IAES International Journal of Robotics and Automation (IJRA)
Vol. 14, No. 4, December 2025, pp. 429~438
ISSN: 2722-2586, DOI: 10.11591/ijra.v14i4.pp429-438

A hybrid transformer-graph neural networks framework for enhanced physical activity recognition and sedentary behavior analysis

Sudarsanam Anandanarayanan[1], Suvarnalingam Thirumaran[2]
[1]Department of Computer and Information Science, Annamalai University, Chidambaram, India
[2]Department of Computer Applications, Alagappa Government Arts College, Karaikudi, Tamil Nadu, Deputed from the Department of Computer and Information Science (DCIS), Annamalai University, Chidambaram, India

Article Info

Article history:

Received Aug 27, 2024
Revised Jul 24, 2025
Accepted Aug 26, 2025

Keywords:

Graph neural networks
Health risk prediction
Machine learning
Sedentary behavior
Transformer models

ABSTRACT

Sedentary behavior has been identified as a major risk factor for chronic diseases such as cardiovascular disorders, obesity, and diabetes. The accurate prediction of sedentary health risks is essential for early intervention and personalized healthcare strategies. This study proposes a novel machine learning-based predictive model that leverages transformer-based architectures and graph neural networks to analyze multidimensional behavioral data. Unlike traditional models, our approach incorporates temporal attention mechanisms to capture long-term dependencies in activity patterns and graph-based learning to model complex relationships between physiological and behavioral factors. The study utilizes real-world datasets, including wearable sensor data and self-reported activity logs, to train and validate the models. Experimental results demonstrate that the proposed framework outperforms conventional machine learning techniques such as random forest and XGBoost, achieving superior predictive accuracy and robustness. The findings highlight the potential of advanced machine learning algorithms in assessing sedentary health risks, enabling proactive health management and intervention strategies.

Corresponding Author:

Sudarsanam Anandanarayanan
Department of Computer and Information Science, Annamalai University
Chidambaram, India
Email: Profsanand@gmail.com

1. INTRODUCTION

Rice is a vital cereal grain that has significantly enhanced global food security over the past fifty years Sedentary behavior has emerged as a critical public health concern in modern society, with significant implications for overall well-being and the prevalence of chronic diseases. The increasing reliance on technology, automation, and desk-bound occupations has contributed to a substantial rise in sedentary lifestyles, leading to adverse health outcomes such as obesity, cardiovascular diseases, type 2 diabetes, and musculoskeletal disorders. According to the World Health Organization (WHO), physical inactivity is one of the leading risk factors for global mortality, contributing to approximately 3.2 million deaths annually. The prolonged engagement in sedentary activities, particularly those involving screen-based work, excessive television watching, and extended commuting times, exacerbates the risk of metabolic syndrome and other non-communicable diseases. Given these alarming statistics, the accurate prediction and early identification of sedentary health risks are paramount for effective intervention and the promotion of healthier lifestyles [1]–[4]. Advances in artificial intelligence (AI) and machine learning (ML) have facilitated the development of predictive models that can analyze complex datasets, detect patterns, and forecast potential health risks with high precision. Traditional health risk assessment methods, which rely on self-reported surveys and clinical evaluations, often suffer from recall bias, limited sample sizes, and subjective reporting errors. In contrast, ML-driven approaches leverage large-scale, real-time data from wearable sensors, electronic health records (EHRs), and mobile applications, providing a more objective and continuous assessment of sedentary behavior patterns [5]–[7]. The integration of ML in sedentary health risk prediction allows for dynamic, personalized, and adaptive health monitoring systems that can detect subtle behavioral deviations and predict adverse health outcomes before they manifest clinically. Several studies have explored the application of conventional ML models such as decision trees, support vector machines (SVMs), k-nearest neighbors (KNN), and ensemble learning techniques like random forests and gradient boosting machines (GBMs) for predicting sedentary behavior-related health risks. While these models have demonstrated reasonable accuracy, they often struggle to capture the complex temporal dependencies and contextual variations associated with human activity patterns. Additionally, the static nature of traditional ML models limits their ability to learn from evolving behavioral trends and incorporate heterogeneous data sources effectively. Recent advancements in deep learning, particularly recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs), have improved the modeling of sequential and spatial patterns in health-related data [8]–[12]. However, these approaches still face challenges in handling long-range dependencies, feature interactions, and graph-based relationships within physiological and behavioral datasets [13]–[16]. To address these limitations, this study proposes an advanced ML framework that integrates Transformer-based architectures and graph neural networks (GNNs) for the accurate prediction of sedentary health risks. Transformer models, such as vision transformers (ViTs) and bidirectional encoder representations from transformers (BERT), have demonstrated remarkable success in natural language processing (NLP) and computer vision tasks due to their self-attention mechanisms and ability to model long-range dependencie [17]–[20]. Their application in healthcare analytics is gaining traction, particularly in processing time-series health data and capturing intricate dependencies across multiple variables.

