

A Transformer-Based Hybrid Model for Human Activity Recognition Using Smart Home Environmental Sensor Data

Ronak Fatahi¹ and Fatemeh Sadat Lesani²

1. M.Sc., Faculty of Electrical and Computer Engineering, Shahab Danesh University, Qom, Iran. Email: rfatahi@shdu.ac.ir
2. Corresponding Author, Assistant Professor, Department of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran. Email: lesani@qut.ac.ir

Article Info	ABSTRACT
Article type: Research Article Article history: Received 16 February 2025 Received in revised form 23 May 2025 Accepted 4 August 2025 Published online 1 December 2025 Keywords: Activity Recognition, Deep Learning, Environmental Sensor, Smart Home, Transformer, GRU.	With the rapid expansion of smart homes, accurate and automated Human Activity Recognition (HAR) has become one of the key challenges in the field of Artificial Intelligence and the Internet of Things (IoT). This technology plays a crucial role in areas such as elderly care, health monitoring, and enhancing smart home security. In the past, traditional methods were used for activity recognition, but they had several limitations. However, with the emergence of deep learning techniques, these limitations have been addressed, leading to improved accuracy and performance of models. In this research, a deep learning-based approach is introduced for HAR, leveraging Transformer and Gated Recurrent Unit (GRU) models. The Transformer model, with its multi-head attention mechanism, can analyze long-term relationships within sensor data and identify behavioral patterns with high accuracy. Meanwhile, GRU significantly enhances activity recognition accuracy due to its ability to learn temporal patterns effectively. Evaluation results show that the model achieves an accuracy of 93.17% on the Aruba dataset and 74.38% on the Milan dataset, demonstrating the proposed model's strong generalization ability and effectiveness in recognizing behavioral patterns.
Cite this article: Fatahi, R. & et al, (2025)., A Transformer-Based Hybrid Model for Human Activity Recognition Using Smart Home Environmental Sensor Data. <i>Enginiiring Management and Soft Computing</i>, 11 (2). 107-141. DOI: https://doi.org/10.22091/jemsc.2025.12363.1257	
	© Fatahi and Lesani (2025) DOI: https://doi.org/10.22091/jemsc.2025.12363.1257
Publisher: University of Qom	

1) INTRODUCTION

Smart homes play a significant role in improving quality of life, security, and health monitoring. Automatic human activity recognition (HAR) [1] is a critical challenge in this field, with applications in elderly care [2] and medical monitoring. HAR systems process data from environmental sensors such as motion, temperature, and door sensors to identify user behaviour patterns. Traditional machine learning methods rely on manually engineered features and have limitations, whereas deep learning models [3], like Recurrent Neural Networks (RNNs), perform better by automatically extracting features. GRU and Transformer models are among the most advanced deep learning architectures for sequential data [4]. GRU efficiently processes long sequences by preserving important information and preventing gradient vanishing [5], while the Transformer extracts long-term dependencies without using recurrent memory, enabling parallel processing.

This research presents a hybrid approach combining Transformer and GRU for human activity recognition and no-activity detection in smart homes. The combination of these two models enhances recognition accuracy and speed while reducing the limitations of traditional methods. The overlap of Transformer's self-attention and GRU's short-term memory effectively extracts temporal dependencies and improves sequential data analysis.

2) Solution MethOd

The proposed architecture is a combination of Bi-GRU and Transformer, which processes data in parallel. Three Bi-GRU layers are used to model both short-term and long-term temporal dependencies, while the Transformer attention layer aids in parallel data processing and extracting long-term dependencies. The features extracted from these two parts are combined using concatenation or weighted averaging and are then passed to a feed-forward neural network to identify human activities and detect inactivity.

This model has several advantages, including improved accuracy and processing speed, combining short-term and long-term temporal dependencies, and high flexibility in processing sequential data. This approach is evaluated using metrics such as accuracy, precision, recall, F1 score, and confusion matrix, showing good performance in Human Activity Recognition (HAR). The proposed architecture consists of three main parts: data preprocessing, sensor data processing and feature extraction, and parallel processing using Bi-GRU and Transformer to model temporal dependencies and long-term relationships. After combining the results from these two parts, they are used for activity recognition. In this research, data related to "inactivity" is also included in the dataset, which helps in more accurately identifying idle times and resident behaviors.

To evaluate the performance of the proposed model and compare it with other deep learning methods, baseline models such as TCN, Transformer, Bi-GRU, LSTM, RNN, GRU, and CNN were examined. These models are crucial for human activity recognition due to their ability to learn temporal dependencies. Hybrid models such as Transformer-GRU and Transformer-TCN were designed to improve long-term data dependencies. The evaluation results show that TCN, Transformer, and Bi-GRU performed better than other models. The TCN model demonstrated the best performance in processing time-series data and feature extraction. The Transformer model also performed well due to its capability to process long-term dependencies. Furthermore, the Bi-GRU model, by enhancing temporal feature learning compared to the regular GRU, achieved better accuracy. Finally, hybrid models using the Transformer architecture significantly improved human activity recognition. The Bi-GRU-Transformer model provided the best performance on the Aruba and Milan datasets. These results suggest that combining different architectures with Transformer can lead to significant improvements in human activity recognition in smart environments.

Transformer-based hybrid models (such as Transformer-GRU and Transformer-TCN) aimed to enhance accuracy and efficiency by combining the strengths of different models. Among these, the Bi-GRU-Transformer model, with an accuracy of 93.17% in Aruba and 74.38% in Milan, along with superior precision and F1 scores, achieved the best performance. Overall, the findings indicate that the

use of hybrid architectures, especially the combination of Bi-GRU with Transformer, offers better performance in human activity recognition, and data quality plays a significant role in the final results.

2-1) Equations

The Transformer model, introduced by Vaswani et al. in 2017, is an advanced architecture in deep learning designed for processing sequential data such as text, speech, and sensor data. This model uses an attention mechanism, making it widely applicable in sequence-based learning tasks. Unlike RNNs and LSTMs, which process data sequentially, the Transformer processes sequences in parallel, enhancing training speed and the model's ability to learn long-term dependencies.

The Transformer model consists of two main parts: the encoder and the decoder. The encoder is responsible for converting the input sequence into a vector representation and is composed of two main sub-layers: a multi-head attention mechanism for extracting temporal dependencies, and a feed-forward neural network to refine the features. The decoder generates the output sequence using the encoder's output and new information, and consists of three main sub-layers: multi-head attention for the decoder, encoder-decoder attention to examine dependencies between the input and output sequences, and a feed-forward neural network for processing the final features. The attention mechanism is the core of the Transformer model. The attention calculation is performed as follows:

$$V \cdot \left(\frac{t_{Q \cdot K}}{\sqrt{k^d}} \right) \text{Softmax} = \text{Attention}(Q, K, V) \quad (1)$$

The GRU model is a type of RNN designed to process sequential data such as sensor signals and time series. This model is widely used due to its simpler structure and better ability to learn long-term sequences, particularly in processing smart home sensor data. GRU addresses gradient instability and long-term information retention issues with two gates (update gate and reset gate), reducing computational complexity.

The reset gate determines how much of the previous state information should be forgotten, while the update gate specifies how much of the previous state information should be transferred to the new state. The candidate state represents new information produced by combining the input and the previous state. The new state is obtained by combining the previous state with the candidate state according to the update gate. Equation 2 represents the combination of the previous state and the candidate state.

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot h_t \quad (2)$$

3) conclusion

This research demonstrates that combining advanced Transformer models with Bi-GRU can improve the performance of human activity recognition systems in smart environments. The results showed that the proposed hybrid model can effectively enhance identification accuracy and other performance metrics, especially in complex and diverse datasets such as Aruba and Milan. The Bi-GRU-Transformer model performed exceptionally well in managing long-term dependencies and extracting complex features from smart sensor data. These results can contribute to the optimization of smart systems in various domains such as healthcare, security, and well-being. However, challenges such as poor data quality, noise, environmental diversity, and high computational resource requirements exist. Additionally, GRU models with simpler architectures may struggle with identifying long-term patterns. To address these issues, it is recommended to use data augmentation methods, lighter architectures, and transfer learning techniques. Also, using wearable sensor data and images could expand the applications of the models.

In the future, improving scalability, real-time processing, and integrating multimodal data could increase the model's accuracy and applicability in various fields such as health monitoring, smart home security, and anomaly detection. Implementing the model within a federated learning framework could help maintain user security and privacy. Finally, this study showed that the Bi-GRU-Transformer model performed well in human activity recognition systems using smart home sensor data and achieved higher

accuracy compared to baseline models like LSTM and GRU. However, challenges such as data quality and the need for labelled data still remain and should be addressed in future research.

3-1) Tables and Figures

In this experiment, all transformer-based hybrid models consist of three layers from the base models. Table 1 shows the performance of these models, with results indicating significant improvement in human activity recognition using the transformer architecture.

Table 1. The results of applying hybrid models in activity recognition for the Aruba and Milan datasets.

Metric (%)	Model	Dataset	
		Aruba	Milan
Accuracy	LSTM-transformer	91.09	72.17
	TCN-transformer	91.22	68.30
	CNN-transformer	91.09	73.53
	RNN-transformer	90.05	72.70
	gru-transformer	89.98	72.18
	Bi-gru-trans	93.17	74.38
Precision	LSTM-transformer	91.56	69.39
	TCN-transformer	91.59	69.12
	CNN-transformer	90.18	71.26
	RNN-transformer	89.65	70.96
	gru-transformer	90.77	73.61
	Bi-gru-trans	93.42	74.34
Recall	LSTM-transformer	91.09	72.17
	TCN-transformer	91.22	68.30
	CNN-transformer	91.09	73.53
	RNN-transformer	90.05	72.70
	gru-transformer	89.98	72.18
	Bi-gru-trans	93.17	74.38
F1 Score	LSTM-transformer	89.70	69.73
	TCN-transformer	89.67	65.94
	CNN-transformer	85.25	71.33
	RNN-transformer	88.71	70.39
	gru-transformer	88.46	70.05
	Bi-gru-trans	91.59	72.31

The bar charts of the results obtained from this experiment for the two datasets, Aruba and Milan, are shown in Figures 1 and 2.

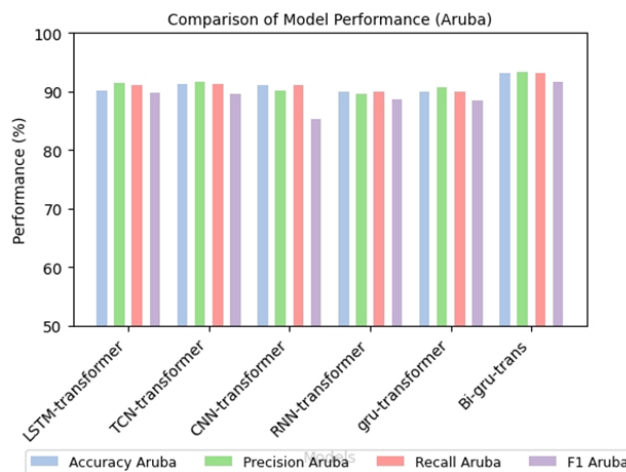


Figure 1: Performance of different models on the Aruba dataset

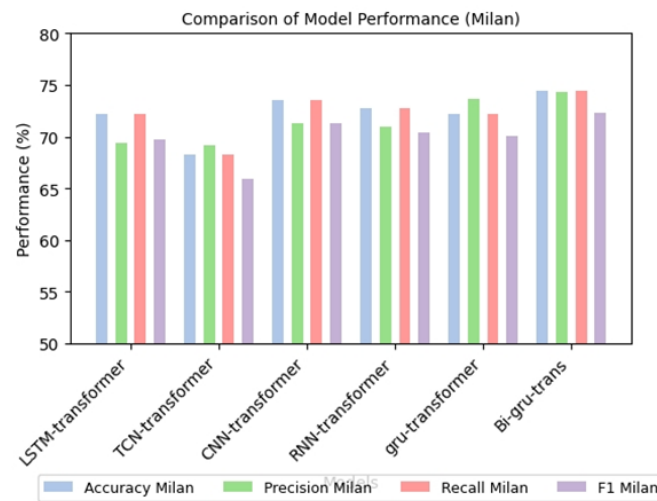


Figure 2: Performance of different models on the Milan dataset

REFERENCES

- [1] N. A. Choudhury and B. Soni, "An Adaptive Batch Size based-CNN-LSTM Framework for Human Activity Recognition in Uncontrolled Environment," *IEEE Transactions on Industrial Informatics*, 2023, doi: <https://doi.org/10.1109/TII.2022.3229522>.
- [2] A. Subasi, K. Khateeb, T. Brahimi, and A. Sarirete, "Human activity recognition using machine learning methods in a smart healthcare environment," in *Innovation in health informatics*: Elsevier, 2020, pp. 123-144.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015/05/01 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [4] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [5] J. Chen, D. Jiang, and Y. Zhang, "A hierarchical bidirectional GRU model with attention for EEG-based emotion classification," *IEEE Access*, vol. 7, pp. 118530-118540, 2019, doi: <https://doi.org/10.1109/ACCESS.2019.2936817>.

مدل ترکیبی مبتنی بر ترنسفورمر برای شناسایی فعالیت‌های انسانی براساس داده‌های حسگرهای محیطی خانه‌های هوشمند

روناک فتاحی^۱ و فاطمه سادات لسانی^۲

۱. کارشناسی ارشد، دانشکده مهندسی برق و کامپیوتر، دانشگاه شهاب دانش، قم، ایران. رایانامه: rfatahi@shdu.ac.ir

۲. نویسنده مسئول، استادیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی قم، قم، ایران. رایانامه: lesani@qut.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی تاریخ دریافت: ۱۴۰۳/۰۴/۰۸ تاریخ بازنگری: ۱۴۰۳/۰۶/۰۱ تاریخ پذیرش: ۱۴۰۴/۰۷/۰۱ تاریخ انتشار: ۱۴۰۴/۰۹/۱۰ کلیدواژه‌ها: الگوریتم اپیوری، داده کاوی، متوسط طول مدت اقامت، مدل‌های خوشه‌بندی.	با گسترش روزافزون خانه‌های هوشمند، شناسایی دقیق و خودکار فعالیت‌های انسانی به یکی از چالش‌های کلیدی در حوزه هوش مصنوعی و اینترنت اشیا تبدیل شده است. این فناوری در حوزه‌هایی نظیر مراقبت از سالمندان، نظارت بر سلامت و ارتقای امنیت خانه‌های هوشمند کاربردهای حیاتی دارد. در این پژوهش، یک روش ترکیبی نوین مبتنی بر یادگیری عمیق برای شناسایی فعالیت‌های انسانی معرفی می‌شود که از مدل‌های ترنسفورمر و واحد بازگشتی دروازه‌ای بهره می‌گیرد. مدل ترنسفورمر با مکانیسم توجه چندسری، توانایی تحلیل روابط طولانی مدت میان داده‌های حسگر را دارد و الگوهای رفتاری را با دقت بیشتری شناسایی می‌کند و واحد بازگشتی دروازه‌ای به دلیل توانایی‌اش در یادگیری الگوهای زمانی به بهبود دقت شناسایی فعالیت‌ها کمک شایانی می‌کند. نتایج ارزیابی‌ها نشان می‌دهد که مدل در مجموعه داده‌های آروبا به دقت ۹۵.۱۹٪ و میلان ۸۰.۰۱٪ دست یافته، که بیانگر توانایی بالای مدل پیشنهادی در تعمیم‌پذیری و شناسایی الگوهای رفتاری است. همچنین، در مقایسات انجام شده با روش‌های مشابه، مدل پیشنهادی بهبود چشمگیری در عملکرد شناسایی فعالیت‌های انسانی از خود نشان داده است.

