Behaviour & Information Technology

ISSN: 0144-929X (Print) 1362-3001 (Online) Journal homepage: https://www.tandfonline.com/loi/tbit20

Using biometric data in software engineering: a systematic mapping study

Juliano Paulo Menzen, Kleinner Farias & Vinicius Bischoff

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2020.1734086

Using biometric data in software engineering: a systematic mapping study

Juliano Paulo Menzen, Kleinner Farias and Vinicius Bischoff

Applied Computing Graduate Program (PPGCA) of University of Vale do Rio dos Sinos (UNISINOS), São Leopoldo, RS, Brazil

ABSTRACT
The use of biometric data (BD) records promises to advance the software engineering field. The rapid adoption of wearable computing technology has widely increased the amount of BD records available. Several aspects about the use of BD records in software engineering field are unknown, such as body measurements used to support daily tasks, and empirical methods that are used to evaluate their benefits. Consequently, a thorough understanding of state-of-the-art techniques still remains limited. This article, therefore, aims at providing a classification and a thematic analysis of studies on the use of BD records in the context of software development. Moreover, it seeks to introduce a classification taxonomy, and pinpoints research gaps, challenges and trends. A systematic mapping of the literature was designed and performed based on well-established practical guidelines. In total, 40 primary studies were analysed and categorised, which were selected by applying a careful filtering process from a sample of 3930 studies to answer seven research questions. Over 77% of articles use more than one biometric aspect to analyse tasks performed by developers; over 47% of articles used eye-track sensor to analyse biometric factors, followed by brain-wearable sensors with 40%, skin sensor with 22%, cardiac sensor with 20%, and others fewer representatives; most studies analysed had their quality assessed as high; most studies were published in journal. This study provides a systematic map of studies that use BD records in software engineering, thereby serving as a basis for future research.

ARTICLE HISTORY
Received 10 August 2019
Accepted 18 February 2020

KEYWORDS
Biometric; psychophysiological; sensors; software development; software engineering; tasks; productivity; systematic mapping study

1. Introduction

The biometric data (BD) records are physiological measures obtained from different parts of human body, including human skin, heart, eyes, brain and among others. These records are composed by data gathered from biometric-related sensors or wearable computing devices, which can be narrowly related to a person's cognitive and emotional states (Kramer 1991). These BD records can be considered as collections of data that can be stored and manipulated to represent the status of humans when performing specific cognitive activities. Software development can be seen as one of these cognitive activities that has been the context of using biometric sensors. Developing software consists of a set of activities performed by people (Brooks 1995), and requires cognition, engagement of different parts of human body, mainly eyes and brain processes.

In this sense, recent studies Gui et al. (2019), Bablani et al. (2019) have shown through surveys how much biometric data has gained attention in the last few years, especially as wearable devices have expanded rapidly. Some researches have also shown that cognitive load (Fritz et al. 2014), stress (Ostberg et al. 2017), emotions (Müller 2016), and attention (Meyer et al. 2017) can impact on daily activities performed by programmers. However, several aspects about the use of BD records in software engineering field are unknown, such as body measurements used to support daily tasks, and empirical methods that are used to evaluate their benefits. Consequently, a thorough understanding of state-of-the-art techniques still remains limited. Thus, practitioners and researchers end up not having a mapping of the literature considering currently available works.

This article, therefore, presents a comprehensive overview and understanding of the current literature, as well as pinpoints research gaps, challenges and trends. A systematic mapping of the literature was designed and performed based on well established practical guidelines (Petersen, Vakkalanka, and Kuzniarz 2015). In total, 40 primary studies were deeply analysed and categorised, which were selected by applying a careful filtering process from a sample of 3023 studies to answer five research questions. Over 77% of articles use more than one biometric aspect to analyse tasks performed by programmers; over 47% of articles used eye-track sensor to

CONTACT Juliano Paulo Menzen juliano.menzen@gmail.com Applied Computing Graduate Program (PPGCA) of University of Vale do Rio dos Sinos (UNISINOS), Av. Unisinos, 950 - Bairro Cristo Rei - CEP: 93.022-750, São Leopoldo, RS, Brazil

analyse the biometric factors, follow by brains sensor with 40%, skin sensor with 22%, cardiac sensor with 20% and others less representative; most studies analysed had their quality assessed as high.

Our article is organised as follows: Section 2 gives a delineation of the main concepts needed to understand this study. Section 3 introduces our review protocol. Section 4 introduces the procedures to filter potentially relevant studies. Section 5 presents the collected results related to research questions explored in our study. Section 6 outlines a discussion of some challenges and trends. Section 7 exposes the threats to the validity of our work. Section 8 describes related work. Section 9 is the conclusion of this paper.

2. Background

This section introduces the chief terms used throughout this paper. Section 2.1 presents the concept of biometric data. Section 2.2 outlines software development tasks that are typically performed by developers.

2.1. Biometric data

Biometric data (BD) are physiological measures collected from parts of human body, including skin, heart, eyes, brain and others, collected through sensors and can be related to a person's cognitive and emotional states (Kramer 1991). In most cases, these sensors related to BD records are non-invasive and collect physiological data produced autonomously by the human body, which can hardly be controlled by us. The measurements are captured in a different way according to the area of body to which they relate.

The variation of the data obtained by these measures in healthy people is also associated with psychological factors, such as emotion and cognitive load. The study of this relationship between biometric data and psychological aspects we call psycho-physiology. Psychophysiology has been receiving attention in the area of Software Engineering because the software development activity involves several cognitive processes such as language comprehension, attention (Siegmund et al. 2017), and also emotional, such as frustration (Wrobel 2013) and stress (Ostberg et al. 2017). Without the use of biometric sensors, the study of psychological aspects could be done only through holistic processes, limiting the accuracy of the analysis (Peitek et al., "Toward Conjoint Analysis," 2018). This area is a promising area in software engineering for promoting a better understanding of the developer's mental process during the performance of their daily activities (Müller 2016; Robillard 1999).

2.2. Software development tasks

The construction of software is an activity of knowledge, which is not limited to programming. The developer must describe and organise the knowledge represented by the program to be coded (Robillard 1999). This process involves a series of activities and interactions as read documentation, write documentation, meetings, browsing on internet, planning (Meyer et al. 2017) and programming. The programming activity can be classified into sub activities such as debugging, versioning and testing (Maalej and Happel 2009). All of these tasks can be developed individually or collaboratively (Gonçalves, de Souza, and González 2011).

Until a few years ago these activities were studied by Software Engineering only with the purpose of making them measurable and logical through the application of methods, tools, symbologies and languages (Robillard 1999). However, in the last years, the number of articles that use psychophysiology has increased in order to analyse the congnitive and mental processes of the developers to better understand how these tasks are performed, to provide better support during their realisation (Müller 2016).

3. Planning

This section aims to describe our review protocol. As previously mentioned in Section 1, we have chosen systematic mapping study as research method. According to Cooper (2016), this methodology tends to produce more reliable findings by reducing bias through a rigourous review process. Figure 1 shows an overview of the adopted systematic review process. Composed by three phases, in which activities are performed to create artefacts, this process serves as a guide on how to advance with the review. The planning phase (Section 3) covers all the procedures for review design. The conduction phase (Section 4) details step by step how a sample of potentially relevant works was obtained, and how this sample was filtered to identify a set of representative works. The reporting phase (Sections 5 and 6) focuses on reporting the results, drawing some trends and highlighting some challenges that can be explored by the scientific community.

3.1. Objective and research question

This section seeks to define the objective and research questions so that a clear *review scope* can be obtained.

Objective. The objective of this study is to produce a systematic map of the literature by classifying the works already published, discussing gaps and drawing

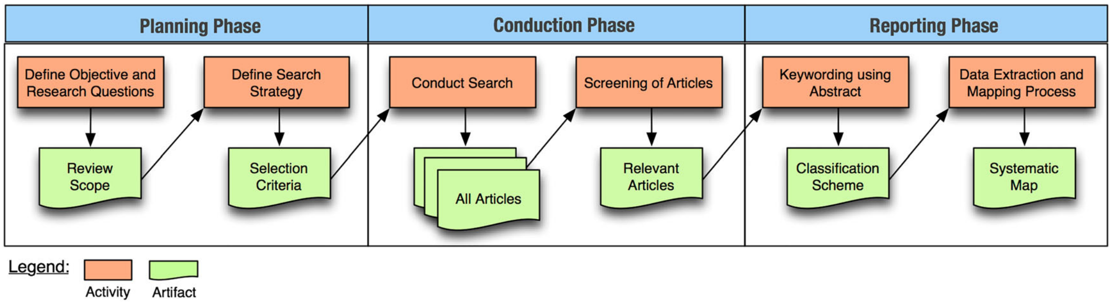

Figure 1. Overview of the systematic mapping process (adapted from Petersen et al. 2008).

some promising research directions, as well as proposing a taxonomy. Lau, Ioannidis, and Schmid (1997) reinforce that, combining findings from a large body of studies makes it possible to increase statistical power and provide answers that no single study can give.

Research questions. We have defined nine research questions (RQ) for properly addressing different facets of our objective. Table 1 presents such RQs, including three Statistical Question (SQ), five General Questions (GQ) and one Focuses Questions (FQ). The purpose of SQs is to disclose statistical data on *where* and *how often* studies have been published over the years. The GQs seek to reveal *what* biometric data are aiding in the execution of software development activities. Lastly, the FQs aim at identifying *how* specific biometric sensors are used to improve the developers' productivity during software-development tasks. Petersen, Vakkalanka, and Kuzniarz (2015) and Petersen et al. (2008) mention that systematic mappings of the literature are not intended to accurately pinpoint available problems and solutions. Rather, they should discover few explored research topics, trends and gaps not yet revealed. Thus, research questions of SMSs cannot be specific, but generic to the point of creating a panoramic view of the research area being explored. Table 1 presents our research questions.

After defining our objective and research questions, the following section determines our search strategy.

Table 1. List of research questions investigated.

Reference	Question
Statistical Questions	
SQ1	What is the number and type of publications by year?
SQ2	Which countries have publications in the researched area?
SQ3	Who are the main authors who publish articles about biometric data in software engineering?
General questions	
GQ1	How would the taxonomy for biometric data classification in software development tasks appear?
GQ2	What are the body measurements used to support software development tasks?
GQ3	What factors related to biometric data can influence software development tasks?
GQ4	What daily tasks have been commonly performed in the context of software development?
GQ5	What are the research methods used to evaluate the studies?
Focused questions	
FQ1	How can biometric data be used to improve developers' productivity?

3.2. Search strategy

Our search strategy is based on automated-search method following well-known empirical guidelines (e.g. Ali and Usman 2018; Petersen, Vakkalanka, and Kuzniarz 2015; Petersen et al. 2008), which helped us to determine an unbiased *search string*, and identify effective *electronic databases*.

Search string. Our search string is formed by terms identified as *main terms*, complemented by their most relevant *synonyms*. The search string plays a pivotal role to retrieve potentially relevant studies. The main terms are Biometric, Device, Software and Tasks. The synonyms of these terms were determined by keywords found in related works, and based on Collins dictionary. Table 2 shows the main terms and their synonyms.

Our search string was produced based on the combination of these terms. Five steps were followed to define our search string: (1) specify the main terms; (2) define alternative words, synonyms or related terms to main terms; (3) check if the main terms are contained in previously published articles; (4) associate synonyms, alternative words or terms related to the main terms with the 'OR' Boolean operator; and (5) link the main terms with Boolean operator 'AND'. The search string that produced the most significant results is presented as follows:

(Biometric OR Psycho-physiological OR Metric) AND (Sensor OR Biosensor OR Device) AND

Table 2. List of the main terms and their synonyms.

Main term	Synonym
Biometric	Biometric, Psychophysiological, Metric
Device	Sensor, Biosensor, Device
Software	Software Development, Software Engineering
Tasks	Task, Assignment, Work, Productivity

(Software Development OR Software Engineering) AND (Task OR Assignment OR Work OR Productivity)

In some electronic databases was necessary adjust the search string to find the number of papers presented in this systematic mapping. Table 3 presents in details adjusted search string and other specific filters for each electronic database selected for this mapping study. In addition, the inclusion criteria presented in Section 3.3 have been added to reduce the number of query returns.

Ali and Usman (2018) highlighted the importance of using a known-set studies to be used as a basis to evaluate the result obtained by the search in each electronic library. For this study five articles were used as know-set papers (Crk, Kluthe, and Stefik 2015; Fritz et al. 2014; Müller and Fritz 2016; Siegmund et al. 2014; Wrobel 2018). These articles were found from background knowledge of the authors on the theme. As a result, all papers used as know-set base were identified during searches performed in electronic libraries.

Table 3. List of adjusted search string in selected electronic database.

