Medical Card Information System for Data Analysis from Fitness Bracelets

Oleksii Bychkov, Oleksandr Gezerdava, Kseniia Dukhnovska, Oksana Kovtun and

Olga Leshchenko

Taras Shevchenko National University of Kyiv, Bohdan Hawrylyshyn str. 24, Kyiv, UA-04116, Ukraine

Abstract

This study explored using fitness tracker data to improve patient health monitoring. Patient health data comes from various sources, including medical records and wearable devices. This study aimed to develop medical record software that analyzes fitness tracker data. This would allow data collection for further analysis (like cluster analysis) and potentially improve the accuracy and features of medical monitoring. We focused on using linear regression to analyze and predict heart rate based on fitness tracker data (Very Active Distance, Fairly Active Minutes, Calories). An information system was developed to perform this analysis. Training and validation were conducted using real fitness tracker data. The study confirmed that linear regression effectively predicts heart rate based on fitness tracker parameters. We compared the model's accuracy with and without data aggregation to determine the optimal conditions for using linear regression with fitness data. Statistical analysis (Student's t-test) confirmed the adequacy of the model ($t = 1.31$, critical value = 2.62). This study demonstrates that linear regression, applied through the developed software, is a valuable tool for personalized monitoring and optimization of physical activity using fitness tracker data. Linear regression may not be ideal for all situations, especially with complex non-linear relationships. Other machine learning methods might be needed in those cases.

Keywords ¹

Medical information system, human health monitoring, linear regression

1. Introduction

The rise of artificial intelligence (AI) has sparked excitement about its potential to revolutionize healthcare. AI can analyze vast amounts of medical data, leading to more precise diagnoses and improved patient outcomes. AI-powered systems can streamline administrative tasks and potentially identify cost-saving treatment options. By analyzing trends and patterns, AI can help predict and prevent the spread of infectious diseases.

However, unlocking AI's full potential hinges on access to high-quality data. Existing data sets often lack crucial details about a patient's medical history. This necessitates the collection of anonymized, "sensitive data" with patient consent, specifically for machine learning applications.

This data not only personalizes patient care but also empowers AI to identify disease outbreaks and impending healthcare crises. The ability to rapidly analyze vast data sets allows for swift responses and resource allocation to prevent epidemics, ultimately strengthening public health surveillance.

©️ 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CEUR Workshop Proceedings (CEUR-WS.org)

 $\overline{}$

Dynamical System Modeling and Stability Investigation (DSMSI-2023), December 19-21, 2023, Kyiv, Ukraine

EMAIL: oleksiibychkov@knu.ua (O. Bychkov)[; oleksandrgezerdava@knu.ua](mailto:oleksandrgezerdava@knu.ua) (O. Gezerdava); duchnov@ukr.net (K. Dukhnovska); oksana.kovtun@knu.ua (O. Kovtun)[; olga.leshchenko@knu.ua](mailto:olga.leshchenko@knu.ua) (O. Leshchenko)

ORCID: 0000-0002-9378-9535 (O. Bychkov)[; 0009-0002-9494-5541](https://orcid.org/0009-0002-9494-5541) (O. Gezerdava); 0000-0002-4539-159X (K. Dukhnovska); [0000-](https://orcid.org/0000-0003-0871-5097) [0003-0871-5097](https://orcid.org/0000-0003-0871-5097) (O. Kovtun)[; 0000-0002-3997-2785](https://orcid.org/0000-0002-3997-2785) (O. Leshchenko)
 \odot 0 2023 Copyright for this paper by its authors.

This research focuses on developing an information technology system capable of analyzing medical information. By identifying common patterns, trends, and templates within the data, the system can aid in both diagnosis and treatment.

In today's era of rapid technological advancement and pervasive information technology, implementing a Medical Information System (MIS) is a crucial step forward for medicine. Developing information systems for processing basic medical data (electronic medical records) is particularly important.

Therefore, this research delves into the development of electronic medical records (EMRs). EMRs are a foundational element for collecting, systematizing, and ultimately processing medical data using AI and other cutting-edge technologies. Research in this area holds significant value as it has the potential to significantly improve healthcare quality, accessibility, and efficiency.

