

# *ScopeSense*: An 8.5-month sport, nutrition, and lifestyle lifelogging dataset

Michael A. Riegler<sup>1,3</sup>, Vajira Thambawita<sup>1</sup>, Ayan Chatterjee<sup>1</sup>,  
Thu Nguyen<sup>1</sup>, Steven A. Hicks<sup>1</sup>, Vibeke Telle-Hansen<sup>2</sup>, Svein Arne Pettersen<sup>3</sup>,  
Dag Johansen<sup>3</sup>, Ramesh Jain<sup>4,1</sup>, and Pål Halvorsen<sup>1,2</sup>

<sup>1</sup> SimulaMet, Norway

<sup>2</sup> Oslo Metropolitan University, Norway

<sup>3</sup> UIT The Arctic University of Norway

<sup>4</sup> University of California Irvine, CA, USA

**Abstract.** Nowadays, most people have a smartphone that can track their everyday activities. Furthermore, a significant number of people wear advanced smartwatches to track several vital biomarkers in addition to activity data. However, it is still unclear how these data can actually be used to improve certain aspects of people’s lives. One of the key challenges is that the collected data is often massive and unstructured. Therefore, a link to other important information (e.g., when, what, and how much food was consumed) is required. It is widely believed that such detailed and structured longitudinal data about a person is essential to model and provide personalized and precise guidance. Despite the strong belief of researchers about the power of such a data-driven approach, respective datasets have been difficult to collect. In this study, we present a unique dataset from two individuals performing a structured data collection over eight and a half months. In addition to the sensor data, we collected their nutrition, training, and well-being data. The availability of nutrition data with many other important objectives and subjective longitudinal data streams may facilitate research related to food for a healthy lifestyle. Thus, we present a sport, nutrition, and lifestyle logging dataset called *ScopeSense* from two individuals and discuss its potential use. The dataset is fully open for researchers, and we consider this study as a potential starting point for developing methods to collect and create knowledge for a larger cohort of people.

## 1 Introduction

Data is the new gold, is a phrase commonly touted. In healthcare, data is critical as their proper collection and analysis can improve the quality of life and even save lives. Nowadays, people record their daily activities digitally through a significant number of devices and sensors. Such digital recording of a person’s activity and lifestyle is referred to as *lifelogging* [15]. Recording and analyzing such lifelog data provide a great opportunity for studying an individual’s life experience. It can help monitor a person’s activity to improve health and well-being [24], help recover memories of past events [31], or analyze social behaviour [21,6]. Lifelogs are also sources of vast rich data for interesting research.

For instance, Chokr et al. [4] described a machine learning approach to predict the number of calories from food images, and De Choudhury et al. [5] described the impact of interaction on social media to influence mental health. Therefore, lifelogs containing various types of information from a person’s daily lifestyle are valuable in numerous contexts.

Although lifelogs might contain data highly valuable for research, they are often not generally available to researchers. A lifelog is typically not stored centrally by a single service that can be easily accessed, but rather exists as the union of data stored in a large number of online and offline data silos [19]. Still, some datasets exist, and existing lifelogging datasets [18] usually contain a person’s daily life activities automatically captured and recorded using smartphone applications, wearable devices, and other sensors. One example is the NTCIR Lifelog test collection [14] consisting of lifelogging datasets for the NTCIR-12/13/14 lifelog tasks, which was first released at the NTCIR-12 conference [13]. The images in this dataset are captured by wearable cameras carried by two different lifeloggers. Some work has been done with similar datasets, for example, retrieving moment of interest [21,8]. However, a key challenge in lifelogging research is the poor availability of test collections [7]. Hence, there is a need for more available lifelog datasets, especially ones collected over longer time periods and with multiple modalities and purposes.