استناد: فتاحی، روناک و همکاران. (۱۴۰۴). «مدل ترکیبی مبتنی بر ترنسفورمر برای شناسایی فعالیت‌های انسانی براساس داده‌های حسگرهای محیطی

خانه‌های هوشمند». *مدیریت مهندسی و رایانش نرم*، دوره ۱۱ (۲). صص: ۱۴۱-۱۰۷. <https://doi.org/10.22091/jemsc.2025.12363.1257>



© فتاحی و لسانی (۱۴۰۴).

ناشر: دانشگاه قم

۱) مقدمه

خانه‌های هوشمند^۱ به عنوان یکی از انقلابی ترین پیشرفت‌های فناوری، نقش مهمی در بهبود کیفیت زندگی، افزایش امنیت، و نظارت بهداشتی ایفا می کنند. این محیط‌های هوشمند با استفاده از اینترنت اشیا (IoT)^۲ و هوش مصنوعی (AI)^۳ قادر به درک نیازهای کاربران و ارائه پاسخ‌های هوشمندانه هستند. یکی از چالش‌های اساسی در این حوزه، شناسایی دقیق و خودکار فعالیت‌های انسانی (HAR)^۴ است، که می تواند در کاربردهایی مانند مراقبت از سالمندان، نظارت پزشکی، افزایش امنیت و مدیریت انرژی خانه‌های هوشمند تأثیر چشمگیری داشته باشد (Sunil et al., 2021)، (Amozadkheli, 2023)، (Choudhury & Soni, 2023)، (Ghasemi et al., 2020).

سیستم‌های HAR داده‌های دریافتی از حسگرهای محیطی مانند حسگرهای حرکت^۵، دما^۶، و درب را پردازش کرده و الگوهای رفتاری کاربران را شناسایی می کنند. HAR شامل پردازش مجموعه داده‌های بزرگ است که می تواند منجر به هزینه‌های محاسباتی بالا و افزایش زمان آموزش مدل‌ها شود. روش‌های سنتی یادگیری ماشین^۷ در شناسایی فعالیت‌های انسانی از طریق مهندسی ویژگی به صورت دستی و الگوریتم‌های دقیق، عملکرد خوبی دارند؛ اما این روش‌ها به دلیل وابستگی به طراحی دستی و فرآیند انتخاب ویژگی‌ها با محدودیت‌هایی مواجه هستند. در مقابل، مدل‌های یادگیری عمیق مانند شبکه‌های عصبی بازگشتی (RNN)^۸ پتانسیل قابل توجهی در HAR نشان داده‌اند و با استخراج خودکار ویژگی‌های معنادار مستقیماً از داده‌های ورودی، بدون نیاز به مهندسی ویژگی‌های دستی، عملکرد بهتری نسبت به تکنیک‌های سنتی دارند (Wang et al., 2019).

با گسترش یادگیری عمیق^۹ در حوزه یادگیری ماشین و داده‌کاوی^{۱۰}، گرایش به سمت سیستم‌های خودکار و خودآموز بیشتر شده است که این امر منجر به بهبود عملکرد و کاهش وابستگی به طراحی ویژگی‌ها توسط متخصصان شده است. در مدل‌های یادگیری عمیق، فرآیند استخراج ویژگی و ساخت مدل معمولاً به طور هم‌زمان انجام می شود و در مقابل روش‌های سنتی که متکی بر طراحی دستی ویژگی‌ها هستند، به صورت خودکار از طریق شبکه یاد گرفته می شوند. همچنین، شبکه‌های عصبی عمیق می توانند در لایه‌های عمیق خود، نمایش‌های سطح بالایی از داده‌ها را استخراج کنند و این امر آن‌ها را برای شناسایی فعالیت‌های پیچیده مناسب می سازد (Kumar et al., 2024)، (ebrahimi, 2023)، (LeCun et al., 2015). مدل‌های واحد بازگشتی دروازه‌ای (GRU)^{۱۱} و ترنسفورمر^{۱۲} از پیشرفته ترین معماری‌های یادگیری عمیق برای پردازش داده‌های توالی محور هستند. GRU، که نوعی شبکه بازگشتی بهینه شده است، با ساختاری ساده تر نسبت به شبکه عصبی حافظه طولانی کوتاه مدت^{۱۳} (LSTM)، قادر به حفظ اطلاعات مهم در دنباله‌های زمانی طولانی بوده و از مسئله

1 Smart home

2 Internet of Things (IOT)

3 Artificial intelligence

4 Human Activity Recognition (HAR)

5 Action

6 Temperature

7 Machine learning

8 Recurrent neural networks

9 Deep learning

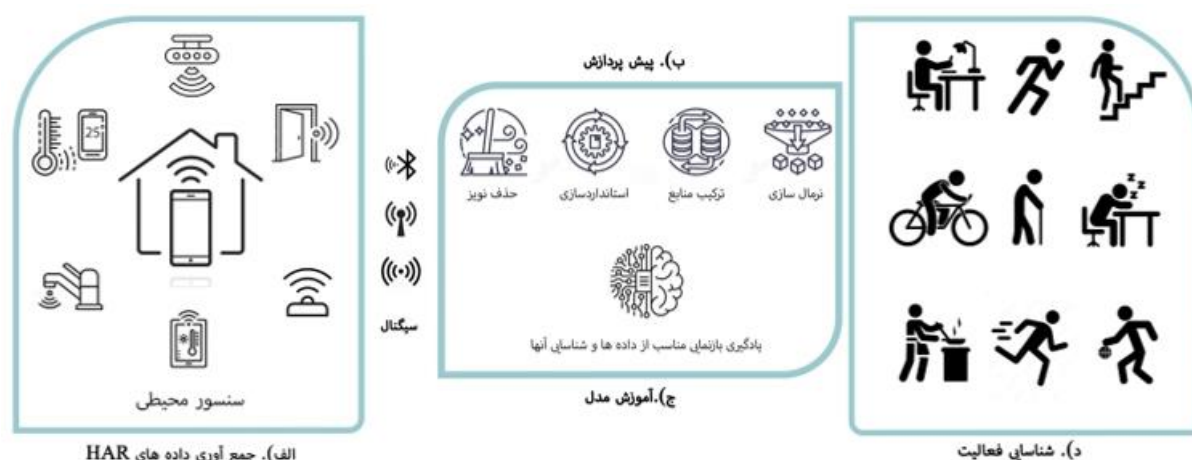
10 Data mining

11 Gated Recurrent Unit

12 Transformers (deep learning architecture)

13 Long Short Term Memory (LSTM)

محو شدن گرادیان^۱ جلوگیری می‌کند. از سوی دیگر، معماری ترنسفورمر که بر پایه مکانیزم خودتوجهی^۲ توسعه یافته است، توانایی پردازش موازی داده‌ها را دارد و وابستگی‌های طولانی‌مدت را بدون استفاده از حافظه بازگشتی استخراج می‌کند. این ویژگی باعث شده است که ترنسفورمر در بسیاری از کاربردهای مرتبط با پردازش زبان طبیعی^۳، بینایی کامپیوتری^۴ و تشخیص فعالیت انسانی عملکرد برجسته‌ای داشته باشد. ترکیب این دو مدل می‌تواند دقت و کارایی سیستم‌های یادگیری عمیق را در تحلیل داده‌های متوالی بهبود ببخشد. شکل ۱ فرآیند کلی HAR را نشان می‌دهد، که شامل جمع‌آوری داده‌های حسگرها، پیش‌پردازش^۵، استخراج ویژگی‌ها^۶ و طبقه‌بندی^۷ فعالیت‌های انسانی با استفاده از مدل‌های یادگیری عمیق است.



شکل ۱. چارچوب شناسایی فعالیت انسانی شامل چهار بخش اصلی است:

(الف) جمع‌آوری داده با استفاده از حسگرهای بینایی، پوشیدنی و محیطی؛ (ب) پیش‌پردازش اولیه شامل نرمال‌سازی و آماده‌سازی داده، بدون انجام استخراج ویژگی دستی؛ (ج) مرحله یادگیری، که در آن فرآیند استخراج ویژگی و یادگیری الگوها به‌طور هم‌زمان و خودکار از طریق شبکه‌های یادگیری عمیق انجام می‌شود؛ و (د) شناسایی فعالیت‌ها بر اساس خروجی مدل آموزش‌دیده.

در این پژوهش، یک روش ترکیبی و نوآورانه مبتنی بر یادگیری عمیق برای شناسایی فعالیت‌های انسانی و تشخیص عدم فعالیت در خانه‌های هوشمند ارائه شده است. این روش با تلفیق هوشمندانه مدل‌های ترنسفورمر و GRU، قابلیت یادگیری روابط طولانی‌مدت را بهبود می‌بخشد و پردازش داده‌های متوالی را به صورت کارآمد انجام می‌دهد. نوآوری این پژوهش در ترکیب مدل ترنسفورمر با واحد بازگشتی دو جهته (Bi-GRU) برای شناسایی فعالیت‌های انسانی در محیط‌های خانه‌های هوشمند است. اگرچه ترکیب ترنسفورمر یا Bi-GRU پیش‌تر در پژوهش‌هایی برای شناسایی فعالیت‌ها استفاده شده است، تمایز اصلی این پژوهش در استفاده از رویکردی منحصر به فرد در مدل‌سازی می‌باشد (Alsarhan et al., 2019).

1 Vanishing Gradient

2 Self-Attention

3 Natural Language Processing (NLP)

4 Computer Vision

5 Preprocessing

6 Feature extraction

7 Classification

به طور خاص، در این روش زمان‌هایی که هیچ فعالیتی انجام نمی‌شود به عنوان یک کلاس مجزا لحاظ شده و مستقیماً به مدل داده می‌شود. این امر باعث می‌شود که مدل بتواند عدم فعالیت را به درستی شناسایی کند و از اشتباه در تشخیص آن به عنوان فعالیت جلوگیری نماید، که در نتیجه دقت کلی مدل در شناسایی انواع فعالیت‌ها افزایش می‌یابد. این رویکرد با بهبود عملکرد مدل در شرایط واقعی، قابلیت تعمیم و کاربردپذیری آن را در محیط‌های هوشمند افزایش می‌دهد. همچنین ترکیب این دو مدل یادگیری عمیق باعث افزایش دقت شناسایی شده و محدودیت‌های روش‌های سنتی را کاهش می‌دهد. هم‌پوشانی قابلیت خودتوجهی مدل ترنسفورمر و GRU، سیستم را قادر می‌سازد وابستگی‌های بلندمدت و ویژگی‌های زمانی را به شکل مؤثری استخراج کند و علاوه بر شناسایی فعالیت‌ها، به تفکیک دقیق وضعیت‌های فاقد فعالیت نیز پردازد.

در ادامه مقاله به صورت زیر سازماندهی شده است: بخش دو به پیشینه تحقیق می‌پردازد، بخش سه روش تحقیق را توضیح می‌دهد، بخش چهار یافته‌ها را ارائه می‌کند و در نهایت، بخش پنج نتیجه‌گیری و پیشنهادات را بیان می‌نماید.

۲) پیشینه تحقیق

در این بخش، ابتدا به معرفی کلی و زمینه‌ای تحقیقات انجام شده در حوزه HAR با استفاده از داده‌های حسگرهای خانه هوشمند پرداخته شده است. سپس، رویکردهای مختلف یادگیری عمیق که در این زمینه کاربرد دارند، از جمله مدل‌های مبتنی بر ترنسفورمر و شبکه‌های GRU به طور ویژه بررسی می‌شوند. معرفی انجام شده بر اساس رویکردهای مهم و مدل‌های شاخص یادگیری عمیق است و هدف آن ارائه مروری منسجم بر پیشرفت‌های اخیر در این حوزه می‌باشد. در سال‌های اخیر، تحقیقات گسترده‌ای در زمینه HAR با استفاده از داده‌های حسگرهای خانه هوشمند انجام شده است. این حسگرها، از جمله حسگرهای دما، نور، فشار و دوربین‌ها، اطلاعات ارزشمندی درباره فعالیت‌های کاربران ارائه می‌دهند. روش‌های یادگیری عمیق، به دلیل توانایی بالا در استخراج ویژگی‌های پیچیده، جایگزین روش‌های سنتی در این حوزه شده‌اند.

برای روشن شدن پیشرفت‌های موجود، ابتدا مدل‌های شاخص یادگیری عمیق در حوزه شناسایی فعالیت‌های انسانی معرفی می‌شوند.

۲-۱) شناسایی فعالیت

انبازهاگان و همکاران در مطالعه خود، یک مدل ترکیبی به نام FCN-LSTM را پیشنهاد داده که ترکیبی از شبکه کانولوشنی (FCN) و حافظه کوتاه‌مدت (LSTM) است. این مدل، همزمان ویژگی‌های مکانی و زمانی فعالیت‌ها را استخراج کرده و امکان شناسایی دقیق فعالیت‌های روزانه را فراهم می‌کند. این روش روی مجموعه داده WISDM، شامل داده‌های شتاب‌سنج وژیروسکوپ برای فعالیت‌هایی نظیر راه رفتن، نشستن و دویدن، آزمایش شده است (Anbazhagan et al., 2024). نتایج نشان می‌دهد که مدل FCN-LSTM در مقایسه با روش‌های سنتی، عملکرد بهتری در تشخیص فعالیت‌های انسانی داشته و از دقت، بازیابی و امتیاز F^۱ بالاتری برخوردار است. این یافته‌ها اهمیت به کارگیری یادگیری عمیق را در بهبود

HAR در خانه‌های هوشمند تأیید می‌کنند.

۲-۲) معماری ترنسفورمر

در این بخش، مدل‌های مختلف مبتنی بر معماری ترنسفورمر که در حوزه شناسایی فعالیت‌های انسانی (HAR) با داده‌های حسگر استفاده شده‌اند، معرفی و بررسی می‌شوند. این مدل‌ها بر اساس ویژگی‌های کلیدی هر مطالعه، مانند نوع داده‌ها، رویکردهای حسگر، روش‌شناسی و معیارهای ارزیابی دسته‌بندی نشده‌اند، بلکه تلاش شده است تا تنوع کاربردها و محدودیت‌های هر مدل به همراه مجموعه داده‌های به کار رفته نشان داده شود تا تصویر کاملی از پیشرفت‌ها و چالش‌های موجود ارائه گردد.

ترنسفورمر برای اولین بار در مقاله "Attention Is All You Need" معرفی شد. این معماری مبتنی بر مکانیزم توجه، ابتدا برای مدل‌سازی توالی‌ها در ترجمه ماشینی توسعه یافت، اما به سرعت در حوزه‌های دیگری مانند پردازش زبان طبیعی (NLP)، تشخیص اشیاء، بینایی کامپیوتر و HAR گسترش یافت (Vaswani et al., 2017)، (Chitty-Venkata et al., 2023).

ویژگی کلیدی ترنسفورمر، مکانیزم توجه سراسری است که امکان یادگیری الگوهای پیچیده را از کل توالی داده‌ها فراهم می‌کند. در سیستم‌های HAR، ترنسفورمرها می‌توانند حجم داده‌ها را کاهش داده و دقت شناسایی فعالیت‌ها را افزایش دهند (Hussain et al., 2022)، (Pareek et al., 2024)، (Lesani & Fatahi, 2024).

ورودی سیستم‌های HAR شامل داده‌هایی است که از حسگرهای مختلف جمع‌آوری شده و به صورت توالی زمانی ارائه می‌شوند. این داده‌ها شامل زمان دقیق^۱ و وضعیت حسگرها در هر لحظه هستند. برخلاف داده‌های تلفن‌های هوشمند^۲ یا ساعت‌های هوشمند^۳ که در فواصل زمانی ثابت ثبت می‌شوند، حسگرهای محیطی در خانه‌های هوشمند داده‌ها را به صورت رویدادمحور و به صورت غیرمنظم تولید می‌کنند.

