Automated Analysis of Human Performance Data: a systematic mapping review.

T.B.Dijkhuis

January 11, 2017

Abstract

Automated Analysis of Human Performance Data could help to understand and possibly predict the performance of the human. To inform future research and enable Automated Analysis of Human Performance Data a systematic mapping study (scoping study)on the state-of-the-art knowledge is performed on three interconnected components(i)Human Performance (ii) Monitoring Human Performance and (iii) Automated Data Analysis. Using a systematic method of Kitchenham and Charters for performing the systematic mapping study, resulted in a comprehensive search for studies and a categorisation the studies using a qualitative method. This systematic mapping review extends the philosophy of Shyr and Spisic, and Knuth and represents the state-of-art knowledge on Human Performance, Monitoring Human Performance and Automated Data Analysis

Keywords: Human Performance, Automated Analysis, Monitoring,Data Science,Big Data, State-of-the-art review

1 Introduction

The focus of the, to be conducted, research is 'Automated Analysis of Human Performance Data'. The three interconnected main components are (i)Human Performance (ii) Monitoring Human Performance and (iii) Automated Data Analysis . Human Performance is both the process and result of the person interacting with context to engage in tasks, whereas the performance range is determined by the interaction between the person and the context[Dunn, W., & Brown, 1994]. Cheap and reliable wearable sensors allow for gathering large amounts of data, which is very useful for understanding, and possibly predicting, the performance of the user. Given the amount of data generated by such sensors, manual analysis becomes infeasible; tools should be devised for performing automated analysis looking for patterns, features, and anomalies. Such tools can help transform wearable sensors into reliable high resolution devices and help experts analyse wearable sensor data in the context of human performance, and use it for diagnosis and intervention purposes. Shyr and Spisic describe Automated Data Analysis as follows: "Automated data analysis provides a systematic process of inspecting, cleaning, transforming, and modelling data with the goal of discovering useful information, suggesting conclusions and supporting decision making for further analysis" [Shyr and Spisic, 2014]. Their philosophy is to do the tedious part of the work automatically, and allow experts to focus on performing their research and applying their domain knowledge. However, automated data analysis means that the system has to teach itself to interpret interim results and do iterations. Knuth stated: "Science is knowledge which we understand so well that we can teach it to a computer; and if we don't fully understand something, it is an art to deal with it."[Knuth, 1974]. The knowledge on Human Performance and its Monitoring is to be 'taught' to the system. To be able to construct automated analysis systems, an overview of the essential processes and components of these systems is needed.Knuth "Since the notion of an algorithm or a computer program provides us with an extremely useful test for the depth of our knowledge about any given subject, the process of going from an art to a science means that we learn how to automate something" [Knuth, 1974].

In the presented work the philosophy of Shyr and Spisic, and Knuth will be extended and research the present state-of-art of about Human Performance, Monitoring Human Performance and Automated Data Analysis.

2 Methods

2.1 Introduction

A comprehensive search for studies relevant to the treatment is to be performed, based on a systematic method [Kitchenham and Charters, 2007]. The method chosen is the systematic mapping study, systematic mapping studies are to provide an extensive overview of a research area. The major phases of the systematic mapping study are: Planning the Review, Conducting the Review, Reporting the Review.

2.2 Planning the review

The need for this systematic mapping review arises from the requirement to inform future research and enable development of novel and effective tools by gaining state-of-the-art knowledge regarding Automated Analysis of Human Behaviour Data.

The contribution to future research are:

1. An overview of studies on Human Performance, Monitoring Human Performance and Automated Data Analysis.

2. Classify the selected studies according to their spectrum of supported dimensions, and analyse each of these dimensions as invoked in the studies.

3. Derive the limitations of the current state-of-the-art automated data analysis within the context.

The main research question of this systematic review is:
"What are the key components of Automated Analysis of Human Performance Data ?"

The components for automated analysis of Human Performance data from the relevant studies were derived using qualitative analysis.

2.3 Conducting the review

Conducting the review is the strategy used to search for primary studies. The resources for the search include the databases:

1. IEEExplore
2. ACM Digital library
3. Google scholar (scholar.google.com)
4. Smart Cat (https://rug.on.worldcat.org)

Also industry resources from IBM, Gartner, SAS and Oracle are used to derive the knowledge about automated analysis.

The research question exists of three components:

1. Human Performance
2. Monitoring Human Performance
3. Automated Analysis of Human Performance Data

The search query is based on these three components, the exact keywords used were an iterative process based on the keywords of the studies found and read. The reference list of relevant studies was used identify other relevant studies. The following inclusion criteria for the studies were used: Publications between 2012-2016, major publications before 2012, English, German or Dutch language, articles published in scientific journals , conference papers, scientific books and influential resources outside science. Exclusion criteria were: Master papers, N=1 studies.

