



SHADE-AD: An LLM-Based Framework for Synthesizing Activity Data of Alzheimer's Patients

Heming Fu^{1,2}, Hongkai Chen², Shan Lin¹✉, Guoliang Xing²✉

¹Stony Brook University, ²The Chinese University of Hong Kong
 {heming.fu, shan.x.lin}@stonybrook.edu, {hkchen, glxing}@ie.cuhk.edu.hk

ABSTRACT

Alzheimer's Disease (AD) has become an increasingly critical global health concern, which necessitates effective monitoring solutions in smart health applications. However, the development of such solutions is significantly hindered by the scarcity of AD-specific activity datasets. To address this challenge, we propose SHADE-AD, a Large Language Model (LLM) framework for **Synthesizing Human Activity Datasets Embedded with AD features**. Leveraging both public datasets and our own collected data from 99 AD patients, SHADE-AD synthesizes human activity videos that specifically represent AD-related behaviors. By employing a three-stage training mechanism, it broadens the range of activities beyond those collected from limited deployment settings. We conducted comprehensive evaluations of the generated dataset, demonstrating significant improvements in downstream tasks such as Human Activity Recognition (HAR) detection, with enhancements of up to 79.69%. Detailed motion metrics between real and synthetic data show strong alignment, validating the realism and utility of the synthesized dataset. These results underscore SHADE-AD's potential to advance smart health applications by providing a cost-effective, privacy-preserving solution for AD monitoring.

CCS CONCEPTS

- Human-centered computing → Ubiquitous and mobile computing;
- Computing methodologies → Machine learning.

KEYWORDS

Large Language Model (LLM), Synthesis dataset, Alzheimer's Disease (AD), Human Action dataset.

ACM Reference Format:

Heming Fu^{1,2}, Hongkai Chen², Shan Lin¹✉, Guoliang Xing²✉. 2025. SHADE-AD: An LLM-Based Framework for Synthesizing Activity Data of Alzheimer's Patients. In *The 23rd ACM Conference on Embedded Networked Sensor Systems (SenSys '25)*, May 6–9, 2025, Irvine, CA, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3715014.3722062>

1 INTRODUCTION

Alzheimer's Disease (AD) is a rapidly escalating global health crisis, affecting approximately 6.9 million Americans age 65 and older in

✉Co-corresponding authors.

2024, representing about 10.9% of that age group [1]. This number is projected to grow to 13.8 million by 2060, highlighting an urgent need for effective monitoring and early intervention strategies [16].

Smart health technologies, powered by artificial intelligence (AI) and sensor systems, offer promising solutions by enabling continuous, non-invasive monitoring of patients' daily activities [22]. These technologies are crucial for facilitating early diagnosis and personalized care, which can significantly improve patient outcomes and reduce healthcare costs [15, 17, 24, 32]. However, the effectiveness of AI models in these applications depends heavily on the availability of high-quality, domain-specific datasets for training [9, 19]. Without such datasets, AI models may fail to accurately recognize AD-specific behaviors, leading to delayed diagnoses, inadequate patient monitoring, and potential safety risks [27, 35].

Despite the prevalence of Alzheimer's Disease, AD-specific behavioral datasets are very limited. Existing datasets often lack the necessary domain-specific features to capture the nuanced behaviors associated with AD patients, such as subtle changes in gait, posture, and movement patterns indicative of cognitive decline [4]. Using generative models to augment the AD behavior dataset presents a potential solution to address the aforementioned limitation [11, 33, 39]. However, though existing synthetic data generation models, such as those leveraging vision-language models like BERT [3], Stable Diffusion [26], and MotionDiffuse [34] have demonstrated strong capabilities in understanding and generating general visual content [14, 18], they are primarily trained on data from healthy individuals and do not capture the specific motor and behavioral anomalies seen in AD patients [21]. For example, these models may generate generic walking patterns but fail to reproduce the shuffling gait or balance issues common in AD. This limitation hinders the development of effective AI models for AD monitoring, as models trained on general datasets may not accurately detect or interpret AD-specific behaviors.

A further challenge lies in validating disease-specific synthetic data, particularly due to the complexity of patient-specific physical characteristics. In the case of AD, there is currently no established model or standardized metrics capable of accurately capturing the wide range of behaviors associated with cognitive and motor decline [10, 16]. Although several studies on gait analysis provide valuable insights [38], they predominantly focus on lower-body movements and lack a comprehensive approach to the subtle, full-body action patterns that characterize AD.

Figure 1 shows that AD patients exhibit oscillatory movements when standing up and maintain a stooped posture afterward, unlike normal elders and other motion generative models, such as MotionDiffuse [34] (prompted with "an Alzheimer's Disease patient standing up from a chair"), which generate a smooth transition to an upright position, much like a healthy subject.



This work is licensed under a Creative Commons Attribution 4.0 International License.
SenSys '25, May 6–9, 2025, Irvine, CA, USA
 © 2025 Copyright held by the owner/authors.

ACM ISBN 979-8-4007-1479-5/25/05.
<https://doi.org/10.1145/3715014.3722062>

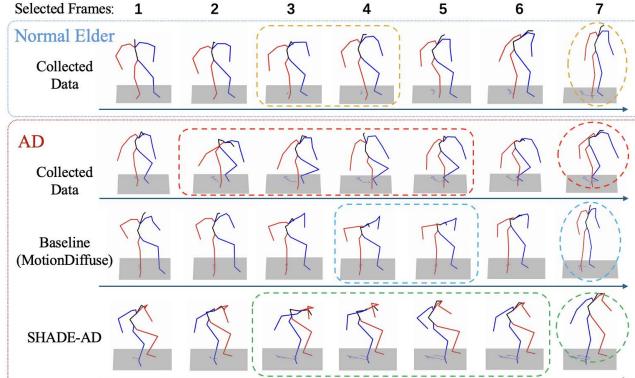


Figure 1: Comparison of “standing up from a chair” motion highlighting two key differences: (1) Standing transition (shown in rectangles): AD patients exhibit oscillation during the standing process (red), while Normal Elder (yellow) and Baseline (blue) both miss this characteristic and demonstrate smooth standing movement. SHADE-AD successfully reproduces this AD-characteristic oscillation in frames 3-6 (green). (2) Final posture (shown in circles): AD patients maintain a forward stooped posture after standing (red), whereas Normal Elder (yellow) and Baseline (blue) both miss this feature and achieve an upright stance. SHADE-AD accurately captures this forward stoop characteristic of AD patients (green).

