The Impact of Synthetic Data on Fall Detection Application

Minakshi Debnath[0009-0008-2777-2437], Md Shahriar Kabir[0009-0004-4908-445X], Jianyuan Ni[0000-0002-6968-6536], and Anne H. H. Ngu[0000-0002-5877-0230]

Texas State University, San Marcos, TX, USA

Abstract. Lack of real-world data in clinical fields poses a major obstacle for training deep learning models.Using data augmentation can increase data volume, making the training of deep learning models more effective. This paper aims to investigate different techniques for generating realistic multivariate synthetic fall data, addressing the challenge of limited fall data availability. We experimented with three traditional time series data augmentation techniques, a generative AI approach with diffusion, and extraction of data from public video recordings of older adults falling. We evaluated the effectiveness of the generated data with both an LSTM model trained offline and using the SmartFall App running the LSTM model in real-time. Initial results indicate a 7-10% increase in the F1-score for the fall detection model when trained with additional data generated through the diffusion method during offline evaluation and a notable improvement of 24% was observed with the real-time evaluation of the model.

Keywords: Time series data generation · Fall detection · Diffusion model · Video data extraction

1 Introduction

Falling poses a significant health risk for older adults globally [9]. In fact, the injury posed by falling in older adults are the leading cause of unintentional death in individuals over 85 years old [15]. Research on wearable device technologies like smartwatches and IMU sensors for fall detection has become popular due to their affordability, portability, and non-intrusiveness. In complex physiological processes like fall onset, deep learning struggles with limited training data as fall events are rare and large data collection is difficult. Researchers have collected simulated fall data in controlled environments, a costly and labor-intensive process. Data augmentation or synthetic data generation techniques are one of the standard approaches to addressing the issue of small datasets[5]. Generative AI, like GANs, VAEs, and Diffusion Models, is prominent in creating synthetic data for images and time series.Diffusion models have become a popular method among deep generative models, showcasing outstanding performance in diverse applications [17]. More recently, virtual IMU signal has been reported as a reliable alternative way for synthetic data. For instance, an engineering pipeline was proposed to generate on-body virtual sensor data utilizing data of a different modality (i.e., video) [6]. Therefore, we have adopted the methodology presented in [8] for the extraction of video fall data publicly available from two long-term care facilities in British Columbia [14].

In this work our contributions include: A) Introducing the Diffusion model for data generation. B) Extracting fall data from videos using pose estimation. C) Validating synthetic data techniques. D) Comparing fall detection model performance with real and synthetic data using the SmartFall App. E) Showing the effectiveness of data generated with the Diffusion model and video extraction in improving fall detection models.

2 Experimental Setup

Datasets: We employed three fall-based datasets as input to different synthetic data generation techniques and one video dataset for extraction of fall data for impact assessment. Those are SmartFallMM's smartwatch data (accelerometer data) (collected in our laboratory) [2], the UniMiB[11], and the K-Fall [18]. All those datasets have various simulated falls and activities of daily life performed by healthy young adults. The video dataset is a real-life video recording of older adults falling in a long-term care facility in British Columbia [14].

Data Preprocessing, Deep learning Model, Training and Evaluation: We used a basic LSTM deep learning model, which is favored for time series data due to its capability to learn temporal dynamics. The detail of the architecture can be found in our technical report[1]. Our model, deployed and tested in our SmartFall App, outperformed 1D CNN, Gradient Boosting, and Random Forest [10].

The input data is pre-processed by segmenting into overlapping windows with a step size of 10, using a window size of 128 across all experiments. Different training scenarios are explored, with baseline models trained solely on original datasets, without any synthetic data. The dataset is split into training, validation, and test sets at a ratio of 70/20/10, and a 5-fold validation method is applied. Baseline models serve as the reference. New LSTM models are trained using combined original and synthetic data, while validation and testing are conducted solely on real data. Performance evaluation includes standard metrics: Precision, Recall, F1-score, and Accuracy, to assess the effectiveness of synthetic data from various methods.