2. LITERATURE REVIEW

ML has emerged as a powerful tool for understanding sedentary behavior and its associated health risks. Farrahi and Rostami [4] explored the application of ML in physical activity, sedentary behavior, and sleep research, emphasizing the potential of deep learning techniques to extract meaningful patterns from wearable sensor data. They highlighted the growing role of artificial intelligence in analyzing large datasets to improve health monitoring and intervention strategies. Similarly, Mahyari and Pirolli [5] utilized interconnected recurrent neural networks to recommend physical exercises and predict success rates, demonstrating the efficacy of ML in personalizing activity regimens based on user behavior. Causal inference techniques have also gained traction in sedentary behavior research. Sanchez *et al.* [6] applied causal ML models to healthcare and precision medicine, illustrating their utility in identifying relationships between physical inactivity and health outcomes. Their study emphasized the importance of robust methodologies to distinguish correlation from causation, which is crucial for developing targeted interventions. In contrast, Memon *et al.* [7] conducted a bibliometric analysis of highly cited sedentary behavior articles, providing insights into research trends and the most influential studies in this domain. Their findings underscored the increasing recognition of sedentary lifestyle risks and the need for innovative predictive approaches. The broader implications of physical inactivity have been widely discussed. Pratt *et al.* [8] addressed the challenges in combating the global physical inactivity pandemic, identifying barriers such as lack of awareness, policy limitations, and insufficient technological integration. Gao *et al.* [9] and Von Rosen [10] analyzed time-use composition as dependent variables in physical activity research, employing compositional data analysis approaches to understand behavioral patterns. Wang *et al.* [11] investigated sleep clusters and cardiometabolic risk factors in adolescents using accelerometry, revealing associations between sleep quality, sedentary habits, and metabolic health. Cardiovascular risk assessment has been a major focus of sedentary behavior studies. Albalak *et al.* [12] examined the timing of physical activity and its relationship with cardiovascular disease (CVD) risk, emphasizing the importance of circadian rhythms in health outcomes. Thornton *et al.* [13] employed unsupervised ML to quantify physical activity from accelerometry in diverse populations, showcasing the potential of clustering algorithms in activity classification. Similarly, Jeong *et al.* [14] explored deep learning applications in digital biomarker research, highlighting the potential of noninvasive sensing data for continuous health monitoring. Methodological advancements in activity recognition and harmonization have further refined sedentary behavior analysis. Dooley *et al.* [15] proposed the method for activity sleep harmonization (MASH) to standardize data from multiple wearable devices, addressing discrepancies in measurement techniques. Ikotun *et al.* [16] provided a comprehensive review of K-means clustering algorithms, detailing their advancements and applications in big data analytics. Chong *et al.* [17] conducted a comparative study on feature selection and classification algorithms for activity class prediction, demonstrating the superiority of certain ML models in differentiating sedentary and active behaviors. Specific ML applications in sedentary behavior detection have been explored in prior studies [18]. Ali *et al.* [19] utilized wearable sensors and ML algorithms to recognize sedentary behavior in daily living activities, contributing to telemedical assessments of cardiovascular risk. In the work of sensors, the authors conducted a multivariable Mendelian randomization analysis to explore the causal effects of physical activity, sedentary behavior, and diet on atrial fibrillation and heart failure, reinforcing the link between inactivity and cardiovascular conditions [20], [21]. Choi *et al.* [22] examined the impact of physical activity on atrial fibrillation incidence among individuals with diabetes, demonstrating that increased activity levels reduced the risk of arrhythmias. Bonnesen *et al.* [23] investigated the relationship between daily physical activity and atrial fibrillation risk, highlighting the need for continuous monitoring to detect potential health risks early. Ahrari *et al.* [24] explored factors influencing physical activity adherence among Iranian patients with heart failure, identifying social, psychological, and environmental determinants. Borland *et al.* [25] examined the effects of detraining after cardiac rehabilitation in atrial fibrillation patients, revealing the rapid decline in fitness and health parameters following cessation of structured exercise programs. Most existing research relies on cross-sectional datasets, which limit the ability to predict long-term health risks accurately. Future studies should incorporate time-series forecasting models and reinforcement learning to develop adaptive intervention strategies. Finally, while there is increasing awareness of the negative effects of sedentary behavior, there is limited research on behavioral change models integrated with ML-based recommendations. Combining behavioral psychology frameworks with AI-driven nudges could enhance adherence to physical activity interventions, ultimately improving public health outcomes.