To address these challenges, we develop SHADE-AD, a generative model to synthesize activity videos of AD patients. SHADE-AD embeds AD features through a three-stage process. The first two stages develop feature extraction abilities for AD terminology and action patterns. In the third stage, SHADE-AD incorporates motion metrics derived from 12 key joints, which validate the realism and clinical relevance of the synthetic dataset.

In summary, the key contributions of our work are:

- We propose a generative model, SHADE-AD, for synthetic 3D skeleton activity videos specific to AD patients, addressing the scarcity of AD behavior data.
- We design a framework that embeds AD-specific behavioral knowledge into a CLIP-based LLM encoder, enhancing the model’s ability to generate the nuanced actions of AD patients.
- We validate the effectiveness of the synthetic dataset generated by SHADE-AD using motion metrics of 12 key joints, and demonstrate its ability to significantly improve performance across multiple downstream tasks in smart health.

2 RELATED WORK

The development of effective AI models for smart health applications, particularly in AD scenarios, relies heavily on the availability of comprehensive behavioral datasets. This section reviews the existing literature on AD behavior datasets, data augmentation methods, and the use of LLMs for data generation, highlighting the gaps that this work aims to address. We will compare these methodologies with our approach in the preliminary results part.

2.1 Behavioral Datasets

Behavioral datasets are crucial in disease detection and patient monitoring applications, as they provide insights into patient activities

that can indicate the onset or progression of various conditions [22]. However, there is a significant lack of behavioral datasets that encompass a wide range of scenarios and activities for disease specific patient groups. Most existing datasets focus on clinical data such as neuroimaging, genetic markers, and cognitive assessments [37], while neglecting real-world behavioral data. Some initiatives have attempted to collect patient-related behavior through sensor data and video recordings [23], but these are limited in scope and often do not cover the diverse activities encountered in daily living. The scarcity of comprehensive behavioral datasets for specific patient populations hampers the development of AI models capable of accurately detecting and monitoring health conditions through patient behavior analysis, especially in AD scenarios.

2.2 Data Augmentation

Data augmentation is widely used to enhance the diversity of training data [28]. Traditional data augmentation methods for activity recognition include spatial transformations (e.g., rotation, scaling), temporal manipulations (e.g., time warping), and adding noise to existing data [30]. While these techniques can introduce variability, they are constrained by the inherent characteristics of the original data and do not add new semantic information. In applications requiring the detection of subtle and complex behaviors, traditional data augmentation techniques may be insufficient. These methods cannot generate new behavior patterns or capture the nuanced differences needed for accurate recognition. For instance, applying simple rotation or noise injection to a normal elder’s “standing up from a chair” action cannot realistically induce the repeated armrest supports or sustained stooping that real AD patients often exhibit (Figure 1). These nuanced shifts in posture or support patterns cannot be generated by basic transformations of healthy data. Consequently, models trained with traditionally augmented data may fail to generalize well to real-world scenarios where such subtle behaviors are present [31].

2.3 LLMs for Data Generation

Recent advancements in generative models, particularly those leveraging LLMs and vision-language models, have shown promise in generating synthetic data [5, 6, 13]. These models are trained on extensive datasets of common daily activities and general knowledge, enabling them to generate realistic content in these domains. For example, RFGenesis [7] proposes a cross-modal diffusion framework to generate mmWave sensing data and improve zero-shot generalization, while UniMTS [36] introduces a unified pre-training procedure for motion time series. These works highlight the growing trend of using generative models to enhance real-world data collection for Human Action Recognition (HAR). However, in disease domains like AD, these generative models perform inadequately due to a lack of domain-specific knowledge [23]. The actions associated with AD patients are often subtle, fine-grained, and context-dependent, making them hard to capture without specialized understanding.

Moreover, the inability to control the generated content at a fine-grained level poses challenges in tailoring the synthetic data to reflect the unique characteristics of AD-related actions. This gap underscores the need for generative models that can incorporate domain-specific knowledge and allow for precise control over the generated behaviors in smart health applications.

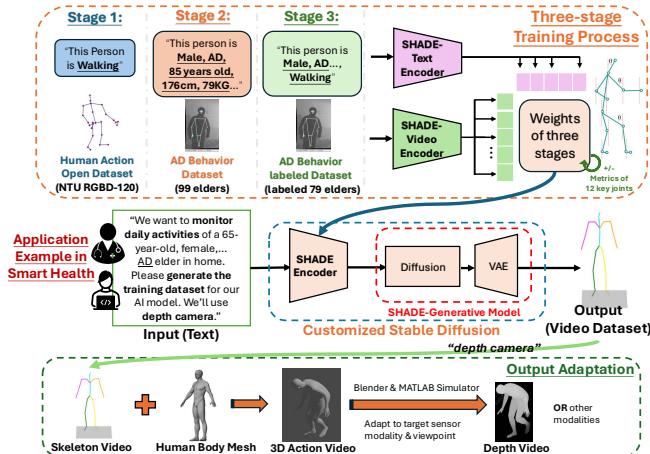


Figure 2: Design of SHADE-AD. The Training Process involves three stages: Stage 1 learns general human actions like “Walking”; Stage 2 embeds AD-specific knowledge like “AD” or “MoCA score”; and Stage 3 fine-tunes with patient-specific motion metrics derived from 12 key joints.

