ISSN 2959-6157

Healthcare with Wearable Devices and Machine Learning

Jessica YiJia Yan[1, *]

[1] Harrow International School Hong Kong, Hongkong, China
*Corresponding author: jessicayijiayan@gmail.com

Abstract:

As health management continues to grow in importance globally, driven by the quest for improved health outcomes, longevity, and personalized care, machine learning (ML) has become a popular technique for detecting and analyzing health indicators, playing an essential role in multiple areas of health detection, such as continuous heart and blood pressure monitoring, blood sugar tracking for diabetics, sleep analysis to improve sleep quality, and even disease prediction. With wearable devices, a particular scale of health detection data has been accumulated at this stage. With this data, machine learning can be predictive analytics to assess real-time health conditions, enabling early intervention and personalized treatment plans. However, data quality remains the most critical issue for machine learning-based health detection systems, as inaccuracies in sensor readings can lead to misdiagnoses or incorrect treatment decisions. In addition, privacy concerns are significant, as the sensitive health information collected by these devices requires strong security measures to protect users' data. Moreover, the interpretability of machine learning algorithms, especially those involving deep learning, can be limited, hindering clinicians' understanding and trust. This paper addresses these challenges by providing a comprehensive overview of relevant data sets used in health monitoring and examining the latest machine learning algorithms tailored for wearables-based health detection. It further delves into the current limitations and challenges facing this emerging field, providing insights into potential solutions and future directions for research and development.

Keywords: Healthcare; Wearable devices; Machine learning; Deep learning.

1. Introduction

With the increasing emphasis on health management worldwide, machine learning has been widely used in detecting health indicators. By 2024, the global machine learning market value is expected to reach 96.7 billion US dollars, showing a solid growth momentum. Especially in health care, the application of machine learning technology provides a more accurate and efficient solution for detecting health indicators.

On the one hand, wearable devices have shown remarkable advantages in the detection of health indicators: convenience: wearable devices are usually designed to be light and portable, and users can wear them at any time without carrying additional professional equipment. This portability enables health monitoring to be easily integrated into daily life and improves the continuity and real-time performance of monitoring. Real-time monitoring: Wearable devices can monitor various physiological indexes of users in real-time, such as heart rate, blood pressure, blood sugar, sleep quality and so on. This real-time monitoring function helps users know their health status in time and take measures when abnormal situations occur. On the other hand, in detecting health indicators, machine learning algorithms can process a large amount of health data, including heart rate, blood pressure, blood sugar, sleep quality, and so on, to realize real-time monitoring and evaluation of personal health status [1-3]. Its main application scenarios include 1) Heart rate and blood pressure monitoring: Modern wearable devices generally adopt photoelectric volume pulse wave recording (PPG) sensors, which can realize accurate monitoring of heart rate and blood pressure by combining machine learning algorithms. These devices can not only provide users with real-time data in daily life but also give an alarm in abnormal situations to prevent potential health risks. 2) Blood sugar monitoring: For diabetic patients, it is essential to monitor their blood sugar levels continuously. Machine learning algorithms combined with continuous glucose monitoring systems (CGM) can sense the blood glucose concentration in real-time and predict the future blood glucose trend according to the data changes, thus allowing doctors to adjust treatment plans. 3) Sleep quality analysis: Wearable devices analyze sleep quality by monitoring the user's sleep cycle, turn-over times, wake-up times and other indicators, combined with machine learning algorithms. This helps users to understand their sleep habits and improve their sleep quality, thus preventing health problems such as sleep disorders. 4) Disease prediction and risk assessment: A machine learning algorithm can find potential disease risk factors by analyzing health data. For example, by analyzing users' exercise habits, eating habits and other data, we can predict their risk of chronic diseases such as cardiovascular diseases and diabetes and intervene and treat them in advance.

Although machine learning has made remarkable progress in the research of health indicators detection, it still faces some challenges. First, data quality and privacy protection are urgent problems that must be solved. Health data involves personal privacy and sensitive information. Data security and privacy protection are essential challenges for machine learning in health indicator detection. Secondly, the interpretability of the algorithm is also a technical problem that needs to be solved urgently. Although machine learning algorithms can process complex data and get accurate prediction results, their decision-making process is often challenging to explain, which limits their application in the medical field. Besides, although wearable devices have made some progress in health monitoring, there still needs to be a significant error in the accuracy of their monitoring data. The data measured under different equipment, users and use conditions may be quite different, which affects the reliability of the data to some extent. Given the above limitations, this paper first introduces the relevant data sets to provide a valuable reference for relevant researchers. Then, it introduces the typical machine learning-related algorithms in this direction. Finally, this paper discusses the challenges and future research direction of health detection systems based on wearable devices.

