

AI-assisted sensor data processing using machine learning and fuzzy logic

János Dávid Balogh, Attila Adamkó

Faculty of Informatics, University of Debrecen
davidkm2@mailbox.unideb.hu
adamko.attila@inf.unideb.hu

Abstract. In today's world, we are surrounded by smart devices, which we increasingly focus on, while we may pay less attention to our health and sports activities. Smart devices, especially those we wear, can collect a lot of data about us, which might be used to maintain and improve our health. This paper is dedicated to creating an AI assistant that creates two-way communication between the user and the components. Two components ensure that the goals are achieved: a service containing machine learning and a service implementing fuzzy logic. The services read the data from a database cluster. Since a lot of data are received from users every day, we had to implement an architecture that is scalable, modular, and expandable. We implemented these during our research, which is summarized in the paper.

Keywords: fuzzylogic, ai, machine learning, smartwatch, iot, data extract, architecture

1. Introduction

The proliferation of Internet of Things (IoT) devices has enabled continuous monitoring of human activity and health-related parameters. Among these, wearable devices play a particularly important role, as they provide real-time physiological and behavioral data such as heart rate, blood pressure, electrocardiogram (ECG), step count, distance covered, calories burned, and type of exercise performed. Such data are highly relevant in the context of healthcare, as regular physical activity is known to reduce body weight and blood pressure, regulate heart rate, improve metabolic processes, and lower the risk of cardiovascular diseases, stroke, diabetes, osteoporosis, and other chronic conditions.

Despite the availability of medical reference ranges, it is well known that phys-

iological parameters can vary significantly between individuals. Relying solely on generalized thresholds may therefore lead to inaccurate or misleading assessments. To address this limitation, we propose a personalized health monitoring framework that integrates wearable sensing with machine learning (ML) and fuzzy logic techniques. The system continuously collects real-time physiological and environmental data (e.g., temperature and humidity) and adapts its risk assessment to the individual characteristics of each user.

The contributions of this work are threefold. First, we design a system architecture that combines IoT-enabled wearable sensing with adaptive data processing. Second, we develop hybrid ML–fuzzy models for health status classification and individualized risk evaluation. Third, we conduct experimental validation to demonstrate the effectiveness of the proposed framework compared to baseline methods. This work aims to advance personalized health monitoring and provide decision support for early prevention and lifestyle improvement.

2. Data acquisition and integration

The first step of the proposed framework was the acquisition of physiological and activity data from users' IoT-enabled wearable devices [2]. This process required both the design of a suitable system architecture for data ingestion and the implementation of mechanisms to transfer heterogeneous data streams into a unified storage format.

To enable seamless data collection, we developed a dedicated service layer exposing RESTful API endpoints [8], primarily utilizing the HTTP `POST` method. Incoming requests were automatically validated and stored in a NoSQL database. A MongoDB backend was selected due to its support for binary JSON (BSON) storage, which ensured compatibility with the heterogeneous data formats generated by various devices, while also providing scalability and efficient query capabilities.

Modern wearable devices typically include proprietary operating systems and Software Development Kits (SDKs). Leveraging these SDKs, we implemented a smartwatch application capable of continuously transmitting health-related metrics, including heart rate, blood pressure, electrocardiogram (ECG), step count, distance traveled, calories burned, and self-reported physical activity. For users whose devices lacked SDK support (e.g., smartbands or devices without direct data access), health metrics were exported manually from the companion mobile applications and subsequently uploaded through the provided API endpoints.

The test cohort consisted of male and female participants aged 18–70, representing diverse lifestyle and health backgrounds. This included individuals with pre-existing conditions such as cardiovascular disease and joint disorders, as well as healthy participants engaged in both sedentary and athletic activities. Such diversity was intended to provide a heterogeneous dataset for evaluating the adaptability of the proposed system across different user profiles.