2. Literature review and problem statement

Modern medical information systems powered by AI are poised to transform medicine and public health. These systems leverage digital patient records, creating a rich data source for AI analysis. This translates to several key benefits for both doctors and patients. AI can analyze vast amounts of medical data, uncovering hidden patterns and trends that may escape the human eye. This translates to more accurate diagnoses and improved patient outcomes. AI can analyze patient data to predict potential health risks and disease progression. This allows doctors to intervene early, preventing complications and improving overall patient well-being.

This shift towards AI-powered medical records offers a glimpse into the future of healthcare, where data-driven insights empower doctors to deliver more personalized and preventive care.

In the work [1], electronic medical records were used to study the quality of care for people with Parkinson's disease. Specialized software was used to collect statistical data, and a sorting method using closed tabs was used to structure specific data.

In the study [2], an evaluation of various data stream processing methods from electronic medical records of patients suffering from chronic obstructive pulmonary disease was conducted. The main goal of this study was to establish similarity between patients, namely highlighting the most representative and clinically significant data from their electronic medical records. For this purpose, the following data were selected as clinical features: gender, body mass index, smoking status, and personal history of atopy. To quantitatively assess the contribution of each clinical feature to the overall similarity of patients, the authors developed a metric of relative variability, which compares the variability of the feature among N nearest neighbors of each observation with the variability of the feature in the entire cluster.

The aim of the work [3] was to develop an electronic medical record for patients with schizophrenia to improve the management of their clinical information. The study shows that such a record gives the medical team access to up-to-date and complete patient information, which, in turn, helps them make more informed and effective treatment decisions. The electronic medical record allows hospital staff to better organize and systematize information about patients, making it more accessible and convenient to use. This can significantly facilitate the treatment process for patients with schizophrenia, as the electronic medical record saves time and resources, and makes the process more clear and organized. Thanks to these advantages, the electronic medical record can help hospitals better manage patient information and, as a result, improve the overall provision of medical care.

The study [4] used data from the "Epic" health information management system. Using electronic clinical data, monitoring was conducted on the condition of patients with juvenile rheumatoid arthritis, revealing that correlates for the overall assessment of arthritis by a physician differ depending on the subtype of the disease.

Also, using data from the "Epic" health information management system, the study [5] analyzed a variety of real-time data streams from patients arriving at the emergency department of the hospital in order to forecast total hospitalization within a short period of time. Hospital bed planners closely collaborated with the research group to determine their requirements. They requested to send daily forecasts four times a day regarding the need for beds in the next four and eight hours for further

comparison with their internal reports. As part of the study, a program was developed that automatically formats and sends emails to bed planners every four reporting hours. This program is based on a machine learning algorithm used to forecast bed availability in the hospital.

Besides medical information systems, more and more attention is being drawn to IoT in the field of medicine, which is capable of filling these systems with information in the future. An example of such research can be found in [6]. In this work, a device for conducting electroencephalograms of the brain was developed. The device records at least 16 channels of signals from the human brain and then transmits electronic clinical data to the patient's history, allowing for automated data analysis in the future.

Also, in study [7], an analysis of the accuracy of sleep registration using wearable devices, including Xiaomi Mi Smart Band 5, was conducted, comparing their data with polysomnography (PSG) data - the gold standard in sleep assessment. The study showed significant differences between sleep indicators obtained from Xiaomi Mi Smart Band 5 and PSG. Xiaomi Mi Smart Band 5 cannot completely replace PSG in determining sleep stages but can be useful as an initial means of assessing sleep quality. Wearable devices can be used to monitor sleep and serve as a basis for consultation with a doctor or adjustment of sleep habits.

In the study [8], the accuracy of sleep recording using wearable devices, including the Xiaomi Mi Smart Band 5, was analyzed by comparing their data with polysomnography (PSG) data - the gold standard in sleep assessment.

The study showed significant differences between the sleep parameters obtained from the Xiaomi Mi Smart Band 5 and PSG. The use of paired t-tests and Bland-Altman plots confirms a large difference between the means of different sleep parameters measured by both methods.

EBE analysis showed different levels of sensitivity, specificity, and agreement between Xiaomi and PSG in determining different sleep stages, such as light sleep, deep sleep, and REM sleep.

