PMDATA: A Sports Logging Dataset

Vajira Thambawita*
Steven Alexander Hicks*
Hanna Borgli†
Håkon Kvale Stensland†
Debesh Jha‡
Martin Kristoffer Svensen†
SimulaMet
Norway

Svein-Arne Pettersen
Dag Johansen
Håvard Dagenborg Johansen
Susann Dahl Pettersen
Simon Nordvang
Sigurd Pedersen
Anders Gjerdrum
UiT The Arctic University of Norway
Norway

Tor-Morten Grønli
Per Morten Fredriksen
Ragnhild Eg
Kjeld Hansen
Siri Fagernes
Christine Claudi
Andreas Biørn-Hansen
Kristiania University College
Norway

Duc Tien Dang Nguyen
University of Bergen
Norway

Tomas Kupka
Forzasys AS
Norway

Hugo Lewi Hammer§
OsloMet
Norway

Ramesh Jain
University of California, Irvine
US

Michael Alexander Riegler
SimulaMet
Norway

Pål Halvorsen*
SimulaMet
Norway

ABSTRACT

In this paper, we present PMDATA: a dataset that combines traditional lifelogging data with sports-activity data. Our dataset enables the development of novel data analysis and machine-learning applications where, for instance, additional sports data is used to predict and analyze everyday developments, like a person's weight and sleep patterns; and applications where traditional lifelog data is used in a sports context to predict athletes' performance. PMDATA combines input from Fitbit Versa 2 smartwatch wristbands, the PMSys sports logging smartphone application, and Google forms. Logging data has been collected from 16 persons for five months. Our initial experiments show that novel analyses are possible, but there is still room for improvement.

CCS CONCEPTS

• **Applied computing** → *Health informatics*; • **Computing methodologies** → *Machine learning*.

KEYWORDS

Multimedia Dataset, Neural Networks, Machine Learning, Sports Logging, Sensor Data, Questionnaires, Food Pictures

*Also affiliated with Oslo Metropolitan University, Norway
†Also affiliated with University of Oslo, Norway
‡Also affiliated with UiT The Arctic University of Norway
§Also affiliated with SimulaMet, Norway

MMSys'20, June 8–11, 2020, Istanbul, Turkey

ACM ISBN 978-1-4503-6845-2/20/06.
https://doi.org/10.1145/3339825.3394926

ACM Reference Format:
Vajira Thambawita, Steven Alexander Hicks, Hanna Borgli, Håkon Kvale Stensland, Debesh Jha, Martin Kristoffer Svensen, Svein-Arne Pettersen, Dag Johansen, Håvard Dagenborg Johansen, Susann Dahl Pettersen, Simon Nordvang, Sigurd Pedersen, Anders Gjerdrum, Tor-Morten Grønli, Per Morten Fredriksen, Ragnhild Eg, Kjeld Hansen, Siri Fagernes, Christine Claudi, Andreas Biørn-Hansen, Duc Tien Dang Nguyen, Tomas Kupka, Hugo Lewi Hammer, Ramesh Jain, Michael Alexander Riegler, and Pål Halvorsen. 2020. PMDATA: A Sports Logging Dataset. In *11th ACM Multimedia Systems Conference (MMSys'20), June 8–11, 2020, Istanbul, Turkey*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3339825.3394926

1 INTRODUCTION

In one way or another, many people are recording parts of their lives digitally. This could, for example, be through sensors in a smartwatch, GPS location tracking in smartphones, pictures from highly portable cameras, or through activities on various online social media services. It is not uncommon to see people posting pictures of their food on platforms such as Instagram or boasting about their workouts on Facebook as the events unfold.

The activity of recording one's life digitally, through various input sources, is often referred to as *lifelogging* [11], and a person who engages consciously in such activities is referred to as a *lifelogger*. Recording and analyzing lifelog data is a great opportunity for studying an individual's life experience. It can help monitor a person's activity to improve health [17], help recover memories of past events [19], or analyze social behaviour [3, 15]. From a multimedia perspective, lifelogs are sources of wast rich data for interesting research. For instance, Chokr and Elbassuoni [1] describe a machine-learning approach for predicting the number of calories from pictures of food, and De Choudhury et al. [2] describe how interaction on social media influence our mental health.