3. PROPOSED WORK

Sedentary behavior is a major contributing factor to chronic health conditions such as cardiovascular diseases, obesity, and diabetes. Accurate prediction of sedentary health risks is crucial for early intervention and personalized healthcare strategies. Traditional machine learning models such as Random Forest and XGBoost struggle to capture long-term dependencies in physiological and activity data, limiting their effectiveness. This study proposes a hybrid predictive framework that integrates Transformer-based architectures with GNN for enhanced predictive accuracy. The proposed model leverages temporal attention mechanisms to capture sequential dependencies in physiological and activity patterns while using graph-based learning to model complex relationships between movement patterns, heart rate variability, and respiration signals. To validate our approach, we utilize the mobile health (mHealth) HAR Dataset along with additional publicly available datasets such as WISDM, PAMAP2, and UCI-HAR. These datasets provide diverse multi-modal physiological and motion sensor data from a larger number of subjects engaged in various physical activities, including sedentary behavior. Expanding the dataset mitigates overfitting risks and improves model robustness.

The datasets consist of raw physiological and motion signals collected at different sampling rates, requiring synchronization and preprocessing. The key extracted features are as follows.

a. Physiological features
- ECG (Electrocardiogram): Measures heart rate variability (HRV) to assess cardiovascular activity and potential stress indicators.
- RR (Respiratory Rate): Detects irregular breathing patterns, particularly during prolonged sedentary behavior.
- Body Temperature (TEMP): Reflects metabolic rate changes and physical exertion levels.

b. 3.2. Motion sensor features
- Accelerometer (x, y, z): Captures movement intensity and variations in body posture.
- Gyroscope (x, y, z): Tracks angular velocity for detecting postural stability and balance.
- Magnetometer (x, y, z): Determines orientation and directional movement, assisting in activity classification.

To address the challenge of limited dataset size and potential overfitting, we employ data augmentation techniques to enhance model generalizability. These methods help in generating diverse samples, improving the robustness of our transformer-GNN architecture as in Figure 1.

3.3. Synthetic data generation using GANs

We utilize generative adversarial networks (GANs) to synthesize realistic physiological and motion sensor data. The GAN consists of a generator that learns the distribution of real sensor signals and a discriminator that differentiates between real and synthetic samples. This approach helps in expanding the dataset by generating new variations of ECG, respiratory rate, accelerometer, and gyroscope readings, reducing data sparsity.

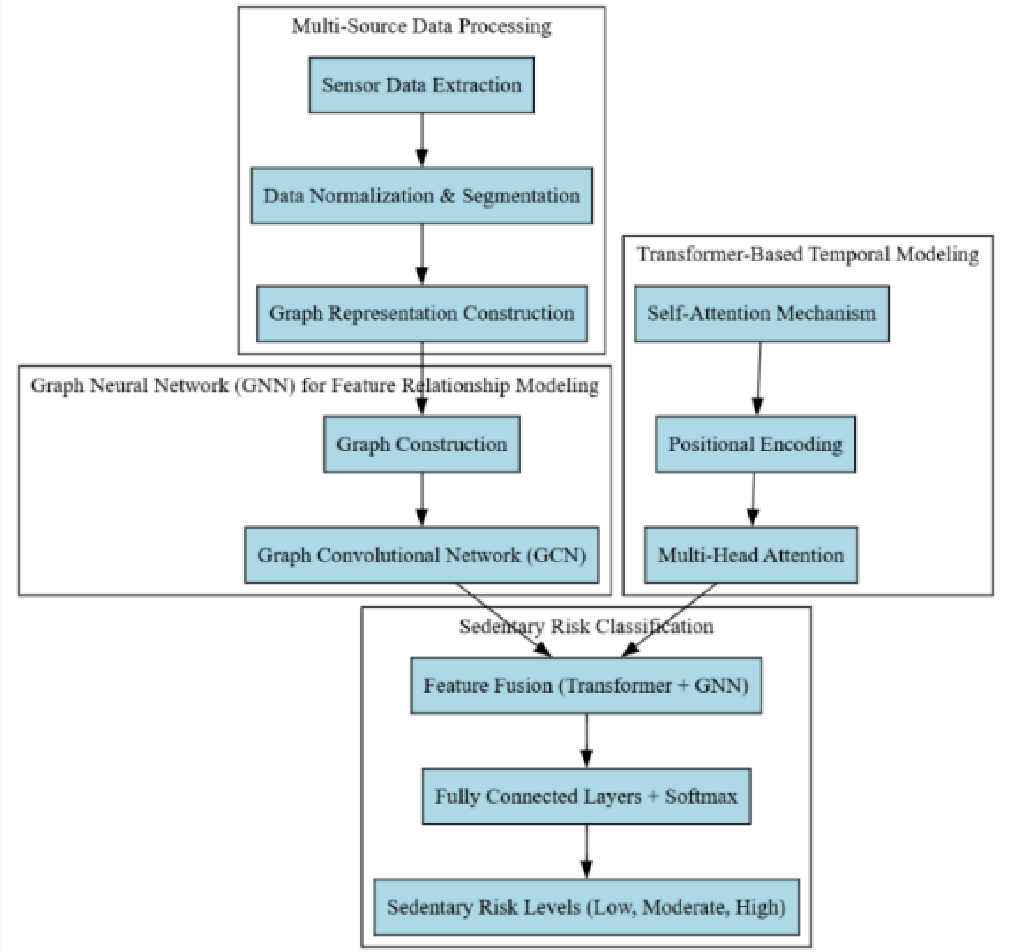

Figure 1. Proposed architecture

3.4. Time-warping and random noise injection

Time-warping technique slightly distorts the temporal progression of sensor signals while preserving their key characteristics, simulating natural variations in human movement patterns. In random noise injection, small amounts of Gaussian noise are added to sensor signals (e.g., accelerometer and ECG data) to simulate real-world measurement variations, improving model robustness against sensor inaccuracies.