کاواپیز و همکاران استفاده از مدل‌های یادگیری عمیق، به‌ویژه مدل ترنسفورمر، را برای شناسایی فعالیت‌های انسانی با استفاده از حسگرهای پوشیدنی بررسی کرده‌اند. در مقاله آن‌ها، مدل‌های یادگیری عمیق مانند RNN، LSTM، BLSTM، CNN یک‌بعدی و DeepConvLSTM تحلیل شده‌اند (Kwapisz et al., 2011).

به طور کلی، نامنظمی رویدادهای حسگر، چالش‌هایی را برای روش‌های سنتی پردازش سری‌های زمانی ایجاد می‌کند که فواصل ثابت را فرض می‌کنند. وان و همکاران، رویکردهای جدیدتری مانند «پنجره‌های لغزنده مبتنی بر فعالیت» برای پردازش این داده‌ها را مناسب‌تر دانستند. الگوریتم پنجره لغزنده توسط وجک و همکاران معرفی شد، که یک روش مبتنی بر فعالیت در تحلیل سری‌های زمانی و پردازش سیگنال است که با حرکت یک پنجره متحرک روی داده‌ها، ویژگی‌های مهم را استخراج و الگوهای رفتاری را شناسایی می‌کند. این رویکرد امکان تحلیل مداوم داده‌ها و به‌روزرسانی الگوهای شناسایی شده را فراهم می‌کند. از این روش در حوزه‌هایی مانند شناسایی فعالیت در سیستم‌های حسگری، تحلیل حرکات

1 Timestamp

2 Smartphones

3 Smart watches

در ویدئوها و تشخیص رویدادهای خاص استفاده می‌شود (Wan et al., 2015)، (Wojek et al., 2008).

ترونک-هیئو و همکارانش، مدل ترنسفورمر را برای HAR از حسگرهای اینرسی بررسی کردند. این مدل توانایی کشف همبستگی‌های زمانی بین ویژگی‌ها را دارد و مزایایی مانند موازی‌سازی محاسبات سری‌های زمانی بزرگ و یادگیری دقیق‌تر زمینه در سری‌های زمانی طولانی را ارائه می‌دهد. در آزمایش آنها، تأثیر اندازه پنجره بر عملکرد مدل ترنسفورمر مورد بررسی قرار گرفته و انتخاب اندازه پنجره مناسب می‌تواند به بهبود شناسایی کمک کند (Le et al., 2022).

یکی از چالش‌های کلیدی در سیستم‌های HAR افزایش دقت و تعمیم مدل، به‌ویژه با داده‌های حسگری که شامل نویز یا تغییرات غیرقابل پیش‌بینی هستند، می‌باشد. بیش‌برازش یکی از مشکلات عمده است که عملکرد مدل‌ها را روی داده‌های دیده‌نشده محدود کرده و به‌ویژه زمانی که داده‌ها با فرکانس پایین جمع‌آوری می‌شوند یا نویز بالایی دارند، عملکرد را کاهش می‌دهد (Wang et al., 2019).

سیدانی و همکارانش، سیستمی را پیشنهاد دادند که از تکنیک‌های افزایش داده برای افزایش تعمیم مدل و جلوگیری از بیش‌برازش استفاده می‌کند. تکنیک‌هایی مانند جابجایی زمانی، تطبیق دامنه و نویز گاوسی برای افزایش داده‌های آموزشی و بهبود تعمیم مدل به کار گرفته شده است. ویژگی‌های ترکیبی پیشنهادی این مدل قابلیت استخراج اطلاعات از داده‌های حسگر در سطح پایین و بالا را دارند که به افزایش قدرت تمایز سیستم کمک می‌کند. در این مدل، ویژگی‌های استخراج شده به عنوان ورودی مدل ترنسفورمر استفاده می‌شوند. استفاده از لایه‌های کمتر برای استخراج وابستگی‌های بلندمدت منجر به بهبود دقت شده و این مدل با موفقیت برای شناسایی الگوهای پیچیده فعالیت‌های انسانی به کار گرفته شده است (Saidani et al., 2023).

تنظیم کارآمد و پویا نرخ یادگیری، توانایی مدل برای بهره‌گیری از داده‌های ورودی و بهینه‌سازی فرآیند یادگیری را افزایش می‌دهد. یکی دیگر از چالش‌های کلیدی در این حوزه، مدیریت مکانیزم‌های توجه و استخراج ویژگی‌های مرتبط از داده‌های حسگر، به‌ویژه در مدل‌های پیچیده با انواع داده متغیر است.

پرامانیک و همکاران، تکنیکی برای HAR مبتنی بر حسگر با استفاده از یادگیری عمیق و مکانیزم توجه معکوس بر پایه ترانسفورمرها معرفی کرده‌اند. این مکانیزم توجه معکوس در طول دوره یادگیری تنظیم شده و ماژول‌های توجه را منظم کرده و نرخ یادگیری را به‌طور پویا تنظیم می‌کند. این رویکرد نسبت به روش‌های پیشرفته دیگر در پنج مجموعه داده HAR مبتنی بر حسگر، عملکرد بهتری دارد. علاوه بر این، معماری جایگزینی شامل شبکه‌های عصبی کانولوشنی (CNN)، حافظه بلندمدت کوتاه (LSTM) و یک رمزگشای توجه معکوس معرفی شده است. روش‌های ارزیابی مختلفی برای ارزیابی عملکرد مدل‌های HAR در این مقاله استفاده شده‌اند (Pramanik et al., 2023).

پردازش بلادرنگ داده‌های حسگر در HAR به دلیل وجود اطلاعات خام، پر نویز و غیرمرتبط، که می‌تواند عملکرد مدل را کاهش دهد، چالش‌برانگیز است. برای رفع این مشکل، لی و همکاران سیستمی آزمایشی با استفاده از ترانسفورمر خودتوجه برای فیلتر کردن داده‌ها و شناسایی فعالیت‌ها به طور کارآمد پیشنهاد کرده‌اند (Lee et al., 2023).

در این مقاله، سیستمی آزمایشی طراحی شده که شامل یک شبکه فیلترکننده و یک طبقه‌بند اولیه است تا تأثیر داده‌های فیلتر نشده بر تشخیص HAR را بدون نیاز به طبقه‌بندی مجدد بررسی کند. شبکه فیلترکننده با استفاده از مکانیزم توجه

ترانسفورمر، ویژگی‌های مهم را از داده‌ها در یک پنجره زمانی استخراج می‌کند. ویژگی‌های استخراج شده به یک طبقه‌بند اولیه مبتنی بر سری‌های زمانی با استفاده از LSTM منتقل می‌شوند. این طبقه‌بندی اولیه تعیین می‌کند که آیا اطلاعات جمع‌آوری شده برای شناسایی فعالیت کافی است یا خیر. اگر کافی باشد، سیستم نتایج شناسایی فعالیت را ارائه می‌دهد. نتایج تجربی نشان می‌دهند که استفاده از شبکه فیلترکننده، عملکرد شناسایی فعالیت‌ها را به طور قابل توجهی بهبود می‌بخشد.

شریفی و همکاران از معماری یادگیری عمیقی به نام BioMAT استفاده کرده‌اند که بر اساس مدل ترانسفورمر ساخته شده است. این مدل برای پیش‌بینی سیستماتیک مفاصل از سیگنال‌های چندین واحد اندازه‌گیری اینرسی (IMUs) به کار رفته و شامل تقسیم داده‌های حرکتی به چرخه‌های با طول ثابت است. معماری BioMAT شامل لایه‌های خودتوجه است که امکان در نظر گرفتن وابستگی‌های جهانی در داده هنگام پردازش دنباله‌ها به صورت موازی را فراهم می‌کنند. این مدل شامل بخش رمزگذار به عنوان لایه‌های توجه و بخش رمزگشا است که خروجی از بردار نمایش به دنباله هدف تبدیل می‌کند (Sharifi-Renani et al., 2023).

برای ارزیابی عملکرد مدل، از معیارهایی مانند خطای میانگین مربعات ریشه‌ای^۱، خطای میانگین مربعات ریشه‌ای نرمال‌شده^۲ و ضریب همبستگی پیرسون (r) بین داده‌های پیش‌بینی شده و اندازه‌گیری شده استفاده شده است. نتایج نشان می‌دهد که مدل BioMAT حتی بدون نیاز به تقسیم‌بندی چرخه‌های حرکتی، دقت بیشتری نسبت به مدل‌های معمولی شبکه عصبی عمیق دارد و دقت را افزایش می‌دهد.

جیانگ و همکاران، روشی نوین برای شناسایی حرکات پیوسته انسان ارائه داده‌اند که شامل دو بخش اصلی است: (۱) شبکه مبتنی بر ترانسفورمر که ویژگی‌های زمانی-فرکانسی میکرو-دوپلر را استخراج و حرکات را در طول زمان پیش‌بینی می‌کند، و (۲) شبکه‌ای جداگانه برای شناسایی نوع حرکت بدون در نظر گرفتن اطلاعات زمانی. این روش با ترکیب پیش‌بینی‌های حرکتی و استفاده از نمودار انتقال حالت، حرکات معتبر را تعیین می‌کند. نتایج نشان می‌دهد که این مدل دقت بالایی در شناسایی حرکات پیوسته داشته و دو استراتژی پیشنهادی را با موفقیت به کار گرفته است (Jiang et al., 2023).

یی‌لیو و همکاران روش جدیدی به نام TransTM برای شناسایی فعالیت‌های انسانی با استفاده از جمع‌آوری داده‌های COTS RFID بدون نیاز به دستگاه ارائه کرده‌اند. این روش شامل ترکیبی از لایه‌های چندمقیاس ترانسفورمر است. در این روش، داده‌های خام RSSI RFID به عنوان ورودی گرفته می‌شود و از مدل TransTM، که ترکیبی از ترانسفورمر چندمقیاس و شبکه‌های کانولوشنی است، برای یادگیری ویژگی‌های رفتاری جهت تشخیص دسته‌های مختلف فعالیت استفاده می‌شود. در این مقاله، ترانسفورمر به عنوان مدل پیش‌آموزشی به کار گرفته شده و به عنوان یک عنصر کلیدی برای درک اطلاعات جهانی در ورودی‌ها استفاده می‌شود (Liu et al., 2023).

1 Root Mean Square Error (RMSE)

2 Normalized Root Mean Square Error (nRMSE)

۲-۳) شبکه‌های GRU

شبکه‌های GRU که توسط چن و همکاران معرفی شد، یکی از انواع شبکه‌های عصبی بازگشتی (RNN) هستند که با ساختار ساده‌تر و قابلیت یادگیری بهتر نسبت به داده‌های دنباله‌دار شناخته می‌شوند. GRUها مشکلات مربوط به ناپایداری گرادین در شبکه‌های RNN کلاسیک را کاهش می‌دهند و در نتیجه، دقت بالاتری در شناسایی فعالیت‌های انسانی دارند. مطالعات متعدد نشان داده‌اند که GRUها می‌توانند به‌طور مؤثری در HAR از داده‌های حسگری استفاده شوند (Chen et al., 2019). پن و همکاران به بررسی استفاده از GRU برای تشخیص فعالیت‌های انسانی با استفاده از داده‌های حسگرهای مختلف می‌پردازد. GRU به دلیل توانایی‌اش در یادگیری وابستگی‌های زمانی در داده‌های سری زمانی و مقابله با مشکل محو شدن گرادین، انتخاب شده است. مدل پیشنهادی داده‌های حسگر را به عنوان ورودی می‌پذیرد و از GRU برای یادگیری الگوهای زمانی و استخراج ویژگی‌های مهم استفاده می‌کند. سپس این ویژگی‌ها به یک لایه کاملاً متصل داده می‌شوند تا کلاس‌های فعالیت مختلف پیش‌بینی شوند (Pan et al., 2022).

نتایج نشان می‌دهند که مدل GRU می‌تواند با دقت بالایی فعالیت‌های مختلف انسانی را تشخیص دهد و عملکرد بهتری نسبت به برخی مدل‌های قدیمی‌تر مانند RNN و LSTM دارد. استفاده از GRU باعث کاهش پیچیدگی محاسباتی و افزایش کارایی مدل شده است.

محسن در این مقاله، از GRU برای دسته‌بندی فعالیت‌های انسانی استفاده کرده‌اند. این الگوریتم با پردازش داده‌های دنباله‌ای، زمان محاسباتی کمی نیاز دارد و مشکل کاهش گرادین در RNN را برطرف می‌کند. مدل پیشنهادی بر روی مجموعه داده WISDM با شش کلاس فعالیت آزمایش شده و نتایج نشان‌دهنده عملکرد بالای آن در تشخیص فعالیت‌های انسانی است (Mohsen, 2023).

با وجود پیشرفت‌های چشمگیر در به کارگیری مدل‌های یادگیری عمیق برای شناسایی فعالیت‌های انسانی، همچنان خلأهای مفهومی و عملیاتی مهمی در طراحی معماری‌های ترکیبی وجود دارد. بسیاری از مطالعات موجود، از معماری‌های مجزا مانند GRU یا ترنسفورمر بهره برده‌اند، اما تلاش‌های اندکی در راستای ادغام هدفمند این دو ساختار با در نظر گرفتن ویژگی‌های مکمل آن‌ها صورت گرفته است. GRU با توانایی بالا در مدل‌سازی وابستگی‌های زمانی محلی و ترنسفورمر با قابلیت تمرکز پویا بر روابط بلندمدت و زمینه‌ای، در صورت ترکیب هوشمندانه، می‌توانند مدلی توانمند و سازگار با پیچیدگی‌های رفتاری کاربران ارائه دهند (Shavit & Klein, 2021).

با این حال، فقدان چارچوبی که این دو معماری را به صورت ساختاریافته و تعاملی با یکدیگر تلفیق کند، به‌ویژه در حضور داده‌های نامتقارن مانند بازه‌های بدون فعالیت، همچنان به عنوان یک شکاف جدی در ادبیات علمی این حوزه باقی مانده است. از این رو، طراحی مدلی ترکیبی که هم‌زمان بتواند از مزایای دو رویکرد بهره‌برداری کرده و به‌طور دقیق به تفکیک بازه‌های فعالیت و عدم فعالیت بپردازد، ضرورتی علمی و کاربردی تلقی می‌شود که این پژوهش درصدد پاسخ‌گویی به آن است.

جدول ۱ خلاصه‌ای از برخی مطالعات اخیر در زمینه HAR و همچنین مدل‌های مبتنی بر ترنسفورمر را در شناسایی فعالیت ارائه می‌دهد. این جدول نشان می‌دهد که ترنسفورمرها با بهره‌گیری از مکانیزم توجه و توانایی پردازش توالی‌های

داده پیچیده، نقش قابل توجهی در بهبود عملکرد سیستم‌های HAR ایفا می‌کنند.