Capturing daily life events is also something many sports professionals do. Athletes have kept written training diaries for a long time using both pen-and-paper and, more recently, digital logging systems. Now, the use of wearables to measure activity and its intensity in both top sport and among the regular physically active population help to improve performance, recovery, and other aspects of health [9]. A challenge is to make sense of the data, and often, the captured data is limited to self-reports since activity logs from smartwatches and phones are hard to understand. Thus, there are still steps needed for integration of data [11] and to find standardized ways to analyze, evaluate, and present data [10]. Another problem in the area of sport is that professional athletes do not control the captured data by themselves, and they need the assistance of coaches, physicians, or technical support staff [19]. This process adds the burden of informed consent, authorization, and privacy. Furthermore, a trainer or team doctor does not have time to properly evaluate the myriads of sensor data from the athletes to possibly detect relevant performance data that could be used to improve training.

Moreover, dietary intake is an important factor affecting metabolic regulation, thereby affecting human health and fitness. However, individual response to dietary components has been known for a long time [22,23,29] where evidence has shown that people eating identical meals show high variability in metabolic response, such as blood glucose and lipids [33,25,2,16,12,27]. Hence, one set of dietary advice may not be equally efficient for all individuals. To improve healthy lifestyle and reduce the risk of diseases, we need to investigate individual variation using multi-dimensional and high-resolution continuous time-series of markers of individual health and fitness. Furthermore, we need to investigate

how various factors correlate. With new technology and analytical tools, we may be able to develop more personalized approaches, but to achieve this, access to relevant data is crucial.

To aid these efforts, automatic methods to analyze sensor data and the quantification of self-reports may play a crucial role in retrieving the information that people may need. To be able to perform these analyses with the increasing volume of data produced by different devices, new methods and tools are needed. *ScopeSense* is made available in an effort to enable the development of such support systems. We provide a starting point by combining the idea of lifelogging data collection with activity logging. Multiple activity-specific analyses can be performed on such data as predicting sports performance, weight loss, or gain, but there is a lack of available datasets. We have therefore logged several objectives and subjective parameters of a person’s daily life together with all food and drink intake. Specifically, we have used the following systems to capture the data:

- The Apple Watch series 6 smartwatch to track 24/7 activity and heart rate, and training sessions, electrocardiogram (ECG), sleep, blood oxygen, etc.
- The PMSys sports logging app to track subjective wellness parameters such as sleep, mood, stress, fatigue, muscle soreness; training load with session rating of perceived exertion (sRPE) and length; and injuries with body location and severity.
- The Lifesum food tracking app where all intake of food and drinks are self-reported and logged, together with a persons weight, and the content of nutrients are calculated.

*ScopeSense* contains logging data between 8. February 2021 and 20. October 2021 (255 days) from two individuals. To the best of our knowledge, *ScopeSense* is the only available dataset to combine both subjective and objective parameters combining daily lifestyle and activities over a long period of time. The dataset is openly available for research.

In the rest of the paper, we describe the data collection procedure in section 2, and in section 3, we describe the dataset in detail. Section 4 presents some initial analysis of the data. Furthermore, we provide examples of possible use-cases and applications of the dataset in section 5, before we conclude the paper in section 6.

## 2 Data collection

Based on experiences from previous datasets [30], we conjecture a need for even more details in a lifelog covering activity, wellness, and nutrition data, collected over longer time periods. In this respect, *ScopeSense* is a 255 days dataset containing the lifelogs of two individuals with objective biometrics and activity data, food and nutrition, and subjective wellness and exercise load. During the data collection phase, there was a goal of regular running and strength workout sessions. The participants gave written consent to publish and distribute the data. Moreover, the data collection has been reported to the Norwegian centre for



Fig. 1: Examples of logging using Apple Watch version 6.

research data (NSD), and been assessed and approved (reference #294827). Additionally, our application to the regional committees for medical and healthcare research ethics (REK) concluded that no approval from was needed for this data collection (reference #506192).