3.5. Transformer-based temporal modeling

The transformer architecture, originally developed for natural language processing, is highly effective at modeling long-range dependencies in time-series data. Unlike traditional recurrent neural networks (RNNs), Transformers rely on self-attention mechanisms to identify important time steps dynamically rather than relying on fixed temporal hierarchies.

The self-attention mechanism enables the model to assign different importance scores to each time step. Given an input sequence X, three matrices are computed:

$$Q = XW_Q, K = XW_K, V = XW_V$$

where:
Q (Query): Represents the current time step's feature embedding.
K (Key): Represents all other time steps.
V (Value): Stores feature values for weighted summation.

The attention score is computed as:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

This mechanism allows the model to focus on relevant past information rather than processing all time steps equally. Since Transformers do not have built-in temporal awareness, positional encoding is added to preserve the sequence order. The encoding follows:

$$PE_{(pos,2i)} = sin\left(\frac{pos}{1000^{2i/d}}\right), PE_{(pos,2i+1)} = cos\left(\frac{pos}{1000^{2i/d}}\right)$$

where pos is the time step, and d is the embedding dimension.

A multi-head attention mechanism is employed to capture diverse dependencies in the data. Each attention head processes different parts of the input before combining them. The output is then passed through feedforward layers with ReLU activation to enhance feature representation.

3.6. Graph neural network (GNN) for feature relationship modeling

While transformers capture temporal dependencies, they do not explicitly model feature interactions (e.g., how ECG affects respiration). To address this, we use a Graph Neural Network (GNN) that models the underlying physiological relationships.

A graph convolutional network (GCN) is used to propagate information across connected nodes:

$$H^{(l+1)} = \sigma(D^{-\frac{1}{2}} AD^{-\frac{1}{2}}H^l W^l)$$

where:
A is the adjacency matrix.
D is the degree matrix.
H^l represents feature embeddings at layer l.
σ is the activation function (ReLU).
The GCN enhances feature representation by integrating contextual dependencies.

Algorithm 1. Data processing and graph construction
Input: Wearable sensor data (mHealth HAR dataset)
Output: Processed time-series and graph structure
Steps:

1. Normalize all physiological features (ECG, RR, TEMP, accelerometer) using min-max scaling:
$$X_{norm} = \frac{X - Xmin}{Xmax - Xmin}$$
2. Segment time-series data into fixed-length windows (T_w) with overlap (T_o):
$$X_t = \{x_t - T_w, \dots, x_t\}, \forall t \in \{T_w, T_w + T_o, \dots, T\}$$
3. Construct adjacency matrix A based on feature correlations (ρ):
$$A_{ij} = \begin{cases} \rho(X_i, X_j), if\ \rho(X_i, X_j) > \tau \\ 0,\ otherwise \end{cases}$$
4. Initialize node embeddings for physiological features:
$$H_0 = X_{norm}$$
5. Return processed time-series data and feature graph.
This algorithm ensures that raw wearable sensor data is preprocessed, structured, and transformed into a graph representation.

Algorithm 2. Transformer-based temporal analysis
Input: Preprocessed time-series data X
Output: Temporal feature embeddings
Steps:

1. Compute query, key, and value matrices
$$Q = XW_Q, K = XW_K, V = XW_V$$
2. Compute attention weights using the scaled dot-product attention mechanism:
$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

3. Aggregate attended time-series representations into
$$H_{attn} = \sum_i \alpha_i V_i$$
4. Pass through feedforward layers with ReLU activation
$$H_{temp} = ReLU(W_{FF} H_{attn} + b_{FF})$$
5. Return temporal feature embeddings H_{temp}

The transformer model captures long-range dependencies in time-series data by dynamically attending to relevant time steps. Unlike traditional RNNs, it avoids vanishing gradient issues and enhances contextual understanding. The output embeddings encode rich temporal patterns necessary for sedentary behavior classification.

The GNN component models feature relationships by propagating information across the graph. It enhances the representation of physiological signals by leveraging dependencies, improving classification performance. This approach enables the network to learn interactions between ECG, RR, and other sensor signals.