We validate the best model using generated data with the SmartFallMM dataset in a real-world setting via the SmartFall App [12]. Three students participated in the evaluation under IRB 7846 at Texas State University. They wore watches on the left wrist with the SmartFall App installed, executing falls on an air mattress and daily activities. Both correct and incorrect predictions were recorded.

3 Synthetic Data Generation

Basic Data Augmentations: We employed three data augmentation techniques, namely Jittering [13], Magnitude Warping [13], and Rotation [16]. Jittering involves augmenting time series data with random Gaussian noise. Magnitude warping is a technique applied to time series datasets where the magnitude of each sample is modified. This modification is achieved by multiplying the original time series with a cubic spline curve. The rotation augmentation technique serves as a means to simulate various sensor placements (e.g. left vs right wrist), introducing the diversity of data patterns without modifying the inherent labels associated with the data.

Diffusion Method: Denoising Diffusion Probabilistic Models (DDPMs) represent a class of generative AI models that have demonstrated remarkable success in synthesizing high-quality data across domains such as images and audio [4, 7]. We have integrated diffusion models with a U-Net architecture adapted from previous work [7]. Originally designed for image analysis, this U-Net architecture has been reconfigured for time-series data using one-dimensional (1D) convolutional layers with a kernel size of 7 and padding of 3, capturing essential temporal dependencies in time series data. Figure 1 represents the architecture used for this work.

Upon receiving the time-series input, the data undergoes normalization with RMSNorm, which stabilizes the training process. The network architecture, comprising ResNet blocks and Linear Attention units, executes downsampling and upsampling operations to refine features and preserve temporal information. Time and sinusoidal positional embeddings are integrated within each block, ensuring the model's responsiveness to the diffusion timesteps and sequence positions. Li's original model [7] posed challenges in

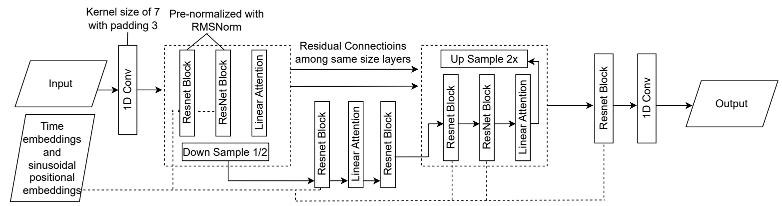


Fig. 1: Schematic of the U-Net Architecture Adapted for Time-Series Data.

handling variable-length accelerometer data for different types of falls and lacked a stable normalization technique. We improved the model by incorporating a padding strategy during preprocessing, enabling consistent input size and avoiding loss of information. Detail of the method can be found in our technical report[1].

Extraction of fall data from video via Pose Estimation: We have adopted the methodology presented in [8] for the extraction of video fall data. To extract the fall data correctly, we edited 34 publicly available videos sourced from [14]. We first isolated the falling person in the video by cropping the frame around them to reduce the time for the extraction process and to zoom in on the most relevant data to extract. We ensure to include 1 to 2 seconds of pre-fall and post-fall segments. Resolution and brightness adjustments are made for each video. The 3D pose estimation extracted 17 joint positions from each video's detected human skeleton. For generating synthetic data, if we aim to add video fall data to the SmartFallMM dataset, we focus on extracting accelerometer data from the left wrist joint position. Alternatively, for UniMiB, we extract accelerometer data from the left and right hips' joint positions. If we are creating synthetic data for UniMiB, we will extract accelerometer data from the left and right hips' joint positions. After pose estimation, we use 3D keypoints to extract acceleration data. Calculating velocity from position changes, then acceleration from velocity changes, we extract about 30 fall samples. Figure 2 outlines this methodology for deriving accelerometer readings from a video capturing an elderly person's fall.