Although lifelogs might contain data highly valuable for research, they are often not available to the researchers. A lifelog is typically

not stored centrally in one single service that can be tapped into, but rather exists as the union of data stored in a large number of online and offline data silos [13]. Still, some datasets exist, and existing lifelogging datasets [12] 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 [10] consisting of lifelogging datasets for the NTCIR-12/13/14 lifelog tasks, which was first released at the NTCIR-12 conference [9]. 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 [5, 15]. However, a key challenge in lifelogging research is the poor availability of test collections [4]. Hence, there is a need for more available lifelog datasets.

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 using 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 [6]. 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 [8] and to find standardized ways to analyze, evaluate, and present data [7]. 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 support staff [13]. This process adds the burden of informed consent, authorization, and privacy. Moreover, a trainer or team doctor does not have time to look at the myriads of sensor data from the athletes to possibly find something that could be used to improve training. Using PMData, we have launched a competition task in ImageCLEF/LifeCLEF[1], where the goal is to predict the participants' weight and run performance at the end of the data collection period.

To aid these efforts, automatic methods to analyze sensor data and the quantification of self-reports will play an important role in retrieving the information that sports athletes may need. To be able to perform these analyses with the increasing volume of data coming from different devices, new methods and tools are needed. PMData is made available in an effort towards enabling development such support systems. We provide a starting point by combining the idea of lifelogging data collection with sports activity logging. Multiple sport-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 objective parameters like heart rate, sleep, calorie consumption, movement distance, activity sessions, weight, and subjective parameters of wellness, training load, injuries, food, and drink intake. We have used the Fitbit Versa 2 smartwatch[2], the PMSys sports logging app,[3] and Google forms for the data collection. For now, the dataset, named PMData, contains logging data for three months from 16 persons. To the best of our knowledge, PMData is the first available dataset to combine both subjective and objective parameters combining both daily life and sports activities.

In the following, we describe the procedure for collecting data and describe the dataset in detail. Furthermore, we present a preliminary experiment using machine learning to predict the possibility of a person gaining, losing, or keeping the current weight from logging. We also provide possible research questions and applications of the dataset.

2 DATA COLLECTION

The goal of PMData has been to gather lifelog data related to the activities of our participants, but without being too invasive. We planned to collect data from the end of November 2019 to the end of March 2020. We log data about the participant's daily activities, similar to a sport lifelog, and encourage them to exercise at least twice a week. We did not set any restrictions or requirements on the type or duration of the exercise participants can engage in.

2.1 Fitbit Versa 2: Objective Biometrics and Activity Data

To log objective biometrics and activity data, we used the Fitbit Versa 2 fitness smartwatch (see Figure 1). Each participant was encouraged to wear the watch as much as possible, also when sleeping. All settings were set to default, i.e., sleep tracking in normal mode and auto-exercise recognition on for all activities longer than 15 minutes. When training, participants were told to log in using the exercise menu option in the watch (e.g., run or treadmill).

Figure 1: Fitbit Versa 2

2.2 PMSys: Subjective Wellness, Training Load, and Injuries

Subjective assessments of each participant's wellness, training load, and injuries have been logged using the PM Reporter Pro smartphone application[4] where Figure 2 shows an example of a reporting sequence. PM Reporter is part of the PMSys online sports logging system that enables athletes to monitor individual training load, daily subjective wellness parameters, and injuries [20]. Wellness is reported typically once a day through a sequence of questionnaires. 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). The 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, where the participants press on a body part to indicate a minor or major injury or pain. To increase the reporting rate, PMSys sends scheduled push messages directly to the participants' smartphones, reminding them to report.

[1] https://www.imageclef.org/2020/lifelog
[2] https://www.fitbit.com/no/versa
[3] https://forzasys.com/pmsys.html
[4] https://bitbucket.org/corporesano/pm-reporter

Figure 2: Entering wellness data into PMSys

2.3 Google Forms: Demographics, Food, Drinking, and Weight

A Google Form questionnaire was used to collect information about food intake and weight development. Every day, the participants were asked to report eaten meals (breakfast, lunch, dinner, evening), the number of glasses of fluid (water, coffee, milk, juice, soda, etc.) that they consumed. They were also asked about their weight and whether they have consumed alcohol or not. To increase the reporting rate, we used the PMSys push-messaging system to send reminders to the participants' smartphones. A one-time questionnaire (see dataset home page) was used to ask for age, gender, height, and whether the person has a Type A or Type B personality [14]. Most participants regard themselves as having a Type A personality, and generally wakes up early (potentially also goes to bed early), rather than one who wakes up late (Type B).