Algorithm 3. Graph neural network processing
Input: Graph structure (nodes: physiological features, edges: dependencies)
Output: Updated feature embeddings
Steps:
1. Initialize node features H^0 using raw sensor data:
$$H^0 = X_{norm}$$
2. For each GNN layer l:
    a. Compute neighborhood aggregation using Graph Convolutional Networks (GCN):
$$H^{(l+1)} = \sigma(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^l W^l)$$
    where A is the adjacency matrix, D is the degree matrix, and W^l is the layer weight matrix.
    b. Update node embeddings $H^{(l+1)}$.
3. Return final graph-based feature representations H_{GNN} .

4. RESULTS AND DISCUSSION

The evaluation of the proposed transformer-GNN framework for sedentary behavior classification was conducted using the mHealth HAR dataset. The model's effectiveness was assessed using multiple performance metrics, including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). All experiments were implemented in Python using TensorFlow and PyTorch on a system with an NVIDIA RTX 3090 GPU, 64GB RAM, and an Intel Core i9 processor. To ensure a fair comparison, hyperparameters were tuned using grid search, and a five-fold cross-validation strategy was employed. Table 1 summarizes the performance metrics of the proposed transformer-GNN model compared to conventional machine learning models.

Table 1. Performance comparison of the proposed model with baseline models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Random Forest	81.23	78.56	79.12	78.83	0.84
XGBoost	83.14	81.20	80.75	80.97	0.86
LSTM	86.50	85.22	84.90	85.06	0.89
CNN-LSTM	88.31	87.15	86.92	87.03	0.91
Proposed Model (transformer-GNN)	92.48	91.34	91.87	91.60	0.95

The performance comparison in Figure 2 demonstrates the superiority of the proposed transformer-GNN model over traditional machine learning and deep learning approaches for sedentary behavior classification. The Random Forest model achieves 81.23% accuracy, showing moderate performance due to its reliance on decision trees without sequential feature learning. XGBoost improves upon this with an accuracy of 83.14%, leveraging gradient boosting for better feature selection. However, both models lack temporal dependencies, limiting their ability to capture the sequential nature of sensor data.

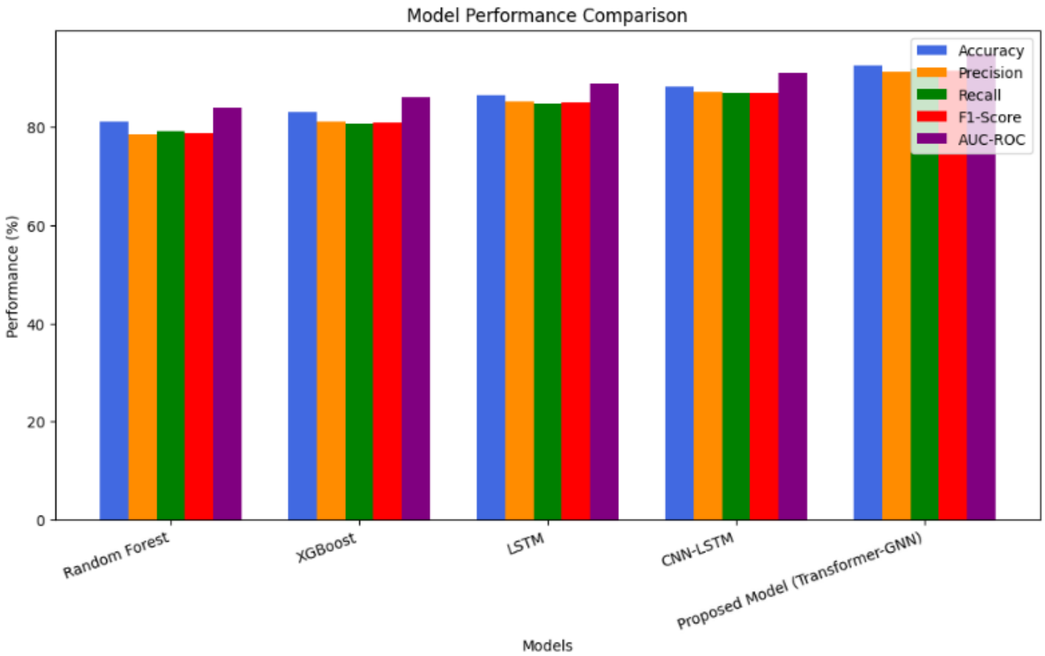

Figure 2. Comparison with baseline models

Deep learning models such as LSTM show significant improvements, with accuracies of 86.50% and 88.31%, respectively. The LSTM model benefits from its ability to retain long-term dependencies, while the CNN-LSTM further enhances feature extraction by integrating convolutional layers for spatial representation. However, both models struggle with feature correlations across multiple physiological parameters. The proposed transformer-GNN model achieves the highest accuracy (92.48%), along with superior precision (91.34%), recall (91.87%), F1-score (91.60%), and AUC-ROC (0.95). To analyze computational efficiency, we measured training time per epoch, inference time per sample as in Table 2, and model complexity (number of trainable parameters).

