Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD30297

A Review of AI Models for Wearable Devices Driven by Micro Machine Learning

Haowei Ding

College of Computer Science, Guangzhou Huaili College, Guangzhou, China
dinghaowei48@gmail.com

Abstract. With the rapid advancement of the Internet of Things (IoT) and artificial intelligence (AI) technologies, wearable devices have evolved to core tools for health management and sports monitoring. AI models enabled by Tiny Machine Learning (TinyML) provide critical support for the intelligent upgrading of wearable devices, owing to their inherent characteristics of low power consumption, high real-time responsiveness, and robust privacy protection. This paper focuses on the application of TinyML in wearable devices and systematically elaborates on its practical accomplishments in the fields of health monitoring (e.g., abnormal heart rate early warning, sleep quality analysis) and motion recognition (e.g., action classification, posture assessment): Specifically, through lightweight design, TinyML models can adapt to the battery-powered constraints of wearable devices, thereby achieving extended battery life; leveraging edge computing capabilities, they enable real-time data processing and feedback; and through local data processing, they reduce data transmission links and mitigate the risk of privacy leakage. Meanwhile, this paper prospects its future development directions, including model structure optimization, multi-modal data fusion, and integration with emerging sensor technologies, so as to provide references for relevant research.

Keywords: TinyML, Wearable Devices, AI Models, Health Monitoring, Motion Recognition

1. Introduction

With the in-depth integration of Internet of Things (IoT) and artificial intelligence (AI) technologies (according to Gartner's prediction, the global shipment volume of wearable devices will exceed 1.1 billion units by 2025 [cited from IoT Industry Report]), and the growing demand for personalized health management and precise sports tracking, the application of Tiny Machine Learning (TinyML) in wearable devices has emerged as a research hotspot. Traditional wearable devices are constrained by hardware resources (e.g., computing power and battery life), which makes it challenging to implement complex intelligent analysis. Nevertheless, TinyML technology has overcome this constraint via lightweight model design and edge computing architecture.

By integrating modules including accelerometers and heart rate sensors, wearable devices can collect physiological and motion data in real time, providing data support for scenarios such as health monitoring and sports guidance [1]. This paper employs a combination of literature review and case analysis methods to sort out the current application status of TinyML in wearable devices,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD30297

and focuses on discussing model optimization strategies under resource constraints (such as low-power design and real-time performance enhancement) as well as data privacy protection mechanisms. The research aims to provide ideas for promoting the development of wearable devices towards intelligence, efficiency and security, and expand their application boundaries in fields such as medical and health care, and sports competitions.

2. Applications of micro machine learning AI models in wearable devices

2.1. Health monitoring application

In the field of health monitoring, wearable devices have become important tools for health management due to their convenience and real-time responsiveness. The integration of Tiny Machine Learning (TinyML)-enabled AI models has significantly expanded their application boundaries. These models can efficiently process multi-source physiological data collected by wearable devices, including heart rate, blood oxygen saturation, and body temperature. By collecting users' gesture data, voice data, voice data of surrounding individuals, arm swing amplitude data, blood pressure data, and heart rate data, and through in-depth analysis of long-term data [1], AI models can accurately identify individuals' health status and abnormal changes. Taking heart rate monitoring as an example, the model can not only feed back the heart rate value in real time, but also assess the user's stress level, sleep quality, and potential risk of heart disease based on Heart Rate Variability (HRV). In sleep monitoring, the model can analyze sleep cycles, evaluate sleep quality, and provide users with targeted improvement recommendations. Furthermore, for patients with chronic diseases, wearable devices integrated with AI models can continuously track disease-related indicators, assist doctors in conducting remote monitoring, achieve more efficient disease management and prevention, and safeguard individuals' health.

2.2. Sports recognition application

In the field of motion recognition, micro-machine learning AI model plays a crucial role in wearable devices. They can collect multidimensional data generated by human movement by leveraging various sensors on wearable devices, such as accelerometers and gyroscopes. Then, the convolution neural network, long-term memory network and other models are used to analyze and process these data, which can accurately identify different sports types, such as walking, running, riding, climbing stairs and so on. Moreover, they can also monitor and evaluate movement postures and trajectories in real time, provide users with sports data statistics, sports effect analysis and personalized sports suggestions, help users exercise scientifically, enhance sports experience, and demonstrate broad application prospects and value in sports fitness, rehabilitation training, sports competition and other scenes.