Information extraction from the studies is done by using the software tool Mendely to classify the study, extracting the abstract, and record the study characteristics journal. Highlighting relevant quotes and adding notes was done manually. For charting the data, the studies were categorized using the key components of the research question and the found relevant dimensions within the three components. Categorisation within Mendely was sufficient to find the relevant dimensions of the components.

2.4 Reporting the review

The dissemination of the results is, for now, limited to the HUAS lectureship New Business and ICT, the RUG research group Distributed Systems and RUG department Human Movement Sciences, with the aim of influencing the future direction of primary research on Automated Analysis of Human Performance Data.

3 Primary results

3.1 Introduction

There are three areas (i) Human Performance (ii) Monitoring Human Performance (iii) Automated Analysis of Human Performance Data each of which are elaborated below:

1. Human Performance
   Human Performance as a result of physical, psychological and social well-being and the interconnected changes in load, recovery and capacity.

2. Monitoring Human Performance
   The monitoring of the Human Performance is the application of methods used to measure the health indicators and the possible changes in load, recovery and performance. Wearables can replace some of the manual monitoring methods like questionnaires. For example the use of the heart rate measured by a wearable, instead of the use of a questionnaire as a monitoring tool of the perceived exertion.

3. Automated Analysis of Human Performance Data
   The goal of Automated Analysis is the discovery of useful information, suggesting conclusions, supporting decision-making on adjustments and further analysis. Automated Analysis is automation and selection of relevant aspects of the generic process of Knowledge Discovery in Databases.

Figure 1 shows the process of Automated Analysis of Human Performance Data. The elements of this model are the interconnected components of load, recovery and capacity, individually centered in the environment of Health, connected to monitoring these components for applying Automated Analysis. The figure is adapted from [Kentta and Hassmen, 1998], the three domains of health [Huber et al., 2011] and Automated Data Analysis [Shyr and Spisic, 2014]

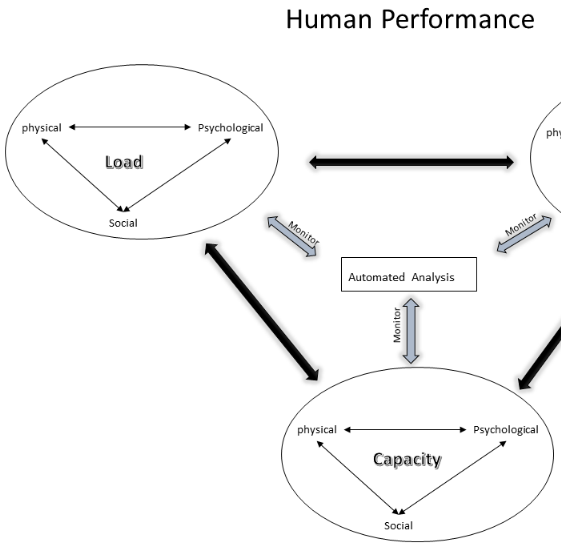

Figure 1: Automated Analysis of Human Performance Data.

Examples to illustrate the possible use of Automated Analysis of Human Performance Data are:

1. Load and Performance
   During the project 'Groningen Monitoring Athletic Performance' [Raak-Pro, 2011] a substantial amount of data and knowledge is gathered in various questions about load and tests on the performance of competitive athletes. The investigation into the factors that influence performance and injury risk of athletes lead to more insight into the information needed for connections between load and performance. Their investigation was done manually. By applying automated analysis possibly new insights in patterns connected to load and performance of athletes can be discovered.

2. Automated monitoring of load and performance
   During the project 'Groningen Monitoring Athletic Performance' [Raak-Pro, 2011] the load of the athletes was monitored by questionnaires. The quality of the data depended on the discipline of the athlete. By the use of wearables the burden of filling in questionnaires on Rated Perceived Exertion (RPE) versus Total Quality Recovery (TQR) and possible risk of loss of relevant data can be reduced. RPE versus TQR is a model developed by Kentta and Hassmen, which attempts to match the rate of perceived exertion with total quality recovery, where the athlete subjectively rates the daily workload and the daily recovery. Also the possibility of timely intervention becomes into reach when anomalies are automatically detected.

3. Occupational Health
   HanzeFit is a project of the Hanzehogeschool. Employees were coached to move more during daytime and wore wearables to gather information on the number of steps during working hours. The data on the amount of steps per individual is analysed by hand to see if the employees did move more during the daytime. When applying automated analysis, timely detection of anomalies in behaviour and the effectiveness of coaching becomes possible.