2.4 Food Images

The reports on eaten meals collected using the Google Forms questionnaire indicate how often and regularly a person consumes food, but leaves out important details about their content, like nutrients and calories. Therefore, selected participants were asked to take photos of everything they have been eating or drinking using their smartphones. This is a time-consuming task and hard to remember activity, i.e., severely influencing the daily behavior of the participants. The collection period is therefore limited to two months.

3 DATASET DETAILS

PMDATA contains data collected from 16 persons: twelve men and three women, in the age range 25–60 years, with an average age of 34 years. The reporting period is from the start of November 2019 to the end of March 2020. The participants range from a broad background with regards to training and exercises. Some are active athletes, some previous athletes, and some rarely exercised at all.

An overview of the participants' demographic information is provided in the *participant-overview.xlsx* file where information like age, height, gender, measured max heart rate, test run results, and walk and run stride lengths are included. Furthermore, there is a directory per participant that contains the data from the Fitbit, PMSys, Google Forms, and Food image data sources. An overview of the dataset ontology can be found in Figure 3. Statistics about the Fitbit JSON-files can be found in Table 1 and statistics about the CSV-Files can be found in Figure 4. Note that all files have timestamps that must be used to connect the data from the different files.

Categories	File	Rate of entries	Number of entries
Calories	calories.json	Per minute.	3377529
Steps	steps.json	Per minute.	1534705
Distance	distance.json	Per minute.	1534705
Sleep	sleep.json	When it happens (usually daily).	2064
Lightly active minutes	lightly_ active_minutes.json	Per day.	2244
Moderately active minutes	moderately_ active_minutes.json	Per day.	2396
Very active minutes	very_ active_minutes.json	Per day.	2396
Sedentary minutes	sedentary_ minutes.json	Per day.	2396
Heart rate	heart_rate.json	Per 5 seconds.	20991392
Time in heart rate zones	time_in_heart_rate_ zones.json	Per day.	2178
Resting Heart Rate	resting_ heart_rate.json	Per day.	1803
Excercise	excercise.json	When it happens. 100 entries per file.	2440
Sleep Score	sleep_score.csv	When it happens (usually daily).	1836
Google Forms reporting	reporting.csv	Per day.	1569
Wellness	wellness.csv	Per day.	1747
Injury	injury.csv	Per week.	225
SRPE	srpe.csv	Per exercise.	783

Fitbit Google Forms PMSys

Figure 3: Overview of the dataset.

All participants have been informed about the collection and publication of the data related to this project and signed a form consenting to this. The Norwegian Centre for Research Data (NSD) has evaluated the project and found it to be in accordance with Norwegian and EU data protection laws.

The dataset is available at the Open Science Framework (OSF) at the following URL: https://osf.io/vx4bk/; or at the Simula datasets site: https://datasets.simula.no/pmdata/. The dataset is free to use for research and teaching purposes under the license Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).[5]

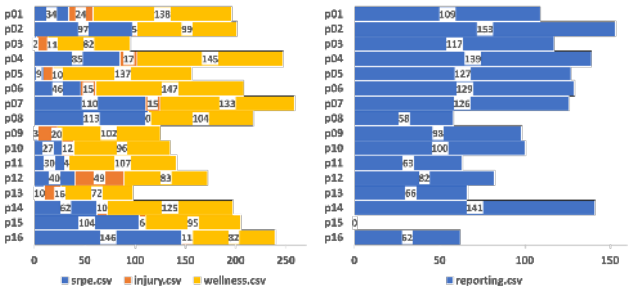
3.1 Fitbit

The data from the Fitbit Versa 2 smartwatch has been extracted into CSV and JSON files. The *fitbit* directory contains the following files:

[5]https://creativecommons.org/licenses/by-nc/4.0/

Table 1: Number of Fitbit entries for each participant.