ACM Digital Library
(Biometric OR Psycho-physiological OR Metric) AND (Sensor OR Biosensor OR Device) AND ('Software Development' OR 'Software Engineering') AND (Task OR Assignment OR Work OR Productivity)
Google Scholar
(Biometric Sensor Task 'Software Engineering' Psychophysiological OR Biosensor OR Productivity OR 'Software Development')
IEEE Xplore Digital Library
(Biometric OR Psycho-physiological OR Metric) AND (Sensor OR Biosensor OR Device) AND ('Software Development' OR 'Software Engineering') AND (Task OR Assignment OR Work OR Productivity)
Inspec
(Biometric OR Psycho-physiological OR Metric) AND (Sensor OR Biosensor OR Device) AND ('Software Development' OR 'Software Engineering') AND (Task OR Assignment OR Work OR Productivity)
Pubmed
(Biometric OR Psycho-physiological OR Metric) AND (Sensor OR Biosensor OR Device) AND ('Software Development' OR 'Software Engineering') AND (Task OR Assignment OR Work OR Productivity)
Science Direct
(Biometric OR Psycho-physiological) (Sensor OR Device) ('Software Development' OR 'Software Engineering') (Task OR Assignment OR Work OR Productivity)
Scopus
(Biometric OR Psycho-physiological OR Metric) AND (Sensor OR Biosensor OR Device) AND ('Software Development' OR 'Software Engineering') AND (Task OR Assignment OR Work OR Productivity)
Taylor & Francis OnLine
(Biometric OR Psycho-physiological OR Metric) AND (Sensor OR Biosensor OR Device) AND ('Software Development' OR 'Software Engineering') AND (Task OR Assignment OR Work OR Productivity)
Wiley Online Library
(Biometric OR Psycho-physiological) AND (Sensor OR Device) AND ('Software Development' OR 'Software Engineering') AND (Task OR Assignment OR Work OR Productivity)
Microsoft Academy
Biometric AND Sensor AND 'Software Engineering' AND Task
Springer Link
Biometric OR Psycho-physiological Sensor OR Device 'Software Development' OR 'Software Engineering' Task OR Assignment OR Work OR Productivity
Semantic Scholar
text:(Biometric Sensor 'Software Engineering' Task) AND abstract:(Biometric Sensor 'Software Engineering' Task)

Electronic databases. After determining our search string, the next step was to identify electronic databases to retrieve potentially relevant studies. Table 4 displays the twelve electronic databases used. These electronic databases were chosen for two reasons. First, these databases have an elevated, representative number of published articles, related to the research topic explored in our SMS. Second, they have been widely used in previous systematic mapping studies (e.g. Rodríguez et al. 2017), which means that their usefulness and effectiveness have already been demonstrated.

3.3. Exclusion and inclusion criteria

This section seeks to establish criteria to support the filtering process of potentially relevant articles, which are retrieved from the selected electronic databases (Table 4) after applying our search string. For this, we define exclusion and inclusion criteria. Inclusion criteria consider characteristics that an article must have to be included in the initial sample of potentially relevant studies. Exclusion criteria in turn consider those characteristics that disqualify an article to be included in our sample of representative studies. Thus, if an article has an inclusion criteria, it is in the initial sample of potentially relevant studies; if an article has an exclusion criteria, it is out the sample of representative studies.

These criteria prescribe rules to make the filtering process as objective and auditable as possible, while seeking to prevent bias typically found in manual tasks performed by humans. There is no criterion that establishes an order to apply the inclusion and exclusion criteria, but for this systematic mapping study the inclusion criteria were applied directly to electronic database advanced search, while the exclusion criteria were applied in an order to favour the filtering process presented in Section 4.

Exclusion criteria. The following list specifies the exclusion criteria (EC) used in this article, which removed studies where:

Table 4. List of the selected electronic databases.

Source	Electronic address
ACM Digital Library	https://dl.acm.org
Semantic Scholar	https://www.semanticscholar.org
Google Scholar	https://scholar.google.com
IEEE Xplore Digital Library	https://ieeexplore.ieee.org/Xplore/home.jsp
Inspec	http://digital-library.theiet.org
Microsoft Academy	https://www.microsoft.com/en-us/research
PubMed	https://www.ncbi.nlm.nih.gov/pmc
Science Direct	https://www.sciencedirect.com
Scopus	https://www.scopus.com
Springer Link	https://link.springer.com
Wiley Online Library	https://onlinelibrary.wiley.com
Taylor & Francis OnLine	https://www.tandfonline.com

- EC1: The title, abstract or even their contents was related to the search keywords, however without any semantic interplay with the scope of this study.
- EC2: A patent had been registered, or studies were not published in English, might be considered as an initial stage, typically presenting an abstract and summary of future steps;
- EC3: No similarity with the research theme, or even the intention of the research is opposite to issues addressed by our research questions.
- EC4: The studies did not address any aspect of the research questions;
- EC5: It was a duplicate; and
- EC6: It did not address issues regarding biometric data in software engineering.

Reasons for choosing the exclusion criteria. Previous studies (Bischoff et al. 2019; Gonçales et al. 2015, 2019; Rodríguez et al. 2017) have demonstrated the usefulness and effectiveness of these criteria. Anyway, our choice can also be explained for five reasons. The first reason would be because of the lack of any sense of taking into account studies without any semantic relation with regard to the addressed research topics. Even though some studies might have their title or abstract matched with our formulated search string, their content may not be directly linked to the exploited content in our study (EC1). Second, early stage work would little help elaborate our primary objective, i.e. a systematic map of current literature. Instead, they could distort the data collected, thereby harming the design of trends and identification of gaps (EC2). Third, we did not seek to select studies that were not minimally related to our study purpose. Rather, we prioritised studies that might contribute to answer the formulated research questions properly (EC3). Fourth, if the abstract of a published article, or even its full text, does not address facts of the topics explored in our research questions, then it does not make sense to consider it (EC4). Finally, duplicate studies were thrown away as it would not make sense to count a study more than once (EC5) or not address issues regarding biometric data in software engineering (EC6). EC4 differs from EC6 in considering the all aspects of research questions, while EC6 focuses only on issues related to the use of biometric data in software engineering.

Application of the exclusion criteria. An iterative and incremental process was established to apply each exclusion criteria as well as to select and extract information from the selected articles. For each exclusion criteria applied, an interaction was made by reading individually and in parallel by the authors, followed by a step of synchronisation and consensus regarding the decisions made by the authors. Thus, the exclusion of an article was always based on peer review and the authors' consensus, so that bias could be minimised. To this end, two review cycles were performed to prevent any unwanted removal of articles.

Inclusion criteria. Furthermore, we have defined four criteria to guide the inclusion of candidate works in our sample to be analysed. The Inclusion Criteria (IC) are presented as follows:

- IC1: Studies should have been published in a conference, workshop, scientific journal, book or educational institutions.
- IC2: The study should be related to the usage of biometric in the context of execution of software development tasks.
- IC3: Works that have been written, published or disseminated in English.
- IC4: Studies published from January 2002 until January 2019.

Reasons for choosing the inclusion criteria. Inclusion criteria operate as a filter, just as exclusion criteria. In this study, four inclusion criteria were applied. The first inclusion criterion (IC1) considers only articles that have been published in a conference, workshop, scientific journal, book or educational institution. This criterion allowed us to disregard articles that have been submitted in non-academic media. In addition, only articles that contain biometric data related in some way to software development tasks were included (IC2). The third criterion (IC3) considers for inclusion in this mapping only articles written in English, while the fourth criterion considers only articles published between the year 2002 and February 2019 (IC4). The period from 2002 to 2019 was used, since no work related to the researched theme was identified before 2002 and the research execution was completed in 2019.

3.4. Data extraction

This section focuses on explaining how we extracted data from the selected studies, which are presented in Section 4.

Extraction procedures. After reading each selected study carefully, we obtained and stored data on a spreadsheet. Figure 2 presents the data extraction form used to feed this spreadsheet, which is inspired on ones already validated (e.g. Fernández-Sáez, Genero, and Chaudron 2013). Table 5 explains each label found in the extraction form.

Reasons for choosing this form. We have used this form because of four reasons. First, it guides the authors, serving as a template, in the data extraction procedures,

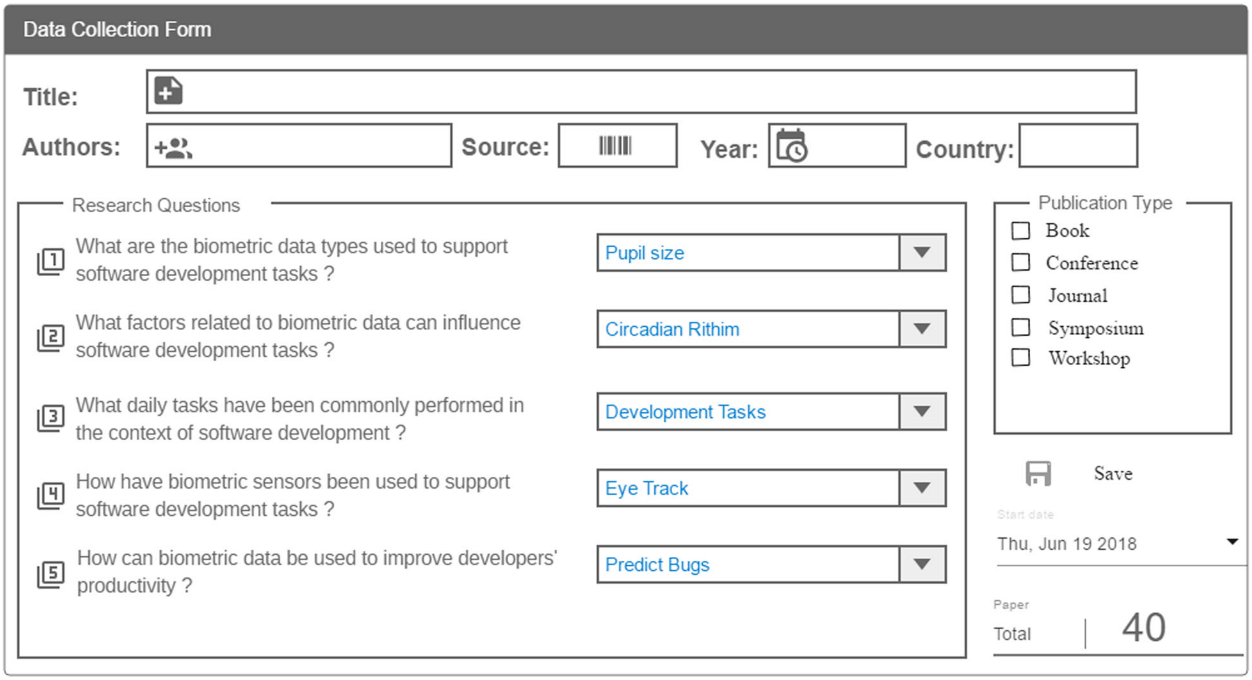

Figure 2. Data extraction form.

avoiding bias during the identification and storing of data. Second, although manual data extraction can be an error prone task, this risk was reduced by standardising how the data must be collected. Third, by concentrating the data, it helps in the synthesis, analysis and plotting of the collected indicators. As a result, better findings on our research questions may be generated and plotted. In addition, our extraction form is supported by a proposed classification scheme, shown in Table 5. This scheme does not only helps us to generate numerical, nominal or ordinal data, but also is essential for any attempt to create a consistent snapshot of the current literature.

Table 5. The proposed scheme to classify the selected studies.

Field	Description
Title	Article title.
Authors	Article authors. This data will be used to respond SQ3.
Source	Database in which the article was found. This information will be required to respond to SQ1.
Year of Publication	Year of publication used to answer the SQ1.
Country	Country related to the educational institution which the main author is linked. This information is used to respond SQ2.
Publication Type	Used as exclusion criterion, allowing only articles published in a conference, journal, book, symposium or workshop.
Research Questions	Place to answer the search questions with the article found.
Start Date	Place where the date on which the article search was executed.
Total	Total number of articles inserted in the mapping.

Justification of the start and end years chosen to collect the articles. The identification of the starting year to collect the articles was determined according to three criteria: (i) the year had up to 5 published articles (or no published articles); (ii) had a reduced number of published articles (less than or equal to 5 articles) within the next 10 years; and (iii) the number of articles published within 10 years after the candidate year should represent up to 30% of the total sample of articles found. As can be seen in Figure 3, the application of such criteria converges to the year 2002, which shows the beginning of a community-based force of interest on the subject of biometric data in Software Engineering. Such objective criteria allowed us to find a start year that graphically indicated a power tail towards 2018. Therefore, the chosen start year was 2002. Moreover, the search and filtering of the articles took place until January 2019. After this date, the authors began writing the article itself.

