BUSINESS INFORMATICS
INFORMATYKA EKONOMICZNA

Business Informatics
Informatyka Ekonomiczna
Year 2023, 1-2(65-66)
ISSN 2450-0003
journals.ue.wroc.pl/ie

Predicting Physical Activity Levels for the Personalization of Well-Being Programmes

Andrzej Małowiecki

Wroclaw University of Economics and Business

e-mail: andrzej.malowiecki@ue.wroc.pl

ORCID: 0009-0000-1264-1436

Julia Jaśkiewicz

Wroclaw University of Health and Sport Sciences

e-mail: julia.jaskiewicz2816@op.pl

ORCID: 0009-0001-7529-4272

Quote as: Małowiecki, A., & Jaśkiewicz, J. (2023). Predicting Physical Activity Levels for the Personalization of Well-Being Programmes. *Business Informatics. Informatyka Ekonomiczna*, 1-2(65-66), 33-41.

DOI: 10.15611/ie.2023.1-2.04

JEL: I12, C53, O33

Abstract

Aim: The aim of this paper is to verify whether a machine learning model can be effectively used to predict people's physical activity levels for personalising employee well-being programmes in companies.

Methodology: The following research methodologies were used in this paper: literature analysis and experiment, in the form of verification of the predictions made by the created machine learning model.

Results: The results obtained from the evaluation of the model showed that the use of human characteristics data to predict physical activity levels for the personalisation of well-being programmes does not guarantee good enough results.

Implications and recommendations: By effectively predicting physical activity levels, well-being programmes can be more effectively personalised to the individual needs of employees, which can contribute to improving their health.

Originality/value: The literature review found that the use of machine learning to predict physical activity levels has not been described in detail in the literature.

Keywords: GPAQ, MET, well-being, machine learning

1. Introduction

In recent years, there has been a growing recognition of the importance of personalized well-being programmes in promoting overall health and preventing lifestyle-related diseases. Physical activity is a key component of these programmes, contributing significantly to both physical and mental health. Despite the general recommendations provided by health organizations, the efficacy of these programmes can be substantially enhanced by tailoring them to the individual needs and behaviours of participants. This personalization is particularly crucial given the diverse nature of physical activity levels among different individuals. Machine learning, with its ability to analyse large datasets and uncover patterns, offers a promising solution for predicting physical activity levels and customizing well-being programmes accordingly.

The primary objective of this study is to develop a machine learning model that predicts physical activity levels, thereby facilitating the personalization of well-being programs. By accurately predicting activity levels, interventions can be designed to be more aligned with individual behaviours and preferences, ultimately enhancing programme adherence and effectiveness. After creation, the model predictions will be validated against actual data to assess whether the model can be used to effectively predict physical activity levels. This study will also contribute to the existing body of literature by exploring the integration of machine learning techniques in the field of health and well-being, an area that holds significant promise yet remains underexplored.

In the following sections, a comprehensive review of the literature will be conducted in order to understand the current state of research in this domain. This will include an examination of different approaches to calculating physical activity levels, possibilities to use physical activity levels to personalise well-being programmes, as well as various machine learning approaches used for predicting physical activity and their respective strengths and limitations. The methodology employed in this study will be then detailed, including data collection, preprocessing, model selection, and evaluation criteria. The results section will present the findings from our predictive models, followed by a discussion on their implications for personalized well-being programmes. Additionally predictions will be verified against factual physical activity levels calculated based on International Physical Activity Questionnaire (IPAQ) in order to evaluate their efficiency.

Ultimately, this research aims to bridge the gap between technological advancements in data collection and the practical application of these insights in personalizing health interventions. By leveraging machine learning, there is a hope to pave the way for more effective and individualized approaches to promoting physical activity and overall well-being. This, in turn, can lead to improved health outcomes, reduced healthcare costs, and a higher quality of life for individuals.

2. Literature Review

2.1. Physical Activity Levels

According to the World Health Organisation (WHO), physical activity can be defined as body movement induced by skeletal muscles, requiring an energy input (World Health Organization, 2024). Regular physical activity has a significant impact on the prevention of cardiovascular disease and other chronic conditions. Malm et al. (2019) indicate that aerobic exercise helps maintain normal body weight, reduces the risk of metabolic syndrome, normalises blood lipid concentrations, and alleviates some of the symptoms of cancer and the effects of its treatment. On the other hand, strength-based exercise reduces the risk of sarcopenia, osteoporosis or reduces the risk of falls in the elderly. Systematic physical activity is beneficial not only physically, but also mentally. Studies show that appropriately selected exercises have a positive effect on preventing or alleviating symptoms of depression. They also contribute to an improved quality of life, well-being, cognitive function, and enhance self-esteem and self-efficacy (Gieroba, 2019).

