# Driving Big Data – Integration and Synchronization of Data Sources for Artificial Intelligence Applications with the Example of Truck Driver Work Stress and Strain Analysis

Matthias Klumpp
*University of Göttingen*, matthias.klumpp@uni-goettingen.de

Benedikt Severin
*University of Duisburg-Essen*, benedikt.severin@uni-due.de

Henrik Lechte
*University of Goettingen*, henrik.lechte@uni-goettingen.de

Jannes Heinrich Diedrich Menck
*University of Goettingen*, jannes.menck@uni-goettingen.de

Maria Keil
*Georg-August-University Goettingen*, maria.keil@uni-goettingen.de

*See next page for additional authors*

## Presenter Information

Matthias Klumpp, Benedikt Severin, Henrik Lechte, Jannes Heinrich Diedrich Menck, Maria Keil, Sarah M. Straub, and Viola Milke

# Driving Big Data – Integration and Synchronization of Data Sources for Artificial Intelligence Applications with the Example of Truck Driver Work Stress and Strain Analysis

*Completed Research Paper*

**Matthias Klumpp**
University of Goettingen, Platz der Goettinger Sieben 3, 37073 Goettingen
matthias.klumpp@uni-goettingen.de

**Benedikt Severin**
University of Duisburg-Essen, Schützenbahn 70, 45127 Essen
benedikt.severin@uni-due.de

**Henrik Lechte**
University of Goettingen, Humboldtallee 3, 37073 Goettingen
henrik.lechte@uni-goettingen.de

**Jannes H. D. Menck**
University of Goettingen, Humboldtallee 3, 37073 Goettingen
jannes.menck@uni-goettingen.de

**Maria Keil**
University of Goettingen, Platz der Goettinger Sieben 3, 37073 Goettingen
maria.keil@uni-goettingen.de

**Sarah M. Straub, Caroline Ruiner**
University of Hohenheim, Wollgrasweg 23, 70599 Stuttgart, {sarah.straub; caroline.ruiner}@uni-hohenheim.de

**Viola Milke, Vera Hagemann**
University of Bremen, Enrique-Schmidt-Straße 1, 28359 Bremen
vhagemann@uni-bremen.de

**Marc Hesenius**
University of Duisburg-Essen, Schützenbahn 70, 45127 Essen
marc.hesenius@uni-due.de

## Abstract

This paper contributes to the issue of Big Data analysis and data quality with the specific field of time synchronization. As a highly relevant use case, Big Data analysis of work stress and strain factors for driving professions is outlined. Drivers experience work stress and strain due to trends like traffic congestion, time pressure or worsening work conditions. Although a large professional group with 2.5 million (US) and 3.5 million (EU) truck drivers, scientific analysis of work stress and strain factors is scarce. Driver shortage is growing into a large-scale economic and societal challenge, especially for small businesses. Empirical investigations require Big Data approaches with sources like physiological and truck, traffic, weather, planning or accident data. For such challenges, accurate data is required, especially regarding time synchronization. Awareness among researchers and practitioners is key and first solution approaches are provided, connecting to many further Machine Learning and Big Data applications.

**Keywords:** Data integration, data synchronization, truck driving, work stress and strain analysis, Big Data analysis, Lab Streaming Layer, Artificial Intelligence

# Introduction

As applications of Artificial Intelligence (AI) and Machine Learning (ML)[1] are dubbed *data-driven applications*, the emphasis on the importance of data – and thus data collection – is immanent. A multitude of very different data sources must be integrated for a meaningful analysis, which in turn requires development teams to take several aspects into account (Moraru et al. 2010). Mistakes during data collection typically cannot be corrected afterwards and will most likely invalidate the results. Even worse, if data collection mistakes remain undetected, they may cause unreliable results that cannot be trusted or replicated, which is in general an arising problem with AI (Morabit, Desaulniers and Lodi 2021). Hutson (2018) discusses the natural randomness of AI training runs with regard to replicating AI results, but we argue that mistakes made during data collection can yield similar problems.

While, especially in Software Engineering, attempts have been made to structure the development of AI applications and thus gain a hold on complexity (Amershi et al. 2019; Hesenius et al. 2019; Sundararaman, Buy and Kshemkalyani 2005), they typically assume that data has already been collected and thus offer little insight into how to structure data collection efforts (Frontoni et al. 2022; Georgieva et al. 2022). Especially for critical application domains such as medical applications, where data is the foundation for diagnoses with a potentially large impact on patient wellbeing – see (Kühnisch et al. 2022; McLennan et al. 2022; Urbina and Ekins 2022) –, dedicated efforts to ensure data collection quality are required. Therefore, data preparation, integration, and synchronization are major issues when developing AI applications, especially as development teams may encounter initially ”hidden problems” when debugging issues and investigating unwanted application behavior. In many cases, collected data cannot be “adjusted” after development or after deployment, emphasizing the need for proper engineering in early project phases, including the notion of integrating data sources outside the focal organization. In this paper, we discuss major challenges for data integration and synchronization from different sources:

- problems arising from variances in timekeeping,
- problems connected to network latencies,
- problems due to diverse data collection and storage location concepts and applications and
- problems induced by different data levels and standards.

Big Data analytics and the Internet of Things (IoT) are popular research areas where time synchronization of data is of high importance (Calyam et al. 2016; Tirado-Andrés, Rozas and Araujo 2019; Yigitler, Badihi and Jäntti 2020). However, in our view, many Big Data and data analytics frameworks in the information systems (IS) domain do not sufficiently consider the setup of complex (potentially hardware-based) data collection processes. For instance, practical methodologies like CRISP-DM (Wirth and Hipp 2000) and Knowledge-Discovery in Databases (KDD) (Fayyad, Piatetsky-Shapiro and Smyth 1996) typically deal with pre-existing data. Similarly, for rather high-level data analytics frameworks from the IS domain, the topic of time synchronization appears to be underrepresented (e.g., frameworks in Phillips-Wren et al. (2015) or Kühl et al. (2021)). However, as time synchronization issues may become apparent when combining multiple data sources, addressing this topic holistically and at an early stage in the data analysis process (i.e., before collecting the data) is beneficial.

We *contribute* to solving these challenges with the analysis of typical hurdles and topics in data integration and synchronization for a Big Data and AI. We review existing literature and summarize solutions focusing on resolving synchronization issues. In addition, we develop a draft scheme for the identification and handling of such issues as a generalized model for application development. This model can serve as a building block that can be used in conjunction with existing Big Data frameworks especially for use cases relating to sensor networks and the Internet of Things. A major goal is to raise awareness of potential problems and to provide guidance to other researchers and developers engaging in data collection as a foundation for components using AI techniques. This paper is geared towards scientists and practitioners who aim to combine multiple data sources that need to be synchronized based on their temporal properties. As a hands-on example, we report from our ongoing work in identifying causes for mental stress and strain with truck drivers, where we had to cope with a variety of challenges during data collection that we will

[1] For the sake of brevity, we consider ML and AI to be synonyms, although technically ML is just a particular subset of AI. However, ML is the predominant and most common AI-technique currently in use.

discuss in detail. The paper is *structured* as follows: Section two outlines the state of the art regarding data integration and synchronization. Section three is presenting a data integration and synchronization model draft. The specific data integration problem setting for the use case of analysing truck driver stress and strain is outlined in section four, discussing the technical and process options connected to that use case to derive general application principles. This is discussed in section five as core contribution by integrating further references and case application insights into an elaborate and generalized model draft regarding data integration and synchronization issues in the context of AI applications. Limitations and an outlook towards further research complete the final section.

## Conceptual Background

Data collection mistakes may cause an AI to yield unreliable results that cannot be trusted or replicated (Morabit, Desaulniers and Lodi 2021). Dorst et al. (2021) showed that even minor synchronization errors in data collection might have negative influences on ML algorithms. Especially when the trained model is planned to be used for other setups than the one the training data is gathered synchronization becomes mandatory for both the training data and the live setup. Time synchronization issues were also found to be problematic in non-AI systems such as inertial navigation systems (Skog and Handel 2008) or LAN-based digital substations (Son, Chang and Kang 2019). This highlights the need for synchronizing data, leading to a first assessment of the general synchronization approaches that can be used in a specific setup (Brahmi et al. 2013; Youn 2013). In general, there are different approaches to synchronize data sources:

- Synchronizing internal clocks of sensor devices: If the clocks of all sensor devices are synchronized, correct timestamps can be assigned to all sensor data before storage in a central system or database.
- Synchronizing data at central system considering known message delay: If the delay between event occurrence and final data storage in a central system is known, the final timestamp of the central system can be used, corrected by the delay.
- Synchronizing after data collection: Regardless of timestamps obtained while collecting data, different data sources may be synchronized through events leaving identifiable traces in all data sources, e.g., as described by Fridman et al. (2016).
- Sensor device calibration as a means to achieve one of the three other approaches: Devices may be calibrated, e.g., by exposing sensors to an external stimulus at a known time.

For a more detailed technical viewpoint, general synchronization strategies and algorithms as listed above can be found in extant literature such as Sundararaman, Buy and Kshemkalyani (2005) or Olson (2010). The choice of synchronization approach and hardware is also related to the *causes* of synchronization errors. Existing literature such as (Ping 2003) or (Sivrikaya and Yener 2004) provide an overview of common contributors to message delay like sender delay, transmit time, and receiver delay, which may be a contributing factor to time synchronization issues. Understanding where and how delays occur may help in assessing *if* and *how* synchronization errors arise. If sensors are synchronized based on analogue or digital stimulus sent to all sensors simultaneously, one decision refers to the timing of marker impulses. Two strategies exist (Brahmi et al. 2013): (a) generating stimulus in periodic time intervals or (b) generating stimulus at significant points of time e.g., at the very beginning and ending of a measurement. In both cases, data is time synchronized during post-processing by shifting the time series until all markers overlap.