4. Study filtering

This section outlines how the conduction phase of our study was performed (Figure 1). This phase is composed by eight steps to filter potentially relevant works. Figure 3 illustrates the filtering process, which is based on the application of exclusion criteria described in Section 3.3. Each step is described as follows:

Step 1: Initial search. The focal point of this step was to apply our search string to the selected electronic

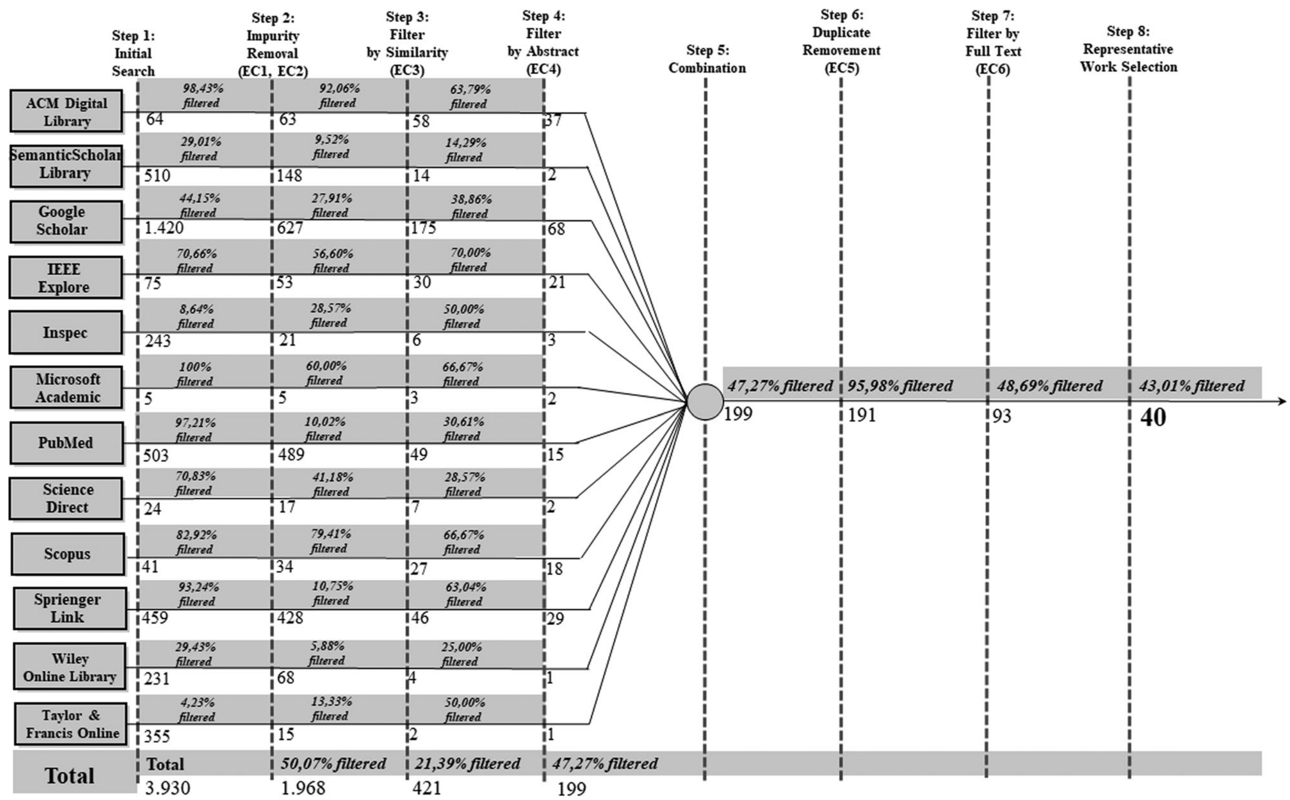

Figure 3. The obtained results after executing the filtering process.

databases described in Section 3.2. In total, 3930 potentially relevant articles were retrieved.

Step 2: Impurity removal (EC1 & EC2). This step aimed to remove impurities, which was realised by applying the exclusion criteria 1 and 2 (EC1, EC2) (detailed in Section 3.3). Many articles were retrieved because the terms of our search string matched with their text. However, these articles were not narrowly related to the purpose of our study. Examples of the works retrieved improperly would be calls for papers to conferences, special issues of journals, patent specifications, research reports, non-peer-reviewed materials, and studies that were not written in English. Therefore, the removal of impurities refers to the process of excluding these articles improperly retrieved. In this step, 50.07% (1968/3930) of the initially researched articles were kept, being 1962 works removed.

Step 3: Filter by similarity (EC3). In this step 1547 articles were removed, as they did not present similarity to the search string. Only 21.39% (421/1968) of the filtered articles were maintained.

Step 4: Filter by abstract (EC4). This step consists of analysing the abstract of the filtered works, verifying if it has relation with at least one of the research questions described in Table 1. In this step, 47.27% (199/421) of the filtered articles were maintained, with 222 studies being removed.

Step 5: Combination. All studies filtered from the last step were then brought together, producing a total of 199 studies.

Step 6: Duplicate removal (EC5). Commonly, articles are made available in more than one electronic database. So, it makes sense to remove the articles in duplicate, thereby assuring ensuring the uniqueness of each study in our sample. Only 8 articles in duplicate were removed in this step, maintaining 95.98% (191/199) of the filtered articles.

Step 7: Filter by full text (EC6). This step filtered studies by applying EC6 to the full text. That is, we read entirely the text of 191 selected articles. Through the application of EC6, 48.69% (93/191) of the articles were kept, excluding the 98 articles that were not narrowly related to the purpose of this study.

Step 8: Representative work selection. In this stage of the process, the 93 previously filtered articles were read again, and 43.01% of them (40/93) were selected as being the most relevant. Although the 53 excluded articles went through all the exclusion criteria, they did not meet the objective of this article related in Section 3.1. Some of these excluded articles use biometric data in activities not related to software engineering, as security and authentication process. Other articles presented literature reviews, and were explored as related work in Section 8. Finally, 40 studies were selected as the most

Table 6. Search stratification with primary studies.

Digital library	Amount	Percentage	List of primary studies
ACM Digital Library	21	52.5%	A01, A03, A05, A10, A14, A16, A17, A20, A22, A23, A24, A25, A26, A28, A29, A30, A31, A33, A35, A37, A39
IEEE Xplore	11	27.5%	A04, A06, A08, A09, A13, A15, A18, A21, A27, A32, A34
Google Scholar	4	10%	A02, A07, A12, A40
Springer Link	2	5%	A11, A36
Pubmed	1	2.5%	A38
Science Direct	1	2.5%	A19

representative ones, hereinafter called *primary studies*. Table 7 presents the final list of the filtered studies.

Search stratification over electronic databases. After discussing the filtering process, the next step is to show how the selected studies were stratified over the electronic databases (Table 4) used in our study. Figure 4 shows the search stratification over the electronic databases. In this stratification process, after applying all inclusion and exclusion criteria, only six of twelve digital libraries initially surveyed returned primary studies. We can highlight that the ACM Digital Library was the electronic database with the largest number of studies (52.5%), followed by IEEE Explore Database with 27.5%. Table 6 presents the same stratification presented in Figure 4, but details the primary studies found in each electronic database. These studies are presented in detail in Table 7.

Although 12 electronic databases have been considered in our study, selected articles from 6 electronic databases were removed throughout the filtering process. Such databases were Semantic Scholar, Inspec, Microsoft Academic, Scopus, Wiley Online Library, Taylor & Fancis Online. For this reason, they were not considered in Figure 6.

5. Results

This section aims to present the findings for our research questions (Table 1) after examining the 40 primary studies (Table 7). Numerical processing and graphical representation of interesting features of our findings are outlined to facilitate understanding.

5.1. SQ1: what is the number and type of publications by year?

The SQ1 reveals the number and type of publications of the primary studies over the years. This classification is based on the year of publication (i.e. from 2002 to 2019), type of publication (i.e. journal, book, conference, workshop, and symposium), and the number of publications per year. Figure 5 presents the data obtained according to the mentioned criteria.

This distribution helps to create a 'big picture' of the literature over years. We can notice that from 2014 there was an increase in the number of publications on the use of biometric data related to Software Engineering and from this then the theme has been continuously researched. This demonstrates that this research area is very active and still growing. Figure 5 reveals that there is a preference for publications in conferences (55%, 22/40), corresponding to more than half of the primary studies. In addition to the conferences, many articles are published in journals (20%, 8/40) and symposiums (15%, 6/40),

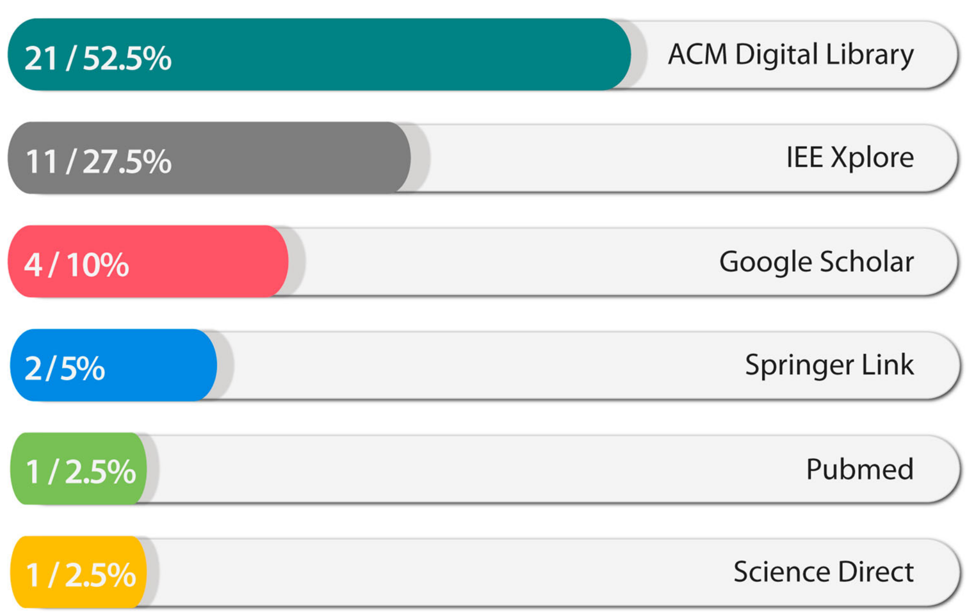

Figure 4. Search stratification based on the used electronic databases.

Table 7. List of primary studies.

Identifier	Author, Year	Publisher	Type
A01	Fritz et al. (2014)	ACM	Conference
A02	Gonçalves, de Souza, and González (2011)	J.UCS	Journal
A03	Kevic et al. (2015)	ACM	Conference
A04	Maalej and Happel (2009)	IEEE	Conference
A05	Meyer et al. (2014)	ACM	Conference
A06	Meyer et al. (2017)	IEEE	Journal
A07	Müller (2016)	ZORA	Book
A08	Müller and Fritz (2015)	IEEE	Conference
A09	Perry, Staudenmayer, and Votta (2002)	IEEE	Journal
A10	Siegmund et al. (2017)	ACM	Conference
A11	Kaklauskas et al. (2011)	Springer	Book
A12	Wrobel (2018)	Applied Science	Journal
A13	Wrobel (2013)	IEEE	Conference
A14	Züger and Fritz (2015)	ACM	Conference
A15	Siegmund (2016)	IEEE	Conference
A16	Müller and Fritz (2016)	ACM	Conference
A17	Crk, Kluthe, and Stefik (2015)	ACM	Journal
A18	Fritz and Muller (2016)	IEEE	Conference
A19	Kevic et al. (2017)	Science Direct	Journal
A20	Nakagawa et al. (2014)	ACM	Conference
A21	Begel (2016)	IEEE	Workshop
A22	Ostberg et al. (2017)	ACM	Workshop
A23	Konopka (2015)	ACM	Conference
A24	Züger et al. (2018)	ACM	Conference
A25	Bednarik and Tukiainen (2006)	ACM	Symposium
A26	Latoza, Venolia and Deline (2006)	ACM	Conference
A27	Sharif and Maletic (2010)	IEEE	Conference
A28	Siegmund et al. (2012)	ACM	Symposium
A29	Glücker et al. (2014)	ACM	Conference
A30	Rodeghero et al. (2014)	ACM	Conference
A31	Siegmund et al. (2014)	ACM	Conference
A32	Busjahn et al. (2015)	IEEE	Conference
A33	Palmer and Sharif (2016)	ACM	Symposium
A34	Duraes et al. (2016)	IEEE	Symposium
A35	Floyd, Santander, and Weimer (2017)	ACM	Conference
A36	Lee et al. (2017)	Springer	Journal
A37	Fakhoury et al. (2018)	ACM	Journal
A38	Castelhano et al. (2018)	Pubmed	Journal
A39	Peitek et al., "Toward Conjoint Analysis" (2018)	ACM	Symposium
A40	Peitek et al., "Simultaneous Measurement of Program" (2018)	ACM	Symposium

and the rest is distributed homogeneously between workshops and books.

5.2. SQ2: which countries have publications in the researched area?

SQ2 presents from which countries the publications of primary studies originated. The analysis of the countries is relevant because it allows identifying in which regions the authors are making most efforts in research of the addressed theme.

Analysing the publications allows us to create a global map presented in Figure 6. This map reveals that most publications occur in the northern hemisphere (97.5%, 39/40). Switzerland is the country with most publications (27.5%, 11/40), followed by United States (25%, 10/40) and Germany (22.5%, 9/40). The remaining countries, together account for 25% of the publications, with 10 publications. Table 8 presents the primary works by country in detail.

5.3. SQ3: who are the main authors who publish articles about biometric data in software engineering?

This question was answered from analysis of the authors related to the primary studies. In this analysis, only the main and second author were accounted as presented in Table 9. The other authors were disregarded because their contribution is usually smaller.

