Human Activity Classification for Patient/Elderly Monitoring

1st Urmi Saha
Department of Computer Science
Wake Forest University
sahau24@wfu.edu

1st Mohammad Marufur Rahman
Department of Computer Science
Wake Forest University
rahmm224@wfu.edu

***Abstract*—Human Activity Recognition (HAR) is crucial for preventive healthcare, in assited living settings, fall prevention, and real-time monitoring of senior citizens. Using the HAR70+ dataset, which consists of dual-sensor accelerometer recordings from older persons, including those who use walking assistance, this work compares three advanced sequence modeling architectures: Mamba, iTransformer, and PatchTST. Activity changes, sensor noise, and small transitions like shuffling or standing make the HAR70+ dataset extremely difficult to use. We set out to assess the robustness and temporal modeling capabilities of these most recent algorithms for patient-centric activity classification. With an F1-score of 91.96 ± 0.02%, the results show that Mamba, a selective state space model, performs noticeably better than the transformer-based baselines, better capturing complex activity patterns and long-term relationships. iTransformer and PatchTST, on the other hand, demonstrated shortcomings in traditional attention-based or patch-wise modeling techniques under the low-frequency, high-variability circumstances characteristic of geriatric movement data, with F1-scores of 41.47 ± 0.03% and 60.34 ± 0.03%, respectively. The performance difference highlights how well Mamba handles non-stationary sensor data and variable temporal patterns, which are common in actual elderly HAR scenarios. For HAR tasks involving clinically relevant populations and complicated locomotor patterns, our study offers evidence in favor of using modern state space models rather than transformer variations.**

***Index Terms*—human activity recognition, mamba, transformer, patchTST, wearable sensors data, time series classification**

I. Introduction

This section gives a general description of the human activity recignition setting, emphasizing the difficulties of keeping an eye on individuals who need assistance due to their age, mobility, or medical concerns, yoga pose, smart home and work place monitoring etc . It believes how early solutions, such as manual or labor work, are reactive and frequently insufficient with the advancement of technology. The main research questions that guide the study to make a comparable study between transformer models on sensor data in this section, along with the research objectives intended to address these issues through the creation of a Transformer based monitoring systems.

II. Background

Human Activity Recognition are designed to provide care and support for individuals who need help with daily activities. These facilities cater to a diverse group of residents, including the elderly who may face age-related challenges such as reduced mobility, cognitive decline, or chronic health conditions, people recovering from surgeries, people with degenerative diseases like Parkinson's or multiple sclerosis, or individuals with severe arthritis [15].

Furthermore, Human Activity Recognition is also an option for people with disabilities or chronic conditions that limit their ability to live independently but who do not need constant medical care. These residents benefit from personalized care plans, social activities, and a supportive environment that promotes safety, independence, and well-being, while offering peace of mind for their families and caregivers [12].

Over the years, the scope of HAR extends to a multitude of domains, HAR may help identify mental health issues such as depression, sleep difficulties, or stress. HAR makes it possible to detect suspicious activity in public or restricted areas, which is useful in security and monitoring. HAR is also increasingly important in the automobile industry for detecting driver weariness, in sports and fitness for performance analysis and posture correction, and in workplace safety for tracking dangerous behavior and fatigue in industrial environments. Since it can modify robotic reactions in real time according to human context, HAR is essential to human–robot interaction. However, such applications are becoming more and more feasible for embedded and edge devices with the advent of transformer-based models that are lightweight and protect privacy, ensure reliability and accurcay.

III. Problem Statement

In Human Activity Recognition environments, residents often require close supervision due to mobility and health issues, increasing the risk of falls and other accidents in a working environment, hospital, home or in a public or private space. Traditional solutions, such as assigning dedicated workers, caregivers, installing alarm systems, or conducting routine patrols, have limitations. These approaches cannot provide real-time and continuous monitoring of all residents, leading to potential gaps in reliability and accuracy,

A. Challenges

The key challenges these early solutions are:

Identify applicable funding agency here. If none, delete this.

- Absence of real-time monitoring: Current systems only react to events after they happen, making them reactive rather than proactive, such as Smart Home.
- Caregiver strain: Due to a lack of human resources, caregivers are unable to keep an eye on every patient/elderly at once.
- Hazards: Unattended activities, such falls or dangerous motions, might cause harm or death in a working or surveillance setting.