In general, time synchronization is a well-known issue in distributed systems. As local clocks may exhibit offsets and drifts (Sivrikaya and Yener 2004), local system times alone cannot be compared for most use cases. In the past, several approaches have been proposed, both for synchronizing clocks and for determining the causal order of events (Sundararaman, Buy and Kshemkalyani 2005). A relatively simple approach for synchronizing clocks is the algorithm proposed by Cristian (1989) where a timestamp is requested from a central server and corrected by half the round-trip-time. Another approach is Berkeley's algorithm which aims to set the clocks of all systems to the average time of systems which have a timestamp within a specified deviation (Gusella and Zatti 1989). In the Internet, a similar concept to Cristian's algorithm is commonly used, the Network Time Protocol (NTP) (Mills 1991). However, in presence of a message delay, a perfect synchronization of local clocks is impossible (Lundelius and Lynch 1984). Therefore, in some distributed systems, the notion of causal ordering is used instead which guarantees the temporal relation between specific sets of events. For this purpose, Lamport (1978) used the concept of logical timestamps where each event is assigned a logical time. This concept has been extended, for example, in the form of vector clocks (Baldoni and Raynal 2002; Liskov and Ladin 1986). In extending the

NTP concept, Precision Time Protocol (PTP) was conceived according to IEEE Standard 1588-2008 (IEEE 2008): The PTP is a protocol designed to synchronize system times in small networks and aims to achieve as high precision as possible. It is based on a master-slave approach, where one specific client is defined as the master (also referred to as the "Grandmaster Clock") that broadcasts its system time to all other clocks for synchronization. PTP-connected systems and subnets can be dynamically reconfigured: If a clock loses contact to the Grandmaster Clock but can still connect to any other (slave) clock in the network, the intermediate clock can serve as a (temporary) master for the cut-off clock. Neagoe, Cristea and Banica (2006) compare PTP and NTP for general architecture and protocol design, stating that PTP allows for a synchronization performance on the sub-microseconds level, while NTP on milliseconds-level (both with regard to local area networks). However, they deem NTP more robust with regard to synchronization errors.

One solution for time synchronization of data is the open-source software Lab Streaming Layer (LSL: https://github.com/sccn/labstreaminglayer). Coming from neurobiological research, the protocol allows synchronizing data streams of sensor devices which are connected to the same network infrastructure. By sending continuous requests to all connected devices and measuring the delay between message and response, time shifts can be identified, and measurements can be temporal re-aligned. The LSL concept differentiates between two roles: (1) the data provider with a *stream outlet* and (2) the data consumer with a *stream inlet*. Providers are typically medical instruments or sensors making measurement data available. Consumers act upon the incoming data e.g., by writing the data to disk or by visualizing the data streams. To use the LSL, a dedicated software library must be integrated into all participating hardware or software components. This requirement becomes problematic when using closed-source systems like consumer-grade pulse sensors or embedded control software in a truck. Data in the LSL network is transmitted using the basic network protocols TCP and UDP. Artoni, Galeasso and Micera (2019) compared the jitter and delays of real-time synchronization *during* measurement as done by the LSL with a more conventional approach which used a wired synchronization channel and their results emphasize that LSL captured data streams quality is comparable to a classic acquisition setting. In another work, Artoni et al. (2017) identified four strategies for synchronization approaches with analog or digital pulses via a dedicated Transistor-Transistor-Logic (TTL) port as alternatives to live LSL alignment. Previous applications of LSL can be found in brain-computer-interaction (BCI) (Wang et al. 2017), human-computer-interaction (HCI) in general (Rozado, Niu and Lochner 2017), muscle-computer-interaction (MuCi) (Karolus et al. 2018; Karolus et al. 2020; Kilian et al. 2021), virtual and augmented reality applications (AR/VR) (Kosuru et al. 2019; Ostrin, Frey and Cauchard 2018; Stepanova et al. 2020; Wang et al. 2021a) and neurobiological experimentation (Manjunatha et al. 2020; Mendonca and Abreu 2019). Blum et al. (2021) developed an android app making step count, accelerometer or environmental light conditions sensor data available via LSL.

## Concept Model Data Integration

Before considering technical solutions, a detailed understanding of the individual goals and restrictions of the data collection with respect to time synchronization needs to be developed. Therefore, based on existing literature and an expert workshop conducted in mid-March 2022 with 17 participants, we created the conceptual framework depicted in Figure 1 below. This framework should serve as a starting point to help scientists and practitioners assess their data collection regarding time synchronization and make an informed choice on sensor appliances, synchronization approaches, and hardware setup.

First, researchers and practitioners should consider *if* the problem of temporal synchronization is applicable to their use case. Many data collection processes do not face the issue of temporal synchronization as data sets are either matched based on other properties (e.g., geolocation, or user identifiers) or because they simply require a granularity where no mismatch is expected as slightly differing clock times are no issue (e.g., based on date or hour). Furthermore, multiple and complex data types may be collected using a single device (e.g., smartphone) so that no network delay is expected and data is timestamped based on the common device time as reference.

Second, it is important to achieve a good understanding of the *general requirements and constraints* that are not related to time synchronization. Researchers should be aware of the different data sources and associated needs regarding data quality and granularity. For example, aspects such as required data accuracy and precision may be relevant determinants for the choice of sensor appliances. Apart from requirements directly relating to data, typical other constraints for sensors can be found in literature relating to IoT, such as budget (Aygün and Cagri Gungor 2011), size (Maksimovic, Vujovic and Perisic 2015), energy consumption (Lin et al. 2010), and reliability (Castaño et al. 2019). For many data collection processes, low cost approaches have been established such as the usage of smartphones or wearables. For example, in driving data collection, smartphone apps can often be used as an alternative to built-in and specialized equipment (Bethge et al. 2021; Engelbrecht et al. 2015; Mantouka et al. 2021). This is also the case for medical data (Esco, Flatt and Nakamura 2017; Hernando et al. 2018). Moreover, specialized sensor appliances may be required, e.g., because of the desired accuracy and resolution of sensor data. In this case, the adaptability of devices may be limited, thereby reducing the availability of some mitigation strategies regarding time synchronization issues. In other cases, data may stem from an external source with limited control and knowledge regarding data collection and possible synchronization errors. Once general requirements are known, the question arises *what* consequences imperfect synchronization has for the researcher's use case, i.e., how resilient the data analysis processes are to misaligned data. For instance, if time synchronization issues are uniform among all data, for example, in case of a fixed delay, no adverse impact on ML may occur. In other cases, ML models may be adversely affected, even by small time shifts (Dorst et al. 2021). This helps to develop an understanding of the threshold and magnitude of possible synchronization errors required for the use case. Depending on the maximum acceptable error, the choice of hardware and synchronization approaches will vary. Besides, the data analysis focus may also influence the criticality of time shifts. If single moments in time are of interest, the synchronization requirement may be more demanding compared to an analysis of a longer period where slightly shifted start and end may have a minor impact on pattern detection.

Third, researchers should identify, on a high level, *which requirements* exist specifically for solving time synchronization—apart from the acceptable magnitude of errors. These constraints can be of technical or non-technical nature and are strongly interlinked with the general requirements for data collection. Typically, one constraint lies in the monetary budget and whether data synchronization requires specialized equipment. Additionally, there may be other high-level requirements for time synchronization that limit the choice of solutions. These requirements call for sensible trade-offs and may comprise aspects like accuracy, scalability, efficiency, robustness, security, lifetime, cost and size (Puttnies et al. 2020; Sivrikaya and Yener 2004). These aspects are also related to the physical setup of data collection devices (e.g., factory or vehicle) and the future accessibility of the system. We think that the importance of these aspects also depends on the goals and lifetime of the sensor setup, i.e., long-lasting IoT sensor networks versus small-scale experimental data collection.