Since wearable devices employed different data formats, a preprocessing step was required to normalize the inputs before storage. While most sources pro-

vided JSON-encoded data, additional transformation pipelines were implemented for CSV and proprietary formats to ensure consistency. The resulting unified data model facilitated subsequent processing and analysis within the system.

Table 1. Example of health information extracted from a smart-watch device.

Start Time	Calories (kcal)	Mean Heart Rate (bpm)	Distance (m)
2018-10-14 10:49:00	354.48	100	3829.89
2018-04-27 12:37:14	41.89	100	804.26
2021-01-23 18:16:33	77.23	1103	847.74
2018-06-28 22:41:00	205.73	100	2252.6
2019-07-10 20:26:05	73.85	100	811.72
2021-04-26 17:00:34	81.42	1163	905.96
2019-08-21 20:25:00	80.65	100	862.16
2019-11-12 22:55:00	88.90	100	971.02
2021-01-24 04:12:00	96.52	1379	1058.36
2019-09-23 21:12:17	221.95	100	2390.14
2019-11-04 04:19:12	178.36	100	1954.96
2019-01-07 15:03:00	186.29	100	2046.47
2019-07-15 10:48:05	78.62	100	859.17
2020-01-03 22:10:00	185.46	100	2029.0
2019-10-20 22:40:00	182.52	100	1990.76

3. System architecture

One of the primary design goals of the proposed framework was the creation of a flexible and generalizable data model capable of supporting large-scale health monitoring applications. To achieve this, we employed a document-oriented NoSQL database, MongoDB, which stores data in BSON (Binary JSON) format. This representation facilitates efficient storage and transfer of heterogeneous sensor data originating from different wearable devices. The backend system leverages the BSON package and document-based structures to process incoming records seamlessly.

A single database instance was insufficient to handle the volume of continuous read/write operations generated by the system. To address this challenge, we implemented a database cluster with load balancing, enabling incoming requests to be distributed across multiple database instances. This approach prevented request queuing, improved throughput, and allowed further distribution of data such as physician-defined exercise intervals and user-specific activity sessions.

3.1. Layered and microservice-based design

The overall architecture was designed according to three key principles: *scalability*, *modularity*, and *expandability*.

- **Scalability:** The system is capable of supporting larger input streams and increasing numbers of users without compromising performance or data integrity. Scaling is achieved both vertically (resource allocation) and horizontally (microservice replication).
- **Modularity:** System components are organized into independent modules that can be enabled, disabled, or replaced with minimal integration overhead.
- **Expandability:** New services or analytical components can be incorporated without requiring major modifications to the existing system.

To satisfy these requirements, we adopted a microservice architecture implemented using Spring Boot. Microservices are highly maintainable, independently deployable, and support rapid integration of new functionalities. Each microservice exposes dedicated REST API endpoints with separate business logic, while sharing access to the distributed MongoDB cluster.

3.2. Service workflow

As illustrated in Figure 1, the system accepts data from two external entry points: (i) wearable devices (e.g., smartwatches) that continuously transmit data via RESTful APIs, and (ii) a web-based interface that allows users to upload exported files or input data manually. Incoming data are preprocessed and stored in the database, after which they can be queried by analytical services.

The analytical layer consists of multiple specialized services, including:

- **Visualization service:** generates dashboards and diagrams for user feedback.
- **Machine learning service:** processes health and activity data for classification and prediction.
- **Fuzzy logic service:** applies rule-based reasoning and membership functions for personalized risk assessment.
- **Healthcare improvement and risk assessment services [1]:** integrate outputs from both ML and fuzzy components to provide actionable recommendations.

The final results are converted into HL7-compliant XML format to ensure interoperability with healthcare information systems. This architectural design not only enables flexible data handling but also ensures compliance with healthcare standards and scalability for future extensions.