جدول ۱. خلاصه‌ای از برخی مطالعات پیشین انجام شده در زمینه شناسایی فعالیت‌های انسانی.

transformer-based HAR							
نام مطالعه	سال	روش شناسی	مدل حسگر	مجموعه داده‌ها	ارزیابی	محدودیت‌ها	یافته‌ها
۱ (Anbazhagan et al.)	۲۰۲۴	مدل ترکیبی به نام FCN-LSTM	اینرسی	WISDM	دقت ۹۵.۹۹	تفسیرپذیری پایین، نیاز به داده برچسب خورده، چالش حفظ حریم خصوصی، عدم اطمینان در شرایط واقعی.	ترکیب خوبی از ویژگی‌های مکانی و زمانی را استخراج می‌کند. عملکرد بهتری نسبت به RNN، Naive Bayes و KNN دارد.
۲ (Le et al.)	۲۰۲۲	پنجره لغزشی	اینرسی	MHD, C.C MDFALL, D ALiAc	دقت ۹۹.۵۶	نیاز به داده‌های گسترده‌تر	شناسایی حرکت مبتنی بر داده‌های شتاب‌سنج و ژيروسکوپ
۳ (Saidani et al.)	۲۰۲۳	تکنیک داده‌های پیشرفته	محیطی، پوشیدن، ی، بی‌سیم (موبایل)	WISDM, P AMPA2, U CIHAR	دقت ۹۸.۶، ۹۸.۲ و ۹۷.۳	مجموعه داده نامتعادل	ترنسفورمر و ترکیب ویژگی‌های طیف‌سنج‌ها
۴ (Pramanik et al.)	۲۰۲۳	شبکه معکوس توجه مبتنی بر ترنسفورمر	محیطی	MHEALTH USC، HAD، WHARF UTD-2 MHAD1،	دقت 0.95، 0.92، 0.64، 0.8421	افزایش تعداد پارامترها به دلیل وجود طرح توجه نیاز به آموزش دارد.	در طول فرآیند یادگیری خود کالیبراسیون، مازول‌های توجه را منظم می‌کند و نرخ یادگیری را به صورت پویا تنظیم می‌کند.
۵ (Lee et al.)	۲۰۲۳	انکودر ترنسفورمر مبتنی بر مکانیزم توجه به عنوان شبکه فیلتر	محیطی	Casas: milan, kyoto8, kyoto11	بهترین نتایج دقت برای مجموعه داده‌های Milan، Kyoto8، Kyoto11: ۰.۸۰، ۰.۹۲، 0.88	پیچیدگی داده‌ها پایین به دلیل تعداد کم ساکنان	فیلتر کردن داده‌های نامتعادل در پنجره انتقال و طبقه‌بندی کننده اولیه ویژگی‌های مربوطه
۶ (Sharifi-Renani et al.)	۲۰۲۳	معماری BERT با انکودر	پوشیدن، ی، اینرسی	مجموعه داده بیومکانیک اندام تحتانی	به طور میانگین، RMSE برابر ۵.۵	آموزش مدل با داده‌های متنوع و جدیدتر	عملکرد بهتر بدون نیاز به تقسیم‌بندی و افزایش دقت.

۷	(Liu & Fang)	۲۰۲۳	شناسایی حرکت‌های مداوم	مبتنی بر رادار	حرکت مداوم بر اساس ویژگی‌های میکرو-داپلر	دقت ۹۴٪	عدم تعادل داده‌ها و وابستگی به تقویت داده‌ها محدودیت در تعمیم‌پذیری را ایجاد می‌کند.	شبکه OCR برای شناسایی انواع حرکت بدون زمان
۸	(Liu et al.)	۲۰۲۳	TransTM، ترنسفورمر چندمقیاس و CNN	محیطی	RFID	دقت ۹۹.۱٪	وابستگی به داده‌های سفارشی و هزینه بالای محاسبات محدودیت در تعمیم‌پذیری مدل را ایجاد می‌کند.	داده‌های کم کیفیت را می‌گیرد و از فرآیند پاکسازی داده‌ها اجتناب می‌کند.
۹	(Pan et al.)	۲۰۲۲	مبتنی بر شبکه GRU با ترکیب دو مکانیزم توجه	اینرسی	مجموعه داده شامل ۱,۸۱۷,۵۸۳ نمونه از ۱۰ داوطلب	دقت ۹۰٪	دقت بالا و مصرف منابع کمی دارد. توجه زمانی برای فعالیت‌های پویا و توجه کانالی برای فعالیت‌های ایستا مؤثرتر است.	توجه کانالی در داده‌های I.I.D عملکرد ضعیفی دارد و برای فعالیت‌های پیچیده نیاز به بررسی بیشتری هست.
۱۰	(Mohsen)	۲۰۲۳	مبتنی بر شبکه GRU	حسگر بی سیم	WISDM	دقت ۹۷.۰۸٪	وابستگی به تنظیمات ابرپارامتر، عدم بررسی روی داده‌ها و محیط‌های دیگر	عملکرد قوی GRU با تنظیم ابرپارامترها

بررسی پیشینه پژوهش‌ها نشان می‌دهد که مدل‌های یادگیری عمیق، به‌ویژه معماری‌های GRU و ترنسفورمر، عملکرد مناسبی در شناسایی فعالیت‌های انسانی بر پایه داده‌های حسگرهای هوشمند دارند. اگرچه ترکیب این دو معماری می‌تواند موجب تقویت مدل در استخراج الگوهای زمانی و زمینه‌ای شود، اما در صورتی که بازه‌های فاقد فعالیت به‌صورت مستقل در فرآیند آموزش مدل در نظر گرفته نشوند، این ترکیب نیز ممکن است نتایج مطلوبی به همراه نداشته باشد. لحاظ کردن داده‌هایی که در مجموعه داده ثبت شده‌اند اما بیانگر وقوع فعالیت خاصی نیستند، می‌تواند نقش بسزایی در افزایش دقت مدل ایفا کند؛ چرا که این کار به مدل امکان می‌دهد تا مرز مشخصی میان فعالیت‌های واقعی و بازه‌های عدم فعالیت قائل شود و از شناسایی نادرست بازه‌های عدم فعالیت به‌عنوان فعالیت جلوگیری کند. این مسئله در نهایت منجر به بهبود قابل توجه عملکرد کلی سیستم خواهد شد.

در این پژوهش، با بهره‌گیری از یک معماری ترکیبی شامل Bi-GRU و ترنسفورمر و استفاده از دو مجموعه داده معروف، مدلی طراحی شده است که علاوه بر شناسایی دقیق فعالیت‌های انسانی، قادر به تشخیص وضعیت‌های بدون فعالیت نیز می‌باشد. این ویژگی، مزیت مهمی در کاربردهای واقعی خانه‌های هوشمند محسوب می‌شود. رویکرد پیشنهادی با مدل‌سازی هم‌زمان وابستگی‌های زمانی کوتاه‌مدت و بلندمدت، می‌تواند بر محدودیت‌های روش‌های پیشین غلبه کرده و عملکرد دقیق‌تری در تحلیل رفتار کاربران ارائه دهد.

(۳) روش تحقیق

در این بخش، روش تحقیق مورد استفاده در پژوهش ارائه می‌شود. ابتدا معماری مدل‌های پایه، شامل ترنسفورمر و Bi-GRU، که مبنای توسعه مدل پیشنهادی هستند، معرفی می‌گردد. در ادامه، مدل پیشنهادی که بر اساس ترکیب این دو معماری و با هدف بهبود دقت در شناسایی فعالیت‌های انسانی طراحی شده است، تشریح خواهد شد.

۳-۱-۳ مدل‌های پایه

۳-۱-۱-۳ مدل ترنسفورمر

مدل ترنسفورمر توسط واسوانی و همکاران معرفی شد و به سرعت به یکی از مهم‌ترین معماری‌ها در یادگیری عمیق تبدیل شد. این مدل با تکیه بر مکانیزم توجه، امکان پردازش هم‌زمان کل توالی را فراهم می‌کند و برخلاف RNN و LSTM که به صورت ترتیبی عمل می‌کنند، قادر به مدل‌سازی وابستگی‌های بلندمدت با سرعت بالاتر است. ترنسفورمر از دو بخش اصلی تشکیل شده است: رمزگذار و رمزگشا هر لایه رمزگذار شامل مکانیزم توجه چندسری برای تمرکز بر بخش‌های کلیدی ورودی و یک شبکه عصبی جلوگرا برای پردازش ویژگی‌هاست. در بسیاری از مسائل مانند شناسایی فعالیت، تنها از بخش رمزگذار برای استخراج ویژگی‌های توالی استفاده می‌شود (Vaswani et al., 2017).

مکانیزم توجه هسته اصلی مدل ترنسفورمر است که به مدل اجازه می‌دهد بر روی بخش‌های خاصی از توالی تمرکز کند. سه عنصر اصلی در این مکانیزم وجود دارد:

۱. Query (Q) نماینده بردار جستجو است.

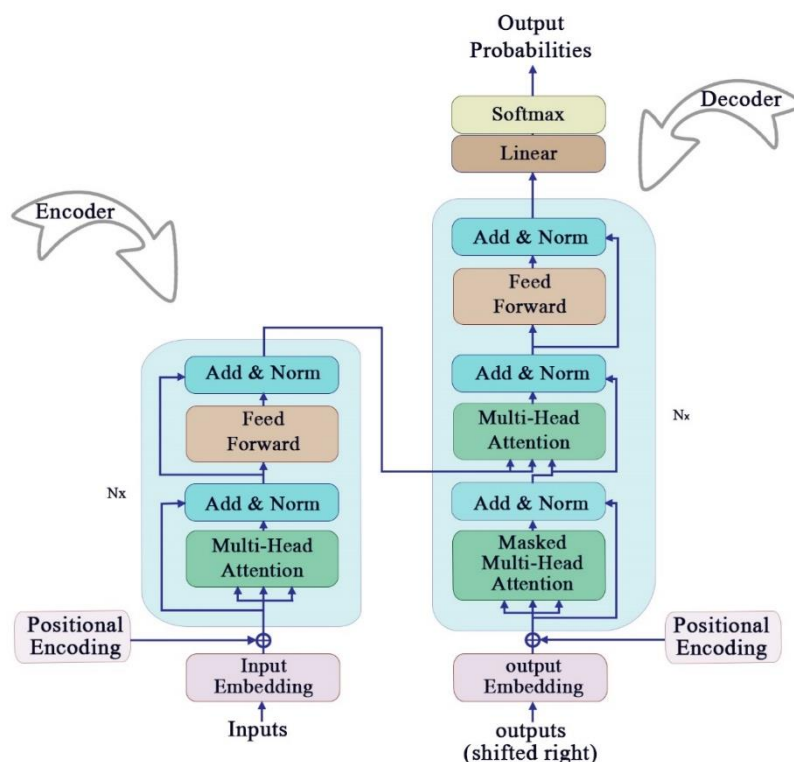
۲. Key (K) نماینده کلیدهایی است که بخش‌های مختلف توالی را نمایندگی می‌کنند.

۳. Value (V) نماینده مقادیر مرتبط با هر کلید است.

محاسبه توجه به صورت زیر انجام می‌شود:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (\text{رابطه ۱})$$

که در آن d_k تعداد ابعاد کلیدها است و به منظور نرمال‌سازی استفاده می‌شود. شکل ۲ معماری کامل ترنسفورمر را نشان می‌دهد.



شکل ۲. مدل ترنسفورمر. این معماری شامل embedding، کدگذار با مکانیزم توجه چندسری و لایه طبقه‌بندی است و برای مدل‌سازی وابستگی‌های زمانی بلندمدت در داده‌های حسگر طراحی شده است.

۳-۱-۲) مدل GRU

مدل GRU نوعی شبکه RNN است که برای پردازش داده‌های دنباله‌دار طراحی شده و نسبت به LSTM ساختار ساده‌تری دارد. این مدل که توسط چن و همکاران معرفی شده، با استفاده از دو دروازه به‌روزرسانی و بازنشانی، اطلاعات مهم را در طول زمان حفظ و داده‌های غیرضروری را حذف می‌کند. سادگی ساختار GRU موجب کاهش پیچیدگی محاسباتی و افزایش سرعت آموزش شده است. به دلیل قابلیت حفظ وابستگی‌های زمانی، GRU در تحلیل داده‌های حسگر و شناسایی فعالیت‌های انسانی در خانه‌های هوشمند عملکرد مؤثری دارد (Chen et al., 2019).

مدل GRU برای هر گام زمانی t به صورت زیر عمل می‌کند:

۱. دروازه بازنشانی T_t مشخص می‌کند که چه مقدار از اطلاعات حالت قبلی h_{t-1} برای به‌روزرسانی در حالت کنونی باید فراموش شود. مقدار دروازه بازنشانی به صورت زیر محاسبه می‌شود:

$$T_t = \sigma(w_T[h_{t-1}, x_t] + b_T) \quad (\text{رابطه ۲})$$

در اینجا σ تابع سیگموید است که مقادیر را بین ۰ و ۱ نگه می‌دارد.

۲. دروازه به‌روزرسانی Z_t مشخص می‌کند که چه مقدار از اطلاعات حالت قبلی h_{t-1} به حالت جدید منتقل شود و چه مقدار از اطلاعات جدید $\tilde{h}_t$ در حالت فعلی لحاظ شود:

$$Z_t = \sigma(w_Z[h_{t-1}, x_t] + b_Z) \quad (\text{رابطه ۳})$$

۳. حالت کاندید ($\tilde{h}_t$) نمایشی از اطلاعات جدید است که با ترکیب ورودی x_t و حالت قبلی کنترل شده توسط

دروازه بازنشانی تولید می شود:

$$\tilde{h}_t = \tanh(w_h \cdot [r_{t-1} h_{t-1}, x_t] + b_h) \quad \text{رابطه ۴}$$

۴. حالت جدید (h_t) با ترکیب حالت قبلی h_{t-1} و حالت کاندید $\tilde{h}_t$ طبق دروازه به روزرسانی تولید می شود:

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad \text{رابطه ۵}$$

۳-۱-۳ شبکه دو طرفه GRU

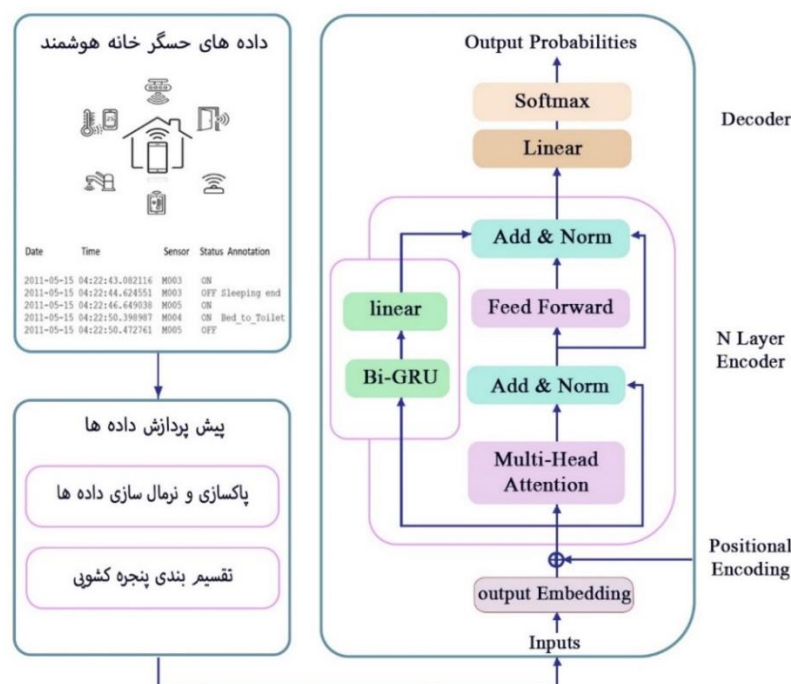
شبکه GRU دو طرفه (Bi-GRU) برای پردازش هم زمان اطلاعات گذشته و آینده در داده های دنباله دار طراحی شده است. این ویژگی کمک می کند تا وابستگی های زمانی پیچیده تر شناسایی شوند. در این پژوهش، سه لایه Bi-GRU با واحدهای مخفی متفاوت به صورت موازی با معماری ترانسفورمر ترکیب شده اند تا مدل بهینه ای برای پردازش داده های حسگری خانه های هوشمند ارائه شود (Chen et al., 2019).