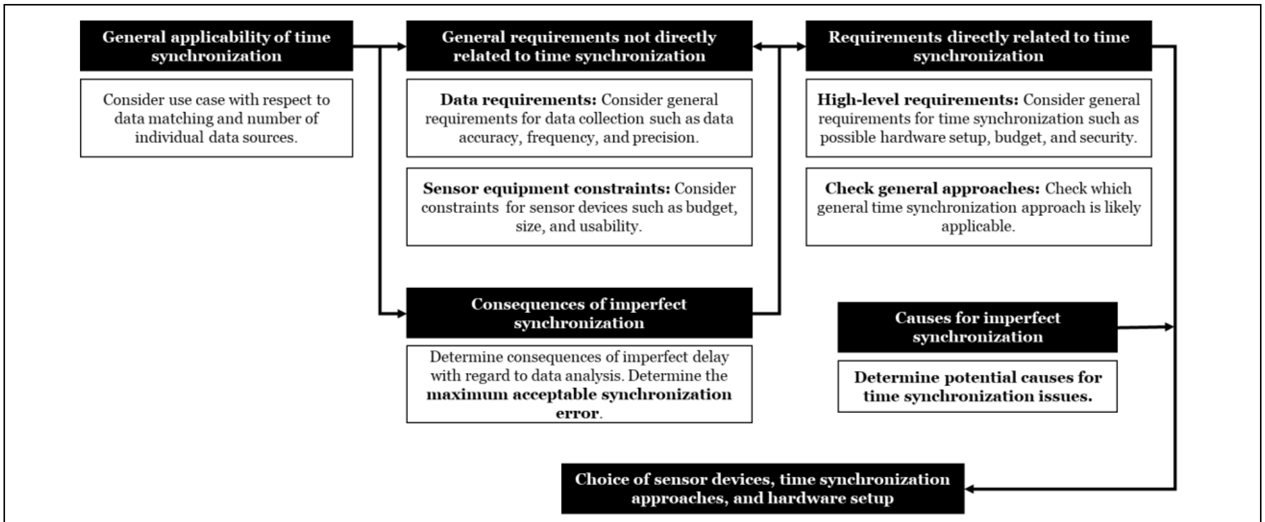


**Figure 1. High-level Framework for Solving Time Synchronization Issues**

## Application Case: Stress and Strain Analysis for Driving Professions

The use case presented here is intended to serve as an example of such data synchronization issues to illustrate the benefits. Drivers as a professional group are susceptible to a high level of work-related pressure and unhealthy working conditions (Hege et al. 2019; Hill and Boyle 2007; Onninen et al. 2021). Nevertheless, in spite of the central backbone role in transportation and economic systems around the globe, empirical research regarding the impacting factors of driver stress and strain is scarce. This is all the more astonishing as these themes are not only relevant for truck drivers themselves or the business environment via global supply chains, but also affect general security issues in public traffic (Cai et al. 2021; Ge et al. 2014; Han and Zhao 2020). Stress in general has been subject to numerous studies (Fischer, Reuter and Riedl 2021; Järvelin-Pasanen, Sinikallio and Tarvainen 2018; Öz, Özkan and Lajunen 2010; Rastgoo et al. 2019; Rowden et al. 2011; Taelman et al. 2011; Thielmann, Pohl and Böckelmann 2021) and several attempts have been made to measure stress levels, e.g. Bleichner and Emkes (2020) or Bleichner and Debener (2017). While these approaches indicate the existence of stress, they cannot be used to determine the causes directly: Determining the causes requires the analysis of additional data sources to introduce meaningful context factors, e.g., weather, traffic, health, or work status. With about 2.5 million truck drivers in the US and 3.5 million in the European Union, the segment of driving professions is one of the largest in most economies – but still, this professions also sports the most prominent shortages in worker supply and the highest levels of safety issues at the same time (Santos and Lu 2016; Sartori, Smet and Vanden Berghe 2021; Sekkay et al. 2021).

Consequently, research has concentrated on safety issues connected to driver stress and strain as well as other factors (Shattell et al. 2010; Useche, Ortiz and Cendales 2017; Wang et al. 2021b). But yet, a comprehensive approach regarding the integration of different data sources from vehicles, drivers and external factors and sources is missing. This is largely due to the fact of a demanding challenge of data integration and synchronization. Individual data sources like for example electroencephalography (EEG) or other physiological parameters are available (Reiser, Wascher and Arnau 2019; Wascher et al. 2021). But they are hardly to access in a common approach with for example data from the vehicles themselves – although there is also ample supply of data. In the working world, people find themselves in a socio-technical system. Each person is exposed to various influencing factors that are processed individually. These factors can result from social conditions, technical influences, environmental influences and also from individual conditions. For this reason, an extensive survey of various objective and subjective data is implemented in the use case, as well as the inclusion of environmental influences and environmental data (see Figure 2, Table 1). Methods and possibilities of data integration presented can also be used in other contexts to answer job-specific questions. It is conceivable to record the stress experienced by nursing staff during the execution of their respective tasks or to optimize workflows in delivery processes.

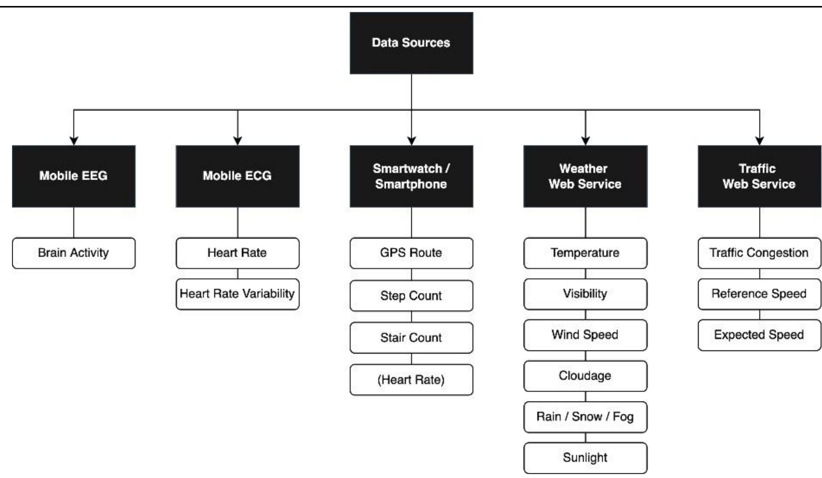


**Figure 2. Data Structure Use Case**

As part of the objective surveys, an electrocardiogram (ECG), heart rate variability (HRV) and electro-encephalography (EEG), among others, will be performed. Furthermore, saliva samples are taken here to determine cortisol levels and attention tests are performed. As part of the subjective measurements, various questionnaires are administered on social demographics, impulsivity, and chronotype. In addition, data from vehicles will be included in the analyses and publicly available data such as traffic situations, weather conditions, routing information and other information about the time and special events during the tour will be considered. The collected data will be processed and analyzed by AI with a focus on stress and strain. In the context of prognostics and health management, this should ideally identify and minimize unfavourable situations. Table 1 shows the collected data.

| Data | Description |
|---|---|
| EEG | Stress can be categorized as a human body response to mental, physical and emotional stimuli (Katmah et al. 2021). Indicators of stress can be quantified objectively using biosignals and markers, like the brain activation (Hou et al. 2015). EEG provides a direct non-invasive measurement and record the brain's electrical activity (Saidatul et al. 2011). This technique has the key advantage of high time resolution, with the possibility to continuously monitor brain states including e.g. human mental workload (Glatz et al. 2017), emotions (Liu and Sourina 2014) and stress levels (Sulaiman et al. 2012). EEG features in the time-domain, frequency-domain and synchronicity-domain can be used to detect and assess human stress levels. Time-domain features capture temporal information using the amplitude. With a focus on mental demand significant correlations between levels of psychological stress and EEG power have been shown (Seo and Lee 2010). Frequency-domain features are obtained from the frequency bands, such as delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (14-30 Hz) and gamma (30-50 Hz) (Gill and Singh 2021). For example, psychological stress is positively correlated with beta EEG power at anterior temporal lobe (Seo and Lee 2010). For collecting this data, we apply an ear EEG, which can be worn on the head of the drivers during the entire tour without any problems. |
| ECG and HRV | If a person enters a stressful situation, the brain and the cardiovascular system react to it. The cardiovascular system is controlled by the autonomic nervous system (ANS) and this is controlled by the brain (Kaur et al. 2015). Underlying the ANS are two subsystems with opposing roles. The Sympathetic Nervous System (SNS), which responds during tension, and the Parasympathetic Nervous System (PNS), which responds during relaxation (Londhe and Atulkar 2018; Taelman et al. 2011). Depending on how well or poorly the heart can respond to incoming stimuli from the ANS, HRV varies (Acharya et al. 2006). From a meta-analysis, HRV is an objective measure in measuring stress and strain (Kim et al. 2018). This marker was even able to show differences in the types of stress, namely whether it was physical or mental or a combination of both (Taelman et al. 2011). Further, HRV has been shown to be associated with specific cortical regions that are relevant in the assessment of stress (Thayer et al. 2012). The suitability of HRV as a marker of occupational stress has also been demonstrated (Järvelin-Pasanen, Sinikallio and Tarvainen 2018). The interaction of occupational stress and lower activation of the PNS could also be demonstrated, supporting the quality of the marker (Clays et al. 2011; Collins and Karasek 2010). HRV thus reflects the collaboration of the heart and brain and thus represents a neuro-cardiac function (Londhe and Atulkar 2018). Therefore, we measure HRV of drivers during tours. |
| Cortisol | During a stressful event, the steroid hormone cortisol is released (Aguilar Cordero et al. 2014). Due to the fact that it is a steroid, the hormone can be measured in all body fluids, including saliva (Smyth et al. 2013). Cortisol is one of the best-known hormones in stress measurement because it responds directly to stress and strain (Clow and Hamer 2010). In healthy individuals, however, cortisol levels are subject to certain fluctuations (Edwards et al. 2001), and for this reason we collect a saliva sample on several occasions. It has been shown that the release of cortisol is a reliable stress marker (Aguilar Cordero et al. 2014), it might even be a better indicator than heart rate and HRV (Nomura et al. 2009). |
| Attention | Another indicator to be considered is attention. When attention is heightened, the mental state shows increased alertness and receptivity. These are factors that are controlled both volitionally and involuntarily (Matthews et al. 2000). In the process, selection is made as to what information is important and what is not important to the organism (Broadbent 1958.). When attention is heightened, the overall responsiveness of the individual also increases. Stress triggers a reaction of the cognitive system in the sense that it increases attention for the brief moment by filtering out mainly unimportant information (Schaub 2012). In relation to the considered group of professional drivers, the decreasing attention with the duration of the working time might lead to wrong behaviour, which causes accidents. |
| Subjective data | Unpleasant situations – including social circumstances – can be reflected individually on a physical and on a psychological level (Evers 2009). The terms stress and strain can be combined in a so-called stress-strain concept (Romert and Rutenfanz 1975). This states in the work-scientific sense that stresses are factors that act on people from the outside. These can be of a physical, mental or psychosocial nature. Strain is then the effect of the stresses on an individual. The relationship between stress and strain is influenced by situational and personal factors, but also by subjective perception. This means that identical stressors can result in different strain levels for different people and that stressors of different types and severity can also result in the same strain level. For this reason, questions like family circumstances, housing situation, level of education, occupation and working conditions are recorded. |
| Chrono-type | The effects of stress are very diverse. Among other things, one reaction to perceived stress can be fatigue (Borbély and Achermann 1992). However, it has also been shown that circadian rhythm disturbances are a factor in increased daytime sleepiness (Laube et al. 2015). Sleepiness can be induced by monotonous tasks such as, in this context, long vehicle journeys (Hartmann 1980). For this reason, the group of professional drivers is particularly affected by the risk of daytime sleepiness. There have been a large number of accidents on highways caused by excessive sleepiness. Individuals who work shifts in which their own circadian rhythm is opposite often suffer from increased sleepiness because their bodies are normally attuned to sleep (Monk and Buysse 2013). This working against one's circadian rhythm represent a temporal stressor. Should these phases accumulate, sleepiness assumes an increased magnitude (Münch, Cajochen and Wirz-Justice 2005). The central nervous activation pattern decreases, lowering conscious control. |