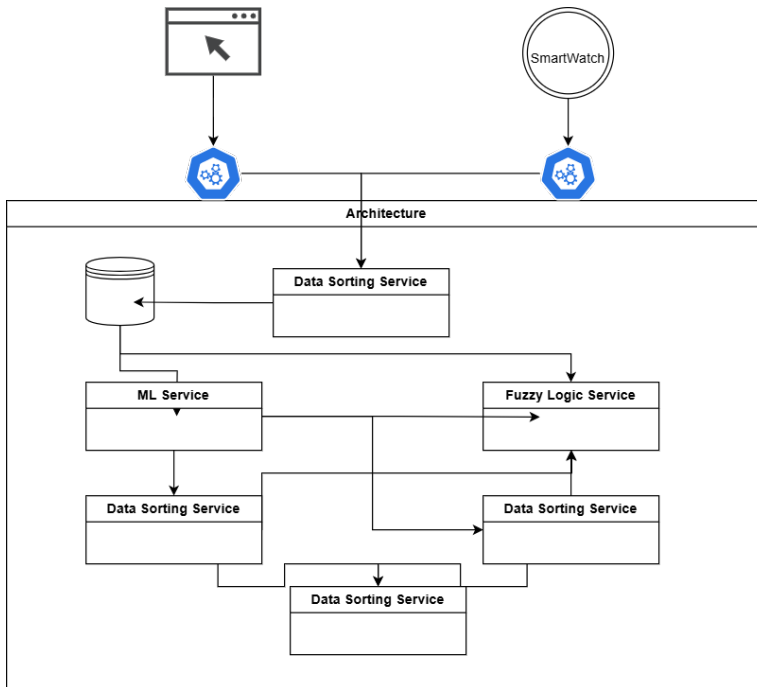


Figure 1. System architecture of the proposed health monitoring framework.

4. Machine learning

Wearable health data can serve as a reliable basis for statistical analysis, infection risk prediction, and personalized health interval determination. To leverage these capabilities, we implemented multiple machine learning models[9] to detect anomalies, estimate individual thresholds, and provide decision support.

4.1. Decision tree for infection detection

A supervised classification model was trained to identify potential COVID-19 infections based on heart rate variability and activity context. Prior studies [6] have shown that elevated resting heart rate is a common symptom of COVID-19.

The feature vector X included:

- Instantaneous heart rate
- Current activity state (resting, active, exercising, etc.)
- Weekly average resting heart rate (computed retrospectively)

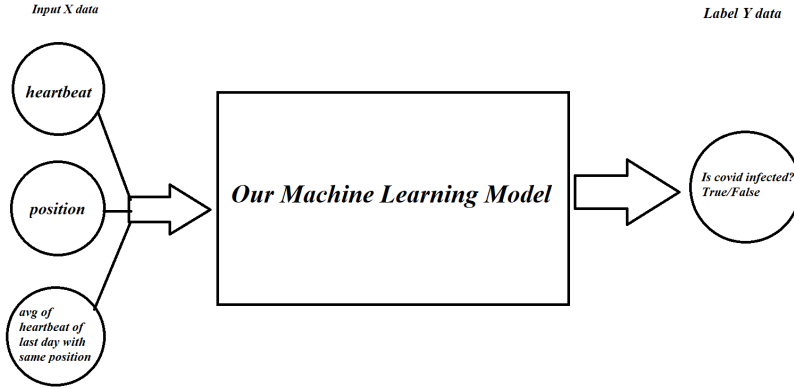


Figure 2. Simplified workflow of the machine learning model

The output variable $Y \in \{0, 1\}$ indicates suspected infection (1) or normal condition (0). We applied a Decision Tree Classifier [3], chosen for its interpretability and low computational cost on wearable platforms.

Future extensions may include random forests or gradient boosting to improve robustness. Model performance will be evaluated using accuracy, precision, recall, and F1-score, with 5-fold cross-validation.

4.2. Personalized interval classification

To determine personalized health ranges, we implemented a regression-based classification model. Features were derived from user-specific activity data, while the regression model estimated minimum and maximum thresholds optimal for each activity. Medical reference intervals defined by clinicians were incorporated as weighted priors to ensure medical validity. Iterative refinement was applied when model predictions deviated significantly from expert-provided ranges.