The objective of this research is to develop an automated system, based on Transformer based models and Human Activity Recognition, capable of recognizing activities from sensor data.

B. Objectives

This study seeks to answer the following research questions as objectives:

- How can we use Transfomer models to accurately identify people activities and predict transitions between activities?
- Can change point detection techniques be applied effectively to monitor postural and behavioral changes in real time?
- How can posture and location data be used to enhance the accuracy of monitoring people?

IV. Related Works

This section reviews the relevant literature on human activity detection and segmentation, focusing on the application of Deep Learning models in human activity recognition. It also explores several change point detection techniques used to identify transitions between activities, and discusses prediction models that enhance the accuracy and timeliness of recognizing abnormal behaviors. The review highlights existing methods, their limitations, and how this research builds upon or addresses these gaps in the context of Human Activity Recognition environments.

A. Scope of HAR

In study [17] This research combined deep convolutional networks (for local feature extraction) and transformer encoders (for global temporal modeling). Significant contribution: Eight transformation techniques applied (e.g., noise, scaling, flipping, rotation) to increase sample diversity. The main advantage is that the DCTCSS model can effectively reduce the dependence on large amounts of labeled data while achieving competitive activity recognition performance. However, As for limitations, the recognition effect on similar activities can further be improved, as well as requires dataset-specific tuning of augmentation strategies.

In another study [18],This research analyzed the relations between 8 architectures accuracies, signal waveforms, signals correlation, sampling rate, exercise duration, and window size. Introduces a hybrid data collection setup using IMUs (accelerometer, gyroscope, magnetometer) and MAS (camera + visual markers). However, Focused on arm-only exercises, excludes full-body or dynamic daily activities, as well as only 10 participants was present, Only a unidirectional encoder was used; bidirectional models may yield better results.

In study [19], This research extracted a diverse set of features to capture a wide range of patterns from the sensor data. The extracted features were used as input to Transformer model, constructed using a fewer number of layers to derive long-range dependencies leading to improved accuracy. The proposed technique was evaluated on three datasets: WISDM, PAMAP2, and UCI HAR. However, The drawback of this approach is that it requires domain expertise to extract and combine a large number of diverse features, as well as class imbalance of dataset can make huge problem for the settings.

Another study by Aminikhanghahi and Cook [8] proposed a method for real-time segmentation of daily human activities using sensor data, focusing on detecting transitions between activities. They introduce two unsupervised algorithms: Relative Unconstrained Least Squares Importance Fitting (RuLSIF) and Bayesian Change Point Detection (BCPD). These methods identify change points in streaming data without prior labeling. Using smart home data sets, both algorithms outperform baseline methods, with RuLSIF showing superior performance. The study demonstrates that change point detection improves the segmentation of continuous activities, helping applications such as healthcare monitoring.

In study [14], Hossen et al. introduced a novel approach for automatic segmentation of human activities using skeleton data through the application of an autoencoder model. Their method aims to address the challenges posed by manual segmentation in human activity recognition, which is both labor-intensive and error-prone. By employing an auto-encoder, the model effectively extracts features from the skeleton data and identifies transition points between different activities, relying on minimal reconstruction errors to detect breakpoints. Additionally, a dynamic thresholding mechanism is incorporated to adaptively adjust the segmentation threshold based on the activity's dynamics, thus improving segmentation accuracy. The method was rigorously evaluated on several publicly available datasets and demonstrated impressive performance by achieving an average precision, recall, and F1-score values of 3.6, 90%, 87%, and 88%, respectively.

In [10] Mutegeki et al. proposed a CNN-LSTM method to contribute both in spatial and temporal terms, lowered the model's complexity and eliminated the requirement for sophisticated feature engineering, along with increasing the predicted accuracy of human activities from raw data. They worked on two datasets and in their own research Ispl dataset they achieved 99% precision and in the UC Irvine HAR public dataset benchmark they acquired 92%.

