

Usage of Psychophysiological Data as an Improvement in the Context of Software Engineering: A Systematic Mapping Study

Roger Denis Vieira

Universidade do Vale do Rio dos Sinos
São Leopoldo, Rio Grande do Sul
rogervi@edu.unisinos.br

Kleinner Farias

Universidade do Vale do Rio dos Sinos
São Leopoldo, Rio Grande do Sul
kleinnerfarias@edu.unisinos.br

RESUMO

A systematic mapping had been developed seeking for answering six research questions regarding using of psychophysiological data in the Software Engineering domain. Employing it was selected 2084 studies from 7 digital databases, reaching a final amount of 27 articles. Mostly of experimental nature, marked by usage of the sensing devices and integrated development environments (IDEs), looking for analysis of developers' psychophysiological attributes applied to comprehension and debugging of source code. It was not identified as native mechanisms to support psychophysiological data integration and analysis within the integrated development environment.

CCS CONCEPTS

• **Applied computing** → **Bioinformatics**; • **Software and its engineering** → **Formal software verification**; *Integrated and visual development environments*; *Software testing and debugging*;
• **Human-centered computing** → Human computer interaction (HCI).

KEYWORDS

Data Analysis, Software Engineering, Neurosciences, Bioinformatics, Data Processing

1 INTRODUÇÃO

O desenvolvimento de *software* é um processo lógico e intelectual, onde cada desenvolvedor aplica o seu conhecimento embasado por premissas lógicas para a produção de soluções voltadas a problemas reais, seja na área corporativa, acadêmica, ou ainda, em problemas corriqueiros do cotidiano. Tal desenvolvimento demanda que o profissional esteja focado e imerso em seu trabalho, atentando a cada detalhe que possa impactar no comportamento da solução que está sendo produzida [7].

Não obstante, os investimentos em torno do processo de desenvolvimento de *software*, suas metodologias e tecnologias, é inevitável a geração de defeitos inerentes ao comportamento do produto a medida que a sua complexidade aumenta. Este fenômeno pode ser atribuído a diversos fatores, como o processo intelectual envolvido na produção de código, o conjunto de ferramentas e técnicas inadequadas, além de problemas intrínsecos às definições de requisitos atrelados ao negócio. Sob a perspectiva em torno dos fatores humanos, autores advogam que o desenvolvedor, enquanto ser humano, é passível de falhas, e que estas são naturais e esperadas durante a execução das suas tarefas [7].

Estudos buscam compreender quais processos cognitivos estão envolvidos durante a produção de *software*, fazendo o uso de tecnologias como Eletroencefalografia (EEG) e Imagem por Ressonância Magnética Funcional (fMRI) [7, 10, 30]. Os autores destes estudos advogam que é possível encontrar relações entre a produção de defeitos em artefatos de *software* e os processos cognitivos envolvidos no momento da escrita de código, visando criar mecanismos capazes de priorizar a refatoração de código de baixa confiabilidade e até a predição na confecção de possíveis artefatos defeituosos.

Com o advento da microeletrônica, processadores menores, mais velozes e de baixo consumo energético têm surgido, possibilitando que dispositivos móveis e acessíveis para a coleta de sinais biológicos sejam difundidos em experimentos acadêmicos. Esses são os casos do NeuroSky Mindwave Mobile 2 [28] e o Emotiv EPOC e EPOC+ [27] que são utilizados para aumentar a compreensão acerca dos processos cognitivos envolvidos durante o processo de codificação. Aliados à extensibilidade dos atuais IDEs, a utilização de tais dispositivos pode permitir o desenvolvimento de novas funcionalidades que permitam a identificação de padrões de comportamento dos desenvolvedores acerca da sua produtividade e qualidade de artefatos produzidos [12, 27, 37].

Na última década, dados psicofisiológicos têm sido amplamente utilizados como recurso de melhoria no contexto de engenharia de *software* [16]. No entanto, a literatura atual carece de uma classificação e um mapeamento sistemático dos trabalhos já realizados e publicados abordando este tema [10, 16]. Consequentemente, um entendimento mais detalhado do uso de tais dados permanece limitado. Apoiado por tal argumento, o presente estudo visa responder à seguinte questão de pesquisa: *como dados psicofisiológicos estão sendo utilizados como recurso de melhoria no contexto de Engenharia de Software?*

Este trabalho, portanto, foca em classificar e fornecer uma análise temática dos estudos já publicados sobre o uso de dados psicofisiológicos como recurso de melhoria no contexto de Engenharia de *Software*. Para isso, um mapeamento sistemático da literatura foi projetado e realizado, seguindo diretrizes bem estabelecidas [2, 5, 17, 26] para responder a seis questões de pesquisa. No total, 2084 estudos foram inicialmente obtidos de 7 bases de dados digitais, sendo 27 trabalhos selecionados para responder às questões de pesquisa formuladas. Como resultados, observou-se que a maioria dos estudos como de natureza experimental, caracterizada pelo uso de dispositivos para coleta de dados psicofisiológicos e IDEs, visando a análise de atributos psicofisiológicos dos desenvolvedores aplicados, principalmente, à compreensão e depuração de código fonte. Ademais, não foram identificados mecanismos de suporte à integração e análise de dados psicofisiológicos de forma nativa nos IDEs. Por fim, este trabalho identifica alguns desafios, os quais

poderão ser explorados pela comunidade científica nos próximos anos.

Inicialmente este estudo trará uma visão geral em torno dos principais trabalhos correlatos à temática proposta (Seção 2). Em seguida, serão detalhadas as etapas escolhidas para o desenvolvimento do mapeamento sistemático (Seção 3), metodologia esta adotada para a condução do estudo, tendo em vista à sua capacidade de fornecer um panorama em acerca de determinado tópico de pesquisa. Posteriormente, a condução do processo de seleção, filtragem e extração dos dados contidos nos estudos selecionados será apresentada (Seção 4). Por fim, os resultados encontrados serão discutidos à luz da atual literatura (Seções 5 e 6) e as conclusões obtidas serão explicitadas, assim como uma visão prévia de estudos futuros (Seção 7).

2 TRABALHOS RELACIONADOS

Esta seção apresenta uma análise comparativa deste artigo com a literatura atual. Para isso, vários artigos foram identificados após aplicar a *string* de busca “Psychophysiological Data AND Software Engineering” na Google Scholar¹, sendo sete selecionados pela similaridade em relação à temática explorada neste estudo. Tais estudos são brevemente analisados a seguir.

