulling, Jamie A Ward, and Hans Gellersen. Multimodal recognition of reading activity in transit using body-worn sensors. *ACM Transactions on Applied Perception (TAP)*, 9(1):1–21, 2012.

[Chavarriaga *et al.*, 2013] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R Millán, and Daniel Roggen. The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters*, 34(15):2033–2042, 2013.

[Chen *et al.*, 2021] Kaixuan Chen, Dalin Zhang, Lina Yao, Bin Guo, Zhiwen Yu, and Yunhao Liu. Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. *ACM Computing Surveys (CSUR)*, 54(4):1–40, 2021.

[Fernando *et al.*, 2013] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE international conference on computer vision*, pages 2960–2967, 2013.

[Flamary *et al.*, 2016] R Flamary, N Courty, D Tuia, and A Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell*, 1:1–40, 2016.

[Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.

[Kingma and Welling, 2014] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[Lentzas and Vrakas, 2020] Athanasios Lentzas and Dimitris Vrakas. Non-intrusive human activity recognition and abnormal behavior detection on elderly people: A review. *Artificial Intelligence Review*, 53(3):1975–2021, 2020.

[Li *et al.*, 2020] Shuang Li, Binhui Xie, Jiashu Wu, Ying Zhao, Chi Harold Liu, and Zhengming Ding. Simultaneous semantic alignment network for heterogeneous domain adaptation. In *Proceedings of the 28th ACM international conference on multimedia*, pages 3866–3874, 2020.

[Lu *et al.*, 2018] Jie Lu, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, and Guangquan Zhang. Learning under concept drift: A review. *IEEE transactions on knowledge and data engineering*, 31(12):2346–2363, 2018.

[Lu *et al.*, 2021] Wang Lu, Yiqiang Chen, Jindong Wang, and Xin Qin. Cross-domain activity recognition via substructural optimal transport. *Neurocomputing*, 454:65–75, 2021.

[Ma *et al.*, 2019] Haojie Ma, Wenzhong Li, Xiao Zhang, Songcheng Gao, and Sanglu Lu. Attnsense: Multi-level attention mechanism for multimodal human activity recognition. In *IJCAI*, pages 3109–3115, 2019.

[Pan and Yang, 2009] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[Patel *et al.*, 2015] Vishal M Patel, Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Visual domain adaptation: A survey of recent advances. *IEEE signal processing magazine*, 32(3):53–69, 2015.

[Qian *et al.*, 2019] Hangwei Qian, Sinno Jialin Pan, Bingshui Da, and Chunyan Miao. A novel distribution-embedded neural network for sensor-based activity recognition. In *IJCAI*, volume 2019, pages 5614–5620, 2019.

[Reiss and Stricker, 2012] Attila Reiss and Didier Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th international symposium on wearable computers*, pages 108–109. IEEE, 2012.

[Rokni and Ghasemzadeh, 2018] Seyed Ali Rokni and Hassan Ghasemzadeh. Autonomous training of activity recognition algorithms in mobile sensors: A transfer learning approach in context-invariant views. *IEEE Transactions on Mobile Computing*, 17(8):1764–1777, 2018.

[Rokni *et al.*, 2018] Seyed Ali Rokni, Marjan Nourollahi, and Hassan Ghasemzadeh. Personalized human activity recognition using convolutional neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[Ruthotto and Haber, 2021] Lars Ruthotto and Eldad Haber. An introduction to deep generative modeling. *GAMM-Mitteilungen*, 44(2):e202100008, 2021.

[Rybkin *et al.*, 2021] Oleh Rybkin, Kostas Daniilidis, and Sergey Levine. Simple and effective vae training with calibrated decoders. In *International Conference on Machine Learning*, pages 9179–9189. PMLR, 2021.

[Sekiguchi *et al.*, 2020] Ryoichi Sekiguchi, Kenji Abe, Takumi Yokoyama, Masayasu Kumano, and Masaki Kawakatsu. Ensemble learning for human activity recognition. In *Adjunct proceedings of the 2020 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2020 ACM international symposium on wearable computers*, pages 335–339, 2020.

[Shen *et al.*, 2022] Qiang Shen, Haotian Feng, Rui Song, Stefano Teso, Fausto Giunchiglia, Hao Xu, et al. Federated multi-task attention for cross-individual human activity recognition. In *IJCAI*, pages 3423–3429. IJCAI, 2022.

[Soleimani and Nazerfard, 2021] Elnaz Soleimani and Ehsan Nazerfard. Cross-subject transfer learning in human activity recognition systems using generative adversarial networks. *Neurocomputing*, 426:26–34, 2021.

[Sukhija and Krishnan, 2019] Sanatan Sukhija and Narayanan C Krishnan. Supervised heterogeneous feature transfer via random forests. *Artificial Intelligence*, 268:30–53, 2019.

[Sun *et al.*, 2016] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

[Van Kasteren *et al.*, 2010] TLM Van Kasteren, Gwenn Englebienne, and Ben JA Kröse. An activity monitoring system for elderly care using generative and discriminative models. *Personal and ubiquitous computing*, 14:489–498, 2010.

[Wang *et al.*, 2018] Gaojing Wang, Qingquan Li, Lei Wang, Wei Wang, Mengqi Wu, and Tao Liu. Impact of sliding window length in indoor human motion modes and pose pattern recognition based on smartphone sensors. *Sensors*, 18(6):1965, 2018.

[Wilson and Cook, 2020] Garrett Wilson and Diane J Cook. A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5):1–46, 2020.

[Wu *et al.*, 2020] Hanrui Wu, Yuguang Yan, Michael K Ng, and Qingyao Wu. Domain-attention conditional wasserstein distance for multi-source domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(4):1–19, 2020.

[Xing *et al.*, 2018] Tianwei Xing, Sandeep Singh Sandha, Bharathan Balaji, Supriyo Chakraborty, and Mani Srivastava. Enabling edge devices that learn from each other: Cross modal training for activity recognition. In *Proceedings of the 1st International Workshop on Edge Systems, Analytics and Networking*, pages 37–42, 2018.

[Ye and Wang, 2023] Xiaozhou Ye and Kevin I-Kai Wang. Cross-user activity recognition via temporal relation optimal transport. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, 2023.

[Zhang *et al.*, 2021] Yi Zhang, Miaomiao Li, Siwei Wang, Sisi Dai, Lei Luo, En Zhu, Huiying Xu, Xinzhong Zhu, Chaoyun Yao, and Haoran Zhou. Gaussian mixture model clustering with incomplete data. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 17(1s):1–14, 2021.