From Table 9 is possible identify that some authors are active in the searched area, such as Janet Siegmund (6 articles), Thomas Fritz (6 articles) and Sebastian Muller (5 articles). These authors have the following characteristics: *Janet Siegmund*: PhD researcher at the University of Passau, Germany. Her main interests are code comprehension, experimental software engineering, human factors in computer science and Functional Magnetic Resonance Imaging (fMRI). *Sebastiam Muller*: Assistant Researcher at the University of Zurich from 2011 until 2016, when he held his pos doc until 2017. Also active in the field of Software Engineering in the industry until 2019. *Thomas Fritz*: Professor at the University of Zurich. His research is related to the empirical software engineering and the use of biometric data to better understand developers and improve their productivity. The other authors present a smaller number of publications, with a maximum of three publications each.

5.4. GQ1: how would the taxonomy for biometric data classification in software development tasks appear?

Our primary studies have revealed wide-ranging and diverse research on the use of biometric data to support software-development tasks. Nevertheless, such research cannot be graded unless accompanied by an intuitive grasp classification scheme. Accordingly, we seek to build a support taxonomy to assist researchers and practitioners in performing these classifications.

For the GQ1 research question, we sought to define a taxonomy, shown in Figure 7, based on the selected and analysed articles. We searched the literature for related taxonomies or even processes to build it, but nothing was found. For this reason, our taxonomy was defined based on the following steps:

Publication Type	Percentage
Conference	55.00% [22 of 40] Primary Studies
Journal	20.00% [8 of 40] Primary Studies
Symposium	15.00% [6 of 40] Primary Studies
Workshop	5.00% [2 of 40] Primary Studies
Book	5.00% [2 of 40] Primary Studies

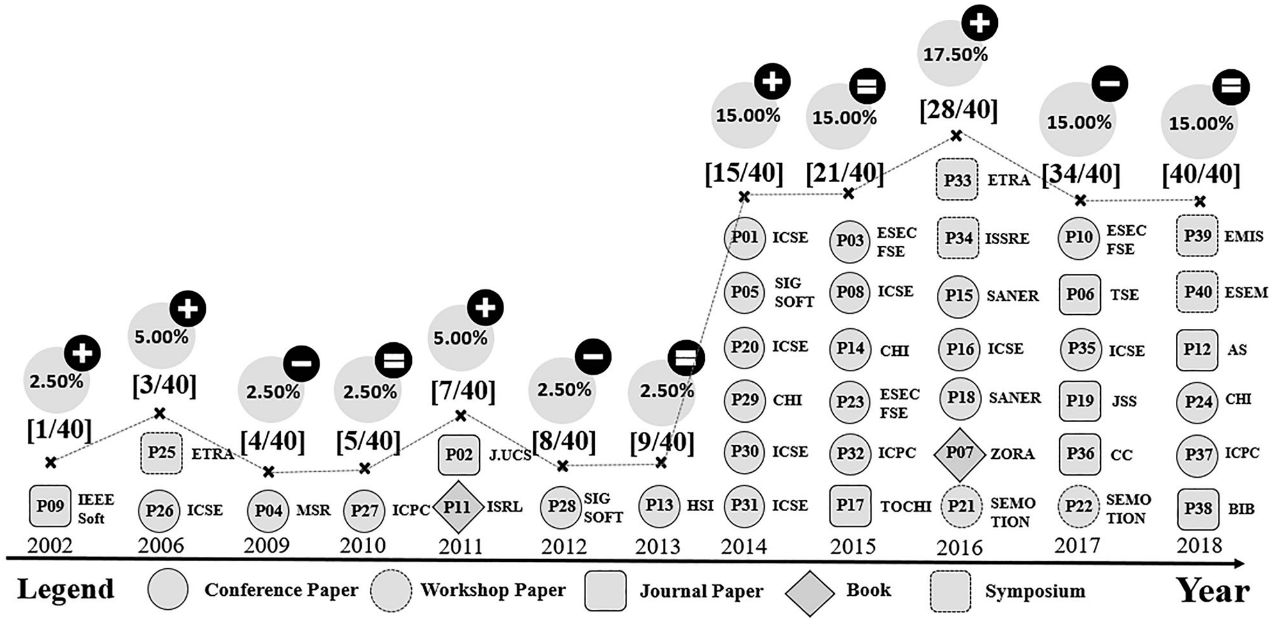

Figure 5. Frequency of publications.

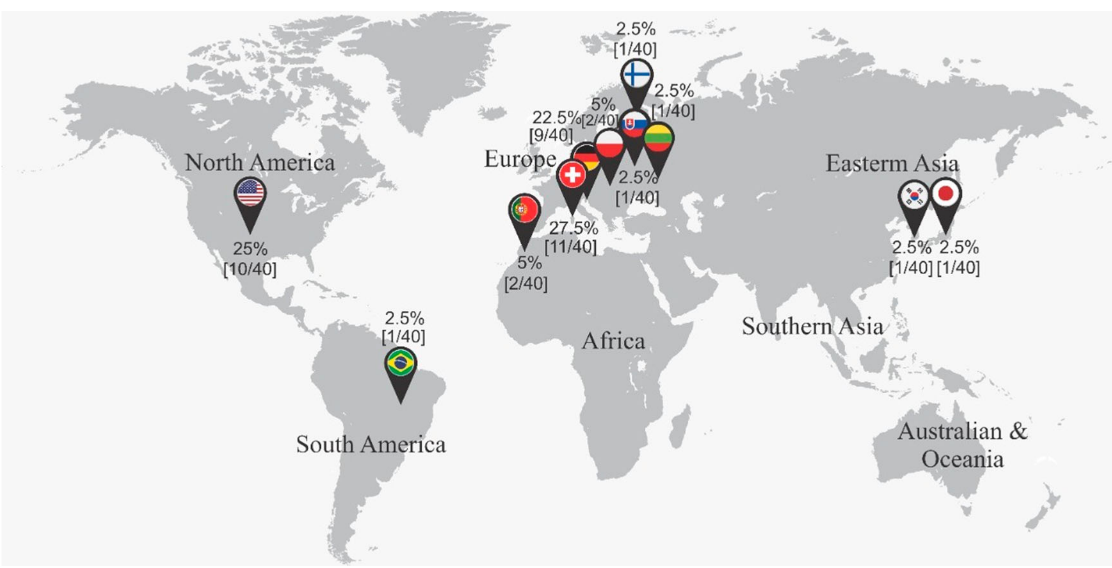

Figure 6. Publications by Country.

- *Step 1: identification of major groups of terms.* We primarily identified five major groups of term (detailed in Table 10) related to studies using biometric data in the context of software engineering tasks. These macro groups, narrowly related to our research questions, would be: (1) body measurements (GQ2), (2) factors (GQ3), (3) tasks (GQ4), (4) research methods (GQ5), and (5) research goals (FQ1).

Table 8. Publications by country in details.

Country	Amount	Percentage	List of primary studies
Switzerland	11	27.5%	A01, A03, A05, A06, A07, A08, A14, A16, A18, A19, A24
United States	10	25%	A09, A17, A21, A26, A27, A30, A33, A35, A37, A40
Germany	9	22.5%	A04, A10, A15, A22, A28, A29, A31, A32, A39
Poland	2	5%	A12, A13
Portugal	2	5%	A34, A38
Brazil	1	2.5%	A02
Finland	1	2.5%	A25
Japan	1	2.5%	A20
Lithuania	1	2.5%	A11
Slovakia	1	2.5%	A23
South Korea	1	2.5%	A36

- *Step 2: detailing the identified macro groups.* From these five groups, we were able to examine the selected studies in depth to grasp each one of them. This detail was made progressively as the research questions (GQ2-5 and FQ1) were answered.
- *Step 3: assembly and stylisation of the taxonomy.* After having the macro groups and their details, the next step was to assemble the taxonomy itself and define style issues including lines, colours, etc. That is, the taxonomy was done in retrospect. The style of our taxonomy was inspired by the study proposed by Aghajani et al. (2019).

Table 9. Authors by publications.

Author	First author			Second author		
	Amount	Percentage	Primary studies	Amount	Percentage	List of primary studies
Janet Siegmund	4	10%	A10, A15, A28, A31	2	5%	A39, A40
Sebastian Muller	3	7.50%	A07, A08, A16	2	5%	A18, A24
Katja Kevic	2	5%	A03, A19	0	0%	–
Manuela Zuger	2	5%	A14, A24	0	0%	–
Michal R.Wrobel	2	5%	A12, A13	0	0%	–
Norman Peitek	2	5%	A39, A40	1	2.5%	A10
Thomas Fritz	2	5%	A01, A18	4	10%	A05, A08, A14, A16
André Meyer	2	5%	A05, A06	0	0%	–
Andrew Begel	1	2.5%	A21	1	2.5%	A01
Arturas Kaklauskas	1	2.5%	A11	0	0%	–
Bonita Sharif	1	2.5%	A27	0	0%	–
Benjamin Floyd	1	2.5%	A35	0	0%	–
Christopher Palmer	1	2.5%	A33	0	0%	–
Dewayaje E. Perry	1	2.5%	A09	0	0%	–
Hartmut Glücker	1	2.5%	A29	0	0%	–
Igor CRK	1	2.5%	A17	0	0%	–
J. Duraes	1	2.5%	A34	0	0%	–
Jan-Peter Ostberg	1	2.5%	A22	0	0%	–
Joao Castelhano	1	2.5%	A38	0	0%	–
Marcio Kuroki Goncalves	1	2.5%	A02	0	0%	–
Martin Konopka	1	2.5%	A23	0	0%	–
Paige Rodeghero	1	2.5%	A30	0	0%	–
Roman Bednarik	1	2.5%	A25	1	2.5%	A32
Sarah Fakhoury	1	2.5%	A37	0	0%	–
Seolhwa Lee	1	2.5%	A36	0	0%	–
Takao Nakagawa	1	2.5%	A20	0	0%	–
Teresa Busjahn	1	2.5%	A32	0	0%	–
Thomas D. LaToza	1	2.5%	A26	0	0%	–
Walid Maalej	1	2.5%	A34	0	0%	–
André Brechmann	0	0%	–	1	2.5%	A28
Bonita Sharif	0	0%	–	1	2.5%	A33
Braden M. Walters	0	0%	–	2	5%	A03, A19
Christian Kästner	0	0%	–	1	2.5%	A31
Cleidson R. B. de Souza	0	0%	–	1	2.5%	A02
Collin McMillan	0	0%	–	1	2.5%	A30
Danial Hooshyar	0	0%	–	1	2.5%	A36
Daniel Graziotin	0	0%	–	1	2.5%	A22
Felix Raab	0	0%	–	1	2.5%	A29
Gina Venolia	0	0%	–	1	2.5%	A27
H. Madeira	0	0%	–	1	2.5%	A34
Hans-Jorg Happel	0	0%	–	1	2.5%	A04
Isabel C. Duarte	0	0%	–	1	2.5%	A38
Laura E. Barton	0	0%	–	1	2.5%	A06
Markku Tukiainen	0	0%	–	1	2.5%	A25
Nancy .A. Staudenmayer	0	0%	–	1	2.5%	A09
Timothy Kluthe	0	0%	–	1	2.5%	A17
Tyler Santander	0	0%	–	1	2.5%	A35
Yasutaka Kamei	0	0%	–	1	2.5%	A20
Yuzhan Ma	0	0%	–	1	2.5%	A37

In this sense, this taxonomy can benefit researchers and practitioners from an overview of possible classifications of the current studies, mainly in terms of body measurements, factors, tasks, research methods, and research goals.

The analysis of Figure 7 together with Table 10 makes possible to understand the relationship between the main elements of the proposed taxonomy in relation to the research questions of this study. Although this taxonomy is the first general research question (GQ1), it was the last to be answered, as it required the answer of the other questions for its correct elaboration.

5.5. GQ2: what are the body measurements that are used to support software development tasks?

The GQ2 seeks to identify the body measurements that are being commonly used to support software development tasks. To answer this research question, we analyse all primary studies that explore body measurements in the context of daily work activities of software developers (summarised in Table 11). Upon examining the primary articles, we were able to locate an assorted list of types of metrics related to body measurements, including physical, behavioural and physiological aspects. Each record usually reflects a partial view of a biometrics-related sensor, usually without allowing researchers to control and interact with these measurements in an integrated way.

There are several types of biometric data involved in quantifying developer characteristics, such as brain waves, heart rate and temperature. Eight categories were proposed for a more refined framing of the data types, including Eye, Skin, Brain, Heart, Breathing, Body Movements, Voice, and Endocrine system. These data types refer to metrics related to physiological and behavioural characteristics for labelling and describing people who develop software.

An interesting finding was that most studies use only one type of biometric data to perform the analysis of activities executed by the developers (42.5%, 17/40), while using information from different parts of the body (35%, 14/40). Most of the studies used biometric data associated with the eyes (47.5%, 19/40), brain (40%, 16/40) and skin (22.5%, 9/40). These measures are the most used by researchers because they are associated with the autonomic nervous system and, therefore, cannot be controlled by us so easily, ensuring greater accuracy in the experiments. Few studies were found using body movements (10%, 4/40) and only one study was found referring to voice and the endocrine system corresponding to 2.5% respectively.