The 2020 WHO recommendations for minimum physical activity that has a positive impact on health and well-being indicate that adults aged 18-64 years should undertake at least 150-300 minutes per week of moderate-intensity aerobic physical activity or 75-150 minutes of high-intensity physical activity (Bull et al., 2020). Moderate-intensity physical activity is understood as an activity that causes a small acceleration of the heartbeat and respiratory rate, whereas we can understand high-intensity physical activity as an activity that significantly accelerates the heartbeat and respiratory rate (Bergier et al., 2019). In addition to aerobic physical activity, muscle-strengthening exercises (involving all major muscle groups) also provide additional health benefits. Research indicates that adults should perform such exercises at least twice a week. The intensity of these exercises should be at a moderate to high level. In addition, the World Health Organisation recommends reducing sedentary lifestyles as much as possible, and recommends that those exposed to such lifestyles increase their physical activity, both aerobic and muscle-strengthening, beyond recommended levels (Bull et al., 2020).

Physical activity is crucial for proper human development, but today there is a growing problem of insufficient levels of physical activity among the population. Studies indicate that approximately 31% of the world's population over 15 years of age have insufficient levels of physical activity, which may contribute to the deaths of approximately 3.2 million people per year (Park et al., 2020). Other studies show that Americans spend 55% of their time during the day (about 7.7 hours per day) inactive, while Europeans spend about 40% of their leisure time (2.7 hours per day) watching TV (Patterson et al., 2018). These observations reveal how a large proportion of the population leads a sedentary lifestyle.

A sedentary lifestyle is defined as time spent lying down or sitting, with low energy expenditure, that is, ≤1.5 metabolic equivalents (METs) (Bystron, 2023). There can be a multitude of causes of a sedentary lifestyle. The development of technology, infrastructure, transport or new areas of the economy has contributed significantly. Nowadays, it seems almost impossible to live without a telephone or computer. Social media play a huge role here. They make it possible to communicate and maintain relationships regardless of the actual distance between users. However, they are often the cause of various addictions, which can also result in reduced physical activity (Goździcki & Tomczyk, 2019). Physical activity levels can also be affected by economic change. In recent years, there has been an increasing displacement of physical work in favour of office work, which is mostly performed in a sedentary position. The development of transport and infrastructure has meant that people are less and less likely to cover a given distance, even a short one, on foot or by bicycle in everyday situations. The first choice in such situations is usually a car or public transport. Studies show that the place of residence can also influence the degree of physical activity. People living in rural areas are often more physically active than those living in urban areas (Martins et al., 2021). These observations demonstrate the importance of adequate promotion of healthy lifestyles today, with a strong emphasis on appropriate levels of physical activity. Well-being programmes significantly contribute to the promotion of physical activity by fostering a holistic approach to health that emphasizes the importance of regular exercise as a key component of overall mental and physical health.

2.2. Well-Being Programmes

Well-being programmes are special initiatives that are designed with a purpose of promoting physical and mental health of people. These programmes are often implemented in the context of environments, i.e. workplaces, schools or communities. In the case of a workplace it relates to all aspects of individuals working life: from safety and quality level of the environment itself to the climate at work and in the organization overall (Zhou et al., 2020).

In the past employee well-being was not as important as it is nowadays. From the managers' perspective, it was merely a useful addition rather than a strategic interest of the organisation (Berry et al., 2020). However, there has been a change in this trend over the last few years, to which the rising cost of healthcare has played a large part. Changes in employer health care costs per employee in the US are presented in Figure 1.