2. The Overview of Dataset

Table 1. The overview of the dataset (Bellabeat)

	# of features/files	Duration (day)	Sample size (users)	Gender
Bellabeat	11	60	33	Co-ed
Bellabeat: dailyActivity	15	~	~	~
Bellabeat: heartrate	3	~	~	~
Bellabeat:HourlyCalories	3	~	~	~
Bellabeat: HourlyIntensities	4	~	~	~
Bellabeat: HourlySteps	3	~	~	~
Bellabeat: minuteCalories	3	~	~	~
Bellabeat: minuteIntensities	3	~	~	~
Bellabeat: minuteMETs	3	~	~	~
Bellabeat: minuteSleep	4	~	~	~
Bellabeat: minuteSteps	3	~	~	~
Bellabeat: WeightLog	8	~	~	~

As shown in Table 1 and Table 2, all five datasets meticulously track various specific activities, each contributing a unique perspective to the broader realm of data analysis in fitness, health, and lifestyle monitoring. 'Bellabeat' and 'Fitabase' stand out as comprehensive resources, embracing a broader scope encompassing not just fitness routines but also various aspects of overall health and wellbeing. They offer insights into users' daily physical activities, dietary habits, and mental health, catering to a diverse audience seeking to improve their holistic wellbeing.

On the other hand, 'Fitness: Healthy Man' adopts a more targeted approach, laser-focusing solely on fitness aspects such as exercise routines, strength training, and cardio workouts. This dataset is ideal for those specifically interested in optimizing their physical fitness or conducting research targeted at male populations. Meanwhile, 'Sleep & Lifestyle' narrows its focus even further, emphasizing solely health practices related to sleep quality and lifestyle habits that contribute to good rest. It could be further classified as a dataset primarily centred around sleep analysis, given its specificity in addressing this vital aspect of health.

Table 2. The overview of the dataset (Others)

	# of features/ files	Duration (day)	Sample size (users)	Gender	Generator	Year of creation
Fitness: Healthy man	21	130	1	Male	Selçuk Yılmaz	2022
Sleep & Lifestyle	13	1	374	Co-ed	LAKSIKA THARMALINGAM	2023
PA during COVID	11	730	113	Co-ed	Garmin/Fitbit	2019/2020
Fitabase (Example set)	28	7	N.A	Co-ed	Fitabase	2017

Table 1 and Table 2 provide a comprehensive overview of the distinguishing features of each dataset, detailing the number of features or files contained within, the duration of data collection, sample size, gender representation, data type, keywords for easy retrieval, the entity that generated the data, the year of creation, file size, and usability considerations. Upon closer inspection, it becomes evident that each dataset possesses unique characteristics, particularly notable in duration, file size, and sample size, which exhibit a considerable range across the board. These dimensions underscore the varying degrees of depth and breadth captured by each dataset.

It is also worth mentioning that some columns in the table contain missing or unavailable data, primarily in usability, file size, and sample size. Notably, the ‘usability’ column seems to be the most affected, highlighting potential gaps in information related to the practical application or accessibility of the datasets. However, despite these inconsistencies, all five datasets maintain a core set of essential characteristics that make them valuable resources for research and analysis.

Amongst the five, ‘Bellabeat’ and ‘Sleep & Lifestyle’ emerge as the most comprehensively documented, boasting no columns in the table that are unavailable. This completeness of information positions them as beautiful options for researchers and analysts seeking a seamless and comprehensive experience when working with fitness and health-related data. Ultimately, the choice of dataset will depend on the specific research objectives and the level of detail required. However, these two stand out as strong contenders for those seeking the most comprehensive datasets within this domain.

3. Methods

Machine learning algorithms in wearable devices are widely used. These devices collect user data through built-in sensors, such as heart rate, sleep quality, exercise data, etc., and use machine learning algorithms to analyze and mine these data to provide more intelligent and personalized services.

3.1 Traditional Machine Learning Method

In wearable devices, a typical application example of the traditional machine learning algorithm is heart rate monitoring and anomaly detection. Wearable devices (such as smart watches, health bracelets, etc.) collect the user‘s heart rate data in real time through the built-in heart rate sensor. These data are then used to monitor the user‘s heart rate status and provide timely alarms or suggestions when abnormalities are found.

To construct an abnormal heart rate detection model, it is necessary to clean the collected heart rate data first, remove noise and abnormal values, and ensure the accuracy and reliability of the data. Then, meaningful features, such as maximum heart rate, average heart rate and heart rate variability, are extracted from the preprocessed heart rate data. These features will be used to train machine learning models. Then, an appropriate machine learning algorithm (such as support vector machine SVM, Random Forest, etc.) is selected to build a heart rate anomaly detection model. The model is trained using historical heart rate data and known anomaly labels (such as doctor‘s diagnosis results) to learn the characteristics of heart rate anomaly. Finally, the model‘s performance, such as accuracy and recall, is evaluated by a verification set or cross-validation, and the model‘s parameters are adjusted and optimized according to the evaluation results to improve the accuracy and generalization ability of the model [4].