B. On HAR70+ Dataset

ABBAS et al.(2024) [1] Advancing Healthcare and Elderly Activity Recognition: Active Machine and Deep Learning for Fine-Grained Heterogeneity Activity Recognition This research explored a diverse array of ML algorithms, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), K-Nearest Neighbors (KNN),

Stochastic Gradient Descent (SGB) and DL methods such as Deep Neural Networks (DNN) and Long Short-Term Memory networks (LSTM) for experimentation. This research trained models on 7 activities: walking, shuffling, climbing stairs (up and down), standing, sitting, and lying down, and 4 activities separately: standing, sitting, walking, and lying down, using the same classifiers integrated with Active Learning. However, F1 score for different activities and for different models was different, in general shuffling, and climbing stairs up and down activities were not classified correctly.
Logacjov et al. (2024) [2] Self-supervised learning with randomized cross-sensor masked reconstruction for human activity recognition This research introduced a new auxiliary task, randomized cross-sensor masked reconstruction (RCSMR), for SSL, and they pre-trained a transformer encoder on the large-scale HUNT4 dataset with RCSMR. The model RCSMR achieved an average F1-score of 74.03%, outperforming other supervised baselines (e.g. DeepConvLSTM, XGB, MLP) However, their model performed poor on those dataset where there is no fixed sensor position and orientation (e.g. MobiAct dataset)

Logacjov et al. (2024) [6] SelfPAB: large-scale pre-training on accelerometer data for human activity recognition This research explored a self-supervised learning method that uses a masked reconstruction objective inspired by the TERA model from speech recognition. Used several preprocessing techniques Time-Domain and Frequency Domain Masking, and Normalization, sampling rate 50 Hz with STFT. Achieved nearly 78.5% F1-score for the HAR70+ dataset. However, Pre-trained a Transformer Encoder on up to 100,000 hours of unlabeled dual-accelerometer data, which is computationally expensive.
Nematallah et al.(2024) [3] Quantitative Analysis of Mother Wavelet Function Selection for Wearable Sensors-Based Human Activity Recognition This research investigated how different mother wavelet (MW) functions affect human activity recognition (HAR) performance using wearable sensor (accelerometer) data, and to identify the optimal wavelet for feature extraction. They used wavelet packet transform (WPT) and a novel energy-to-Shannon entropy ratio to select optimal mother wavelets and compared performance using Decision Tree (DT) and Support Vector Machine (SVM) classifiers. They worked on Eight public datasets: WISDM, HARSense, HARTH, HAR70+, MHEALTH, PAMAP2, REALDISP, DaLiAc. All signals were downsampled to 50 Hz and segmented into 2.5-second windows. However, The choice of mother wavelet significantly affects HAR accuracy.
Ustad et al.(2024) [4] Validation of an Activity Type Recognition Model Classifying Daily Physical Behavior in Older Adults: The HAR70+ Model This research evaluated the performance of an existing activity type recognition ML model (HARTH), based on data from healthy young adults for classifying older adults. Compared the performance with a ML model (HAR70+) that included training data from older adults, using Leave-One-Subject-Out (LOSO) cross-validation. Evaluated the ML models on older adults with and without walking aid. The accuracy of overall activities was 94% and without walking aids 93%, however their model best F1 score was found for sitting and laying.

V. Motivation

As the demand for care in Human Activity Recognition facilities grows, traditional monitoring methods such as alarm systems and routine checks are becoming inadequate, often leading to delayed responses in critical situations such as falls or health or professional emergencies. This proposal is motivated by the need for a proactive, scalable solution that leverages Transformer models and sensor technology to track in general people activities to detect normal or abnormal activities. By improving model accuracy and reliability, this project aims to recognize human activities in any environments.

A. Preliminary Questions

- How can we identify and predict patient/elderly activities accurately?
- How can detect activities from sensor data to improve patient/elderly monitoring?
- Can we effectively detect transitions between activities?

VI. Methodology

This section outlines the study's methodology, including data collection, the development of Transformer Models for real-time activity recognition. In this study we plan to use multiple Human Activity Recognition (HAR) datasets to train transformer based models (Mamba, iTransformer, PatchTST) for activity classification and perform comparative performance analysis.