Bi-GRU با حفظ اطلاعات زمانی در هر دو جهت، مشکلات ناپایداری گرادینان را کاهش داده و عملکرد بهتری در تحلیل داده های حسگری دارد. برخلاف GRU استاندارد که توالی را تنها در یک جهت پردازش می کند، Bi-GRU با افزودن یک لایه معکوس، از اطلاعات آینده و گذشته بهره می برد و دقت مدل را بهبود می بخشد.

۳-۲ معماری مدل پیشنهادی

معماری پیشنهادی ترکیبی از Bi-GRU و ترانسفورمر است که به صورت موازی پردازش داده ها را انجام می دهد. سه لایه Bi-GRU برای مدل سازی وابستگی های زمانی کوتاه مدت و بلند مدت استفاده شده و لایه توجه ترانسفورمر به پردازش موازی داده ها و استخراج وابستگی های بلند مدت کمک می کند. برای ترکیب ویژگی های استخراج شده، از عملیات الحاق یا میانگین گیری وزنی استفاده می شود تا نمایشی یکپارچه از داده ها ایجاد گردد. در مرحله نهایی، ویژگی های ترکیبی به یک شبکه عصبی پیش رو داده شده و با استفاده از تابع Softmax، کلاس نهایی فعالیت های انسانی شناسایی و عدم فعالیت نیز تشخیص داده می شود.

مزایای این مدل شامل بهبود دقت و سرعت پردازش، ترکیب وابستگی های زمانی کوتاه مدت و بلند مدت، و انعطاف پذیری بالا در پردازش داده های توالی دار است. همچنین، این رویکرد به کمک معیارهایی مانند دقت، صحت، بازخوانی، امتیاز F1 و ماتریس سردرگمی ارزیابی شده و عملکرد بالایی را در HAR نشان داده است.



شکل ۳. بلوک دیاگرام معماری پیشنهادی.

معماری پیشنهادی برای شناسایی فعالیت‌های انسانی مبتنی بر داده‌های حسگرهای محیطی در خانه‌های هوشمند طراحی شده است. این مدل ترکیبی از قابلیت‌های مدل‌های توالی‌محور^۱ مانند GRU و مدل‌های مبتنی بر توجه مانند ترنسفورمر است و تلاش دارد تا وابستگی‌های زمانی کوتاه مدت و بلندمدت را به صورت هم‌زمان مدل‌سازی کند. ساختار این مدل شامل سه مرحله اصلی است: ۱: پیش‌پردازش داده‌ها، ۲: استخراج ویژگی به کمک Bi-GRU و ترنسفورمر، ۳: ادغام خروجی‌ها و دسته‌بندی فعالیت‌ها.

الف) پیش‌پردازش داده‌ها

داده‌های ورودی شامل اطلاعات ثبت شده از حسگرهای محیطی نصب شده در خانه هوشمند هستند که شامل زمان، شناسه حسگر، نوع حسگر، دما و وضعیت فعالیت می‌باشند. ابتدا داده‌ها پاک‌سازی شده و نمونه‌های ناقص یا دارای نویز حذف می‌شوند. سپس، عملیات نرمال‌سازی، کدگذاری عددی ویژگی‌ها (نظیر شناسه حسگر و نوع فعالیت) و مرتب‌سازی زمانی داده‌ها انجام می‌گیرد تا آماده ورود به مدل شوند.

برای حفظ ساختار توالی زمانی در داده‌ها، از بخش‌بندی مبتنی بر پنجره‌های زمانی ثابت^۲ استفاده می‌شود. در این مرحله، توالی داده‌ها به بازه‌های زمانی با طول مشخص تقسیم می‌گردند، به طوری که هر پنجره به عنوان یک نمونه ورودی مستقل برای مدل در نظر گرفته می‌شود. این روش، ضمن حفظ توالی و ترتیب رخدادها، امکان یادگیری بهتر الگوهای زمانی و بهبود دقت شناسایی فعالیت‌ها را فراهم می‌سازد.

1 Sequential Models

2 Sliding Windows

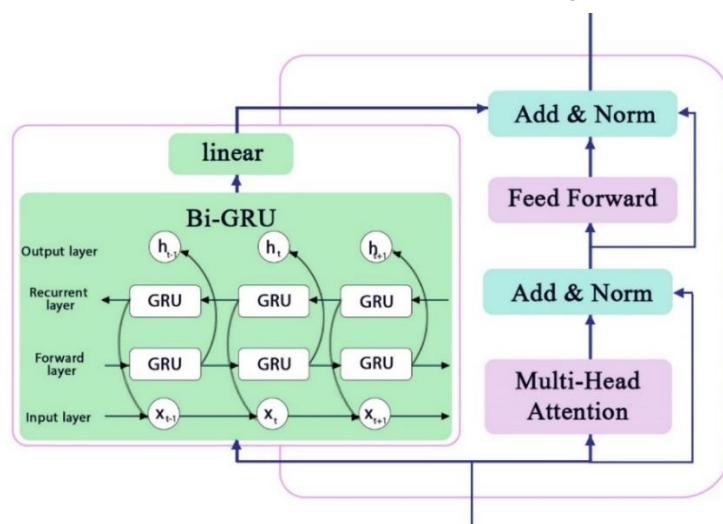
ب) پردازش مدل‌های Bi-GRU و ترنسفورمر

مدل پیشنهادی شامل دو مسیر پردازشی موازی است که به‌طور همزمان داده‌ها را از دو جنبه زمانی و معنایی تجزیه و تحلیل می‌کنند. این معماری به‌گونه‌ای طراحی شده که هر مسیر به‌طور مستقل و هم‌زمان اطلاعات را پردازش می‌کند تا نهایتاً ترکیب این دو مسیر منجر به استخراج ویژگی‌های پیچیده‌تر و دقیق‌تری از داده‌ها شود.

مسیر اول، شامل سه لایه GRU دوطرفه است که به‌صورت سری به یکدیگر متصل شده‌اند. این لایه‌ها با هدف استخراج ویژگی‌های زمانی و مدل‌سازی وابستگی‌های محلی و ترتیبی از داده‌ها به کار می‌روند. استفاده از GRU دوطرفه این امکان را فراهم می‌آورد که مدل بتواند اطلاعات را از هر دو جهت زمانی (گذشته و آینده) پردازش کند و ویژگی‌های غنی‌تری از توالی داده‌ها استخراج کند.

مسیر دوم، از ساختار ترنسفورمر با لایه‌های توجه چندسری^۱ و شبکه‌های تغذیه پیش‌رو^۲ استفاده می‌کند. در این مسیر، مکانیزم Self-Attention به مدل کمک می‌کند تا تعاملات پیچیده و وابستگی‌های بلندمدت بین رویدادهای مختلف زمانی را شناسایی کند. این بخش از مدل توانایی تحلیل روابط معنایی و توجه به قسمت‌های مختلف داده‌ها در مقیاس‌های زمانی متفاوت را فراهم می‌آورد.

همان‌گونه که در شکل ۴ مشاهده می‌شود، سه لایه GRU دوطرفه به‌طور موازی با لایه‌های توجه ترنسفورمر قرار گرفته‌اند و هر دو مسیر به‌طور همزمان داده‌های ورودی را دریافت و پردازش می‌کنند. این موازی‌سازی پردازش داده‌ها به مدل اجازه می‌دهد که ویژگی‌های مختلف داده‌ها را به‌طور هم‌زمان و از جنبه‌های گوناگون (زمانی و معنایی) استخراج کرده و نهایتاً با ترکیب این اطلاعات، نتایج دقیق‌تری به‌دست آورد. این معماری موازی به‌طور چشمگیری سرعت پردازش را افزایش می‌دهد و دقت مدل را در تحلیل‌های پیچیده زمانی و معنایی ارتقاء می‌بخشد.



شکل ۴. معماری مدل ترکیبی Bi-GRU و ترنسفورمر

مدل شامل سه لایه GRU دوطرفه (Bi-GRU) برای استخراج ویژگی‌های زمانی در دو جهت است. خروجی این بخش به‌صورت موازی با ماژول ترنسفورمر شامل لایه‌های Attention چندسری و Feed Forward پردازش می‌شود. این ساختار ترکیبی، وابستگی‌های زمانی و معنایی را به‌طور هم‌زمان مدل‌سازی می‌کند.

1 Multi-head Attention

2 Feedforward

ج) ترکیب خروجی و شناسایی فعالیت‌ها

در این مرحله، ویژگی‌های استخراج شده از دو مسیر پردازش GRU دوطرفه و ترنسفورمر به صورت دو بردار مجزا ترکیب شده و به لایه تمام متصل ارسال می‌شوند. این ترکیب به مدل این امکان را می‌دهد که وابستگی‌های زمانی را از دو جنبه مختلف (گذشته و آینده) به طور همزمان درک کند. به طور خاص، ویژگی‌های زمانی استخراج شده توسط GRU به مدل کمک می‌کنند تا الگوهای ترتیبی کوتاه مدت را شبیه سازی کند، در حالی که ویژگی‌های معنایی و بلندمدت توسط ترنسفورمر شبیه سازی می‌شوند. این فرایند ترکیب بهبود قابل توجهی در دقت شناسایی فعالیت‌ها ایجاد می‌کند و مدل را قادر می‌سازد تا الگوهای پیچیده‌ای از رفتارهای انسانی را شناسایی کرده و فعالیت‌ها را با دقت بیشتری تشخیص دهد.

یکی از ویژگی‌های منحصر به فرد این تحقیق، در نظر گرفتن داده‌های مربوط به «عدم فعالیت» است. در این تحقیق، علاوه بر فعالیت‌های انسانی که به طور معمول در محیط‌های هوشمند شناسایی می‌شوند، زمان‌هایی که ساکنان هیچ فعالیتی انجام نمی‌دهند نیز در مجموعه داده‌ها گنجانده شده است. این کلاس که بیشترین تعداد نمونه را نسبت به سایر فعالیت‌ها دارد، برای مدل سازی زمان‌های بی حرکتی و رفتارهای ساکنان بسیار مهم است. شناسایی دقیق تر این دوره‌های عدم فعالیت، به مدل کمک می‌کند تا بتواند زمان‌هایی که افراد در حال استراحت یا عدم فعالیت هستند را نیز شناسایی کند و در نهایت، دقت کلی مدل در پیش بینی رفتارها و فعالیت‌های ساکنان افزایش می‌یابد.

اضافه کردن این کلاس به مجموعه داده، مدل را قادر می‌سازد تا به طور مؤثری با زمان‌های بی حرکتی برخورد کند و در نتیجه باعث بهبود عملکرد مدل‌های یادگیری عمیق در تحلیل داده‌های پیچیده و شناسایی دقیق تر فعالیت‌های انسانی می‌شود. این رویکرد، به ویژه برای کاربردهایی که در آن شناسایی رفتارهای غیرفعال افراد ضروری است (مانند نظارت بر سالمندان یا بهبود سیستم‌های هشدار در خانه‌های هوشمند)، بسیار کارآمد است.

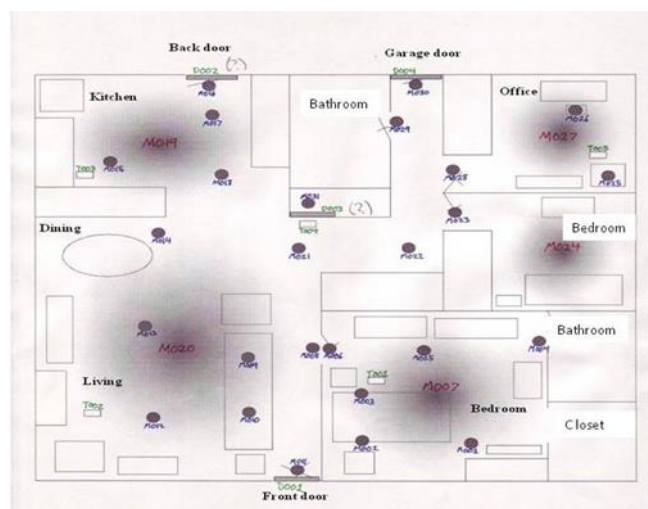
۳-۳) مجموعه داده‌ها

در این تحقیق، از مجموعه داده‌های CASAS، که شامل داده‌های جمع آوری شده از انواع مختلف حسگرها در خانه‌های هوشمند است، برای HAR استفاده شده است. این مجموعه داده‌ها در دانشگاه ایالتی واشنگتن توسعه یافته‌اند، به ویژه در حوزه شناسایی فعالیت‌های انسانی و تحلیل داده‌های خانه‌های هوشمند کاربرد دارند. تاکنون بیش از ۲۰ مجموعه داده مختلف توسط CASAS منتشر شده است، که برخی از آن‌ها به خانه‌های چندساکن اختصاص دارند (Cook, 2010).

- مجموعه داده اروپا شامل اطلاعات فعالیت‌های فردی است که توسط حسگرهای مختلف مانند حسگرهای حرکتی، درب، و دما جمع آوری شده‌اند. این مجموعه داده شامل فعالیت‌هایی مانند خوابیدن، تهیه غذا، استراحت، خوردن، کار، شستن ظروف و فعالیت‌های دیگر است که به طور دقیق برچسب گذاری شده‌اند. فعالیت‌ها به طور دقیق شمارش شده و مجموعه داده شامل ۱۱ نوع فعالیت مختلف و یک عدم فعالیت است. بخشی از مجموعه داده اروپا در شکل ۵ نشان داده شده است. همچنین نقشه خانه این مجموعه داده در شکل ۶ مشاهده می‌شود.

Date	Time	Sensor	Status	Annotation
2011-05-15	04:22:43.082116	M003	ON	
2011-05-15	04:22:44.624551	M003	OFF	Sleeping end
2011-05-15	04:22:46.649038	M005	ON	
2011-05-15	04:22:50.398987	M004	ON	Bed_to_Toilet begin
2011-05-15	04:22:50.472761	M005	OFF	
2011-05-15	04:22:57.782528	T005	21	
2011-05-15	04:22:58.533178	M004	OFF	
2011-05-15	04:22:58.623508	M007	OFF	
2011-05-15	04:25:44.619463	M004	ON	

شکل ۵. نمونه‌ای از مجموعه داده آروبا.

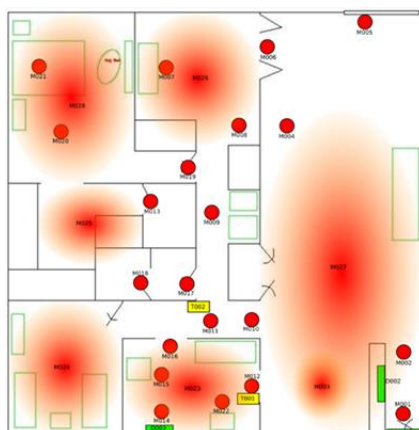


شکل ۶. حسگرهای خانه مورد آزمایش برای مجموعه داده آروبا

- مجموعه داده میلان، مشابه آروبا، شامل داده‌های جمع‌آوری شده از حسگرهای محیطی در خانه‌های هوشمند است، اما با تفاوت‌هایی در نوع فعالیت‌ها و محیط‌های مورد نظر. در این مجموعه داده، فعالیت‌هایی مانند مدیتیشن، تماشای تلویزیون، انجام کارهای خانه، مصرف دارو و سایر فعالیت‌های مشابه ثبت شده است. این داده‌ها شامل فعالیت‌های مختلف و محیط‌های متنوع‌تری نسبت به آروبا هستند. بخشی از مجموعه داده میلان در شکل ۷ نشان داده شده است. همچنین نقشه خانه این مجموعه داده در شکل ۸ مشاهده می‌شود.

