CONTEXT-AWARE PREDICTION OF USER PERFORMANCE PROBLEMS CAUSED BY THE SITUATIONALLY-INDUCED IMPAIRMENTS AND DISABILITIES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MEHMET ELGIN AKPINAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

JULY 2022

Approval of the thesis:

CONTEXT-AWARE PREDICTION OF USER PERFORMANCE PROBLEMS CAUSED BY THE SITUATIONALLY-INDUCED IMPAIRMENTS AND DISABILITIES

submitted by **MEHMET ELGIN AKPINAR** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ______________

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering** ______________

Prof. Dr. Pınar Karagöz
Supervisor, **Computer Engineering, METU** ______________

Assoc. Prof. Dr. Yeliz Yeşilada
Co-supervisor, **Computer Engineering, METU NCC** ______________

Examining Committee Members:

Prof. Dr. Kürşat Çağıltay
Computer Education and Instructional Technology, METU ______________

Prof. Dr. Pınar Karagöz
Computer Engineering, METU ______________

Prof. Dr. Tolga Can
Computer Engineering, METU ______________

Prof. Dr. Hacer Karacan
Computer Engineering, Gazi University ______________

Assist. Prof. Dr. Cemil Zalluhoğlu
Computer Engineering, Hacettepe University ______________

Date: 28.07.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Mehmet Elgin Akpınar

Signature :

ABSTRACT

CONTEXT-AWARE PREDICTION OF USER PERFORMANCE PROBLEMS CAUSED BY THE SITUATIONALLY-INDUCED IMPAIRMENTS AND DISABILITIES

Akpınar, Mehmet Elgin
Ph.D., Department of Computer Engineering
Supervisor: Prof. Dr. Pınar Karagöz
Co-Supervisor: Assoc. Prof. Dr. Yeliz Yeşilada

July 2022, 239 pages

When we interact with small screen devices, we can make errors due to our abilities/disabilities, contextual factors that distract our attention, or problems related to the interface. Predicting and learning these errors based on the previous user interaction and contextual factors and adapting the user interface to prevent these errors can improve user performance and satisfaction. This thesis aims to understand how a system can be developed that monitors user performance and contextual changes and predicts performance problems based on context. In this thesis, we first conducted a systematic review to understand the context and its effect on user performance. Then, we conducted a user study in the wild to collect text entry interactions, sensor data, and context labels. First, we used this data to measure user performance regarding typing speed and error rate in an automated system without a predefined task model. Moreover, we investigated how different context dimensions affect user performance. Our findings showed that context affects users differently; therefore, user-specific adaptations should be considered. We also investigated whether different context factors can be sensed using available sensors. Our experiments with machine learning algorithms using available smartphone sensors show that we can differentiate different context factors, particularly mobility and environment. Moreover, we used sensor and user performance data to predict performance problems. The regression model to predict typing errors outperformed the random baseline. Finally, based on these informative studies, we propose adaptation techniques and design guidelines for developers to support user interaction in different contexts.

Keywords: context, smartphones, text entry, user study

ÖZ

DURUMSAL KAYNAKLI BOZUKLUKLARIN VE YETERSİZLİKLERİN NEDEN OLDUĞU KULLANICI PERFORMANS SORUNLARININ BAĞLAMA DUYARLI TAHMİNİ

Akpınar, Mehmet Elgin
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Pınar Karagöz
Ortak Tez Yöneticisi: Doç. Dr. Yeliz Yeşilada

Temmuz 2022 , 239 sayfa

Küçük ekranlı cihazlarla etkileşime girdiğimizde yeteneklerimiz/engellerimiz, dikkatimizi dağıtan bağlamsal faktörler veya arayüzle ilgili sorunlar nedeniyle hatalar yapabiliriz. Bu hataları önceki kullanıcı etkileşimi ve bağlamsal faktörlere dayalı olarak tahmin etmek, öğrenmek ve bu hataları önlemek için kullanıcı arayüzünü uyarlamak, kullanıcı performansını ve memnuniyetini artırabilir. Bu tez çalışması, kullanıcı performansını ve bağlamsal değişiklikleri izleyen ve bağlama dayalı olarak performans sorunlarını öngören bir sistemin nasıl geliştirilebileceğini araştırmaktadır. Bu çalışma kapsamında, öncelikle bağlamı ve bunun kullanıcıların performansı üzerindeki etkisini anlamak için sistematik bir literatür taraması yapıldı. Ardından, metin girişi etkileşimlerini, sensör verilerini ve kullanıcıların mevcut bağlam sınıflarını toplamak amacıyla, kullanıcıların kendi ortamlarında katıldıkları bir kullanıcı çalışması gerçekleştirildi. İlk olarak, toplanan bu veriler, önceden tanımlanmış bir görev modeli olmadan otomatik bir sistem aracılığıyla kullanıcı performansını yazma hızı ve hata oranı bazında ölçmek için kullanıldı. Hesaplanan kullanıcı performansı verileri

ve kullanıcıların bağlam atamaları kullanılarak, farklı bağlam boyutlarının kullanıcı performansını nasıl etkilediği araştırıldı. Bulgularımız bağlamın kullanıcıları farklı şekilde etkilediğini gösterdi; bu nedenle, kullanıcıya özel uyarlamalar dikkate alınmalıdır. Mevcut sensörler kullanılarak farklı bağlam faktörlerinin algılanıp algılanamayacağı da araştırıldı. Ayrıca, performans sorunlarını tahmin etmek için sensör ve kullanıcı performans verileri kullanıldı. Yazma hatalarını tahmin etmeye yönelik regresyon modelinin, rastgele yapılan tahminlere göre daha iyi performans gösterdiği görüldü. Son olarak, yapılan çalışmalara dayanarak, farklı bağlamlarda kullanıcının etkileşimini desteklemek için uyarlama teknikleri araştırıldı, derlendi ve geliştiriciler için tasarım önerileri olarak sunuldu.

Anahtar Kelimeler: bağlam, akıllı telefonlar, metin girişi, kullanıcı çalışması

To my dearest daughter, Nehir

ACKNOWLEDGMENTS

I would like to thank my supervisors, Yeliz Yeşilada and Pınar Karagöz, for their continuous support, patience, and motivation. Their guidance and positive behavior helped me in the research and writing this thesis.

Besides my advisors, I would like to thank the thesis monitoring committee members, Kürşat Çağıltay and Tolga Can, for their constructive comments. I would also like to thank my former supervisor, Selim Temizer, for all his support. I also thank the examining committee members Hacer Karacan and Cemil Zalluhoğlu for their precious time reading my thesis and for their constructive comments.

I would like to thank my former employer in TechNarts, Taha Yaycı; former manager in Comodo, Faruk Önen; former manager in Mobiliz, Ozan Ali Kaya and current manager in Mobiliz, Zeynep Yelboğa, for providing me with time to study whenever I needed it.

My special thanks go to Emre Baykal, Alperen Ustaömer, Özerk Sarıoğlu, İrem Durmaz, Yoonji Jung Demir, Onur Çinar, İlke Ozan Öztürk, Bekir Atay, Çağdaş Baş, Funda İpek Atay, Mehmet Serkan Narlı, Gülgün Pekmezci, Hasan Deniz, Oğuz Süllü, Damla Süllü, Merve İyras, Melda Arslan, Kübra Kaya Aydan and all other participants who voluntarily participated in our user study. I also thank Hüseyin Ateş and Beytullah Özkan for their technical support.

Finally, I would like to deeply thank my family: Nihan Özdemirler Akpınar, Nehir Akpınar, Zübeyde Akpınar, Hasan Akpınar, Ekin Akpınar, Necla Özdemirler, Selçuk Özdemirler, and Akın Özdemirler, for their unconditional love and support. My deepest thanks are to Nihan Özdemirler Akpınar and Nehir Akpınar. Throughout this study, they motivated me with their never-ending patience, tolerance, and understanding.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ANOVA	ANalysis Of VAriance
API	Application Programming Interface
CNN	Convolutional Neural Network
DT	Decision Tree
ER	Error Rate
ESM	Experience Sampling Method
GPS	Global Positioning System
HCI	Human-Computer Interaction
ID	Insignificant Difference
IoT	Internet of Things
KNN	K-Nearest Neighbors
KSPC	Keystroke Per Character
KSPS	Keystrokes Per Second
LSTM	Long Short Term Memory
MLP	Multilayer Perceptron
NLP	Natural Language Processing
PDA	Personal Digital Assistants
RF	Random Forest
SD	Significant Difference
SIID	Situationally Induced Impairments And Disabilities
SVM	Support-Vector Machine
UMPC	Ultra-Mobile PCs
URL	Uniform Resource Locator
WPM	Words Per Minute

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

Over the last decade, smartphones started to play a significant role in our daily lives. Their use reached almost 3.8 billion users in 2021 with a drastic increase [1]. An average smartphone user checks their device 58 times and spends about three hours with his/her smartphone daily [2]. They are essential in people's lives so that people can no longer leave their homes without having their smartphones with them [3]. Smartphones are no longer just used for communication but also to perform most of the daily tasks [4]. With a few taps, we can read the news, watch a movie, chat with friends, or spend time on social media [5]. Asynchronous communication has been widespread recently [6], as a result, smartphones are widely used for text entry tasks, such as writing text messages or emails [7].

The small screen sizes and portability of smartphones and recent advances in Internet technologies enable smartphones to be used almost everywhere. Moreover, while using smartphones, the users might be engaged with different and parallel tasks [8]. A user can check the news on his/her smartphone while having breakfast. Similarly, a user can send the location of a restaurant to his/her friends while walking to that restaurant. The user's primary task with the smartphone, the parallel task user is engaged with, the environment and the surroundings are different in these examples. As a result, the context around smartphones includes more complex and diverse dimensions than desktop computers [9].

There have been many attempts to define the term "context" [10]. In this thesis, we refer to context as "any information that characterizes a situation related to the interaction between humans, applications, and the surrounding environment" [11] (p. 106). Based on this definition, Greenberg [12] highlights the dynamic nature of context and the challenges of determining the information required to identify a contextual state.

Using a smartphone itself can cause performance problems similar to those experienced by users with motor impairments due to its small screen [13]. Furthermore, some factors related to the user's current context can cause performance problems. In the literature, temporary reductions in user performance due to context are referred to as situationally-induced impairments and disabilities (SIIDs) [14]. This phenomenon was defined as "difficulty accessing computers due to the context or situation one is in, as opposed to a physical impairment" [15, 16]. Unfortunately, the application designers and developers still do not consider these difficulties due to context and SIIDs [9].

The effect of these situational impairments can be reduced with adaptive systems, which change themselves for context or user behavior [14]. According to this approach, users do not adapt to a system; instead, the system adapts itself based on their performance. For example, suppose a user has problems with clicking on a target. In that case, the system may increase the target size or adjust the mouse settings to prevent the error. For this purpose, a continuous approach to observing user performance changes and contextual factors might be used.

The primary purpose of this thesis is to understand the effect of context on user performance, implement a sensing mechanism to collect performance and context data and predict user performance problems related to the context. Moreover, based on these, this thesis also aims to propose techniques to adapt the system to realize an ability-based system with the support of our prediction mechanism. The system can suggest adaptations whenever a performance problem related to the context is predicted. Mobile operating systems or specific applications can adapt themselves to the users to help them maintain their performance. As a result, these adaptations can reduce the negative effect of SIIDs on users' performance.

This thesis work started with a systematic review to understand the context and its effect on users' performance. In contrast with the broad definition of context, the research on the effect of context on users' performance has been limited to a few contextual factors, such as mobility. Our systematic review identified five contextual dimensions: environment, mobility, social, multitasking, and distraction. Most existing studies have been based on experimental tasks conducted in controlled environments. Participants are typically asked to complete predefined tasks under different contextual factors in controlled laboratory settings. This approach, of course, provides a consistent way of measuring speed and errors made. However, this approach can also miss some difficulties in real-world usage [17]. Furthermore, in controlled studies, users can only use specific interaction methods, and this restriction may also jeopardize the validity of these studies [18].

Collecting data from actual users' contexts where the users' are not restricted to performing a specific task can address these kinds of issues. This approach is referred to as "in the wild" (or in-situ) studies. It aims to observe users' behavior in their natural contexts [19]. However, collecting data unobtrusively also has some challenges. Reproducibility is an issue [18]. Since the users' intentions are not fully known, the reliability of performance measurement can also be questioned [17]. Even though these are essential issues to consider, the existing literature also shows that it is possible to conduct an in-situ user study without a specific task model and still detect errors with a good accuracy [18].

After our systematic review, we conducted an in-situ remote user study to investigate context's effect on users' text entry performance in real-world settings. Real-world text entry data is collected from 48 participants during their everyday interactions. During this study, we collected the participants' text entry data, sensor data, and context labels. To compare the user's performance under different conditions, we needed to interpret user performance in several metrics. One crucial metric was typing errors to measure users' performance. Several studies have identified typing errors in the wild; however, these studies had some limitations. For instance, daily texting language was not considered in these approaches. Therefore, we combined several existing approaches to detect typing errors and distinguish between edits and corrections using the text entry data. Finally, we investigated the effect of context on user performance by combining text entry data and context labels in five dimensions: environment, mobility, social, multitasking and distraction.

Our user study collected data from available sensors in participants' smartphones. We asked them to label the current context in five dimensions. Using this sensor dataset and context labels, we investigated which contextual factors can be identified using available sensor data. For this purpose, we compared a set of classification models with different parameters. This comparison helped us to identify relevant sensors to distinguish contextual factors. Finally, combining the user performance and sensor data, we investigated how we can predict user performance problems. In this part, we applied regression models to individual user data.

1.2 Research Questions

This thesis focuses on the following research questions:

R1. *"Which contextual factors have been examined in the literature for smartphone interaction that can cause SIIDs?"* – This question aims to identify and bring together all the contextual factors that have been examined in the literature centered around SIIDs. Context is one of the components defining interaction along with user, task, and technology. This research question would enable other researchers and us to see what kind of factors need to be considered or the factors that have not been considered at all.

R2. *"What are the most effective techniques to measure users' typing performance regarding speed and error rate?"* – This question aims to develop a system that automatically measures the typing speed and error rate using an input stream. There are existing error detection approaches in the literature [17, 18]. However, these approaches do not consider some specific cases, such as using text-speak. These approaches can be combined and extended with additional rules to detect typing errors more accurately in daily settings. It would enable other researchers and us to detect performance problems without manual investigation.

R3. *"How do different contextual factors affect smartphone users' performance?"* – This question aims to see the overall effect of the contextual factors on the users' performance. Our literature review revealed that previous studies investigating the SIIDs had been conducted in controlled settings and mainly focused on mobility conditions. In this thesis, we collect data from users in their daily settings. Moreover, we cover different dimensions of context. The findings of this research question could be helpful for people who would like to conduct usability studies or would like to develop intelligent applications to improve users' performance in a specific context.

R4. *"Can we predict users' performance problems caused by SIIDs using the available smartphone sensors?"* – This question compares different regression models to predict user performance problems related to context. It would enable other researchers and us to develop systems that can adapt themselves to users to overcome the adverse effects of SIIDs.

R5. *"What are the possible adaptation techniques based on context and automated performance prediction?"* – This question reviews and discusses possible adaptation techniques to reduce the adverse effects of SIIDs. A system can deploy these adaptations using the available smartphone sensors and automatically predicted performance measures. It would help other researchers and us decide on the adaptation approaches for specific application purposes.