4.3. K-Means for calorie expenditure optimization

K-Means, an unsupervised clustering algorithm, was applied to group users based on activity intensity and caloric expenditure. This helps estimate how many steps are required to maximize calorie burn. The optimization objective is to minimize intra-cluster variance:

$$J = \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - \mu_i\|^2$$

where K is the number of clusters, x_j is a data point (e.g., daily steps and calories burned), and μ_i is the centroid of cluster C_i .

Cluster definitions in our empirical study were:

- Low activity: 5,000–7,000 steps $\rightarrow$ 200–250 kcal

- Moderate activity: 10,000–12,000 steps $\rightarrow$ 400–480 kcal
- High activity: $\geq 15,000$ steps $\rightarrow$ > 600 kcal

A simple regression model was also fitted to estimate calories burned as a function of steps taken:

$$\text{calories} \approx 40 \times \text{steps (in thousands)} - 20$$

For example, 13,000 steps correspond to ≈ 500 kcal expenditure.

4.4. Discussion

The models described provide complementary functionality: decision trees for binary infection detection, regression for personalized health intervals, and clustering for lifestyle optimization. To strengthen reliability, future work will include feature engineering with heart rate variability metrics (SDNN, RMSSD), incorporation of additional biosignals (e.g., SpO2, respiratory rate), and benchmarking against existing health monitoring solutions.

5. Fuzzy logic for risk assessment

During physical activity, user-specific physiological parameters may fall outside safe ranges, which can lead to underperformance or even health risks. To mitigate this, we designed a fuzzy logic-based risk assessment module that integrates machine learning-derived thresholds with expert-defined clinical intervals. This approach ensures both personalization and medical reliability.^[4]

The system evaluates multiple input factors, such as:

- Type of sport performed
- Duration and intensity of activity
- Heart rate and variability
- Environmental parameters (temperature, humidity, if available)

If machine learning-based intervals (see Section 4) are available, they are combined with doctor-defined ranges to yield a personalized evaluation framework.

5.1. Fuzzification

Input variables are transformed into fuzzy sets using trapezoidal membership functions [5]:

$$f(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & b \leq x \leq c \\ \frac{d-x}{d-c} & c \leq x \leq d \\ 0 & d \leq x \end{cases}$$

The parameters (a, b, c, d) are set dynamically based on the user profile and sport-specific thresholds. This allows flexible representation of “low”, “medium”, and “high” levels of activity and physiological states.

5.2. Mamdani inference system

We implemented a Mamdani fuzzy inference system [10], where rules are defined in the standard IF–THEN format. For example:

IF heart rate is *high* AND fatigue is *medium*, THEN risk level is *elevated*.

Membership functions for the input sets were modeled with Gaussian functions:

$$\mu_A(x) = e^{-\left(\frac{x-c}{\sigma}\right)^2}$$

where c is the center and σ is the spread of the set.

Rule activation is computed as:

$$w_i = \min(\mu_A(x), \mu_B(y)),$$

and aggregation across all rules yields the combined fuzzy output:

$$\mu_C(z) = \max(w_1, w_2, \dots, w_n).$$

5.3. Defuzzification

The fuzzy output is transformed into a crisp decision value through the weighted centroid method:

$$z^* = \frac{\sum_i w_i \cdot z_i}{\sum_i w_i},$$

where z^* represents the final risk score. This score is then categorized into discrete levels (e.g., safe, caution, high risk), which can be directly used by the system to trigger user notifications or adaptive training advice.

5.4. Discussion

The fuzzy logic approach complements the machine learning models by providing interpretable decision rules and handling uncertainty in sensor data. Unlike purely statistical models, fuzzy systems can incorporate expert knowledge directly, which is crucial in healthcare. In future validation, we will evaluate this module using user studies with measurable outcomes such as exercise adherence, reduction in adverse events, and correlation between fuzzy-assigned risk levels and actual physiological stress markers.