4. Work environment and stress
   Another possibility to apply automated analysis is the work environment and the amount of stress office workers experience. The amount of stress office workers experience in their work environment can be monitored with wearables and as well as the performance of the office workers. Automated Analysis can help to find patterns and make suggestions for further analysis of the factors influencing the level of stress.

3.2 Human Performance

Human Performance is both the process and result of the person interacting with context to engage in tasks, whereas the performance range is determined by the interaction between the person and the context[Dunn, W., & Brown, 1994]. The performance in figure 1 is defined by the dimension health and the interaction with the interconnected changes in load, recovery and capacity. There are several definitions of Health. The definition of health given by [Huber et al., 2011] is “health as the ability to adapt and to self-manage ” replaces the 1948 definition of WHO: “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.” [Huber et al., 2011] is grounded in the three domains of health: physical, mental, and social. The paradigm to look at these three domains are:

1. physical health is the capability to the maintenance of the physiological homoeostasis through changing circumstances.
2. mental health is the capability to cope, recover from strong psychological stress and prevent post-traumatic stress disorders.
3. social health is the capability to despite a medical condition, to participate in social activities.

This is a broad definition of health, the domain where automated analysis is to be applied is focussed on improving the physical health or preventing decline of the physical health. Improving physical health is the main topic of Human Movement Sciences. Balance is, as for every human, the key word. Not only the physical balance, Stress and recovery can also influence the athletic balance and capacity [Kentta and Hassmen, 1998]. The amount of psychosocial stress and recovery is dependent on the appraisal of stressful situations by the individual [Otter, 2016]. Examples of sources of psychosocial stress are pressure at school or work [Perna and McDowell, 1995]. Examples of psychosocial recovery are having a good time or feeling happy. Lamberts, for instance, claims that in high performance cycling, it is important to maintain a healthy balance between training load and recovery [Lamberts, 2014]. Also the training parameters and the occurrence of injuries of competitive athletes are related. It was found that injury rate is increased several weeks after a week of high training load or training intensity[Otter, 2016]. To improve capacity balancing between load and recovery the connection with the health aspects physical, mental and social is crucial.

3.3 Monitoring Human Performance

Monitoring Human Performance is done in different ways: using tests, (wearable) sensors, questionnaires and observations. Using tests and sensors in health is the domain of the digital medicine. Elenko et al.: "we define digital medicine technology and products as that technology and those products that are undergoing rigorous clinical validation and/or that ultimately will have a direct impact on diagnosing, preventing, monitoring or treating a disease, condition or syndrome. "[Elenko et al., 2015].
Wearables are lightweight, sensor-based devices which are worn close to and/or on the surface of the skin, where they detect, analyze, and transmit information concerning several internal and/or external variables to an external device[Düking et al., 2016]. As the line between consumer health wearables and medical devices begins to blur, it is now possible for a single wearable device to monitor a range of medical risk factors [Düking et al., 2016].
The use of wearables as tooling for health is a part of digital health . Sonnier defines "Digital health is the convergence of the digital and genomic revolutions with health, healthcare, living, and society. As we are seeing and experiencing, digital health is empowering us to better track, manage, and improve our own and our family's health, live better, more productive lives, and improve society."[Sonnier, 2013]. (i.e. Performance).
Commercial Activity trackers are developed to increase an individual's awareness about physical activity behaviour throughout the day. It is well known that regular physical activity decreases the risk of many chronic diseases and can improve quality of life and performance[Kooiman et al., 2015].
Wearables are becoming of more interest to athletes. On the one hand to monitor their health and on the other hand to monitor the training load. By measuring diverse trainings load parameters like distance, moist, heart rate (HR), hear rate recovery (HRR) , the athlete can measure himself/herself to adjust training load and methods. By measuring various health parameters like hydration, temperature and sleep the athlete can gain insights to help to adjust his or her behaviour [Halson, 2016, Düking et al., 2016]. Wearables alone are not enough to monitor load and health. Also, questionnaires are needed to assess subjective variables like mood disturbance or perceived stress and inadequate recovery [Saw et al., 2014]. Other elements of athlete self-report measures are RPE and TQR. To monitor Human Performance; wearables, questionnaires and tests are to be combined to get a complete picture.

3.4 Automated Analysis of Data

Large amounts of data are to arrive at a speed, that purely manual analysis is not possible for the human. A solution is automated data analysis Shyr and Spisic “Automated data analysis is a system that performs a systematic process of inspecting, cleaning, transforming, and modelling data with the goal of discovering useful information, suggesting conclusions and supporting decision making for further analysis ” [Shyr and Spisic, 2014].