1.3 Contributions and Novelties

This thesis contributes to the literature on several points:

- We conducted a systematic review on the effect of context on user performance. In this review, we reported the different context factors investigated, how different context factors affect user performance, and possible research gaps in the literature (R1). Although different contextual factors have been widely investigated in the literature, the corresponding studies have been conducted in controlled environments, and the effect of context on users' performance in their daily settings had little attention.
- Recent approaches to detecting typing errors using free text have been based on lookup approaches. We combined the approaches of Nicolau et al. [17], Evans

and Wobbrock [18], and Torunoğlu and Eryiğit [20] to cover daily texting language and detect typing errors in both English and Turkish (R2). Our performance evaluation with the participants showed that the combined approach is more successful, especially in detecting typing errors in the daily Turkish language.

- Most of the text entry studies have been conducted in controlled laboratory environments. In this study, we collected text entry and sensor data in the wild. We extended an existing framework to capture the participants' keyboard interactions, a set of sensor data, and context labels submitted by the participants (R3).

- The effect of context on users' text entry performance has been primarily investigated for different mobility conditions in the literature. This study considered the context in five dimensions: environment, mobility, social, multitasking, and distraction. According to our findings, being in an outdoor environment, being mobile, the presence of other people, multitasking, and having distractions increase error rate but have no effect on typing speed. This study provides the first empirical evidence on the effect of context on users' typing performance in an in-situ study (R3).

- Human activity recognition has been widely studied with the advance in the available sensor in smartphones. However, research to date has not yet considered predicting user performance using available smartphone sensors. This study first investigates the sensors used to identify contextual factors. Then, using these sensors and user performance metrics, this thesis shows how smartphone sensors can be used to predict user performance problems (R4).

- Previous studies proposed many adaptation techniques for both disabled users and smartphone users under SIIDs. This study first reviews these adaptation techniques and proposes a model by combining the findings of this study (R5).

1.4 Publications

The contributions of this thesis are published in the following publications:

- Elgin Akpinar, Yeliz Yeşilada, and Selim Temizer. 2019. Ability and Context Based Adaptive System: A Proposal for Machine Learning Approach. In *Proceedings of the CHI'19 Workshop: Addressing the Challenges of Situationally-Induced Impairments and Disabilities in Mobile Interaction*, 8 pages. https://arxiv.org/abs/1904.06118

- Elgin Akpinar, Yeliz Yeşilada, and Selim Temizer. 2020. The Effect of Context on Small Screen and Wearable Device Users' Performance - A Systematic Review. *ACM Computing Surveys* 53, 3, Article 52 (May 2021), 44 pages. https://doi.org/10.1145/3386370

- Elgin Akpinar, Yeliz Yeşilada, and Pınar Karagöz. 2022. Effect of Context on Smartphone Users' Typing Performance in the Wild *ACM Transactions on Computer-Human Interaction*, 43 pages. In submission (minor revision received)

1.5 Online Repository

All the materials and data of this study (instructions and consent form of the user study and individual performance comparisons) are available in our external online repository at https://iam.ncc.metu.edu.tr/cabas/. Moreover, we collected user performance measurements, context labels, and sensor data from available smartphone sensors. We published our dataset in our public repository: https://github.com/melgin/cabas-dataset

1.6 The Outline of the Thesis

The rest of the thesis has been organized as follows:

Chapter 2: Literature review This chapter reviews the user studies that focused on the effect of context on SIIDs. In particular, this review answers two research questions. The first question investigates which contextual factors have been examined in the literature that can cause SIIDs. The other question focuses on how different contextual factors affect small screen and wearable device users' performance. In this chapter, 187 publications were systematically reviewed under a framework with five factors for context analysis: physical, temporal, social, task, and technical contexts.

Chapter 3: Sensing the context and user performance This chapter presents a remote user study in the wild with 48 participants. In this study, we collected smartphone keyboard interactions and context details. We first propose an approach for error detection by combining approaches introduced in the literature. A follow-up study shows that the accuracy of error detection is improved.

Chapter 4: The effect of the context on user performance We investigate the effect of context on typing performance based on five dimensions: environment, mobility, social, multitasking and distraction, and reveal that the context affects participants' error rate significantly but with individual differences.

Chapter 5: Modelling context and user performance This chapter explains all of the steps in the classification pipeline, including sensor data collected during the user study and features used in the literature. We validate the sensor data and context labels with classification methods to identify the context. We compare the performance of each method with different parameters. Then, we applied regression models to smartphone sensors and user performance data. Finally, we compared individual and overall regression results.

Chapter 6: Adaptation and discussion This chapter discusses the findings and implications of this thesis work. We present strategies to overcome SIIDs and adaptations employed to solve performance problems in the literature. Then, we discuss two possible use cases of our findings for adaptation.

Chapter 7: Conclusion and future work Finally, in this chapter, we conclude the thesis, explain the limitations of the study, and discuss our future work.

CHAPTER 2

LITERATURE REVIEW

Small screen and wearable devices play a key role in most of our daily tasks and activities. However, depending on the context, users can easily experience situationally-induced impairments and disabilities (SIIDs). Previous studies have defined SIIDs as a new type of impairment in which an able-bodied user's behaviour is impaired by the context including the characteristics of a device and the environment. This chapter[1] systematically reviews the empirical studies on the effect of context on SIIDs. A significant amount of empirical studies have been conducted focusing on some factors such as mobility but there still are some factors such as social factors that need to be further considered for SIIDs. Finally, some factors have shown to have significant impact on users' performance such as multitasking but not all factors has been empirically demonstrated to have an effect on users' performance.

2.1 Introduction

Small screen and wearable devices play a significant role in our daily lives. It is expected that the number of small screen devices will increase from 1.9 billion to 5.6 billion between 2013 and 2019 [21]. Even though these predictions might not be so accurate, they still show that the number of small screen devices will be quite significant. Some studies also make predictions that with the evolution of Internet of Things (IoT), these numbers can even become higher[2]. These devices are not

[1] Elgin Akpinar, Yeliz Yeşilada, and Selim Temizer. 2020. The Effect of Context on Small Screen and Wearable Device Users' Performance - A Systematic Review. *ACM Computing Surveys* 53, 3, Article 52 (May 2021), 44 pages. `https://doi.org/10.1145/3386370` Reprinted by editor's permission

[2] `https://www.intel.com/content/dam/www/public/us/en/images/iot/guide-to-iot-infographic.png`, Last access: 02.11.2018

used to communicate only anymore, but they are also used to perform most of our daily tasks [4]. Small screen devices include tablets, smartphones, phablets[3] and wearable devices such as smart watches and glasses [22]. In this article, we refer to all these devices as small screen devices. One of the common characteristics of these devices is their small screen size relative to desktop computers. Although this feature increases portability, it can also have significant impact on interaction, in particular usability [23, 24].

Unlike desktop computers which are typically used in a fixed and stable environment (*e.g.* a typical setting would be the user seated with no excessive light or weather conditions, etc.) small screen devices can be used in different environments including indoors, outdoors, noisy, quite, crowded, etc. Furthermore, while using small screen devices, the users might be engaged with different and parallel tasks, for example messaging while walking on a busy street [8]. In the literature, these types of temporary reductions in user performance due to context are referred to as situationally-induced impairments and disabilities (SIIDs) [14]. This phenomenon was defined as "difficulty accessing computers due to the context or situation one is in, as opposed to a physical impairment" [15, 16]. There can be many factors causing SIIDs and the main observation is that both the environment and the current context can cause SIIDs. In this article, we use context to refer to both environment, situation and context which is defined as "any information that characterizes a situation related to the interaction between humans, applications and the surrounding environment" [11].

In the literature, there has been many empirical studies on investigating the effect of different contextual factors on SIIDs with varying findings. In this article, we aim to provide a systematic review of the work that has been done and the impact they showed on the user performance. In particular, we ask the following two research questions: (1) "Which contextual factors have been examined in the literature for small screen or wearable device interaction that can cause SIIDs?" and (2) "How do different contextual factors affect small screen or wearable device users' performances?". Answering these questions would enable us and other researchers to see what has been investigated so far and what the factors that still need to be investigated are. Knowing these would be useful for building smarter applications and would also

[3] devices with capabilities of both tablet and smartphones

be useful for conducting usability studies or user interaction pattern mining under different contextual factors.

There are other systematic reviews in the field of mobile device interaction. First of all, Jumisko-Pyykkö and Vainio [25] reviewed the literature surrounding mobile contexts of use. They explained the characteristics of social, physical, technical, temporal, task and transitions contexts. Although mobility and environmental conditions were briefly mentioned in physical context, this study does not provide a deep understanding of contextual factors and situational conditions which have been examined in the literature for SIIDs. Coursaris and Kim [26] focused on mobile usability evaluation studies in the dimensions of user, task, technology, environment, research methodology, usability dimensions and key findings. Although they covered environment, task and technology dimensions, they only focused on usability studies without focusing specifically on SIIDs. Motti et al. [27] reviewed touchscreen interaction techniques and input devices only with a specific group of older adults. Liu et al. [28] applied keyword analysis on the papers published between 1999 and 2013 on ubiquitous computing field. Rather than contextual factors, they focused on the evaluation of the field. In more recent reviews, Sarsenbayeva et al. [29] provided a brief overview of the factors causing SIIDs as well as approaches to detect and overcome them. In particular, they focused only on ambient temperature, mobility and encumbrance. Finally, Wobbrock [10] discussed the definition of SIID and provided a list of factors that can cause SIIDs. However, only three contextual factors were discussed in detail: walking, cold temperature and divided attention/distraction. Unfortunately, recent reviews on SIIDs have not covered all situational context.

Compared to these, in this article, we present a much broader systematic review of the work that has been conducted to investigate the effect of contextual factors on SIIDs (see Section 2.2). In order to present the contextual factors, we used the context framework proposed by Jumisko-Pyykkö and Vainio [25] as the backbone of our review. This article is organized into two main parts guided by the research questions asked above: first part explains the contextual factors that have been considered to have an effect on SIIDs and in the second part we present the findings of the studies focusing on these contextual factors and their effect on the users' performance. We first present the metrics used in the performance assessment and then we present their findings.

This article systematically reviews 187 publications in Section 2.2 under the context framework that has five factors: physical, temporal, social, task and technical context. This review shows that some factors such as location and mobility as part of physical context have been widely studied but some contextual factors such as interpersonal interaction and culture as part of social context have not been widely studied (see Section 2.3). Furthermore, in order to investigate the effect of context on SIIDs, this review also shows in Section 2.4 that many performance metrics have been used in the literature. Some of these metrics are very typical such as task completion time, but quite a lot of context specific metrics have been introduced; for example, metrics related to navigation or gait and posture related metrics (see Section 2.4.1). Finally, this review also shows that there have been a relatively small set of studies investigating the effect of some contextual factors in user's performance (see Section 2.4.2). In addition, it also shows that for some popular metrics, existing studies have no consensus on some contexts having significant impact on users' performance. Therefore, this article presents the existing work in detail and shows the gaps in the literature where further studies can be conducted (see Section 2.6).

2.2 Research Method

We have conducted our systematic review by following the steps specified by Ghezzi-Kopel [30] and this section provides a summary of those steps.

2.2.1 Research Questions

In our systematic review, we mainly asked the following two research questions:

1. "Which contextual factors have been examined in the literature for small screen or wearable device interaction that can cause SIIDs?" – This question aims to identify and bring together all the contextual factors that have been examined in the literature centred around SIIDs. This would enable us and other researchers to see what kind of factors need to be considered or the factors that have not

been considered at all.

2. "How do different contextual factors affect small screen or wearable device users' performance?" – This question aims to see the overall effect of the contextual factors on the users' performance and how user performance was articulated in the literature. There are some standard performance measures such as task completion time, error rate and workload, but this review will enable us to see all the performance metrics used in the literature. Answering this research question would again enable us and other researchers to see the kinds of effects the contextual factors might have on the users' performance and how they can articulate the users' performance. This could be useful for people who would like to conduct usability studies or people who would like to develop smart applications to improve users' performance in a specific context.

2.2.2 Inclusion and Exclusion Criteria

In our systematic review, we mainly included papers which focus on interaction with small screen devices or wearable devices under different contextual factors. We excluded desktop interaction, or large displays such as wall-mounted displays or tabletop displays. Furthermore, we excluded papers that focus on disabled users, for example blind users. This is mainly because they have specific requirements such as specific assistive technologies used due to their disabilities. We also excluded papers if they are late-breaking results, work-in-progress, posters, student research competitions or adjuncts. Finally, we only reviewed publications in English. We did not limit the publications with time criteria.

2.2.3 Searching for Studies

We mainly searched three sources for relevant research: (1) online libraries and search engines, (2) references and citations of the papers reviewed, (3) publication archives of specific conferences and journals. We started our review with search queries on the following online platforms: ACM Digital Library, METUnique Search, Google Scholar, ScienceDirect and Scopus. We mainly used the following queries with their possible combinations: "mobile, context, walk, situational impairment, cell phone, texting, typing, pointing, touchscreen, siid, eyes free, wearable, smartphone". For instance, we searched for "mobile and walking", "mobile and context", "situational impairment and context", etc. These keywords were chosen because they were commonly referred in the related systematic reviews [25, 27] and they were the most relevant keywords to our research questions explained above. From these online libraries and search engines, we retrieved and reviewed 496 papers, and marked 89 papers as relevant. Then, we reviewed references of these relevant papers as well as papers that cite them. From references and citations, we reviewed 2285 papers and marked 52 papers as relevant. Finally, we manually reviewed all the volumes/issues between 2007 and 2019 of the following key Human-Computer Interaction (HCI) venues[4]:

- Computer Human Interaction (CHI) (5213 papers);
- ACM Conference on Pervasive and Ubiquitous Computing (UbiComp) (1152 papers);
- International Journal of Human-Computer Studies (949 papers);
- ACM Transactions on Computer-Human Interaction (TOCHI) (364 papers);
- Behaviour & Information Technology (932 papers);
- Conference on Designing Interactive Systems (807 papers);
- Mobile HCI (740 papers);
- International Journal of Human-Computer Interaction (827 papers);
- IEEE International Symposium on Mixed and Augmented Reality (525 papers);
- ACM Transactions on Interactive Intelligent Systems (TiiS) (217 papers);
- Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (471 papers);

From these HCI conferences and journals, we reviewed 12,197 papers in total and we marked 46 as relevant. Therefore, from the three main sources ((1) online libraries and

[4] dl.acm.org, www.journals.elsevier.com, www.tandfonline.com, ieeexplore.ieee.org

search engines, (2) references and citations of the papers reviewed, (3) publication archives of specific conferences and journals), we reviewed in total 14,978 paper and marked 187 as relevant.

2.2.4 Data Extraction

In order to systematically review all the contextual factors, we used the framework proposed by Jumisko-Pyykkö and Vainio [25] to identify contextual factors in the literature surrounding situational disabilities and impairments of the small screen or wearable device users. This framework has five main dimensions: physical, temporal, social, task and technical context. The reviewed papers were analyzed based on these dimensions. This allowed us to systematically review all the relevant papers.

2.2.5 Paper Selection

During our review process, we manually checked if the paper was relevant. We started with title and abstract screening to label the papers which are clearly not related to our topic as irrelevant. We also skimmed the full text and searched for specific keywords based on our inclusion and exclusion criteria. We created a citation graph where citation relations were visually represented. Based on this graph, we applied a breadth first search approach to analyze papers or retrieve more candidates.