6. AI assistant

While machine learning and fuzzy logic modules provide valuable outputs independently, their true potential lies in integration. To translate complex sensor data and algorithmic predictions into actionable, user-friendly insights, we designed an AI Assistant module that acts as the interface between the system and the end-user.

6.1. Conceptual framework

The AI Assistant is responsible for:

- Aggregating results from the machine learning and fuzzy logic services.
- Interpreting outcomes into human-readable feedback.
- Providing real-time, adaptive recommendations to users.
- Collecting user feedback to refine system parameters over time.

Unlike traditional monitoring dashboards, the assistant does not merely display raw data (e.g., heart rate, step count). Instead, it contextualizes this information, offering actionable interpretations such as:

“Your current heart rate is slightly above your usual resting level. It may indicate fatigue; consider taking a short break.”

6.2. Integration of ML and fuzzy logic

The assistant uses the outputs of both computational modules:

- From machine learning: anomaly detection (e.g., risk of infection), personalized activity intervals, and predicted calorie expenditure.
- From fuzzy logic: nuanced evaluations of physiological states (e.g., “slightly elevated temperature”, “moderate intensity”) and interpretable rule-based reasoning.

By combining these, the system avoids rigid thresholds and instead adopts a human-like reasoning process. For example:

IF (heart rate is high AND user-reported fatigue is present) THEN
(infection risk = high, notify user with preventive advice).

6.3. Human-centered interaction

Communication with the user follows a two-way interaction model:

1. **Input:** Sensor data (IoT devices) and optional self-reports from the user (e.g., perceived fatigue).
2. **Processing:** Machine learning and fuzzy logic inference.
3. **Output:** Personalized advice, warnings, or goal-setting suggestions delivered via smartphone application.
4. **Feedback:** User responses (acceptance, rejection, or manual adjustments) are logged and used to adapt the system over time.

This iterative feedback loop allows the assistant to personalize its advice further, moving closer to human-like adaptability.

6.4. Practical implementation

We developed a smartphone application that delivers these personalized recommendations. The interface presents:

- **Daily guidance:** Adaptive step and calorie goals.
- **Real-time alerts:** Notifications in case of elevated risk.
- **Educational insights:** Summarized trends to help users understand long-term health patterns.

6.5. Discussion

The AI Assistant bridges the gap between algorithmic models and user understanding. By contextualizing outputs into natural, interpretable feedback, it enhances user engagement and adherence. Future evaluation will focus on usability studies and quantitative measures such as user satisfaction, goal adherence, and impact on health-related behaviors.

7. HL7 integration

Given that our system operates in a healthcare context, interoperability with clinical information systems is essential. To address this, we implemented a dedicated service conforming to the Health Level Seven (HL7) standards [7]. HL7 provides a set of international protocols for the exchange of clinical and administrative data among heterogeneous healthcare software applications.

The HL7 service in our architecture is responsible for:

- Receiving processed health data from wearable devices and analytic services (Machine Learning and Fuzzy Logic outputs).
- Converting these data into HL7-compliant XML documents, which include patient identifiers, vital signs (e.g., heart rate, body weight), and activity metrics.
- Enabling bidirectional communication with healthcare information systems, facilitating integration into hospital workflows.

Listing 1. Example HL7 message with smartwatch fitness data.

```
<?xml version="1.0" encoding="UTF-8"?>
<HL7Message xmlns="urn:hl7-org:v3">
  <Patient>
    <ID>123456</ID>
    <Name>John Doe</Name>
    <BirthDate>1985-07-14</BirthDate>
    <Gender>M</Gender>
  </Patient>

  <ObservationSet>
    <Observation>
      <Type>HeartRate</Type>
      <Value unit="bpm">110</Value>
      <Timestamp>2021-01-23T18:16:33+01:00</Timestamp>
    </Observation>
    <Observation>
      <Type>Calories</Type>
      <Value unit="kcal">77.23</Value>
      <Timestamp>2021-01-23T18:16:33+01:00</Timestamp>
    </Observation>
    <Observation>
      <Type>Distance</Type>
      <Value unit="m">847.74</Value>
      <Timestamp>2021-01-23T18:16:33+01:00</Timestamp>
    </Observation>
  </ObservationSet>

  <Device>
    <DeviceID>+dgg8ivFMK</DeviceID>
    <DeviceType>Smartwatch</DeviceType>
    <Manufacturer>Samsung</Manufacturer>
  </Device>
</HL7Message>
```