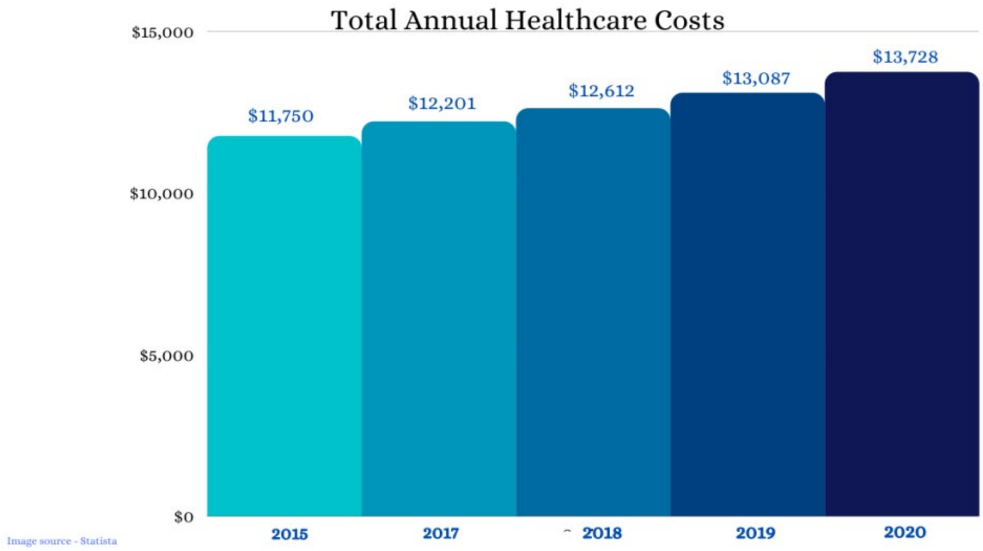

Figure 1. Changes in employer health care costs per employee in the US from 2015 to 2020 (US dollars)

Source: elaboration based on (Wellness360, 2022).

In working environments the main goal of well-being programmes is to help employees establish a healthy work-life balance, which can result in increased job satisfaction and reduced absenteeism, and thus also in reduced healthcare costs for companies. Although companies independently determine the components of their well-being programmes, there are some common elements that occur regularly in them. These include:

- on-site fitness facilities,
- access to mental health professionals,
- flexible working hours,
- initiatives that promotes healthy lifestyle (including physical activity).

In large companies, well-being programmes are a generic set of elements that take little account of the personal needs of the individual employee. However, the development of mobile and IoT technologies, which enable the continuous monitoring and collection of data on people's current health status, gives access to data the effective analysis of which allows for better personalisation of the programme for the individual employee (Oyebode et al., 2022). Machine learning algorithms are able to make effective use of this information by leveraging data to tailor health and wellness initiatives to individual needs and preferences.

2.3. Machine Learning

The proliferation of large-scale language models (LLMs) in everyday life, spearheaded by the release of ChatGPT from OpenAI, has brought artificial intelligence into the mainstream. However, it is worth remembering that the beginning of this field dates back to the middle of the previous century. In the 1950s, the first visions of autonomous thinking machines began to emerge. Naturally, many decades passed before there was a chance to turn them into reality. One of the constraints hindering the development of this field has been the problem of access to data that could be used for model development. The spread of Internet access initiated in the first decade of the 21st century has helped to prevent this problem and, at the same time, enhanced the development of sub-fields of artificial intelligence such as machine learning.

American computer scientist and artificial intelligence pioneer Arthur Samuel is credited with coining the term 'machine learning'. The term was first used in his work *Some studies in machine learning* *using the game of checkers* from 1959. Samuel defined machine learning as a scientific area in which a computer acquires the ability to learn without being programmed first (Samuel, 1959). The purpose of acquiring this learning ability is to perform a set task for which the models use appropriate algorithms and statistical models (Mahesh, 2020). Based on the type of algorithms and models used, machine learning can be divided into:

- Supervised learning – a type of machine learning in which the function used maps the input data against the output data using actual mappings from a so-called training set as an example. Supervised learning is characterised by dividing the data set into at least two sets: a training set and a test set. The training set contains features together with labels, representing the result to be achieved by the model being trained. The test set is used to test and validate the model. Examples of algorithms of this type are: decision trees, naive Bayes and Support Vectore Machine (SVM). The tasks carried out by supervised learning include: classification, which makes class predictions for each observation, and regression, which makes value predictions for each observation.
- Unsupervised learning – a type of machine learning in which the dataset is not labelled, so that the algorithm used has to organise the data itself. Examples of algorithms of this type are: K-Means Clustering and Principal Component Analysis (PCA). The tasks carried out by unsupervised learning include: clustering, where observations are assigned to clusters identified on the basis of similarity between their features, and feature reduction.
- Reinforcement learning – a type of machine learning in which an agent – in this example a learner or decision-maker that interacts with the environment – performs actions in order to learn how to make decisions. The purpose of these actions is to maximize some notion of cumulative reward (Mahesh, 2020).