2.1 Análise dos Trabalhos Relacionados

Fritz *et al* (2014) [11]: investigou-se uma nova abordagem para a classificação da dificuldade na compreensão de código-fonte, utilizando dados biométricos associados a estados psicofisiológicos. Por meio de um experimento envolvendo 15 desenvolvedores, utilizou-se *Eye-trackers*, sensores de atividade eletrodermal (EDA) e eletroencefalógrafos (EEG) a fim de verificar como os dados coletados poderiam prover uma forma de predição quando um desenvolvedor consideraria uma tarefa como sendo difícil.

Fritz *et al* (2016) [13]: dissertou-se em torno de uma série de experimentos controlados realizados em estudos anteriores, os quais aplicavam sensores diversos para a coleta de dados. Discorreram também sobre as estratégias adotadas para coleta, preparação, transformação, carga e análise dos dados, além de elucidarem os algoritmos de aprendizado de máquina aplicados para a extração de *features* relevantes aos objetivos de cada estudo.

Fucci *et al* (2019) [14]: os autores replicaram experimentos previamente apresentados na literatura [10], mediante o uso de dispositivos sensores de baixo custo. Para tanto, realizaram medições utilizando EEG, EDA e BVP (*Blood Volume Pulse*) em uma amostra de 28 estudantes de Ciência da Computação enquanto estes executavam tarefas envolvendo a compreensão de textos escritos em linguagem natural e código-fonte.

Gonçalves *et al* (2019) [16]: foi desenvolvido um mapeamento sistemático com o objetivo de identificar lacunas nos trabalhos que abordavam o estudo de carga cognitiva no contexto de Engenharia de *Software*. Foram analisado 33 artigos relacionados que interseccionavam as temáticas de carga cognitiva, engenharia de *software*, uso de dispositivos sensores para coleta de dados psicofisiológicos.

Gui *et al* (2019) [18]: realizou-se um levantamento literário entre os anos de 2007 e 2017 no que diz respeito ao estado da arte acerca da utilização de dados biométricos cerebrais, os tipos de sensores

utilizados, como esses dados eram coletados, os seus contextos de aplicação e as abordagens estatísticas e de aprendizado de máquina para a extração de *features* em torno da utilização de tais dados.

Radveski *et al* (2015) [27]: propôs-se um modelo conceitual de arquitetura objetivando a aquisição de dados em tempo real provenientes de EEGs de baixo custo. Realizou-se também um teste piloto, no qual foram utilizados os dispositivos Emotiv EPOC e EPOC+ a fim de avaliar critérios de qualidade, tais como usabilidade e confiabilidade. Ademais, os autores dissertaram em torno das implicações éticas envolvidas na coleta de dados psicofisiológicos em tempo real para a avaliação da produtividade de desenvolvedores.

Rostami *et al* (2015) [28]: por meio de um estudo experimental, foi mensurada a carga cognitiva e a familiaridade de um grupo de estudantes do curso de Ciência da Computação no que se refere à interação destes com IDEs, utilizando um dispositivo EEG de baixo custo para a coleta de dados psicofisiológicos. Os autores apontaram ainda, as vantagens do EEG em relação à técnicas similares, e as limitações técnicas dos dispositivos sensores de baixo custo.

2.2 Análise Comparativa e Oportunidades

Foram definidos cinco Critérios de Comparação (CC) para auxiliar no processo de identificação das similaridades e diferenças entre o trabalho proposto e os artigos selecionados. Esta comparação é crucial para tornar o processo de identificação de oportunidades de pesquisa utilizando critérios objetivos, ao invés de subjetivos. Os critérios são descritos a seguir:

- **Mapeamento Sistemático (CC01):** estudos que desenvolveram um mapeamento sistemático da literatura visando adquirir uma visão geral do estado da arte.
- **Integração de Dados (CC02):** trabalhos que pesquisaram como os dados psicofisiológicos estão sendo suportados pelos IDEs.
- **Métricas Psicofisiológicas (CC03):** nesta categoria estão os estudos que pesquisaram sobre métricas ou procedimentos utilizados para mensurar, seja qualitativamente ou quantitativamente, aspectos psicofisiológicos.
- **Aplicações na Engenharia de Software (CC04):** estudos que buscaram avaliar como dados psicofisiológicos são aplicados na Engenharia de *Software*.
- **Ambientes Integrados de Desenvolvimento (CC05):** estudos que exploraram ou melhoraram IDEs através do uso de dados psicofisiológico.

Tabela 1: Análise comparativa dos Trabalhos Relacionados selecionados

Trabalho Relacionado	Critério de Comparação				
	CC1	CC2	CC3	CC4	CC5
Trabalho proposto	●	●	●	●	●
Fritz <i>et al</i> (2014) [11]	○	○	●	●	○
Fritz <i>et al</i> (2016) [13]	○	○	●	●	○
Fucci <i>et al</i> (2019) [14]	○	○	●	●	○
Gonçalves <i>et al</i> (2019) [16]	●	○	●	●	○
Gui <i>et al</i> (2019) [18]	○	○	●	○	○
Radveski <i>et al</i> (2015) [27]	○	●	●	●	○
Rostami <i>et al</i> (2015) [28]	○	●	●	●	●

● Atende Completamente ● Atende Parcialmente ○ Não Atende

¹<https://scholar.google.com/>

A Tabela 1 apresenta a comparação dos trabalhos selecionados, contrastando os mesmos com o trabalho proposto. Algumas lacunas e oportunidades de pesquisas são observadas:

- (1) apenas o trabalho proposto foi o único a atender completamente todos os critérios de comparação;
- (2) métricas psicofisiológicas e aplicações em Engenharia de *Software* foram os temas mais explorados, enquanto mapeamento sistemático e IDEs os menos explorados;
- (3) nenhum estudo explorou detalhadamente o uso de dados psicofisiológicos no contexto de ambientes integrados de desenvolvimento.

Portanto, a seguinte oportunidade de pesquisa foi identificada: uma análise holística da literatura que contemple os cinco critérios de comparação estabelecidos. Esta oportunidade é explorada nas próximas seções.

3 PROCESSO DE MAPEAMENTO

O mapeamento sistemático, enquanto método para revisão da literatura, visa identificar um conjunto de estudos capazes de indicar lacunas na literatura acerca de determinada temática, trazendo oportunidades de pesquisa [2, 26].