| | |
|---|---|
| Impul-sivity | A literature review examined the relationship between aggressive behaviour and problem driving. Both clinical and nonclinical studies were included. The characteristic of impulsivity was also considered. A study was able to confirm the relationship (Witthöft, Hofmann and Petermann 2011). A study by Raio et al. (2013) was able to show that under stressful situations, emotion regulation could not take place as well as in a control condition. This is probably because the prefrontal cortex and an intact executive function, i.e., action control and behavioural control, must be functional in cognitive regulation and these are damaged under stressful conditions. This dysregulation of the prefrontal cortex and thus of executive functions is caused on the one hand by biological factors and on the other hand by the environment. |
| Operating data | There are a number of factors that can trigger stress reactions in professional drivers from a technical background. In 1980, Hartmann was already able to point out a number of factors (Hartmann 1980). He was able to show, for example, that driving in city traffic exerts a higher stress on drivers than driving on the highway. This is due to the acceleration and braking processes, the numerous risky situations and the presence of a flood of environmental stimuli. For this reason, vehicle data will also be included in the analyses. |
| Environ-mental data | Environmental factors are those factors which can be influenced by the drivers and those which cannot be influenced. An example of a factor that can be influenced is temperature. Extreme temperatures are perceived as a burden by the group of professional drivers, but they can regulate this themselves as long as the technology works (Plänitz 1983). Moreover, there are also factors which cannot be influenced. These include, for example, weather conditions, which can only be reacted to to a limited extent, and the traffic situation, which can also be influenced only to a limited extent. In their 2002 study, Ellinghaus and Steinbrecher (2002) were able to show that slippery roads are perceived as unpleasant by 94% of professional drivers, followed by fog with 90%. Traffic conditions also showed extreme stress: 56% of drivers perceive dense traffic as very stressful, half of them find traffic jams unpleasant. |

**Table 1. Indicator Data for Driver Stress Analysis**

In the research project case addressed, an implementation in six steps is planned (see Figure 3). The process of implementation is shown schematically in the following diagram. The study begins by obtaining consent from the drivers to participate. In the second stage, a large data set consisting of qualitative and quantitative data is collected, including both objective and subjective measurements. An online questionnaire will be conducted in which the items demographics, education, family situation, housing situation, work activity, and current sleep will be highlighted. Subsequently, the questions from the Morningness-Eveningness-Questionnaire are asked in a German translation (D-MEQ) (Griefahn et al. 2001) to assess the subjective circadian phasing of the drivers – the chronotype. Following is Barratt's impulsivity scale (Patton, Stanford and Barratt 1995). This is the most widely used scale for self-assessment of one's impulsivity (Stanford et al. 2009). In order to consider or exclude health influences or influences due to the use of alcohol, medication, drugs, etc., questions about health are asked. After that, the collection of objective data begins. In order to measure the attention of the drivers before starting to drive, the Frankfurt Attention Inventory (FAIR-2) in a revised version is used (Moosbrugger and Oehlschlägel 2011; Petermann 2011). Subsequently, the first saliva sample is taken by salivette to determine the cortisol content before the shift. Afterwards, the measuring instruments are put on to record the physiological and neurological reactions during the journey. The drivers start their tours and give a saliva sample again in the middle of the driving time. In addition, special incidents during the tour are documented by the drivers themselves, what helps to identify possible causes of fluctuations and interruptions in signal quality. After a tour, the fourth step of the procedure begins, a third saliva sample is taken and a further attention test is performed.

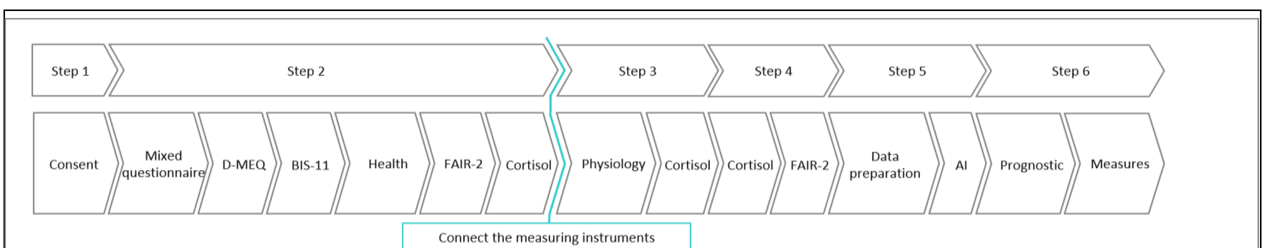


**Figure 3. Data Acquisition Process (Use Case Example)**

Using our model for decision support, we can now choose an appropriate data synchronization strategy. First, we assessed the criticality of temporal synchronization for our data collection. As we use physiological and neurological measurement devices working with high sampling rates (250-500 Hz), synchronization becomes necessary for useful data analysis. As we use several independent devices, we cannot rely on already synchronized internal device clocks. Nevertheless, medical devices should maintain standards regarding minimal internal clock drift i.e., differences of less milliseconds or seconds for a whole day. Beside the medical instruments, we gather further data from smartwatch and smartphone including GPS and driving data as well as measures of activity like the step count or body rotation. The medical devices are closed systems and we cannot implement additional time synchronization interfaces on the device e.g., for the NTP. Smartwatches allow the installation of custom apps but the programming interfaces available to developers offer only limited access e.g., to low-level network protocols as UDP/IP or TCP/IP. Nevertheless, as smartwatches measure several aspects at once, these data series are already internally time synchronized as the same device clock was used. Thereby, the requirement in our setting is to temporally synchronize all used devices assuming correct intra-device alignment (see Figure 4).

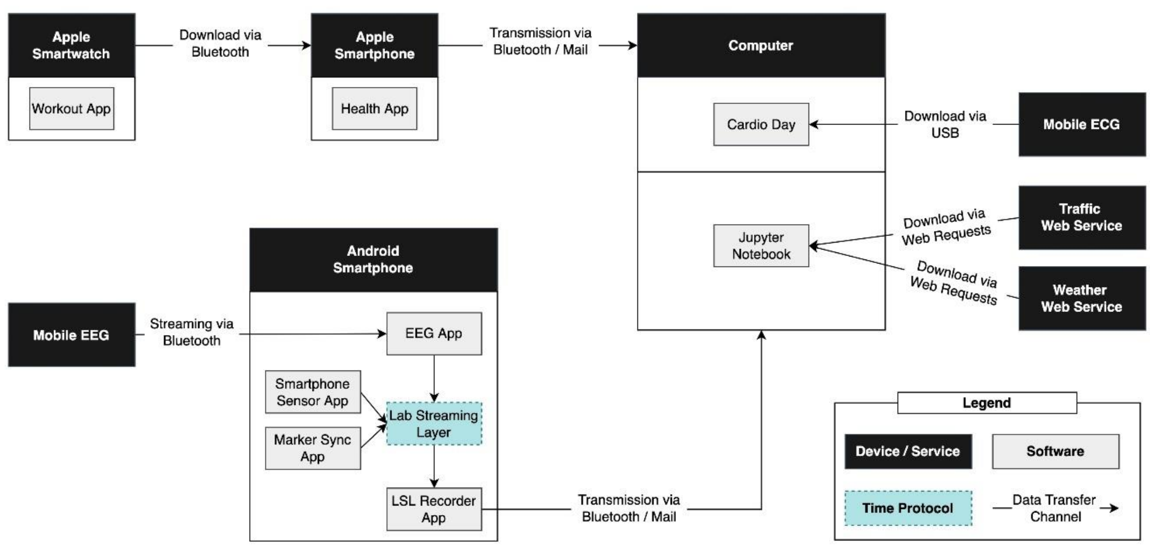


**Figure 4. Measurement Setup Architecture**

The next question in our decision support model is about consequences of imperfect synchronization. The internal clock drift of medical devices is assumed to be neglectable. The smartwatch acquires GPS signals that are coupled with high accuracy timestamps. Smartphones use the highly precise cellular network clock as point of reference. Therefore, we assume a low internal clock drift over the day. We want to measure comprehensive working days between 8 and 12 hours. Thereby, the relevant clock drift is only half or a third of the complete day drift and post-acquisition time shifting might be acceptable. As not all devices offer the streaming capabilities, we cannot make use of live time synchronization during the data collection using protocols as the LSL. Nevertheless, no network latencies must be considered as potential sources of inaccuracy. Our analysis focuses on the identification of factors for stress and strain of professional drivers. As pointed out before, stress is a phenomenon that typically remains for a longer period and so we are interested in time spans and no precise moments in time. Nevertheless, to map the start point of stress to the driver's context, the timing requirement is assumed to be less seconds. Our decision model supports in choosing an appropriate time synchronization strategy for the data collection. We measure data in a field setting and as driving is a safety-critical task, the measurement devices must not influence the driver during his work. Devices shall work without any cables as they might hinder the driver. Besides, we want to measure a time span of 8 to 12 hours and so, the battery life of the devices must be saved. We cannot assume that drivers change or charge batteries by their own. In addition, permanent data streaming over a wireless network of all involved devices might be no option for this setting due to the battery demand. As many drivers are required to load and unload their trucks manually, the measurement devices must be robust and shock insensitive. For our setting, we decide for a pre-post-stimulus synchronization strategy combined with live time synchronization using the LSL. The stimulus strategy fits best for devices with a low internal clock drift, constraints regarding battery and closed system as well as accuracy requirements on the level of seconds and time span analysis. The stimulus itself will be either digital or analogue. The time synchronization takes place after the measurements ends: the unsynchronized data streams are shifted until their marker positions match. To minimize the synchronization efforts, we utilize the LSL for as much data sources as possible. EEG data is acquired in a smartphone app, streamed into the LSL and synchronized with time marker data as well as smartphone sensor data. Although these sources origin from one smartphone device, using the LSL provides the flexibility, e.g., to generate marker data on a different device or integrate further sensor data sources like an eye tracker. For the data architecture, the smartphone clock is used as reference as this device is the source for several data streams and we assume that the GPS or cellular network-based clock is sufficiently precise. All other data sources are aligned to this reference clock.