Date	Time	Sensor ID	Value	Label
2010-01-05	08:25:37.000026	M003	OFF	
2010-01-05	08:25:45.000001	M004	ON	Read begin
...	...	...	...	...
2010-01-05	08:35:09.000069	M004	ON	
2010-01-05	08:35:12.000054	M027	ON	
2010-01-05	08:35:13.000032	M004	OFF	(Read should end)
2010-01-05	08:35:18.000020	M027	OFF	
2010-01-05	08:35:18.000064	M027	ON	

شکل ۷. نمونه‌ای از مجموعه داده میلان.



شکل ۸. حسگرهای خانه مورد آزمایش برای مجموعه داده میلان.

هر دو دیتاست شامل اطلاعاتی از سه نوع حسگر هستند:

- حسگرهای حرکتی^۱: ثبت حرکات در مناطق مختلف خانه.
- حسگرهای درب^۲: وضعیت باز و بسته بودن درب‌ها.
- حسگرهای دما^۳: مقادیر دمای محیطی. داده‌ها به صورت توالی زمانی ثبت شده‌اند و شامل ستون‌هایی مانند زمان رویداد، نوع حسگر، و وضعیت حسگر هستند.

جدول ۲ ویژگی‌های اصلی مجموعه داده‌های مورد استفاده در این پژوهش را نمایش می‌دهد. این جدول شامل اطلاعاتی مانند سال جمع‌آوری داده‌ها، تعداد کل نمونه‌ها، نوع حسگرها، تعداد حسگرها، تعداد فعالیت‌های ثبت شده و ویژگی‌های هر مجموعه داده است.

جدول ۲. ویژگی‌های مجموعه داده‌های استفاده شده

مجموعه داده	سال	کل نمونه‌ها	نوع حسگرها	تعداد حسگرها	تعداد فعالیت‌ها	ویژگی‌های ثبت شده در مجموعه داده
آروبا	۲۰۱۱	۱۰۴۸۵۶۷	محیطی	۳۹	۱۲	زمان، نوع حسگر، وضعیت حسگر، فعالیت
میلان	۲۰۱۸	۴۳۳۶۶۵	محیطی	۳۳	۱۵	زمان، فاصله زمانی، نوع حسگر، وضعیت حسگر، فعالیت

هدف این تحقیق شناسایی فعالیت‌های پیچیده ساکنان با استفاده از داده‌های حسگرهای مختلف است که می‌تواند به درک بهتر رفتارهای انسانی در محیط‌های هوشمند کمک کند. این پژوهش با استفاده از این داده‌ها، به‌ویژه برای ارزیابی و توسعه مدل‌های یادگیری عمیق که قادر به شناسایی و پیش‌بینی فعالیت‌های انسانی به‌طور دقیق و زمان‌بندی شده هستند، اهمیت دارد.

1 Motion Sensors

2 Door Sensors

3 Temperature Sensors

۳-۴) آماده‌سازی و پیش‌پردازش داده‌های مجموعه داده

آماده‌سازی داده‌ها برای شناسایی فعالیت‌های انسانی از اهمیت زیادی برخوردار است و شامل مراحل مختلفی برای تبدیل داده‌های خام به فرمت قابل استفاده برای مدل‌های یادگیری ماشین و یادگیری عمیق است. این فرآیند شامل پاک‌سازی داده‌ها، مدیریت مقادیر گمشده، هماهنگ‌سازی داده‌ها و تبدیل آن‌ها به فرمت مناسب است. هدف اصلی، اطمینان از کیفیت و یکپارچگی داده‌ها قبل از اعمال مراحل پیچیده‌تری مانند نرمال‌سازی و پنجره‌گذاری زمانی است.

داده‌ها به توالی‌های زمانی با طول ثابت تقسیم می‌شوند تا برای ورودی مدل‌های یادگیری عمیق آماده شوند. برای جلوگیری از عدم توازن کلاس‌ها، از تکنیک‌های داده‌افزایی و افزودن نویز به کلاس‌های کم‌تکرار استفاده شده است. همچنین، برای هر گروه داده، برچسب فعالیت از اولین نمونه آن گروه دریافت می‌شود و پنجره‌ها به اندازه ثابت (۳۰ مقدار) تنظیم می‌شوند.

در مجموعه داده آروبا، به دلیل حجم بیشتر داده‌ها از پنجره‌های بدون همپوشانی استفاده شد تا سرعت پردازش افزایش یابد. در مجموعه داده میلان، به علت حجم کمتر داده‌ها از پنجره‌های همپوشان و داده‌افزایی با اسموت استفاده شد تا تعداد داده‌ها افزایش یافته و دقت مدل بهبود یابد. این رویکردهای متفاوت باعث بهینه‌سازی آماده‌سازی داده‌ها برای هر مجموعه داده شده است.

۳-۵) معیارهای ارزیابی

برای ارزیابی عملکرد مدل پیشنهادی در شناسایی فعالیت‌های انسانی، از چندین معیار ارزیابی آماری مبتنی بر ماتریس سردرگمی (Confusion Matrix) استفاده شده است. این معیارها شامل دقت (Accuracy)، حساسیت (Recall)، صحت (Precision) و امتیاز F1 (F1-score) می‌باشند. در ادامه به توضیح هر یک از این معیارها و اهمیت آن‌ها در ارزیابی مدل پرداخته می‌شود:

• دقت^۱:

دقت نشان‌دهنده درصد کلی پیش‌بینی‌های درست مدل (چه مثبت و چه منفی) نسبت به کل نمونه‌هاست. به عبارت دیگر، دقت نشان می‌دهد مدل چه میزان توانسته به درستی خروجی‌ها را پیش‌بینی کند.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad \text{(رابطه ۶)}$$

که در آن TP پیش‌بینی صحیح برای کلاس مثبت، TN پیش‌بینی صحیح برای کلاس منفی، FP پیش‌بینی اشتباه به عنوان کلاس مثبت و FN پیش‌بینی اشتباه به عنوان کلاس منفی است.

• حساسیت^۲:

حساسیت، که با نام **Recall** نیز شناخته می‌شود، نشان‌دهنده‌ی توانایی مدل در شناسایی صحیح نمونه‌های واقعی مثبت است. این معیار مخصوصاً در کاربردهایی که از دست رفتن نمونه‌های مثبت (مانند رخداد یک فعالیت مهم) اهمیت دارد، حیاتی است. نحوه محاسبه آن در رابطه ۷ بیان شده است.

$$recall = \frac{TP}{TP+TN} \quad \text{رابطه ۷}$$

• صحت^۱:

صحت مشخص می‌کند چه درصدی از نمونه‌هایی که مدل به عنوان مثبت پیش‌بینی کرده، واقعاً مثبت بوده‌اند. این معیار زمانی اهمیت دارد که اشتباه در پیش‌بینی مثبت‌ها هزینه‌بر باشد. نحوه محاسبه صحت در رابطه ۸ بیان شده است.

$$precision = \frac{TP}{FP+TP} \quad \text{رابطه ۸}$$

• امتیاز F1^۲:

امتیاز F1 میانگینی از Precision و Recall است که تعادلی میان این دو معیار برقرار می‌کند. این معیار زمانی اهمیت دارد که توازن بین حساسیت و دقت پیش‌بینی اهمیت داشته باشد. رابطه ۹ نحوه محاسبه را بیان می‌کند.

$$F1 \text{ Score} = 2 \frac{Precision \cdot Recall}{Precision+Recall} \quad \text{رابطه ۹}$$

استفاده از این معیارهای ارزیابی به دلیل پوشش کامل جنبه‌های مختلف عملکرد مدل، از جمله دقت کلی، توانایی شناسایی درست فعالیت‌ها، جلوگیری از پیش‌بینی اشتباه، و توان تفکیک کلاس‌ها، باعث می‌شود که تحلیل جامعی از عملکرد مدل GRU در شناسایی فعالیت‌های انسانی فراهم شود.

۴) نتایج پیاده‌سازی مدل پیشنهادی و بحث

در این فصل، نتایج حاصل از پیاده‌سازی مدل پیشنهادی ارائه و تحلیل می‌گردد. ابتدا عملکرد مدل بر اساس معیارهای ارزیابی تعریف شده بررسی می‌شود و سپس نتایج به‌دست آمده با هدف ارزیابی دقت، کارایی و توانایی مدل در تفکیک فعالیت‌ها و بازه‌های عدم فعالیت تحلیل می‌شوند. در ادامه، یافته‌های به‌دست آمده با نتایج مدل‌های پایه مقایسه شده و نقاط قوت و ضعف مدل پیشنهادی مورد بحث قرار می‌گیرد.

در این پژوهش، برای بخش‌بندی توالی‌های رویدادهای حسگر از روش پنجره کشویی با فواصل ۳۰ ثانیه استفاده شد. مدل پیشنهادی Bi-GRU-Transformer با مدل‌های پایه و ترکیبی شامل ترنسفورمر مقایسه و با معیارهای استاندارد دقت، یادآوری، صحت و F1 ارزیابی شد. ورودی مدل شامل هشت ویژگی برای داده‌های Aruba و هفت ویژگی برای Milan است. در تنظیم مدل، ابعاد ویژگی‌های پنهان (۶۴، ۱۲۸، ۲۵۶) و سه لایه BiGRU به کار رفته‌اند. جزئیات هایپرپارامترها در جدول ۳ آمده است.

جدول ۳. هایپرپارامتر تیونینگ

هایپرپارامتر	فضای جستجو	مقدار انتخاب شده
نرخ یادگیری	1e-3, 5e-4, 1e-4, 5e-5	1e-4
اندازه دسته	۱۶,۳۲,۶۴,۱۲۸	۳۲
بهینه ساز	'Adam', 'SGD', 'RMSprop'	Adam
نرخ حذف تصادفی	۰.۱, ۰.۳, ۰.۵	۰.۳, ۰.۵
تعداد دوره‌های آموزشی	۱۰,۲۰,۵۰,۱۰۰,۰۰۰	۵۰۰
تعداد نوروں‌های لایه پنهان	۳۲,۶۴,۱۲۸,۲۵۶	۶۴,۱۲۸,۲۵۶
تعداد لایه‌ها	[1, 2, 3, 4]	۳
وزن دهی	۰.۰۱, ۰.۰۰۱, ۰.۰۰۵	۰.۰۱, ۰.۰۰۵
تابع خطا یا تابع هزینه	'CrossEntropy', 'MSE', 'MAE'	CrossEntropy

۴-۱) نتایج پیاده سازی

برای ارزیابی مدل پیشنهادی و مقایسه با سایر روش‌ها، مدل‌های پایه‌ای مانند LSTM، Bi-GRU، Transformer، TCN، GRU و CNN مورد بررسی قرار گرفتند. این مدل‌ها به دلیل توانایی در یادگیری وابستگی‌های زمانی، گزینه‌های مناسبی برای تحلیل داده‌های حسگرهای هوشمند هستند.

مدل‌های ترکیبی مانند Transformer-GRU و Transformer-TCN با هدف بهبود استخراج ویژگی‌ها و مدیریت وابستگی‌های طولانی‌مدت طراحی شدند. این ترکیب‌ها دقت مدل، کاهش نیاز به داده‌های برچسب‌خورده و پایداری عملکرد را افزایش می‌دهند.

در آزمایش‌ها، مدل TCN با دقت ۹۲.۸۴٪ در آروبا و ۷۲.۳۲٪ در میلان بهترین عملکرد را در همه معیارها داشته است. پس از آن، Transformer و Bi-GRU نیز نتایج قابل قبولی ارائه داده‌اند. مدل Bi-GRU نسبت به GRU عملکرد بهتری ارائه داده است، به‌طوری‌که دقت آن در مجموعه داده آروبا ۸۹.۹۸٪ و در مجموعه داده میلان ۶۹.۳۱٪ گزارش شده است. این بهبود نشان می‌دهد که استفاده از GRU دوطرفه می‌تواند وابستگی‌های زمانی را بهتر مدیریت کرده و دقت مدل را افزایش دهد.

در مقابل، مدل‌های LSTM و RNN عملکرد ضعیف‌تری به‌ویژه در داده‌های میلان نشان داده‌اند. داده‌های آروبا به دلیل تنوع و کیفیت بالاتر، شرایط بهتری برای آموزش مدل‌ها فراهم کرده‌اند. به‌طور کلی، نتایج نشان می‌دهد که مدل‌های TCN، Transformer و Bi-GRU به‌واسطه توانایی در یادگیری الگوهای رفتاری پیچیده، گزینه‌های مناسبی برای شناسایی فعالیت‌های انسانی در محیط‌های هوشمند هستند. در جدول ۴، عملکرد مدل‌های پایه سه‌لایه‌ای بدون ترکیب با ترنسفورمر ارائه شده است.

جدول ۴. نتایج اعمال مدل‌های پایه در شناسایی فعالیت‌های دو مجموعه داده آروبا و میلان

معیار ارزیابی (%)	مدل	مجموعه داده	
		آروبا	میلان
دقت	LSTM	۷۹.۴۴	۶۲.۲۳
	TCN	۹۲.۸۴	۷۲.۳۲

۶۱.۷۶	۸۹.۰۷	CNN	
۶۳.۳۲	۷۹.۷۰	RNN	
۶۲.۰۷	۸۵.۳۰	GRU	
۶۹.۳۱	۸۹.۹۸	Bi-GRU	
۶۹.۶۱	۹۰.۷۶	Transformer	
۶۲.۵۶	۸۰.۶۶	LSTM	صحت
۷۵.۵۲	۹۳.۹۹	TCN	
۷۳.۸۷	۸۹.۲۶	CNN	
۶۶.۹۶	۸۱.۵۸	RNN	
۵۸.۸۰	۸۶.۱۷	GRU	
۶۸.۸۵	۹۰.۴۵	Bi-GRU	
۶۹.۲۵	۹۰.۸۵	Transformer	
۶۲.۲۳	۷۹.۴۴	LSTM	
۷۲.۳۲	۹۲.۸۴	TCN	
۶۱.۷۶	۸۹.۰۷	CNN	
۶۳.۳۲	۷۹.۷۰	RNN	یادآوری
۶۲.۰۷	۸۵.۳۰	GRU	
۶۹.۳۱	۸۹.۹۸	Bi-GRU	
۶۹.۶۱	۹۰.۷۶	Transformer	
۵۶.۳۳	۷۶.۱۸	LSTM	
۷۰.۲۸	۹۱.۸۲	TCN	
۵۴.۲۱	۸۷.۶۵	CNN	
۵۷.۳۷	۷۶.۳۶	RNN	
۵۵.۷۹	۸۳.۱۷	GRU	
۶۶.۴۳	۸۸.۳۶	Bi-GRU	
۶۶.۷۴	۸۹.۹۶	Transformer	امتیاز F1

تمامی مدل‌های ترکیبی مورد بررسی از سه لایه پایه تشکیل شده‌اند. در جدول ۵ عملکرد مدل‌ها ارائه شده است. نتایج این آزمایش نشان می‌دهند که ترکیب مدل‌ها با ترنسفورمر موجب بهبود قابل توجه در شناسایی فعالیت‌های انسانی شده است. در این میان، مدل Bi-GRU-Transformer با دقت ۹۵.۱۹٪ در آروبا و ۸۰.۰۱٪ در میلان، بهترین عملکرد را در بین تمامی مدل‌ها داشته است. در شرایطی که فعالیت «عدم فعالیت» (No Activity) از مجموعه داده‌ها حذف شده است، مدل پیشنهادی با وجود کاهش نسبی در دقت کلی، همچنان عملکرد بهتری نسبت به سایر مدل‌های پایه از خود نشان داده است. این موضوع نشان‌دهنده توانایی بالای مدل در شناسایی فعالیت‌های انسانی حتی در غیاب کلاس غالب است. این مدل در معیارهای دیگری مانند صحت و F1 نیز برتری خود را حفظ کرده و توانسته از قدرت ترنسفورمر در مدیریت وابستگی‌های طولانی و Bi-GRU در درک توالی زمانی بهره‌برد. در مقایسه، مدل‌هایی مانند CNN-

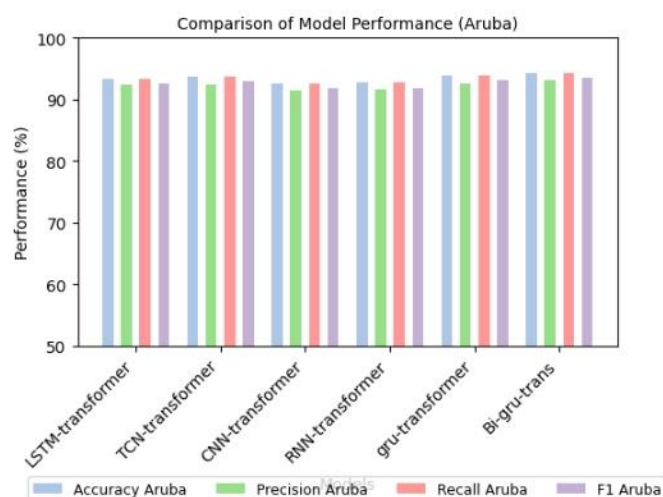
Transformer و RNN-Transformer نیز عملکرد خوبی داشتند، اما دقت کمتری نسبت به Bi-GRU-Transformer نشان دادند. در مجموع، ترکیب Bi-GRU با Transformer مؤثرترین معماری در این پژوهش بوده و نتایج، اهمیت کیفیت و تنوع داده‌ها را نیز در بهبود عملکرد مدل‌ها برجسته می‌سازد.