This paper introduces a method and process of motion recognition through wearable devices using machine learning method: step one, construct a data acquisition platform for human motion states, and collect acceleration and angular velocity data information in human motion via sensors worn in wearable devices at different body parts; Step 2, perform filtering and denoising the collected human motion data and data fusion processing through the DMP data management platform; Step 3, employ a threshold algorithm to identify and classify different human motion states based on the calculated motion data, and obtain discrimination and classification results; Step 4, the recognition result of the current human motion state is displayed in real time through the

DOI: 10.54254/2755-2721/2025.LD30297

front-end design of the upper computer, and the human motion data information is displayed in the form of a dynamic data line [2].

3. The low-power characteristics of micro-machine learning AI models in wearable devices

3.1. The low-power characteristic of micro-machine learning

This emerging technology can simplify the implementation, deployment, and utilization of machine learning methods in smartphones, gadgets, and certain edge computing-based applications. Meanwhile, it consumes significantly less energy compared with traditional processors, microcontrollers, and computers. Thus, it serves as an ideal solution for applications constrained by energy consumption, such as Wireless Sensor Networks (WSNs). In general, this technology consists of chips as hardware and algorithms as software, which are usually embedded in a compact package.

3.2. Model design and power consumption optimization

In the application of wearable devices, the low-power characteristics of Tiny Machine Learning (TinyML) AI models is of crucial importance, which largely depend on scientific model design and power consumption optimization strategies. From the perspective of model design, streamlining the model structure is the key. By adopting lightweight neural network architectures such as MobileNet and ShuffleNet, and reducing the number of parameters and computational load via depthwise separable convolution, computational complexity is reduced, thereby lowering power consumption during operation. In the model training phase, the selection of optimization algorithms should also not be overlooked either. Stochastic Gradient Descent (SGD) and its improved variants can achieve model convergence with relatively low computational overhead, thereby avoiding unnecessary energy consumption [3].

In terms of power consumption optimization, Dynamic Voltage and Frequency Scaling (DVFS) technology can dynamically adjust the voltage and frequency of wearable device chips based on the load during model operation, thereby reducing power consumption while meeting computational requirements. Quantization training (e.g., INT8 quantization) is employed to convert floating-point operations into integer operations, which reduces energy consumption for storage and computation. Knowledge distillation technology is utilized to transfer the "knowledge" of complex models to lightweight models, reducing the computing power requirement under the premise that the accuracy loss is less than 5% [3]. In addition, data quantization technology reduces the data precision in the model; within an acceptable range of accuracy loss, it decreases the energy consumption for data storage and computation. The combination of these measures enables the low-power operation of models in wearable devices.

3.3. Model adaptation and power consumption management

In the application scenarios of wearable devices, the low-power characteristics of Tiny Machine Learning (TinyML) AI models rely on the collaborative effect of model adaptation and power consumption management. In terms of model adaptation, it is necessary to fully consider the constraints of hardware resources in wearable devices and perform targeted optimization on the models. Specifically, knowledge distillation technology transfers the knowledge of complex models to lightweight models, enabling these lightweight models to reduce computational complexity and

DOI: 10.54254/2755-2721/2025.LD30297

adapt to the low-computing-power hardware of wearable devices, and keep performance loss under control.

Power consumption management is another key aspect to achieve low power consumption. On the one hand, a dynamic power management strategy can be employed: based on the operating status of the model (such as idle, light load, heavy load, etc.), the working mode of the device and the allocation of hardware resources are dynamically adjusted, and modules that are not temporarily in use are turned off to reduce energy consumption. On the other hand, through the collaborative optimization of hardware and software, task scheduling algorithms are optimized at the operating system level, enabling accurate matching between model operation and hardware resource supply. This reduces unnecessary power consumption overhead, and ultimately ensures the stable and low-power operation of TinyML AI models in wearable devices, as well as extends the battery life of the devices.