Among the 14978 publications reviewed, we selected 187 publications as relevant, which focus on different mobility or contextual conditions as well as eyes-free interaction. Figure 2.1 and Figure 2.2 illustrate the tag clouds for the keywords of the papers published in two different periods. While walking and lighting were popular conditions studied on mobile devices and PDAs between 2002 and 2009 (Figure 2.1); the focus has been switched to other contextual factors such as eyes-free interaction, encumbrance, stress, attention and distraction between 2010 and 2019 (Figure 2.2). Similarly, Figure 2.3 illustrates the distribution of the relevant papers published. Although this figure shows that the highest numbers of publications (18 papers) were in 2011 and 2017, it also shows that the topic is still very popular. In fact there were 17 papers published in 2018. Later in this paper, we will see that the focus is still SIIDs

Figure 2.1: Tag cloud for papers between 2002 and 2009.

Figure 2.2: Tag cloud for papers between 2010 and 2019.

but with different devices and interaction techniques.

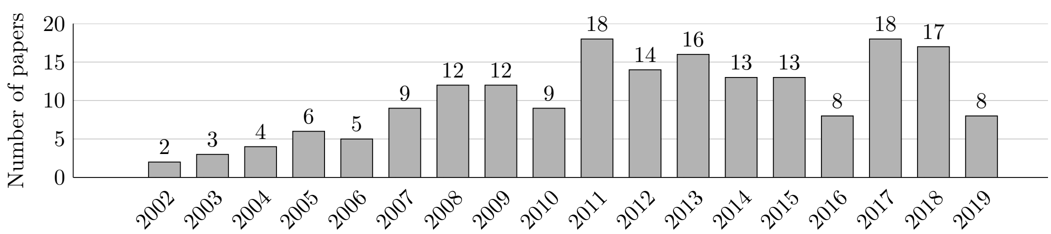


Figure 2.3: Number of publications per year.

In this section, we described our research methodology in conducting a systematic review of the contextual factors causing situational impairments and disabilities studied in the literature. This methodology can easily be applied to update the review presented here with the recently published papers. In the following sections, we present the findings of our systematic review. We organized our findings under the two research questions we asked (see Section 2.2.1).

2.3 Contextual Factors Explored for SIIDs

This section mainly investigates our first research question:

> "Which contextual factors have been examined in the literature for small screen or wearable device interaction that can cause situationally-induced disabilities and impairments?"

In order to present the contextual factors studied in the literature which can cause situationally induced disabilities and impairments in a systematic way, we organized our findings based on the dimensions given by Jumisko-Pyykkö and Vainio [25]. Before we present each context review in detail, we first briefly summarise the participant characteristics in the studies we reviewed.

2.3.1 Participant Characteristics

In 96 papers out of 187 reviewed, participants were recruited from graduate/undergraduate students and university staff. Some authors recruited equal number of male and female participants [31–52]. In general, participants were familiar with small screen devices; however, in some studies, participants were not familiar with wearable devices. In experiments and interviews, 46 users participated on average (minimum: 4, maximum: 3338, standard deviation: 255.39)[5]. In observational studies, number of people observed are relatively higher than experimental studies: 347 [55], 357 [56], 431 [57], 2668 [58] and 4129 [59]. Hiniker et al. [60] observed and inter-

[5] Due to these two studies [53, 54], standard deviation is quite large as [53] has 1669 and [54] has 3338 participants in their studies which are not typical.

viewed adult caregivers. Tigwell et al. [9] applied a questionnaire to mobile content designers. Fitton et al. [61] conducted experiments with teenagers.

2.3.2 Physical Context

Physical context [25] is defined as:

> "The apparent features of situation in which the human-mobile computer interaction takes place, including spatial location, functional place and space, sensed environmental attributes, movements and mobility, and artifacts present."

Based on this definition, we grouped our physical context as location, mobility, artifacts, and sensed environmental attributes. Even though the definition above also refers to functional place, in our review, we did not refer to that, since most experiments have been conducted in a controlled environment with specific interaction tasks where the function of the interaction (entertainment, work, etc.) was ignored. Table 2.1 presents the physical context studied in the literature. As can be seen from this table, location and mobility are more widely studied compared to other factors.

People can use their small screen devices and wearable devices in various locations and in the literature we can see that different **locations** have been studied. The most popular location for the experiments conducted related to SIIDs is the lab environment. One advantage of conducting experiments in the lab environment is that, participants and experimental conditions can be easily isolated from the external factors, such as other people or weather conditions. However, in this case, the experiments might not reflect realistic conditions. In order to address this problem, some studies were conducted outside of the laboratory, especially in public indoor environments. However, they still might not reflect the requirements of using the small screen devices and wearable devices outdoors where some other external factors such as weather conditions need to be considered. Therefore, some studies have also been conducted in outdoor locations in the literature. Our literature review revealed that there are several studies that aimed to compare lab environment to indoor or outdoor public environments.

Table 2.1: Physical contexts investigated in the literature

Context	**Types**	**Papers**
Location	Lab environment	[8, 62–66]
	Indoor environment	[8, 66]
	Outdoor environment	[62, 65]
	Pedestrian street or public area	[8, 63, 64]
Functional place	Home, work, outdoors, public transit, restaurant, other	[67]
Mobility	Sitting	[36, 61, 68–100]
	Standing	[33, 81, 88, 90, 91, 93, 101–118]
	Walking on a route	[33, 41, 61, 68, 69, 74–76, 78–81, 84, 86, 89–95, 98–109, 111–113, 115, 117, 119, 120]
	Walking on a treadmill	[41, 68, 70, 72, 77, 78, 82, 83, 87, 94–96, 99, 110, 114, 116, 118, 119, 121, 122]
	Walking after a researcher	[36, 88]
	Walking on a straight path	[92]
	Walking through a street or public area	[71]
	Public transportation	[97]
	Walking	[73, 85]
Sensed Environmental Attributes	Lighting levels (low and high)	[75, 79, 80, 119]
	Weather (cloudy, partly cloudy, sunny)	[106]
	Vibration	[123]
	Environmental noise	[50, 123]
	Temperature (cold, warm)	[47, 48]

Similar to location, **mobility** aspect has also been widely studied in the literature and the effect of mobility was demonstrated in various ways. One way of simulating mobility in the experiments is to walk on a treadmill or on a mini-stepper. In general, when participants were walking on a treadmill, they were isolated from any other factors if the experimental design did not include artificial distractors. This mobility condition may be effective if the aim of the study is to observe the effect of walking apart from any other factors; however, it does not reflect realistic scenarios. Alternatively, many studies have examined mobility conditions where participants walked on a straight path, walked on a predefined route and walked freely.

When we look at the sensed environmental attributes, there are very few studies that focus on conducting experiments under different **sensed environmental attributes**. Only a few studies observed contextual factors such as lighting levels [75,79,80,119], temperature [47,48], weather [106], vibration and noise levels [123]. Although they are not related to environmental attributes, some authors used various motion sensors in the experiment. These motion sensors consist of acceleration sensors, gyroscope, magnetic field sensors, motion sensors and heart rate monitor.

There are other physical context factors that have been used as experimental conditions rather than the main focus of the studies. Table A.1 in Appendix A.1 also gives a broad overview of the physical contextual factors used in the studies. Several examples for public indoor environments include corridors [36,66,85,102,106,124–126], stairs [127, 128], university cafeteria [129], quiet hallways [89, 108, 130, 131], an empty seminar room [93, 115], public areas [66, 71] and others [43, 86, 132]. Some studies have also been conducted in outdoor locations in the literature: quiet roads and paths [62, 65], pedestrian street [63, 64, 68, 133], both uncrowded and crowded areas in pedestrian zone [134], sports field [44], city center [35, 135], city forest [136] and others [33, 43, 55, 106, 137–140]. To consider the location differences, there are also several studies that were conducted in multiple places, for example including busy street, escalator, quiet street, bus, metro platform, metro car, railway station, cafeteria, laboratory [8], different parts of a building (entrance, lift, corridor, office, meeting room, etc.) [42], train station, shopping center, university bus stop, business area and market street [57]. Besides those, there are also some unique locations studied in the literature including a warm and a cold room at an arctic medical facility [47], vehicle and public transportation [97, 123] and virtual environment [141, 142]. In the studies presented so far, at least one experimenter was present during the experiments. This experimenter was mainly monitoring or guiding people participating in the studies. There are also studies that were conducted in the wild (in-situ studies) [54, 143–146]. In these studies, participants completed the tasks in their daily routine without having an experimenter observing their task completion.

In order to reflect realistic mobility, during the experiments in some studies, there were also physical objects around participants which might have caused collisions or attention switches. We grouped these under **artifacts** used. Some authors intentionally placed obstacles with respect to their experimental design in the laboratory environment, such as physical obstacles or furniture in the lab environment. In some studies, on the other hand, other people or obstacles were present due to the nature of outdoor or public environments. In the studies that were conducted in virtual environments, virtual obstacles and virtual vehicles were used. Researchers placed computer moni-tors in front of participants to simulate pedestrian crossing behaviour while interacting with mobile devices. With these experimental setups, they aimed to create a sense of experiment in real-world and eliminate possible injury risks [141, 142]. Finally, a unicycling clown was used in Hyman et al. [55] to observe inattentional blindness while talking on the phone in a natural environment.

2.3.2.1 Summary and Discussion

One of the most significant discussions in physical context has been on location. Chamberlain and Kalawsky [101] states that, the environmental conditions for each participant would be unique when the experiments were conducted outside; therefore, the environment must be controlled to ensure a uniform set of experiences. Many other researchers have conducted experiments in a controlled lab environment. On the other hand, many researchers have drawn attention to the unrealistic environmental settings in the lab environments [66, 147]. A lab environment might not reflect real world cases due to lack of various lighting levels [61], complex obstacles or disruptions [90, 102, 108, 126, 148–150], environmental noise [74], pedestrians [102] and any factor that requires attentional resources for safety reasons [62]. Similarly, Stavrinos et al. [142] argued that using a real world setting might result differently than using a virtual environment for the experiments. Kane et al. [106] highlighted the challenges of conducting experiments outside; such as weather conditions, unexpected interruptions and distractions, and safety concerns; however, they stated that those experiments conducted outside are more realistic. Suggested open research areas are conducting experiments in dynamically changing paths [150], crossing roads [126] and music concerts to observe the effect of vibration and noise levels [123].

Another debate in physical context has been on mobility. Barnard et al. [119] and Ng et al. [41] stated that, using treadmill to simulate walking was simpler to control and maintain walking speed. On the other hand, Ng et al. [41] highlighted the possible input problems about treadmills with safety bars when conducting experiments with encumbrance. Barnard et al. [119] indicates that, ground walking is more realistic than using a treadmill. Crossan et al. [81] argued that resting, sitting, standing and walking postures in a lab environment do not reflect real world cases such as walking on a busy street. Kane et al. [106] suggested transitions between mobility conditions as a research area, such as starting the experiment with standing and continuing with walking.

2.3.3 Temporal Context

Temporal context [25] is defined as:

> "The user's interaction with the mobile computer in relation to time in multiple ways such as duration, from time of day to years, the situation before and after use, actions in relation to time, and synchronism."

Table 2.2 presents the studies that investigated the effect of temporal context. As can be seen from this table, only the effect of session and walking speed have been investigated in the experiments. The experiment session was used to observe performance changes across different sessions. Moreover, some authors aimed to observe how walking speed affected participants' performance in walking conditions. Compared to these though, synchronism aspect, different time of day to years and other action relation to time aspects have not been so widely studied.

Table 2.2: Temporal contexts investigated in the literature

Temporal Context	**Types**	**Papers**
Duration	Multiple sessions on different days	[37, 90, 153, 155]
	Multiple experiments	[156]
Actions' relation to time	Walking speed	[41, 65, 157, 158]

Majority of temporal context factors have been used as experimental conditions. Table A.2 in Appendix A.2 also summarizes the factors that have been used related to temporal context. The length of the interaction session depends on the task to be completed, number of sessions and experimental settings. **Duration** of the overall experiments ranges from 15 minutes [45] to 90 minutes [47, 134, 151]. The length of a single session also varies from three minutes [69, 89, 152] to 40 minutes [153] or 45 minutes [36]. On the other hand, Jongil et al. [154] limited the length of a single trial to one minute. In some studies, multiple sessions were arranged with the participants. For example, Banovic et al. [37] conducted 16 sessions which took around 40 minutes; whereas Clawson et al. [90] completed 15 sessions where each session took 20 minutes. Conradi et al. [114] arranged four sessions in consecutive days. Unlike others, Reyal et al. [144] conducted experiments for four weeks in the wild, and asked participants to complete corresponding tasks 10 times in a day. The same experimental settings may also take different amounts of time. For instance, indoor sessions took around 30 minutes in [106]; while outdoor sessions took 90 minutes. On the other hand, in [43], indoor sessions took 50 - 60 minutes where outdoor sessions lasted in 20 minutes. Finally, experiments in stationary condition took 15 minutes and walking condition in 90 minutes in [116]. These intervals give ideas about the duration of the overall experiments. Task completion time for a single task, on the other hand, has been used to compare performance of the participants within various conditions or approaches.

For most of the studies, authors did not mention the **time of day, week and year** of the experiments especially if they conducted the experiments in the lab environment. This may be due to the fact that, the same experimental conditions may be repeated regardless of the time. On the other hand, it may be challenging to conduct two experiments with similar parameters in outdoor environments since weather conditions or number of people around the participants change. To address this challenge, Pielot and Boll [35] repeated the experiments on Saturdays in May. Similarly, Kane et al. [106] conducted experiments in the afternoons. On the other hand, Harper et al. [57] conducted their observations at different hours of a day; so that they could see the changes with respect to time. Finally, some authors aimed to prepare experimental setups in which participants also deal with pedestrians on their route. For this purpose, Wenig et al. [43] and Pielot et al. [134] conducted their experiments in summer, when many tourists visited the city. MacKay et al. [71] also set their experiment time to busiest time of the day (11:00 a.m. to 4:00 p.m.).

In terms of **events before or after the experiments**, in [144], participants were sent periodic notifications which asked them to complete text entry tasks. These notifications were sent 10 times in a day. Similarly, Aliannejadi et al. [146] sent search tasks to participants based on a pre-defined schedule. There is no experimental setup in the remaining studies if they were conducted in the wild.

Regarding the **actions' relation to time**, along with multitasking, some studies aimed to put more pressure on the participants during the experiments. These studies aimed to simulate cases such that a user is late for a meeting and has to send a text message to a colleague while he/she is walking to his/her office. In some studies, participants were asked to walk in different walking speeds [41, 65, 78, 157–159]. Similarly, Oulasvirta et al. [8] simulated hurrying, normal and waiting conditions in their experiments. Finally, Conradi [157] changed presentation time in the experiments to analyze minimum time for users to perceive short words on the phone screen.

Regarding the **synchronism**, generally, participants were asked to complete the tasks individually and synchronously in existing studies in the literature. There are several exceptions for this condition. In some experiments, participants talked on the phone with an experimenter [142, 160]; while some experimental setup consisted of both talking on the phone and texting with an experimenter [44, 141, 147]. Harper et al. [57] observed people while they were texting and walking in the wild. Similarly, Hyman et al. [55] checked whether people paid attention to their surroundings while they were walking and talking on the phone. In synchronism context, talking on the phone is a synchronous task; while texting is asynchronous.