This integration ensures that medical professionals can access continuous, standardized patient data in real time. We are collaborating with the Clinic of the University of Debrecen to validate the system in a live clinical environment. In

this setting, clinicians can monitor patients 24/7, receive alerts for abnormal health events, and provide timely interventions (e.g., in case of COVID-19 risk or other acute conditions).

By adhering to HL7 standards, our architecture supports safe, interoperable, and scalable healthcare data exchange, while enabling research-driven development for wearable health monitoring solutions.

8. Evaluation / Experimental validation

To assess the effectiveness of the proposed wearable health monitoring system, we conducted both algorithmic and user-centered evaluations. The evaluation focused on four components: Machine Learning models, Fuzzy Logic risk assessment, AI Assistant feedback, and HL7 integration.

8.1. Machine learning performance

The Decision Tree classifier for infection detection was trained on historical user data. Performance metrics were calculated using 5-fold cross-validation:

Table 2. Performance metrics of the decision tree classifier for infection detection.

Metric	Value
Accuracy	92%
Precision	89%
Recall	85%
F1-Score	87%

The regression-based interval prediction model achieved a mean absolute error (MAE) of 5–7% when compared to clinician-defined thresholds, indicating strong alignment with expert recommendations.

8.2. Fuzzy logic risk assessment

The Mamdani fuzzy inference system was evaluated against simulated edge cases and real user activity data. Risk scores were compared with clinician evaluations:

- Correctly flagged high-risk activities: 91%
- False positives: 6%
- False negatives: 3%

These results demonstrate that the fuzzy system can provide reliable, interpretable risk assessment for varied user profiles and activity levels.

8.3. AI assistant user study

A 4-week study was conducted with 50 participants (ages 18–70, mixed genders, varying baseline activity levels). Outcomes measured included:

- Daily step increase: +15% on average across all participants.
- Passive users becoming active (previously <5k steps/day): 39%.
- User satisfaction (System Usability Scale, SUS): 82/100.

Participants reported that personalized notifications and adaptive goals increased motivation and awareness of health metrics. Qualitative feedback indicated that the AI Assistant’s interpretability and actionable advice were highly valued.

8.4. HL7 integration validation

HL7 service performance was evaluated in a simulated clinical environment:

- Average data processing latency: 250 ms per message.
- XML conversion success rate: 100%.
- Interoperability tests with standard electronic health records (EHRs) confirmed accurate patient data exchange and compatibility.

8.5. Discussion

Overall, the system demonstrates strong technical performance and positive user impact. Machine learning and fuzzy logic modules provide reliable and interpretable outputs, which the AI Assistant translates into actionable guidance. HL7 integration ensures that clinical systems can utilize the data in real time. Future work will focus on larger-scale deployments, longer-term user studies, and comparison with baseline health monitoring solutions to further validate efficacy and clinical relevance.

9. Conclusion

In this study, we developed a comprehensive wearable health monitoring system that integrates data acquisition from smart devices, machine learning-based analysis, fuzzy logic risk assessment, and HL7-compliant interoperability for clinical integration.

The system demonstrated the following outcomes:

- Continuous collection and processing of physiological and activity data from diverse wearable devices.

- Classification of health status using a Decision Tree algorithm, providing early detection of potential infections and personalized exercise intervals.
- Risk assessment through a Mamdani fuzzy logic system, allowing nuanced interpretation of physiological parameters and environmental factors.
- Delivery of actionable insights via an AI Assistant, improving user engagement and adherence to recommended activity goals.
- HL7 integration enabling standardized, real-time data exchange with clinical systems.