3 SYSTEM DESIGN

SHADE-AD integrates advanced components to model both general human activities and the subtle nuances of AD-specific behaviors. The framework and the operational workflow of SHADE-AD are illustrated in Figure 2, highlighting the interactions between its modules and the three-stage training procedure.

3.1 System Components

The SHADE-AD framework consists of three primary components: SHADE-Text Encoder, SHADE-Video Encoder and SHADE-Generative Model. Each component is meticulously designed to capture and integrate both general human actions and the subtle behavioral characteristic of AD.

SHADE-Text Encoder: This module is responsible for processing textual descriptions and converting them into rich semantic embeddings. It utilizes a pre-trained OpenCLIP text encoder [25], tailored specifically for our application. To embed the encoder with sensitivity to AD-specific terminology and nuances, we fine-tune it on a specialized corpus containing medical assessments, cognitive descriptors (e.g., Montreal Cognitive Assessment (MoCA) score, Zarit Burden Interview (ZBI) score, etc.), and behavioral observations related to AD we collected from 99 patients. This fine-tuning enables the encoder to grasp the intricate language patterns and context associated with AD behaviors, ensuring that generated videos accurately reflect the specified conditions.

SHADE-Video Encoder: The video encoder processes input activity videos to extract meaningful spatiotemporal representations. It combines CNNs for spatial feature extraction with temporal modeling [3] to capture motion dynamics over time. By training on both general action datasets and our AD-behavior dataset, the encoder learns to distinguish and represent the subtle differences in activity patterns associated with AD. The three-stage training approach enhances the encoder’s ability to capture a wide range of human actions while being sensitive to AD-specific characteristics.

SHADE-Generative Model: Serving as the core of the SHADE-AD framework, the generative model synthesizes skeleton-based human action videos conditioned on textual prompts. We have developed a customized diffusion-based generative model [26] that operates efficiently in a latent space, allowing for skeleton action generation. By integrating the embeddings from the text encoder and leveraging the spatiotemporal representations from the video encoder, the generative model produces videos that are both semantically aligned with the input text and exhibit activity patterns indicative of AD.

Output Adaptation: The generated skeleton video can be adapted to different sensor modalities using simulation tools like MATLAB and Blender, making it suitable for various smart health applications.

3.2 Training Procedure

Stage 1 - Pre-training on General Action Data. In the first stage, we pre-train SHADE Encoder using a large-scale dataset of general human actions. Specifically, we utilize the NTU RGB+D 120 dataset [20], which contains over 120,000 video samples across 120 different action classes. This dataset provides a comprehensive representation of common human activities, allowing our encoders to learn rich semantic and spatiotemporal features. To preserve the broad linguistic knowledge within the OpenCLIP text encoder, most layers remain fixed. We enable gradient updates only on its final transformer block and projection layer to minimally adapt the encoder to action-specific terminology. Meanwhile, the video encoder is fully trained on open dataset, ensuring the two encoders remain well-aligned. SHADE-Text Encoder is trained to capture the semantic meanings of action descriptions, employing contrastive learning to align textual embeddings with corresponding video embeddings produced by SHADE-Video Encoder.

Stage 2 - Fine-tuning on AD Behavior Dataset. We have collected a dataset involving 99 elders in various AD conditions, each monitored over a period of one month. Using privacy-preserving depth cameras installed in their primary living spaces, we obtained approximately 72,000 hours of video data, totaling around 100 TB.

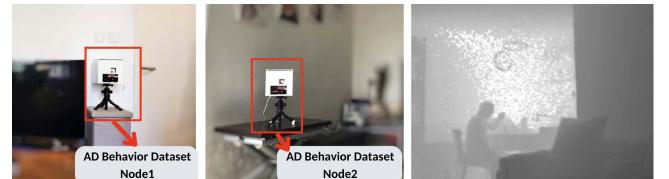


Figure 3: Examples of the layout of our data collection nodes and depth data in elders’ homes.

This dataset is enriched with detailed annotations, including cognitive health statuses (e.g., AD, Mild Cognitive Impairment (MCI) and Normal Cognition (NC)) and medical assessments like MoCA scores. The annotations provide valuable context for understanding the severity and progression of AD-related behaviors. To process the raw video data into a format suitable for training and reduce computational overhead, we employ pose estimation algorithms [8] to extract skeleton representations of the individuals.

During fine-tuning, we apply domain adaptation technologies [12, 29] to bridge the gap between the general action domain and the AD-specific domain. By introducing a domain discriminator and utilizing adversarial learning, the encoders learn domain-invariant features

that are sensitive to AD-specific characteristics. The domain classification loss, denoted as $\mathcal{L}_{\text{domain}}$, measures the accuracy of predicting the domain (general or AD-specific) of the extracted features.

$$\mathcal{L}_{\text{domain}} = \text{CrossEntropy}(d_{\text{true}}, d_{\text{pred}})$$

$$\mathcal{L}_{\text{Stage2}} = \mathcal{L}_{\text{main}} - \lambda \cdot \mathcal{L}_{\text{domain}}$$

The total loss for the encoders, $\mathcal{L}_{\text{total}}$, combines the main task loss (diffusion-based generation loss that ensures image quality and coherence) and the adversarial domain loss with a weighting factor λ . Here, the negative sign before $\mathcal{L}_{\text{domain}}$ indicates the adversarial nature of the training, as the encoders aim to maximize this loss to make it harder for the domain discriminator to correctly identify the origin domain of the features, while the domain discriminator aims to minimize it by becoming more accurate at distinguishing between general and AD-specific domains. Here, we adopt a smaller learning rate (5×10^{-5}) to avoid catastrophic forgetting of general language features while embedding AD-specific concepts (e.g., “MoCA Score= 15”). Concurrently, the video encoder undergoes adversarial alignment, ensuring it captures subtle AD cues, such as slower joint velocities. The domain alignment weight (λ) is set to 0.25.