2.3.3.1 Summary and Discussion

As can be seen from Table A.2, durations widely studied can be considered short given the amount of time people spend with their small screen and wearable devices these days. Therefore, one criticism about temporal context is that, some interaction techniques may require longer learning curves; however, in experiments participants had limited time to learn and perform the tasks [161]. Therefore, some people argue that such as Arif et al. [150] longitudinal studies should be conducted to give time to participants to be familiar with the input device or interaction technique during the experiments. Furthermore, in real world, people do have actions related to time and also do use their devices in different times of the day and year. Even though there are a few studies focusing on these, there can be still many more studies with varying factors to better understand the effect of the temporal context.

2.3.4 Task Context

Task context is defined as follows [25] :

> "The surrounding tasks in relation to user's task of interacting with mobile computer containing the sub-components of multitasking, interruptions and task domain. Task context is related to the demands of the entire situation upon one's attention."

Table 2.3 presents the studies that investigated the effect of task context. As can be seen from this table, walking is the most widely studied multitasking aspect. Table A.3 in Appendix A.3 provides a summary of the task context used in the literature under three factors: multitasking, interruptions and task domain. According to this table, navigation, reading, text entry and target selection are the most popular task domains studied. However, as presented in this table, there are many other domains which are not widely studied and of course there may be many others that can be studied in the future.

Table 2.3: Task contexts investigated in the literature

Context	**Types**	**Papers**
Multitasking	Walking	[33,36,41,61,68–96,98,101–117,119,121, 122]
	Encumbrance	[41, 111, 113, 115, 118, 151, 159]
	Collision/hazard avoidance	[162]
	Distraction or cognitive tasks	[52, 148]
	Presence of dual task while walking	[40,51,66,120,135,147,149,152,154,158, 162–165]
	Presence of dual task while crossing street	[53, 141, 142]
Interruptions	Eyes-free interaction	[166]
	Stressor tasks	[52]
	Distraction tasks	[121, 153]
	Visual disruptions	[167]
	Incoming phone calls	[39]

Multitasking has been considered as an effective way of fragmenting attentional or cognitive resources during small screen and wearable device interaction. This is considered to be one of the major causes of SIIDs. In experimental studies, one of the most common techniques used in the literature is to ask participants to walk while completing a set of predefined tasks. Mobility conditions that have been used to achieve this are presented in Section 2.3.2. In different mobility conditions, a user needs to maintain her walking speed, check for route in order not to get lost and watch for obstacles, vehicles and other pedestrians to avoid collisions. As a result both mental and physical workload increase in such conditions. Another multitasking condition is encumbrance. In the literature, participants were asked to hold an object (box, bag, etc.) during some experiments [41, 111, 113, 151, 159]. Alternatively, Oulasvirta and Bergstrom-Lehtovirta [168] simulated several conditions such as use of non-prefered hand and occupation of whole or some parts of hand by asking participants to hold objects with different sizes (such as box, basketball, coffee mug, tongs or scissors) while they were entering text on mobile devices. Moreover, users' attention to the possible risks while interacting with a small screen or wearable device was also examined. In these studies, different cases were simulated including artificial hazard notifications [76, 169], injury risk in virtual environment [142] and collision cases with obstacles and other pedestrians [129]. In such experiments, participant safety is critical; as a result, such experiments were conducted in virtual or controlled environments. Similarly, Conradi et al. [114] and Jongil et al. [154] asked participants to watch for environmental changes while they were interacting with small screen or wearable devices. Kjeldskov and Stage [68] asked participants to play Jungle Book Groove Party game. Although talking on the phone has been used as the main task in most studies, in [160] participants were asked to complete several tasks such as numeric text entry or calendar checking while they were talking on the phone and playing a driving game. This experimental condition illustrates the cases in which users interact with a cell phone in an eyes-free fashion. Finally, cognitive tasks such as note recalling [82], attention saturating tasks [148] and mathematical calculations [40] have also been used in the literature.

Temporary **interruptions** that break users' attention have been covered with different methods in the literature. One of these methods is to place obstacles in the participants' walking paths. Similarly, conducting the experiments in a public area in which other people were presented causes interruptions in the interaction with small screen and wearable device. In both cases, participants need to divide their attention between completing current task and avoiding collisions. The studies designed to include obstacles or other people in their experimental settings are provided in Section 2.3.2. In [76] and [169], participants were asked to check for artificial hazard notifications. Unlike a normal walking path, going up or down the stairs may also interrupt participants [127, 128]. Using virtual objects such as vehicles is another technique to interrupt participants [141, 142]. Jain and Balakrishnan [153] used visual distractions in forms of changing numbers, while Yang et al. [170] placed stop signs on the route to interrupt participants' attention on the tasks. We considered eyes-free interaction in which participants complete a set of tasks without looking at the device screen as an interruption condition. Such case of interaction was also covered in the literature [31, 37, 45, 46, 88, 136, 145, 156, 160, 166, 171–178].

A considerable amount of the studies in the literature aim to compare the performance of the user under various **task** domains. As a result, majority of the tasks are highly goal oriented. In these tasks, participants were given instructions and asked to complete the tasks by considering several performance metrics such as task completion time or accuracy. One popular task domain is text entry. In this domain, participants were presented a set of phrases and asked to type it as it is. As a result, in text entry domain participants did not type free text or have conversation with another person. Another popular task domain is target selection. The participants were shown several targets and asked to select them by using different techniques or under different conditions. Reading is another popular task domain. Along with reading, scrolling and searching tasks were also used. In recent years, gesture based interaction and navigation have been popular task domains. Some task domains consist of multiple realistic tasks such as numeric text entry, recording phone number, checking calendar [160] and reading messages, replying with message templates, answering calls, sharing fitness information online [44]. Other goal oriented task domains are tapping on buttons [62], visual acuity [65], visual search [112], web search [8, 94, 95, 140], zooming [125], RSS reading [145], cognitive tasks [89], crossmodal icon identification [77], dealing with incoming notifications [135], drag and drop tasks [130], remembering symbols [179], responding to alerts [105], sliding [45], speech based text entry [72, 169], sports tracking [137], menu navigation [177] and menu selection [132, 133]. Although the original framework categorizes mobile-gaming as an action oriented task domain, we considered game playing tasks in [129] and [143] as goal oriented since main purposes of both studies were to compare efficiency and effectiveness of different approaches. As a result, performance and preference had higher priority than entertainment in these studies. The studies reviewed so far in this section took specified tasks as the main task domain and compared participants' performance under different factors. There are also studies that investigated the effect of small screen device usage on walking or posture. In these studies, the following task domains have been used: talking on the phone [141, 142, 147, 149], texting [40, 141, 149, 152, 164], listening to music [141], reading [147, 165] and text entry [66, 147, 154, 165]. It is important to note that, in text entry task domain, participants retyped given phrases; while in texting task domain they had conversation with experimenters by sending or receiving text messages. We consider these tasks as goal oriented; since performance related metrics such as task completion time and accuracy were still important during the experiments.

2.3.4.1 Summary and Discussion

When we look at the literature, in the majority of the studies, participants have interacted with experimental applications which were developed to simulate a particular interaction method. However, these are much more simplistic than real world applications and tasks [70]. Realistic tasks that require continuous attention [8] or are cognitively demanding [71] may reflect real world use cases. In text entry tasks, participants have been asked to type predefined sets of phrases. In the majority of the studies, the text phrases have been in English. The effect of other languages may be examined in text entry task domain. Moreover, instead of typing standard text, Vadas et al. [74] argued that user performance may be affected if they type content which they are interested in, such as personal emails. Lucero and Vetek [135], on the other hand, stated that, using participants' online accounts may cause privacy concerns and participants may have different experiences due to various content lengths. Finally, people normally use their devices with autocorrect/autofill functions enabled. Plummer et al. [66] stated that, the experimental setups that disable these functions may not reflect real world usage.

2.3.5 Social Context

[25] defines social context as follows:

> "The other persons present, their characteristics and roles, the interpersonal interactions and the surrounding culture that influence the user's interaction with a mobile computer."

Based on this definition, we group the contextual factors under this dimension under the following categories: persons present shows if there is another person during the experiment, interpersonal interaction shows if there is an interpersonal interaction during the experiment and culture shows if specific culture elements have been considered or not during the experiment. Table 2.4 presents the social factors considered in the literature for causing SIIDs. Table A.4 in Appendix A.3 also provides a summary of the social factors used in the literature. As can be seen from this table, most studies were conducted with individuals, and interpersonal communication aspects

Table 2.4: Social contexts investigated in the literature

Social Context	Types	Papers
Persons present	Self	[8, 55, 57, 67]
	Accompanied	[55, 57]
	Other pedestrians	[57]
Culture	Users from UK and India	[143]

were mainly considered one to one.

Our review shows that experiments were typically conducted individually in separate sessions. Only Hoggan et al. [123] conducted an experiment in which all participants completed tasks together in the same environment. The reason for this was to ensure that participants interacted with device under the same vibration and noise levels. However, participants did not interact with each other. In two observational studies, some observed people were in pairs or groups, while some of them were individuals [55, 57]. In order to simulate conditions that require interpersonal interaction such as talking on the phone or texting with someone, the participants interacted with an experimenter [44, 141, 142, 147, 152, 160]. On the other hand, Harper et al. [57] suggested that, presence of the experimenter with the participants may affect their behavior. Hyman et al. [55] used a unicycling clown in the experiments and observed whether people could recognize him while they were talking on the phone and walking. Finally, although there were no interaction, other people and pedestrians were also present in the environmental settings of some studies conducted in public areas [35, 57, 66, 106, 129, 134, 135].

Regarding the **interpersonal interaction**, in the majority of studies in the literature, participants were given instructions to complete a set of tasks individually. This type of interaction can be categorized in the framework as one-to-myself; since participants only interact with device. On the other hand, experimenters interacted with participants in several studies that include tasks on talking on the phone or texting. In these one-to-one interactions, conversation topic was predefined [44, 141, 142, 147, 152, 160]. We have not encountered any work in the literature that examined one-to-many or many-to-many interactions.

Culture has not been one of the major social factors considered in studies that investigated the effect of SIIDs. Only Williamson et al. [143] conducted experiments with two specific user groups from the UK and India. Moreover, other attitudes of culture, such as work and organizational culture have been ignored in the research field.

2.3.5.1 Summary and Discussion

Research around the social context has been mostly restricted to single participants interacting with a small screen or wearable device, and only a few studies consisted of interpersonal interaction or other people around the interaction. In our literature review, we identified three open research areas related with social context. First, some task domains require interpersonal interaction due to their asynchronous nature such as talking on the phone or texting. Stavrinos et al. [142] argued that, they could have different results if participants had interacted with someone who they are familiar with instead of researchers during the experiments. Chen et al. [53] also suggested that, participants may ignore messages or phone calls from strangers; as a result, they asked participants to bring a friend to the experiments. They could conduct experiments that reflect real world interactions if they had not restricted the conversation topic. Moreover, the studies that aim to observe user interaction in public areas or in the wild may ask participants to take videos during interaction. However, McMillan et al. [67] stated that, ethical concerns may arise from this since permission from those people around the experimenter is not taken. Although they asked participants to turn off video recording in inappropriate situations; a more effective solution may be proposed to prevent such cases from affecting experiment. The final open research area related with the social context is social acceptance [68, 109, 143], especially for gesture-based interaction [81] and kick-based interaction [180, 181]. Any unusual device or interaction technique during experiments may engage other people's attention, and as a result, participants' performances may be affected [135].

2.3.6 Technical Context

Technical context is defined as [25]:

Table 2.5: Technical contexts investigated in the literature

Context	Types	Papers
Device	Smartphone with touchscreen	[35, 49, 94, 95, 179]
	Tablet	[94, 95]
	Wearable device	[32, 101, 182]
	Smartwatch	[179]
	Smartphone with physical keyboard	[76]
	Smartphone with touchscreen and physical keyboard	[168]
	UMPC	[74]
	Twiddler, trackball, gyroscopic mouse and touchpad	[130]
	Head mounted display, e-book reader	[74]

> "Relation of other relevant systems and services including devices, applications and networks, their interoperability, informational artifacts or access, and mixed reality to the user's interaction with the mobile computer."

Table 2.5 presents the studies that are conducted addressing SIIDs related to this context. Table A.5 in Appendix A.5 also provides a broad overview of the technical factors used in the studies that addressed SIIDs. Based on the definition given by Jumisko-Pyykkö and Vainio [25], we group the technical factors in four categories: devices used, information artifacts, interoperability, and mixed reality systems. As can be seen from this table, based on the maturity of the devices, there are more studies around them. For example, there are quite a lot of studies centred around smartphones with touchscreen but not many studies around wristband/smart bracelet. Other factors compared to devices are less explored in the field. A variety of **device** types have been examined in the literature including Personal Digital Assistants (PDA), Ultra-Mobile PCs (UMPC), wearable devices, smartphones, smartwatches, media devices, tablets and head mounted displays. These device types have different characteristics and modalities. For example, some smartphones have either a touchscreen or a physical keyboard whereas some of them have both. Figure 2.4 shows the devices used in the literature with respect to years. Although PDAs had been used widely in early studies; smartphones, smartwatches and tablets have been very popular recently.

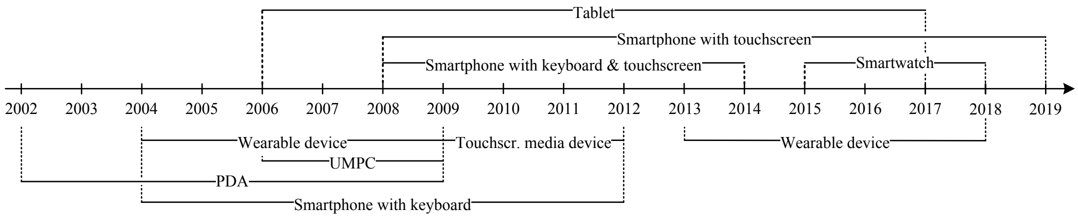


Figure 2.4: Distribution over the years of the surveyed papers according to the device type used

Interoperability between devices has been used for two main purposes. One of the reasons is to provide participants a user interface during the interaction. For example, when participants interact with a wearable device which does not have a user interface, another connected device for representing tasks and feedback is required. For this purpose, wearable devices have been connected with PDA [132], laptop computer [87], desktop computer [91] and Google Glass [93]. In some studies, on the other hand, interoperability helped to represent tasks and feedback if experimental design required eyes-free interaction. Such kind of interoperability was achieved with different devices, including PDA/Twiddler and desktop computer [31], smartphone and laptop computer [88, 171, 172], smartphone and headphone [173], wearable device/smartphone and external monitor [166]. All interoperability cases in the literature are between unequal resources. As an example for interoperability between applications, Zhang and Rau [44] asked participants to share their fitness information on a social media platform. Accessibility on different platforms has not been studied in the context of SIIDs.

We haven't encountered any work in the literature which used multiple devices accessing to the same content. However, in some studies, some **informational artifacts** have been used to represent tasks, including projection screen [69, 72], computer display [91, 168], external monitor [37, 166], LCD display [89], wall display [40, 91] and tablet [45]. In one study by Bragdon et al. [148], a large monitor was used for distraction tasks.

In order to simulate the experiment scenarios in which participants had injury risks, **virtual reality systems** have been used [141, 142]. Crease et al. [76] and Lumsden and Drost [169], on the other hand, used projection screens to indicate hazards.