Automated data analysis the automation of the process of Knowledge Discovery in Databases (KDD). KDD is defined by Fayyad et al. “KDD has evolved, and continues to evolve, from the intersection of research fields such as machine learning, pattern recognition, databases, statistics, AI, knowledge acquisition for expert systems, data visualization, and high-performance computing. The unifying goal is extracting high-level knowledge from low-level data in the context of large data sets. KDD places a special emphasis on finding understandable patterns that can be interpreted as useful or interesting knowledge. KDD comprises many steps, which involve data preparation, search for patterns, knowledge evaluation, and refinement, all repeated in multiple iterations. ”[Fayyad et al., 1996].

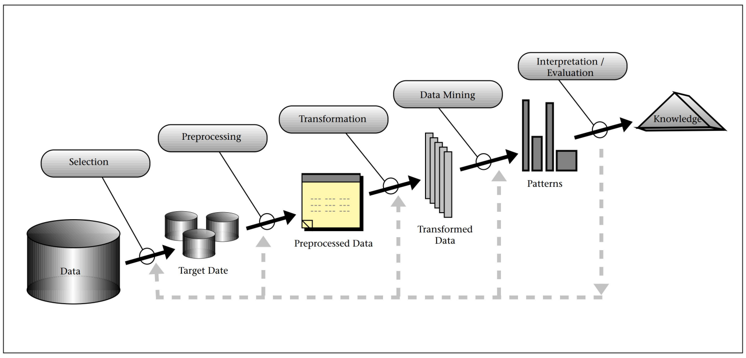

Figure 2: Knowledge Discovery in Databases

The five phases of the KDD process are:

1. Data selection: the aim is to understand the application domain, to define the goal of the data mining and as a result identify (possible) relevant data sources and collections of data.
2. Data pre-processing: the data is collected and stored in a way that in a way that allows for subsequent analysis. Typical actions in this phase, though not extensive, are cleaning data, handling missing values, errors, and noise.
3. Data transformation: the data is transformed into a format suitable for data mining. Transformation ranges from converting complex data (e.g. images) into a set of useful features to dimensionality reduction techniques, feature selection, and data sampling.
4. Data Mining: searching for relevant patterns and models, matching the initial goal, in this

step, data mining, statistical modelling techniques and/or machine learning techniques are applied to models or patterns.

5. Evaluation: the patterns and models derived in the data mining phase are examined on their validity, suitability and applicability.

3.4.1 The concepts of Data, Information and Knowledge

To be able to do automated analysis, the concepts of Data, Information and Knowledge are defined. The fourth element Wisdom, although important for mankind, is not applicable for automated analysis. “Data without interpretation have no value and only attain value when they are assessed and interpreted with existing knowledge. Information is the core element in forming knowledge and knowledge is essential to the processing of data and information.”[Sato and Huang, 2015]. Data is an object without meaning, it is only perceived as information once the data can be interpreted existing knowledge.“knowledge is defined by the possibilities of use and appropriation offered by discourse” [Foucault, 1995]. It depends on possibilities of use and appropriation what information can be extracted from the data.

3.4.2 Big Data

The framework of KDD stems out of 1996, in recent years Big Data and Data Science has become a main topic in the discovery of knowledge in data. Big Data is the technological evolution of the initial KDD. The definition of Big Data is still evolving where the three V’s Volume, Velocity and Variety are the standard dimensions to describe big data. The first definition of Big Data stems from Gartner,inc “Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.”[Laney, 2002]. Additional terms to define Big Data are:

1. Veracity is added by IBM “Managing the reliability and predictability of inherently imprecise data types”[Schroeck et al., 2012].

2. Value is added by Gantz and Reinsel Big Data as a technology designed to economically extract value from a wide variety of data and where data has intrinsic value which has to be discovered by a range of analysis techniques [Oracle, 2016, Gantz and Reinsel, 2011].

3. Variability is mentioned by SAS "In addition to the increasing velocities and varieties of data, data flows can be highly inconsistent with periodic peaks."[SAS, 2016].

But there are also more technical definitions for example: “Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning.” [Ward and Barker, 2013]. Ward and Barker states also that the high-volume aspect can be left out[Ward and Barker, 2013]. The definition of Big Data is diverse and as pointed out by Gandomi and Haiderthe relativity of volume applies for all dimensions, there are no universal benchmarks for volume, variety and velocity that define Big Data [Gandomi and Haider, 2015]. What is common ground for the definitions of Big Data is the overall recognition that extracting insight is what adds value.