4. Real-time analysis of micro machine learning AI models in wearable devices

4.1. Data processing and model inference speed

In the real-time analysis of Tiny Machine Learning (TinyML) AI models applied in wearable devices, the speed of data processing and model inference is crucial. Wearable devices are required to continuously collect multi-source physiological data (e.g., heart rate and acceleration), which is characterized by high frequency and dynamic variability. Fast and accurate data processing is a prerequisite for the efficacy of subsequent analysis; it is necessary to perform preprocessing operations such as denoising and normalization on the raw data to conform to model input requirements while reducing the resource consumption caused by redundant data.

The speed of model inference directly dictates whether the device can respond in real time. TinyML models need to rapidly infer predictions from processed data under limited hardware resources. For instance, when determining the user's movement state, the model must quickly generate results to enable timely feedback to the user. Optimizing inference algorithms, adopting lightweight model architectures, and applying hardware acceleration technologies constitute key strategies to improve model inference speed. These measures ensure the efficient collaboration between data processing and model inference in real-time scenarios of wearable devices, thereby meeting users' real-time requirements.

4.2. Balancing real-time performance and model accuracy

Real-Time Processing and Analysis:Acquire real-time data streams from diverse data sources, map these real-time data streams to a preset data structure framework to obtain initial data streams, and process the initial data streams via a real-time data fusion engine to generate target real-time data streams. Then, use data mining and machine learning algorithms to perform pattern recognition and trend analysis on the target real-time data streams, produce target insight reports, and optimize the target insight reports using an intelligent decision-making unit. Finally, generate and execute strategy information to obtain the optimal analysis results [4]. Research on using approximate query processing to address the response time challenge in real-time interactive analysis has gradually emerged as a critical research trend. By providing approximate but quality-guaranteed query results, such systems significantly reduce query latency and improve system scalability through technologies like sampling and summarization [5].

DOI: 10.54254/2755-2721/2025.LD30297

In the real-time analysis of Tiny Machine Learning (TinyML) AI models applied to wearable devices, achieving a balance between real-time performance and model accuracy is crucial. On one hand, wearable devices require AI models to process data quickly to meet the demand for real-time feedback—for example, in real-time monitoring of data such as heart rate and step count. If the model takes too long to process data, feedback will be delayed, which severely impairs user experience. On the other hand, model accuracy is the cornerstone of accurate judgment and analysis; low-accuracy models may provide erroneous health condition assessments or motion analyses, thereby misleading users.

To strike this balance, multi-dimensional considerations are necessary. From the perspective of algorithm optimization, lightweight algorithms are adopted to reduce computational load while ensuring a certain level of accuracy. In terms of hardware adaptation, model architectures are customized based on the hardware characteristics of wearable devices, such as constrained computing resources and low-power requirements. Additionally, data preprocessing can be used to streamline data volume and improve processing speed without losing key information, thereby finding the optimal solution between real-time performance and model accuracy.

4.3. Real-time analysis technology for human body detection in wearable devices

Human activity monitoring based on wearable devices is a process that utilizes wearable devices and computer technology to obtain users' movement status by sensing human activity data. In the research on human activity monitoring methods, the main technologies adopted include data acquisition and activity recognition. At the data acquisition level, most traditional activity monitoring methods use video image-based approaches; however, these methods are subject to numerous environmental constraints and raise significant privacy concerns, failing to fully meet the requirements of activity monitoring.

First, a data preprocessing method is proposed for the data collected by a single three-axis accelerometer. For noise data that does not overlap with the signal, a Butterworth filter is first used to perform low-pass filtering on the data to remove such noise. Subsequently, the KGA algorithm—an approach that integrates the genetic algorithm into the Kalman filter algorithm—is applied. This algorithm uses the genetic algorithm to encode the parameters in the Kalman filter algorithm, thereby optimizing the Kalman filter algorithm.

Second, a human activity recognition method based on an improved one-dimensional convolutional neural network (1D-CNN) is presented. According to the characteristics of human movement, feature values that can distinguish between different activity types are extracted. Meanwhile, to address the overfitting problem of traditional convolutional neural networks, a sample self-learning method is employed to identify the optimal sample training set, thus avoiding the occurrence of overfitting In the recognition of 11 types of human activities, this method achieves a high average accuracy rate. When compared with other activity recognition methods on the same dataset, it delivers better classification performance.