While the Xiaomi Mi Smart Band 5 cannot fully replace PSG in sleep stage determination, they can be useful for consumers as an initial tool for assessing sleep quality. These devices can be used to monitor sleep and serve as a basis for consultation with a doctor or for adjusting one's sleep habits.

Thus, the study confirms that wearable devices, including the Xiaomi Mi Smart Band 5, may be suitable for simplified sleep quality monitoring, but their accuracy does not allow them to fully replace professional PSG studies in sleep phase assessment.

In turn, the data format and their compatibility become a key element for further systematization and processing of data. Study [9] proposes a model for compatibility of medical data standards, privacy protection methods, and medical image measurements. This model applies Health Level Seven (HL7) and Digital Imaging and Communications in Medicine (DICOM) standards to medical image data standards. This approach ensures increased access to medical image data in accordance with privacy laws through de-identification methods. This study focuses on proposing a standard for the measurement values of standard materials, which eliminates the uncertainty in measurements that was not foreseen by previously existing standards for medical image analysis. The study found that medical image data standards are consistent with existing standards and also provide privacy protection for any medical images using de-identification methods.

In medical practice, it is important not only information technologies that help in the treatment of a patient or in the diagnosis of his disease. Information technologies that predict threats to human health are also important. Thus, in study [10, 11], a model for predicting the thermal state of a person during physical activity in a hot environment is proposed. The model took into account the characteristics of the environment, namely air temperature, relative humidity and wind speed. The intensity of the load and the duration of its impact on the person are also taken into account here. As a result of the work of such a model, a dynamic of characteristics is obtained that reflects the physiological state of a person in hot conditions during physical exertion.

Overall, these studies confirm the potential of IoT and modern medical technologies to improve the collection, processing, and analysis of medical data. They demonstrate promising opportunities for the use of wearable devices to monitor patients and improve access to medical information. However, at the same time, they emphasize the importance of further research and improvement of technologies to ensure their reliability and accuracy in clinical use.

Information technology for disease analytics can combine elements of machine learning for data analysis and risk forecasting, data processing from wearable devices such as fitness trackers for realtime health monitoring, electronic medical records for accessing a patient's medical history and test results. This will provide advantages such as improving diagnostics through early detection and disease forecasting, personalizing treatment and disease prevention, increasing efficiency, and saving costs in the healthcare system.

3. The purpose and objectives of the research

The aim of the research is to develop an information system for medical records that provides analysis of data obtained from fitness trackers. To achieve this goal, the following tasks were set:

- development and training of a mathematical model in the form of linear regression to predict heart rate;
- development of an electronic medical card with a module for predicting changes in heart rate based on data obtained from the fitness tracker.

The expected results of the research are the creation of an information system that will collect data from fitness trackers and use them to predict heart rate. As a result, there will be an improvement in the accuracy of medical monitoring through the use of data collected from fitness trackers. The information system will enable early detection of cardiovascular problems.

4. Research materials and methods

The dataset used in this study was taken from the official Kaggle website [12]. The data was collected through Amazon Mechanical Turk between December 3, 2016 and December 5, 2016. In total, the dataset contains data from Fitbit users who, in accordance with legal requirements, consented to the provision of personal data from the tracker, including physical activity, heart rate, and sleep monitoring at the minute level. Individual reports in the dataset can be analyzed by export session identifier (column A) or timestamp (column B). The difference in output means the use of different types of Fitbit trackers and individual tracking behaviors/preferences.

In fig. 1 shows an example of data elements from daily Activity merged.csv:

Figure 1: Examples of datasets

The complete data set consists of 18 csv files (Fig. 2).

Figure 2: Complete data set

The analysis of unique identifiers turns out to be a key stage in understanding the properties of data and their distribution in the specified set (Fig. 3).

Figure 3: Identified data

Identifying and comparing the number of unique IDs to the total number of records can help reveal potential anomalies, such as duplicates or data loss. This procedure allows for the detection of potential data quality issues, which is critical for the reliability and correctness of further analysis.

In addition, unique identifiers can be used as keys to join data from different sources or tables. This facilitates the identification and matching of information about the same objects or observations in different sources.

As can be seen from Fig. 3, not all activity indicators are present for the individuals who provided their data for analysis. This will be taken into account in the study.