Considering the biometric data individually, the fixation of the eyes corresponds to the most used biometric data in the analysis of the activities of the developer (35%, 14/40), followed by the blood oxygen level of the brain (22.50%, 9/40) and phasic and tone of the skin (22.50%, 9/40).

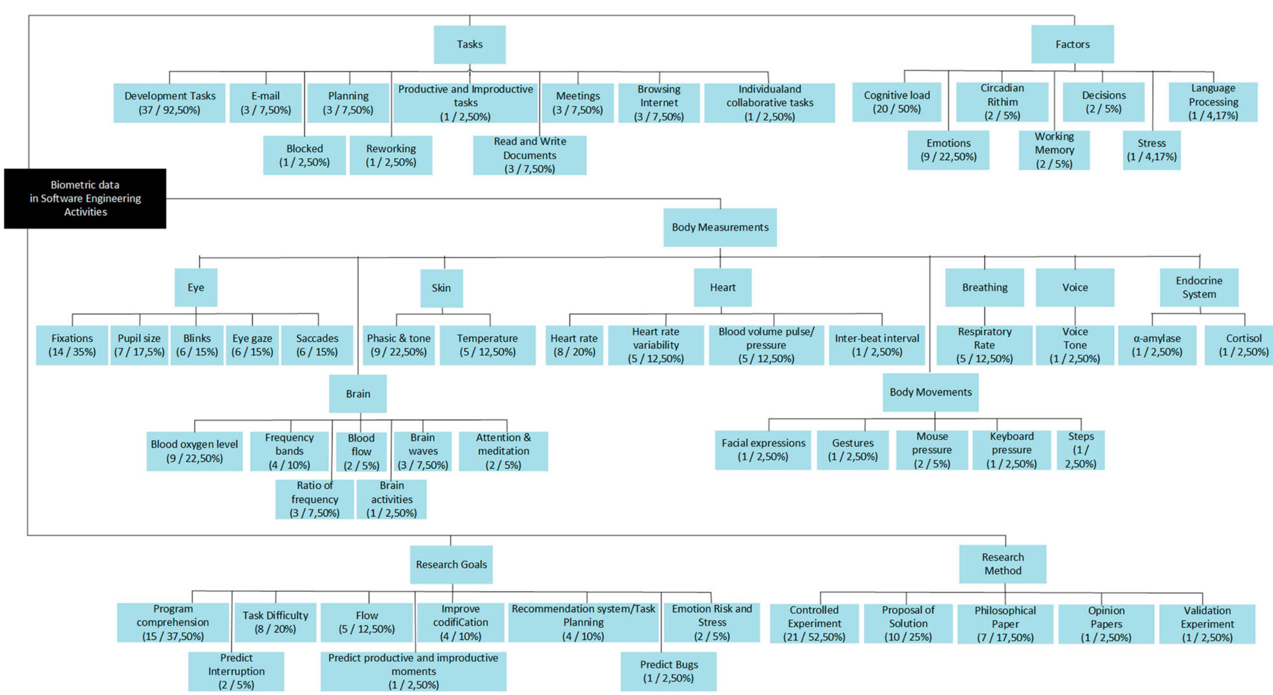

Figure 7. A taxonomy to support the classification of studies that use biometric data in software development tasks.

Table 10. Biometric data classification taxonomy.

Group	Description
Biometric measurements	Provide metrics related to human characteristics used to support software developers on daily work activities (see GQ2 in Section 5.5).
Factors	Important influential factors that end up impacting on the execution of daily work activities performed by developers (see GQ3 in Section 5.6).
Tasks	Main daily work tasks performed by developers (see GQ4 in Section 5.7).
Research method	Empirical strategies used to plan, execute and evaluate studies (see GQ5 in Section 5.8).
Research goals	Main purpose of the work undertaken to increase the knowledge about how to increase developer's productivity by managing risks, mainly related to problems encompassing program comprehension, stress and emotion risk, among others (see FQ1 in Section 5.9).

5.6. GQ3: what factors related to biometric data can influence software development tasks?

The GQ3 focuses on identifying the factors that can be measured by the biometric data collected and identified by GQ2 that influence to the tasks performed by software developers. Table 12 presents the factors that are related to the activities performed by developers.

Cognitive load corresponds to 50% of the articles researched (20/40), being the most used physiological aspect along with the emotions (22.5%, 9/40). The aspect of cognitive load is widely used in the primary studies because the coding activity is an intellectual task that requires reasoning and adequate understanding of the programming to be performed.

The studies report that a high cognitive load is associated with difficulties of understanding and performing a task (Crk, Kluthe, and Stefik 2015; Fritz et al. 2014). Emotions are also studied because they are associated with the effort employed in the execution of a task (Müller and Fritz 2015). The circadian rhythm is receiving attention only in recent papers, association greater attention and productivity at different times of the day. Although stress is an emotional state, it has been categorised emotionally separately, because a study has been elaborated only by analysing the influence of stress on programming tasks (Ostberg et al. 2017). Like decision making, working memory and language processing are cognitive processes, but were categorised separately because there are studies specifically targeted to these areas.

5.7. GQ4: what daily tasks have been commonly performed in the context of software development?

The GQ4 identifies daily work activities performed by developers. These activities are presented in Table 13. To assist in the identification of these tasks, Meyer et al. (2014) conducted a survey to identify the perception of productivity by developers. This survey allowed to categorise developers' tasks with the representativeness of each type of task during a work day. It was also possible to categorise these tasks as productive and non-productive from the point of view of the developers.

Table 11. Body measurements used to support software development tasks.

Group	Body measurements	Amount	Percentage	List of primary studies
Eye	Fixations	14	35%	A01, A07, A08, A18, A21, A25, A27, A29, A30, A32, A33, A37, A39, A40
	Pupil size	7	17.5%	A01, A07, A08, A18, A21, A36, A39
	Blinks	6	15%	A01, A07, A08, A14, A18, A39
	Eye gaze	6	15%	A03, A19, A23, A30, A32, A39
	Saccades	6	15%	A01, A21, A27, A36, A39, A40
Skin	Phasic & tone	9	22.5%	A01, A07, A08, A11, A12, A14, A16, A18, A21
	Temperature	5	12.5%	A07, A08, A11, A16, A18
Brain	Blood oxygen level	9	22.5%	A10, A15, A28, A31, A34, A35, A38, A39, A40
	Frequency bands	4	10%	A07, A08, A18, A36
	Ratio of frequency	3	7.5%	A07, A08, A18
	Brain waves	3	7.5%	A01, A15, A17
	Attention & meditation	2	5%	A08, A14
	Blood flow	2	5%	A20, A37
	Brain activities	1	2.5%	A14
Heart	Heart rate	8	20%	A07, A08, A11, A12, A14, A16, A18, A24
	Heart rate variability	5	12.5%	A07, A08, A16, A18, A24
	Blood volume pulse/pressure	5	12.5%	A07, A08, A11, A14, A18
	Inter-beat interval	1	2.5%	A14
Breathing	Respiratory rate	5	12.5%	A07, A12, A16, A18, A24
Body Movements	Facial expressions	1	2.5%	A12
	Gestures	1	2.5%	A12
	Mouse pressure	2	5%	A11, A21
	Keyboard pressure	1	2.5%	A12
	Steps	1	2.5%	A24
Voice	Voice tone	1	2.5%	A12
Endocrine System	Cortisol	1	2.5%	A22
	α-amylase	1	2.5%	A22

Table 12. Factors related to biometric data can influence software development tasks.

Physiological factors	Amount	Percentage	List of primary studies
Cognitive load	20	50%	A01, A06, A10, A14, A15, A16, A17, A18, A19, A20, A22, A25, A27, A28, A31, A35, A36, A37, A39, A40
Emotions	9	22.5%	A07, A08, A11, A12, A13, A14, A18, A21, A22
Circadian rithim	2	5%	A06, A24
Working memory	2	5%	A17, A34
Decisions	2	5%	A34, A38
Stress	1	4.17%	A22
Language Processing	1	4.17%	A34

Table 13. Daily tasks commonly performed in the context of software development.

Task category	Amount	Percentage	List of primary studies
Development Tasks	37	92.5%	A01, A03, A04, A06, A07, A08, A10, A11, A12, A13, A14, A15, A16, A17, A18, A19, A20, A21, A22, A23, A24, A25, A26, A27, A28, A29, A30, A31, A32, A33, A34, A35, A36, A37, A38, A39, A40
E-mail	3	7.5%	A04, A06, A26
Planning	3	7.5%	A04, A06, A26
Meetings	3	7.5%	A04, A06, A26
Browsing internet	3	7.5%	A04, A06, A26
Read and Write documents	3	7.5%	A04, A06, A26
Individual and collaborative tasks	1	2.5%	A02
Productive and unproductive tasks	1	2.5%	A05
Blocked tasks	1	2.5%	A09
Reworking tasks	1	2.5%	A09

Most of the articles studied (95.5%, 37/40) only generalises the development activity, without detailing it. However, Meyer developed an application to monitor daily tasks (Meyer et al. 2017), and came to a categorisation of activities similar to those obtained in previous surveys (Meyer et al. 2014). Meyer also noticed a large fragmentation in daily activities, besides of identifying the influence of the circadian rhythm of each developer in relation to their productivity higher or lower at certain times of the day. Similarly, Maalej and Happel (2009) perceived, after reviewing commits, that developers' tasks are the same ones found by Meyer.

5.8. GQ5: what are the research methods used to evaluate the studies?

The GQ5 investigate research methods used by the primary studies. According to Wieringa et al. (2006), the research methods can be classified as: Controlled Experiment, Proposal of Solution, Philosophical Paper, Opinion Paper and Validation Experiment. As many other studies (Bischoff et al. 2019; Gonçales et al. 2015, 2019) have demonstrated the usefulness of this classification, we have adopted it. Table 14 presents the study classification.

Controlled experiment was the most frequently adopted research method (52.5%, 21/40), followed by Proposed Solutions (25%, 10/40). Only one of the studies was an experiment validation associated with cognitive process identification during the execution of programming tasks (Siegmund et al. 2017). This demonstrates that the research area is still new. An interesting research direction would be to conduct case studies with practitioners in real-world software development environments.

5.9. FQ1: how can biometric data be used to improve developers' productivity?

The FQ1 explores how biometric data have been used to improve the execution of software development tasks, such as code comprehension, code smells, and bug prediction. Table 15 presents a summary of the collected data. Code comprehension is the subject most addressed by researchers (37.5%, 15/40). The measurement of difficulty related to the execution of development tasks has also received considerable attention (20%, 8/40). This can be explained because, as shown in Table 13, they are narrowly related to development tasks. The coding activity is also related keep the developer focused on activity with no execution blocks (flow) (12.5%, 5/40), coding improvement (10%, 4/40), bug prediction (2.5%, 1/40) and emotions and stress measurement (5%, 2/40). Emotions and stress associated with

Table 14. Classification of primary studies based on research method.

Research methods	Number of studies	Percentage	Studies
Controlled Experiment	21	52.5%	A01, A06, A08, A11, A12, A07, A17, A18, A20, A22, A23, A24 A27, A28, A32, A34, A35, A37, A38, A39, A40
Proposal of Solution	10	25%	A01, A03, A07, A14, A25, A29, A30, A31, A33, A36
Philosophical Paper	7	17.5%	A02, A04, A09, A13, A15, A19, A26
Opinion Paper	1	2.5%	A21
Validation Experiment	1	2.5%	A10

Table 15. Biometric data used to improve developers' productivity.

Development tasks	Amount	Percentage	List of primary studies
Program comprehension	15	37.5%	A10, A15, A20, A25, A27, A28, A30, A31, A32, A34, A35, A37, A38, A39, A40
Task difficulty	8	20%	A01, A08, A16, A17, A18, A19, A22, A36
Flow	5	12.50%	A05, A06, A08, A18, A24
Improve codification	4	10%	A03, A23, A29, A32
Recommendation system/Task Planning	4	10%	A02, A07, A11, A36
Emotion risk and stress	2	5%	A13, A22
Predict interruption	2	5%	A14, A18
Predict (im)productive moments	1	2.5%	A06
Predict Bugs	1	2.5%	A34

development tasks impact the greater risk of the developer adding bugs to code due to the influence of these feelings on cognitive activity.

6. Discussion and challenges for future research

This section introduces an additional discussion when cross-referencing the research questions GQ2, GQ3 and FQ1. Moreover, we also present some challenges that can be explored by research community.

Most explored research topics. Figure 8 shows the relationship between the biometric data and the psycho-physiological factors that influence productivity improvement. Each bubble presents a triple ($t1$,$t2$,$t3$) with the obtained data, where $t1$ represents the research questions GQ2 and FQ1, $t2$ is the GQ3 research question, and $t3$ consists of the number of studies that explore the crossing of these questions. From the quantitative evidence highlighted in this figure, we try to understand which are the most researched psycho-physiological aspects in Software Engineering, identified through which biometric data and applied on which characteristics can improve the productivity of the daily tasks performed by developers.