Stage 3 - Fine-Grained Fine-Tuning with Patient-Specific Motion Metrics. We perform a fine-grained fine-tuning of the model using a subset of our AD behavior dataset that has been meticulously labeled. We selected data from 79 patients with rich action classes and labeled approximately 1% of the data focusing on 22 AD-related activities.

We not only use conventional training losses but also incorporate motion metrics of 12 key joints as part of the loss function. These joints include the left and right hips, knees, ankles, elbows, shoulders, neck, and back. The goal is to ensure that the generated actions accurately reflect patient-specific physical characteristics. A detailed explanation of our motion metrics is given in the next section.

We introduce a motion metric loss $\mathcal{L}_{\text{metric}}$ that measures the discrepancy between the motion metrics of the generated data and those of the real patient data:

$$\mathcal{L}_{\text{metric}} = \frac{1}{N} \sum_{i=1}^N \|M_{\text{real}}^{(i)} - M_{\text{synthetic}}^{(i)}\|_2^2$$

where $M_{\text{real}}^{(i)}$ and $M_{\text{synthetic}}^{(i)}$ represent the motion metrics of the i -th joint in the real and synthetic data, respectively, and N is the number of key joints. The overall loss function during Stage 3 combines the generative loss \mathcal{L}_{gen} (e.g., adversarial loss, reconstruction loss) and the motion metric loss:

$$\mathcal{L}_{\text{Stage3}} = \mathcal{L}_{\text{gen}} + \alpha \mathcal{L}_{\text{metric}}$$

where α is a weighting factor that balances the contribution of the motion metric loss. We use a modest weighting factor (0.5) to balance AD realism and generative quality, preventing the model from overlooking subtle AD behaviors such as repeated support movements, while also achieving consistent performance as observed during training.

By integrating the motion metrics into the training process, the model learns to generate synthetic data that not only appears realistic but also faithfully replicates the specific activity patterns and physical characteristics of AD patients.

Moreover, we implement a curriculum learning strategy [2] by gradually increasing the complexity of training samples. Starting with simpler actions commonly affected by AD (e.g., walking, sitting),

we progressively introduce more complex behaviors. This strategy helps the model to stabilize during training and improves its capacity to capture subtle AD-related movement deviations.

3.3 Advantages of the Design

The SHADE-AD framework offers several advantages: 1) *AD-Specific Knowledge Embedding*: three-stage training mechanism embeds domain-specific knowledge, capturing AD patients' subtle characteristics. 2) *Adaptability to Sensor Modalities*: Synthetic skeleton videos can be adapted for various applications. 3) *Privacy-Preserving and Cost-Effective*: Reduces the labeling work for real-world data.

An example can be seen in Figure 1, where our approach (prompted with “an Alzheimer’s Disease patient standing up from a chair”) demonstrates both the repeated support movements and the stooped posture, effectively capturing the motion characteristics observed in real AD patient data. SHADE-AD framework effectively addresses the challenges of synthesizing AD-specific human activity videos, providing a valuable tool for enhancing smart health.

4 MOTION METRICS FOR VALIDATION

We performed a comprehensive comparative analysis focusing on 12 key joints of the human body to assess the similarities and differences between the real and synthetic data. The analysis employed various motion metrics, including joint speeds, joint angles, range of motion (ROM), and correlation measures, to evaluate how effectively the synthetic data replicates the motion patterns of real AD patients. We compare 200 “Walking” action video clips between real elders’ data and synthetic data generated using the elders’ same health condition. Representative results are shown in Table 1 and illustrated by the correlation heatmap in Figure 4.

Joint Speed Distribution. The distributions of joint speeds between the real and synthetic data show a high degree of overlap across the analyzed joints. For instance, the histograms of *Left Hip Speed* and *Right Wrist Speed* (Figures 5) illustrate that the synthetic data effectively replicates the speed characteristics of AD patients. This close match suggests that SHADE-AD successfully learned the slower movement speeds typical of AD-related motor decline. This is critical for recognizing actions like *walking*, where speed variations are significant indicators of motor impairment.

Joint Angle Trajectories. The time series of joint angles, after DTW alignment, exhibit similar patterns and trends between the real and

Feature	Mean Ratio (S/R)	ROM Ratio (S/R)	Correlation	t-test p-value
Left Hip Angle	0.72	1.24	0.8626	7.18×10^{-3}
Right Hip Angle	0.84	0.88	0.7455	0.4487
Left Knee Angle	0.68	1.70	0.5860	2.07×10^{-4}
Right Knee Angle	0.81	0.84	0.7686	0.0655
Left Shoulder Angle	1.17	0.49	0.8026	0.0009
Right Shoulder Angle	0.91	0.60	0.7130	0.0502
Neck Angle	0.99	0.63	0.8710	0.5028
Back Angle	0.99	1.23	0.4323	0.3769
Neck Speed	0.95	0.84	0.6590	0.0029
Left Ankle Speed	1.10	1.00	0.6444	0.6541
Right Ankle Speed	0.83	0.83	0.6453	0.5045
Left Wrist Speed	0.83	0.87	0.7241	0.7618
Right Wrist Speed	0.99	0.76	0.6389	0.5651
Left Knee Speed	1.18	1.01	0.5260	0.4507
Right Knee Speed	0.73	1.00	0.5000	0.3064
Left Hip Speed	1.12	1.52	0.8167	0.6335
Right Hip Speed	1.00	0.98	0.7214	0.9538

Table 1: Comparison of Motion Metrics Between Real and SHADE-AD Synthetic Data on “Walking” action.