Based on an article of [Labrinidis and Jagadish, 2012],[Gandomi and Haider, 2015] made a model of the overall process of extracting insights for Big Data

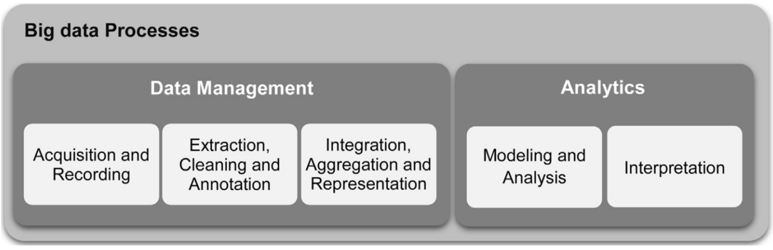

Figure 3: Big Data Processes

There's an new categorization of the processes in two components: Data Management and Analytics. The subjects of the processes for extracting insights from big data cover the same dimensions of the process of Knowledge Discovery in Databases.
1. Selection=>Acquisition and Recording
2. Preprocessing=>Extraction, Cleaning and Annotation
3. Transformation=>Integration,Aggregation and Representation
4. Data Mining=> Modelling and Analysis
5. Interpretation=> Interpretation and Evaluation

The processes for Big Data.
In a more technical survey on 'how' to handle Big Data and gain insight, the Big Data EcoSystem is defined. The Ecosystem is composed of many overlaid components that need to be integrated together and covers the possible technical choices between solutions which can be used to support the processes of Knowledge Discovery in Databases [Khalifa et al., 2016]. The techniques in the Big Data Eco System are divided into six pillars:

1. Storage: the handeling of high volume, high speed arrival and multiple formats of the data.
2. Processing: the processing depending on the nature of the need.
3. Orchestration: the orchestration of complex analytic jobs and work flows to achieve the user's goals
4. Assistance: help users with decisions when selecting operations and and building analytical processes
5. Interfacing: provide users with a familiar environment to build and run analytics.
6. Deployment: deployment of the Ecosystem

4 Automated Analysis of Human Performance Data

The five phases of the KDD process cover the main phases of Automated Analysis of Human Performance Data. In the section below an oversight of the possibilities within each phase is described shortly.

1. Data selection:the aim is to understand the application domain, to define the goal of the data mining and as a result identify (possible) relevant data sources and collections of data. This is the initiating phase of the process and corresponds to the overall goal: developing a set of tools to perform Automated Analysis of Human Performance Data. The application domain is Human Performance and the goal of the data mining is to discover useful information, suggesting conclusions and supporting decision making for further analysis out of Wearable and contextual data. Relevant data sources are those data sources which help to research the influences of load, recovery and capacity on the dimensions of social ,psychological and physical Health. Depending on the application of the data different data sources become feasible. For instance for the monitoring of the load and influence on the capacity of athletes a combination of wearables, dairy's, personal information and periodical performance tests are to be used. Whereas wearable data can be combined with Ecological Momentary Assessments (EMA) to inform on the mental state of humans [Blaauw et al., 2016].

2. Data pre-processing: the data is collected and stored in a way that in a way that allows for subsequent analysis.
   The collection of the data is a combination of monitoring the load, recovery and capacity by wearables and the use of other sources like questionnaires and tests.
   The monitoring and the use of other sources have implications on the human delivering the data. Filling in Questionnaires can be a burden, wearing a wearable can lead to stress [Halson, 2016]. To secure the safe and secure deliverance of the data,the collection of this data has to have as little impact as possible on the human [Piwek et al., 2016, Düking et al., 2016].
   In modern times, to store the data, not only relational databases are considered as an option, selecting data storage should be done considering the nature and characteristics of the data. Hereby entering a world of Polyglot Persistence where individual applications uses different data management systems [Sadalage and Fowler, 2012].
   There's no cookbook which is the perfect solution for a project. The nature of the data is an indication of the management system to be used[Hwang et al., 2015]. The nature and characteristics of the data can align to one of the four data model approaches of NoSQL: Key-Value, Document, Column Family and Graph. Next to the data model NoSQL Databases provide scalable,distributed, schema flexible storage of data.
   An example of an logical choice for a document store is the data from Fitbit which is in JSON format, what is actually an document. Although you have to convert the JSON document, to store the document database MongoDB the nature of the data model is the same.
   While the development of NoSQL databases provides scalable and distributed storage of data, new query tools and analytic platforms have been proposed to explore and smoothly process the large volumes of data currently being generated. The most popular platform is HADOOP [Anagnostopoulos et al., 2016]. HADOOP is an eco-landscape where other platforms are used to extract data from and perform analytics. There are many open

source solutions for performing data handling tasks. Examples are Mahout, Pig, Hive, Hbase, SPARK. It's landscape is very diverse, it is out of scope for this study to explicitly investigate all of these solutions. To store the data in a way that subsequent analysis is possible, traditional techniques should be used to data cleansing and data curation. The preferable solution is to set strict rules from the very beginning of data creation (data entry, submission, acquisition, etc.)[Anagnostopoulos et al., 2016].