## 2.1 Objective biometrics and activity data

We used an Apple Watch (series 6, hardware version 6.2, software version 7.6.1), in combination with an iPhone, to collect objective biometric and activity data. This watch was chosen because it was equipped with the most available sensors in one device at that time. Most of the biometric data were recorded automatically by wearing the smartwatch 24/7; however, for some measurements, we manually started a recording using the various apps on the smartwatch (figure 1):

- All exercise and training data were logged using the Apple workout app. It mainly starts the type of exercise one specifies and ends the session when it is finished.
- An ECG recording was made once every day to measure the heart’s rhythm and electrical activity. This is a 30 second measurement holding in the clock button.
- Blood oxygen is measured by starting the app and allowing the sensors to work for 15 seconds.
- All other parameters are recorded automatically, like heart rate, sleep, number of hours where the person has been in a vertical position, flights (floors) climbed, counted steps walked/ran, energy used, etc., but one may use the apps to monitor your values.

All the data collected by the smartwatch is stored and extracted from the Apple Health application which works like a central data hub for the Apple devices.

## 2.2 Food and nutrition

To collect dietary intake and nutrition data, we used the Lifesum<sup>1</sup> mobile app on iPhones with a premium subscription. Basically, figure 2 shows the procedure used for collecting dietary intake data. The participants selected meal or snack

<sup>1</sup> <https://lifesum.com>

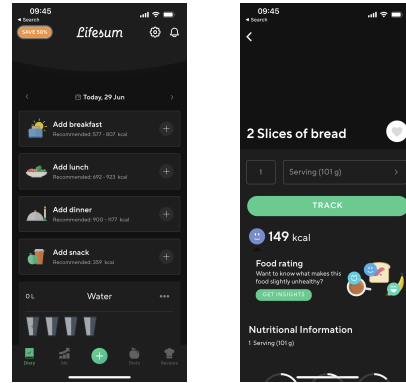
(figure 2a), and then, they inserted the portion of each components (figure 2b), either searching in the existing database or scanning the barcode printed on the packaging. Thereafter, they added the estimated amount (e.g., weight in grams). When known, the exact type of a component was inserted, like the type of bread, to get the exact types of nutrients. However, the exact type is sometimes unknown, e.g., when eating out. To quantify the amounts, we used the information provided on the packaging, e.g., while consuming a box of yogurt, we measured the amount of liquid a cup or

a glass can contain. Moreover, we weighted the amounts included in a portion of a particular meal, e.g., the weight of a slice of cheese or a spoon of jam. A challenge was when eating out where it is practically infeasible to estimate accurately the individual ingredients used in a dish and the total volume, and if so, an approximate number is reported. To ease the logging, the app supports making favorites and bookmarks of the ingredients, and also to make predefined meals and dishes. Here, we also logged the weight of the participants every day. Moreover, the app sent push-messages to remind participants to log both dietary intake and weight. Data from Lifesum is included in the Apple Health data, but also extracted the data in own comma separated value (CSV) files directly from the Lifesum system.

### 2.3 Subjective wellness, training load, and injuries

We used the PMSys system<sup>2</sup> to collect subjective data regarding wellness, training load, and injuries. Figure 3 depicts an example of normal reporting sequences. The PMSys system is an online sports logging system where athletes can monitor for example individual training load, daily subjective wellness parameters, and injuries [26,32]. Wellness has typically been reported once a day through a sequence of questionnaires, as shown in figure 3. The wellness data was reported in the morning.

Training load or Session Rating of Perceived Exertion (sRPE) is a metric calculated from the product of the session length and the reported Rating of Perceived Exertion (RPE), i.e., reported similarly as shown in figure 3. The perceived training load is reported after every training session. Finally, the injuries questionnaire is recommended completed once a week, regardless of having an injury or not, but here the participants mainly reported the injury when one occurred, where the participants press on a body part to indicate a minor or



(a) Select meal. (b) Type & amount.

Fig. 2: Logging food using Lifesum.