4 Results

Offline Evaluation of Fall Detection Model: Figure 3(a) compares datasets and synthetic data using three methods. Results from 5-fold validation, including precision, recall, F1-score, and accuracy, are shown for each. Each colored line represents the variation of each metric across different datasets. Abbreviations SF, UM, and KF represent SmartFallMM, UniMiB, and K-Fall datasets, while DF, VE, Jit, MW, and Ro stand for Diffusion, Video Extraction, Jittering, Magnitude Warping, and Rotation. Only SmartFallMM dataset (SF) achieves F1 score of 0.72 and accuracy of 0.77. With synthetic data, especially using diffusion, SF's F1-score improves to 0.80, nearly 10% better. Pose estimation-based data extraction also boosts performance. We additionally assessed and compared results across two other public datasets, UniMiB and K-Fall. The baseline F1 score and accuracy for UniMiB (UM) are 0.79 and 0.78, respectively. We noted a enhancement in performance by incorporating synthetic data generated via the diffusion

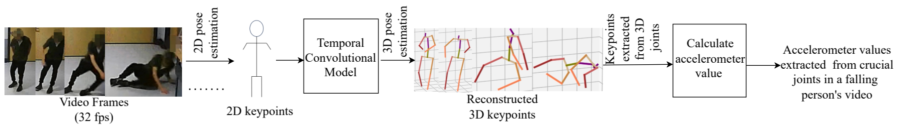
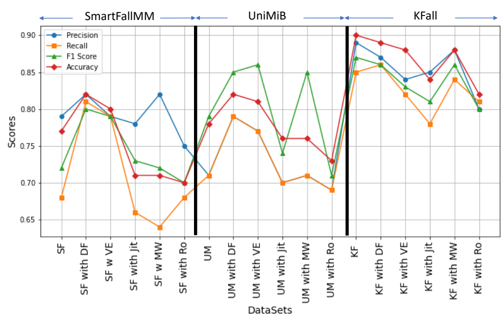
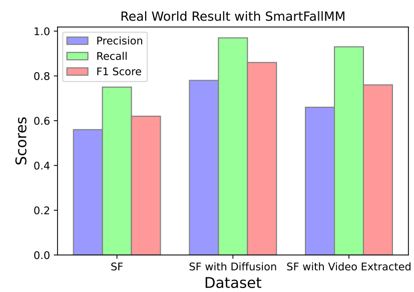


Fig. 2: Accelerometer data extraction process from video frames.

method. The F1 score increased from 0.79 to 0.85, reflecting an improvement rate of nearly 7%. Despite incorporating diffusion-generated and video-extracted data, there was no improvement observed for K-Fall (KF). This could be attributed to the larger size of the K-Fall dataset compared to the other two datasets, the added data does not lead to more generalization with the simple LSTM architecture.



(a)



(b)

Fig. 3: **(a)**Evaluation of SmartafallMM, UniMiB, and K-Fall with Synthetic Data Generated using three different methods. **(b)**Real world result with SmartFallMM.

The better performance gap of synthetic data from diffusion and pose extraction methods as compared to basic augmentation likely stems from the quality of the added information. Data from diffusion and pose extraction enriches the dataset with meaningful patterns and the generated data aligns better with real data.

Real-time Evaluation of Fall Detection Model Figure 3(b) showcases the real-time evaluation result for the top-performing offline model. We only tested the offline model with SmartFallMM watch data because our SmartFall app exclusively uses watch-sensed data. We share results from testing the SmartFall App across three participants, starting with an initial F1 score of 0.62 using basic LSTM model for SF. Next, we evaluated top models trained with a mix of synthetic and real data: SF with Diffusion and SF with Video Extracted. The top SmartFall App model, trained with diffusion-generated data, achieved an F1 score of 0.86 (24% improvement), while the video-extracted data model reached 0.76 (14% improvement). Real-time testing confirms synthetic data's effectiveness in enhancing fall detection methods.