Conradi et al. [114] also used virtual reality to examine participants' attention to distractors in the virtual environment. Similarly, Jongil et al. [154] aimed to simulate environmental changes by using a projection floor and a monitor.

2.3.6.1 Summary and Discussion

Our literature review revealed that, not only different types of devices, but also different models of these devices have been used in the experiments. As manufacturers launched new small screen and wearable devices, researchers continued to experiment with these devices. Technological advances also enabled researches to simulate various use case scenarios by using different informational artifacts, interoperability between devices and virtual systems. As a research challenge, Wiliamson et al. [145] highlighted the importance of charging and correct placement of the wearable devices for in-situ studies. Moreover, they stated that, some participants may require customization on the devices and there is a trade-off between participant frustration and equally prepared experimental conditions. They suggested that, an in-situ study must be flexible to handle such kind of customization while the study continues. As an open research area, Zhang and Rau [44] suggested to study attractiveness of wearable devices for different genders.

2.4 The Effect of Contextual Factors on Users' Performance

This section investigates our second research question:

> "How do different contextual factors affect small screen or wearable device users' performance?"

In the literature, different metrics have been used to assess the users' performance. In order to explain the impact of the context, in this section we first present the metrics used in the literature and then we explain how these metrics are used to show the impact of context on users' performance.

2.4.1 Performance Metrics

Majority of the studies reviewed in this document aim to observe the main effect of an interaction technique or contextual factor on how users interact with the system or how they behave. For this purpose, they used one or more evaluation metrics for measuring users' performance. Table 2.6 provides a summary of the performance metrics used. As can be seen from this table, performance metrics in terms of task completion time and error rate, and perceived workload in terms of NASA TLX [183] have been widely used. For text entry tasks, another popular performance metric is text entry speed in terms of words per minute [31, 36, 37, 63, 64, 86, 90, 92, 123, 144, 150, 153, 171, 175], characters per second [102, 155], characters per minute [88] or number of codes entered [62].

Besides those metrics, some other metrics are also used that focus on the effects of contextual factors on user behavior. These can be summarised as follows:

Gait and posture related metrics: spatio-temporal parameters, ankle and knee kinematics [152], minimum margin of stability [40] velocity, lateral deviation, linear distance [149], stride length [158, 163, 165], stride frequency, abs path lateral direction, delta right foot position [165], maximum displacement, interference duration, maximum deviation from normal movement during right steps [179], synchronization of gait [203], toe clearance, toe velocity, step length, foot position, foot contact [147], dynamic walking stability [154], walking deviations [106], average walking speed, total walking distance, total steps, strides per minute [131], gait speed [51, 120, 158, 163], stride time [163], error steps [89], head flexion angle [209].

Attention related metrics: accurate change detection, response time, detection rate [167], attentional resources, attention switched, switch back duration [8], number of attention switches [57], noticing unicycling clown [55], field-of-regard loss [154], number of glances [106], number of slow-downs and head-ups [129], collisions [150], wrong turns [150], secondary task performance, duration and subjective distraction [136], number of stop-signs missed and step-outs, unnecessary stop-time [170], situational awareness [120].

Table 2.6: Evaluation metrics

Metric	Papers
Task completion time	[32, 33, 35, 39, 41–43, 45–47, 50, 52, 61, 68, 70, 71, 73–76, 78–81, 83–85, 87, 89, 91, 93, 97, 98, 101, 103, 104, 106–109, 111–119, 124–126, 129–134, 136–139, 148, 151, 156, 159–161, 166, 170, 172, 176–178, 182, 184–194]
Error rate / accuracy	[31, 33, 35–37, 41, 43, 46–48, 50, 52, 61, 63, 64, 66, 68, 69, 72–82, 82–90, 92, 93, 96–99, 101–116, 118–120, 122, 124–130, 132–134, 136, 137, 144, 148, 150, 151, 153–156, 159–162, 169–176, 178, 181, 185, 187–190, 192, 193, 195–197]
Workload	[33–35, 37, 39, 43, 44, 62, 68, 69, 72, 74–76, 78–80, 82, 84, 94, 95, 97, 98, 101, 103, 112, 119, 121, 125–128, 131, 132, 134, 155, 161, 169, 170, 177, 179, 184, 189, 192, 194]
Walking speed	[32, 66, 89, 98, 102, 124, 126, 132–134, 150, 161, 164, 165, 165, 169, 194]
Text entry speed	[31, 36, 37, 52, 62–64, 66, 82, 86, 88, 90, 92, 102, 120, 123, 127, 128, 144, 150, 153, 155, 171, 175]
Subjective evaluation	[32, 33, 42, 44, 45, 71, 85, 96, 107, 117, 125, 131, 131, 140, 150, 160, 166, 167, 184, 186, 188, 192, 195, 198–201]
Interaction related metrics	[45, 67, 79, 79, 80, 80, 96, 97, 105, 123, 131, 133, 134, 140, 145, 147, 170, 171, 175–177, 193, 197, 202]
Gait & posture rel. metrics	[40, 51, 89, 106, 120, 131, 147, 147, 149, 152, 154, 158, 163, 165, 179, 203]
Navigation metrics	[32, 34, 38, 54, 134, 136, 138, 139, 190, 200, 201, 204]
Usability metrics	[43, 44, 61, 68, 71, 133, 136, 174, 182, 194]
Attention related metrics	[8, 55, 57, 106, 129, 136, 150, 154, 167, 170]
Gesture related metrics	[143, 186, 194, 196, 205]
Pedestrian variables	[53, 55, 141, 142]

Others: reaction/response time [122, 154], point of subjective equality [65, 157], just noticeable difference [65], manual Multitasking Index [168], multimodal flexibility index [173], search related metrics [94, 95], cost of error correction [150], system performance [138, 139], heart rate classification accuracy [206], finger and battery temperature [49], reading speed [121], valence, energy, stress [197, 207], user interviews [57, 60], social reactions [135, 145], gaze behaviour [208]

Pedestrian variables: looks away, looks left and right [141], missed opportunities, attention to traffic, hits/close calls [142], time left to spare [141, 142], head turning frequency, unsafe crossing behavior [53], crossing time [53, 55].

Navigation metrics: number of references to route directions, number of self-position references [204], traveled distance, time looking at the map, disorientation events [32, 136], navigation performance [54], arrive time, position accuracy, orientation accuracy [200], distance [191], walking time [201], step and time differences [34], number and length of sessions [38], orientation loss, number of targets found, orientation phases [134], stationary time, time and distance for incorrect route, portion of time in which participants paid attention to the display [138, 139], number of steps [190].

Interaction related metrics: number of swipes on menu [96], total time fingers were in contact with touchscreen [45], time looking at the phone [202], amount of interaction, time device was held in the hand, scanning, time the screen was turned off [54], number and duration of fixations on the phone [147], interaction type (touch vs no touch), interaction time, application, length of interaction [67], interactions per minute [145], list navigation time [193], tap features [197], amount of interaction [134], number of operations [140], field change error, incorrect field error, drawing time [131], scrolls [79, 80], number of commands [176], number of pages viewed [170], average swiping time [177], answer time, access time, pocket time [105], reading time [79, 80], reading speed and text comprehension [100], number of breakdowns when participants take device out of their pockets [133], keystrokes per character [97, 123, 171, 175].

Gesture related metrics: number of incorrect nods, accuracy of gestures, number of gestures aborted [194], gesture mode and technique [143], character recognition rate [205], distribution of pseudo impulse [196], articulation time, gesture size, aperture, corner shape distance, angular difference, shape distance [186].

Usability metrics: ability to answer questions [182], usability problems identified [68], ease of use [44, 71], usefulness [44], efficiency [61], System Usability Scale [43], subjective usability [133], pleasantness [174], difficulty [174], comfort [194], confidence [136].

This section has reviewed a comprehensive set of evaluation metrics surrounding SIIDs. The following section will focus on how contextual factors given in Section 2.3 affect these evaluation metrics.

2.4.2 The Effect of Contextual Factors on Users' Performance

In order to present the effects in a systematic way we again use the five contextual factors as the backbone for our discussion. In Sections 2.4.2.1–2.4.2.5, we presented quantitative results of the papers which give the significance of the results explicitly (either significant or insignificant) or provide a p-value for the statistical tests presented. Then, in Section 2.4.2.6, we outlined qualitative findings.

2.4.2.1 Physical Context

Table 2.7 and Table 2.8 present the effect of physical factors on the performance metrics discussed above. Environment has been used to compare indoor and outdoor environments or public and private places in terms of user performance and attentional metrics. While Brewster [62] showed main effect of indoor and outdoor environment on text entry performance in terms of speed and error rate, a more recent study by Plummer et al. [66] did not reveal significant difference in terms of texting speed and accuracy between private and public environments. The effect of different mobility conditions on user performance has been widely investigated for different task models. Although the literature agrees on influence of the mobility on perceived workload; there are contradicting results on task completion time and error rate. In traveling condition, on the other hand, Hoggan et al. [97] compared sitting conditions in laboratory environment and subway train; and showed that mobility condition affected task completion time; however, it had no main effect on either error rate or workload.

When it comes to sensed environmental variables, there are relatively few studies in the literature which investigate their effects on user performance. According to Kane

Table 2.7: The effect of mobility (SD: Significant diff., ID: Insignificant diff.)

Physical Context	**Task Domain**	**Evaluation Metric**	**SD**	**ID**
Mobility (stationary, walking on a treadmill	Target sel.	Task completion time	[83, 114, 116]	[70, 87]
		Error rate / accuracy	[83, 87, 110, 114, 116]	
	Other tasks	Task completion time	[118]	
		Error rate / accuracy	[72, 82, 96, 118]	[77, 122]
		Workload	[82, 121]	
Mobility (stationary, walking on a treadmill, walking on a route)	Target sel.	Task time & err. rate	[78]	
	Text entry	Task time & err. rate		[68]
		Workload	[68]	
	Read./search.	Workload	[94, 95]	
	Searching	Hits per query/errors	[99]	
Mobility (sitting in lab, sitting in subway train)	Text entry	Task completion time	[97]	
		Error rate, workload		[97]
Mobility (stationary, walking on a route)	Target sel.	Task completion time	[71, 81, 84, 91, 93, 101, 103, 109, 111, 113, 115]	[104, 106, 107] [73, 108]
		Error rate / accuracy	[73, 84, 93, 103, 115] [81, 104, 107, 109, 111]	[101, 106, 108]
		Workload	[84, 101, 103]	
	Text entry	Task completion time	[61, 76]	
		Error rate / accuracy	[36, 61, 86, 120] [76, 102]	[88, 90, 92]
		Text entry speed	[36, 90, 92, 120]	[86, 88, 102]
		Workload	[69, 76]	
	Reading	Task completion time	[33, 75, 117]	[74, 79, 80, 98]
		Error rate / accuracy	[74, 75, 79, 80]	[33, 98]
		Workload	[33, 74, 75, 79, 80]	
		Reading speed	[100]	
	Other tasks	Task completion time	[33, 75, 89, 112]	
		Error rate / accuracy	[33, 75, 85, 89]	[105, 112]
		Workload	[33, 75, 112]	

et al. [106], weather condition has no effect on target selection time. Sarsenbayeva et al. [50] clearly indicated that different types of ambient noise influenced target selection, visual search and text entry time. Similarly, Sarsenbayeva et al. [47] and Goncalves et al. [48] showed the effect of environmental temperature and finger tem-

Table 2.8: The effect of environment and sensed environmental attributes (SD: Significant difference, ID: Insignificant difference)

Physical Context	**Task Domain**	**Evaluation Metric**	**SD**	**ID**
Environment	Visual acuity	Psychophysical metrics		[65]
(indoor, outdoor)	Tap. on buttons	Workload	[62]	
	Text entry	Error rate, text ent. speed	[63, 64]	
Environment (lab.,in-	Text entry	Gait speed	[66]	
door real-world)		Texting speed & accuracy		[66]
Environment (indoor,	Web search	Attention metrics	[8]	
outdoor, transport.)		Switch-back duration		[8]
Functional place	Smartwatch use	Length of interaction		[67]
Ambient Noise (mu-	Target acq.	Task comp. time, err. rate	[50]	
sic: fast, slow, silence)	Visual search	Task completion time	[50]	
Ambient Noise	Target selection	Task completion time	[50]	
(urban noise: indoor,	Visual search	Task comp. time, err. rate	[50]	
outdoor)	Text entry	Task completion time	[50]	
		Error rate		[50]
Ambient Noise (speech:	Target sel., vis.	Task completion time	[50]	
meaningful/meaningless)	search, text ent.	Error rate		[50]
Weather (cloudy, sunny)	Target selection	Task completion time		[106]
Vibration level /	Text entry	Text entry speed	[123]	
Noise level		Keystrokes per char.	[123]	
Temperature	Target selection	Task completion time	[47]	
(cold, warm)		Error rate		[47]
Thumb finger temp.	Target selection	Movement time, err. rate	[48]	
Index finger temp.	Target selection	Movement time	[48]	
		Error rate		[48]
Lighting level	Reading	Task time, workload	[75, 79, 80, 119]	
(low, high)		Error rate		[75, 79, 119] [80]
	Searching	Task completion time	[75, 119]	
	words	Error rate		[75, 119]
		Workload	[119]	[75]

Table 2.9: The effect of temporal context (SD: Significant difference, ID: Insignificant difference)

Temp. Cont.	Task Domain	Evaluation Metric	SD	ID
Task length	Speech based text entry	Error rate	[72]	
Session	Text entry	Text entry speed	[37, 90, 153, 155]	
		Error rate / Accuracy	[90, 153, 155]	[37]
		Workload	[37]	
	Sight-free text entry	Text ent. speed, accuracy, workload	[37]	
	Target selection	Task comp. time & error rate	[156]	
Walking speed	Reading/Visual acuity	Point of subjective equality	[157]	[65]
		Just noticeable difference	[65]	
	Texting, talking on phone	Gait parameters	[158]	
	Target selection	Error rate	[41, 110]	
		Task comp. time & accuracy	[41]	

perature on target selection time. The studies comparing low and high lighting levels for reading and searching tasks have indicated that there is a main effect of lighting level on task completion time and workload but not on error rate.

2.4.2.2 Temporal Context

Table 2.9 presents the effect of temporal factors on the performance metrics discussed before. There is a relatively small body of literature that investigated the effect of session or temporal tensions on user performance. Session was used to observe the changes in user performance through overall experimental process if it contains multiple sessions. So far, a number of studies have reported that error rate, task completion time and perceived workload decrease as the experiment progresses [37, 90, 153, 155, 156]. Different levels of walking speed have also been used to simulate temporal tensions like hurrying. Several lines of evidence suggest that users make more errors when they walk at a higher speed [41, 110]. Taken together, these studies support the main effect of session and walking speed on user performance. Unfortunately, we encountered no studies that compared different times of day, week and year or synchronism conditions.

2.4.2.3 Task Context

Table 2.10 summarizes the effect of task context on the performance metrics. The effect of walking as a multitask was given in Section 2.4.2.1. Böhmer et al. [39] reported that task completion time increased in case of phone call interruptions during question answering tasks. Similarly, Mariakakis et al. [121] suggested that perceived workload increased with the presence of distraction tasks. Jain and Balakrishnan [153] conducted experiments with three levels of distraction including no distraction, low distraction and high distraction. They reported that there was a significant difference in terms of text entry speed between no distraction, low distraction and high distraction. Moreover, highest text entry speed was observed in high distraction condition. They commented that, text entry speed increased due to the attention demand of high distraction condition. Sarsenbayeva et al. [52] showed that, stress decreased target selection time and increased touch offset size; however, it did not have effect on text entry in terms of texting speed and error rate.