The Pandas library in the Python programming language was used to process and transform the data contained in different sets (heartrate, intensities, calories, activity, sleep) (Fig. 4). This analysis is aimed at processing the date and time in text format for further analysis and use of this data in relevant research.

```
# intensities
intensities['ActivityHour'] = pd.to_datetime(intensities['ActivityHour'], format="%m/%d/%Y
 %I:%M:%S %p")
intensities['time'] = intensities['ActivityHour'].dt.strftime("%H:%M:%S")
intensities['date'] = intensities['ActivityHour'].dt.strftime("%m/%d/%y")
# calories
calories['ActivityHour'] = pd.to_datetime(calories['ActivityHour'], format="%m/%d/%Y %I:%M:%S
 %p'calories['time'] = calories['ActivityHour'].dt.strftime("%H:%M:%S")
calories['date'] = calories['ActivityHour'].dt.strftime("%m/%d/%y")
# activity
activity['ActivityDate'] = pd.to_datetime(activity['ActivityDate'], format="%m/%d/%Y")
activity['date'] = activity['ActivityDate'].dt.strftime("%m/%d/%y")
# sleep
sleep['SleepDay'] = pd.to_datetime(sleep['SleepDay'], format="%m/%d/%Y %I:%M:%S %p")
sleep['date'] = sleep['SleepDay'].dt.strftime("%m/%d/%y")
```
Figure 4: Date and time format standardization for further analysis: Python code

These actions are performed to standardize the representation of date and time into a special date/time object type, which simplifies further operations with time series analysis, sorting, filtering, and data visualization. This approach facilitates ease of processing and makes analysis more accessible for further scientific conclusions and use in research.

In addition, it is worth noting that for the convenience of this study, the data of all individuals was also summarized by each available parameter and presented in the form of graphs. First of all, the results of the study of the distribution of heart rate (HR) in the heartrate_selected_columns dataset were selected. This dataset contains HR values in beats per second, collected using a fitness bracelet to track health.

Thus, the data were prepared for model training by processing and standardizing information on physical activity, heart rate, and sleep monitoring collected using a Fitbit tracker.

5. Mathematical model in the form of linear regression for predicting heart rate

This study uses a mathematical model of linear regression, which considers the dependence of the heart rate (labeled as "Value") on three main variables:

 $Value = \beta_0 + \beta_1 \times VeryActiveDistance + \beta_2 \times FairlyActiveMinutes + \beta_3 \times Calories + \epsilon$ (1)

where:

Value – heart rate (dependent variable),

VeryActiveDistance – variable representing the distance traveled by the user during very physical activities such as running or vigorous aerobics. It is measured, for example, in kilometers or meters,

FairlyActiveMinutes – the variable reflects the number of minutes during which the user spent some physical activity. This may include walking or light aerobics. Measured in minutes,

Calories – the variable represents the number of calories consumed by the user during the activity. It can take into account calories consumed during exercise and basal metabolism. It is measured in kilocalories (kcal) or joules (J).

 $β₀ - a constant (transition term),$

 β_l , β_2 , β_3 – regression coefficients that determine the influence of the relevant variables,

ε – reflects random errors originating from unaccounted factors or random deviations from expectations.

This model allows you to quantify the effect of each variable on the heart rate and predict its value. Random error (ϵ) takes into account unknown or unpredictable factors that can affect the heart rate and introduce random deviations into predictions.

6. Development of system

The presented medical card system uses data from a fitness bracelet to monitor and diagnose the patient's health (Fig. 5). Fitness bracelets collect data such as the number of steps, heart rate, sleep quality and other parameters characterizing the state of health. The information system synchronizes data in real time.

In particular, one of the important directions is the use of data obtained from fitness bracelets for objective assessment of patients' condition. Thanks to the use of methods of machine learning and data analysis, it becomes possible not only to diagnose the current state of health, but also to predict its possible risks and trends. The presented information system uses a module for predicting a person's heart rate based on data obtained from a fitness bracelet. The module is built on the basis of the mathematical model (1).

To develop this module, popular Python libraries for working with data, modeling and visualization are used: pandas for data processing, scikit-learn (sklearn) for modeling linear regression, and matplotlib for visualization of results.