# Contributions, Limitations and Outlook

## *Conceptual Contributions*

This paper makes a theoretical contribution by serving as an extension to existing data analytics frameworks from the IS domain focusing on time synchronization as part of the data acquisition, e.g. (Phillips-Wren et al. 2015). The objective is to raise awareness for time synchronization issues regarding sensor-based data collection. We contribute by providing a topic overview regarding time synchronization and by deriving a high-level concept model serving as a starting point for researchers and practitioners unfamiliar with the issue. Furthermore, we apply and refine this concept model by considering a specific use case for time synchronization: truck driver work stress and strain analysis. From a conceptual point of view, our research relates to well-established models in data science (Shafique and Qaiser 2014) like Cross-Industry Standard Process for Data Mining CRISP-DM (Wirth and Hipp 2000) and Knowledge Discovery in Databases KDD (Fayyad, Piatetsky-Shapiro and Smyth 1996). However, these models are often applied based on pre-existing data (Wiemer, Drowatzky and Ihlenfeldt 2019). In the practice-oriented CRISP-DM, the initial data acquisition is not a focal part and is hidden in a "data understanding" step and the "database" in the center, leading to researchers extending the model (Martinez-Plumed et al. 2021). There are also specific models for sensor-based data such as Automated PRE-Processing for Data Mining APREP-DM (Nagashima and Kato 2019). For data requiring temporal synchronization, we argue that dealing with issues a-priori is crucial for fully realizing the potential of data analysis in the case of sensor-based data. Figure 5 provides an extended concept draft for data integration and synchronization issues in big data and AI analyses: While applying our framework to the drivers use case, we gathered multiple learnings. The data collection from various devices required an iterative refinement of the setup architecture. In our case, the data set consisted of many different time formats with varying precision and points of reference (see red notes in Figure 5).

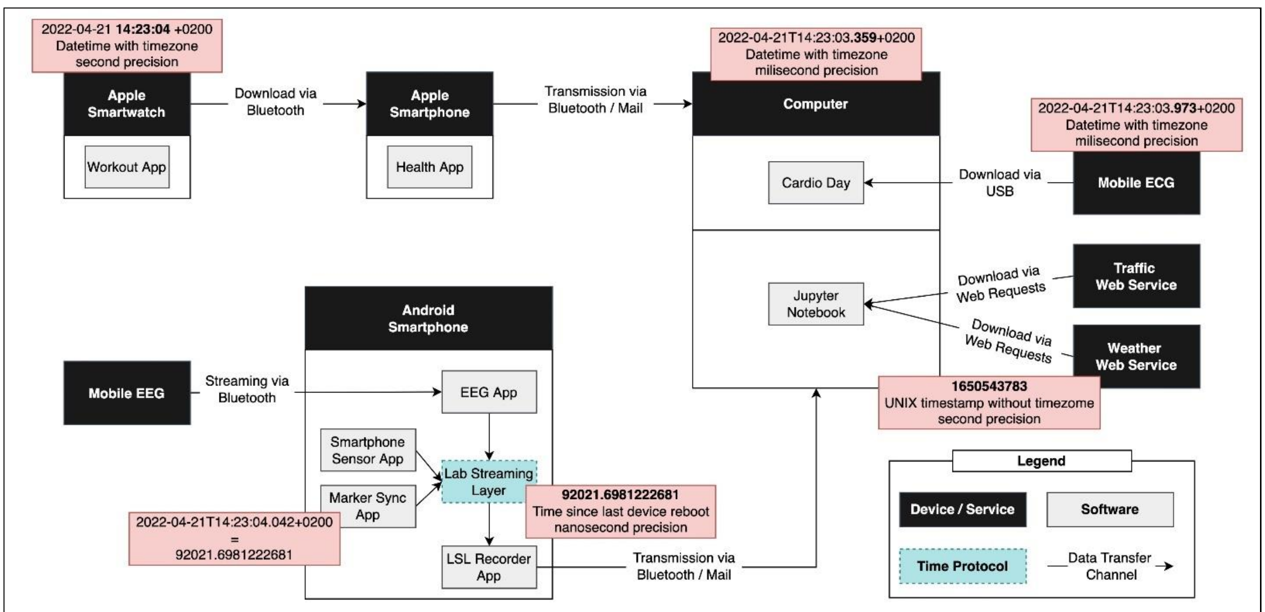


**Figure 5. Data Synchronization Issues Concept**

Generally, time formats can either use an absolute or relative point of reference. Absolute time might be the wall clock time or in case of UNIX timestamps, the seconds from 1970-01-01. Relative timestamps were used inside the LSL and referenced the recording device uptime which was reset after a device restart due to low energy. Without proper data inspection or preventive time synchronization strategies like pre-post stimulus (see the red note at "Marker Sync App" in Figure 5), these inconsistent timestamps might hinder data analysis. Another factor in time synchronization is the used time format. Some devices have explicit time zone information included while others require manual insertion. This fact might become problematic e.g., during changes to daylight saving time in summer or truck rides crossing time zones. Precision of time formats might also complicate proper data handling. Our medical devices captured data 500 times a second, while weather data was only captured few times an hour. It must be judged per data type if a single data point only counts for the capturing time itself or also fits for a time span. Gathering the driving direction every 10 seconds might be handled different than traffic density due to their varying change frequency. Regarding our measurement architecture, the portability requirement of the devices became more demanding than initially assumed. During truck unloading, the EEG wireless connection to the acquisition tablet disconnected several times without automatic reconnection and the data set became unusable. We extracted three learnings: early observations of a typical driver's workday would have influenced the setup from the very beginning. Second, data measurements in everyday settings set high requirements on reliability, and failure-tolerance and the acquisition software should be hardened to handle e.g., disconnections. Our last learning refers to the portability requirement. If the tested subjects cannot be observed during their whole workday, fitness tracker might give valuable insights on the activity level and the required setup flexibility. In different scenarios, the overarching quality requirement might be changed: collecting data in smart cities might require less portability and energy efficiency but more scalability to handle vast amounts of collecting devices. Nevertheless, issues like heterogeneous time formats and sampling rates or differing internal device clocks are likely to occur in every scenario.

## *Application Contributions*

The drafted data integration and synchronization framework could be applied to other use cases aiming at synchronizing different data, especially in a Big Data and AI application context. This could refer either to a transfer of measuring stress and strain in different work settings – with complex and diverse data source structures as in the described use case of truck drivers. Specifically, for truck drivers, a smartphone application could integrate all data and provide individual suggestions, e.g. to make a break, based on the analysis of the Big Data approach described and then subsequently working with smaller datasets. In a larger transfer perspective, the detailed data sources approach could be used to develop chatbots for work environment support, similarly requiring the analysis of a diversity of relevant data sources. For the continuation of the operations and logistics context presented herein, Big Data analytics and AI tools could be developed to manage and coordinate processes in networks on a large and complex supply chain scale. Regarding generalizability, the findings presented are expected to further support research in contexts of medical applications to ensure data quality and else, such as security management (e.g., traffic safety or societal security) and disaster management in crises with crucial relevance of data time synchronization.

## *Limitations and Outlook*

An important limitation must be considered: AI development requires an experimental approach, meaning that one cannot state before training whether results will be achieved as intended. While ensuring data quality as described is an important step towards functioning AI-models, still no guarantee can be given that a trained model will perform as desired. However, we argue that considering all discussed aspects will yield a deeper understanding of the underlying data and thus support development efforts. For the same reasons, developers face a dilemma: While a dedicated data collection requires effort, it cannot be determined in advance how strong the effect of, e.g., network latencies during data collection on the trained model will be. Maybe the desired patterns in the data are strong enough to train robust algorithms. Thus, a dedicated weighing of options is required regarding the resulting costs, i.e., whether collecting new data in case model training fails is easier than going through all the effort to ensure data validity. We see this as a limitation to our approach, because answering this question depends on application specifics that – from our current experiences – cannot be generalized. Nevertheless, we think that our proposed approach can support developers by creating awareness for potential problems and thus improve development efforts.

Furthermore, our approach is based on recent literature and the set of sensors we use for acquiring data. We cannot exclude that other hardware (other sensors or the same sensors from different vendors) will require some adaptations or will introduce new aspects. This study set out to explore the work-related stress and strain that truck drivers experience in their daily lives. We argue that to understand the root causes of stress better, we need to access additional data sources. The arising challenge of data integration and synchronization among different data sources is pertinent to understanding causal relationships. Based on a detailed account of integration synchronization challenges, we introduce the concept of lab streaming layers and derive at a high-level framework for solving time synchronization issues in data collection. We adopt this framework to a real-world case with stress and strain analyses of truck drivers, leading to a revised data integration and synchronization framework. This allows researchers and practitioners alike to address integration and synchronization issues. The framework provides a solid ground for AI-based Big Data endeavours and a valid contribution to better understand stress and its root causes. This framework can inform multiple stakeholders and levels, from driving assistant systems to work-related optimizations in different fields and industries.

## References


Aguilar Cordero, M. J., Sánchez López, A. M., Mur Villar, N., García García, I., Rodríguez López, M. A., Ortegón Piñero, A., and Cortés Castell, E. 2014. "Cortisol salival como indicador de estrés fisiológico en niños y adultos; revisión sistemática," *Nutricion Hospitalaria* (29:5), pp. 960-968.