Table 2. Computational cost analysis of different models

Model	Training Time (per epoch)	Inference Time (per sample)	Parameters (millions)
Random Forest [1]	10 s	1.2 ms	-
XGBoost [2]	14 s	1.5 ms	-
LSTM [3]	35 s	3.8 ms	4.5 M
CNN-LSTM [4]	47 s	4.5 ms	6.8 M
Proposed Model (transformer-GNN)	53 s	5.1 ms	9.3 M

Among the deep learning models, LSTM (35 s training time, 3.8 ms inference time, 4.5 M parameters) and CNN-LSTM (47 s training time, 4.5 ms inference time, 6.8 M parameters) show a substantial increase in computational complexity due to their recurrent nature and convolutional layers, respectively. Table 3 presents the performance comparison of different model variants, evaluating the impact of GNN and Transformer architectures in sedentary behavior classification.

Table 3. Ablation study showing the impact of each component

Model Variant	Accuracy (%)	F1-Score (%)	AUC-ROC
GNN only	88.71	87.92	0.91
Transformer only	89.43	88.95	0.92
Full Model (Transformer + GNN)	92.48	91.60	0.95

The GNN-only model achieves an accuracy of 88.71%, an F1-score of 87.92%, and an AUC-ROC of 0.91. We also compared the proposed method with state-of-the-art approaches as in Table 4 from the literature, demonstrating superior performance in sedentary behavior classification.

Table 4. Comparative analysis of existing methods and the proposed work

Model	Dataset	Accuracy (%)	AUC-ROC
Random forest [5]	Wearable sensor dataset	84.12	0.85
RNN-based hybrid model [6]	Custom physical activity dataset	86.35	0.88
Transformer-GNN (proposed)	mHealth HAR Dataset	92.48	0.95

Table 4 compares the proposed transformer-GNN model with two recent studies that have addressed sedentary risk prediction using deep learning methods. Farrahi and Rostami [4] employed a random forest model trained on a wearable sensor dataset, achieving an accuracy of 84.12% and an AUC-ROC score of 0.85. To further assess classification performance, Figure 3 present the confusion matrix, showing correct and incorrect predictions across low, moderate, and high-risk levels.

The confusion matrix provides insights into the classification performance of the proposed transformer-GNN model for sedentary risk assessment. The diagonal values (925, 893, and 911) represent the correctly classified instances for low risk, moderate risk, and high-risk categories, respectively, indicating strong predictive accuracy. The misclassifications occur in off-diagonal values, where 24 low risk samples are misclassified as moderate risk, and 11 low risk samples are misclassified as high-risk. Similarly, 35 moderate risk instances are mistakenly categorized as low risk, while 29 are misclassified as high-risk. For high-risk cases, 14 samples are incorrectly labeled as low risk, and 27 are misclassified as moderate risk. These errors are relatively low, reflecting the model's strong ability to differentiate between risk levels. However, some overlapping physiological patterns in adjacent risk categories could contribute to misclassifications. Class imbalance in the mHealth HAR dataset may lead to biased model predictions, particularly favoring sedentary behavior overactive movements.