Table 2.10: The effect of task context (SD: Significant diff., ID: Insignificant diff.)

Task Context	Task Domain	Evaluation Metric	SD	ID
Multitasking (encumbrance)	Target selection	Accuracy	[41, 111, 113] [159]	
		Speed	[113]	
		Task completion time	[41, 159]	[111]
	Gestures based int.†	Task comp. time, error rate	[118, 151] [115]	[151] [115]
Presence of interruption	Question answering	Task completion time	[39]	
Presence of vis. disruption	Target detection	Accurate change detection	[167]	
Presence of distraction	Reading	Workload	[121]	
	Text entry	Text entry speed	[153]	
		Accuracy		[153]
Pres. of motor act. & dist.	Gesture based int.	Task comp. time & accuracy	[148]	
Object negotiation	Texting	Game score	[162]	

Continued on next page

Table 2.10 – *Continued from previous page*

Task Context	**Task Domain**	**Evaluation Metric**	**SD**	**ID**
Sightedness (sight./blind.)	Target searching	Task completion time	[166]	
Presence of dual task while walking	Texting	Spatio-temporal parameters	[152]	
		Ankle and knee kinematics		[152]
		Gait parameters	[40, 164]	
			[51, 66, 162]	
			[120, 149]	
	Talking on phone	Velocity	[149]	
		Lateral dev. & linear dist.		[149]
	Texting, visual task	Resp. rate, field of reg. loss	[154]	
		Dyn. walk. stab., resp. time		[154]
	Texting, talking on phone	Time & fixations	[147]	
		Gait parameters	[147, 158]	
	Texting & reading	Walking speed	[165]	
	Dealing with notif.	Time	[135]	
	Other	Gait parameters	[163]	
Presence of dual task while crossing street	Texting	Time, head turning freq.	[53]	
	Talking on phone	Attention parameters	[142]	
	Texting, talking on phone, listening to music	Looks away	[141]	
		Time to spare, looks left-right		[141]
Presence of stress task	Target selection	Task comp. time & accuracy	[52]	
	Visual search	Time to memorise item	[52]	
		Time to find item		[52]
	Text entry	Texting speed, error rate		[52]
Task type (working memory tasks, vigilance task)	Cognitive tasks	Task comp. time, walking speed, error steps	[89]	

† [151] found significant differences in terms of error rate for all gestures. They also found significant difference in terms of task completion times for tapping and dragging gestures but not for spreading, pinching or rotating gestures. [115] found significant differences in terms of task completion time and error rate for tapping and wrist flicking gestures but not for swiping gesture.

According to Gustafson et al. [166], task completion time increased when participants were blindfolded. Unfortunately, there is not a consensus among scientists on the effect of encumbrance on user performance. Finally, it is now well established

Table 2.11: The effect of social context (SD: Significant difference, ID: Insignificant difference)

Social Context	Task Domain	Evaluation Metric	SD	ID
Persons present	Text entry	Number of attention switches		[57]
	Talking on the phone	Time to cross	[55]	
	Smartwatch usage	Interaction type & length	[67]	
	Web search	Duration of continuous attention to the phone	[8]	
Culture	Playing a game	Gesture mode		[143]

from a variety of studies that interacting with a small device or wearable device has a main effect on gait and attention metrics. Overall, these studies highlight that such dual tasks reduce walking speed [55, 66, 147, 149, 163–165] and cause divided attention [149] or inattentional blindness [55]. As a result, they introduce pedestrian safety problems [53, 141, 142, 149].

2.4.2.4 Social Context

Table 2.11 presents the effect of social factors on the performance metrics. Similar to temporal context, there are relatively few studies in the literature which investigate the effect of social context on attentional metrics and interaction techniques. Overall, there seems to be some evidence to indicate that presence of others has a main effect on how people interact with smartwatches. On the other hand, one study by Harper et al. [57] indicated that other people do not affect the number of attention switches; while Oulasvirta et al. [8] showed that users attended to the environment significantly more when the environment was crowded. Culture was only used to compare gesture mode preferences of users from UK and India; according to Williamson et al. [143], it does not have a main effect on user interaction.

2.4.2.5 Technical Context

In previous studies which included factors related to technical context, either the effect of using various device types or use of a wearable device has been compared on

Table 2.12: The effect of technical context (SD: Significant difference, ID: Insignificant difference)

Technical Context	**Task Domain**	**Evaluation Metric**	**SD**	**ID**
Wearable, smartphone	Cognitive task	Workload & gait parameters	[179]	
Head-mounted display, e-book reader, UMPC	Reading	Task comp. time & workload	[74]	
		Error rate		[74]
Trackball, mouse, touchpad	Drag and drop	Task comp. time & error rate	[130]	
Different smartphone types	Target selection	Battery temperature	[49]	
	Text entry	Manual Multitasking Index	[168]	
		Task comp. time & workload	[76]	
		Accuracy		[76]
Wearable device	Information retrieval	Task completion time	[182]	
		Ability to answer questions	[182]	
	Target selection	Task time, err. rate, workload	[101]	
	Navigation	Task comp. time & subj. eval.		[32]
		Walk. speed & navigation param.	[32]	
Personal navigation device, wearable device, smartphone	Navigation	Task comp. time & workload		[35]
		Navigation errors	[35]	
Tablet, smartphone	Web searching	Workload	[94, 95]	
		Number of hits per query		[94, 95]

user performance. Table 2.12 presents the effect of technical context on evaluation metrics. The majority of the studies indicate that user performance is influenced by device; however, several lines of evidence suggest that device has no significant effect on some of the user performance metrics. Using a wearable device improves user performance in terms of task completion time for information retrieval [182] and target selection [101] tasks; but does not have a main effect for navigation task [32, 35]. On the other hand, it increases target selection errors [101] and navigation errors [35]. Pielot et al. [32] reported that, using a wearable device for navigation purpose decreased traveled distance, walking speed and distraction compared to using only a map.

Finally, Table 2.13 represents the interactions between contextual factors. As can be

Table 2.13: Interactions between contextual factors (SD: Significant difference, ID: Insignificant difference)

Interaction	**Task Domain**	**Time / Speed**		**Error / Accuracy**		**Workload**	
		SD	**ID**	**SD**	**ID**	**SD**	**ID**
Mob. x Light	Reading		[75, 119] [79, 80]		[75, 119] [79, 80]		[75, 79] [80]
Mob. x Light	Word search	[75, 119]			[75, 119]	[75]	
Mob. x Device	Target selection		[101]		[101]		[101]
Mob. x Encumb.	Tapping		[115]	[115]			
Mob. x Encumb.	Target selection	[113]		[113]			
Mob. x Task	Cognitive tasks	[89]		[89]			
Mob. x Task	Reading	[121]					
Env. x Task	Text entry		[66]		[66]		

seen in the table, there are both significant and insignificant interactions on performance metrics depending on the task domain.

2.4.2.6 Qualitative Results

One third of the studies conducted subjective evaluations to investigate participants' preference on their proposed interaction mechanism over existing methods. Lucero and Vetek [135], Williamson et al. [143] and McMillan et al. [67] focused on the social acceptibility of gesture based interactions. According to Lucero and Vetek [135] participants had problems with gesture and speech based interactions while using smartglasses in public. Similarly, Williamson et al. [143] reported that some participants felt uncomfortable due to social reactions while they were playing a game on their mobile phone with gestures. McMillan et al. [67] gave examples about how smartwatch notifications affect social context when users are in conversations with others. Sarsenbayeva et al. [52] stated that presence of a stress task psychologically and physically affected participants' perceived task performance. Oulasvirta et al. [8] observed participants' strategies to reduce fragmented attentional resources. Ranasinghe et al. [210] questioned users' trust on GPS based navigation apps while they are having GPS problems. Participants indicated a distrust for such cases and provided strategies to overcome the problem. Ioannidou et al. [208] reported that, although a considerable number of participants had at least one fall from stairs, majority of them continue to use their phone while walking down or up stairs. Most of the participants commented that light conditions affected mobility. Applying a questionnaire, Piazza et al. [211] investigated the intentions of participants to cross a street while using a mobile phone. Hiniker et al. [60] observed and interviewed adult caregivers. They investigated the apps caregivers used while they were parenting and strategies to reduce phone absorption. Although some caregivers thought that it is acceptable to use a mobile phone while parenting if the children are safe; others thought that mobile phone use should be minimised. However, most caregivers agreed that using a mobile phone makes it more difficult to pay attention to the children. Tigwell et al. [9] interviewed mobile content designers and reported that majority of them did not consider situational visual impairments in their designs due to several reasons including limited resources, restricted design scope or unawareness on situational impairments. Finally, Mäntyjärvi and Seppänen [199] stated that, adaptive behaviour is acceptable for participants; however, participants highlighted the importance of accuracy and control over adaptations.

A few studies aimed to investigate the effect of context on attentional and gait walking behaviour by using observational data. Hyman et al. [55] placed a unicycling clown on a pedestrian walking path to illustrate inattentional blindness and observed that cell phone users were less likely to see the clown. Similarly, Chen and Pai [58] evaluated situational blindness, situational deafness and situational awareness with a clown walking at the opposite direction with the pedestrians and playing national anthem. They concluded that pedestrians who were using their smartphones for the tasks such as playing a game or listening to music failed to see the clown or hear the anthem more than non-smartphone users. Harper et al. [57] observed that, small device users who were standing or sitting had less attention switches than walking small device users. Alsaleh et al. [56] suggested that pedestrians who were distracted with smartphones had slower walking speed than non-distracted pedestrians. Finally, Horberry et al. [59] observed pedestrians crossing streets and revealed that pedestrian smartphone users had higher risk of colliding with another pedestrian or vehicle, and crossing at the wrong time or place than non-smartphone users.

2.4.3 Summary and Discussion

Mobility has been a popular contextual factor in the literature and there are two research trends on mobility. One group of researchers consider interacting with a small screen device or a wearable device as the main task and walking as a secondary task that may affect user performance. These researchers have asked participants to complete goal-oriented tasks under various mobility conditions. The others consider walking as the main task and interacting with a small screen device or a wearable device as a dual task that may affect gait or distract users. These researchers asked participants to walk on a treadmill or cross a street in virtual environment while either completing action-oriented tasks or without using a device. Previous research findings on the effect of walking on user performance have been inconsistent and contradictory. On the other hand, there is a consensus among scientists that interacting with a small screen device or a wearable device affects how we walk or pay attention to our surroundings.

Most of the previous research on the effect of contextual factors on SIIDs have focused on mobility conditions, while a relatively small body of literature has covered other contextual factors. This may be explained by the fact that, different mobility conditions can be easily simulated by ensuring identical experimental settings across all sessions. On the other hand, other contextual factors such as lighting level, temperature, ambient noise or people around interaction are hard to control and they can easily differ between sessions. However, a more comprehensive study would include all the contextual factors that may cause SIIDs.

There are several in-situ studies that have been conducted in users' own environment. However, such studies remain narrow in focus dealing only with interaction techniques. Unfortunately, our systematic review did not reveal any studies that have collected context and performance data in the wild and had drawn some implications on the interaction between context and user performance.

The researchers have faced several problems with automatic data collection mechanisms. Yatani and Truong [127] argued that, additional sensors such as accelerometers to observe physical workload might cause disruption during the experiments.

Agostini et al. [152] suggested that, recording eye movements might give insights about when participants looked at screen or path in walking experiments. Kjeldskov and Stage [68] highlighted the difficulty of screen capturing while participants are walking. Finally, Reyal et al. [144] stated that, collecting data from participants' devices outside of experiment scope might cause privacy concerns.

2.5 Discussion

Our systematic review shows that there are significant research on investigating the contextual factors for SIIDs and also investigating the effect of contextual factors on users' performance. In our systematic review, we investigated 14978 publications and we identified as 187 relevant. Our review also revealed that SIIDs are first started to be discussed in the literature around 2000 and after almost 20 years there are still significant number of publications around this phenomenon. In fact, our review shows that the devices and interaction styles studied changes over time but we still do not have a good understanding of SIID itself and contextual factors (see Figure 2.3). Our review has also shown that early studies tend to focus on walking and the effect of light conditions, but the recent studies do focus on larger contextual factors as discussed in our paper. In fact, in Figure 2.1 and Figure 2.2, we show two different tag clouds for papers published in 2000s and 2010s which clearly show the trends in the papers we reviewed. Our review also shows that most of the user studies are conducted with students at universities and staff, and also they are mainly done in controlled, mainly indoor, lab environments. It is quite understandable that finding participants for studies is not easy therefore it is quite normal that participants are chosen from close proximity. However, more studies can be conducted with users from different backgrounds. Furthermore, our review also shows that more studies can be conducted in the *wild* to better reflect the context and cause of SIIDs. With the recent developments in sensors and small screen devices and their capabilities, we expect that more studies will be conducted in the wild.

When we look at contextual factors, our review shows that mobility and location are the key contextual factors that are studied as the physical context dimensions. However, there are less studies considering functional place and also the sensed environmental attributes such as lighting levels, vibration levels, temperature, etc. Therefore, in the future more studies can focus on these contextual dimensions. Regarding temporal context, our review also shows that studies designed in the literature tend to be shorter studies and longitudinal studies are rare. This can easily be explained that it is much easier to control shorter studies. However, with the new technological advances, we are hoping that more longitudinal studies can be conducted to investigate SIIDs. Furthermore, our review also reveals that synchronisation aspect, time of the day or year aspects are not so widely investigated. In the literature the focus so far has been on the walking speed and multitasking. In fact, as a task context, multitasking is widely studied. However, interruptions of users including stressor tasks, etc. are not investigated much. Further studies can be conducted to better understand the impact of such task-based contextual factors on SIIDs. Our review also shows that much of the studies investigating contextual factors were conducted in very much controlled environments with simplified tasks. Therefore, in the future more studies can be conducted in the wild with more complex tasks which will of course require different systematic methods for collecting and analysing data. Compared to physical, task and temporal contexts, our review shows that social context is the least studied context in the literature. Most of the studies focus on investigating SIIDs when people are alone. However, further studies can be conducted to investigate the SIIDs experienced when people are in different social environments. Of course such studies would require better merging and combination of social science methods and technical user studies.

Our systematic review reveals that many metrics are used to investigate SIIDs. These metrics range from very-well known ones such as task completion time to very custom metrics such as metrics to assess stress-level. When we look at effect of the context on SIIDs, we can see that in the literature we have studies confirming the findings and studies that contradict each other. For example, the effect of walking on user performance has been inconsistent. However, it is consistently shown in the literature that presence of dual task while walking is an important cause of SIIDs. Our review fully shows all the effects presented in the literature which are summarised in Tables 2.7–2.12.

In brief, our systematic review shows that SIID is an important phenomenon with a significant number of publications surrounding the context and the effect of context. Overall, this review shows that ability-based design could be a good approach to design applications for small screen and wearable devices that take into account users' context better and could cause less SIIDs. Ability-based design is described as identifying and exploiting users' abilities rather than their disabilities to enhance interaction by using available resources. It recommends systems to sense context that may affect users' abilities [14]. Further research is needed to show what would be the actual effect of ability-based designed applications to the users' performance.

2.6 Conclusion

In this article, we conducted a systematic review of the literature on investigating the effect of different contextual factors on SIIDs. For this purpose, we reviewed a wide range of articles from online platforms, popular HCI conferences and journals.