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., and Zimmermann, T. 2019. "Software Engineering for Machine Learning: A Case Study," in *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP),* Montreal, QC, Canada. 25.05.2019-31.05.2019, IEEE, pp. 291-300.

Artoni, F., Barsotti, A., Guanziroli, E., Micera, S., Landi, A., and Molteni, F. 2017. "Effective Synchronization of EEG and EMG for Mobile Brain/Body Imaging in Clinical Settings," *Frontiers in Human Neuroscience* (11), p. 652.

Artoni, F., Galeasso, E., and Micera, S. 2019. "P300 in the park: feasibility of online data acquisition and integration in a Mobile Brain/Body Imaging setting," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER),* San Francisco, CA, USA. 20.03.2019-23.03.2019, IEEE, pp. 319-322.

Aygün, B., and Cagri Gungor, V. 2011. "Wireless sensor networks for structure health monitoring: recent advances and future research directions," *Sensor Review* (31:3), pp. 261-276.

Baldoni, R., and Raynal, M. 2002. "Fundamentals of distributed computing: A practical tour of vector clock systems," *IEEE Distributed Systems Online* (3:2), p. 12.

Bethge, D., Kosch, T., Grosse-Puppendahl, T., Chuang, L. L., Kari, M., Jagaciak, A., and Schmidt, A. 2021. "VEmotion: Using Driving Context for Indirect Emotion Prediction in Real-Time," in *The 34th Annual ACM Symposium on User Interface Software and Technology*, J. Nichols, R. Kumar and M. Nebeling (eds.), Virtual Event USA. 10.10.2021-14.10.2021, New York, NY, USA: ACM, pp. 638-651.

Bleichner, M. G., and Debener, S. 2017. "Concealed, Unobtrusive Ear-Centered EEG Acquisition: cEEGrids for Transparent EEG," *Frontiers in Human Neuroscience* (11), p. 163.

Bleichner, M. G., and Emkes, R. 2020. "Building an Ear-EEG System by Hacking a Commercial Neck Speaker and a Commercial EEG Amplifier to Record Brain Activity Beyond the Lab," *Journal of Open Hardware* (4:1).

Blum, S., Hölle, D., Bleichner, M. G., and Debener, S. 2021. "Pocketable Labs for Everyone: Synchronized Multi-Sensor Data Streaming and Recording on Smartphones with the Lab Streaming Layer," *Sensors* (21:23), p. 8135.

Borbély, and Achermann. 1992. "Concepts and models of sleep regulation: An overview," *Journal of Sleep Research* (1:2), pp. 63-79.

Brahmi, M., Schueler, K., Bouzouraa, S., Maurer, M., Siedersberger, K.-H., and Hofmann, U. 2013. "Timestamping and latency analysis for multi-sensor perception systems," in *2013 IEEE SENSORS,* Baltimore, MD, USA. 03.11.2013-06.11.2013, IEEE, pp. 1-4.

Broadbent, D. E. 1958. *Perception and Communication*, Cambridge: Pergamon.

Cai, M., Yazdi, M. A. A., Mehdizadeh, A., Hu, Q., Vinel, A., Davis, K., Xian, H., Megahed, F. M., and Rigdon, S. E. 2021. "The association between crashes and safety-critical events: Synthesized evidence from crash reports and naturalistic driving data among commercial truck drivers," *Transportation Research Part C: Emerging Technologies* (126), p. 103016.

Calyam, P., Mishra, A., Antequera, R. B., Chemodanov, D., Berryman, A., Zhu, K., Abbott, C., and Skubic, M. 2016. "Synchronous Big Data analytics for personalized and remote physical therapy," *Pervasive and Mobile Computing* (28), pp. 3-20.

Castaño, F., Strzelczak, S., Villalonga, A., Haber, R. E., and Kossakowska, J. 2019. "Sensor Reliability in Cyber-Physical Systems Using Internet-of-Things Data: A Review and Case Study," *Remote Sensing* (11:19), p. 2252.
Clays, E., Bacquer, D. de, Crasset, V., Kittel, F., Smet, P. de, Kornitzer, M., Karasek, R., and Backer, G. de. 2011. "The perception of work stressors is related to reduced parasympathetic activity," *International Archives of Occupational and Environmental Health* (84:2), pp. 185-191.
Clow, A., and Hamer, M. 2010. "The iceberg of social disadvantage and chronic stress: implications for public health," *Neuroscience and Biobehavioral Reviews* (35:1), p. 1.
Collins, S., and Karasek, R. 2010. "Reduced vagal cardiac control variance in exhausted and high strain job subjects," *International Journal of Occupational Medicine and Environmental Health* (23:3), pp. 267-278.
Cristian, F. 1989. "Probabilistic clock synchronization," *Distributed Computing* (3:3), pp. 146-158.
Dorst, T., Robin, Y., Eichstädt, S., Schütze, A., and Schneider, T. 2021. "Influence of synchronization within a sensor network on machine learning results," *Journal of Sensors and Sensor Systems* (10:2), pp. 233-245.
Ellinghaus, D., and Steinbrecher, J. 2002. *LKW im Strassenverkehr. Eine Untersuchung über die Beziehung zwischen LKW und PKW-Fahrern*, Hannover.
Engelbrecht, J., Booysen, M. J., Rooyen, G.-J., and Bruwer, F. J. 2015. "Survey of smartphone-based sensing in vehicles for intelligent transportation system applications," *IET Intelligent Transport Systems* (9:10), pp. 924-935.
Esco, M. R., Flatt, A. A., and Nakamura, F. Y. 2017. "Agreement Between a Smartphone Pulse Sensor Application and Electrocardiography for Determining lnRMSSD," *Journal of Strength and Conditioning Research* (31:2), pp. 380-385.
Evers, C. 2009. *Auswirkungen von Belastungen und Stress auf das Verkehrsverhalten von Lkw-Fahrern*, University of Bonn (Dissertation): Bonn.
Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. 1996. "From Data Mining to Knowledge Discovery in Databases," *AI Magazine* (17:3), p. 37.
Fischer, T., Reuter, M., and Riedl, R. 2021. "The Digital Stressors Scale: Development and Validation of a New Survey Instrument to Measure Digital Stress Perceptions in the Workplace Context," *Frontiers in Psychology* (12), p. 607598.
Fridman, L., Brown, D. E., Angell, W., Abdić, I., Reimer, B., and Noh, H. Y. 2016. "Automated synchronization of driving data using vibration and steering events," *Pattern Recognition Letters* (75), pp. 9-15.
Frontoni, E., Paolanti, M., Lauriault, T. P., Stiber, M., Duranti, L., and Muhammad, A.-M. 2022. "Trusted Data Forever: Is AI the Answer?" *Workshop Proceedings EDBT/ICDT 2022,* Edinburgh, UK.
Ge, Y., Qu, W., Jiang, C., Du, F., Sun, X., and Zhang, K. 2014. "The effect of stress and personality on dangerous driving behavior among Chinese drivers," *Accident Analysis and Prevention* (73), pp. 34-40.
Georgieva, I., Lazo, C., Timan, T., and van Veenstra, A. F. 2022. "From AI ethics principles to data science practice: a reflection and a gap analysis based on recent frameworks and practical experience," *AI and Ethics*.
Gill, R., and Singh, J. 2021. "A Review of Feature Extraction Techniques for EEG-Based Emotion Recognition System," in *Soft Computing: Theories and Applications*, T. K. Sharma, C. W. Ahn, O. P. Verma and B. K. Panigrahi (eds.), Singapore: Springer, pp. 83-95.
Glatz, C., Ditz, J., Kosch, T., Schmidt, A., Lahmer, M., and Chuang, L. L. 2017. "Reading the mobile brain," in *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, N. Henze, P. Woźniak, K. Väänänen, J. Williamson and S. Schneegass (eds.), Stuttgart. 26.11.2017-29.11.2017, New York, NY, USA: ACM, pp. 573-579.
Griefahn, B., Kunemund, C., Brode, P., and Mehnert, P. 2001. "Zur Validitat der deutschen Ubersetzung des Morningness-Eveningness-Questionnaires von Horne und Ostberg. [The Validity of a German Version of the Morningness-Eveningness-Questionnaire Developed by Horne and Ostberg]," *Somnologie* (5:2), pp. 71-80.
Gusella, R., and Zatti, S. 1989. "The accuracy of the clock synchronization achieved by TEMPO in Berkeley UNIX 4.3BSD," *IEEE Transactions on Software Engineering* (15:7), pp. 847-853.
Han, W., and Zhao, J. 2020. "Driver behaviour and traffic accident involvement among professional urban bus drivers in China," *Transportation Research Part F: Traffic Psychology and Behaviour* (74), pp. 184-197.
Hartmann, H. P. 1980. "Streß und Übermüdung beim Fahrzeuglenker," in *Der Kranke als Fahrzeuglenker*, H. P. Hartmann (ed.), Berlin, Heidelberg: Springer, pp. 37-41.
Hege, A., Lemke, M. K., Apostolopoulos, Y., Whitaker, B., and Sönmez, S. 2019. "Work-Life Conflict among U.S. Long-Haul Truck Drivers: Influences of Work Organization, Perceived Job Stress, Sleep, and Organizational Support," *International Journal of Environmental Research and Public Health* (16:6), p. 984.
Hernando, D., Roca, S., Sancho, J., Alesanco, Á., and Bailón, R. 2018. "Validation of the Apple Watch for Heart Rate Variability Measurements during Relax and Mental Stress in Healthy Subjects," *Sensors* (18:8), p. 2619.
Hesenius, M., Schwenzfeier, N., Meyer, O., Koop, W., and Gruhn, V. 2019. "Towards a Software Engineering Process for Developing Data-Driven Applications," in *2019 IEEE/ACM 7th International Workshop on*

*Realizing Artificial Intelligence Synergies in Software Engineering (RAISE),* Montreal, QC, Canada. 28.05.2019-28.05.2019, IEEE, pp. 35-41.