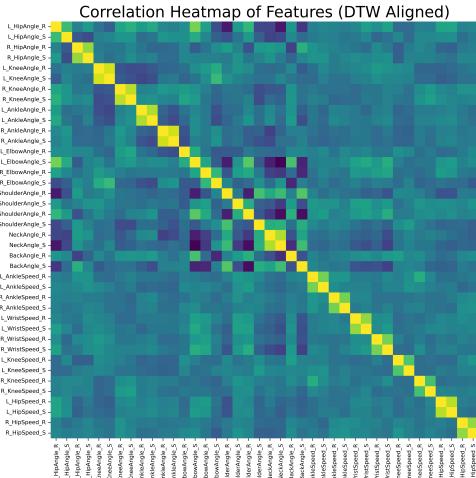


Figure 4: Correlation heatmap of joint features between real and synthetic data (DTW aligned), demonstrating strong positive correlations for most features. Suffixes *_R* and *_S* represent Real and Synthetic data respectively.

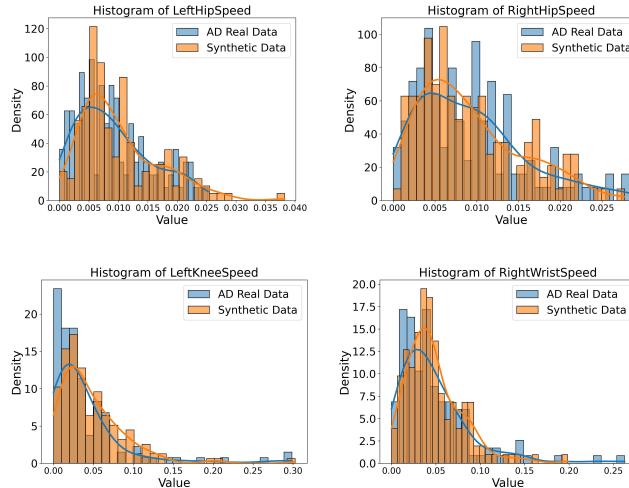


Figure 5: Histograms of various features for real and synthetic data, showing overlapping distributions and indicating similar movement dynamics.

synthetic data in one activity period. As shown in Figure 6, the *Right Knee Angle* over time demonstrates consistent cyclical movements corresponding to gait patterns. Although there are minor discrepancies at the initiation of actions, likely due to differences in starting postures, the overall alignment indicates that the synthetic data captures the essential movement dynamics.

Mean and Range of Motion (ROM). Most Mean and ROM Ratios of synthetic and real data are around 1.0 in Table 1, which confirms that the synthetic data accurately represents the limitations in joint mobility associated with AD.

Correlation. The correlation coefficients between real and synthetic data are generally high for most features, with several exceeding 0.7, indicating strong positive relationships (Table 1). The correlation

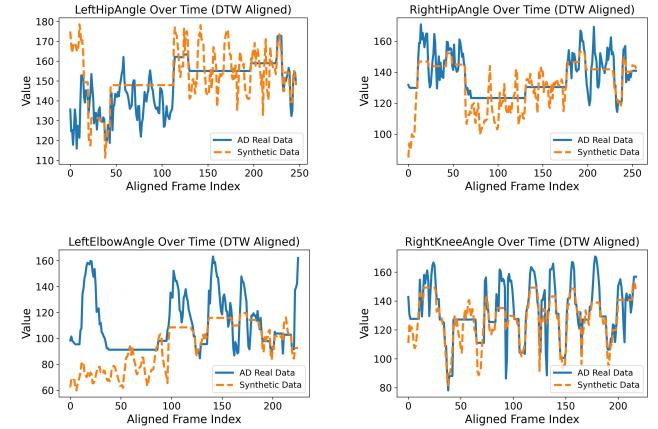


Figure 6: Time series of joint speeds and angles over time (DTW aligned) for real and synthetic data, demonstrating consistent trends within action cycles.

heatmap (Figure 4) visually reinforces this finding, showing strong correlations across multiple joint features.

Statistical Tests. The t-test p-values provide statistical evidence regarding the differences between the datasets. For features like *Right Hip Angle*, high p-values ($p > 0.05$) indicate no significant difference between the real and synthetic data. This statistical similarity further supports the conclusion that the gap between the datasets is minimal.

5 PRELIMINARY RESULTS

To assess the effectiveness of the proposed SHADE-AD framework, we first compare our method with four existing data enhancement approaches, and subsequently evaluate the adaptability of SHADE-AD across another data modality.

5.1 Experiment 1: Human Activity Recognition Task (Skeleton-based)

Objective: Assess the performance of HAR models trained with our synthetic data and other data enhancement methods in recognizing activities in AD patients.

Model and Methods: We trained a lightweight HAR model composed of five convolutional layers and two recurrent layers, suitable for deployment on embedded systems in smart health applications. The model was trained with following methods:

- 1) *Vanilla Model without data augmentation (Model V)*. Trained on AD behavior dataset containing 21,910 labeled skeleton videos from 79 elderly individuals, with a biased distribution heavily favoring the *sitting* action (*sitting*: 16,596; *standing*: 3,121; *walking*: 1,735; *turning*: 272; *lying*: 186).
- 2) *Model enhanced with common Data Augmentation (Model DA)*. Trained on the AD behavior dataset using common data augmentation techniques (rotation, scaling and noise addition), increasing the dataset size by 1.5 times.
- 3) *Model enhanced with common Statistical Method (Model SM)*. Trained on the AD behavior dataset using bootstrapping strategy that resamples skeleton sequences, increasing the dataset size by 1.5 times.