3.1 Objetivo e Questões de Pesquisa

O objetivo deste trabalho é classificar e fornecer uma análise temática dos estudos publicados sobre o uso de dados psicofisiológicos como recurso de melhoria no contexto de Engenharia de *Software*. Para atingir este objetivo, seis Questões de Pesquisa (QPs) foram formuladas para explorar diferentes facetas deste objetivo. A Tabela 2 apresenta tais QPs, suas motivações e as variáveis exploradas.

Tabela 2: Questões de Pesquisa (QPs), suas descrições e variáveis relacionadas

Questão de Pesquisa	Motivação	Variável
QP01: Quais tipos de dispositivos têm sido utilizados na coleta de dados psicofisiológicos?	Identificar os dispositivos utilizados nos estudos selecionados para a coleta de dados.	Dispositivos
QP02: Quais tipos de atributos são extraídos pelos dispositivos?	Elucidar os tipos de atributos extraídos e o que eles representam no escopo dos estudos analisados.	Atributos
QP03: Como os dados extraídos são suportados pelos IDEs?	Explicar como os dados extraídos podem ser integrados nos IDEs.	Integração
QP04: Com quais propósitos os dados extraídos têm sido utilizados em IDEs?	Revelar os propósitos de uso de dados psicofisiológicos em IDEs.	Propósitos
QP05: Quais são as principais contribuições dos estudos?	Explicitar as contribuições dos estudos acerca da utilização de dados psicofisiológicos no contexto de Engenharia de <i>Software</i> .	Contribuições
QP06: Qual a natureza dos estudos analisados?	Compreender a natureza do estudo quanto ao seu objetivo.	Natureza

3.2 Estratégia de Busca

A Estratégia de Busca utilizada neste estudo foi concebida em duas etapas: (1) seleção das bases de dados digitais relacionadas à temática do estudo; e (2) construção de uma *string* de busca e suas

variantes, respeitando as peculiaridades dos mecanismos de busca presentes em cada base de dados.

3.2.1 Bases de Dados. A Tabela 3 apresenta bases de dados eletrônicas utilizadas. Elas foram escolhidas por serem amplamente utilizadas e por estudos anteriores de mapeamento da literatura terem demonstrado efetividade das mesmas [5, 17]. Ademais, a seleção de bases de dados foi embasada na cobertura dos seus mecanismo de busca, os quais retornaram trabalhos dos principais meios de publicação relacionados à Engenharia de *Software*.

Tabela 3: Bases de dados e seus endereços digitais

Base de Dados	Endereço Eletrônico
ACM Digital Library	http://portal.acm.org/
DBLP	https://dblp.uni-trier.de/
IEEE Xplore	http://ieeexplore.ieee.org
Science Direct	https://www.sciencedirect.com/
Scopus	https://www.scopus.com/
Semantic Scholar	https://www.semanticscholar.org/
Springer Link	http://www.springerlink.com/

3.2.2 String de Busca. Para a confecção da *string* de busca, foram adotados os seguintes passos para o desenvolvimento da mesma: (1) definir os termos principais (*Major Terms*); (2) pesquisar pelos sinônimos relacionados ou ainda, palavras com equivalência semântica (*Minor Terms*); e (3) combinar os termos principais aos seus sinônimos utilizando operadores lógicos, tais como “AND” e “OR”. Aplicando-se as etapas citadas, foram definidos os termos descritos na Tabela 4, chegando-se assim à seguinte *string* de busca:

(eeg OR electroencephalography OR neural OR brain OR cognitive OR psychometric OR bioinformatic) AND (“software OR developer” OR programmer OR professional OR “software engineering”) AND (ide OR code OR editor OR “integrated development environment”)

Tabela 4: Termos primários da string de busca e os seus sinônimos

Termo Principal	Sinônimos
Neuroscience	eeg, fmri, electroencephalography, “functional magnetic resonance imaging”, neural, brain, cognitive, psychometric, bioinformatic
Software Engineering IDE	software, developer, programmer, professional, “software engineering”
Data Analysis	ide, code, editor, “integrated development environment” “data analysis”, “signal analysis”, analytic, “signal processing”

3.2.3 Critérios de Inclusão e Exclusão. Tais critérios buscam definir um método objetivo para filtrar os artigos inicialmente recuperados das bases de dados. Estes critérios auxiliam no processo de construção de uma amostra de artigos relevantes, os quais são fundamentais para responder às QPs formuladas.

Enquanto os Critérios de Exclusão (CEs) ajudam no descarte de estudos não relevantes para as QPs [26], os Critérios de Inclusão (CIs) ajudam na construção de uma amostra consistente de artigos relevantes [26]. Sendo assim, foram definidos como Critérios de Inclusão os seguintes itens:

- **CI01:** Publicação entre os anos de 2009 e 2019;
- **CI02:** Escrita na língua inglesa;
- **CI03:** Disponível de forma completa nas bases de dados digitais;
- **CI04:** Relação com a *string* de busca e Questões de Pesquisa;

Ademais, para a filtragem dos estudos previamente selecionados por meio dos Critérios de Inclusão, foram aplicados os seguintes Critérios de Exclusão:

- **CE01:** Coerência com a *string* de busca mas fora do contexto da pesquisa;
- **CE02:** Inexistência de resumo;
- **CE03:** Apenas um resumo ou *Call for Paper* de conferências;
- **CE04:** Sem relação com o domínio de Engenharia de *Software*;
- **CE05:** Cópias/duplicatas ou versões antigas de estudos já selecionados;
- **CE06:** Impossibilidade de acesso ao artigo completo; e
- **CE07:** Em desacordo com a motivação da pesquisa.

3.3 Estratégias para a Extração de Dados

Para a etapa de extração de dados, foram definidos os mecanismos: (1) esquemas de classificação e (2) formulário para a extração de dados. Tais mecanismos visam prover, de forma padronizada, o processo de extração e classificação dos dados contidos nos trabalhos selecionados, possibilitando assim uma identificação confiável de potenciais respostas às Questões de Pesquisa [26].