Figure 8 shows us that most of the primary studies of this mapping capture the biometric data obtained by the eyes, heart, skin and brain to analyse emotions and cognitive load in order to understand the process of code comprehension and tasks difficulty. Sensors associated with the brain and eyes were primarily used to identify cognitive load (Crk, Kluthe, and Stefik 2015; Fritz et al. 2014; Müller 2016; Siegmund et al. 2017; Züger and Fritz 2015). In many studies, these two sensors were used together producing greater accuracy in the results (Fritz et al. 2014; Lee et al. 2017; Müller 2016; Peitek et al., "Toward Conjoint Analysis," 2018; Peitek et al.,

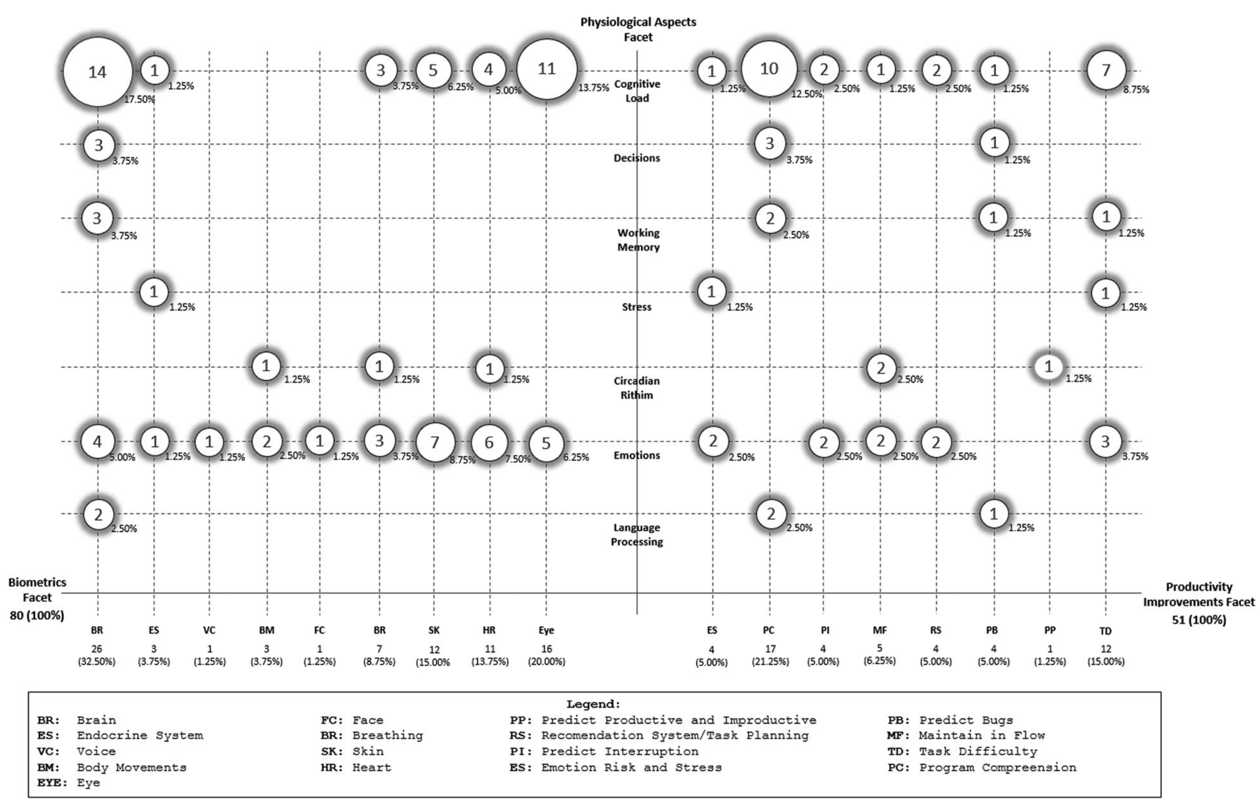

Figure 8. Relation between physiological factors, biometrics and productivity improvements.

"Simultaneous Measurement of Program," 2018; Züger and Fritz 2015). To analyse the emotions were related sensors to all parts of the body as seen in Figure 8, however, most studies used measurements obtained by the skin, heart, brain and eyes (Müller 2016; Müller and Fritz 2015; Ostberg et al. 2017; Wrobel 2018; Züger and Fritz 2015). While cognitive load was related to code understanding and difficulty (Fritz et al. 2014; Fritz and Muller 2016; Müller and Fritz 2016; Nakagawa et al. 2014; Siegmund et al. 2017), emotions and circadian rhythm were related to interruption, focus and planning of developmental activities (Fritz and Muller 2016; Kaklauskas et al. 2011; Meyer et al. 2017; Müller 2016; Züger et al. 2018).

This result suggests that most of the research in the area focuses on the understanding of code and difficulty of tasks, associated with the biometric data obtained by the brain and eyes, that assess the individual's cognitive load. The analysis of emotions during development activities is also research objective, but with less intensity. In the other physiological aspects and processes of productivity improvement research is still scarce.

The SQ1 presents two interesting aspects regarding the place and year of publication. Through Figure 5 it is possible to identify that from 2014 there was a significant advance in the number of articles published in the area. It is attributed that this increase is related to the increase of the processing capacity of the computers, larger number of wearable equipments and improvements in the artificial intelligence techniques. All these aspects are related to the topics covered by articles. This statistical question also points out that most articles are published at conferences. Conferences are preferred by computer science authors for the following reasons: (1) Conferences feature greater status, visibility, and impact; (2) Conferences have stricter selection criteria than most journals; (3) Conferences have a shorter time to publish than most journals. In certain journals, a publication may take years to be published and (4) Conferences have a higher novelty factor than journals.

Challenges for future research. From the analysis of these articles it is possible to identify future works and further challenges. The following study suggestions are presented below:

(1) *Use of sentiment and emotions:* Emotions are sentiments that affect the performance of our daily activities. The detection of emotions through the biometric data collected by different sensors presented in the primary studies of this article could allow the developers to perform their tasks in a more assertive way. Ostberg (A22) (Ostberg et al. 2017) proposes the use of methods and tools that alleviate the stress of the developer and with that increase his/her effectiveness in the work and performance in the execution of tasks. Wrobel (A13) (Wrobel 2013) conducted a survey to identify which emotions the developer felt during the work day. In this article, the author has identified the transformation of a positive emotional state into a negative state, which may negatively impact the outcome of the work performed. The author reports that the possibility of identifying a transformation of emotional state before it occurs and from this to provide means to avoid its transformation could improve the result of the activities performed.

Future research may use biometric data to identify emotions in order to predict negative emotions before they occur, promoting support to the developer to avoid them. Emotion detection can also be used by task recommendation systems to suggest tasks to developer perform according to his/her mental state to minimise failures during it execution.

(2) *Detection of quality concerns based on cognitive indicators:* Researches in Software Engineering indicate that cognitive processes can be used to detect quality concerns about source code. Fritz (A01) (Fritz et al. 2014) proposed a technique to identify when a developer was experiencing difficulty in his work, stopping his execution, before he could introduce a bug in the code. Fritz used in his technique biometric data collected by an eye-track, an electrodermal activity sensor and an electroencephalography sensor. The collected data were processed through a classifier to predict the difficulty of the tasks. Crk (A17) (Crk, Kluthe, and Stefik 2015) proposes a direct way to measure a developer's expertise by analysing the brain's electrical activity with the use of electroencephalography (EEG). The author classifies the expertise by analysing the programming comprehension, assigning classes of levels to carry out the classification. Parallel to studies that relate cognitive load to code comprehension and difficulty performing a task, Meyer (A06) (Meyer et al. 2017) conducted a survey to identify the activities that developers do during the day. From this research, the author was able to identify the activities performed by the developer during his/her work day.

Future research could identify the activities that one or several developers have and through the identification of their expertise regarding the tasks that need to be performed, recommending the tasks collaboratively aiming the quality concern.

(3) *A quality model for biometric data records:* Software Engineering has many quality models proposed in the last decade (Farias 2010; Farias et al. 2014); however, these models focus on software modelling in general (Lange 2007) and not on biometric data recording processes. A quality model for biometric data records designed to connect biometric data with quality

attributes is still lacking. The existence of a quality model that provides attributes that allow defining a process of acquisition of biometric data for the developers cognitive load can allow the researchers to focus their efforts in researches that evaluate the impact of the cognitive load in the development process as the influence of the working memory and language processing during development (A34) (Duraes et al. 2016), or the recommendation of task based on the difficulty rather than directing efforts to the process of collecting and extracting data.

Future research should answer questions as: (1) What attributes should the quality model aggregate to record biometric data to assist the developer in their activities? (2) Can only one quality model be used for all types of biometric data records in experiments in software engineering? (3) Or should specific models be proposed?

7. Threats to validity

This section discusses measures adopted to reduce threats to the validity of our results. Some aspects can threaten the validity of this study, including the validity of the construct, internal validity, and the validity of the statistical conclusions. In part, these threats are due to decisions made during the process of identification and review of the literature. For example, defining a search string in databases that does not result in relevant papers to the area, or the lack of adoption of criteria to select articles that bring greater contributions to the research. In this sense, we have analysed some threats related to internal, construct and conclusion validity.

Internal Validity. The major threat was the difficulty in establishing a relationship between the daily activities performed by the developers and the biometric data and psycho-physiological aspects that influence these activities. This threat was characterised by the lack of detail of the activities performed by developers, on the selected articles. To reduce this threat, the articles were thoroughly filtered by the inclusion and exclusion criteria presented in Section 3.3 and reviewed in detail by the authors of this research.

Construct validity. Incorrect classification and exclusion of relevant articles are imminent threats in the literature review. We try to minimise this problem by establishing a review protocol, with well-defined and auditable inclusion/exclusion criteria and also following the framework proposed by Ali and Usman (2018). As presented by Ali the number of systematic literature reviews in Software Engineering has increased. The reliability of these reviews depends on the rigour applied during the search strategy process. The lack of rigour during this process causes discrepancy in the primary studies included in the review. This discrepancy promotes revisions based on the same subject but performed by independent researchers result in different primary studies. Table 16 presents the check-list proposed by Ali highlighting which of the items are followed by this systematic mapping.

Conclusion validity. This threat is related to problems that may affect the reliability of our findings. In this context, the selected articles can be influenced by personal interests with the results of the study. In order to avoid this situation, the authors applied inclusion and exclusion criteria during the filtering process of the articles and took care to monitor any personal influence that could negatively impact the results. In addition to this threat, another threat to be considered is the article classification. In this case, a search string with the main synonyms of each search term was elaborated and whenever conflicting or doubtful classifications were found, the authors made a collaborative evaluation to find a consensus. Finally, all conclusions were made after the results were collected, avoiding any possibility of error (Wohlin et al. 2012).

External validity. He reflects on the validity of results obtained in other contexts. In practical terms, we should be concerned with understanding to what extent the findings of this study can be generalised to other realities or study configurations. Could some questions in this regard arise such as if other electronic databases, different search strings, and exclusion and inclusion criteria had been used would the same findings be produced? In this sense, we analysed whether our sample of studies could be changed significantly if, for example, search engines were changed. As we use 12 electronic databases, we believe that our sample would not be significantly altered as we take into account the main ones.

In addition, we could use the theory of proximal similarity (proposed by Campbell and Russo 1999) to draw the degree of generalisation of our results. Basically, the focal point is to define criteria that can be used for similar situations or configurations in which the found findings would be the same or even similar. Some criteria are shown as follows. First, the authors might make use of other electronic databases for retrieving potentially-relevant articles. Second, the biometrics data should be used as a resource to improve software engineering practices, including source code comprehension. Next, the search string should be similar to one used in our study, where changes in terms by using other synonyms would be allowed. Given that these changes may happen

Table 16. SLR validity check-list (adapated from Ali and Usman 2018).

Step	Reporting guideline	Guideline applied?
Choosing search strategy	[R1] Describe the search strategy	Yes
	Report the reason(s) for selecting a strategy e.g. is completeness critical for your topic of investigation.	No
Identifying known-set	[R3] Describe which of the approaches were used to identify a known-set of papers.	Yes
	[R4] In case manual-search was used for creating a 'known-set' document which journals and conference proceedings were considered for the search.	Do not apply
	[R5] Document the papers in the known-set.	Yes
	[R6] Document the reason(s) for choosing the source (s) of the known-set.	Yes
Search string construction	[R7] Document the keywords, alternative terms, spellings and synonyms.	Yes
	[R8] Document the sources and rationale for choosing and dropping keywords.	In parts
	[R9] Document the general search string.	Yes
	[R10] Document refinements (and the reasons for them) in the general search string based on trials.	In parts
	[R11] Document the time frame for literature covered in the search.	Yes
	[R12] Document the rationale for the selected time span.	Yes
	[R13] Document whether the search was done in the title, metadata or full-text.	Yes
	[R14] Document the reason for searching in title, metadata and or full-text.	No
	[R15] Document the level of recall and precision that were achieved with the final search string.	No
	[R16] Document why this level of precision/recall was considered acceptable. In particular, with recall less than 100% reflect on why the search string is good enough as is.	No
Source selection	[R17] Document the databases used for the search.	Yes
	[R18] Document the reasons why the selected databases are appropriate, and ensure coverage of the topic of interest.	Yes
	[R19] In case of using manual-search to supplement the search, report the names and years of the venues searched.	Do not apply
	[R20] Document the attempts to reduce publication bias.	No
	[R21] Document the gray literature considered for the study.	No
	[R22] Report the experts who were consulted for their unpublished work and what material was identified in this manner.	Do not apply
Protocol validation	[R23] Document if the protocol was independently validated and by whom.	No
	[R24] Document the suggestions made by the reviewer regarding the search strategy, and the actions that are taken to address these suggestions in the search protocol.	Yes
Conducting search	[R25] Document the database specific search strings.	Yes
	[R26] Document any additional filters used and the reasons for using them.	No
	[R27] Document any deviations in the database specific string from the general search string and the reasons for the change.	Yes
	[R28] Document the database specific number of search hits.	Yes
	[R29] Document the database specific search results.	Yes

in other studies, our findings may be generalised or accepted, at some point, to other contexts that are more similar to these requirements.