Our first research question targeted the contextual factors that have been examined in the literature for small screen or wearable device interaction that may cause SIIDs. We classified the factors used in the literature under the context framework that has five main dimensions: physical, temporal, social, task and technical context. Our review has shown that, physical context has been widely studied to observe the effect of mobility or location. On the other hand, further studies regarding the effect of social or temporal factors would be worthwhile.

The other research question was related to the effect of contextual factors on small screen or wearable device users' performance. For this purpose, we first identified the evaluation metrics which helped researchers compare different contextual factors. Then, we reviewed the literature to see whether contextual factors have significantly affected corresponding evaluation metrics. Our results have shown that, there is no consensus among the scientist that have worked on popular contextual factors such as mobility or encumbrance. On the other hand, there are relatively few studies considering each of the other contextual factors. Further experiments using a broader range of contextual factors, could help us understand how we interact with small screen or wearable devices under SIIDs. Technological advances in these devices enable collecting more precise and functional data with the help of available sensors.

Small screen and wearable devices have an important role in our everyday lives. As they transform into new forms and provide new functionalities, we start to use them for new purposes in different environments. Our study has not only shown that researchers have made great progress in understanding SIIDs and the factors that may cause them, it has also shown that we have yet a lot to learn on the effect of context on users' performance.

CHAPTER 3

SENSING THE CONTEXT AND USER PERFORMANCE

Our literature review showed that most of the studies had focused on specific contextual factors, such as mobility, and the experiments had been conducted in controlled environments. This chapter investigates the effect of context on users' typing performance in their everyday settings. For this purpose, we conduct a remote user study in the wild with 48 participants. We collect smartphone keyboard interactions and context details in this user study. Using the dataset collected, we implement an error detection mechanism.

Section 3.1 starts with a summary of literature around text entry metrics, the effect of context on user performance in the text entry task domain, and text entry studies conducted in the wild. Section 3.2 explains our user study's material, decisions and procedure. Finally, Section 3.3 presents our automatic error detection approach that combines several approaches in the literature.

3.1 Literature Summary

The main aim of this study is to investigate the effect of context on users' typing performance. We start our literature review by identifying the metrics to measure typing performance. We continue our review to identify the typing errors and typing behaviour in daily life. Then, we review the literature surrounding how context affects these performance metrics in the text entry task domain. The studies reviewed have been conducted in controlled settings, and a systematic understanding of how context affects users' typing performance in their daily life is still lacking. Finally, we review the studies that automatically measure typing performance in the wild.

3.1.1 Text Entry Metrics

Several metrics are used to measure typing performance. In terms of typing speed, words per minute (WPM) and keystrokes per second (KSPS) are the most popular metrics. WPM considers only the length of transcribed text and how long it takes to produce it. It considers a word every five characters entered and measures the number of words entered in a minute [212]. KSPS is used to measure the number of keystrokes made in a second. It is useful when taking error corrections into account [212]. Keystroke per character (KSPC) and error rate (ER) are widely used for accuracy. KSPC is the ratio of the total entered character count to the length of the transcribed string [213]. ER is the ratio of incorrect characters to all characters entered [213]. Minimum string distance between intended and transcribed text can also be used for ER [212]. Error rates can be assessed in several ways especially for the studies conducted in the wild without a predefined task model.

3.1.1.1 Unintentional Errors

Text entry errors can be classified into unintentional and intentional typing errors. For unintentional errors, Durham et al. [214] identified four types of word-level text errors as follows: transposition, the wrong letter, extra letter, and missing letter. According to Chen et al. [215], mobile device users experience character ambiguity, missing or additional character, bounce (repeating characters), long-press, and transposition errors. Greene et al. [216] also reported extra or missing character, incorrect shifting, wrong character, adjacent character, transposition, and misplaced character errors. A word in a text can contain many errors, and the number of errors even can exceed the number of correct characters [217].

3.1.1.2 Intentional Errors

The intentional typing errors are referred to as “text-speak" [218] and consists of intentional corruptions on the words [219] for several reasons such as mirroring positive and negative emotions [220], increasing perceived playfulness [221], typing faster to reduce latency in a synchronized way of communication [219, 222], or common words in communication slang [223, 224]. Using text-speak, users compress the text by employing abbreviations, phonetic substitutions, and character strategies [222]. Table B.1 in Appendix B illustrates common text-speak techniques in daily texting use and examples for these techniques. Since the users are intentionally typing in this way, they should not be associated with a performance problem.

3.1.1.3 Corrected/Uncorrected Errors

Wobbrock and Myers [225] classified errors into insertions, omissions, and substitutions and considered whether these errors were corrected or uncorrected. The corrected errors do not appear in the transcribed text; however, they can be traced using the input stream and can help measure the text entry performance better. There might also be cases when a user did not make an error but somehow thought that he/she did and deleted the corresponding text to rewrite it (corrected no-errors). Uncorrected errors are the errors that remain in the final transcribed text. The total ER is then can be calculated by the sum of the corrected ER and uncorrected ER [225].

3.1.2 The Effect of Context on Users' Text Entry Performance

Table 3.1 presents a summary of the research on the effect of contextual factors on text entry performance. Instead of a character level entry rate, Hoggan et al. [97] used time to enter phrases. They showed that sitting in a subway train decreased the entry time than sitting in the laboratory. Similarly, Crease et al. [76] used task completion time and showed that walking and avoiding hazards together decreased the task completion time.

Most of the previous research has focused on mobility conditions, while a relatively small body of literature has covered other contextual factors. The popularity of mobility conditions may be explained by the fact that different mobility conditions can be easily simulated by ensuring identical experimental settings across all sessions. On the other hand, other contextual factors such as lighting level, temperature, ambient noise, or social context are hard to control, and they can easily differ between sessions.

Table 3.1: Literature summary (↓: decreased, ↑: increased, ⊘: no significant effect, -: NA)
WPM: Words per minute, KSPS: Keystrokes per second, KSPC: Keystroke per character, ER: Error rate

Ref.	Context	Factor	WPM	KSPS	KSPC	ER
[66]	Environment (lab/indoor real-world)	Being in a public place	-	⊘†	-	⊘
[102]	Mobility (stable/mobile)	Walking	-	⊘	-	↑
[97]	Mobility (stable/mobile)	Being in a subway train	-	-	⊘	⊘
[82]	Mobility (stable/mobile)	Walking	⊘	-	-	↑
[36]	Mobility (stable/mobile)	Walking	⊘	-	-	↑
[86]	Mobility (stable/mobile)	Walking	⊘	-	-	↑
[61]	Mobility (stable/mobile)	Walking	-	-	-	↑
[90]	Mobility (stable/mobile)	Walking	↓	-	-	⊘
[92]	Mobility (stable/mobile)	Walking	↓	-	-	⊘
[66]	Mobility (stable/mobile)	Walking	-	↓†	-	↑
[120]	Mobility (stable/mobile)	Walking	↓	-	-	↑
[50]	Urban noise (indoor/outdoor)	Outdoor noise	-	↓∗	-	-
[50]	Speech (meaningful/meaningless)	Meaningful noise	-	↓∗	-	-
[226]	Ambient light	Dimmed light or sunglasses	-	⊘∗	-	⊘
[52]	Multitasking	Presence of stress task	-	⊘∗	-	⊘
[76]	Multitasking	Avoiding hazards	-	-	-	↑
[153]	Distractions	Presence of distraction	↑	-	↓	↓

The metrics used in corresponding studies were (∗) time per character entry, and (†) character per minute which can be interpreted as KSPS.

Although mobility has been a popular contextual factor, there have been inconsistent findings on its effect on typing speed and error rate. Several studies have shown that environment [66], ambient light [226], and multitasking [52] did not affect typing speed and error rate. Jain and Balakrishnan [153] demonstrated that the presence of distraction increased typing speed. They commented that the increase in typing speed might be related to higher attention caused by higher distraction.

Table 3.1 only encloses the most relevant studies to text entry. However, there is considerable research on the effect of different contextual factors on the other task domains. Sarsenbayeva et al. [47] and Goncalves et al. [48] showed the effect of ambient temperature on target selection time. Barnard et al. [75, 119] compared low and high lighting levels for reading and searching tasks and indicated that there is a main effect of lighting level on task completion time and workload but not on error rate. Encumbrance also has a main effect on target selection accuracy and time [41, 111, 113, 159]. Further detailed review on the effect of context on users' performance can be found in our systematic review in Chapter 2.

3.1.3 Text Entry Studies in the Wild

Several methods are used to detect texting errors in the wild which include the following:

Using Transcribed Text

Palin et al. [7] conducted a study with considerably large number of participants. However, instead of allowing participants to enter free text during their daily activities, they presented texts for participants to transcribe. Similarly, Reyal et al. [144] compared two different keyboard methods in the wild. Although participants used their own devices during their daily activities, they performed transcription tasks. Schlögl et al. [227] and Wimmer et al. [228] used game-based approaches to measure a large number of text entry metrics for different soft keyboards.

Using an Offline Lexicon

Evans and Wobbrock [18] aimed to measure desktop text entry performance in the wild. They used WPM, uncorrected, and corrected error rates. To detect errors and distinguish between corrections and edits, they used an offline lexicon (English Lexicon Project). If a word was in the lexicon, it was considered correct. Nicolau et al. [17] conducted a study with blind users to observe their everyday typing behaviour on mobile devices. They used the Hunspell lexicon for error detection.

Using a Spell-Checker

Komninos et al. [229] observed typing error and correction behaviour in the wild. They used a spell-checker to classify errors as slight and severe concerning the suggestions for entered text. Wong et al. [230] used Aspell for spelling error detection in chat records.

Using an Online Query Service

Evans and Wobbrock [18] used Bing API in addition to the offline lexicon. The API returned suggestions if the word is incorrect. They considered these suggestions the intended words. Wong et al. [230] used an online resource to expand abbreviations. Varnhagen et al. [224] used NetLingo and UrbanDictionary as helper services.

Manual Analysis

Battestini et al. [231] conducted an in-situ study to analyze text message topics. They analyzed whom the participants texted with, why they sent text messages and their thoughts on text messaging. They manually categorized topics of conversation. Nicolau et al [17] also manually analyzed words that do not appear in the offline lexicon to detect text entry errors.

Rodrigues et al. [6] compared transcription, composition, and passive sensing approaches in terms of the effort of the participants, the effect on the typing behavior, and the participants' perception of privacy by conducting a study in the wild. They observed that the amount of effort put on the participants was the least for passive sensing and the most for composition tasks. Moreover, they ensured a policy that no raw data was collected during the study and provided a mechanism to pause capturing data. These helped to create a perception of privacy and trust among the participants. On the other hand, the composition task, in which participants were asked to compose a text describing their daily activities, caused more cognitive effort and privacy issues.

Using transcribed text in a controlled environment may increase the consistency of a study; however, these studies fail to cover real-world cases. On the other hand, detecting users' intention when there is no task model, and users enter free text in daily settings is challenging [17]. Evans and Wobbrock [18], and Nicolau et al. [17] used offline lexicons to detect typing errors along with other resources such as an online search API or manual analysis. However, this method may not be practical due to many out-of-vocabulary words for morphologically rich languages, such as Turkish [20]. Using offline lexicons fail when the text contains words changed with text-speak for daily language. Torunoğlu and Eryiğit [20] carried out a study to normalize Turkish text on social media. The transformations they applied on out-of-vocabulary tokens include letter case transformation, removal of character repetitions, transformations on emo style writing, proper noun detection, Deasciification, vowel restoration, and accent normalization.

According to Evans and Wobbrock [18], if a participant deletes some characters and enters text again, there are two possibilities: participant either corrects an error or changes his/her mind to enter a new word. They used a straightforward approach. If an online query returned suggestions for removed words and reentered words matched with one of these suggestions, it was identified as an error correction. Otherwise, it was considered an edit. Nicolau et al. [17] noted that blind users tend to correct errors as soon as possible. As a result, they needed to check incomplete words with final words to distinguish between errors and edits. First, they checked whether removed and reentered characters were adjacent. If all characters were adjacent, it was considered an error correction. Then, they used Hunspell to retrieve spelling suggestions for the removed text. It was considered an error correction if the final text was in the spelling suggestions. Finally, if the minimum string distance between deleted and final word was more than half of the words' length, they considered it an edit. Otherwise, it was considered an error correction.

In summary, our literature review showed that the effect of the context on users' typing performance had been investigated mainly in controlled settings. Conducting studies in the wild is essential to collect more realistic data on the tasks users do in their daily lives. Processing the text entries in daily lives requires a mechanism to measure typing performance automatically. There have been several attempts to achieve this; however, such studies remain narrow in focus dealing only with formal writing. Morphologically rich languages and daily texting language should also be considered.

3.2 User Study – in the Wild

We conducted a user study in which we aimed to collect user performance data and corresponding context factors in the wild. In general, we adopted the Experience Sampling Method (ESM) [232] for context labels and automated collecting performance data. This section explains the methodology of our study in full detail.

3.2.1 Data Collection Framework

We used the AWARE Framework for data collection [233]. AWARE is a framework that provides logging mechanisms for various available sensors in Android devices. It also enables data collection using ESM. One of the significant advantages of AWARE is that it is an open-source framework so that anyone can extend it for specific purposes. Moreover, it provides mechanisms to register and unregister to studies, pause and resume the data collection, disable data synchronization when the battery level is low or the smartphone is not connected to Wi-Fi, and monitor the studies.

AWARE is a general-purpose framework and did not have certain features required within our study. Therefore, we implemented several features on the AWARE framework for our study. First, we embedded the informed consent form in the app and made it the opening page after installation (see “Online Repository” Section on page 7). The participants could participate in the study only if they read and accepted the informed consent form. We also created a demographics form (see Section 3.2.6 for the questions and available options). After participants registered in the study, we asked them to complete this form once. The app retrieved sensor configurations in JSON format from a web service and configured the study automatically. Since we were interested in sensor data only when participants entered text, the app disabled all sensors and stopped recording when the screen was off or locked. When the screen was on or unlocked, it again enabled sensors. This optimization helped us reduce bandwidth use and storage required for overall study data. We also ignored the sensor data for the sessions that participants did not enter text. If a participant entered text longer than five characters, the app asked the participant to answer five questions about the context. To not interrupt the participants during a task, the app showed these questions when the participants returned to the home screen. We removed all unnecessary permission requests by disabling irrelevant modules, such as cameras or contacts.

During the study, the app collected data from the following sensors: accelerometer, applications, barometer, battery, communication, gravity, gyroscope, light, linear accelerometer, locations, magnetometer, proximity, rotation, screen, significant motion, telephony, and Wi-Fi. Moreover, after each keystroke, the text before and after the keystroke was recorded. The app did not take pictures, capture videos or audio, collect passwords, or collect screen content. It also did not send messages on behalf of the participants.

We deployed the AWARE server application on METU NCC servers. The interaction between the app and the server was handled with the HTTPS protocol. We used this application for monitoring and data collection purposes. Finally, we created a web page for the study [1].