With its ability to detect patterns in large, structured datasets, machine learning is widely used in many everyday fields. The situation is no different with well-being programmes. Examples of tasks that can be supported by this area include personalisation, predictive analysis, monitoring and optimisation. In terms of personalisation, machine learning models are able to support the customisation of fitness and nutrition plans. Predictive analysis enables the earlier detection of potential risks to the health or life of workers by analysing their health data (Tate et al., 2020).

3. Methodology

The study used quantitative research methods: literature review and experiment. In the experiment, a machine learning model was trained on a collected dataset. Since the aim was to predict the value of physical activity levels, a regression model was chosen with random forest algorithm. This model was then used to create predictions for the data from the specified test set. Based on the predictions, the model was evaluated to verify its effectiveness in predicting physical activity levels. The following metrics were used for the purpose of evaluation:

- Mean Absolute Error (MAE),
- Root Mean Squared Error (RMSE),
- Coefficient of Determination (R^2).

3.1. Description of the Examined Dataset

The dataset contains information on 78 individuals who underwent voluntary surveys that elicited information on their features and level of physical activity (calculated based on answers to the GPAQ questionnaire). These features were used as a training dataset for machine learning regression model whose task was to use this data to predict the total level of physical activity. The calculated total level of physical activity, expressed in metabolic equivalent (MET) min/week, was used to verify the predictions.

The dataset consisted of the following features:

- gender,
- age (in years),
- body weight (in kilograms),
- body height (in centimetres),
- BMI,
- waist circumference (in centimetres),
- hip circumference (in centimetres),
- waist-to-hip ratio (WHR),
- waist-to-height ratio (WHtR),
- grip power of the dominant hand (in kilograms).

The statistical characteristics of the examined group are presented in Table 1.

Table 1. The statistical characteristics of the examined group

Features	Gender	Mean	Std	Min	Max
Age (in years)	female	39.0	7.0	26.0	50.0
	male	36.0	7.5	20.0	50.0
Body height (in cm)	female	163.8	5.7	153.0	173.0
	male	179.6	7.0	168.0	201.0
Body weight (in kg)	female	58.7	7.7	48.0	74.0
	male	77.3	9.5	58.0	97.0
BMI (in kg/m^2)	female	21.8	2.2	18.6	26.6
	male	23.9	2.3	19.5	29.3

Source: own elaboration.

3.2. Model Construction

The first step was to subject the dataset simple pre-processing to better prepare the data for model training and testing. The 'gender' feature was converted to a numeric type using label encoding (0 – female, 1 – male). Other features were standardised using the StandardScaler function.

In the next step, a correlation matrix was created between the characteristics in order to detect potential multicollinearity between them. The correlation matrix created is presented in Table 2.

Table 2. The correlation matrix between features in the dataset

	Age	Body weight	Body height	BMI	Waist circumference	Hip circumference	WHR	WHtR	Grip power of the dominant hand	Gender
Age	1.00	–0.10	–0.17	0.01	0.07	–0.03	0.12	0.18	–0.15	–0.19
Body weight	–0.10	1.00	0.82	0.80	0.88	0.84	0.58	0.60	0.74	0.70
Body height	–0.17	0.82	1.00	0.33	0.60	0.63	0.35	0.15	0.82	0.74
BMI	0.01	0.80	0.33	1.00	0.84	0.74	0.60	0.85	0.38	0.40
Waist circumference	0.07	0.88	0.60	0.84	1.00	0.76	0.82	0.88	0.58	0.65
Hip circumference	–0.03	0.84	0.63	0.74	0.76	1.00	0.26	0.57	0.52	0.47
WHR	0.12	0.58	0.35	0.60	0.82	0.26	1.00	0.80	0.42	0.57
WHtR	0.18	0.60	0.15	0.85	0.88	0.57	0.80	1.00	0.23	0.36
Grip power of the dominant hand	–0.15	0.74	0.82	0.38	0.58	0.52	0.42	0.23	1.00	0.81
Gender	–0.19	0.70	0.74	0.40	0.65	0.47	0.57	0.36	0.81	1.00

Source: own elaboration.