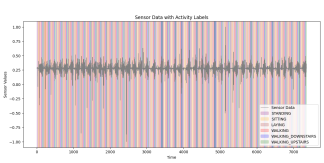
A. Data Collection

1) UCI HAR: So far we performed EDA of UCI Human Activity Recognition (HAR) dataset [**?**], created from recordings of 30 participants performing six daily activities (WALKING, WALKING UPSTAIRS, WALKING DOWNSTAIRS, SITTING, STANDING, LAYING) while carrying a waist-mounted smartphone equipped with accelerometer and gyroscope sensors. The participants, aged between 19 and 48 years, performed these activities while 3-axial linear acceleration and angular velocity data were recorded at 50Hz. The experiments were video-recorded for manual activity labeling, and the dataset was divided randomly into training (70%) and test (30%) sets. Pre-processing included noise filtering, segmenting the data into 2.56-second sliding windows with 50% overlap (128 readings per window), and separating body and gravitational acceleration using a Butterworth low-pass filter with a 0.3 Hz cutoff. From each window, features were extracted in both time and frequency domains, yielding a 561-feature vector for each record. Each entry in the dataset includes triaxial acceleration (total and body acceleration), triaxial angular velocity, the 561-feature vector, the activity label, and an identifier for the subject performing the activity.

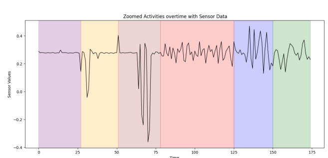
2) HAR70+: Every CSV file (501.csv, for example) represents a subject, and includes accelerometer time-series data, Timestamps, 6-axis (3 from lower back, 3 from thigh), and activity labels are the columns, 50 Hz is the sampling rate. The dataset was collected over 18 older adult participants wearing two 3-axial Axivity AX3 accelerometers for around 40 minutes in a semi-structured free-living setting. One sensor was attached to the right front thigh and the other to the lower back, containing six features in the dataset.Containing Seven Activities with assigned labels:
1: Walking
3: Shuffling(Movement resembling walking, but with short, dragging steps and reduced foot clearance; Visual: Feet do not lift completely off the ground.)
4: Stairs (ascending)
5: Stairs (descending)
6: Standing
7: Sitting
8: Lying

B. Data Exploration

In the Fig.1(a), the overall training data was plot and for more deeper insights for the consecutive six activities general pattern analysis have been shown in Fig. 1(b)



((a)) EDA1: Train Data Analyis.



((b)) EDA2: Data Analysis for Six Activities.

Fig. 1: Exploratory Data Analysis for EDA1 and EDA2.

Among the six activities, for the three static activities: Sitting, Standing, Lying, they have similar distributions, and for the dynamic activities: Walking, Walking Up and Wallking Down have similar data distribution.

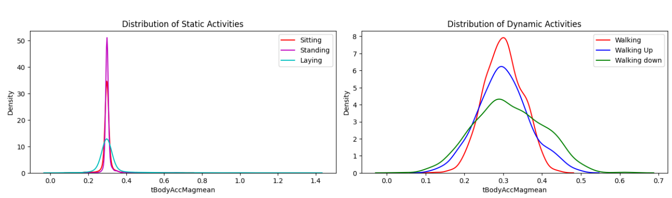
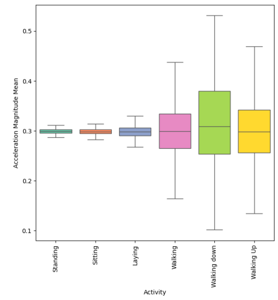
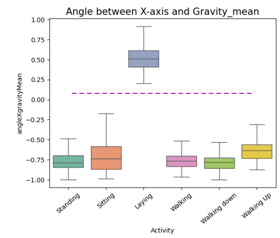


Fig. 2: Data Distribution of Static and Dynamic Activities.

In Fig.3, acceleration magnitude mean for six activities and the gravity mean around the six activities show that for static activity Laying does not fluctuate much and stayed in positive section. In Fig.7, the segmented activity data is shown to better understand how the two different activities may overlap in the segment of the data.



((a)) Acceleration Magnitude Mean of Six Activities.



((b)) Angular Gravity mean.

Fig. 3: Data Analysis on Acceleration Magnitude and Gravity Mean.