<sup>2</sup> <https://forzasys.com/pmSys.html>

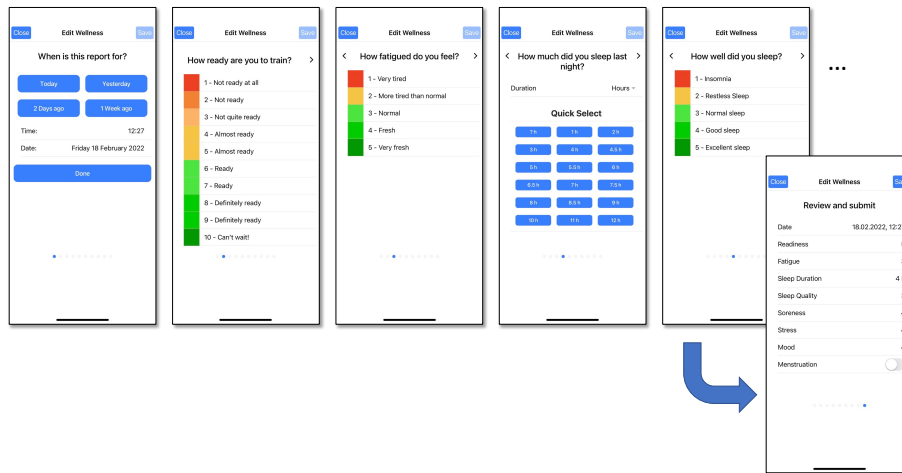


Fig. 3: Subjective parameters logged using PMSys, exemplified by wellness.

major injury or pain. To ensure timely reporting, PMSys sends scheduled push notifications directly to the participants' smartphones. All data was extracted at the end of the logging period from the system into CSV files.

#### 2.4 Data anonymization

To prevent any identity disclosure of the participants, the data has been processed to remove all ids. All occurrences of names in devices, like GPS coordinates, have been changed to a random value or removed completely to make it impossible to re-identify the participants. Exact GPS position data is also not important for the sport-related analysis besides features that can be extracted from the data, such as speed, distance, and elevation level which are part of the dataset.

### 3 Dataset overview

*ScopeSense* contains logging data from two male persons, in the age range of 40-50 years, between 8. February 2021 and 20. October 2021 (255 days) in Norway. The participants have followed no particular food regime and have been exercising regularly during the period. The dataset is organized according to the way it is collected, even though one could alternatively organize the data according to the type of data. We divided and organized data into three folders for each participant as shown in figure 4.

The dataset is fully available and open for free use for researchers at the well-known Open Science Framework (OSF)<sup>3</sup> and at the Simula dataset site<sup>4</sup>.

<sup>3</sup> <https://osf.io/v5acr/>

<sup>4</sup> <https://datasets.simula.no/scopesense>

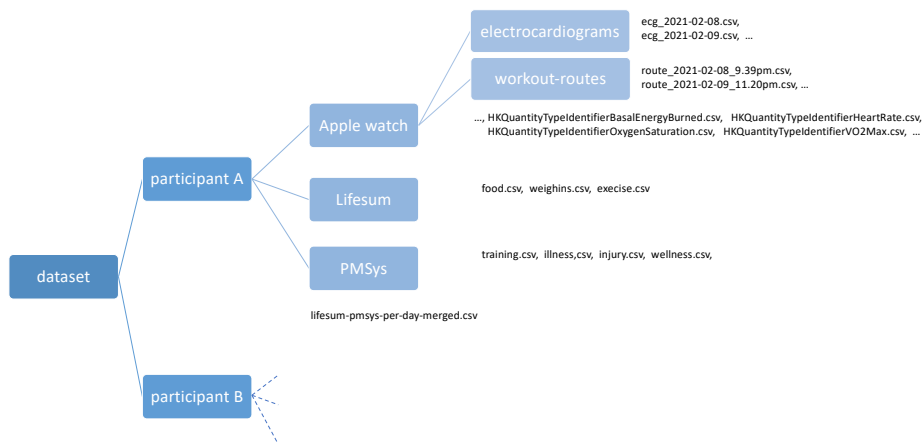


Fig. 4: Dataset organization. The structure for participants A and B is the same. Blue boxes represent the folder structure of the dataset.

The dataset is free to use for research and teaching purposes under the license Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)<sup>5</sup>.