5 Discussion and Future Work

This study explores methods to generate synthetic fall data to overcome data scarcity. Enhanced performance is observed in offline evaluation for SmartFallMM and UniMiB with diffusion and video-extracted synthetic fall data. Additionally, promising real-time performance is demonstrated for SmartFallMM with synthetic data from diffusion and video extraction. In the future, we aim to identify the ideal balance of real and synthetic data for training robust models, alongside exploring video extraction methods via AI platforms like Sora [3], which generate videos from textual descriptions.

Acknowledgment

We thank the National Science Foundation for funding the research under the NSF-SCH (21223749).

References

1. Enhancing fall detection: The role of synthetic data, https://drive.google.com/file/d/1WLcxcjwg1d_t1i0T2RLr930Cpx-j64tt/view .
2. Smartfallmm watch accelerometer dataset, https://drive.google.com/file/d/10tOrG7zgbLOgBJTFj0PDl3DYFjz7kkTO/view .
3. Sora. https://openai.com/sora .
4. Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *NeurIPS*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.
5. Brian K. Iwana and Seiichi Uchida. An empirical survey of data augmentation for time series classification with neural networks. *PLoS ONE*, 16(7):e0254841, 2021.
6. Hyeokhyen Kwon, Catherine Tong, Harish Haresamudram, Yan Gao, Gregory D Abowd, Nicholas D Lane, and Thomas Ploetz. Imutube: Automatic extraction of virtual on-body accelerometry from video for human activity recognition. *IMWUT*, 4(3):1–29, 2020.
7. Xiaomin Li. Mitigating Data Shortage in Biomedical Signal Analysis: An Investigation into Transfer Learning and Generative Models. Ph.D. dissertation, Texas State University, Texas, June 2023.
8. Junfa Liu, Juan Rojas, Yihui Li, Zhijun Liang, Yisheng Guan, Ning Xi, and Haifei Zhu. A graph attention spatio-temporal convolutional network for 3d human pose estimation in video. In *ICRA*, pages 3374–3380. IEEE, 2021.
9. João P. Matos-Carvalho, Sérgio D. Correia, and Slavisa Tomic. Sensitivity analysis of lstm networks for fall detection wearable sensors. In *CIoT*, pages 112–118, 2023.
10. Taylor R. Mauldin, Anne H. Ngu, Vangelis Metsis, and Marc E. Canby. Ensemble deep learning on wearables using small datasets. *ACM Trans. Comput. Healthcare*, 2(1), December 2021.
11. Daniela Micucci, Marco Mobilio, and Paolo Napoletano. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, 7(10), 2017.
12. Anne H. Ngu, Awatif Yasmin, Tarek Mahmud, Adnan Mahmood, and Quan Z. Sheng. Demo: P-fall: Personalization pipeline for fall detection. In *Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies*, CHASE '23, page 173–174, New York, NY, USA, 2024. Association for Computing Machinery.
13. Alexander Nikitin, Letizia Iannucci, and Samuel Kaski. TSGM — a flexible framework for synthetic time series generative modeling. 2023.
14. Stephen Robinovitch. Falls experienced by older adult residents in long-term care homes, 2018.
15. S Sherin Zebiah, Ancy Vetha Shalomy, Jyotsana Kachhap, Nikita Tete, R Nancy, A. Ananthi, J. Prasanna, M.S.P. Subathra, and S.Thomas George. Human fall detection using machine learning and deep learning techniques: A survey. In *ICSPC*, pages 253–257, 2023.
16. Terry T. Um, Franz M. J. Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *ICMI*, pages 216–220, New York, NY, USA, 2017. ACM.
17. Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Comput. Surv.*, 56(4), nov 2023.
18. Xiaoqun Yu, Jaehyuk Jang, and Shuping Xiong. A large-scale open motion dataset (kfall) and benchmark algorithms for detecting pre-impact fall of the elderly using wearable inertial sensors. *Frontiers in Aging Neuroscience*, 13:692865, 2021.