Figure 5: Medical card interface

Real and predicted heart rate values are displayed on the graph (Fig. 6), which allows you to visually compare how well the model adapts to real data. This makes it easier to understand the effectiveness and scope of the linear regression model in this context.

Many points on the graph are closely clustered together, indicating a high degree of correlation between the predicted and actual values of heart rate. This suggests that the linear regression model effectively captures the relationship between the selected independent variables and heart rate, confirming its adequacy for these specific data.

However, attention should be paid to some outliers or deviations from the main cluster of points. These cases may indicate situations where the model encounters difficulties in accurately predicting heart rate. Perhaps they reflect the influence of other factors or anomalies in the data, which should be further examined and taken into account in the analysis of the results.

Figure 6: Graphic presentation of forecasting results

The graph also shows three horizontal lines, which may represent the mean or median values for different data groups. This may indicate the presence of subgroups or peculiarities in the data that could affect the model's results. Investigating these sections could lead to a deeper understanding of the dynamics of the interaction between the selected variables and heart rate.

7. Discussion of the results of using the information system of the medical record for the analysis of data from fitness bracelets

The use of linear regression in the architecture of information systems for analyzing and predicting heart rate based on data from fitness trackers is an important research direction. Linear regression is a well-known statistical method widely used in solving numerous tasks of statistical data processing. It works well on large and sparse datasets without complex trends.

In building this model, we assumed an approximate linear relationship between heart rate and characteristics of physical activities. This has a simple explanation. During physical activities, a person's muscles require more oxygen. Oxygen is used to produce energy. During inhalation, oxygen enters the arterial blood, which is transported by the heart's work. During physical exertion, the heart works harder to supply oxygen to working muscles, which can affect heart rate. Typically, an increase in physical activity leads to an increase in heart rate. This explains the model's results.

The modeling showed that the expected heart rate will increase by 0.3996623 beats per minute with each meter walked. At the same time, since we use characteristics of a single physical activity, the expected heart rate will decrease by 0.11206182 with each second of this activity and by 0.00063471 with each calorie burned.

The model's adequacy has been proven, as the Student's criterion calculated is significantly smaller than the critical value.

From Figure 2, it can be seen that the standard error, which shows how far the actual heart rate value can be from the expected value, is negligible.

Comparative analysis of the loss function helps to understand how different optimization methods affect the quality and speed of model training.

Forecasting heart rate has been studied for a long time and by various methods. However, unlike other models, the proposed model is the most accessible to users. You only need to have a fitness tracker and software for analyzing data from fitness trackers. The model predicts heart rate at any time during physical activity, unlike the model described in study [13], which predicts only for five seconds.

With the developed model, the user can plan physical workouts without harming their health. Since this model is aimed at predicting heart rate during exercise, which distinguishes it from the LSTM-BiLSTM-Att model ([14]), where the expected heart rate is calculated for a person at rest.

The study was conducted on data from young people of almost the same physical fitness level. How model (1) will behave for other age groups was not investigated.

Among the disadvantages of the model, it should be noted that linear regression typically uses independent variables. However, here, characteristics of a single event are considered, which is unlikely to prove that they are independent. However, the adequacy of the model has been proven.

8. Conclusions

This research explored the use of fitness tracker data to personalize heart health monitoring and physical activity optimization. We systematized user data and developed a mathematical model based on linear regression.

This model effectively analyzes data from parameters like Very Active Distance, Fairly Active Minutes, and Calories to predict heart rate. This allows for individual monitoring and optimization of physical activity, ultimately promoting cardiovascular health. The model's adequacy is confirmed by a Student's t-test ($t = 1.31$, critical value = 2.62).

Building upon this model, we developed and implemented an information system. This system utilizes linear regression to analyze the impact of these fitness tracker metrics on heart rate.

The key feature of this system is its ability to not only record health data but also to monitor and forecast health trends. This empowers individuals to take a more proactive role in managing their cardiovascular health.