participant	very and moderately active minutes	sleep	sleep score	calories	heart rate	steps and distance	sedentary minutes	exercise	light active minutes	time in heart rate zones	resting heart rate
P01	152	155	150	218880	1573165	218836	152	190	152	152	152
P02	152	158	138	218880	1472629	107326	152	324	152	148	91
P03	152	84	74	218880	808341	53042	152	57	152	117	152
P04	152	188	140	218473	1571315	86457	152	161	152	146	35
P05	152	133	117	218880	1370967	111231	152	145	152	145	95
P06	152	165	147	218880	1579882	117780	152	161	152	152	152
P07	148	161	140	212816	1581947	108048	148	176	148	147	148
P08	143	143	132	205920	1613326	100451	143	261	143	139	143
P09	152	142	132	218880	1305520	85271	152	54	152	150	152
P10	148	103	98	213120	1083257	75427	148	140	148	114	148
P11	152	128	119	218880	1383149	92982	152	96	152	123	98
P12	152	8	1	218880	801264	83752	152	93	0	134	0
P13	152	57	47	218880	634746	48629	152	50	152	80	0
P14	140	138	115	129600	1251156	68703	140	270	140	135	140
P15	145	148	140	208800	1563024	98198	145	243	145	144	145
P16	152	153	146	218880	1397704	78572	152	19	152	152	152
Mean	150	129	115	211096	1311962	95919	150	153	150	136	129
All	2396	2064	1836	3377529	20991392	1534705	2396	2440	2244	2178	1803



(a) PMSys **(b) Google Docs**

Figure 4: Number of self-reports.

calories.json shows how many calories the person has burned the last minute.

distance.json gives the distance moved per minute. Distance is in centimeters.

exercise.json describes each activity in more detail. It contains the date with start and stop time, time in different activity levels, type of activity, and various performance metrics depending on the type of exercise, e.g., for running, it contains distance, time, steps, calories, speed, and pace.

heart_rate.json shows the number of heartbeats per minute (bpm) at a given time.

sedentary_minutes.json sums up the number of sedentary minutes per day.

lightly_active_minutes.json sums up the number of lightly active minutes per day.

moderately_active_minutes.json sums up the number of moderately active minutes per day.

very_active_minutes.json sums up the number of very active minutes per day.

resting_heart_rate.json gives the resting heart rate per day.

sleep_score.csv helps understand the sleep each night so you can see trends in the sleep patterns. It contains an overall 0-100 score calculated from the composition, revitalization and duration scores, the number of deep sleep minutes, the resting heart rate, and a restlessness score.

sleep.json is a per sleep breakdown of the sleep into periods of light, deep, REM sleeps, and time awake.

steps.json displays the number of steps per minute.

time_in_heart_rate_zones.json gives the number of minutes in different heart rate zones. Using the common formula of 220 minus your age to find the max heart rate, Fitbit[6] will calculate your maximum heart rate and then create three target heart rate zones — fat burn (50 to 69 percent of your max heart rate), cardio (70 to 84 percent of your max heart rate), and peak (85 to 100 percent of your max heart rate).

As can be observed, there are various parameters included. For example, as we can see in Table 1, in total, there are 2,440 activity sessions (manual and 15-min-auto reports), 20,991,392 heart rate measurements, and 1,836 days of sleep scores included. It can, of course, be discussed how accurate data from a smartwatch can be. For example, we have observations that indicate that the Versa step-counter is influenced by other activities than walking or running and that the estimated distances are slightly inaccurate. For heart rates, the watch seems to be surprisingly accurate when we performed small comparisons using several devices at the same time. Thus, the Fitbit Versa 2 is not the best watch on the market, and the absolute values might be slightly off. However, the collected data should give reasonable indications of activities, and the relative differences between logs at least show if there have been positive or negative changes.

3.2 PMSys

In terms of subjective PMSys reporting, there are three CSV-files:

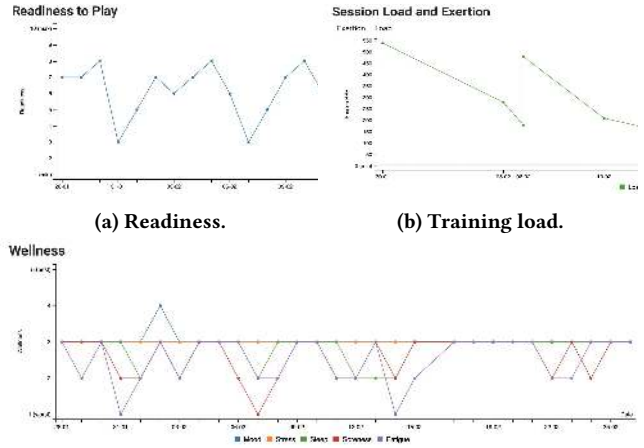
srpe.csv contains a training session's end-time, type of activity, the perceived exertion (RPE), and the duration in the number of minutes. This is, for example, used to calculate the session's training load or sRPE (RPE × duration).

wellness.csv 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, and 1-2 are scores 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 0-10) is an overall subjective measure of how ready you are to exercise, i.e., 0 means not ready at all, and 10 indicates that you cannot feel any better and are ready for anything!

injury.csv shows injuries with a time and date and corresponding injury locations and a minor and major severity.