جدول ۵. نتایج اعمال مدل‌های ترکیبی در شناسایی فعالیت‌های دو مجموعه داده آروبا و میلان

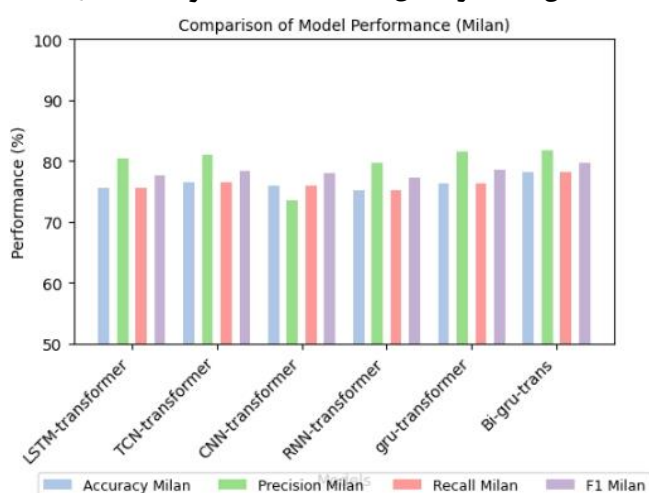
مجموعه داده		مدل	معیار ارزیابی (%)
میلان	آروبا		
۷۱.۷۳	۹۳.۰۶	LSTM-transformer	دقت
۷۲.۲۰	۹۲.۶۴	TCN-transformer	
۷۵.۱۲	۹۱.۵۱	CNN-transformer	
۷۵.۰۱	۹۲.۸۵	RNN-transformer	
۷۶.۶۳	۹۲.۹۲	GRU-Transformer	
۷۸.۹۶	۹۳.۴۰	Bi-GRU-Transformer-without noactivity	
۸۰.۰۱	۹۵.۱۹	Bi-GRU-Transformer	
۷۹.۰۵	۹۲.۹۷	LSTM-transformer	صحت
۷۸.۰۵	۹۲.۸۹	TCN-transformer	
۸۰.۲۰	۸۹.۷۴	CNN-transformer	
۸۱.۰۵	۹۲.۹۸	RNN-transformer	
۸۱.۲۷	۹۲.۷۲	GRU-Transformer	
۸۲.۴۴	۹۳.۴۷	Bi-GRU-Transformer-without noactivity	
۸۲.۷۶	۹۵.۲۳	Bi-GRU-Transformer	
۷۱.۷۳	۹۳.۰۶	LSTM-transformer	یادآوری
۷۲.۲۰	۹۲.۶۴	TCN-transformer	
۷۵.۱۲	۹۱.۵۱	CNN-transformer	
۷۵.۰۱	۹۲.۸۵	RNN-transformer	
۷۶.۶۳	۹۲.۹۲	GRU-Transformer	
۷۸.۹۶	۹۳.۴۰	Bi-GRU-Transformer-without noactivity	
۸۰.۰۱	۹۵.۱۹	Bi-GRU-Transformer	
۷۴.۹۰	۹۲.۳۵	LSTM-transformer	امتیاز F1
۷۴.۴۵	۹۲.۰۵	TCN-transformer	
۷۶.۸۷	۸۹.۶۱	CNN-transformer	
۷۷.۷۵	۹۲.۲۹	RNN-transformer	
۷۸.۵۸	۹۱.۷۳	GRU-Transformer	
۸۰.۵۰	۹۲.۶۷	Bi-GRU-Transformer-without noactivity	
۸۱.۲۶	۹۴.۵۹	Bi-GRU-Transformer	

مدل‌های ترکیبی عملکرد بهتری نسبت به مدل‌های پایه داشتند. در میان مدل‌های پایه، Bi-GRU به دلیل توانایی در یادگیری وابستگی‌های دوطرفه، دقت بالاتری نسبت به سایر مدل‌های سری‌زمانی نشان داد. همچنین، ترنسفورمر با بهره‌گیری از مکانیزم توجه و پردازش موازی، عملکرد قابل توجهی ارائه کرد.

ترکیب مدل‌های پایه با ترنسفورمر، به‌ویژه مدل Bi-GRU-Transformer، موجب بهبود چشمگیر در معیارهایی چون دقت، یادآوری و امتیاز F1 شد. این مدل در تمامی مجموعه‌داده‌ها، به‌ویژه در آروبا، بهترین عملکرد را داشت و نشان داد که ترکیب وابستگی‌های زمانی با توجه سراسری می‌تواند روابط پیچیده فعالیت‌های انسانی را بهتر شناسایی کند. مدل‌های TCN-Transformer و CNN-Transformer نیز در برخی داده‌ها عملکرد رقابتی داشتند. نمودارهای تحلیلی نمایش داده شده در شکل ۹ و ۱۰ نیز برتری Bi-GRU-Transformer را در دقت و F1 نسبت به سایر مدل‌ها تأیید می‌کنند، در حالی که در مجموعه میلان تفاوت عملکرد مدل‌ها به ویژگی‌های خاص داده‌ها برمی‌گردد.

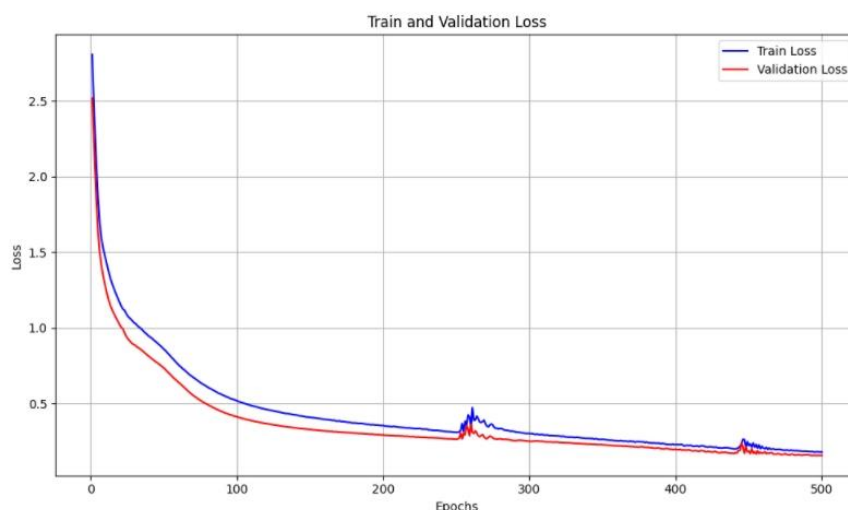


شکل ۹. عملکرد مدل‌های مختلف در مجموعه داده آروبا.

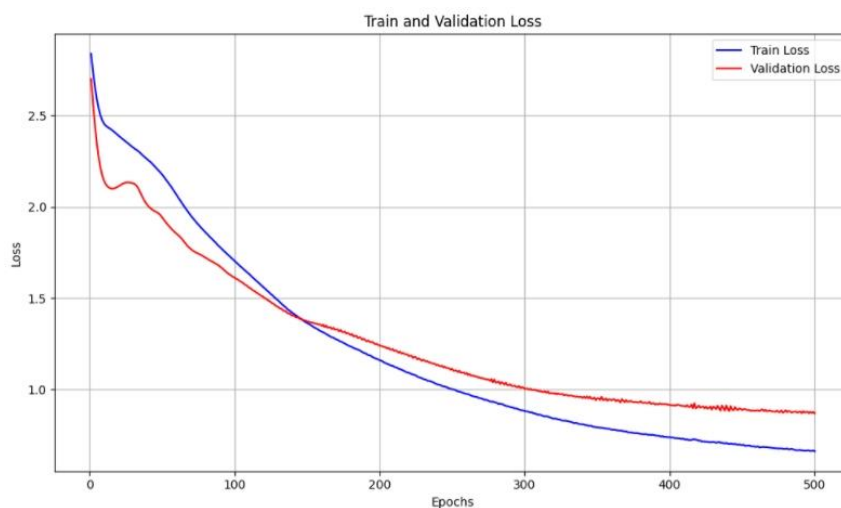


شکل ۱۰. عملکرد مدل‌های مختلف در مجموعه داده میلان.

نمودارهای رسم شده در شکل های ۱۱ و ۱۲ روند تغییرات تابع هزینه^۱ را در مراحل آموزش و آزمون مدل برای مجموعه داده ها نمایش می دهند. همان طور که مشاهده می شود، مقدار خطای آموزش، اعتبارسنجی و آزمون به تدریج کاهش یافته اند که نشان دهنده یادگیری مؤثر مدل است. همگرایی این مقادیر بیانگر آن است که مدل به خوبی آموزش دیده و دچار بیش برآزش^۲ نشده است.



شکل ۱۱. نمودار خطای مجموعه داده اروپا.



شکل ۱۲. نمودار خطای مجموعه داده میلان.

نمودار ماتریس درهم ریختگی برای مدل، نتایج پیش بینی ها را در قالب ماتریسی از تعداد پیش بینی های صحیح و نادرست به نمایش می گذارد. این نمودار که در شکل های ۱۳ و ۱۴ برای مجموعه داده های اروپا و میلان رسم شده است، امکان تحلیل دقیق عملکرد مدل در شناسایی هر کلاس را فراهم می سازد.

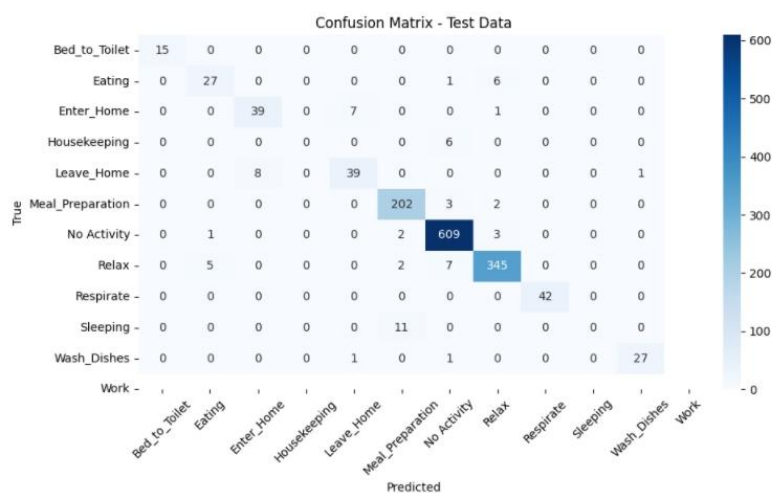
1 Loss

2 Overfitting

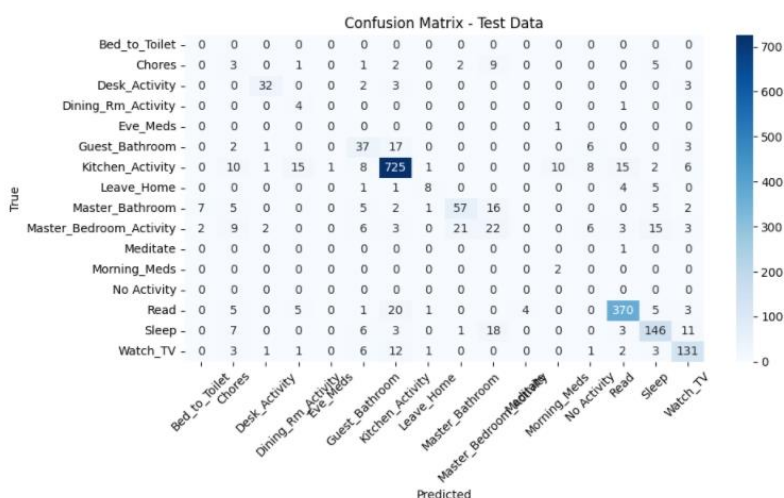
مطابق این نمودارها، کلاس‌هایی با فراوانی بالاتر، به‌ویژه کلاس "No Activity" بدون فعالیت، با دقت بیشتری شناسایی شده‌اند. مدل در هر دو مجموعه داده توانسته است این کلاس را با صحت بالا از سایر فعالیت‌ها تفکیک کند، که نشان‌دهنده کارایی آن در تشخیص وضعیت‌های فاقد فعالیت خاص است. این موضوع از نظر کاربردی اهمیت زیادی دارد، چرا که مدل باید قادر باشد بازه‌های فاقد فعالیت را به‌درستی شناسایی و از دیگر کلاس‌ها متمایز کند.

در مقابل، برخی کلاس‌ها مانند "ورود به خانه" در آروبا و "کارهای روزانه" در میلان، به‌دلیل تعداد محدود نمونه‌های آموزشی، با دقت کمتری شناسایی شده‌اند. همچنین، فعالیت‌هایی با مدت‌زمان کوتاه یا ویژگی‌های ناکافی نیز به‌طور کامل شناسایی نشده یا با کلاس‌های دیگر اشتباه گرفته شده‌اند. این موارد در ماتریس، به‌صورت مقادیر صفر یا نرخ بالایی پیش‌بینی نادرست نمایان است.

در مجموع، تحلیل این ماتریس‌ها نشان می‌دهد که عملکرد مدل در شناسایی کلاس‌ها تا حد زیادی تحت تأثیر توازن داده‌ها و کیفیت ویژگی‌های استخراج شده است. بنابراین، به‌منظور بهبود عملکرد کلی مدل، به‌ویژه در شناسایی فعالیت‌های پیچیده یا کم‌تکرار، استفاده از داده‌های متوازن‌تر و استخراج ویژگی‌های غنی‌تر ضروری است.