Hill, J. D., and Boyle, L. N. 2007. "Driver stress as influenced by driving maneuvers and roadway conditions," *Transportation Research Part F: Traffic Psychology and Behaviour* (10:3), pp. 177-186.

Hou, X., Liu, Y., Sourina, O., Tan, Y. R. E., Wang, L., and Mueller-Wittig, W. 2015. "EEG Based Stress Monitoring," in *2015 IEEE International Conference on Systems, Man, and Cybernetics,* Hong Kong, China. 09.10.2015-12.10.2015, IEEE, pp. 3110-3115.

Hutson, M. 2018. "Artificial intelligence faces reproducibility crisis," *Science* (359:6377), pp. 725-726.

IEEE. 2008. "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems: IEEE Std 1588-2008 (Revision Std 1588-2002)," *IEEE Standards,* IEEE (ed.), IEEE.

Järvelin-Pasanen, S., Sinikallio, S., and Tarvainen, M. P. 2018. "Heart rate variability and occupational stress-systematic review," *Industrial Health* (56:6), pp. 500-511.

Karolus, J., Kilian, A., Kosch, T., Schmidt, A., and Wozniak, P. W. 2020. "Hit the Thumb Jack! Using Electromyography to Augment the Piano Keyboard," in *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, R. Wakkary, K. Andersen, W. Odom, A. Desjardins and M. G. Petersen (eds.), Eindhoven Netherlands. 06.07.2020-10.07.2020, New York, NY, USA: ACM, pp. 429-440.

Karolus, J., Schuff, H., Kosch, T., Wozniak, P. W., and Schmidt, A. 2018. "EMGuitar," in *Proceedings of the 2018 Designing Interactive Systems Conference*, I. Koskinen, Y. Lim, T. Cerratto-Pargman, K. Chow and W. Odom (eds.), Hong Kong, China. 09.06.2018-13.06.2018, New York, NY, USA: ACM, pp. 651-655.

Katmah, R., Al-Shargie, F., Tariq, U., Babiloni, F., Al-Mughairbi, F., and Al-Nashash, H. 2021. "A Review on Mental Stress Assessment Methods Using EEG Signals," *Sensors* (21:15), p. 5043.

Kilian, A., Karolus, J., Kosch, T., Schmidt, A., and Paweł, P. W. 2021. "EMPiano: Electromyographic Pitch Control on the Piano Keyboard," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, Y. Kitamura, A. Quigley, K. Isbister and T. Igarashi (eds.), Yokohama, Japan. 08.05.2021-13.05.2021, New York, NY, USA: ACM, pp. 1-4.

Kosuru, R. K., Lingelbach, K., Bui, M., and Vukelić, M. 2019. "MindTrain," in *Proceedings of Mensch und Computer 2019*, F. Alt, A. Bulling and T. Döring (eds.), Hamburg, Germany. 08.09.2019-11.09.2019, New York, NY, USA: ACM, pp. 643-647.

Kühl, N., Hirt, R., Baier, L., Schmitz, B., and Satzger, G. 2021. "How to Conduct Rigorous Supervised Machine Learning in Information Systems Research: The Supervised Machine Learning Report Card," *Communications of the Association for Information Systems* (48:1), pp. 589-615.

Kühnisch, J., Meyer, O., Hesenius, M., Hickel, R., and Gruhn, V. 2022. "Caries Detection on Intraoral Images Using Artificial Intelligence," *Journal of Dental Research* (101:2), pp. 158-165.

Lamport, L. 1978. "Time, clocks, and the ordering of events in a distributed system," *Communications of the ACM* (21:7), pp. 558-565.

Laube, E., Bonetti, N. R., Bächli, E., and Beer, J. 2015. "Die etwas andere Chefarztumfrage: Somatische Ursachen der Müdigkeit," *Swiss Medical Forum – Schweizerisches Medizin-Forum* (15:17), pp. 374-381.

Lin, K., Kansal, A., Lymberopoulos, D., and Zhao, F. 2010. "Energy-accuracy trade-off for continuous mobile device location," in *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services - MobiSys '10*, S. Banerjee, S. Keshav and A. Wolman (eds.), San Francisco, CA, USA. 15.06.2010-18.06.2010, New York, NY, USA: ACM Press, p. 285-298.

Liskov, B., and Ladin, R. 1986. "Highly available distributed services and fault-tolerant distributed garbage collection," in *Proceedings of the Fifth Annual ACM Symposium on Principles of Distributed Computing - PODC '86*, J. Halpern (ed.), Calgary, Alberta, Canada. 11.08.1986-13.08.1986, New York, NY, USA: ACM Press, pp. 29-39.

Liu, Y., and Sourina, O. 2014. "Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm," in *Transactions on Computational Science XXIII*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, M. L. Gavrilova, C. J. K. Tan, X. Mao and L. Hong (eds.), Berlin, Heidelberg: Springer, pp. 199-223.

Londhe, A. N., and Atulkar, M. 2018. "Heart Rate Variability Analysis: Application Overview," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT),* Coimbatore. 20.04.2018-21.04.2018, IEEE, pp. 1518-1523.

Lundelius, J., and Lynch, N. 1984. "An Upper and Lower Bound for Clock Synchronization," *Information and Control* (62), pp. 190-204.

Maksimovic, M., Vujovic, V., and Perisic, B. 2015. "A custom Internet of Things healthcare system," in *2015 10th Iberian Conference on Information Systems and Technologies (CISTI),* Aveiro, Portugal. 17.06.2015-20.06.2015, IEEE, pp. 1-6.

Manjunatha, H., Distefano, J. P., Jani, A., Ghassemi, P., Chowdhury, S., Dantu, K., Doermann, D., and Esfahani, E. T. 2020. "Using Physiological Measurements to Analyze the Tactical Decisions in Human Swarm Teams," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* Toronto, ON, Canada. 11.10.2020-14.10.2020, IEEE, pp. 256-261.

Mantouka, E., Barmpounakis, E., Vlahogianni, E., and Golias, J. 2021. "Smartphone sensing for understanding driving behavior: Current practice and challenges," *International Journal of Transportation Science and Technology* (10:3), pp. 266-282.

Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., and Flach, P. 2021. "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Transactions on Knowledge and Data Engineering* (33:8), pp. 3048-3061.

Matthews, G., Davies, D. R., Stammers, R. B., and Westerman, S. J. 2000. *Human Performance: Cognition, Stress, and Individual Differences*, Hove: Psychology Press.

McLennan, S., Fiske, A., Tigard, D., Müller, R., Haddadin, S., and Buyx, A. 2022. "Embedded ethics: a proposal for integrating ethics into the development of medical AI," *BMC Medical Ethics* (23:6).

Mendonca, P., and Abreu, M. A. 2019. "A Hybrid System for Assessing Mental Workload," in *2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG),* Lisbon, Portugal. 22.02.2019-23.02.2019, IEEE, pp. 1-4.

Mills, D. L. 1991. "Internet time synchronization: the network time protocol," *IEEE Transactions on Communications* (39:10), pp. 1482-1493.

Monk, T. H., and Buysse, D. J. 2013. "Exposure to shift work as a risk factor for diabetes," *Journal of biological rhythms* (28:5), pp. 356-359.

Moosbrugger, H., and Oehlschlägel, J. 2011. *Frankfurter Aufmerksamkeits-Inventar 2 FAIR-2*.

Morabit, M., Desaulniers, G., and Lodi, A. 2021. "Machine-Learning–Based Column Selection for Column Generation," *Transportation Science* (55:4), pp. 815-831.

Moraru, A., Pesko, M., Porcius, M., Fortuna, C., and Mladenic, D. 2010. "Using Machine Learning on Sensor Data," *Journal of Computing and Information Technology* (18:4), p. 341.

Münch, M., Cajochen, C., and Wirz-Justice, A. 2005. "Schlaf und zirkadiane Rhythmik im Alter," *Zeitschrift fur Gerontologie und Geriatrie* (38), pp. i21-i23.

Nagashima, H., and Kato, Y. 2019. "APREP-DM: A Framework for Automating the Pre-Processing of a Sensor Data Analysis based on CRISP-DM," in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops),* Kyoto, Japan. 11.03.2019-15.03.2019, IEEE, pp. 555-560.

Neagoe, T., Cristea, V., and Banica, L. 2006. "NTP versus PTP in Computer Networks Clock Synchronization," in *2006 IEEE International Symposium on Industrial Electronics,* Montreal, Quebeck, Canada. 09.07.2006-13.07.2006, IEEE, pp. 317-362.

Nomura, S., Mizuno, T., Nozawa, A., Asano, H., and Ide, H. 2009. "Salivary Cortisol as a New Biomarker for a Mild Mental Workload," in *2009 International Conference on Biometrics and Kansei Engineering,* Cieszyn, Poland. 25.06.2009-28.06.2009, IEEE, pp. 127-131.

Olson, E. 2010. "A passive solution to the sensor synchronization problem," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems,* Taipei, Taiwan. 18.10.2010-22.10.2010, IEEE, pp. 1059-1064.

Onninen, J., Pylkkönen, M., Hakola, T., Puttonen, S., Virkkala, J., Tolvanen, A., and Sallinen, M. 2021. "The self-reported causes of sleepiness in shift-working tram and truck drivers," *Transportation Research Part F: Traffic Psychology and Behaviour* (78), pp. 153-163.

Ostrin, G., Frey, J., and Cauchard, J. R. 2018. "Interactive Narrative in Virtual Reality," in *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, S. Abdennadher and F. Alt (eds.), Cairo Egypt. 25.11.2018-28.11.2018, New York, NY, USA: ACM, pp. 463-467.

Öz, B., Özkan, T., and Lajunen, T. 2010. "Professional and non-professional drivers' stress reactions and risky driving," *Transportation Research Part F: Traffic Psychology and Behaviour* (13:1), pp. 32-40.