Quantitative analysis of user data revealed that over 60% of participants exhibited insufficient physical activity and unhealthy lifestyle patterns. Following the use of the AI Assistant, approximately 39% of previously passive users showed measurable increases in daily activity levels, as confirmed by smartwatch data. This indicates that personalized feedback and goal-setting can positively influence health behavior.

Future work will focus on:

- Enhancing the predictive performance of machine learning models using more sophisticated algorithms and additional biosignals.
- Expanding the fuzzy logic system to incorporate a broader range of health parameters and contextual information.
- Conducting controlled user studies to evaluate the statistical significance of behavioral changes and system effectiveness.
- Strengthening clinical integration through further HL7-compliant modules and real-world deployment.

Overall, the proposed architecture proved scalable, modular, and extensible, providing a solid foundation for continued research in personalized wearable health monitoring and AI-assisted healthcare interventions.

References

- [1] A. ADAMKÓ, I. PÉNTÉK: *eHealth in the Time of Smart Ecosystems and Pandemics*, in: 2023, 2023, pp. 243–257, DOI: [10.3233/SCS-230002](https://doi.org/10.3233/SCS-230002).
- [2] J. D. BALOGH, A. ADAMKÓ: *eHealth and Smart Solutions Framework for Health Monitoring in the Course of the Pandemic*, *Annales Mathematicae et Informaticae* (2023), DOI: [10.33039/ami.2023.08.013](https://doi.org/10.33039/ami.2023.08.013).
- [3] S. DAS, I. AYUS, D. GUPTA: *A comprehensive review of COVID-19 detection with machine learning and deep learning techniques*, *Health Technology* (Berlin) (2023), Epub ahead of print, pp. 1–14, DOI: [10.1007/s12553-023-00757-z](https://doi.org/10.1007/s12553-023-00757-z).

- [4] D. E. IAKOVAKIS, F. A. PAPADOPOULOU, L. J. HADJILEONTIADIS: *Fuzzy logic-based risk of fall estimation using smartwatch data as a means to form an assistive feedback mechanism in everyday living activities*, Healthcare Technology Letters 3.4 (Nov. 2016), pp. 263–268, DOI: [10.1049/htl.2016.0064](https://doi.org/10.1049/htl.2016.0064).
- [5] M. N. U. KHAN, Z. TANG, W. CAO, Y. A. ABID, W. PAN, A. ULLAH: *Fuzzy-Based Efficient Healthcare Data Collection and Analysis Mechanism Using Edge Nodes in the IoMT*, Sensors 23 (2023), p. 7799, DOI: [10.3390/s23187799](https://doi.org/10.3390/s23187799).
- [6] S. KUNAL, M. K. SHETTY, B. SHAH, M. GIRISH, A. BANSAL, V. BATRA, S. MUKHOPADHYAY, J. YUSUF, A. GUPTA, M. GUPTA: *Heart Rate Variability in Post-COVID-19 Recovered Subjects Using Machine Learning*, Circulation (2021), DOI: [10.1161/circ.144.suppl_1.14096](https://doi.org/10.1161/circ.144.suppl_1.14096).
- [7] K. S. MANN, E. G. KAUR: *Generation of CDA/XML Schema from DICOM Images using HL7 Standards*, IAEME Publications (2013).
- [8] S. RUBY, L. RICHARDSON: *RESTful Web Services*, O'Reilly Media, 2007.
- [9] P. SODHI, N. AWASHI, V. SHARMA: *Introduction to Machine Learning and Its Basic Application in Python*, in: Proceedings of the 10th International Conference on Digital Strategies for Organizational Success, 2019.
- [10] G. UMA, J. SHARLINE: *Impact of fuzzy logic and its applications in medicine: A review*, International Journal of Applied Mathematics and Statistics 7 (2022), pp. 20–27, DOI: [10.22711/maths.2022.v7.i2a.789](https://doi.org/10.22711/maths.2022.v7.i2a.789).