Objetivando a confecção de um mecanismo efetivo para a classificação de estudos, entre as possíveis abordagens para a extração de categorias utilizadas no esquema de classificação, neste estudo foram adotadas as seguintes:

- **Esquema Independente de Tópico:** categorias genéricas são utilizadas de forma a prover uma visão geral em torno dos estudos selecionados e a sua temática. Usualmente os autores recomendam a utilização de categorias já adotadas em outros mapeamento sistemáticos;
- **Esquema Dependente de Tópico:** questões específicas emergem enquanto os estudos são filtrados, trazendo assim um visão detalhada do seu conteúdo, sendo uma peculiaridade presente em cada mapeamento sistemático;

Em posse das categorias que seriam utilizadas para a classificação dos estudos, realizou-se a confecção do formulário de extração de dados. Essa ferramenta tinha por objetivo padronizar o processo de extração dos dados, permitindo assim que os dados fossem analisados de forma quantitativa posteriormente.

4 EXECUÇÃO DO PROCESSO DE FILTRAGEM

O processo de filtragem utilizado neste trabalho deu-se em oito etapas distintas, apresentadas por meio da Figura 1. Excetuando-se a etapa de Pesquisa Inicial, a qual utilizou apenas os Critérios de Inclusão na triagem, as demais utilizaram agrupamentos dos Critérios de Exclusão como diretrizes para a remoção dos estudos inicialmente selecionados.

A seguir, estão descritas as etapas adotadas, assim como as suas diretrizes de condução:

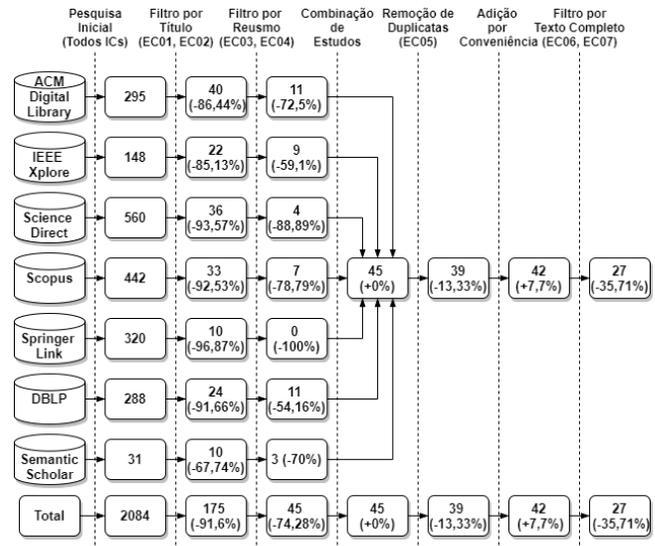


Figura 1: Processo de seleção dos estudos

- (1) **Pesquisa Inicial:** caracterizada pela inserção da *string* de busca (Seção 3.2.2) nas bases de dados escolhidas (Seção 3.2.1), combinada aos filtros compostos parametrizados segundo os Critérios de Inclusão (Seção 3.2.3), com isto, a busca retornou um montante de 2084 artigos.
- (2) **Filtro por Título:** foram analisados os títulos dos artigos resultantes. A filtragem se deu pelos Critérios de Exclusão CE01 e CE02 (Seção 3.2.3), reduzindo o montante em 91,6%, restando o total de 175 trabalhos.
- (3) **Filtro por Resumo:** realizada a leitura dos resumos dos 175 artigos resultantes. Foram aplicados como os Critérios de Exclusão CE03 e CE04 (Seção 3.2.3), reduzindo o volume de estudos em 74,28%, chegando assim a um total de 45 artigos.
- (4) **Combinação de Estudos:** conforme ilustrado pela Figura 1, os rótulos das bases de dados foram removidos e os estudos, combinados, fazendo com que estes compusessem um único *corpus* literário.
- (5) **Remoção de Duplicatas:** aplicado o Critério de Exclusão CE05, foram removidos 13,39% de estudos duplicados e/ou versões antigas dos mesmos, resultando assim em 39 artigos (Figura 1).
- (6) **Adição por Conveniência:** foram adicionados por conveniência 3 artigos relacionados com a temática e dentro dos Critérios de Inclusão e Exclusão, implicando em um aumento de 7,7% do total, ou ainda, 42 artigos (Figura 1). Tal adição deu-se por conhecimento dos autores e por meio de análises encadeada das referências presentes nos artigos já analisados (*snowballing*) [26].
- (7) **Filtro pelo Texto Completo:** os artigos foram lidos integralmente, aplicando-se os Critérios de Exclusão CE06 e CE07 (Seção 3.2.3), acarretando na remoção de 35,71% dos artigos e gerando um montante final de 27 trabalhos (Figura 1).
- (8) **Seleção dos Estudos:** a fase final do processo resultou na seleção de 27 artigos, nomeados neste trabalho como Estudos

Primários, os quais serviram de insumo para a extração de dados, segundo os critérios apresentados na Seção 3.3.

5 RESULTADOS

Para apresentação e descrição dos resultados referentes a cada Questão de Pesquisa, foram utilizadas tabelas contendo a distribuição dos Estudos Primários em relação às respectivas categorias de respostas, conforme elucidados pelas seções a posteriores.

5.1 QP01: Dispositivos

A Tabela 5 apresenta os resultados referentes à QP01. É possível observar que na maioria dos estudos envolvendo a coleta de dados psicofisiológicos (29,63%), são utilizados eletroencefalógrafos (EEG) como dispositivo de coleta de sinais. Tal fato pode ser atribuído à redução do custo dos aparelhos de eletroencefalografia, além do desenvolvimento de versões acessíveis pelo público não especializado, democratizando o seu uso em diversos contextos [28].

Tabela 5: Distribuição dos estudos por dispositivos utilizados (QP01)

Dispositivo	Trabalhos	Percentual	Estudos Primários
EEG	8	29,63%	[1, 6, 15, 22, 27, 28, 32, 35]
fMRI	5	18,52%	[7, 10, 25, 30, 31]
Eye-tracker	5	18,52%	[3, 4, 19, 24, 29]
Combinação de Dispositivos	4	14,81%	[8, 9, 11, 12]
Sem Dispositivos	3	11,11%	[23, 36, 38]
fNIRS	2	7,41%	[20, 33]

Em seguida, *Eye-trackers* e *Imagem por Ressonância Magnética Funcional* representaram, igualmente, 18,52% dos dispositivos utilizados nos Estudos Primários. Nota-se ainda o uso de *Espectroscopia Funcional de Infravermelho Próximo (fNIRS)* como dispositivo de coleta (7,41%).