### 3.1 Apple watch data

We have exported all data from the Apple health system. The raw data comes in XML-format<sup>6</sup>, but we have extracted the various parameters as CSV-files. There is one record for each of the measurements, where each record contains the parameter type, the source (which is anonymized), the type of device (hardware and software), the unit, dates, and the value(s). The Apple health system stores a large number of health parameters, including step count, heart rate, resting heart rate, and blood oxygen saturation. Other parameters include sleep, energy burned (basal/active), time standing, distance (walked/ran), and walking speed. The ECG data-series captured once a day is stored in its own sub-folder.

Each training session contains a training record in Apple health. Specifically, combining information from various files, one can extract information like type of training, duration, start and end time, total active energy (kcal), total distance (km), average heart rate and heart rate during training, step count, and speed. Each run session is also collected as a series of points along the route as shown in figure 5 giving, for example, time, speed, and elevation, but where the actual GPS coordinates are removed for privacy issues. All these routes are stored in a separate sub-folder.

### 3.2 Lifesum data

The Lifesum system collects as much as possible of consumed dietary (eating and drinking) information. Every report in the system is contained in one row of

<sup>5</sup> <https://creativecommons.org/licenses/by-nc/4.0/>

<sup>6</sup> <https://developer.apple.com/documentation/healthkit>

```

time,lon,lat,elevation,speed,course,hAcc,vAcc
...
2021-02-19T17:15:15Z,43.448116,94.243819,224.995163,1.859252,231.84137,0.851156,0.798664
2021-02-19T17:15:16Z,43.448125,94.243802,225.067017,1.99339,230.884521,0.860994,0.803635
2021-02-19T17:15:17Z,43.448133,94.243784,225.131119,2.111444,232.588455,0.874346,0.810753
2021-02-19T17:15:18Z,43.44814,94.243765,225.185623,2.213889,232.957123,0.891013,0.81974
...

```

Fig. 5: Example records a run route session.

the table with the date of the report, which meal it is connected to (or a snack), the “name” of the food (including a title, potential manufacturer, and content), and the amount (both for the used metric and in grams). Subsequently, as can be viewed in the last part of each row, the reported intake is used to calculate the number of calories, various types of carbohydrates, various types of fat, protein, sodium, potassium, and cholesterol.

In addition, the Lifesum system has logged the participant’s daily weight. The participants’ weight is given in each row, and also the calculated body mass index (BMI) based on the participants’ height.

Finally, we have also merged Lifesum and PMSys data (see below) in a day-by-day manner. These records are stored in the `lifesum-pmsys-per-day-merged.csv` file which contains one row per day.

### 3.3 PMSys data

In terms of subjective PMSys reporting, the raw data is contained in the CSV files:

- Wellness includes parameters like time and date, fatigue, mood, readiness, sleep duration (number of hours), sleep quality, soreness (and soreness area), and stress. Fatigue, sleep quality, soreness, stress, and mood all have a 1-5 scale. Score 3 is normal, scores 1-2 are below normal, and 4-5 are scores above normal. Sleep length is just a measure of how long the sleep was in hours, and readiness (scale 1-10) is an overall subjective measure of how ready you are to exercise, i.e., 1 means not ready at all, and 10 indicates that you are exceptionally ready. In elite sport, readiness is used to tune if the athlete should push or pull the training load. In the CSV file, there are columns for each wellness parameter, which contain columns for the date and for the wellness parameter value for each of the participants.
- A training load report contains a training session’s time, type of activity, perceived exertion (RPE), and duration in the number of minutes. This is, for example, used to calculate the session’s training load or sRPE ( $RPE \times \text{duration}$ ). The data is stored as one tab in the spreadsheet for each of the participants. There is one line for each session with the date, daily summarized load, the calculated session RPE (sRPE), the experienced RPE, the session length, and then various calculated metrics like weekly load, monotony, strain, acute chronic workload ration, and chronic training load (over 28 and 42 days). If more than one session per day has been logged, the additional lines will have just the sRPE, RPE, and length of the session, which are then used to calculate the total load parameters in the first line.
- Injury and illness are reported in separate files having one line per incident with date and symptom/place of pain.