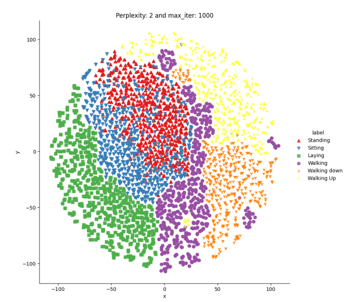


Fig. 4: Segment of activities.

C. Architecture

Transformer-based models have shown superior performance in almost every domain and tasks. Various architectures are being introduced to farther improve efficiency and performance. This study will focus on three special architectures that were optimized for handling time series data. Following are short descriptions of these architectures.

1) Mamba: Mamba is a state space model (SSM) designed for efficient sequence modeling. Unlike transformers, which rely on attention mechanisms, Mamba uses structured state space layers to capture long-range dependencies linearly in time and memory, making it ideal for long sequences and sensor data [20].

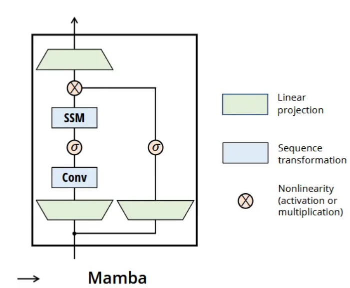


Fig. 5: Mamba architecture.

2) iTransformer: Inverted Transformer (iTransformer) is a time series forecasting model based on transformer architecture. It applies attention to inverted dimension which is feature dimension unlike transformer that applies on time dimension. iTransformer regards independent time series as variate tokens to capture multivariate correlations by attention and utilize layernorm and feed-forward networks to learn series representations.

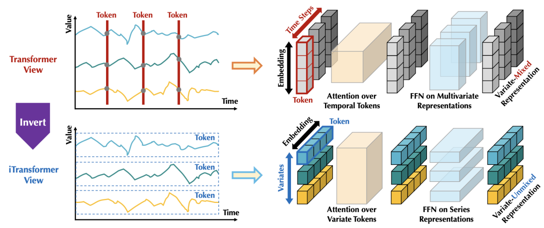


Fig. 6: Inverted Transformer (iTransformer) for Multivariate Time Series Forecasting .

3) PatchTST: Patch Time Series Transformer (PatchTST) is a transformer based model that exhibited improved efficiency and performance on time-series forecasting and classification. The idea is to divide the time series into patches that are flattened and linear projection is applied to get patch embeddings. Then these patch embeddings are feed into the standard transformer encoder.

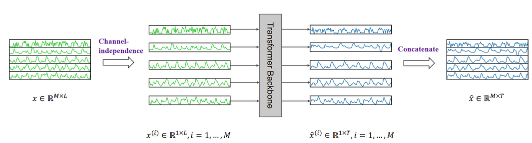


Fig. 7: Patch Time Series Transformer.

VII. Performance of Models and Discussions

A. Mamba

The Mamba model performs exceptionally well in classification across all activity classes when tested on the HAR70+ dataset. All seven activity classes have Area Under the Curve (AUC) values around or exactly 1.0000, according to the multi-class ROC curve, suggesting nearly flawless discrimination. In particular, all other classes, including Classes 2 through 6, obtained flawless AUC values of 1.0000, although Class 0 (the dominant activity) obtained an AUC of 0.9998 shown in 8. This illustrates how well the model detects complex activities in an older population with high sensitivity and specificity.

- Accuracy: 99.26 ± 0.00%
- Precision: 93.02 ± 0.02%
- Recall: 91.37 ± 0.01%
- F1 Score: 91.96 ± 0.02%

Class 0 predictions were made with a very high degree of accuracy (3045 out of 3054).With 1482 and 625 accurately predicted examples, respectively, Classes 5 and 6 likewise demonstrated nearly flawless classification. Class 1 shows some misclassification into Classes 0 and 4, and Class 2 shows a few scattered errors, possibly as a result of labeling noise or motion similarity shown in 9.