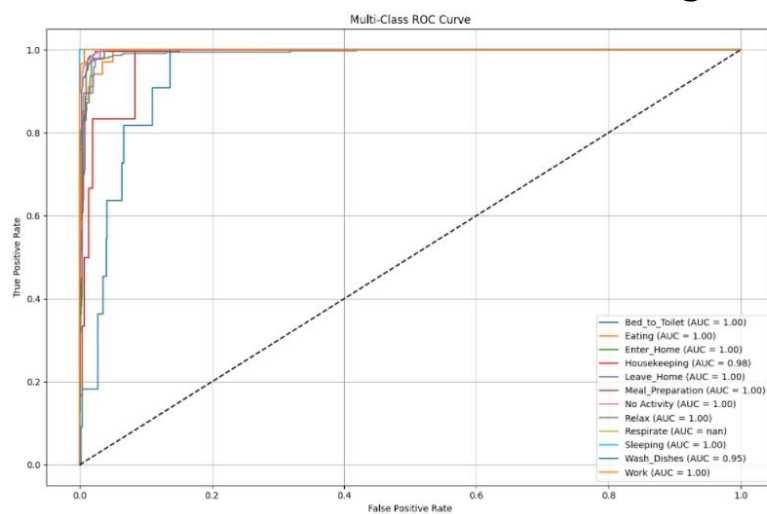
شکل ۱۳. ماتریس درهم ریختگی مجموعه داده آروبا.



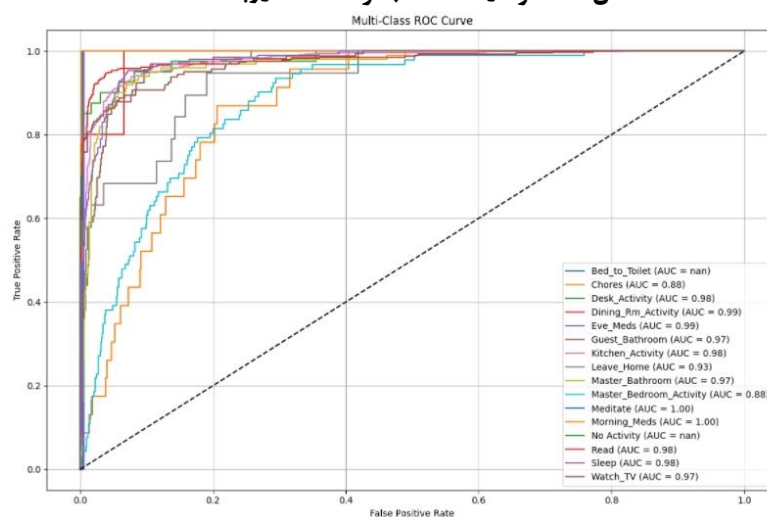
شکل ۱۴. ماتریس درهم ریختگی مجموعه داده میلان.

برای ارزیابی دقیق تر عملکرد مدل در تفکیک فعالیت‌ها، از نمودار ROC^۱ استفاده شده است. این نمودار رابطه میان نرخ مثبت واقعی^۲ و نرخ مثبت کاذب^۳ را در آستانه‌های تصمیم‌گیری مختلف نشان می‌دهد. مساحت زیر منحنی ROC^۴ که با AUC^۴ مشخص می‌شود، معیار مهمی برای سنجش توان مدل در تشخیص صحیح کلاس‌ها است؛ به طوری که مقدار AUC برابر با ۱ بیانگر عملکرد ایده‌آل، و مقدار نزدیک به ۰.۵ نشان‌دهنده عملکرد تصادفی است.

در شکل‌های ۱۵ و ۱۶، منحنی‌های ROC برای مدل پیشنهادی در دیتاست آروبا و میلان ترسیم شده‌اند و نشان می‌دهند که مدل در تفکیک صحیح کلاس‌های مختلف، عملکرد قابل قبولی دارد.



شکل ۱۵. نمودار ROC مجموعه داده آروبا.



شکل ۱۶. نمودار ROC مجموعه داده میلان.

1 Receiver Operating Characteristic
2 True Positive Rate
3 False Positive Rate
4 Area Under Curve

جدول ۶، دقت مدل پیشنهادی را با پژوهش‌های پیشین مقایسه می‌کند. همان‌طور که مشاهده می‌شود، مدل پیشنهادی عملکرد بهتری نسبت به روش‌های قبلی داشته است. این بهبود حاصل ترکیب ساختارهای زمانی و توجه سراسری در مدل Bi-GRU-Transformer و در نظر گرفتن کلاس عدم فعالیت جهت جداسازی و تشخیص بازه‌های شبه‌فعالیتی است که ممکن است به اشتباه به عنوان یک فعالیت موجود شناسایی شوند. همچنین با تنظیم دقیق فرآیندهای مدل بهینه‌ترین مدل ترکیبی حاصل یاد گرفته شده است.

جدول ۶. مقایسه دقت مدل پیشنهادی با روش‌های پیشین

مقاله	معماری	توضیح روش	دیتاست	
			آروبا	میلان
(Chen et al., 2022)	TRANS-BiGRU	ترکیب لایه‌های BiGRU با لایه توجه (Attention) در معماری Transformer	۹۱.۵۳	۷۸.۹۱
(Wang et al., 2024)	CNN-GRU-Transformer	استفاده از CNN برای استخراج ویژگی و ترکیب GRU و Transformer برای شناسایی	۹۳.۱۴	۶۱.۹۵
(Mohsen, 2023)	GRU	استفاده از دو لایه GRU ساده	۹۳.۰۶	۴۴.۹۲
(Liu et al., 2023)	TransTM	استفاده از تکنیک‌های pooling و pushing در پیش‌پردازش و ترکیب CNN و Transformer	۹۱.۱۵	۴۱.۶۴
مدل پیشنهادی	Bi-GRU-Transformer	ترکیب معماری Bi-GRU با لایه توجه سراسری جهت بهره‌گیری از وابستگی‌های زمانی و با در نظر گرفتن بازه‌های عدم فعالیت به عنوان کلاس جداگانه	۹۵.۱۹	۸۰.۰۱

تحلیل نتایج نشان می‌دهد که مدل پیشنهادی Bi-GRU-Transformer، با بهره‌گیری هم‌زمان از قابلیت مدل‌های Bi-GRU و ترنسفورمر، عملکرد بهتری در شناسایی فعالیت‌های انسانی نسبت به مدل‌های پایه از خود نشان داده است. ترکیب وابستگی‌های زمانی (کوتاه‌مدت و بلندمدت) با مکانیزم توجه سراسری، موجب بهبود قابل توجهی در دقت شده است.

۵) نتیجه‌گیری و پیشنهادها

این پژوهش اثربخشی مدل ترکیبی ترنسفورمر و Bi-GRU را در بهبود شناسایی فعالیت‌های انسانی (HAR) در محیط‌های هوشمند نشان می‌دهد. مدل پیشنهادی با بهره‌گیری از توانایی ترنسفورمر در مدل‌سازی وابستگی‌های بلندمدت و قدرت Bi-GRU در استخراج الگوهای زمانی کوتاه‌مدت، در مواجهه با داده‌های پیچیده و واقعی مانند مجموعه‌های آروبا و میلان، عملکرد قابل توجهی ارائه داده است. به‌طور خاص، دقت ۹۵.۱۹٪ برای آروبا و ۸۰.۰۱٪ برای میلان حاصل شد که نشان‌دهنده توانایی مدل در تعمیم‌پذیری به محیط‌های مختلف و داده‌های متنوع است.

با این وجود، ترکیب این دو معماری چالش‌هایی را نیز به همراه دارد؛ از جمله افزایش پیچیدگی محاسباتی، دشواری در هم‌راستاسازی خروجی Bi-GRU با ورودی ترنسفورمر، و حساسیت مدل به کیفیت پایین یا نویزی داده‌ها. همچنین،

نیاز به داده‌های برجسته‌دار زیاد، یکی از موانع مهم در توسعه مدل‌های مبتنی بر یادگیری عمیق برای HAR محسوب می‌شود. برای مقابله با این چالش‌ها، به‌کارگیری روش‌هایی نظیر داده‌افزایی^۱، یادگیری انتقالی^۲، و طراحی معماری‌های سبک‌تر و بهینه‌شده از نظر مصرف منابع پیشنهاد می‌شود. همچنین، استفاده از داده‌های چندوجهی شامل حسگرهای پوشیدنی، داده‌های موبایل و تصاویر می‌تواند تنوع ورودی را افزایش داده و دقت شناسایی فعالیت‌ها را در شرایط واقعی ارتقا دهد.

در مسیر توسعه آتی، تمرکز بر طراحی مدل‌های مقیاس‌پذیر و قابل اجرا در زمان واقعی، ترکیب داده‌های چندمنبعی^۳، و بهره‌گیری از چارچوب‌هایی مانند یادگیری فدرال می‌تواند باعث افزایش اعتماد، حریم خصوصی و کاربردپذیری این سیستم‌ها در حوزه‌هایی مانند سلامت هوشمند، پایش سالمندان، امنیت خانگی و تشخیص رفتارهای غیرعادی شود.

منابع

- Alsarhan, T., Alawneh, L., Al-Zinati, M., & Al-Ayyoub, M. (2019, 27-30 Oct. 2019). Bidirectional Gated Recurrent Units For Human Activity Recognition Using Accelerometer Data. 2019 IEEE SENSORS , 23716-23727. <https://doi.org/https://doi.org/10.1109/JIOT.2022.3190307>
- Anbazhagan, K., Swamy, G., Janani, R., & Farakte, A. (2024, 22-23 March 2024). Deep Learning based Human Activity Recognition in Smart Home. 2024 4th International Conference on Data Engineering and Communication Systems (ICDECS) ,
- Chen, D., Yongchareon, S., Lai, E. M.-K., Yu, J., Sheng, Q. Z., & Li, Y. (2022). Transformer with bidirectional GRU for nonintrusive, sensor-based activity recognition in a multiresident environment. IEEE Internet of Things Journal, 9(23), 23716-23727. <https://doi.org/https://doi.org/10.1109/ACCESS.2019.2936817>
- Chen, J., Jiang, D., & Zhang, Y. (2019). A hierarchical bidirectional GRU model with attention for EEG-based emotion classification. IEEE Access, 7, 118530-118540. <https://doi.org/https://doi.org/10.1109/ACCESS.2019.2936817>
- Chitty-Venkata, K. T., Mittal, S., Emani, M., Vishwanath, V., & Somani, A. K. (2023). A survey of techniques for optimizing transformer inference. Journal of Systems Architecture, 102990. <https://doi.org/https://doi.org/10.1016/j.sysarc.2023.102990>
- Choudhury, N. A., & Soni, B. (2023). An Adaptive Batch Size based-CNN-LSTM Framework for Human Activity Recognition in Uncontrolled Environment. IEEE Transactions on Industrial Informatics. <https://doi.org/https://doi.org/10.1109/TII.2022.3229522>
- Cook, D. J. (2010). Learning Setting-Generalized Activity Models for Smart Spaces. IEEE Intell Syst, 2010. ۱, (۹۹) <https://doi.org/10.1109/mis.2010.112>
- Hussain, A., Hussain, T., Ullah, W., & Baik, S. W. (2022). Vision Transformer and Deep Sequence Learning for Human Activity Recognition in Surveillance Videos. Computational Intelligence and Neuroscience, 2022, 3454167. <https://doi.org/10.1155/2022/3454167>
- Jiang, L., Wu, M., Che, L., Xu, X., Mu, Y., & Wu, Y. (2023). Continuous Human Motion Recognition Based on FMCW Radar and Transformer. Journal of Sensors, 2023. <https://doi.org/https://doi.org/10.1155/2023/2951812>
- Kumar, P., Chauhan, S., & Awasthi, L. K. (2024). Human Activity Recognition (HAR) (Using Deep Learning: Review, Methodologies, Progress and Future Research Directions. Archives of Computational Methods in Engineering, 31(1), 179-219. <https://doi.org/https://doi.org/10.1007/s11831-023-09986-x>
- Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. ACM SigKDD Explorations Newsletter, 12(2), 74-82. <https://doi.org/https://doi.org/10.1145/1964897.1964918>
- Le, T.-H., Tran, T.-H., & Pham, C. (2022). Human action recognition from inertial sensors with Transformer. 2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR) ,
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Lee, T.-H., Kim, H & Lee, D. (2023). Transformer based Early Classification for Real-time Human Activity Recognition in Smart Homes. Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing ,

1 data augmentation

2 transfer learning

3 multimodal data integration

- Lesani, F. S., & Fatahi, R. (2024). A Review on Transformer-Based Methods for Human Activity Recognition. *International Journal of Web Research*, 7(4), 81-100. <https://doi.org/10.22133/ijwr.2024.485291.1244>
- Liu, B., & Fang, S. (2023). Multi-level wavelet network based on CNN-Transformer hybrid attention for single image deraining. *Neural Computing and Applications*, 35(30), 22387-22404. <https://doi.org/10.1007/s00521-023-08899-x>
- Liu, Y., Huang, W., Jiang, S., Zhao, B., Wang, S., Wang, S., & Zhang, Y. (2023). TransTM: A device-free method based on time-streaming multiscale transformer for human activity recognition. *Defence Technology*. <https://doi.org/https://doi.org/10.1016/j.dt.2023.02.021>
- Mohsen, S. (2023). Recognition of human activity using GRU deep learning algorithm. *Multimedia Tools and Applications*, 82(30), 47733-477. <https://doi.org/10.1007/s11042-023-15571-y>
- Pan, J., Hu, Z., Yin, S., & Li, M. (2022). GRU with dual attentions for sensor-based human activity recognition. *Electronics*, 11(11), 1797.
- Pareek, G., Nigam, S., & Singh, R. (2024). Modeling transformer architecture with attention layer for human activity recognition. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-023-09362-7>
- Pramanik, R., Sikdar, R., & Sarkar, R. (2023). Transformer-based deep reverse attention network for multi-sensory human activity recognition. *Engineering Applications of Artificial Intelligence*, 122, 106150. <https://doi.org/https://doi.org/10.1016/j.engappai.2023.106150>
- Saidani, O., Alsafyani, M., Alroobaea, R., Alturki, N., Jahangir, R., & Menzli, L. J. (2023). An Efficient Human Activity Recognition using Hybrid Features and Transformer Model. *IEEE Access*. <https://doi.org/https://doi.org/10.1109/ACCESS.2023.3314492>
- Sharifi-Renani, M., Mahoor, M. H., & Clary, C. W. (2023). BioMAT: An Open-Source Biomechanics Multi-Activity Transformer for Joint Kinematic Predictions Using Wearable Sensors. *Sensors*, 23(13), 5778. <https://doi.org/https://doi.org/10.3390/s23135778>
- Shavit, Y., & Klein, I. (2021). Boosting inertial-based human activity recognition with transformers. *IEEE Access*, 9, 53540-53547. <https://doi.org/https://doi.org/10.1109/ACCESS.2021.3070646>
- Sunil, A., Sheth, M. H., & Shreyas, E. (2021). Usual and unusual human activity recognition in video using deep learning and artificial intelligence for security applications. 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT),
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Wan, J., O'grady, M. J., & O'Hare, G. M. (2015). Dynamic sensor event segmentation for real-time activity recognition in a smart home context. *Personal and Ubiquitous Computing*, 19, 287-301. <https://doi.org/https://doi.org/10.1007/s00779-014-0824-x>
- Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern recognition letters*, 119, 3-11. <https://doi.org/https://doi.org/10.1016/j.patrec.2018.02.010>
- Wang, L., Zhang, Z., Wei, L., & Zhou, Y. (2024). CNN-GRU-Transformer Human Activity Recognition Model Based on Feature Fusion. 2024 6th International Conference on Frontier Technologies of Information and Computer (ICFTIC),
- Wojek, C., Dorkó, G., Schulz, A., & Schiele, B. (2008). Sliding-windows for rapid object class localization: A parallel technique. *Joint Pattern Recognition Symposium*.