Readiness	Mood	Fatigue	Stress	Sleep Quality	Sleep Duration	Soreness	Daily Load
calories burned, fatigue, soreness, mood	fat, stress, sleep duration, readiness	potassium, soreness, stress, sleep quality, readiness	carbs, fatigue, mood, sleep quality	bmi, proteins, fatigue, mood, stress, sleep duration	soreness, mood, stress, sleep quality	weight, fatigue, readiness	weight, bmi, calories burned, carbs (fiber), fat unsaturated, proteins, fatigue, soreness, readiness

Table 1: Important features for the different self reported values in PMSys.

## 4 Initial experiments

We provided some simple baseline experiments to provide an initial idea about how the dataset can be used and test its usefulness. For all experiments, we used 60% of the data for training and validation, and the remaining 40% as a test dataset. We preprocessed the data into a-value-per-day records (also included in the dataset) and used a subset of the data (weight\_kg, bmi, calories\_burned per day, calories per day, carbs per day, carbs\_fiber per day, carbs\_sugar per day, fat per day, fat\_saturated per day, fat\_unsaturated per day, potassium per day, protein per day, sodium per day, daily load, fatigue, soreness, mood, stress, sleep quality, sleep duration and readiness) where one instance represents one day of the collection period. The split between the train and test datasets was performed randomly and an equal number of instances per participant was included in the splits.

For the first experiment, we explored which features are important to predict different self-reported values. Specifically, we explored the important features for Readiness, Mood, Fatigue, Stress, Sleep Quality, Sleep Duration, Soreness, and Daily Load. We used Correlation-based Feature Subset Selection with Best First Ranker to select and rank the features. Table 1 shows the results where we observe that different features affect different aspects of ones well-being. For example, we can see that potassium is an important feature to predict fatigue, which is reasonable as a symptom is to feel tired/fatigue if a person has a too low level of this chemical element [28]. This shows that the dataset holds the potential to discover different aspects between self-reported and measured values.

For the second experiment, we trained a simple machine learning model to predict readiness (similar to [32,20]) which is seen as one of the most important factors in a sport context. As a baseline, we used ZeroR which predicts the average of the data. In addition, we trained two regressors using RandomForest and the SMOReg support vector machine. For ZeroR, the mean absolute error and root mean squared error were 1.29 and 1.6103, respectively, compared to 0.9944 and 1.3599 for SMOReg, and 0.9897 and 1.298 for RandomForest. From this initial simple analysis, we observe that some of the subjective measurements can be predicted using just a subset of features.

## 5 Example applications of the dataset

Healthcare systems are undergoing a major transformation. It used to be reactive and symptom-based, where a doctor primarily was consulted when being ill. Doctors diagnosed the sickness followed by proper medical treatment. Recently,

system biologists and healthcare researchers started adopting a different perspective. This perspective was introduced as a P4 approach to medicine [1,17]. The P4 approach is based on a predictive, preventive, personalized, and participatory approach that uses data about a person as the main driver in devising an approach that is not reactive but proactive. This approach requires that each person is considered a unique system and is modeled using longitudinal data. This approach was not practical until smartphones, wearable devices, and associated advances in machine learning and cloud computing came along. Therefore, ubiquitous health monitoring is required for seamless data collection and subsequent applications of artificial intelligence and advanced healthcare technologies.

In medicine, an emerging concept is N-of-1 approach, where longitudinal data about a person is used to model the person rather than collecting population data and considering a person a sample of this population. P4 approaches for all aspects of health and wellness require collecting objective as well as subjective data. As discussed in this paper, some of the data is relatively easy to collect, while others require careful planning, collection, and structured organization. Such an approach is inherently multimodal and requires applications of traditional multimedia systems competence in addition to emerging and novel machine learning approaches.