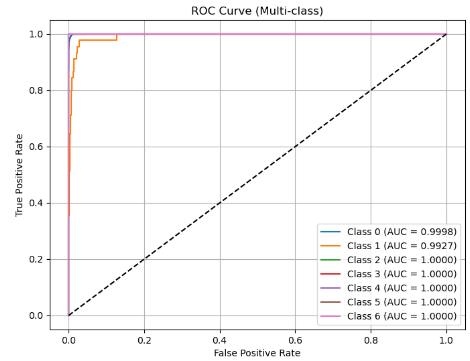


Fig. 8: Mamba: ROC.

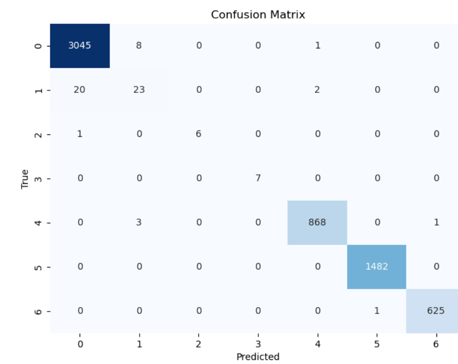


Fig. 9: Mamba: Confusion Matrix

B. itransformer

On the HAR70+ dataset, the iTransformer model performed somewhat well, with inconsistent outcomes for various activity classifications. Class 0 (0.9954) has a comparatively high AUC on the multi-class ROC curve, suggesting that dominant or frequent behaviors are reliably detected. Performance declines for other classes, though, including Class 3 (AUC = 0.6367) and Class 6 (AUC = 0.8686), indicating difficulties differentiating less frequent or delicate movements shown in 10.

- Accuracy: 77.78 ± 0.00%
- Precision: 43.18 ± 0.04%
- Recall: 40.75 ± 0.03%
- F1 Score: 41.47 ± 0.03%

Significant class confusion is also shown by the confusion matrix in classes related to overlapping or transitional motions. For instance: There is significant misclassification of Class 5 into Classes 0, 4, and 6. With 188 cases incorrectly classified as Class 5 and 69 as Class 6, Class 4 also exhibits leaking into neighboring classes shown in 11. Nearly all cases of Class 1 (shuffling) are incorrectly classified as other activities, resulting in a serious underclassification.

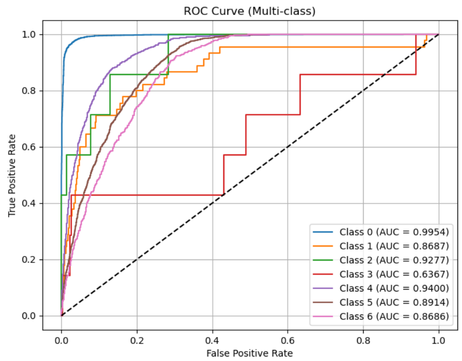


Fig. 10: iTransformer: ROC.

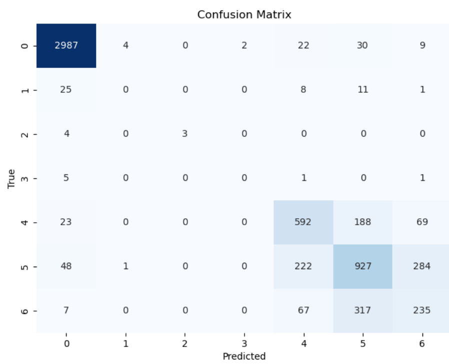


Fig. 11: iTransformer: Confusion Matrix

C. PatchTST

With a significant improvement over iTransformer but still lagging behind Mamba, the PatchTST model demonstrated moderate to strong performance in identifying patient activities from the HAR70+ dataset. Strong performance is shown by the multi-class ROC curve in important classes including Class 0 (AUC = 0.9986), Class 1 (0.9595), and Class 6 (0.9945). With AUCs ranging from 0.8657 to 0.9836, in 12 other classes likewise exhibit substantial separability, indicating that PatchTST's patch-wise representation approach may successfully capture mid- to long-range temporal correlations.