8. Related work

The use of biometric data to improve software engineering practices has increased and gained prominence in recent years. Academia recognises this importance, according to literature review studies that deal with this subject, such as Bateson et al. (2017), Cutmore and James (2007), Sharafi, Soh, and Guéhéneuc (2015), Blasco et al. (2016), Gui et al. (2019) and Bablani et al. (2019). Table 17 summarises the main features of these related works (RWs). These features were chosen because they are narrowly related to our research questions (described in Section 3.1).

RW1 (Bateson et al. 2017) presents a Categorization of Mobile EEG (CoME), based on the device mobility score due to the ambiguity of the mobility concept provided by the EEG devices. The categorisation assigns a score based on the mobility of the device in relation to the use off-body, assembly without additional equipment, mobility of the participant, movement restrictions, system specifications and number of channels. Twenty-nine studies were used to validate this categorisation. The result presented allows the comparison of EEG mobility in a standardised way between studies.

RW2 (Cutmore and James 2007) explored a review focusing on the technology used in the sensors currently used by psycho-physiologists for the collection of signals that inform us about cognitive processes, mental states and behavioural patterns. The review covered the classes of sensors: electric, magnetic, electrochemical, mechanical, thermal and optical. By evaluating how biometric data can be collected, this review helps to decide which sensor to use and facilitates the development of new sensors.

RW3 (Sharafi, Soh, and Guéhéneuc 2015) introduced a review of literature researching studies that use eye-track in Software Engineering. The review investigated studies in the period 1990 to the end of 2014 and identified 36 publications distributed in nine articles, two workshop papers and 25 conference articles. The review identified limitations of current technology that may be a

Table 17. Summarisation of related works.

ID	Title & Reference	Research method	Number of studies	Search protocol	Domain	Goal(s)	Period	Question/Dimensions addressed
RW01	Categorization of Mobile EEG: A researcher's Perspective (Bateson et al. 2017)	Literature Review	29	Yes	EEG data capture	Categorize Mobile or ambulatory EEG devices	Until 2017	A development of a scheme to classify EEG Devices, Discussions of results and shortcomings on the scheme for classification
RW02	Sensors and sensor systems for psychophysiological monitoring: A review of current trends (Cutmore and James 2007)	Literature Review	Does not Specify	No	Psychophysiological Monitoring	Review Psychophysiological sensors technology to help the choices for acquiring signals about cognitive processes	Until 2007	An analysis of sensor classes: electric, magnetic, electrochemical, mechanical, thermal, and optical.
RW03	A systematic literature review on the usage of eye-tracking in software engineering (Sharafi, Soh, and Guéhéneuc 2015)	Literature Review	36	Yes	Eye Tracking in Software Engineering	Evaluates the current state of the art of using eye-trackers in software engineering and provides evidence on the uses and contributions of eye-trackers to empirical studies in software engineering.	1990 to 2014	Eye-tracking measures, method for data-analysis, limitations and applications of eye-trackers in Software Engineering
RW04	A Survey of Wearable Biometric Recognition Systems (Blasco et al. 2016)	Survey	Does not Specify	No	Biometric Recognition	Explore specific issues that resides on wearable biometric recognition systems	Until 2016	Categorization of wearable sensors, technique for processing raw signals, machine learning techniques, and discussion of issues such as the biosignal quality, and lack of public datasets.
RW05	A Survey on Brain Biometrics (Gui et al. 2019)	Literature Review	Does not Specify	No	Brain Biometrics in Recognize Systems	Performs a detailed analysis of all factors necessary for the use of biometric brain data for authentication systems	2007 to 2017	Review of devices and methods for the acquisition of biometric data, review of public databases for studies, techniques for processing biometric data for applications, use of machine learning for applications that use biometric data, multimodal biometric data systems for recognition and security issues
RW06	Survey on Brain-Computer Interface: An Emerging Computational Intelligence Paradigm (Bablani et al. 2019)	Survey	Does not Specify	No	Brain data capture	Contextualize the anatomy and functioning of the brain, BCI systems and techniques that could be used for brain data	Until 2018	Brain regions, modes of signal, types of signal, feature extractions methods, classification algorithms used in developing BCI

threat to the validity of the conducted studies. It also identified the existence of different metrics based on eye movements used in the studies. The study concludes that even with limitations and threats, the use of eye-track is an easy and non-invasive way to obtain biometric data for Software Engineering.

RW4 (Blasco et al. 2016) defined a categorisation of portable sensors to capture biometric signals. For this categorisation the computational cost of different biometric signal processing techniques was analysed, which is an important factor in restricted devices. The proposed categorisation also reviewed the structure of the proposed biometric systems, configuration and results. Finally, the study presents a critique regarding the evaluation and viability of the sensors, as well as reflections and directions that can be approached in future works.

RW5 (Gui et al. 2019) conducted a literature review to evaluate studies on biometric systems based on brain activity. Recently brain electrical activity has been studied as a promising biometric approach because of the unique advantages of confidentiality, resistance to falsification, emotional and mental sensitivity. The study reviews the process of acquiring, collecting, processing and analysing brainwave signals, as well as public databases and classifiers. The emerging issues of the area such as multi-modality, safety, permanence and stability of the signal are also considered by the author.

RW6 (Bablani et al. 2019) examined the techniques used in brain-computer interfaces (BCI) such as electrocorticography (ECoG), electroencephalography (EEG), magnetoencephalography (MEG) and magnetic resonance imaging (MRI). This examination was done through a study on the techniques used for the data collection and also the extraction of characteristics and classification algorithms applied to this data. The result of the study was a comparative analysis of existing techniques and directions for future research.

Our analysis suggests that recent, extensive reviews have been conducted on the use of biometric data to support software engineering practices. Our study differs from them when introducing a support taxonomy to classify the use of biometric data in software development tasks, conducting a longitudinal analysis on the use of biometric data, sensors and psycho-physiological factors in software development practices, and outlining interesting discussions and challenges for future research.

9. Conclusion and future work

This article reported an SMS on the use of biometrics to support the daily activities of developers. We carried out a careful literature review using 12 widely used electronic databases. In total, 40 primary articles were selected after a careful filtering process applied to a sample of 3,023 potentially relevant studies. Our initial motivation was the lack of a comprehensive overview and understanding of past results in the field of biometrics hampers surveys concerning research gaps, challenges, and trends. Academia and industry may benefit from our findings when starting a new research, adopting a methodology to run a study considering a broader software engineering topic than was considered in an initial similar study, elaborating techniques aware of the emotional state of software developers, and choosing reusable research schemes, like the presented proposed SMS planning.

There are several future works that could be performed in the context of this systematic mapping. The main ones would be: (1) to consider new search engines to try to find new studies, allowing to create a larger map by taking into account possibly more studies; (2) define new research questions to increase the scope of the analysis; (3) outline new challenges that could be explored by the scientific community; and (4) seek more related work, which can be compared with ours through comparative criteria.

Finally, we also expect that the discussions pointed out here can motivate researchers to explore biometrics applied to software engineering further. This work can be considered an initial step on improving surveys on the use of biometric data records to improve software engineering practices.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES1) – Finance Code 001. We also thank CNPq [grant 313285/2018-7] for partially funding this research.

References

Aghajani, E., C. Nagy, O. L. Vega-Marquez, M. Linares-Vasquez, L. Moreno, G. Bavota, and M. Lanza. 2019. “Software Documentation Issues Unveiled.” *Proceedings – International Conference on Software Engineering 2019-May*, 1199–1210. doi:10.1109/ICSE.2019.00122

Ali, N. B., and M. Usman. 2018. “Re Liability of Search in Systematic Reviews: Towards a Quality Assessment Framework for the Automated-Search Strategy.” *Information and Software Technology* 99: 133–147. doi:10.1016/j.infsof.2018.02.002

Bablani, A., D. R. Edla, D. Tripathi, and R. Cheruku. 2019. “Survey on Brain-Computer Interface: An Emerging Computational Intelligence Paradigm.” *ACM Computing Surveys* 52 (1): 1–32. doi:10.1145/3297713

Bateson, A. D., H. A. Baseler, K. S. Paulson, F. Ahmed, and A. U. R. Asghar. 2017. “Categorisation of Mobile EEG: A

Researcher's Perspective." *BioMed Research International* 2017: 1.15
Bednarik, R., and M. Tukiainen. 2006. "An Eye-Tracking Methodology for Characterizing Program Comprehension Processes." *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications, ETRA '06*, 125–132. New York, NY; ACM. doi:10.1145/1117309.1117356.
Begel, A. 2016. "Invited Talk: Fun with Software Developers and Biometrics." *2016 IEEE/ACM 1st International Workshop on Emotional Awareness in Software Engineering (SEmotion)*, 1–2. doi:10.1109/SEmotion.2016.009
Bischoff, V., K. Farias, L. J. Gonçales, and J. L. V. Barbosa. 2019. "In Tegration of Feature Models: A Systematic Mapping Study." *Information and Software Technology* 105: 209–225. doi:10.1016/j.infsof.2018.08.016. http://www.sciencedirect.com/science/article/pii/S0950584916302178
Blasco, J., T. M. Chen, J. Tapiador, and P. Peris-Lopez. 2016. "A Survey of Wearable Biometric Recognition Systems." *ACM Computing Surveys* 49 (3): 1–35. doi:10.1145/2968215
Brooks Jr., F. P. 1995. *The Mythical Man-month (Anniversary Ed.)*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc.
Busjahn, T., R. Bednarik, A. Begel, M. Crosby, J. H. Paterson, C. Schulte, B. Sharif, and S. Tamm. 2015. "Eye Movements in Code Reading: Relaxing the Linear Order." *IEEE International Conference on Program Comprehension 2015-August*, 255–265. doi:10.1109/ICPC.2015.36
Campbell, D. T., and M. J. Russo. 1999. *Social Experimentation*. Ann Arbor, MI: University of Michigan.
Castelhano, J., I. C. Duarte, C. Ferreira, J. Duraes, H. Madeira, and M. Castelo-Branco. 2018. "The Role of the Insula in Intuitive Expert Bug Detection in Computer Code: An fMRI Study." *Brain Imaging and Behavior (CompTIA 2016)*, 1–15. doi:10.1007/s11682-018-9885-1
Cooper, I. D. 2016. "What is a 'Mapping Study?'." *Journal of the Medical Library Association: JMLA*104: 76–78. doi:10.3163/1536-5050.104.1.013
Crk, I., T. Kluthe, and A. Stefik. 2015. "Understanding Programming Expertise: An Empirical Study of Phasic Brain Wave Changes." *ACM Transactions on Computer-Human Interaction* 23 (1): 2:1–2:29. doi:10.1145/2829945
Cutmore, T. R. H., and D. A. James. 2007. "Sensors and Sensor Systems for Psychophysiological Monitoring: A Review of Current Trends." *Journal of Psychophysiology* 21 (1): 51–71. doi:10.1027/0269-8803.21.1.51
Duraes, J., H. Madeira, J. Castelhano, C. Duarte, and M. C. Branco. 2016. "WAP: Understanding the Brain at Software Debugging." *Proceedings – International Symposium on Software Reliability Engineering, ISSRE*, 87–92. doi:10.1109/ISSRE.2016.53
Fakhoury, S., Y. Ma, V. Arnaoudova, and O. Adesope. 2018. "The Effect of Poor Source Code Lexicon and Readability on Developers' Cognitive Load." *Proceedings of the 26th Conference on Program Comprehension, ICPC '18*,, 286–296. New York, NY; ACM. doi:10.1145/3196321.3196347
Farias, K. 2010. "Empirical Evaluation of Effort on Composing Design Models. *2010 ACM/IEEE 32nd International Conference on Software Engineering*, Vol. 2, 405–408. IEEE.
Farias, K., A. Garcia, C. Lucena, L. Gonzaga Jr., C. A. da Costa, R. da Rosa Righi, F. Basso, and T. Oliveira. 2014. "Towards a Quality Model for Model Composition Effort." *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, 1181–1183. ACM.
Fernández-Sáez, A. M., M. Genero, and M. R. V. Chaudron. 2013. "Empirical Studies Concerning the Maintenance of UML Diagrams and Their Use in the Maintenance of Code: A Systematic Mapping Study." *Information and Software Technology* 55 (7): 1119–1142.
Floyd, B., T. Santander, and W. Weimer. 2017. "Decoding the Representation of Code in the Brain: An fMRI Study of Code Review and Expertise." *Proceedings – 2017 IEEE/ACM 39th International Conference on Software Engineering, ICSE 2017*, 175–186. doi:10.1109/ICSE.2017.24
Fritz, T., A. Begel, S. C. Müller, S. Yigit-Elliott, and M. Züger. 2014. "Using Psycho-Physiological Measures to Assess Task Difficulty in Software Development." *Proceedings of the 36th International Conference on Software Engineering – ICSE 2014*, 402–413. doi:10.1145/2568225.2568266.
Fritz, T., and S. C. Muller. 2016. "Leveraging Biometric Data to Boost Software Developer Productivity." *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, 66–77. doi:10.1109/SANER.2016.107. http://ieeexplore.ieee.org/document/7476774/
Glücker, H., F. Raab, F. Echtler, and C. Wolff. 2014. "Eyede: Gaze-Enhanced Software Development Environments." *Proceedings of the Extended Abstracts of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI EA '14*, 1555–1560. New York, NY: ACM. doi:10.1145/2559206.2581217
Gonçales, L. J., K. Farias, M. Scholl, M. Roberto Veronez, and T. C. de Oliveira. 2015. "Comparison of Design Models: A Systematic Mapping Study." *International Journal of Software Engineering and Knowledge Engineering* 25 (09n10): 1765–1769.
Gonçales, L., K. Farias, B. d. Silva, and J. Fessler. 2019. "Measuring the Cognitive Load of Software Developers: A Systematic Mapping Study." *Proceedings of the 27th International Conference on Program Comprehension, ICPC '19*, 42–52. Piscataway, NJ: IEEE Press. doi:10.1109/ICPC.2019.00018
Gonçalves, M., L. de Souza, and V. González. 2011. "Collaboration, Information Seeking and Communication: An Observational Study of Software Developers' Work Practices." *Journal of Universal Computer Science* 17 (14): 1913–1930.
Gui, Q., M. V. Ruiz-Blondet, S. Laszlo, and Z. Jin. 2019. "A Survey on Brain Biometrics." *ACM Computing Surveys* 51 (6): 1–38. doi:10.1145/3230632
Kaklauskas, A., E. Zavadskas, M. Seniut, G. Dzemyda, V. Stankevic, C. Simkevičius, T. Stankevic, R. Paliskiene, A. Matuliauskaite, S. Kildiene, L. Bartkiene, S. Ivanikovas, and V. Gribniak. 2011. "We B-based Biometric Computer Mouse Advisory System to Analyze a User's Emotions and Work Productivity." *Engineering Applications of Artificial Intelligence* 24 (6): 928–945. doi:https://doi.org/10.1016/j.engappai.2011.04.006. http://www.sciencedirect.com/science/article/pii/S0952197611000662
Kevic, K., B. M. Walters, T. R. Shaffer, B. Sharif, D. C. Shepherd, and T. Fritz. 2015. "Tracing Software Developers' Eyes and Interactions for Change Tasks." *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering – ESEC/FSE 2015*, 202–213. doi:10.1145/2786805.2786864