Discussions in many fora are about the accuracy of subjective reports, as one is completely dependent on the truthfulness of the reporter. However, sport is not only a physical activity, and an athlete's psychological "state-of-mind" may greatly influence the performance. Thus, if reported correctly, the subjective information may be of huge value, and there may be important information to be found and predicted [18, 21]. In total, as seen in Figures 3 and 4a, there are 783 training sessions, 1,747 wellness reports, and 225

[6] https://blog.fitbit.com/max-heart-rate-by-age/



(a) Readiness. (b) Training load.

(c) Wellness parameters.

Figure 5: Examples of PMSys data that can be extracted.

Figure 6: Weight development of a few selected participants.

injury reports submitted. From the figure, we can see the difference in reporting activity among the participants. The plots from the PMSys trainer pages in Figure 5 show examples of data that can be retrieved.

3.3 Google Forms

The *googledocs* directory contains the *reporting.csv* data file, which contains daily reporting data. The data file contains one line per report, including the date reported for, timestamp of the report submission time, the eaten meals (breakfast, lunch, dinner, and evening meal), the participants weigh this day, the number of glasses drunk, and whether one has consumed alcohol.

In total, there are 1,569 reports (Figures 3 and 4b). Similarly to the PMSys data, these reports are also subjective, and some data points are missing. Nevertheless, the submitted data gives some indications of consumed food and drinks and might give important insights into calorie intake. Together with reported activity, this can indicate weight loss or gain, as shown in Figure 6.

3.4 Food Images Details

Participants 1, 3, and 5 took pictures of everything they consumed, except water, for two months (February and March 2020). Some example images can be seen in Figure 7. There are 644 images included, divided between the participants. Information about the day and time of capture can be found in the Exif image headers. The participants used their mobile phone cameras to collect the images (iPhone 6s, iPhone X, and iPhone XS). MacOS Photos was used to export the photos in full quality.

Figure 7: Examples of the captured images of food and drink.

4 INITIAL EXPERIMENTS

To demonstrate how the PMDATA dataset can be used, this section shows how machine learning can be applied to the data to predict weight gain or loss. More precisely, we define this as the problem of predicting weight change for the next day based on what was reported by the user the previous day. We model this as a classification problem, where we try to classify data from one day into three possible weight change classes for the next day. The three classes are: (0) weight goes down, (1) weight goes up, and (2) no weight change. For these experiments, we are using the following data sources from the PMDATA dataset: (i) Google doc reports, (ii) PMys wellness reports, and (iii) Fitbit sleep scores. We chose these three to show how the different data within the dataset can be combined and because we also had an intuition that well-being and sleep might correlate with weight change. The exact features used are weight_previous_day, water, alcohol, breakfast, lunch, dinner, evening, fatigue, mood, readiness, sleep_d, sleep_q, soreness, stress, overall_score, composition_score, and revitalization_score. We used only entries from the dataset that had at least the weight reported. Some of the data instances are missing values due to not being reported. We replaced the missing values with zeros[7]. This lead to a total of 1578 data instances. The distributions between the classes are 229 with weight goes down, 247 with weight goes up, and 1102 with no change of weight.

All experiments are performed using 10-fold cross-validation. The experiments are performed using two different algorithms: Random Forest and Classification Decision Tree (CDT). As a baseline, we provide ZeroR (majority class baseline). For all tested algorithms, we report the following metrics: false positive rate, precision, recall, F1-score, and Matthew Correlation Coefficient (MCC).