- Accuracy: 91.78 ± 0.01%
- Precision: 69.93 ± 0.05%
- Recall: 58.14 ± 0.03%
- F1 Score: 60.34 ± 0.03%

According to the confusion matrix, some misclassification still occurs between adjacent or semantically related activities, even if PatchTST does a good job of capturing the main activity classes (Class 0, for example, has 3015 correct out of 3054). Take, for example: Class 5 (158 samples) is commonly mistaken for Class 4. Class 6 influences Class 5 (123 samples) shown in

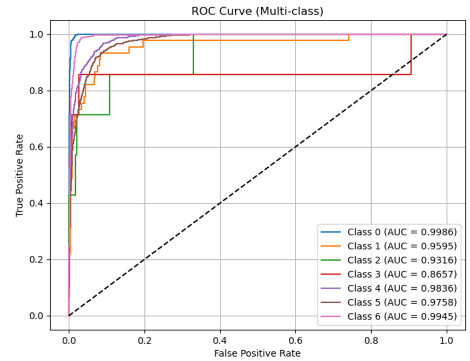


Fig. 12: PatchTST: ROC.

TABLE I: F1-Score Comparison on HAR70+ Dataset Across Models

Author	**Model Used**	**F1-Score (HAR70+)**
(SelfPAB) Logacjov et al. (2024 [6])	Transformer Encoder + MLP	78.5%
(RCSMR) Logacjov et al. (2024) [2]	MonoSelfPAB Transformer	74.03%
Nematallah et al. (2024) [3]	Wavelet Transform + Entropy Ratio	84%
ABBAS et al. (2024 [1])	Diverse ML Models	66%
Our Experiment (2025)[a]	**Mamba**	**91.96 ± 0.02%**
	iTransformer	41.47 ± 0.03%
	PatchTST	60.34 ± 0.03%

[a]Performed on HAR70+ with 50Hz dual-sensor accelerometer data.

VIII. Conclusion

This paper compares three advanced time-series models—Mamba, iTransformer, and PatchTST—on the HAR70+ dataset, which is intended for dual wearable accelerometer monitoring of elderly patients. The Mamba model clearly outperformed both iTransformer (41.47±0.03%) and PatchTST (60.34±0.03%), as evidenced by the testing findings, which showed an F1-score of 91.96±0.02%. Its capacity to effectively handle long-range temporal dependencies and adjust to the noisy, non-stationary character of real-world sensor data in elderly populations is responsible for this performance advantage.
On the other hand, standard transformer variations' poor performance highlights their shortcomings in simulating irregular transitions and small changes in activity that are typical of older persons, including shuffling or sit-to-stand occurrences. These results imply that state-space-based architectures, such as Mamba, provide a stable and expandable basis for developing context-aware, real-time activity identification systems in assistive and therapeutic settings. Future research will examine real-world implementation in assisted living facilities, integration with multi-modal sensor streams, and model generalization across unseen participants.

IX. Future Work

Future research will concentrate on expanding the current findings in a number of important ways. In order to examine cross-domain generalization and flexibility across various sensor setups and activity methods, the performance of the assessed models—including Mamba, iTransformer, and PatchTST—will first be further tested on other benchmark datasets, such as HUNT4, HARTH, and WISDM. Second, the recently suggested RLinear model—which adds recurrent linear layers for effective time-series learning—will be put into practice and compared. To guarantee the best possible setup for every dataset, a thorough hyperparameter tuning process will be applied to every model. Third, to allow for a more thorough interpretation of our findings, a comprehensive comparison analysis will be carried out, combining findings from all recent research that make use of the four datasets discussed in this study. Finally, to evaluate these models' potential for real-world deployment, we want to test their adaptability and transferability across completely diverse sensor combinations and unexplored datasets. These expansions will offer more profound understandings of the clinical significance, generalizability, and scalability of modern time-series models for identifying human activity in both healthy and at-risk populations.

References

[1] S. Abbas *et al.*, "Advancing healthcare and elderly activity recognition: Active machine and deep learning for fine-grained heterogeneity activity recognition," *IEEE Access*, 2024.

[2] A. Logacjov and K. Bach, "Self-supervised learning with randomized cross-sensor masked reconstruction for human activity recognition," *Eng. Appl. Artif. Intell.*, vol. 128, p. 107478, 2024.

[3] H. Nematallah and S. Rajan, "Quantitative analysis of mother wavelet function selection for wearable sensors-based human activity recognition," *Sensors*, vol. 24, no. 7, p. 2119, 2024.