3. Data transformation: the data is transformed into a format suitable for data mining. Transformation ranges from converting complex data (e.g. images) into a set of useful features to dimensionality reduction techniques, feature selection, and data sampling. Use for data mining requires data transformation depending on the goal of data mining: 1. Predictive, descriptive, machine learning[ad literature]

4. Data Mining:
   The lack of data is not the problem, getting information out of the data is the challenge of today. Data Mining is exploring the data from different aspects, classify it and summarize it. There are many data mining techniques and tools. The challenge is to apply the correct technique in accordance with the data to identify possible interesting patterns,predictions,correlations and/or models. The steps of Data Mining are partly overlapping with the definition of the phases of Knowledge Discovery in Databases. Studies on Data mining define the following steps like preprocessing the data (remove the noisy data, replace the missing values etc.), which differs from the definition of Fayyad et al. and corresponds with the Transformation Phase of KDD, feature selection (select the relevant features and discard the irrelevant and redundant features), classification and evaluation of different classifiers. There are several major data mining techniques have been developing and using in data mining projects including association, classification, clustering, prediction, pattern recognition.[Bhojani and Bhatt, 2016]

5. Evaluation: the patterns and models derived in the data mining phase are examined with respect to their validity, suitability and applicability. **Under construction**

Literature	Author	Subject	Domain
s1	Sprint and Cook [2016]	Detect changes in timeseries everyday physical activity data	Monitoring
s2	Blaauw et al. [2016]	Integration sensor data and ema	Monitoring
s3	Fawcett [2015]	QS Problem connect data to decisions of interest user	Monitoring
s4	Grossglauser and Saner [2014]	Data driven healthcare:autonomous decision-making	Automated Analysis
s5	Shyr and Spisic [2014]	Automated Data Analysis	Automated Analysis
s6	Khalifa et al. [2016]	Big Data Analytics Ecosystems	Automated Analysis
s7	Grosz et al. [2016]	AI overview and its influences as the field advances	Automated Analysis
s8	Huber et al. [2011]	Definition of Health	Human Performance
s9	Elenko et al. [2015]	Definition of digital medicine	Human Performance
s10	Sael et al. [2015]	Scalable Tensor Mining	Automated Analysis
s11	Gandomi and Haider [2015]	Big Data and analytics	Automated Analysis
s12	Lamberts [2014]	LSCT test to predict performance	Human Performance
s13	Piwek et al. [2016]	Consumer Health wearables	Monitoring
s14	Knuth [1974]	Computer programming from art to science	Computer Science
s15	Bhojani and Bhatt [2016]	Review Data mining Techniques	Data mining
textbf Under construction			

Table 1: Literature

References

I. Anagnostopoulos, S. Zeadally, and E. Exposito. Handling big data: research challenges and future directions. *The Journal of Supercomputing*, 72(4):1494–1516, 2016. ISSN 0920-8542. doi: 10.1007/s11227-016-1677-z. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-84962978213{&}partnerID=tZOtx3y1.

Shital H Bhojani and Nirav Bhatt. Commerce Computer Science Data Mining Techniques and Trends – A Review Assistant Professor-Anand Agriculture University-Anand - Gujarat Assistant Professor – RK University , Rajkot - Gujarat KEYWORDS : Knowledge Discovery in Databases (KDD), Data Mining. *Global Journal for Research Analysis*, (5):252–254, 2016.

F J Blaauw, H M Schenk, B F Jeronimus, L Van Der Krieke, P De Jonge, and M Aiello. Let ' s get Physiqual – An intuitive and generic method to combine sensor technology with ecological momentary assessments. *Journal of Biomedical Informatics*, 63:141–149, 2016. ISSN 1532-0464. doi: 10.1016/j.jbi.2016.08.001. URL http://dx.doi.org/10.1016/j.jbi.2016.08.001.

Peter Düking, Andreas Hotho, Hans Christer Holmberg, Franz Konstantin Fuss, and Billy Sperlich. Comparison of non-invasive individual monitoring of the training and health of athletes with commercially available wearable technologies. *Frontiers in Physiology*, 7(MAR), 2016. ISSN 1664042X. doi: 10.3389/fphys.2016.00071.

Catana. Dunn, W., & Brown. The Ecology of Human Framework for Considering the Effect of Context. *American Journal of Occupational Therapy*, 48:595–607, 1994. ISSN 0272-9490. doi: 10.5014/ajot.48.7.595.