Table 2 shows the results for the experiments using all features. We can see that both Random Forest and CDT outperform the ZeroR baseline. The best classifier is CDT, with an MCC of weighted average MCC of 0.450. Predicting that weight goes down or up seems equally difficult. One might think that using the previous day's weight is a very important feature. To test this, we also conducted experiments with the two best working classifiers where the previous day's weight is removed as a feature. The results are presented in Table 3, where we can observe that the performance drops significantly. Both methods are having problems beating the majority class baseline significantly if the weight of the previous day is excluded as a feature. For this scenario, Random Forest is better than CDT, with an MCC of 0.259.

[7]We also tested to remove them, but replacing with zeros got overall a better score than removing the entries completely.

Table 2: Classification performance (10-fold cross-validation) including weight previous day feature.

Classifier	Class	False-Positive Rate	Precision	Recall	F1-Score	MCC
ZeroR baseline	weighted average	0.698	0.000	0.698	0.000	0.000
Random Forest	weight up	0.062	0.468	0.296	0.362	0.284
Random Forest	weight down	0.060	0.426	0.262	0.324	0.249
Random Forest	no change	0.532	0.802	0.933	0.863	0.471
Random Forest	weighted average	0.390	0.695	0.736	0.706	0.410
CDT	weight up	0.056	0.513	0.316	0.391	0.320
CDT	weight down	0.056	0.503	0.336	0.403	0.333
CDT	no change	0.504	0.811	0.937	0.870	0.504
CDT	weighted average	0.369	0.720	0.753	0.727	0.450

Table 3: Classification performance (10-fold cross-validation) excluding weight previous day feature.

Classifier	Class	False-Positive Rate	Precision	Recall	F1-Score	MCC
ZeroR baseline	weighted average	0.698	0.000	0.698	0.000	0.000
Random Forest	weight up	0.043	0.387	0.146	0.212	0.159
Random Forest	weight down	0.050	0.299	0.127	0.178	0.112
Random Forest	no change	0.725	0.751	0.946	0.838	0.313
Random Forest	weighted average	0.520	0.629	0.702	0.644	0.259
CDT	weight up	0.032	0.276	0.065	0.105	0.064
CDT	weight down	0.033	0.318	0.092	0.142	0.103
CDT	no change	0.821	0.731	0.965	0.832	0.244
CDT	weighted average	0.583	0.600	0.697	0.618	0.195

5 APPLICATIONS OF THE DATASET

PMData contains a large number of logged parameters that can be used for various analyzes like classification and prediction of a person's well-being and sports performance. Some examples using various selections of parameters include predicting a person's readiness to train for training planning, selecting the best team for the next competition, differences between genders or age, the results of the next competition, etc. The combination of the various parameters gives a unique opportunity to better find, for example, the total training load of a person, at an individual level, including data from even outside the training sessions. Thus, it is of large interest from the sports science point of view. Additionally, from a technical point of view, the time-series dataset is noisy, making it a challenge to analyze where one must handle missing data and find outliers, and the possibility to fuse various data sources raises diverse challenges. We plan to use the dataset for future projects, one being a system using PMData to estimate health states [16].

6 CONCLUSION

We have presented the PMData sports logging dataset, containing both objective and subjective parameters from sport and lifelogging, enabling the development of several interesting analysis applications. Our initial experiments show that such analyses are possible, but the dataset has great potential beyond what we have demonstrated in this paper. Other researchers using the dataset might want to look into some of the applications described in the application of the dataset section or come up with entirely new experiments and hypotheses.

REFERENCES

[1] Manal Chokr and Shady Elbassuoni. 2017. Calories prediction from food images. In *Proc. of the 29th Innovative Applications of Artificial Intelligence (IAAI) Conference.*

[2] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proc. of the 2016 Conference on Human Factors in Computing Systems (CHI)*. 2098–2110.

[3] Tung Duy Dinh, Dinh-Hieu Nguyen, and Minh-Triet Tran. 2018. Social Relation Trait Discovery from Visual LifeLog Data with Facial Multi-Attribute Framework. In *Proc. of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. 665–674.

[4] Aaron Duane and Cathal Gurrin. 2019. User interaction for visual lifelog retrieval in a virtual environment. In *Multimedia Modeling (LNCS)*, Vol. 11295. Springer, 239–250.

[5] Aaron Duane and Cathal Gurrin. 2020. Baseline Analysis of a Conventional and Virtual Reality Lifelog Retrieval System. In *MultiMedia Modeling (LNCS)*, Vol. 11962. Springer, 412–423.

[6] Peter Düking, Silvia Achtzehn, Hans-Christer Holmberg, and Billy Sperlich. 2018. 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).