Representaram 14,81% dos estudos analisados, aqueles em que os experimentos executados empregaram mais de um dispositivo para coleta de dados, usualmente EEGs combinados com *Eye-trackers* ou dispositivos complementares, como pulseiras e faixas peitorais para medição da frequência cardiopulmonar. Tais dispositivos foram responsáveis por complementar as medições em cenários onde o *Eye-tracker* e o EEGs não foram capazes de indicar, de forma isolada, atributos como *stress* e atenção focada [4].

Dada a natureza de alguns dos Estudos Primários, assim como a metodologia adotada, não foi constatada a utilização de dispositivos para a coleta de dados. Tal gama de estudos representaram 11,11% do volume total.

5.2 QP02: Atributos

Segundo os dados apresentados na Tabela 6, observa-se a predominância (29,63%) de trabalhos que tiveram como objeto de estudo a atividade neural dos indivíduos. Tais trabalhos fizeram uso de dispositivos como EEGs, fNIRS e fMRI para avaliar quais regiões cerebrais eram ativadas durante a execução de experimentos, além de relacionar semânticas às atividades cerebrais observadas [21].

Estudos que dissertaram sobre *Estados Mentais* representaram 25,93%. Nestes, optou-se primordialmente pelo uso de EEGs para a

Tabela 6: Distribuição dos estudos por atributos analisados (QP02)

Atributo	Trabalhos	Percentual	Estudos Primários
Atividade Neural	8	29,63%	[7, 10, 15, 20, 25, 30, 31, 33]
Estados Mentais	7	25,93%	[1, 6, 22, 27, 28, 32, 35]
Fixação/Atenção	5	18,52%	[3, 4, 19, 24, 29]
Sem Atributos	3	11,12%	[23, 36, 38]
Atributos Combinados	2	7,40%	[11, 12]
Fixação/Atenção, Atividade Neural	2	7,40%	[8, 9]

observação dos estados mentais, tendo em vista a ligação destes a ondas cerebrais (Alfa, Beta, Gama, Delta e Theta) [32].

A *Fixação/Atenção* foi o atributo estudado em 18,52% dos artigos analisados. A fixação foi atribuída pelos estudos ao foco ocular em pontos de interesse, predominantemente mensurado por meio de *Eye-trackers* (Tabela 5), nos casos em que a atenção era objeto de estudos, fez-se o uso de sensores complementares de forma concomitante.

Em 7,40% do trabalhos, realizou-se a análise combinada dos atributos de *Fixação/Atenção* e *Atividade Neural*. Neste estudos, observou-se que o enfoque dos autores permeava a compreensão de código-fonte, para tanto, foram utilizados *Eye-trackers* para a coleta de dados relativos ao posicionamento e/ou foco ocular e fNIRS, a fim de identificar a ativação neural de áreas cerebrais ligadas ao aprendizado e leitura [8, 9].

Mais de três atributos foram analisados em 7,40% dos estudos, nos quais experimentos com maior especificidade foram realizados, como por exemplo, a mensuração de *Fixação/Atenção*, *Estados Mentais* e atributos complementares durante a execução de tarefas de desenvolvimento [11].

5.3 QP03: Integração

Relativo ao total de 27 estudos analisados, 11,1% destes foram classificados como *Não Aplicável*, ou seja, foram estudos que fizeram algum tipo de uso de dados psicofisiológicos sem a utilização de IDEs em seus experimentos. Por outro lado, cerca de 88,9% dos estudos foram enquadrados na categoria *Não Suportado*, isto é, realizaram experimentos envolvendo IDEs e também a análise de dados psicofisiológicos. Ainda sim, nenhum destes estudos especificou mecanismos capazes de integrar os dados coletados aos IDEs, implicando em análises de dados realizadas de forma externa e por meio de ferramentas especializadas de terceiros.

5.4 QP04: Propósitos

Mediante os dados elucidados pela Tabela 7, constatou-se que, em sua maioria (66,70%), os Estudos Primários tinham como enfoque analisar a *Compreensão de Código-fonte* e as *Estratégias de Depuração* utilizadas pelos desenvolvedores nos experimentos realizados.

Observou-se ainda que, em 14,81% dos Estudos Primários, o propósito focal do trabalho foi a avaliação do desenvolvedor enquanto este executava atividades relacionadas ao seu contexto. É válido ressaltar que, apesar da possibilidade de avaliar a Compreensão de Código-fonte e a estratégia de depuração de forma individual, as

Tabela 7: Distribuição dos estudos por propósito adotado (QP04)

Propósito	Trabalhos	Percentual	Estudos Primários
Compreensão de Código-fonte e Estratégias de Depuração	18	66,70%	[3, 4, 7–11, 19, 20, 22, 24, 25, 29–31, 35, 36, 38]
Carga Cognitiva	4	14,81%	[6, 15, 28, 33]
Interfaces Cérebro-Computador	2	7,41%	[1, 32]
Melhoria na Produtividade	2	7,41%	[12, 27]
Fora do Contexto	1	3,70%	[23]

medições relacionadas a estes propósitos possuem uma maior validade quando atreladas à *Carga Cognitiva*, segundo levantamentos desenvolvidos por outros autores [16].

Constituindo 7,41% dos Estudos Primários estavam àqueles cujo propósito principal era a busca pela *Melhoria na Produtividade* dos desenvolvedores mediante o uso de dados psicofisiológicos e IDEs.

Trabalhos envolvendo o desenvolvimento de *Interfaces Cérebro-Computador* corresponderam a também 7,41% dos Estudos Primários. Esta categoria engloba a construção de um protótipo de dispositivo para o controle de periféricos domésticos por meio de ondas cerebrais [32] e a idealização de um EEG portátil capaz de comunicar-se com *smartphones* [1].

5.5 QP05: Contribuições

A Tabela 8 traz a disposição das categorias de respostas relacionadas à QP05. Nota-se que a maior parte dos estudos primários contribuiu com alguma *Métrica* (37,04%), seja aplicado aos dados psicofisiológicos, como no processo de desenvolvimento de *software*. Em seguida, as categorias *Modelo* e *Método* representaram ambas 22,22% do total de estudos primários selecionados. Contribuições do tipo *Processo* compuseram 11,11% do total, e por último, apenas 2 estudos selecionados (7,41%) trouxeram algum tipo de *Ferramenta*.