Eric Elenko, Lindsay Underwood, and Daphne Zohar. Defining digital medicine. *Nature Biotechnology*, 33(5):456–461, 2015. ISSN 1087-0156. doi: 10.1038/nbt.3222. URL http://www.nature.com/doifinder/10.1038/nbt.3222.

Tom Fawcett. Mining the Quantified Self: Personal Knowledge Discovery as a Challenge for Data Science. *Big Data*, 3(4):249 –266, 2015. ISSN 2167-6461. doi: 10.1089/big.2015.0049. URL http://dx.doi.org/10.1089/big.2015.0049.

Usama Fayyad, Gregory Piatetsky-shapiro, and Padhraic Smyth. From Data Mining to Knowledge Discovery in. 17(3):37–54, 1996.

Michel Foucault. Archaeology of Knowledge. *Routledge classics*, page 239, 1995. ISSN 0539-0184. doi: 10.1177/053901847000900108. URL http://eprints.uwe.ac.uk/3530/$\delimiter"026E30F$nhttp://www.loc.gov/catdir/toc/fy0614/2002067999.html$\delimiter"026E30F$nhttp://www.loc.gov/catdir/enhancements/fy0650/2002067999-d.html$\delimiter"026E30F$nhttp://www.loc.gov/catdir/enhancements/fy1113/2002067999-b.html.

Amir Gandomi and Murtaza Haider. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2):137–144, 2015. ISSN 02684012. doi: 10.1016/j.ijinfomgt.2014.10.007.

John Gantz and David Reinsel. Extracting Value from Chaos State of the Universe : An Executive Summary. *IDC iView*, (June):1–12, 2011. URL http://idcdocserv.com/1142.

M. Grossglauser and H. Saner. Data-driven healthcare: from patterns to actions. *European Journal of Preventive Cardiology*, 21(2 suppl):14–17, 2014. ISSN 2047-4873, 2047-4881. doi: 10.1177/2047487314552755. URL `http://cpr.sagepub.com/content/21/2{_}suppl/14$\delimiter"026E30F$nhttp://cpr.sagepub.com/content/21/2{_}suppl/14.short$\delimiter"026E30F$nhttp://www.ncbi.nlm.nih.gov/pubmed/25354949`.

Barbara J Grosz, Russ Altman, Eric Horvitz, Alan Mackworth, Tom Mitchell, Deirdre Mulligan, and Yoav Shoham. Artificial intelligence and life in 2030: One hundred year study on artificial intelligence. Technical report, Standford University, 2016.

Shona Halson. Wearable Technology for Athletes : Information Overload and Pseudoscience ? *International Journal of Sports Physiology & Performance*, 11:705–706, 2016. doi: http://dx.doi.org/10.1123/IJSPP.2016-0486.

Machteld Huber, J André Knottnerus, Lawrence Green, Henriëtte van der Horst, Alejandro R Jadad, Daan Kromhout, Brian Leonard, Kate Lorig, Maria Isabel Loureiro, Jos W M van der Meer, Paul Schnabel, Richard Smith, Chris van Weel, and Henk Smid. How should we define health? *BMJ (Clinical research ed.)*, 343(9666):d4163, 2011. ISSN 1756-1833. doi: 10.1136/bmj.d4163. URL `http://www.ncbi.nlm.nih.gov/pubmed/21791490`.

Jong Sung Hwang, Sangwon Lee, Yeonwoo Lee, and Sungbum Park. A selection method of database system in bigdata environment: A case study from smart education service in Korea. *International Journal of Advances in Soft Computing and its Applications*, 7(1):9–21, 2015. ISSN 20748523.

G Kentta and P Hassmen. Overtraining and recovery: a conceptual model Surentrainement et recuperation: un modele conceptuel. *Sports Medicine*, 26(1):1–16, 1998. ISSN 01121642. doi: 10.2165/00007256-199826010-00001.

Shadi Khalifa, Yehia Elshater, Kiran Sundaravarathan, Aparna Bhat, Patrick Martin, Fahim Imam, Dan Rope, Mike Mcroberts, and Craig Statchuk. The Six Pillars for Building Big Data Analytics Ecosystems. *ACM Computing Surveys*, 49(2):1–36, 2016. ISSN 03600300. doi: 10.1145/2963143. URL `http://dl.acm.org/citation.cfm?doid=2966278.2963143`.

Barbara Kitchenham and Stuart Charters. Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3. *Engineering*, 45(4ve):1051, 2007. ISSN 00010782. doi: 10.1145/1134285.1134500. URL `http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Guidelines+for+performing+Systematic+Literature+Reviews+in+Software+Engineering{#}0$\delimiter"026E30F$nhttp://www.dur.ac.uk/ebse/resources/Systematic-reviews-5-8.pdf`.