[7] Peter Düking, Franz Konstantin Fuss, Hans-Christer Holmberg, and Billy Sperlich. 2018. 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 (30 April 2018), e102.

[8] Peter Düking, Christian Stammel, Billy Sperlich, Shaun Sutehall, Borja Muñiz-Pardos, Giscard Lima, Liam P Kilduff, Iphigenia Keramitsoglou, Guoping Li, Fabio Pigozzi, and Yannis P. Pitsiladis. 2018. Necessary Steps to Accelerate the Integration of Wearable Sensors Into Recreation and Competitive Sports. *Current sports medicine reports* 17, 6 (2018), 178–182.

[9] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. NTCIR Lifelog: The first test collection for lifelog research. In *Proc. of the International ACM SIGIR conference on Research and Development in Information Retrieval*. 705–708.

[10] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rashmi Gupta, Rami Albatal, Dang Nguyen, and Duc Tien. 2017. Overview of NTCIR-13 Lifelog-2 task. In *Proc. of the 13th NTCIR Conference on Evaluation of Information Access Technologies*. NTCIR.

[11] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval* 8, 1 (June 2014), 1–125.

[12] Bogdan Ionescu, Henning Müller, Renaud Péteri, Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Minh-Triet Tran, Mathias Lux, Cathal Gurrin, Yashin Dicente Cid, Vitali Liauchuk, Vassili Kovalev, Asma Ben Abacha, Sadid A. Hasan, Vivek Datla, Joey Liu, Dina Demner-Fushman, Obioma Pelka, Christoph M. Friedrich, Jon Chamberlain, Adrian Clark, Alba García Seco de Herrera, Narciso Garcia, Ergina Kavallieratou, Carlos Roberto del Blanco, Carlos Cuevas Rodríguez, Nikos Vasillopoulos, and Konstantinos Karampidis. 2019. ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction (LNCS)*, Vol. 11696. Springer, 358–386.

[13] Håvard Johansen, Cathal Gurrin, and Dag Johansen. 2015. Towards consent-based lifelogging in sport analytic. In *Multimedia Modeling (LNCS)*, Vol. 8936. Springer, 335–344.

[14] D. W. Johnston. 1993. The current status of the coronary prone behaviour pattern. *Journal of the Royal Society of Medicine* 86, 7 (1993), 406.

[15] Nguyen-Khang Le, Dieu-Hien Nguyen, Trung-Hieu Hoang, Thanh-An Nguyen, Thanh-Dat Truong, Duy-Tung Dinh, Quoc-An Luong, Viet-Khoa Vo-Ho, Vinh-Tiep Nguyen, and Minh-Triet Tran. 2019. HCMUS at the NTCIR-14 Lifelog-3 Task. In *Proc. of the NTCIR Conference on Evaluation of Information Access Technologies*. 48–60.

[16] Nitish Nag. 2020. *Health State Estimation*. Ph.D. Dissertation. University of California, Irvine.

[17] N. Nag and R. Jain. 2019. A Navigational Approach to Health: Actionable Guidance for Improved Quality of Life. *Computer* 52, 4 (April 2019), 12–20.

[18] Svein A. Pettersen, Håvard D. Johansen, Ivan A. M. Baptista, Pål Halvorsen, and Dag Johansen. 2018. Quantified Soccer Using Positional Data: A Case Study. *Frontiers in Physiology* 9 (2018).

[19] Thanh-Dat Truong, Tung Dinh-Duy, Vinh-Tiep Nguyen, and Minh-Triet Tran. 2018. Lifelogging retrieval based on semantic concepts fusion. In *Proceedings of the ACM Workshop on The Lifelog Search Challenge (LSC)*. 24–29.

[20] Kennet Vuong. 2015. *PmSys: a monitoring system for sports athlete load, wellness & injury monitoring*. Master's thesis. University of Oslo.

[21] Theodor Wiik, Håvard D Johansen, Svein-Arne Pettersen, Ivan Baptista, Tomas Kupka, Dag Johansen, Michael Riegler, and Pål Halvorsen. 2019. Predicting Peek Readiness-to-Train of Soccer Players Using Long Short-Term Memory Recurrent Neural Networks. In *Proc. of the 2019 International Conference on Content-Based Multimedia Indexing (CBMI)*. IEEE, 1–6.