Tabela 8: Distribuição dos estudos por tipo de Contribuições (QP05)

Contribuição	Trabalhos	Percentual	Estudos Primários
Métrica	10	37,04%	[4, 6, 7, 12, 19, 20, 22, 29–31, 36]
Modelo	6	22,22%	[1, 24, 27–29, 32]
Método	6	22,22%	[3, 8–10, 33, 35]
Processo	3	11,11%	[15, 25, 38]
Ferramenta	2	7,41%	[11, 23]

5.6 QP06: Natureza

Nesta Questão de Pesquisa buscou-se trazer um panorama acerca das naturezas dos selecionados, utilizando-se as estratégias de extração de categorias e classificação [34] apresentadas na Seção 3.3. Os resultados obtidos pelo processo de condução do mapeamento sistemático estão organizados e dispostos na Tabela 9.

Observa-se que predominantemente os estudos selecionados (55,56%) são de natureza *Experimento Controlado*, ou seja, estudos em que os autores realizaram algum tipo de experimento de forma a

Tabela 9: Distribuição dos estudos por Natureza (QP06)

Natureza	Trabalhos	Percentual	Estudos Primários
Experimento Controlado	15	55,56%	[3, 4, 6, 7, 9–12, 15, 19, 20, 22, 30, 31, 35]
Proposta de Solução	7	25,93%	[8, 23, 27, 32, 33, 36]
Artigo de Opinião	4	14,81%	[24, 25, 29, 38]
Pesquisa Avaliação	1	3,70%	[28]

validar uma técnica já existente ou ainda, corroborar algum conceito encontrado na literatura. Em seguida estão os estudos de natureza *Proposta de Solução*, representando 25,93% do montante total. Nesta natureza de estudos, estão aqueles em que o principal objetivo foi o de apresentar uma solução para um problema já existente ou uma versão melhorada de uma solução já apresentada, por meio de uma prova de conceito. Os *Artigos de Opinião* compuseram 14,81%, trazendo trabalhos que expressam a opinião dos autores acerca de determinado tópico, instigando a discussão e abrindo oportunidade para novos trabalhos. Por fim, 3,70% dos trabalhos foram de natureza *Pesquisa Avaliação*, onde buscou-se avaliar uma ferramenta, metodologia ou ideia apresentada por outros autores.

É possível constatar que, no decorrer da última década – período avaliado no desenvolvimento deste trabalho – houve um aumento no número de publicações acerca da temática.

6 DISCUSSÃO E DESAFIOS

A Figura 2 apresenta um gráfico de bolhas onde os estudos primários são organizados em três facetas: contribuição (eixo horizontal esquerdo), propósitos (eixo vertical) e natureza (eixo horizontal direito). Tal gráfico traz uma visão combinada dos estudos primários, objetivando identificar as relações entre as contribuições de cada estudo, as suas naturezas e os propósitos adotados. Sendo assim, fazendo uso de tal recurso, foi possível identificar os seguintes aspectos de discussão e desafios:

- *Carência de Suporte à Integração de Dados Psicofisiológicos*: os estudos que realizaram experimentos utilizando IDEs e dados psicofisiológico, mostraram que não existem mecanismos nativos de suporte a tais tipos de dados por parte dos IDEs, tampouco por meio de extensões que adicionem tal funcionalidade. Neste aspecto, os dados eram coletados por ferramentas proprietárias que comunicavam-se com os dispositivos sensores, para posteriormente serem pré-processados e analisados. Tal carência de suporte à integração mostra-se oportuna para a utilização de ferramentas de *Big Data* que trabalhem com grandes volumes de dados e/ou processamento de fluxos de dados em tempo real. Ademais, a produção de uma tecnologia infraestrutura para a integração de dados psicofisiológicos aos IDEs por meio de funcionalidades extensíveis, mostra-se como uma necessidade presente em estudos experimentais aplicados à Engenharia de *Software*, tanto na indústria como sob o viés acadêmico.
- *Experimentos de Compreensão de Código-fonte e Estratégias de Depuração Aplicadas à Engenharia de Software*: os autores dos estudos revisados neste trabalho adotaram como propósito principal, em sua maioria, o uso dos dados psicofisiológicos para avaliação da compreensão de código-fonte por parte

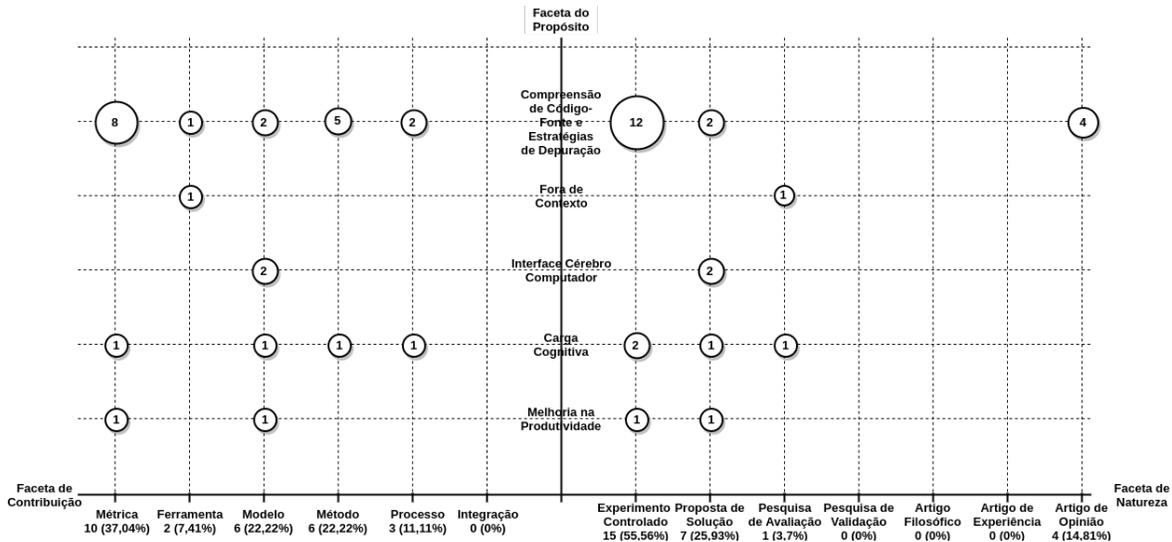


Figura 2: Densidade de estudos pela sua natureza, propósito e contribuições.

dos desenvolvedores, além das suas estratégias adotadas para o processo de depuração. Esta preferência elucida-se pelo domínio dos estudos analisados, os quais usualmente buscavam aplicabilidade no contexto de Engenharia de *Software*. Fica evidente que este tipo de propósito foi adotado nos estudos que fizeram uso de experimentos controlados como metodologia principal.