Donald E Knuth. Computer programming as an art. *Communications of the ACM*, 17(12):667–673, 1974. ISSN 00010782. doi: 10.1145/361604.361612.

Thea J M Kooiman, Manon L Dontje, Siska R Sprenger, Wim P Krijnen, Cees P van der Schans, and Martijn de Groot. Reliability and validity of ten consumer activity trackers. *BMC sports science, medicine and rehabilitation*, 7:24, 2015. ISSN 2052-1847. doi: 10.1186/s13102-015-0018-5. URL `/pmc/articles/PMC4603296/?report=abstract`.

Alexandros Labrinidis and Hosagrahar V Jagadish. Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12):2032–2033, 2012.

Robert P. Lamberts. Predicting cycling performance in trained to elite male and female cyclists. *International Journal of Sports Physiology and Performance*, 9(4):610–614, 2014. ISSN 15550265. doi: 10.1123/IJSPP.2013-0040a.

Doug Laney. What is big data, 2002. URL `http://www.gartner.com/it-glossary/big-data/`.

Oracle. What is big data, 2016. URL `https://www.oracle.com/big-data/index.html`.

Ruby T A Otter. *Monitoring endurance athletes A multidisciplinary approach.* 2016. ISBN 9789036787284.

Frank M. Perna and Sharon L McDowell. Role of psychological stress in cortisol recovery from exhaustive exercise among elite athletes. *International Journal of Behavioral Medicine*, 2(1): 13–26, 1995. ISSN 1070-5503. doi: 10.1207/s15327558ijbm0201_2.

Lukasz Piwek, David A. Ellis, Sally Andrews, and Adam Joinson. The Rise of Consumer Health Wearables: Promises and Barriers. *PLoS Medicine*, 13(2):1–9, 2016. ISSN 15491676. doi: 10.1371/journal.pmed.1001953.

SIA Raak-Pro. Belasting en belastbaarheid van topsporters, 2011. URL `http://www.sia-projecten.nl/projectenbank/project/belasting-en-belastbaarheid-van-topsporters`.

Pramod J. Sadalage and Martin Fowler. *NoSQL Distilled.* Addison-Wesley Professional, 2012. ISBN 0321826620. doi: 10.1007/s13398-014-0173-7.2.

Lee Sael, Inah Jeon, and U. Kang. Scalable Tensor Mining. *Big Data Research*, 2(2):82–86, 2015. ISSN 22145796. doi: 10.1016/j.bdr.2015.01.004.

SAS. What is big data, 2016. URL `http://www.sas.com/en_us/insights/big-data/what-is-big-data.html`.

Atsushi Sato and Runhe Huang. A Generic Formulated KID Model for Pragmatic Processing of Data, Information, and Knowledge. *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, pages 609–616, 2015. doi: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.120. URL `http://ieeexplore.ieee.org/document/7518299/`.

Anna E. Saw, Luana C. Main, and Paul B. Gastin. Monitoring athletes through self-report: Factors influencing implementation. *Journal of Sports Science and Medicine*, 14(1):137–146, 2014. ISSN 13032968. doi: 10.1519/JSC.0000000000000499.

Michael Schroeck, Rebecca Shockley, Janet Smart, Dolores Romero-Morales, and Peter Tufano. Analytics: The real-world use of big data. *IBM Global Business Services Saïd Business School at the University of Oxford*, pages 1–20, 2012. URL `http://www-03.ibm.com/systems/hu/resources/the{_}real{_}word{_}use{_}of{_}big{_}data.pdf$\delimiter"026E30F$nhttp://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Analytics+:+The+real-world+use+of+big+data{#}0`.

Jing Shyr and Damir Spisic. Automated data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(5):359–366, 2014. ISSN 19390068. doi: 10.1002/wics.1318.

Paul Sonnier. Definition of digital health, 2013. URL http://storyofdigitalhealth.com/definition/.

Gina Sprint and Diane J Cook. Unsupervised detection and analysis of changes in everyday physical activity data. *Journal of Biomedical Informatics*, 63:54–65, 2016. ISSN 15320464. doi: 10.1016/j.jbi.2016.07.020. URL http://dx.doi.org/10.1016/j.jbi.2016.07.020.

Jonathan Stuart Ward and Adam Barker. Undefined By Data: A Survey of Big Data Definitions. *arXiv.org*, page 2, 2013. URL http://arxiv.org/abs/1309.5821$\delimiter"026E30F$npapers3://publication/uuid/63831F5F-B214-46D5-8A86-671042BE993F.