Patton, J. H., Stanford, M. S., and Barratt, E. S. 1995. "Factor structure of the Barratt impulsiveness scale," *Journal of Clinical Psychology* (51:6), pp. 768-774.

Petermann, F. 2011. "Frankfurter Aufmerksamkeits-Inventar 2 (FAIR-2)," *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* (59:4), pp. 325-326.

Phillips-Wren, G., Iyer, L. S., Kulkarni, U., and Aricachandra, T. 2015. "Business Analytics in the Context of Big Data: A Roadmap for Research," *Communications of the Association for Information Systems* (37).

Ping, S. 2003. "Delay Measurement Time Synchronization for Wireless Sensor Networks," IRB-TR-03-013, Berkeley: Intel Research.

Plänitz, G. 1983. *Das bißchen Fahren – Arbeits- und Lebensbedingungen von Fernfahrern*, Berlin: VSA.

Puttnies, H., Danielis, P., Sharif, A. R., and Timmermann, D. 2020. "Estimators for Time Synchronization—Survey, Analysis, and Outlook," *IoT* (1:2), pp. 398-435.

Raio, C. M., Orederu, T. A., Palazzolo, L., Shurick, A. A., and Phelps, E. A. 2013. "Cognitive emotion regulation fails the stress test," *Proceedings of the National Academy of Sciences of the United States of America* (110:37), pp. 15139-15144.

Rastgoo, M. N., Nakisa, B., Maire, F., Rakotonirainy, A., and Chandran, V. 2019. "Automatic driver stress level classification using multimodal deep learning," *Expert Systems with Applications* (138), p. 112793.

Reiser, J. E., Wascher, E., and Arnau, S. 2019. "Recording mobile EEG in an outdoor environment reveals cognitive-motor interference dependent on movement complexity," *Scientific Reports* (9:1), p. 13086.

Romert, W., and Rutenfanz, J. 1975. *Arbeitswissenschaftliche Beurteilung der Belastung und Beanspruchung an unterschiedlichen industriellen Arbeitsplätzen*, Bonn.

Rowden, P., Matthews, G., Watson, B., and Biggs, H. 2011. "The relative impact of work-related stress, life stress and driving environment stress on driving outcomes," *Accident Analysis and Prevention* (43:4), pp. 1332-1340.

Rozado, D., Niu, J., and Lochner, M. 2017. "Fast Human-Computer Interaction by Combining Gaze Pointing and Face Gestures," *ACM Transactions on Accessible Computing* (10:3), pp. 1-18.

Saidatul, A., Paulraj, M. P., Yaacob, S., and Yusnita, M. A. 2011. "Analysis of EEG signals during relaxation and mental stress condition using AR modeling techniques," in *2011 IEEE International Conference on Control System, Computing and Engineering,* Penang, Malaysia. 25.11.2011-27.11.2011, IEEE, pp. 477-481.

Santos, J. A., and Lu, J. L. 2016. "Occupational safety conditions of bus drivers in Metro Manila, the Philippines," *International Journal of Occupational Safety and Ergonomics* (22:4), pp. 508-513.

Sartori, C. S., Smet, P., and Vanden Berghe, G. 2021. "Scheduling truck drivers with interdependent routes under European Union regulations," *European Journal of Operational Research* (298:1), pp. 76-88.

Schaub, H. 2012. "Wahrnehmung, Aufmerksamkeit und »Situation Awareness« (SA)," in *Human Factors*, P. Badke-Schaub, G. Hofinger and K. Lauche (eds.), Berlin, Heidelberg: Springer, pp. 63-81.

Sekkay, F., Imbeau, D., Dubé, P.-A., Chinniah, Y., Marcellis-Warin, N. de, Beauregard, N., and Trépanier, M. 2021. "Assessment of physical work demands of long-distance industrial gas delivery truck drivers," *Applied Ergonomics* (90), p. 103224.

Seo, S.-H., and Lee, J.-T. 2010. "Stress and EEG," in *Convergence and Hybrid Information Technologies*, M. Crisan (ed.), London, UK: InTech, p. 426.

Shafique, U., and Qaiser, H. 2014. "A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)," *International Journal of Innovation and Scientific Research* (12:1), pp. 217-222.

Shattell, M., Apostolopoulos, Y., Sönmez, S., and Griffin, M. 2010. "Occupational stressors and the mental health of truckers," *Issues in Mental Health Nursing* (31:9), pp. 561-568.

Sivrikaya, F., and Yener, B. 2004. "Time synchronization in sensor networks: a survey," *IEEE Network* (18:4), pp. 45-50.

Skog, I., and Handel, P. 2008. "Effects of time synchronization errors in GNSS-aided INS," in *Effects of time synchronization errors in GNSS-aided INS*, I. Skog and P. Handel (eds.), Monterey, CA, USA. 05.05.2008-08.05.2008, IEEE, pp. 82-88.

Smyth, N., Hucklebridge, F., Thorn, L., Evans, P., and Clow, A. 2013. "Salivary Cortisol as a Biomarker in Social Science Research," *Social and Personality Psychology Compass* (7:9), pp. 605-625.

Son, K. J., Chang, T. G., and Kang, S.-H. 2019. "The Effect of Time Synchronization Error in LAN-Based Digital Substation," *Sensors* (19:9), p. 2044.

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., and Patton, J. H. 2009. "Fifty years of the Barratt Impulsiveness Scale: An update and review," *Personality and Individual Differences* (47:5), pp. 385-395.

Stepanova, E. R., Desnoyers-Stewart, J., Pasquier, P., and Riecke, B. E. 2020. "JeL," in *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, R. Wakkary, K. Andersen, W. Odom, A. Desjardins and M. G. Petersen (eds.), Eindhoven, Netherlands. 06.07.2020-10.07.2020, New York, NY, USA: ACM, pp. 641-654.

Sulaiman, N., Taib, M. N., Lias, S., Murat, Z. H., Mohd Aris, S. A., Mustafa, M., and Rashid, N. A. 2012. "Development of EEG-based stress index," in *2012 International Conference on Biomedical Engineering (ICoBE),* Penang, Malaysia. 27.02.2012 - 28.02.2012, IEEE, pp. 461-466.

Sundararaman, B., Buy, U., and Kshemkalyani, A. D. 2005. "Clock synchronization for wireless sensor networks: a survey," *Ad Hoc Networks* (3:3), pp. 281-323.

Taelman, J., Vandeput, S., Gligorijević, I., Spaepen, A., and van Huffel, S. 2011. "Time-frequency heart rate variability characteristics of young adults during physical, mental and combined stress in laboratory environment," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* (2011), pp. 1973-1976.

Thielmann, B., Pohl, R., and Böckelmann, I. 2021. "Heart rate variability as a strain indicator for psychological stress for emergency physicians during work and alert intervention: a systematic review," *Journal of Occupational Medicine and Toxicology* (16:1), p. 24.

Tirado-Andrés, F., Rozas, A., and Araujo, A. 2019. “A Methodology for Choosing Time Synchronization Strategies for Wireless IoT Networks,” *Sensors* (19:16), p. 3476.

Urbina, F., and Ekins, S. 2022. “The commoditization of AI for molecule design,” *Artificial Intelligence in the Life Sciences* (2), p. 100031.

Useche, S. A., Ortiz, V. G., and Cendales, B. E. 2017. “Stress-related psychosocial factors at work, fatigue, and risky driving behavior in bus rapid transport (BRT) drivers,” *Accident Analysis and Prevention* (104), pp. 106-114.

Wang, K.-J., Zhang, L., Luan, B., Tung, H.-W., Liu, Q., Wei, J., Sun, M., and Mao, Z.-H. 2017. “Brain-computer interface combining eye saccade two-electrode EEG signals and voice cues to improve the maneuverability of wheelchair,” *IEEE International Conference on Rehabilitation Robotics* 2017, pp. 1073-1078.

Wang, Q., Beardsley, V. J., Zhang, Q., Kim, K., and Barmaki, R. 2021a. “An LSL-Middleware Prototype for VR/AR Data Collection,” in *Symposium on Spatial User Interaction*, F. Ortega, R. Teather, G. Bruder, T. Piumsomboon, B. Weyers, A. U. Batmaz, K. Johnsen and C. Borst (eds.), Virtual Event. 09.11.2021-10.11.2021, New York, NY, USA: ACM, pp. 1-2.

Wang, X., Jiao, Y., Huo, J., Li, R., Zhou, C., Pan, H., and Chai, C. 2021b. “Analysis of safety climate and individual factors affecting bus drivers’ crash involvement using a two-level logit model,” *Accident Analysis and Prevention* (154), p. 106087.

Wascher, E., Reiser, J., Rinkenauer, G., Larrá, M., Dreger, F. A., Schneider, D., Karthaus, M., Getzmann, S., Gutberlet, M., and Arnau, S. 2021. “Neuroergonomics on the Go: An Evaluation of the Potential of Mobile EEG for Workplace Assessment and Design,” *Human Factors*.

Wiemer, H., Drowatzky, L., and Ihlenfeldt, S. 2019. “Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model,” *Applied Sciences* (9:12), p. 2407.

Wirth, R., and Hipp, J. 2000. “CRISP-DM: Towards a Standard Process Model for Data Mining,” Tübingen: University of Tübingen.

Witthöft, J., Hofmann, M., and Petermann, F. 2011. “Aggression im Strassenverkehr,” *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* (59:4), pp. 311-323.

Yigitler, H. N., Badihi, B., and Jäntti, R. 2020. “Overview of Time Synchronization for IoT Deployments: Clock Discipline Algorithms and Protocols,” *Sensors* (20:20), p. 5928.

Youn, S. 2013. “A Comparison of Clock Synchronization in Wireless Sensor Networks,” *International Journal of Distributed Sensor Networks* (9:12), p. 532986.