- *Análises de Dados por Meio de Ferramentas Externas*: apesar dos estudos terem realizado análises de dados para obtenção dos resultados e desenvolvimento das suas discussões e conclusões, observou-se a necessidade de análises externas ao ambiente controlado do experimento, mediante o uso de ferramentas como MATLAB, ou ainda, soluções proprietárias desenvolvidas pelos fornecedores dos dispositivos sensores utilizados nos experimentos. Esta característica evidenciou-se nos experimentos em que foram utilizados IDEs. Nestes experimentos, além dos dados coletados pelos dispositivos, demandou-se que as métricas obtidas em IDEs fossem correlacionadas aos dados psicofisiológicos de forma independente. No aspecto apresentado, encontra-se uma oportunidade para o desenvolvimento de extensões às IDEs que possibilitem a utilização de técnicas de *Data Analytics* de forma nativa, permitindo assim a extração de *insights* em torno dos dados psicofisiológicos de forma integrada ao contexto de uso das IDEs.
- *Crescimento do Interesse por Dados Psicofisiológicos Aplicados à Engenharia de Software*: é possível observar um interesse emergente por este tópico de pesquisa, dado o aumento de publicações nos últimos anos (Figura 3). Não obstante, as integrações multidisciplinares entre Neurociências e Engenharia de *Software* abrem margem para a produção de estudos diversificados e inovadores no que tange a aplicabilidade da psicofisiologia aliada ao processo de desenvolvimento de *software*. Este fenômeno torna-se observável por meio da pluralidade de contribuições oriundas dos artigos analisados,

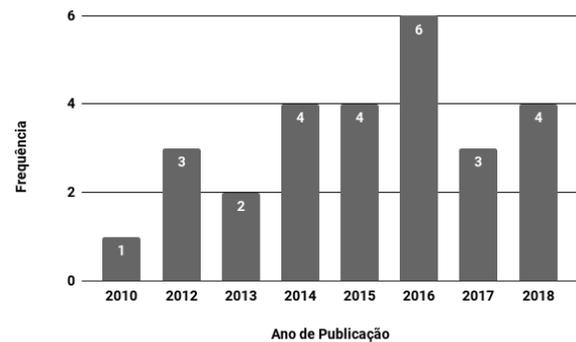


Figura 3: Distribuição dos estudos pelos anos de publicação

assim como pelos trabalhos relacionados utilizados (Item 2), os quais serviram de embasamento teórico no decorrer da produção deste mapeamento sistemático.

7 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresentou um mapeamento sistemático da literatura, o qual buscou classificar e fornecer uma análise temática dos estudos já publicados sobre o uso de dados psicofisiológicos como recurso de melhoria no contexto da engenharia de *software*. No total, 27 estudos foram selecionados após um processo de filtragem de 2084 estudos inicialmente obtidos de 7 bases de dados digitais. Seis questões de pesquisa foram exploradas, sendo os dados obtidos analisados e discutidos. Além disso, desafios foram identificados, os quais poderão ser explorados pela comunidade científica.

Em trabalhos futuros, pretende-se desenvolver mecanismos que possibilitem a coleta e processamentos de dados psicofisiológicos oriundos de dispositivos sensores, assim como a sua integração aos IDEs, de forma a possibilitar a geração de metadados que possam ser relacionados aos eventos gerados dentro dos IDEs. Em posse de tais dados, ainda será possível aplicar técnicas de aprendizagem de

máquina para a obtenção de padrões que possam sugerir potenciais falhas na produção de *software* de acordo com o padrão psicofisiológico de cada desenvolvedor, possibilitando a criação de IDEs inteligentes e adaptáveis ao usuário, além de recursos táticos para a priorização de refatoração em unidades de *software*.