Valid, privacy-preserved, relevant, and accurate data is fundamental to this approach. In this paper, we presented our initial experience with relatively invasive data collection with data collected over a long period of time. This paper presents this data for researchers with different interests to explore approaches that may be suitable. We hope that this data is the beginning of a collection of such data and sharing in the community to enable exploration of approaches related to health. We want to emphasize the role of food data as a very effective data stream. Food containing essential nutrients has always been considered vital for a healthy human being, and the relation between food and exercise has been investigated for a long time showing the importance of eating correctly to perform best [3]. Yet, approaches for understanding the long-time effect of what, when, where, and how much food affects different aspects of physical and psychological health are only partly explored due to the unavailability of data. Just exploring if and how food intake affects physical activity or exercise patterns can give us simple, useful, and interesting information, especially on an individual basis. Moreover, subjective readiness and other wellness parameters can be validated for “real” coinciding changes against objective heart rate variability data from the Apple watch. We hope that the dataset presented in this paper will jump-start this process, and possibly enable insights into how food, readiness for exercise, execution of exercise, and general wellness are related. Moreover, since the data is collected over a long period of time, it’s worth analyzing it locally in shorter time intervals to consider if changes in some features, such as dining preferences and activity, may change between seasons with a potential causal effect on wellness.

## 6 Conclusion

We have presented the *ScopeSense* dataset, containing both objective and subjective longitudinal parameters from activity, wellness, food, and biometrics, potentially enabling the development of several interesting analysis applications. Our initial experiments show that such analyses are possible, but we conjecture the dataset has greater potential beyond what we have demonstrated in this paper. It is for our peer colleagues to use to expand and support this line of technology support for proactive medicine.

We are using our initial experiences with this dataset collection to currently expand further work to include a larger and more diverse cohort with even more parameters monitored. Also, we are developing a wide collection of machine learning applications that can sift through and analyze the data while reporting to enable next-generation proactive medicine to handle massive amounts of heterogeneous data in close to real-time. The goal is to have a personalized digital health screening service analyzing data and, e.g., detecting anomalies and concerning deviations providing potential for rapid response and intervention.

## References

1. Auffray, C., Charron, D., Hood, L.: Predictive, preventive, personalized and participatory medicine: back to the future. *Genome Medicine* **2**, article 57 (2010)
2. Berry, S.E., Valdes, A.M., et al.: Human postprandial responses to food and potential for precision nutrition. *Nature Medicine* **26**(6), 964–973 (2020)
3. Burke, L.M., Cox, G.R.: *The complete guide to food for sports performance: Peak nutrition for your sport*. Allen & Unwin (2010)
4. Chokr, M., Elbassuoni, S.: Calories prediction from food images. In: *Proc. of the 29th Innovative Applications of Artificial Intelligence (IAAI) Conference* (2017)
5. De Choudhury, M., Kiciman, E., et al.: Discovering shifts to suicidal ideation from mental health content in social media. In: *Proc. of the Conference on Human Factors in Computing Systems (CHI)*. pp. 2098–2110 (2016)
6. Dinh, T.D., Nguyen, D.H., Tran, M.T.: Social relation trait discovery from visual lifelog data with facial multi-attribute framework. In: *Proc. of the International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. pp. 665–674 (2018)
7. Duane, A., Gurrin, C.: User interaction for visual lifelog retrieval in a virtual environment. In: *Multimedia Modeling. LNCS*, vol. 11295, pp. 239–250. Springer (2019)
8. Duane, A., Gurrin, C.: Baseline analysis of a conventional and virtual reality lifelog retrieval system. In: *MultiMedia Modeling. LNCS*, vol. 11962, pp. 412–423. Springer (2020)
9. Düking, P., Achtzehn, S., et al.: Integrated framework of load monitoring by a combination of smartphone applications, wearables and point-of-care testing provides feedback that allows individual responsive adjustments to activities of daily living. *Sensors* **18**(5) (2018)
10. Düking, P., Fuss, F.K., et al.: Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. *JMIR mHealth and uHealth* **6**(4), e102 (Apr 2018)
11. Düking, P., Stammel, C., et al.: Necessary steps to accelerate the integration of wearable sensors into recreation and competitive sports. *Current sports medicine reports* **17**(6), 178–182 (2018)