- 4) *Model enhanced with Open Dataset (Model OD)*. Trained on the AD behavior dataset combined with data from open action dataset. Incorporated 5,000 skeleton videos from the NTU RGB+D 120 dataset [20], covering similar action classes performed by healthy individuals.
- 5) *Model enhanced with LLM-Generated Data (Model LLM)*. Trained using synthetic data generated by MotionDiffuse [34], containing 20,000 videos with balanced action distribution.
- 6) *Model enhanced with LLM and Text Conditioning (Model LLM-AD)*. Similar to 5) but with prompts explicitly mentioning “AD elder”.
- 7) *Model enhanced with Our Synthetic Data (Model OS)*. Trained using synthetic data generated by SHADE-AD, which contains 20,000 videos with balanced action distribution, evenly health condition representation (AD, MCI, and NC), and 360-degree viewpoints around subjects. Prompts are structured like, “Generate a HAR training dataset for an elder with AD, MoCA Score= 9, ..., of walking action, from a front view angle.”

Testing Data: We used 3,200 action videos from another 20 elders. To test the real-world performances of the seven approaches more precisely, we select this cross-subject dataset with a balanced action distribution to mitigate the problems of high variance and high bias. **Analysis:** The results of Experiment 1 shown in Table 2 highlight the limitations of existing methods and the effectiveness of our SHADE-AD approach: *Model V* achieves 100.00% accuracy for the *sitting* action due to severe data imbalance but performs poorly on other actions, indicating overfitting and inability to generalize; *Model DA* slightly improves recognition of underrepresented actions, but the enhancements are minimal because these techniques do not introduce new semantic information or address the complexity of AD-specific behaviors; *Model SM* performs marginally better than *Model DA*. By re-sampling skeleton sequences, it diversifies the training data, yet still fails to generate fundamentally new disease-relevant behavior patterns (e.g., subtle stooping); *Model OD* yields better performance on some common actions; however, discrepancies in action execution and context—such as different gait patterns and movement nuances between healthy individuals and AD patients—result in poor recognition; *Model LLM* shows moderate improvements across actions, yet it lacks AD-specific features since current generative models do not capture the subtle, disease-related action patterns; *Model LLM-AD* improves upon *Model LLM* by including “AD elder” in the prompts, leading to better recognition on walking and turning—two actions that can manifest subtle AD characteristics; *Model OS*, trained with synthetic data generated by SHADE-AD that embeds AD-specific behavioral knowledge, achieves balanced and significantly high accuracies across all action classes. This result demonstrates that our approach effectively addresses data imbalance, captures complex AD-related behaviors, and enhances model generalization by providing realistic and diverse training examples tailored to the specific characteristics of AD patients.

Model	Sitting (%)	Standing (%)	Walking (%)	Turning (%)	Lying (%)
Model V (base)	100%	0%	3.13%	0%	0%
Model DA	96.25%	10.63%	8.75%	2.03%	1.56%
Model SM	97.34%	14.22%	10.63%	4.84%	2.50%
Model OD	83.28%	25.16%	12.19%	15.16%	8.44%
Model LLM	69.06%	66.25%	64.38%	63.28%	58.75%
Model LLM-AD	72.34%	67.97%	68.91%	70.16%	60.47%
Model OS	76.56%	73.44%	78.13%	79.69%	75.00%

Table 2: HAR Accuracies using different data enhancement methods on testing data.

5.2 Experiment 2: Test the Adaptability to Different Modality (Depth Videos)

Objective: Evaluate the adaptability of our approach to another data modalities beyond skeleton data.

Model and Methods: A Depth-based HAR model is used.

1) *Vanilla Depth Model without data augmentation (Model V-Depth)*. Trained on AD behavior depth videos (corresponding to the skeleton data in AD behavior dataset used in Exp. 1).

2) *Our Enhanced Depth Model (Model OS-Depth)*. Trained on synthetic depth videos simulated from our synthetic skeleton data in Experiment. 1 using Matlab and Blender.

Model	Sitting (%)	Standing (%)	Walking (%)	Turning (%)	Lying (%)
Model V-Depth	97.97%	5.0%	0%	2.03%	0%
Model OS-Depth	73.75%	70.94%	66.56%	67.34%	65.63%

Table 3: Depth-Based HAR Accuracies on the testing data of 3,200 raw depth videos same in Experiment 1.

Analysis: *Model OS-Depth* significantly improves recognition, which demonstrates that SHADE-AD can be adapted to other sensor modalities, starting from skeleton data and processing it to match various data types required by edge devices. The ability to generate synthetic data suitable for different modalities highlights the flexibility and practicality of SHADE-AD in real-world applications.

5.3 Highlights

The enhanced performance of our models can be attributed to: 1) *Class Balance*: The synthetic dataset provided equal representation of all action classes, mitigating the bias present in the real-world data. 2) *Physiological Variation*: Inclusion of diverse physiological characteristics (e.g., AD, MCI and NC) in the synthetic data allowed the models to learn variations in action patterns associated with different health conditions. 3) *Viewpoint Diversity*: Synthetic data included multiple viewpoints, improving the models’ ability to recognize activities from different angles.

5.4 Future Work

While our approach enhances AD-specific HAR, further improvements can be explored. 1) *Performance Improvement*: Incorporating biomechanical constraints and broader datasets may enhance movement fidelity and generalization. 2) *Comprehensive Motion Validation*: Beyond key-joint metrics, multi-joint coordination analysis can provide a more comprehensive assessment. 3) *Semantic Consistency and AD Relevance*: Future refinements will focus on ensuring semantic consistency and clinical relevance in realistic disease progression.

6 CONCLUSION

We present SHADE-AD as a privacy-preserving and low-cost tool for synthesizing human activity datasets for AD patients. SHADE-AD demonstrates a significant improvement in the accuracy of human activity recognition when enhancing embedded models using a synthesis dataset, which shows a promising potential in smart health.

7 ACKNOWLEDGMENTS

This work is supported by the Research Grants Council (RGC) of Hong Kong, China, under grants GRF 14201924 and CRF C4034-21G, CUHK Direct Grant 4055216, and the Alzheimer’s Drug Discovery Foundation, under Grant RDADB-201906-2019049.