3.2 Deep Learning Method

In wearable devices, the application advantages of deep learning algorithms are mainly reflected in human motion recognition. Deep learning models (such as convolutional neural network CNN, circular neural network RNN, long-term and short-term memory network LSTM, etc.) can learn human motion patterns in sensor data, thus accurately identifying various actions of users, such as walking, running, jumping, etc. These models realize the accurate recognition of human movements by extracting the spatial and temporal features in time series data. For example, based on the motion data of smartwatches or smart bracelets, the deep learning model can reconstruct the trajectory of human arms and realize real-time 3D arm bone tracking and gesture inference. In the field of elderly care, deep learning algorithms can monitor the activities of the elderly in real time and prevent accidents such as falls [5].

4. Discussion

Applying health detection based on machine learning can promote the shortage of medical resources, disease prevention and many other problems. However, due to the privacy protection of data and the accuracy of wearable devices, three problems still need to be solved urgently. This section discusses the reasons for these problems and the corresponding solutions.

Health testing data contains a lot of personal privacy information. In data transmission, sharing and application, any node‘s negligence or improper operation may lead to data leakage. The risk is more prominent, especially in the multiple utilization and inter-agency transmission of data. To protect data privacy, we need to adopt advanced technical means, such as encryption technology and anonymity technology. However, these technologies may have performance bottlenecks or security risks in the application process. In recent years, federated learning has shown remarkable advantages and unique methods in protecting data privacy. The core idea of federated learning is distributing the model training process to all data owners. Each data owner only uses local data to train the model, shares model parameter updates with other data owners and finally gets a global model. In this way, the data is not concentrated in a central server for training, thus reducing the risk of data privacy disclosure. Federated learning often combines differential privacy technology to protect data privacy. Differential privacy blurs the precise value of model parameters by adding appropriate noise in the aggregation stage so that even if the attacker intercepts the noisy model parameters, he cannot recover the original data of any participant. This technology effectively prevents the disclosure of personal privacy.

Limited by wearable devices' computing power and storage capacity, powerful machine learning methods are challenging to embed in wearable devices, seriously affecting their performance. Model compression of mobile devices is an important research direction in machine learning. Model compression technology refers to reducing the size, complexity and calculation of machine learning models through various technical means so as to deploy and run on equipment with limited resources. This technology can effectively reduce the model's storage and computing resource requirements of the model, improve the reasoning speed and efficiency, and thus realize efficient machine learning applications in resource-constrained scenes such as mobile devices and edge devices.

The second section of this paper introduces the currently available health detection data sets. A few sample learnings are essential modules for health detection systems based on machine learning. The core goal of learning with few samples is to train a model that can accurately identify sample categories with small data. This method realizes efficient learning under limited data by using prior knowledge, simplifying the model structure and optimizing the algorithm search path. Unlike zero-shot learning, small sample learning still needs a small amount of labelled data to train the model, but compared with traditional machine learning, the amount of data required is significantly reduced.

5. Conclusion

Machine learning technology is revolutionizing the healthcare industry. Machine learning algorithms process massive amounts of health data and apply complex statistical models and predictive analytics to assess real-time health conditions, enabling early intervention and personalized treatment planning. Data quality is essential for machine learning-based health monitoring systems, as inaccuracies in sensor readings can lead to misdiagnosis or incorrect treatment decisions. Privacy concerns are also significant because the sensitive health information collected by these devices requires strong security measures to protect users' data. In addition, the interpretability of machine learning algorithms, especially those involving deep learning, can be limited, which hinders clinicians' ability to understand and trust entirely generated insights. This paper addresses these challenges by providing a comprehensive overview of relevant data sets used in health monitoring and examining the latest machine learning algorithms tailored for wearables-based health detection. It further delves into the current limitations and challenges facing this emerging field, providing insights into potential solutions and future directions for research and development.

References

[1] José Gabriel Carrasco Ramírez, Md.Mafiqul Islam, ASM Ibnul Hasan Even. Machine Learning Applications in Healthcare: Current Trends and Future Prospects. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 2024, 1(1).

[2] Ubaid Ullah, Begonya Garcia-Zapirain. Quantum machine learning revolution in healthcare: a systematic review of emerging perspectives and applications. IEEE Access, 2024.

[3] Han Lin, Jiatong Han, Pingping Wu, et al. Machine learning and human-machine trust in healthcare: A systematic survey.

CAAI Transactions on Intelligence Technology, 2024, 9(2): 286-302.
[4] Modhugu Venugopal Reddy, Ponnusamy Sivakumar. Comparative Analysis of Machine Learning Algorithms for Liver Disease Prediction: SVM, Logistic Regression, and Decision Tree. Asian Journal of Research in Computer Science, 2024, 17(6): 188-201.
[5] R Kishore Kanna, Bhawani Sankar Panigrahi, Susanta Kumar Sahoo, et al. CNN Based Face Emotion Recognition System for Healthcare Application. EAI Endorsed Transactions on Pervasive Health and Technology, 2024, 10.