12. Gaundal, L., Myhrstad, M.C.W., et al.: Beneficial effect on serum cholesterol levels, but not glycaemic regulation, after replacing sfa with pufa for 3 d: a randomised crossover trial. *British Journal of Nutrition* **125**(8), 915–925 (2021)
13. Gurrin, C., Joho, H., et al.: NTCIR Lifelog: The first test collection for lifelog research. In: Proc. of the International ACM SIGIR conference on Research and Development in Information Retrieval. pp. 705–708 (2016)
14. Gurrin, C., Joho, H., et al.: Overview of NTCIR-13 Lifelog-2 task. In: Proc. of the 13th NTCIR Conference on Evaluation of Information Access Technologies. NTCIR (2017)
15. Gurrin, C., Smeaton, A.F., Doherty, A.R.: Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval* **8**(1), 1–125 (Jun 2014)
16. Hansson, P., Holven, K.B., et al.: Meals with similar fat content from different dairy products induce different postprandial triglyceride responses in healthy adults: A randomized controlled cross-over trial. *The Journal of Nutrition* **149**(3), 422–431 (2019)
17. Hood, L., Friend, S.H.: Predictive, personalized, preventive, participatory (p4) cancer medicine. *Nature Reviews Clinical Oncology* **8**(3), 184–7 (2011)
18. Ionescu, B., Müller, H., et al.: ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. LNCS, vol. 11696, pp. 358–386. Springer (2019)
19. Johansen, H., Gurrin, C., Johansen, D.: Towards consent-based lifelogging in sport analytic. In: Multimedia Modeling. LNCS, vol. 8936, pp. 335–344. Springer (2015)
20. Kulakou, S., Ragab, N., et al.: Exploration of different time series models for soccer athlete performance prediction. *Engineering Proceedings* **18**(1) (2022)
21. Le, N.K., Nguyen, D.H., et al.: HCMUS at the NTCIR-14 Lifelog-3 Task. In: Proc. of the NTCIR Conference on Evaluation of Information Access Technologies. pp. 48–60 (2019)
22. Miller, J.Z., Weinberger, M.H., et al.: Blood pressure response to dietary sodium restriction in healthy normotensive children. *The American Journal of Clinical Nutrition* **47**(1), 113–119 (1988)
23. Morris, C., O’Grada, C., et al.: Identification of differential responses to an oral glucose tolerance test in healthy adults. *PLoS ONE* **8**(8), e72890 (2013)
24. Nag, N., Jain, R.: A navigational approach to health: Actionable guidance for improved quality of life. *Computer* **52**(4), 12–20 (Apr 2019)
25. Ordovas, J.M., Ferguson, L.R., et al.: Personalised nutrition and health. *BMJ* **361** (2018)
26. Pettersen, S.A., Johansen, H.D., et al.: Quantified soccer using positional data: A case study. *Frontiers in Physiology* **9** (2018)
27. Retterstøl, K., Svendsen, M., et al.: Effect of low carbohydrate high fat diet on ldl cholesterol and gene expression in normal-weight, young adults: A randomized controlled study. *Atherosclerosis* **279**, 52–61 (2018)
28. Sjøgaard, G.: Potassium and fatigue: the pros and cons. *Acta Physiologica Scandinavica* **156**(3), 257–264 (1996)
29. Sweeney, J.S.: Dietary factors that influence the dextrose tolerance test: a preliminary study. *Archives of Internal Medicine* **40**(6), 818–830 (1927)
30. Thambawita, V., Hicks, S.A., et al.: Pmdata: A sports logging dataset. In: Proc. of the ACM Multimedia Systems Conference (MMSys). p. 231–236 (2020). <https://doi.org/10.1145/3339825.3394926>
31. Truong, T.D., Dinh-Duy, T., et al.: Lifelogging retrieval based on semantic concepts fusion. In: Proceedings of the ACM Workshop on The Lifelog Search Challenge (LSC). pp. 24–29 (2018)
32. Wiik, T., Johansen, H.D., et al.: Predicting peek readiness-to-train of soccer players using long short-term memory recurrent neural networks. In: Proc. of the IEEE International Conference on Content-Based Multimedia Indexing (CBMI). pp. 1–6 (2019)
33. Zeevi, D., Korem, T., et al.: Personalized nutrition by prediction of glycemic responses. *Cell* **13**(5), 1079–1094 (2015)