Third, regarding fall activities— which pose significant risks to the elderly—a threshold-based human fall state recognition method is proposed, focusing on distinguishing fall activities from daily activities. Based on the characteristics of human fall movements, feature values that can effectively differentiate between daily activities and falls are extracted. The cross-validation method is utilized for verification: in the analysis of 11 types of daily human activities and 15 types of fall activities, this method successfully discriminates the 15 types of fall activities and achieves a high recognition recall rate.

DOI: 10.54254/2755-2721/2025.LD30297

Finally, real-time verification is conducted on the human activity monitoring system based on wearable devices. The human activity data collected by the sensor is transmitted to a cloud platform; the backend server performs real-time analysis on the data to identify the current human activity and triggers an alarm upon detecting a fall This system achieves a high average recall rate [6].

A New Type of Remote Health Monitoring System Based on Wearable Devices:The system integrates physical sign parameter monitoring and an elderly fall alarm mechanism. It is designed with a wristband terminal as the master device and a waistband terminal as the slave device. The master device of the system mainly collects human physical sign parameters, locates the current position, and sends alarm short messages; the slave device mainly performs human fall detection. Low-power Bluetooth 4.0 is employed for data communication between the master and slave devices.

Low-Power Bluetooth 4.0 Communication Technology:On July 7, 2010, the Bluetooth Special Interest Group (Bluetooth SIG) announced the official adoption of the Bluetooth Core Specification Version 4.0 and the launch of the corresponding certification program. Bluetooth BLE 4.0 technology is a short-range, low-cost, and highly robust wireless communication technology that integrates multiple technologies, including legacy Bluetooth, low-energy technology, and high-speed data transmission Bluetooth 4.0 operates in the 2.4GHz ISM band—a globally free public frequency band—supports ultra-short data packets at a data transmission rate of 1Mbps, and has a maximum transmission range of over 100 meters, enabling real-time wireless data and voice transmission. A key feature of this technology is its ultra-low operating and standby power consumption, allowing it to operate continuously for several years using only a single button battery. Thanks to this feature, it has become an ideal choice for wireless transmission solutions with strict requirements on power consumption and cost [7].

5. Data privacy protection for micro-machine learning AI models in wearable devices

5.1. Privacy in machine learning: definitions and categories

1) Training Data Privacy

Training data privacy pertains to the personally identifiable information (PII) and sensitive information embedded in user data utilized for machine learning.

2) Model Privacy

Model privacy concerns the privacy-related information associated with machine learning models, including model training algorithms, model topological structures, model weight parameters, activation functions, and hyperparameters in machine learning.

3) Prediction Result Privacy

Prediction result privacy refers to the sensitive information that users are unwilling to have disclosed, which is generated by the model in response to users' prediction input requests in machine learning [8].

5.2. Local data processing and privacy advantages

Data Privacy Protection for TinyML AI Models in Wearable Devices:Amid the widespread adoption of wearable devices, data privacy protection for Tiny Machine Learning (TinyML) AI models has garnered significant attention, among which local data processing exhibits notable advantages. In the traditional model, wearable devices often upload data to the cloud for processing—during

DOI: 10.54254/2755-2721/2025.LD30297

transmission, data is vulnerable to leakage risks, and user data stored in the cloud may also face security threats.

TinyML AI models that adopt local data processing can effectively mitigate these issues. Devices perform data analysis and processing exclusively locally, without transmitting raw data to external servers, thereby fundamentally reducing the risk of data leakage. For instance, when monitoring users' sleep data, the model directly analyzes information such as sleep stages and heart rate changes on the wearable device itself, rather than sending this sensitive data to the cloud. This approach not only safeguards user privacy but also fosters greater user trust in wearable devices, promoting the safer and more widespread application of wearable devices in fields such as health monitoring and sports tracking.

5.3. Privacy protection technology and algorithm application

Data Privacy Protection and Deployment Algorithms for TinyML AI Models in Wearable Devices

5.3.1. Data privacy protection technologies

In the field of wearable devices, data privacy protection for Tiny Machine Learning (TinyML) AI models is of vital importance, and a variety of privacy-preserving technologies and algorithms are widely applied.