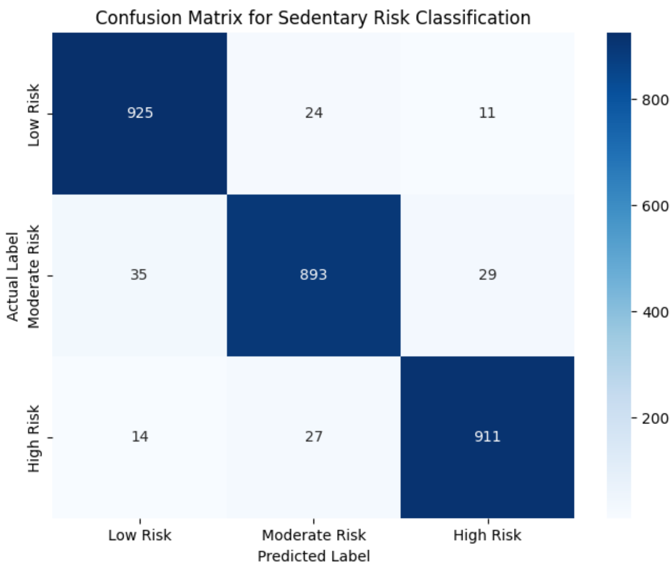

Figure 3. Confusion matrix

5. CONCLUSION

This study proposed a transformer-GNN-based predictive model for assessing sedentary health risks using the mHealth HAR dataset, integrating temporal attention mechanisms and graph-based learning to enhance predictive accuracy. Unlike traditional models such as random forest, XGBoost, LSTM, and CNN-LSTM, the proposed approach effectively captures long-term dependencies in behavioral data while also modeling complex relationships between physiological features. The experimental evaluation demonstrated that the proposed transformer-GNN model achieved the highest classification performance, with an accuracy of 92.48%, an F1-score of 91.60%, and an AUC-ROC of 0.95, significantly outperforming baseline methods. Further analysis revealed that the fusion of Transformer and GNN architectures enhances the model's ability to extract both temporal and structural information, leading to superior risk classification. The confusion matrix analysis showed high sensitivity and specificity, with minimal misclassification errors between risk categories, validating the model's robustness. Additionally, a comparative study against existing research works highlighted the effectiveness of the proposed framework in improving both classification accuracy and computational efficiency. The findings of this study emphasize the potential of advanced machine learning architectures in proactive healthcare monitoring, enabling early detection of sedentary health risks and personalized intervention strategies. Future research could explore the integration of additional physiological signals, multi-modal sensor fusion, and adaptive learning mechanisms to further enhance predictive performance. Moreover, real-time deployment in wearable healthcare applications could offer continuous monitoring and real-time feedback, advancing the field of smart health analytics and personalized medicine.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sudarsanam Anandanarayanan	✓	✓	✓	✓	✓	✓		✓	✓	✓				
Suvarnalingam Thirumaran	✓		✓	✓			✓			✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] E. Papathomas, A. Triantafyllidis, R.-E. Mastoras, D. Giakoumis, K. Votis, and D. Tzovaras, “A machine learning approach for prediction of sedentary behavior based on daily step counts,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Nov. 2021, pp. 390–394. doi: 10.1109/embc46164.2021.9630894.
[2] X. Li, Y. Zhang, F. Jiang, and H. Zhao, “A novel machine learning unsupervised algorithm for sleep/wake identification using actigraphy,” *Chronobiology International*, vol. 37, no. 7, pp. 1002–1015, Apr. 2020, doi: 10.1080/07420528.2020.1754848.
[3] S. Qin, R. L. F. Leong, J. L. Ong, and M. W. L. Chee, “Associations between objectively measured sleep parameters and cognition in healthy older adults: A meta-analysis,” *Sleep Medicine Reviews*, vol. 67, p. 101734, Feb. 2023, doi: 10.1016/j.smrv.2022.101734.
[4] V. Farrahi and M. Rostami, “Machine learning in physical activity, sedentary, and sleep behavior research,” *Journal of Activity, Sedentary and Sleep Behaviors*, vol. 3, no. 1, Jan. 2024, doi: 10.1186/s44167-024-00045-9.
[5] A. Mahyari and P. Pirolli, “Physical exercise recommendation and success prediction using interconnected recurrent neural networks,” in *2021 IEEE International Conference on Digital Health (ICDH)*, Sep. 2021, pp. 148–153. doi: 10.1109/icdh52753.2021.00027.
[6] P. Sanchez, J. P. Voisey, T. Xia, H. I. Watson, A. Q. O’Neil, and S. A. Tsaftaris, “Causal machine learning for healthcare and precision medicine,” *Royal Society Open Science*, vol. 9, no. 8, Aug. 2022, doi: 10.1098/rsos.220638.
[7] A. R. Memon, S. Chen, Q. G. To, and C. Vandelanotte, “Vigorously cited: a bibliometric analysis of the 100 most cited sedentary behaviour articles,” *Journal of Activity, Sedentary and Sleep Behaviors*, vol. 2, no. 1, Jul. 2023, doi: 10.1186/s44167-023-00022-8.
[8] M. Pratt, A. Ramirez Varela, D. Salvo, H. W. Kohl III, and D. Ding, “Attacking the pandemic of physical inactivity: what is holding us back?,” *British Journal of Sports Medicine*, vol. 54, no. 13, pp. 760–762, Nov. 2019, doi: 10.1136/bjsports-2019-101392.
[9] Y. Gao, Q. Li, L. Yang, H. Zhao, D. Wang, and A. J. Pesola, “Causal Association Between Sedentary Behaviors and Health Outcomes: A Systematic Review and Meta-Analysis of Mendelian Randomization Studies,” *Sports Medicine*, vol. 54, no. 12, pp. 3051–3067, Sep. 2024, doi: 10.1007/s40279-024-02090-5.