REFERENCES

- [1] Alzheimer's Association. 2024. *2024 Alzheimer's Disease Facts and Figures*. Annual Report. Alzheimer's Association. <https://alz.org/media/Documents/Facts-And-Figures-2024-Executive-Summary.pdf>
- [2] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 41–48.
- [3] Gedas Bertasius, Heng Wang, and Lorenzo Torresani. 2021. Is Space-Time Attention All You Need for Video Understanding?. In *Proceedings of the International Conference on Machine Learning (ICML)*. 813–824.
- [4] Christopher Buckley, Lisa Alcock, Riona McArdle, Rana Zia Ur Rehman, Silvia Del Din, Claudia Mazzà, Alison J. Yarnall, and Lynn Rochester. 2019. The Role of Movement Analysis in Diagnosing and Monitoring Neurodegenerative Conditions: Insights from Gait and Postural Control. *Brain Sciences* 9, 2 (2019). <https://doi.org/10.3390/brainsci9020034>
- [5] Shenggu Cai, Duygu Ceylan, Matheus Gadelha, Chun-Hao Paul Huang, Tuan-feng Yang Wang, and Gordon Wetzstein. 2023. Generative Rendering: Controllable 4D-Guided Video Generation with 2D Diffusion Models. arXiv:2312.01409 [cs.CV] <https://arxiv.org/abs/2312.01409>
- [6] Haixin Chen, Yong Zhang, Xiaodong Cun, Menghan Xia, Xintao Wang, Chao Weng, and Ying Shan. 2024. VideoCrafter2: Overcoming Data Limitations for High-Quality Video Diffusion Models. arXiv:2401.09047 [cs.CV] <https://arxiv.org/abs/2401.09047>
- [7] Xingyu Chen and Xinyu Zhang. 2024. RF Genesis: Zero-Shot Generalization of mmWave Sensing through Simulation-Based Data Synthesis and Generative Diffusion Models. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems (Istanbul, Turkiye) (SenSys '23)*. Association for Computing Machinery, New York, NY, USA, 28–42. <https://doi.org/10.1145/3625687.3625798>
- [8] MMPose Contributors. 2020. OpenMMLab Pose Estimation Toolbox and Benchmark. <https://github.com/open-mmlab/mmpose>.
- [9] Andre Esteve, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. 2019. A guide to deep learning in healthcare. *Nature Medicine* 25, 1 (2019), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
- [10] Yi Fan, Zhang Yaoyun, Yuan Jing, Liu Ziyue, Zhai Feifei, Hao Ankai, Wu Fei, Somekh Judith, Peleg Mor, Zhu Yi-Cheng, and Huang Zhengxing. 2023. Identifying underlying patterns in Alzheimer's disease trajectory: a deep learning approach and Mendelian randomization analysis. *EClinicalMedicine* (2023). <https://doi.org/10.1016/j.eclimn.2023.101024>
- [11] Heming Fu, Hongkai Chen, and Guoliang Xing. 2024. Demo Abstract: AD-CLIP: Privacy-Preserving, Low-Cost Synthetic Human Action Dataset for Alzheimer's Patients via CLIP-based Models. In *2024 23rd ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 257–258. <https://doi.org/10.1109/IPSN61024.2024.00029>
- [12] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-Adversarial Training of Neural Networks. In *Journal of Machine Learning Research (JMLR)*, Vol. 17, 1–35.
- [13] Yunqi Guo, Kaiyuan Hou, Heming Fu, Hongkai Chen, Zhenyu Yan, Guoliang Xing, and Xiaofan Jiang. 2024. Vivar: A Generative AR System for Intuitive Multi-Modal Sensor Data Presentation. arXiv:2412.13509 [cs.HC] <https://arxiv.org/abs/2412.13509>
- [14] Kaiyuan Hou, Yunqi Guo, Heming Fu, Hongkai Chen, Zhenyu Yan, Guoliang Xing, and Xiaofan Jiang. 2024. Improving On-Device LLMs' Sensory Understanding with Embedding Interpolations. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking (Washington D.C., DC, USA) (ACM MobiCom '24)*. Association for Computing Machinery, New York, NY, USA, 1674–1676. <https://doi.org/10.1145/3636534.3697456>
- [15] Siyang Jiang, Xian Shuai, and Guoliang Xing. 2024. ArtFL: Exploiting Data Resolution in Federated Learning for Dynamic Runtime Inference via Multi-Scale Training. In *2024 23rd ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 27–38. <https://doi.org/10.1109/IPSN61024.2024.00007>
- [16] Lamprós C. Kourtis, Oliver B. Regele, Justin M. Wright, and Graham B. Jones. 2019. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. *npj Digital Medicine* 2 (2019), 9. <https://doi.org/10.1038/s41746-019-0084-2>
- [17] Yang Li, Doris Sau Fung Yu, Shuangzhou Chen, Guoliang Xing, and Hongkai Chen. 2024. Demo: EmoMarker: A Privacy-Preserving, Multi-Modal Sensing System for Dyadic Digital Biomarkers of Expressed Emotions for Patients with Dementia. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services (Minato-ku, Tokyo, Japan) (MOBISYS '24)*. Association for Computing Machinery, New York, NY, USA, 614–615. <https://doi.org/10.1145/3643832.3661844>
- [18] Zhiyuan Li, Hangxiao Zhu, Zhioran Lu, and Ming Yin. 2023. Synthetic Data Generation with Large Language Models for Text Classification: Potential and Limitations. arXiv:2310.07849 [cs.CL]
- [19] Neiwen Ling, Yuze He, Nan Guan, Heming Fu, and Guoliang Xing. 2023. An Indoor Smart Traffic Dataset and Data Collection System: Dataset. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems (Boston, Massachusetts) (SenSys '22)*. Association for Computing Machinery, New York, NY, USA, 1007–1010. <https://doi.org/10.1145/3560905.3567772>