1) Differential Privacy Algorithm

The differential privacy algorithm is a typical representative in this context. As a probability theory-based privacy protection method, it perturbs local data via mechanisms such as the Laplace noise mechanism and Gaussian noise mechanism. This perturbation ensures that changes in individual data do not affect the overall statistical characteristics, thereby defending against differential attacks [9].

2) Homomorphic Encryption Technology

Homomorphic encryption technology also plays a key role. This technology enables computations to be performed directly on ciphertext without decrypting the data. Consequently, once the data collected by wearable devices is encrypted, the model can perform computations directly on this ciphertext, ultimately yielding an encrypted result.. Only the device user possesses the decryption key, which ensures end-to-end privacy and security of the data.

3) Federated Learning

Federated learning facilitates model training through multi-party collaboration. Each participant conducts local training and exclusively uploads model parameters (rather than raw data), thereby realizing the "data usable but invisible" paradigm. It demonstrates strong application potential in data privacy protection for wearable devices and effectively balances the relationship between data utilization and privacy protection.

5.3.2. Deployment algorithm based on neural architecture search

1) Overview of the Algorithm

A deployment algorithm based on Neural Architecture Search (NAS) is proposed to study how to deploy neural network algorithms such as Convolutional Neural Networks (CNNs) on microcontrollers. NAS enables automated model design: by setting constraints and stacking search units, it identifies a model that meets the requirements. This approach does not require extensive

DOI: 10.54254/2755-2721/2025.LD30297

experience in neural network design, but it takes a significant amount of time for the system to test different network structures [10].

2) NAS as the Foundation for Hardware-Aware Automated Quantization (HAQ)

Neural Architecture Search lays the foundation for Hardware-Aware Automated Quantization (HAQ): it can identify neural network architectures that are better suited for specific tasks and hardware platforms. When these architectures run on hardware, their optimized structure provides a more robust foundation for hardware-aware automated quantization—making the quantization process more efficient and accurate, and reducing accuracy loss caused by quantization. The hardware-aware automated quantization framework uses reinforcement learning to explore the strategy space of accuracy and quantization for different neural network structures and hardware structures [10].

6. The future development direction of micro-machine learning AI models in wearable devices

6.1. Technological innovation and model evolution

Future Development of TinyML AI Models in Wearable Device Scenarios:In the application scenarios of wearable devices, the future development of Tiny Machine Learning (TinyML) AI models will focus deeply on technological innovation and model evolution.

1) Technological Innovation

With the continuous advancements in hardware technology, chips featuring lower energy consumption and enhanced computing capabilities are expected to be integrated into wearable devices, thereby providing a more robust hardware foundation for model operation. Meanwhile, the emergence of new sensor technologies can acquire richer and more accurate physiological and environmental data, expanding the data input dimensions of the model and enhancing the accuracy and comprehensiveness of analysis.

2) Model Evolution

From the perspective of model evolution, models will develop in a more lightweight and adaptive direction. On one hand, researchers will continuously optimize the model structure and reduce the number of parameters—while ensuring performance, they will lower the demand for computing resources to adapt to the limited hardware conditions of wearable devices. On the other hand, models will exhibit enhanced adaptive capabilities: they can dynamically adjust learning strategies based on the individual characteristics of different users and usage scenarios, realizing more personalized and accurate services. This will further tap into the application potential of wearable devices in fields such as health monitoring and sports assistance.

6.2. Application expansion and market prospects

Application Expansion and Market Prospects of TinyML AI Models in Wearable Devices:With the continuous advancement of technology, the application expansion and market prospects of Tiny Machine Learning (TinyML) AI models in wearable devices are extremely broad.

1) Application Expansion

In terms of specific applications, machine learning is predominantly used for behavior analysis, recognition and prediction, as well as the automatic scoring of text-like content. However, in-depth exploration remains insufficient in domains such as adaptive process support, resource recommendation, path navigation, and personalized feedback.

DOI: 10.54254/2755-2721/2025.LD30297

In the application of specific algorithms, the utilization of machine learning algorithms is relatively homogeneous: a large number of studies have adopted supervised machine learning algorithms, while a small number have utilized unsupervised machine learning algorithms, and most of these studies have yielded favorable results. Nevertheless, semi-supervised machine learning algorithms have not been well applied [11].