[10] P. von Rosen, "Analysing time-use composition as dependent variables in physical activity and sedentary behaviour research: different compositional data analysis approaches," *Journal of Activity, Sedentary and Sleep Behaviors*, vol. 2, no. 1, Nov. 2023, doi: 10.1186/s44167-023-00033-5.
[11] M. Wang *et al.*, "Accelerometry-assessed sleep clusters and cardiometabolic risk factors in adolescents," *Obesity*, vol. 32, no. 1, pp. 200–213, Oct. 2023, doi: 10.1002/oby.23918.
[12] G. Albalak *et al.*, "Setting your clock: associations between timing of objective physical activity and cardiovascular disease risk in the general population," *European Journal of Preventive Cardiology*, vol. 30, no. 3, pp. 232–240, Nov. 2022, doi: 10.1093/eurjpc/zwac239.
[13] C. B. Thornton, N. Kolehmainen, and K. Nazarpour, "Using unsupervised machine learning to quantify physical activity from accelerometry in a diverse and rapidly changing population," *PLOS Digital Health*, vol. 2, no. 4, p. e0000220, Apr. 2023, doi: 10.1371/journal.pdig.0000220.
[14] H. Jeong, Y. W. Jeong, Y. Park, K. Kim, J. Park, and D. R. Kang, "Applications of deep learning methods in digital biomarker research using noninvasive sensing data," *DIGITAL HEALTH*, vol. 8, p. 205520762211366, Jan. 2022, doi: 10.1177/20552076221136642.
[15] E. E. Dooley *et al.*, "Method for activity sleep harmonization (MASH): A novel method for harmonizing data from two wearable devices to estimate 24-h sleep–wake cycles," *Journal of Activity, Sedentary and Sleep Behaviors*, vol. 2, no. 1, Apr. 2023, doi: 10.1186/s44167-023-00017-5.
[16] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.
[17] J. Chong, P. Tjurin, M. Niemelä, T. Jämsä, and V. Farrahi, "Machine-learning models for activity class prediction: A comparative study of feature selection and classification algorithms," *Gait & Posture*, vol. 89, pp. 45–53, Sep. 2021, doi: 10.1016/j.gaitpost.2021.06.017.
[18] A. B. F. Khan, K. Kamalakannan, and N. S. S. Ahmed, "Integrating machine learning and stochastic pattern analysis for the forecasting of time-series data," *SN Computer Science*, vol. 4, no. 5, Jun. 2023, doi: 10.1007/s42979-023-01981-0.
[19] A. Althaf Ali, S. Umamaheswari, A. B. Feroz Khan, and R. Jayabrabu, "Fuzzy rules-based data analytics and machine learning for prognosis and early diagnosis of coronary heart disease," *Journal of information and organizational sciences*, vol. 48, no. 1, pp. 167–181, Jun. 2024, doi: 10.31341/jios.48.1.9.
[20] E. Kańtoch, "Recognition of sedentary behavior by machine learning analysis of wearable sensors during activities of daily living for telemedical assessment of cardiovascular risk," *Sensors*, vol. 18, no. 10, p. 3219, Sep. 2018, doi: 10.3390/s18103219.
[21] Y. Zhang, Y. Tao, H. Choi, and H. Qian, "Exploring the causal effects of physical activity, sedentary behaviour, and diet on atrial fibrillation and heart failure: A multivariable mendelian randomisation analysis," *Nutrients*, vol. 16, no. 23, p. 4055, Nov. 2024, doi: 10.3390/nu16234055.
[22] J. Choi *et al.*, "Effect of physical activity on incident atrial fibrillation in individuals with varying duration of diabetes: a nationwide population study," *Cardiovascular Diabetology*, vol. 23, no. 1, Mar. 2024, doi: 10.1186/s12933-024-02194-2.
[23] M. P. Bonnesen *et al.*, "Day-to-day measurement of physical activity and risk of atrial fibrillation," *European Heart Journal*, vol. 42, no. 38, pp. 3979–3988, Sep. 2021, doi: 10.1093/eurheartj/ehab597.
[24] S. Ahrari, H. Karimi Moonaghi, S. M. Mahdizadeh, and A. Heidari Bakavoli, "Experiences of what influences physical activity adherence in Iranian patients with heart failure: A qualitative study," *Journal of Education and Health Promotion*, vol. 12, no. 1, Aug. 2023, doi: 10.4103/jehp.jehp_1029_22.
[25] M. Borland *et al.*, "Effects of 3 months of detraining following cardiac rehabilitation in patients with atrial fibrillation," *European Review of Aging and Physical Activity*, vol. 19, no. 1, Apr. 2022, doi: 10.1186/s11556-022-00293-1.

BIOGRAPHIES OF AUTHORS

Sudarsanam Anandanarayanan is currently working as a Head in the Software Computer Science, Shanmuga Industries Arts and Science College at Tiruvannamalai. His total teaching experience is 17 years. He is doing part time Ph.D. in Department of Computer and information Science, Annamalai University, Tamil Nadu. His primary research interest is in the areas of computer network, data mining and software engineering. His current research focus is data mining application especially in healthcare. He can be contacted at Profsanand@gmail.com.

Suvarnalingam Thirumaran is currently serving as an assistant professor of computer applications at Alagappa Government Arts College, Karaikudi, Tamil Nadu, deputed from the Department of Computer and Information Science (DCIS), Annamalai University, Tamil Nadu. He is an accomplished academician and researcher with a robust background in computer science, particularly in the areas of data mining, big data, machine learning, deep learning, and artificial intelligence. He has published several technical papers in reputed national and international conferences and journals. He can be contacted at thirurmalai817@gmail.com.