Kevic, K., B. Walters, T. Shaffer, B. Sharif, D. Shepherd, and T. Fritz. 2017. “Eye Gaze and Interaction Contexts for Change Tasks – Observations and Potential.” *Journal of Systems and Software* 128: 252–266. doi:10.1016/J.JSS.2016.03.030. https://www.sciencedirect.com/science/article/pii/S0164121216000881

Konopka, M. 2015. “Combining Eye Tracking with Navigation Paths for Identification of Cross-language Code Dependencies.” *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering – ESEC/FSE 2015*, 1057–1059. doi:10.1145/2786805.2807561

Kramer, A. F. 1991. “Physiological Metrics of Mental Workload: A Review of Recent Progress.” *Multiple-task Performance*, 279–328. https://ci.nii.ac.jp/naid/10011685403/en/

Lange, C. 2007. “Assessing and Improving the Quality of Modeling: A Series of Empirical Studies about the UML.” PhD thesis, Department of Mathematics and Computer Science, proefschrift. doi:10.6100/IR629604.

Latoza, T. D., G. Venolia, and R. Deline. 2006. “Maintaining Mental Models: A Study of Developer Work Habits.” *Proceedings of the 28th International Conference on Software Engineering*, 492–501, doi:10.1145/1134285.1134355.

Lau, J., J. P. Ioannidis, and C. H. Schmid. 1997. “Quantitative Synthesis in Systematic Reviews.” *Annals of Internal Medicine* 127 (9): 820–826.

Lee, S., D. Hooshyar, H. Ji, K. Nam, and H. Lim. 2017. “Mining Biometric Data to Predict Programmer Expertise and Task Difficulty.” *Cluster Computing* 21: 1–11. doi:10.1007/s10586-017-0746-2.

Maalej, W., and H.-J. Happel. 2009. “From Work to Word: How Do Software Developers Describe Their Work?.” *2009 6th IEEE International Working Conference on Mining Software Repositories*, 121–130. IEEE. doi:10.1109/MSR.2009.5069490. http://ieeexplore.ieee.org/document/5069490/

Meyer, A. N., L. E. Barton, G. C. Murphy, T. Zimmermann, and T. Fritz. 2017. “The Work Life of Developers: Activities, Switches and Perceived Productivity.” *IEEE Transactions on Software Engineering* 43 (12): 1178–1193. doi:10.1109/TSE.2017.2656886. http://ieeexplore.ieee.org/document/7829407/

Meyer, A. N., T. Fritz, G. C. Murphy, and T. Zimmermann. 2014. “Software Developers’ Perceptions of Productivity.” *Proceedings of the 22Nd ACM SIGSOFT International Symposium on Foundations of Software Engineering, FSE 2014*, 19–29. New York, NY; ACM. doi:10.1145/2635868.2635892

Müller, S. 2016. “Using Biometric Sensors to Increase Developers’ Productivity.” PhD thesis, University of Zurich. doi:10.5167/uzh-126890

Müller, S. C., and T. Fritz. 2015. “Stuck and Frustrated Or in Flow and Happy: Sensing Developers’ Emotions and Progress.” *Proceedings – International Conference on Software Engineering* 1: 688–699. doi:10.1109/ICSE.2015.334

Müller, S. C., and T. Fritz. 2016. “Using (Bio)metrics to Predict Code Quality Online.” *Proceedings of the 38th International Conference on Software Engineering – ICSE ’16*, 452–463, New York, NY: ACM Press. doi:10.1145/2884781.2884803

Nakagawa, T., Y. Kamei, H. Uwano, A. Monden, K. Matsumoto, and D. M. German. 2014. “Quantifying Programmers’ Mental Workload During Program Comprehension based on Cerebral Blood Flow Measurement: A Controlled Experiment.” *Companion Proceedings of the 36th International Conference on Software Engineering – ICSE Companion 2014*, 448–451. doi:10.1145/2591062.2591098

Ostberg, J.-P., D. Graziotin, S. Wagner, and B. Derntl. 2017. “Towards the Assessment of Stress and Emotional Responses of a Salutogenesis-Enhanced Software Tool Using Psychophysiological Measurements.” Proceedings – 2017 IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering, SEmotion 2017 (SEmotion), 2–5. doi:10.1109/SEmotion.2017...4

Palmer, C., and B. Sharif. 2016. “Towards Automating Fixation Correction for Source Code.” *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications – ETRA ’16*, 65–68. doi:10.1145/2857491.2857544

Peitek, N., J. Siegmund, C. Parnin, S. Apel, and A. Brechmann. 2018. “Toward Conjoint Analysis of Simultaneous Eye-Tracking and fMRI Data for Program-Comprehension Studies.” Proceedings of the Workshop on Eye Movements in Programming – EMIP ’18, 1–5. doi:10.1145/3216723.3216725

Peitek, N., J. Siegmund, C. Parnin, S. Apel, J. C. Hofmeister, and A. Brechmann. 2018. “Simultaneous Measurement of Program Comprehension with fMRI and Eye Tracking: A Case Study.” *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM ’18*, 24:1–24:10. New York, NY: ACM. doi:10.1145/3239235.3240495

Perry, D. E., N. A. Staudenmayer, and L. G. Votta. 2002. “People, Organizations, and Process Improvement.” *IEEE Software* 11 (4): 36–45. doi:10.1109/52.300082

Petersen, K., R. Feldt, S. Mujtaba, and M. Mattsson. 2008. “Systematic Mapping Studies in Software Engineering.” Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, Vol. 17, 68–77.

Petersen, K., S. Vakkalanka, and L. Kuzniarz. 2015. “Guidelines for Conducting Systematic Mapping Studies in Software Engineering: An Update.” *Information and Software Technology* 64: 1–18.

Robillard, P. N. 1999. “The Role of Knowledge in Software Development.” *Communications of the ACM*42 (1): 87–92. doi:10.1145/291469.291476

Rodeghero, P., C. McMillan, P. W. McBurney, N. Bosch, and S. D’Mello. 2014. “Improving Automated Source Code Summarization via an Eye-tracking Study of Programmers.” *Proceedings of the 36th International Conference on Software Engineering – ICSE 2014*, 390–401. doi:10.1145/2568225.2568247

Rodríguez, P., A. Haghighatkhah, L. E. Lwakatare, S. Teppola, T. Suomalainen, J. Eskeli, T. Karvonen, P. Kuvaja, J. M. Verner, and M. Oivo. 2017. “Continuous Deployment of Software Intensive Products and Services: A Systematic Mapping Study.” *Journal of Systems and Software* 123: 263–291.

Sharafi, Z., Z. Soh, and Y.-G. Guéhéneuc. 2015. “A Systematic Literature Review on the Usage of Eye-tracking in Software Engineering.” *Information and Software Technology* 67: 79–107. doi:10.1016/j.infsof.2015.06.008. http://www.sciencedirect.com/science/article/pii/S0950584915001196

Sharif, B., and J. I. Maletic. 2010. "An Eye Tracking Study on Camelcase and Under-Score Identifier Styles." *IEEE International Conference on Program Comprehension*, 196–205. doi:10.1109/ICPC.2010.41.

Siegmund, J. 2016. "Program Comprehension: Past, Present, and Future." *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, 13–20. doi:10.1109/SANER.2016.35. http://ieeexplore.ieee.org/document/7476769/

Siegmund, J., A. Brechmann, S. Apel, C. Kastner, J. Liebig, T. Leich, and G. Saake. 2012. "Toward Measuring Program Comprehension with Functional Magnetic Resonance Imaging." *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering*, 24:1–24:4. doi:10.1145/2393596.2393624

Siegmund, J., C. Kästner, S. Apel, C. Parnin, A. Bethmann, T. Leich, G. Saake, and A. Brechmann. 2014. "Understanding Understanding Source Code with Functional Magnetic Resonance Imaging." *Proceedings of the 36th International Conference on Software Engineering – ICSE 2014*, 378–389. doi:10.1145/2568225.2568252

Siegmund, J., N. Peitek, C. Parnin, S. Apel, J. Hofmeister, C. Kästner, A. Begel, A. Bethmann, and A. Brechmann. 2017. "Measuring Neural Efficiency of Program Comprehension." *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering – ESEC/FSE 2017 11*, 140–150. doi:10.1145/3106237.3106268

Wieringa, R., N. Maiden, N. Mead, and C. Rolland. 2006. "Requirements Engineering Paper Classification and Evaluation Criteria: A Proposal and a Discussion." *Requirements Engineering* 11 (1): 102–107. doi:10.1007/s00766-005-0021-6

Wohlin, C., P. Runeson, M. Hst, M. C. Ohlsson, B. Regnell, and A. Wessln. 2012. *Experimentation in Software Engineering*. New York, NY: Springer Publishing Company, Incorporated.

Wrobel, M. R. 2013. "Emotions in the Software Development Process." 2013 6th International Conference on Human System Interactions (HSI) (June 2013), 518–523. doi:10.1109/HSI.2013.6577875. http://ieeexplore.ieee.org/document/6577875/.

Wrobel, M. 2018. "Applicability of Emotion Recognition and Induction Methods to Study the Behavior of Programmers." *Applied Sciences* 8 (3): 323. doi:10.3390/app8030323. http://www.mdpi.com/2076-3417/8/3/323

Züger, M., and T. Fritz. 2015. "Interruptibility of Software Developers and its Prediction Using Psycho-Physiological Sensors." *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems – CHI '15*, 2981–2990. New York, NY: ACM Press. doi:10.1145/2702123.2702593

Züger, M., S. C. Müller, A. N. Meyer, and T. Fritz. 2018. "Sensing Interruptibility in the Office: A Field Study on the Use of Biometric and Computer Interaction Sensors." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18*, 1–14. doi:10.1145/3173574.3174165