In the Healthcare Field: This model can be used to monitor users' physiological data (such as heart rate, blood pressure, and sleep) in real time, providing long-term health management for patients with chronic diseases. It can also enable early warning of diseases, promoting the transformation of the medical model from traditional disease treatment to disease prevention.

In the Sports and Fitness Field: Based on users' sports data, it can provide personalized training plans and real-time guidance, improving exercise effectiveness and safety.

2) Market Prospects

From the perspective of market prospects, consumers' demand for health management and personalized services is growing increasingly. Driven by this, the market size of wearable devices continues to expand. Devices equipped with TinyML AI models, leveraging their intelligent advantages, are more likely to gain consumer preference—attracting more capital and enterprises to deploy in this field and promoting industrial innovation and development. In the future, such devices are expected to continuously penetrate various segmented fields, creating enormous economic and social benefits.

7. Conclusion

This paper reviews the research progress of wearable device AI models driven by Tiny Machine Learning (TinyML) and demonstrates their great potential in fields such as health and sports. However, existing research still exhibits certain limitations: for instance, the optimization of TinyML algorithms for wearable devices remains inadequate, and the mechanisms for user data security and privacy protection require further refinement.

To address these deficiencies, future research should prioritize algorithm optimization to explore more efficient and energy-saving TinyML models. Simultaneously, it is essential to enhance data encryption and privacy protection technologies to safeguard the security and privacy of user data.

In addition, with the continuous development of Internet of Things (IoT) technology and the ongoing innovation in hardware technology—such as the emergence of more efficient chips and sensors—it is expected to provide stronger support for the operation of AI models. Meanwhile, interdisciplinary integration will be further advanced: by integrating knowledge from domains such as medicine and materials science, more user-centric AI models for wearable devices with higher reliability and security will be developed, thereby delivering greater convenience and value to human life.

References

[1] Xikang (Wuxi) Internet of Things Technology Co., Ltd. A mental health detection method and device based on wearable devices: cn117918839a [p/ol] .2024-04-26 [2025-02-06]. https: //www.cqvip.com/doc/patent.

[2] Changchun University of Technology. A wearable device-based human motion state recognition method: CN118924287A [P/OL]. 2024-11-12 [2025-02-06]. https: //www.cqvip.com/doc/patent/3474068266

[3] ARASH KHAJOOEI, MOHAMMAD (BEHDAD) JAMSHIDI, SHAHRIAR B. SHOKOUHI. A Super-Efficient TinyML Processor for the Edge Metaverse [J]. Information, 2023, 14(4). DOI: 10.3390/info14040235.

[4] Beijing Shuyang Wisdom Technology Co., Ltd. A method and system for real-time processing and analysis of AI big data: CN119168075A [P/OL] .2024-12-20 [2025-02-06]. https: //www.cqvip.com/doc/patent/34755.

DOI: 10.54254/2755-2721/2025.LD30297

[5] Yuan Zhe, Wen Jirong, Wei Zhewei, Liu Jiajun, Yao Bin, Arqie. Real-time interactive analysis of big data. Journal of Software, 2020, 31 (1): 162-182. http: //www.jos.org.cn/1000-9825/5886.htm
[6] Gao Xiangyu. Research on human behavior monitoring method based on wearable devices [D]. China University of Mining and Technology, 2019.
[7] Wu Peizhi. Design of remote health monitoring system based on wearable devices [D]. Shaanxi: Xi 'an Engineering University, 2018.
[8] Tan Zuowen, Zhang Lianfu. Summary of research on machine learning privacy protection [J]. Journal of Software, 2020, 31 (07): 2127-2156. DOI: 10.13328/j.cnki.jos.006052.
[9] Gou cong. research on data privacy protection of wearable devices based on deep learning [D]. Chongqing: Chongqing university, 2021. doi: 10.27670/d.cnki.gcqdu.20010.000000000105
[10] Wu Jianbang, Qiu Tian, Zhang Xin, et al. Research Status and Prospect of TinyML [J]. Microcontroller and Embedded System Application, 2023, 23(02): 7-11.
[11] Sun Shebing, Zhang Qiaorong, Li Tianqi. Summary of machine learning applied to online scientific inquiry [J]. Modern Information Technology, 2023, 7 (14): 103-110. DOI: 10.19850/J.CNKI.2096-4706.