[4] A. Ustad *et al.*, "Validation of an activity type recognition model classifying daily physical behavior in older adults: the HAR70+ model," *Sensors*, vol. 23, no. 5, p. 2368, 2023.

[5] F. Luo *et al.*, "ActivityMamba: a CNN-Mamba hybrid neural network for efficient human activity recognition," *IEEE Trans. Mobile Comput.*, 2025.

[6] A. Logacjov *et al.*, "SelfPAB: large-scale pre-training on accelerometer data for human activity recognition," *Appl. Intell.*, vol. 54, no. 6, pp. 4545–4563, 2024.

[7] "Human Activity Recognition Using Smartphones," UCI Machine Learning Repository. [Online]. Available: https://archive.ics.uci.edu/dataset/240/human+activity+recognition+using+smartphones. [Accessed: Apr. 6, 2025].

[8] S. Aminikhanghahi and D. J. Cook, "Using change point detection to automate daily activity segmentation," in *Proc. IEEE Int. Conf. PerCom Workshops*, 2017, pp. 262–267.

[9] A. Cooper and S. Meyn, "Reinforcement Learning Design for Quickest Change Detection," *arXiv preprint arXiv:2403.14109*, 2024.

[10] R. Mutegeki and D. S. Han, "A CNN-LSTM approach to human activity recognition," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, 2020, pp. 362–366.

[11] S. M. Soomar and Z. Dhalla, "[RETRACTED ARTICLE] Injuries and outcomes resulting due to falls in elderly patients...," *BMC Emerg. Med.*, vol. 23, no. 1, p. 14, 2023.

[12] S. Zimmerman *et al.*, "The imperative to reimagine assisted living," *J. Amer. Med. Directors Assoc.*, vol. 23, no. 2, pp. 225–234, 2022.

[13] C. Harrington *et al.*, "Appropriate nurse staffing levels for US nursing homes," *Health Serv. Insights*, vol. 13, pp. 1–8, 2020.

[14] M. A. Hossen, A. G. Naim, and P. E. Abas, "Deep Learning for Skeleton-Based Human Activity Segmentation: An Autoencoder Approach," *Technologies*, vol. 12, no. 7, p. 96, 2024.

[15] I. D. Cameron *et al.*, "Interventions for preventing falls in older people in care facilities and hospitals," *Cochrane Database Syst. Rev.*, no. 9, 2018.

[16] Y. Weng and L. Liu, "A sequence anomaly detection approach based on isolation forest algorithm for time-series," in *Proc. ACM ICS Workshops*, 2018, pp. 198–207.

[17] Y. Sun *et al.*, "Efficient human activity recognition: A deep convolutional transformer-based contrastive self-supervised approach using wearable sensors," *Eng. Appl. Artif. Intell.*, vol. 135, p. 108705, 2024.

[18] M. F. Trujillo-Guerrero *et al.*, "Accuracy comparison of CNN, LSTM, and transformer for activity recognition using IMU and visual markers," *IEEE Access*, vol. 11, pp. 106650–106669, 2023.

[19] O. Saidani *et al.*, "An efficient human activity recognition using hybrid features and transformer model," *IEEE Access*, vol. 11, pp. 101373–101386, 2023.

[20] Y. Zhang and J. Wang, "Human behavior recognition based on sparse transformer with channel attention," *Front. Physiol.*, vol. 14, p. 1239453, 2023.

[21] X. Li, Y. Zhou, and Z. Wang, "A lightweight hybrid vision transformer network for radar-based human activity recognition," *Sci. Rep.*, vol. 13, no. 1, p. 17996, 2023.

[22] S. Huan *et al.*, "A lightweight hybrid vision transformer network for radar-based human activity recognition," *Sci. Rep.*, vol. 13, no. 1, p. 17996, 2023.

[23] K. Cao and M. Wang, "Human behavior recognition based on sparse transformer with channel attention mechanism," *Front. Physiol.*, vol. 14, p. 1239453, 2023.

[24] Z. Wang *et al.*, "Is Mamba effective for time series forecasting?," *Neurocomputing*, vol. 619, p. 129178, 2025.