8 AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), código 313285/2018-7.

REFERÊNCIAS

- [1] Abdel Ilah N. Alshbatat, Peter J. Vial, Prashan Premaratne, and Le C. Tran. 2014. EEG-based Brain-computer Interface for Automating Home Appliances. *Journal of Computers* 9, 9 (2014), 2159–2166.
- [2] Kitchenham Barbara and Charters Stuart. 2007. Guidelines for performing Systematic Literature Reviews in Software Engineering.
- [3] Roman Bednarik. 2012. Expertise-dependent visual attention strategies develop over time during debugging with multiple code representations. *International Journal of Human Computer Studies* 70, 2 (2012), 143–155.
- [4] Tanya Beelders and Jean-Pierre du Plessis. 2016. The Influence of Syntax Highlighting on Scanning and Reading Behaviour for Source Code. Association for Computing Machinery (ACM), 1–10.
- [5] Vinicius Bischoff, Kleinner Farias, Lucian José Gonçalves, and Jorge Luis Victória Barbosa. 2019. Integration of feature models: A systematic mapping study. *Information and Software Technology* 105, October 2016 (2019), 209–225.
- [6] Igor Crk and Timothy Kluthe. 2014. Toward using alpha and theta brain waves to quantify programmer expertise. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014* (2014), 5373–5376.
- [7] J. Duraes, H. Madeira, J. Castelhana, C. Duarte, and M. Castelo Branco. 2016. WAP: Understanding the Brain at Software Debugging. *Proceedings - International Symposium on Software Reliability Engineering, ISSRE* (2016), 87–92.
- [8] Sarah Fakhoury. 2018. Moving towards objective measures of program comprehension. *26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (2018), 936–939.
- [9] Sarah Fakhoury, Yuzhan Ma, Venera Arnaudova, and Olusola Adesope. 2018. The effect of poor source code lexicon and readability on developers' cognitive load. *26th Conference on Program Comprehension* (2018), 286–296.
- [10] Benjamin Floyd, Tyler Santander, and Westley 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* (2017), 175–186.
- [11] Thomas Fritz, Andrew Begel, Sebastian C. Müller, Serap Yigit-Elliott, and Manuela Züger. 2014. Using psycho-physiological measures to assess task difficulty in software development. *36th International Conference on Software Engineering* (2014), 402–413.
- [12] Thomas Fritz and Sebastian C. Müller. 2016. Leveraging Biometric Data to Boost Software Developer Productivity. *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering Leveraging* (2016), 66–77.
- [13] Thomas Fritz and Sebastian C. Müller. 2016. Leveraging Biometric Data to Boost Software Developer Productivity. *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)* 5 (2016), 66–77.
- [14] Davide Fucci, Daniela Girardi, Nicole Novielli, Luigi Quaranta, and Filippo Lanubile. 2019. A Replication Study on Code Comprehension and Expertise using Lightweight Biometric Sensors. *27th International Conference on Program Comprehension* (2019). arXiv:1903.03426
- [15] Gregory Funke, Benjamin Knott, Vincent F. Mancuso, Adam Strang, Justin Estep, Lauren Menke, Rebecca Brown, Allen Dukes, and Brent Miller. 2013. Evaluation of subjective and EEG-based measures of mental workload. *Communications in Computer and Information Science* 373, PART I (2013), 412–416.
- [16] Lucian Gonçalves, Kleinner Farias, Bruno Silva, and Jonathan Fessler. 2019. Measuring the Cognitive Load of Software Developers: A Systematic Mapping Study. *27th International Conference on Program Comprehension* March (2019).
- [17] Lucian José Gonçalves, Kleinner Farias, Toacy Cavalcante De Oliveira, and Murilo Scholl. 2019. Comparison of Software Design Models: An Extended Systematic Mapping Study. *ACM Computing Surveys (CSUR)* 52, 3 (2019), 48.
- [18] Qiong Gui, Maria V. Ruiz-Blondet, Sarah Laszlo, and Zhanpeng Jin. 2019. A Survey on Brain Biometrics. *Comput. Surveys* 51, 6 (2019), 1–38.
- [19] Prateek Hejmady and N Hari Narayanan. 2012. Visual attention switching patterns of programmers debugging with an IDE. *ETRA '12 Symposium on Eye Tracking Research and Applications* (2012), 197–200.
- [20] Yoshiharu Ikutani and Hidetake Uwano. 2014. Brain activity measurement during program comprehension with NIRS. *2014 IEEE/ACIS 15th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2014 - Proceedings* (2014), 1–6.
- [21] Richard D Lane, Gereon R Fink, Phyllis M-L Chau, and Raymond J Dolan. 1997. Neural activation during selective attention to subjective emotional responses. *Neuroreport* 8, 18 (1997), 3969–3972.
- [22] Seolhwa Lee, Andrew Matteson, Danial Hooshyar, Songhyun Kim, Jaebum Jung, Gichun Nam, and Heuseok Lim. 2016. Comparing Programming Language Comprehension between Novice and Expert Programmers Using EEG Analysis. *Proceedings - 2016 IEEE 16th International Conference on Bioinformatics and Bioengineering, BIBE 2016* (2016), 350–355.
- [23] Francesco Logozzo, Michael Barnett, F Manuel, and Patrick Cousot. 2012. A Semantic Integrated Development Environment. *3rd annual conference on Systems, programming, and applications: software for humanity* (2012), 15–16.
- [24] Unaiyah Obaidallah, Mohammed Al Haek, and Peter C.-H. Cheng. 2018. A Survey on the Usage of Eye-Tracking in Computer Programming. *Comput. Surveys* 51, 1 (2018), 1–58.
- [25] Norman Peitek. 2018. A neuro-cognitive perspective of program comprehension. *2018 ACM/IEEE 40th International Conference on Software Engineering: Companion Proceedings* (2018), 496–499.
- [26] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology* 64 (2015), 1–18.
- [27] Stevche Radevski, Hideaki Hata, and Kenichi Matsumoto. 2015. Real-time monitoring of neural state in assessing and improving software developers' productivity. *Proceedings - 8th International Workshop on Cooperative and Human Aspects of Software Engineering, CHASE 2015* (2015), 93–96.
- [28] Shahin Rostami, Alex Shenfield, Stephen Sigurnjak, and Oluwatoyin Fakorede. 2015. Evaluation of Mental Workload and Familiarity in Human Computer Interaction with Integrated Development Environments using Single-Channel EEG. *Psychology of Programming Interest Group 2015 - 26th Annual Workshop*. 1, July (2015), 7–22.
- [29] Zohreh Sharafi, Zéphyrin Soh, and Yann-gaël Gaël Guéhéneuc. 2015. A systematic literature review on the usage of eye-tracking in software engineering. *Information and Software Technology* 67 (2015), 79–107.
- [30] Janet Siegmund, Christian Kästner, Sven Apel, Chris Parnin, Anja Bethmann, Thomas Leich, Gunter Saake, and André Brechmann. 2014. Understanding understanding source code with functional magnetic resonance imaging. *36th International Conference on Software Engineering* (2014), 378–389.
- [31] Janet Siegmund, Norman Peitek, Chris Parnin, Sven Apel, Johannes Hofmeister, Christian Kästner, Andrew Begel, Anja Bethmann, and André Brechmann. 2017. Measuring neural efficiency of program comprehension. *11th Joint Meeting on Foundations of Software Engineering* (2017), 140–150.
- [32] Harold Szu, Charles Hsu, Gyu Moon, Takeshi Yamakawa, Binh Q. Tran, Tzyy Ping Jung, and Joseph Landa. 2013. Smartphone Household Wireless Electroencephalogram Hat. *Applied Computational Intelligence and Soft Computing* 2013 (2013), 1–8.
- [33] Murat Pasa Uysal. 2016. Evaluation of learning environments for object-oriented programming: measuring cognitive load with a novel measurement technique. *Interactive Learning Environments* 24 4820, May 2015 (2016).
- [34] Roel Wieringa, Neil Maiden, Nancy Mead, and Colette Rolland. 2006. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements engineering* 11, 1 (2006), 102–107.
- [35] Martin K.C. Yeh, Dan Gopstein, Yu Yan, and Yanyan Zhuang. 2017. Detecting and comparing brain activity in short program comprehension using EEG. *Proceedings - Frontiers in Education Conference, FIE 2017-October*, 1444827 (2017), 1–5.
- [36] M Yin, B Li, and C Tao. 2010. Using cognitive easiness metric for program comprehension. *2nd International Conference on Software Engineering and Data Mining, SEDM 2010* (2010), 134–139.
- [37] Iyad Zayour and Hassan Hajjidiab. 2013. How much integrated development environments (IDEs) improve productivity? *Journal of Software* 8, 10 (2013), 2425–2431.
- [38] Iyad Zayour and Ali Hamdar. 2016. A qualitative study on debugging under an enterprise IDE. *Information and Software Technology* 70 (2016), 130–139.