After pre-processing, the dataset was split into training and test sets in a 2 : 1 ratio. The training set was used to train a regression model using a random forest algorithm (number of estimators = 100).

4. Research Results

Once the learning process was complete, the model was used to create predictions based on the data from the test set. The predictions created were evaluated to determine the effectiveness of the constructed model. The values of the evaluation metrics for the model are presented in Table 3.

Table 3. The values of the evaluation metrics of the model

Metric	Value
Mean Absolute Error	2602.16
Root Mean Squared Error	4304.86
Coefficient of Determination	–0.21

Source: own elaboration.

The model achieved the MAE metric value of 2602.16. As this is the average value of the absolute differences between the model's predictions and the actual values, the lower the value of this metric, the more effective the model is. A value of 2602.16 means that the MAE is greater than the standard deviation of the physical activity level values, indicating that the model predicts with greater error than the natural variation in the data.

The model achieved the RMSE metric value of 4304.86. This metric is used to measure the mean magnitude of the prediction error of a regression model by squaring and then averaging and rooting. Value of RMSE is much higher than that of MAE, which means that there are large errors in the data that affect the value of the first metric.

The value of the R^2 metric was less than 0 and was at the level of –0.21. By definition, the coefficient of determination indicates how well a linear regression model explains variation in the dependent data (i.e. the response variable) on the basis of the independent variables (i.e. the predictors). This means that the value of –0.21 indicates that the model has no predictive value.

Based on the achieved values of the evaluation metrics, it can be concluded that predicting physical activity levels using machine learning from data describing human characteristics is not effective enough to be used in personalising well-being plans.

5. Conclusions

In order to determine whether machine learning can be used to predict physical activity levels from data describing people's characteristics for personalising well-being programmes, a regression model using the random forest algorithm was trained and evaluated. During the evaluation, the metrics used indicated that the model's predictions were not effective. Based on this, it can be concluded that using only the characteristics used in the dataset to predict physical activity levels is not a good idea, since the context of the information collected on individuals should be more extended rather than limited to their characteristics. It is also worth adding that the size of the dataset may also have played a role in the poor performance of the model. For future research in this context, it would be useful to provide a set containing more observations than 78.

In future studies, it would be worthwhile for the dataset to be expanded to include other information about the people surveyed that could be useful in predicting physical activity levels. Besides, it would be worthwhile to broaden the scope of the study by training several machine learning models using different regression algorithms. The results they achieve through evaluation could be compared to see which algorithm would perform best in the task of predicting physical activity levels for the personalisation of well-being programmes.

References

Bergier, J., Wasilewska, M., & Szepeluk, A. (2019). Global Physical Activity Questionnaire (GPAQ) – the Polish Version. *Health Problems of Civilization*, *13*(1), 1-8. https://doi.org/10.5114/hpc.2018.80229

Berry, L., Mirabito, A. M., & Baun, W. B. (2020). What's the Hard Return on Employee Wellness Programs? *Harvard Business Review* (December), Mays Business School Research Paper No. 2012-68, 2012-2068. https://ssrn.com/abstract=2064874

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., Carty, C., Chaput, J.-P., Chastin, S., Chou, R., Dempsey, P. C., DiPietro, L., Ekelund, U., Firth, J., Friedenreich, C. M., Garcia, L., Gichu, M., Jago, R., Katzmarzyk, P. T., & Willumsen, J. F. (2020). World Health Organization 2020 Guidelines on Physical Activity and Sedentary Behaviour. *British Journal of Sports Medicine*, *54*(24), 1451-1462. https://doi.org/10.1136/bjsports-2020-102955

Bystron, J. M. (2023). Siedzący tryb życia oraz praca zdalna jako czynniki predysponujące do rozwoju nowotworów złośliwych. *Medycyna Środowiskowa*, *26*(3-4), 114-118. https://doi.org/10.26444/ms/175315

Gieroba, B. (2019). Wpływ aktywności fizycznej na zdrowie psychiczne i funkcje poznawcze. *Medycyna Ogólna i Nauki o Zdrowiu*, *25*(3), 153-161.