9. References

- [1] Connor, K. I., Siebens, H. C., Mittman, B. S., Ganz, D. A., Barry, F., Ernst, E. J., ... & Vickrey, B. G. (2020). Quality and extent of implementation of a nurse-led care management intervention: care coordination for health promotion and activities in Parkinson's disease (CHAPS). BMC health services research, 20, 1-17. [https://doaj.org/article/1a86bc727d50495e968f63d952da1982.](https://doaj.org/article/1a86bc727d50495e968f63d952da1982)
- [2] Pikoula, M., Kallis, C., Madjiheurem, S., Quint, J. K., Bafadhel, M., & Denaxas, S. (2023). Evaluation of data processing pipelines on real-world electronic health records data for the purpose of measuring patient similarity. Plos one, 18(6), e0287264. [https://doi.org/10.1371/journal.pone.0287264.](https://doi.org/10.1371/journal.pone.0287264)
- [3] Safdari, R., Hamidi, M., Aghaee, M., & Ghazi Saeedi, M. (2017). Designing electronic card of health for schizophrenic patients. Payavard Salamat, 10(6), 479-487.
- [4] Miller, M. L., Ruprecht, J., Wang, D., Zhou, Y., Lales, G., McKenna, S., & Klein-Gitelman, M. (2011). Physician assessment of disease activity in JIA subtypes. Analysis of data extracted from electronic medical records. Pediatric Rheumatology, 9, 1-7.<https://doi.org/10.1186/1546-0096-9-9> .
- [5] King, Z., Farrington, J., Utley, M., Kung, E., Elkhodair, S., Harris, S., ... & Crowe, S. (2022). Machine learning for real-time aggregated prediction of hospital admission for emergency patients. NPJ Digital Medicine, 5(1), 104. pp. 1 – 1[2 https://doi.org/10.1038/s41746-022-00649-y.](https://doi.org/10.1038/s41746-022-00649-y)
- [6] Malakhov, K. S. (2023). Insight into the digital health system of Ukraine (ehealth): Trends, definitions, standards, and legislative revisions. International Journal of Telerehabilitation, 15(2). <https://doi.org/10.5195/ijt.2023.6599>
- [7] Khiani, S., Iqbal, M. M., Dhakne, A., Thrinath, B. S., Gayathri, P. G., & Thiagarajan, R. (2022). An effectual IOT coupled EEG analysing model for continuous patient monitoring. measurement: sensors, 24, 100597. [https://doi.org/10.1016/j.measen.2022.100597A](https://doi.org/10.1016/j.measen.2022.100597)NSYS Fluent. URL: <https://www.ansys.soften.com.ua/products/fluids/ansys-fluent.html>
- [8] Martínez-Martínez, F. J., Concheiro-Moscoso, P., Miranda-Duro, M. D. C., Boedo, F. D., Muiño, F. J. M., & Groba, B. (2020, August). Validation of self-quantification Xiaomi band in a clinical sleep unit. In *Proceedings* (Vol. 54, No. 1, p. 29). MDPI. <https://doi.org/10.3390/proceedings2020054029>
- [9] Kwon, O., & Yoo, S. K. (2021). Interoperability Reference Models for Applications of Artificial Intelligence in Medical Imaging. *Applied Sciences*, *11*(6), 2704. <https://doi.org/10.3390/app11062704>
- [10] Yermakova I., Nikolaienko A., Tadeieva J., Bogatonkova A., Solopchuk Y., Gandhi O. "Computer model for heat stress prediction during physical activity." 2020 IEEE 40th International scientific conference electronics and nanotechnology (ELNANO). Kyiv, Ukraine, 2020, April 22-24, pp. 569-573. <https://doi.org/10.1109/ELNANO50318.2020.9088846>
- [11] V. Kraevsky, O. Kostenko, O. Kalivoshko, N. Kiktev and I. Lyutyy, "Financial Infrastructure of Telecommunication Space: Accounting Information Attributive of Syntalytical Submission," *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)*, Kyiv, Ukraine, 2019, pp. 873-876, doi: 10.1109/PICST47496.2019.9061494.
- [12] https://www.kaggle.com/datasets/arashnic/fitbit
- [13] Zhu, Z., Li, H., Xiao, J., Xu, W., & Huang, M. C. (2022). A fitness training optimization system based on heart rate prediction under different activities. *Methods*, *205*, 89-96.
- [14] Lin, H., Zhang, S., Li, Q., Li, Y., Li, J., & Yang, Y. (2023). A new method for heart rate prediction based on LSTM-BiLSTM-Att. *Measurement*, *207*, 112384.