- [20] Jie Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C Kot. 2019. NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 42. IEEE, 2684–2701.
- [21] Yang Liu, Jiahuan Cao, Chongyu Liu, Kai Ding, and Lianwen Jin. 2024. Datasets for Large Language Models: A Comprehensive Survey. arXiv:2402.18041 [cs.CL] <https://arxiv.org/abs/2402.18041>
- [22] Xiaomin Ouyang, Xian Shuai, Yang Li, Li Pan, Xifan Zhang, Heming Fu, Xinyan Wang, Shihua Cao, Jiang Xin, Hazel Mok, Zhenyu Yan, Doris Sau Fung Yu, Timothy Kwok, and Guoliang Xing. 2024. ADMarker: A Multi-Modal Federated Learning System for Monitoring Digital Biomarkers of Alzheimer's Disease. arXiv:2310.15301 [cs.LG]
- [23] Xiaomin Ouyang, Zhiyuan Xie, Heming Fu, Sitong Cheng, Li Pan, Neiwen Ling, Guoliang Xing, Jiayu Zhou, and Jianwei Huang. 2023. Harmony: Heterogeneous Multi-Modal Federated Learning through Disentangled Model Training. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services (Helsinki, Finland) (MobiSys '23)*. Association for Computing Machinery, New York, NY, USA, 530–543. <https://doi.org/10.1145/3581791.3596844>
- [24] Martin Prince, Emiliano Albanez, Maélen Guerchet, and Matthew Prina. 2014. World Alzheimer Report 2014: Dementia and Risk Reduction.
- [25] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020 [cs.CV]
- [26] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 10684–10695.
- [27] Joeky T Senders, Maryam M Zaki, Aditya V Karhade, Benjamin Chang, William B Gormley, Marike LD Broekman, Timothy R Smith, and Omar Arnaout. 2018. An introduction and overview of machine learning in neurosurgical care. *Acta Neurochirurgica* 160, 1 (2018), 29–38. <https://doi.org/10.1007/s00701-017-3385-8>
- [28] Connor Shorten and Taghi M Khoshgoftaar. 2019. A Survey on Deep Learning: Techniques, Applications, and Future Directions. *Journal of Big Data* 6, 1 (2019), 1–20.
- [29] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. 2017. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7167–7176.
- [30] Limin Wang, Yuanjun Xiong, Zhe Wang, and Others. 2018. Temporal Segment Networks for Action Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 1228–1237.
- [31] Xuai Hu, Xin Liu, Han Zhang, Weichen Wang, Subigya Nepal, Yasaman Sefidgar, Woosuk Seo, Kevin S. Kuehn, Jeremy F. Huckins, Margaret E. Morris, Paula S. Nurius, Eve A. Riskin, Shwetak Patel, Tim Althoff, Andrew Campbell, Anind K. Dey, and Jennifer Mankoff. 2023. GLOBEM: Cross-Dataset Generalization of Longitudinal Human Behavior Modeling. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 190 (Jan. 2023), 34 pages. <https://doi.org/10.1145/3569485>
- [32] Bufang Yang, Siyang Jiang, Lilin Xu, Kaiwei Liu, Hai Li, Guoliang Xing, Hongkai Chen, Xiaofan Jiang, and Zhenyu Yan. 2024. DrHouse: An LLM-empowered Diagnostic Reasoning System through Harnessing Outcomes from Sensor Data and Expert Knowledge. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 153 (Nov. 2024), 29 pages. <https://doi.org/10.1145/3699765>
- [33] Emi Yuda, Tomoki Ando, Itaru Kaneko, Yutaka Yoshida, and Daisuke Hirahara. 2024. Comprehensive Data Augmentation Approach Using WGAN-GP and UMAP for Enhancing Alzheimer's Disease Diagnosis. *Electronics* 13, 18 (2024). <https://doi.org/10.3390/electronics13183671>
- [34] Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. 2022. MotionDiffuse: Text-Driven Human Motion Generation with Diffusion Model. *arXiv preprint arXiv:2208.15001* (2022).
- [35] Wenli Zhang, Yifan Li, Wentao Ren, and Bo Liu. 2023. Artificial intelligence technology in Alzheimer's disease research. *Intractable & Rare Diseases Research* 12, 4 (2023), 208–212. <https://doi.org/10.5582/irdr.2023.01091>
- [36] Xiyuan Zhang, Diyan Teng, Ranak Roy Chowdhury, Shuheng Li, Dezhi Hong, Rajesh K Gupta, and Jingbo Wang. 2024. UniMTS: Unified Pre-training for Motion Time Series. *arXiv preprint arXiv:2410.19818* (2024).
- [37] Zhihao Zhang, Xiangtao Liu, Suixia Zhang, Zhixin Song, Ke Lu, and Wenzhong Yang. 2024. A review and analysis of key biomarkers in Alzheimer's disease. *Frontiers in Neuroscience* 18 (2024). <https://doi.org/10.3389/fnins.2024.1358998>
- [38] Li Zhaoying, Zhu Jingyi, Liu Junyan, Shi Min, Liu Pan, Guo Junjie, Hu Zhenzhu, Liu Shanyu, and Yang Dongdong. 2023. Using dual-task gait to recognize Alzheimer's disease and mild cognitive impairment: a cross-sectional study. *Frontiers in Human Neuroscience* 17 (2023), 1284805. <https://doi.org/10.3389/fnhum.2023.1284805>
- [39] Tong Zhou, Xuhang Chen, Yanyan Shen, Martin Nieuwoudt, Chi-Man Pun, and Shu-Qiang Wang. 2023. Generative AI Enables EEG Data Augmentation for Alzheimer's Disease Detection Via Diffusion Model. 1–6. <https://doi.org/10.1109/ISPCE-ASIA60405.2023.10365931>