Goździcki, T., & Tomczyk, W. (2019). Zagrożenia mediów społecznościowych w opinii studentów LAW – w świetle wyników badań własnych. *Annales Universitatis Mariae Curie-Skłodowska, Sectio J*, *32*(3), 233-248. https://doi.org/10.17951/j.2019.32.3.233-248

Mahesh, B. (2020). Machine Learning Algorithms – A Review. *International Journal of Science and Research*, *9*(1), 381-386. https://doi.org/10.21275/ART20203995

Malm, C., Jakobsson, J., & Isaksson, A. (2019). Physical Activity and Sports – Real Health Benefits: A Review with Insight into the Public Health of Sweden. *Sports*, *7*(5), Article 127. https://doi.org/10.3390/sports7050127

Martins, L. C. G., Lopes, M. V. D. O., Diniz, C. M., & Guedes, N. G. (2021). The Factors Related to a Sedentary Lifestyle: A Meta-Analysis Review. *Journal of Advanced Nursing*, *77*(3), 1188-1205.

Oyebode, O., Fowles, J., Steeves, D., & Orji, R. (2023). Machine Learning Techniques in Adaptive and Personalized Systems for Health and Wellness. *International Journal of Human-Computer Interaction*, *39*(9), 1938-1962. https://doi.org/10.1080/10447318.2022.2089085

Park, J. H., Moon, J. H., Kim, H. J., Kong, M. H., & Oh, Y. H. (2020). Sedentary Lifestyle: Overview of Updated Evidence of Potential Health Risks. *Korean Journal of Family Medicine*, *41*(6), 365. https://doi.org/10.4082/kjfm.20.0165

Patterson, R., McNamara, E., Tainio, M., de Sá, T. H., Smith, A. D., Sharp, S. J., Edwards, P., Woodcock, J., Brage, S., & Wijndaele, K. (2018). Sedentary Behaviour and Risk of All-Cause, Cardiovascular and Cancer Mortality, and Incident Type 2 Diabetes: A Systematic Review and Dose Response Meta-Analysis. *European Journal of Epidemiology*, *33*, 811--829. https://doi.org/10.1007/s10654-018-0380-1

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210-229. https://doi.org/10.1147/rd.33.0210

Tate, A. E., McCabe, R. C., Larsson, H., Lundström, S., Lichtenstein, P., & Kuja-Halkola, R. (2020). Predicting Mental Health Problems in Adolescence Using Machine Learning Techniques. *PLoS One*, *15*(4), Article e0230389. https://doi.org/10.1371/journal.pone.0230389

Wellness360. (2022). *Study Shows the Healthcare Cost Trends and US Firms Offering Health and Wellness Programs*. https://blog.wellness360.co/study-shows-the-healthcare-cost-trends-and-us-firms-offering-health-and-wellness-programs/

World Health Organization. (2024). *Physical Activity*. https://www.who.int/news-room/fact-sheets/detail/physical-activity

Zhou, X., Rasool, S. F., & Ma, D. (2020). The Relationship between Workplace Violence and Innovative Work Behavior: The Mediating Roles of Employee Well-Being. *Healthcare*, *8*(3), Article 332. https://doi.org/10.3390/healthcare8030332

Przewidywanie poziomu aktywności fizycznej na potrzeby personalizacji programów *well-being*

Streszczenie

Cel: Celem artykułu jest sprawdzenie, czy model uczenia maszynowego może być skutecznie wykorzystywany do przewidywania poziomów aktywności fizycznej ludzi w celu personalizacji programów *well-being* dla pracowników w firmach.

Metodyka: W artykule wykorzystano następujące metodologie badawcze: analizę literatury oraz eksperyment w postaci weryfikacji przewidywań dokonanych przez stworzony model uczenia maszynowego.

Wyniki: Wyniki uzyskane z oceny modelu wykazały, że wykorzystanie danych dotyczących cech ludzkich do przewidywania poziomów aktywności fizycznej w celu personalizacji programów *well-being* nie gwarantuje wystarczająco dobrych wyników.

Implikacje i rekomendacje: Dzięki skutecznemu przewidywaniu poziomów aktywności fizycznej, programy *well-being* mogą być lepiej spersonalizowane pod kątem indywidualnych potrzeb pracowników, co może się przyczynić do poprawy ich stanu zdrowia.

Oryginalność/wartość: Przegląd literatury wykazał, że wykorzystanie uczenia maszynowego do przewidywania poziomów aktywności fizycznej nie zostało szczegółowo opisane w literaturze.

Słowa kluczowe: GPAQ, MET, *well-being*, *machine learning*