3.2.2 Methodological Decisions

In general, we followed the guidelines provided by van Berkel et al. [234] and focused on having an unobtrusive study as much as possible. We aimed to minimize participants' burden; therefore, we presented a set of options for each context dimension (details are given in the following section) and asked participants to select only one option for each question. With this approach, we avoided free text entry inputs. Each notification was triggered after a text entry event. If participants did not respond to questionnaires, they expired in 30 seconds and were removed from the notification panel. This notification timeout aimed to ensure that the participants answered the questions within the context of the text entry. We put at least 15 minutes between two questionnaires to not overload participants, and participants received these questionnaires at most eight times a day. We asked participants to keep the app installed

[1] `https://iam.ncc.metu.edu.tr/cabas-user-study/`, last access: 21.01.2022

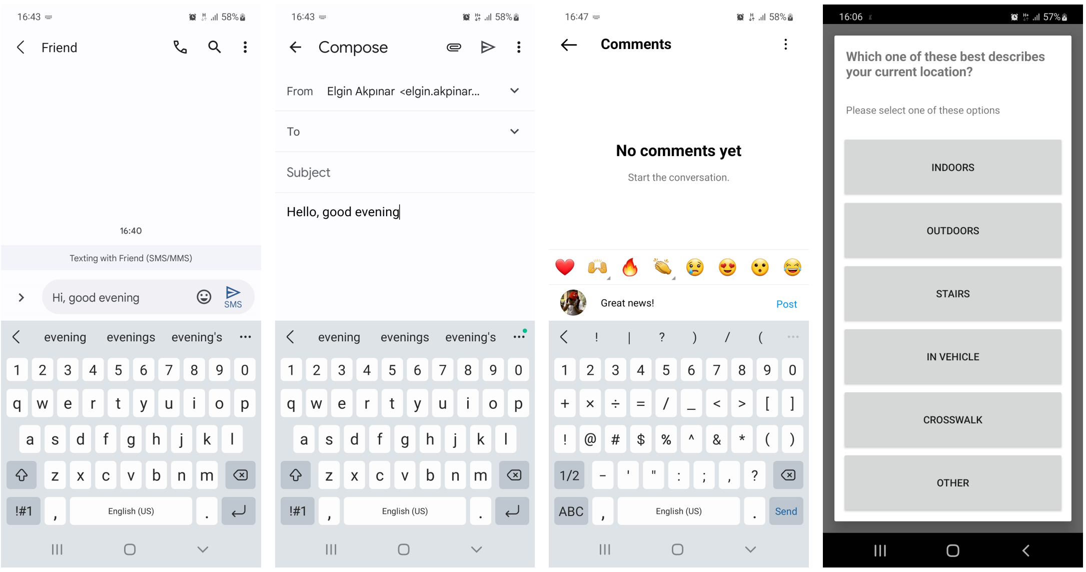
for at least one week to capture context data during different daily activities. Finally, participants were informed that they could pause data collection any time they felt uncomfortable sharing their private data.

3.2.3 Context Labeling Questions

In our systematic review in Chapter 2, we investigated the effect of the context under five dimensions: physical, temporal, social, task, and technical contexts. We also reviewed the relevant ESM-based research and collected the contextual factors used. Then, we combined our findings with our systematic review. In this study, we investigated the effect of context based on the following dimensions: environment (physical context), mobility (physical context), social, multitasking (task context), and distraction (task context). We used the following questions to collect context labels in participants' perspectives:

- Which one of these best describes your current location? *(environment)*
- Which one of these best describes your mobility condition? *(mobility)*
- Which one of these best describes people around you? *(social)*
- Did you handle any other task along with text entry? *(multitasking)*
- Is there anything that interrupted/distracted your interaction with your mobile device? *(distractions)*

We provided a set of options for each question and asked participants to select only one option at a time. For instance, the options for the environment consisted of *indoors*, *outdoors*, *stairs*, *in a vehicle*, *crosswalk*, and *others*. These options are created based on our findings from the systematic review in Chapter 2 and previously conducted ESM-based research studies. Overall options available for these questions are listed in Appendix C.



(a) Sending a text message (b) Composing an email (c) Posting on social media (d) Sample ESM question

Figure 3.1: Sample text entry activities captured and ESM question

3.2.4 Procedure

The participants were provided with a set of instructions for installing the app and registering for the study. These instructions were published online on the user study page [2]. The participants had to confirm that they read the consent form and voluntarily signed up for the study. Then, they were asked to fill a demographics form. After they completed this step, the app was activated to collect data. There was no specific task model; the participants interacted with their smartphones like they usually do. The app captured any text entered by the participants, such as while sending a text message (i.e., Samsung Messages, Figure 3.1a), composing an email (i.e., Gmail, Figure 3.1b), or posting comments on social media (i.e., Instagram, Figure 3.1c). During their interactions, the app asked them to answer a set of questions regarding the current context (Figure 3.1d). The data synchronization process was fully automated; background services posted the data to the server after the interaction was completed. To quit the study, participants removed the app from their smartphones. The participants were rewarded with $10/70TL worth of a gift card from Amazon or a preferred

[2] https://iam.ncc.metu.edu.tr/cabas-user-study-instructions/, last access: 21.01.2022

local shopping site if they completed the study for at least a week.

3.2.5 Administration

This study was approved by the METU Applied Ethics Research Center with 516 ODTU 2019 protocol number[3]. In the consent form, it was clearly stated that the participation was voluntary. Moreover, we also stated that we would not collect the content of the password fields and share or publish the textual content collected during the study. We indicated that the data would be evaluated with an automated process for academic purposes only. We ensured that the questions used during the study would not include questions that would cause personal discomfort. We stated that any participant could leave the study for any reason by just removing the app. Finally, we explained how to pause and resume data collection if the participants had any privacy concerns.

We adopted the Snowball Sampling technique and started our user experiment with personal contacts on July 27th, 2020. Then we announced the study on social media including Facebook, Instagram and LinkedIn, and via various email groups. The study was designed to be conducted fully remotely. We instructed participants if they had problems with the setup and warned them if there was a problem with the data flow. We also notified them when one week period of the study was over. The study was conducted and administered for 58 days and completed on September 22nd, 2020.

3.2.6 Participation and Demographics

Overall, 55 participants downloaded and installed our app on their devices. Seven participants either had a technical problem or decided not to participate in the study; thus, they uninstalled the app within the same day of installation. Other 48 participants kept the app installed from three days to 10 days (mean is 7.3 days and median is 7 days). In our data analysis, we did not exclude any of these 48 participants' data.

Figure 3.2 shows the demographics of the participants. Among 48 participants, 29

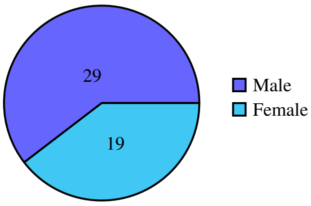
[3] `http://ueam.metu.edu.tr/`, last access: 20.12.2021

were male, and 19 were female. 23 participants were aged between 25-34, 19 participants were 18-24, five participants were 35-54, and one participant was over 55. The majority of the participants (40) used their right hands as their dominant hands. 29 participants had Bachelor's Degree, nine had a Master's Degree, six completed high-/secondary school, and four had a Doctorate. 43 participants have been using mobile devices for more than four years.

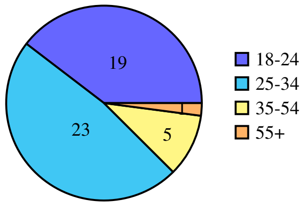
Reported occupations included student (20), software engineer (7), teacher (5), biologist (2), architect (2), data analytics manager (2), pilot (1), QA (1), business analyst (1), network admin (1), communications manager (1), machine engineer (1), game designer (1), doctor (1), researcher (1), DB admin (1). The majority of the participants (37) reported Turkish as their native language. Other native languages were Turkmen (3), Arabic (2), Persian (1), Urdu (1), Korean (1), English (1), Hindi (1), Dutch (1). Finally, participants sent data from different countries, including the United States, Senegal, Mauritania, Germany, Netherlands, Belgium, Greece, Russia, Turkmenistan, India, Pakistan, Kyrgyzstan, Kazakhstan.

All participants were smartphone users. The device sizes ranged between 5.1 and 6.67 inches (mean: 5.9, median: 6.0). None of the participants was excluded due to the device size. The diversity of the device types was unexpectedly high. Participants used 37 different models of eight brands and five different SDK versions. The keyboard apps used by the participants were Samsung Keyboard (18), Gboard – the Google Keyboard (17), and Microsoft SwiftKey Keyboard (12). Table D.1 in Appendix D provides a summary of participants' devices.

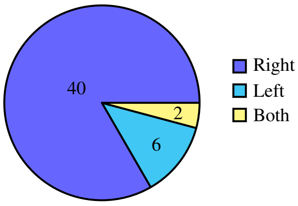
We asked our participants to ignore all of the context labeling questions whenever they felt that paying attention to the questions would cause safety problems, such as while driving. The overall compliance rate to the context labeling questions is 55.32%. Minimum and maximum compliance rates among 48 participants are 2.34% and 100.00%, respectively, and the mean compliance rate among the participants is 58.68% (standard deviation is 27.56%, and median is 65.22%). Figure 3.3 illustrates the histogram for participants' context labels.



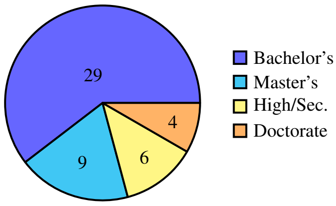
(a) Gender



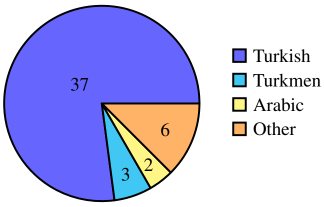
(b) Age



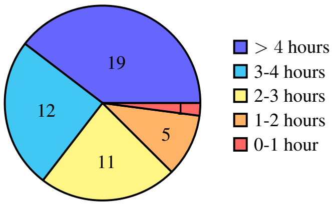
(c) Preferred hand



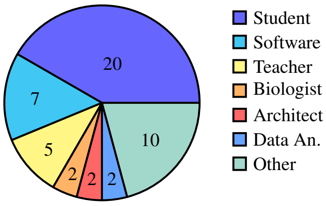
(d) Education level



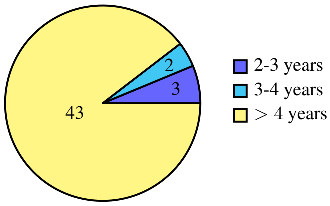
(e) Native language



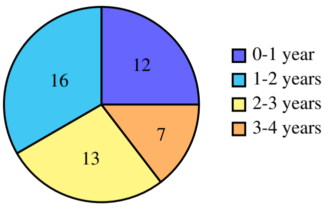
(f) Daily screen time



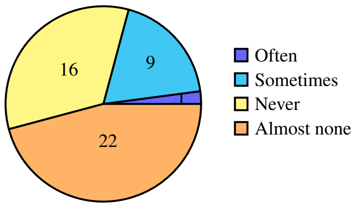
(g) Occupation



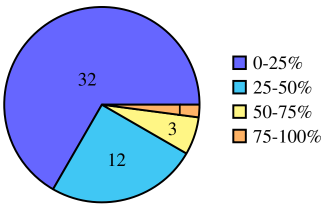
(h) Experience with a mobile device



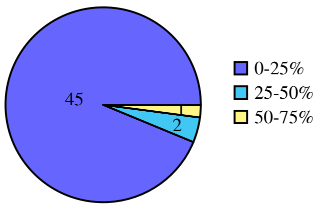
(i) Experience with current mobile device



(j) Colliding freq. while occupied with a mob. dev.



(k) Percent of mobile device use while walking



(l) Percent of mobile device use while crossing

Figure 3.2: Demographic data

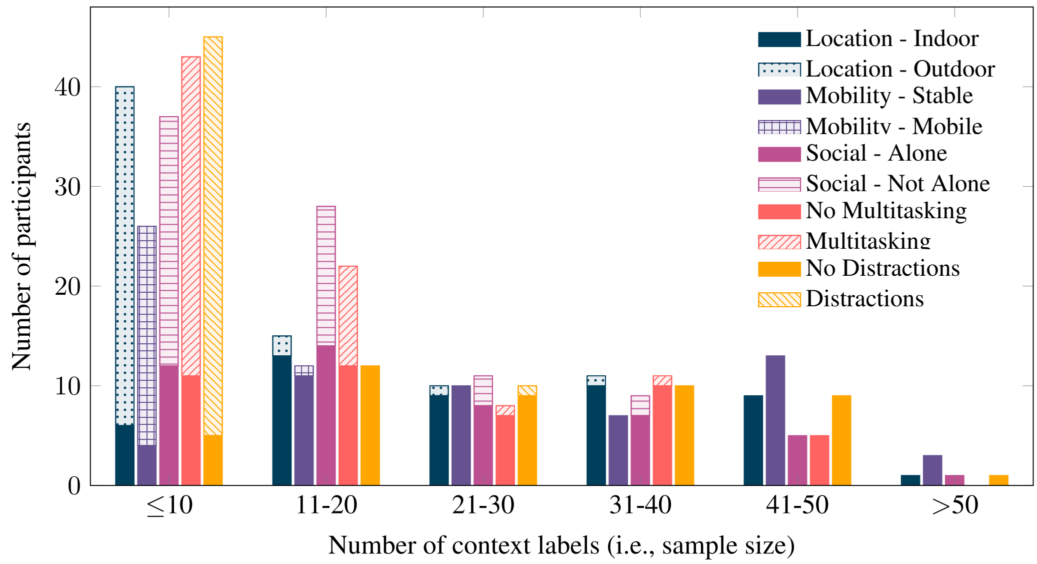


Figure 3.3: Histogram for participants' context labels

3.2.7 Participants' Self Evaluation on Typing Errors

After context labelling questions, we asked participants who had deleted any text during the current session whether they had made a typing error. Table 3.2 presents the participants' responses to this question. According to this table, the majority of text removals were not caused by a typing error.

If the participants selected yes or maybe options, we asked a further question regarding the cause of this typing error. Table 3.3 illustrates the participants' responses to this question. Surprisingly, the participants indicated that there was no particular reason for their typing error in the majority of the cases.

3.3 User Performance Modelling: Detection and Correction

As can be seen from the previous section, instead of transcribing the given text, in our study, participants entered text to complete their daily tasks without having a predefined task model. This section explains the techniques employed to process user data, and evaluate the users' performance and in particular, the techniques used to assess the users' typing errors and corrections.

Table 3.2: Participants' responses to whether they made a typing error in the current session

Participants' response	Count	Percent (%)
No	787	76.93
Yes	158	15.44
Maybe	78	7.62

Table 3.3: Responses to typing error causes

Cause of Typing Error	Count	Percent (%)
No particular reason	142	60.17
Something that interrupts me	16	6.78
Other task I am busy with	15	6.36
My current mobility situation	15	6.36
People around me	11	4.66
My current location	10	4.24
Multiple of these	5	2.12
Other	18	7.63
No response	4	1.69

3.3.1 Data Model

The AWARE Framework logs the keyboard events at the character level and adopts the transcription sequence paradigm by capturing the entire transcription after every keyboard action [235]. A keyboard log is recorded for every keyboard interaction that either inserts or removes a character. Figure 3.4 illustrates the table columns for the keyboard logs and sample data for single character insertion (Figure 3.4a), single character deletion (Figure 3.4b), multiple character insertion (auto-completion, Figure 3.4c), and substitution (auto-correction, Figure 3.4d). The columns include the timestamp of the keyboard event, an ID for interacting users, and the package name of the app used during the keyboard event. Moreover, the text just before the keyboard event and the text just after the keyboard event are also logged in two separate columns. A boolean field indicates if the field that the text entered is a password field.

Action	User enters “t" character
Timestamp	1595852196097
Device ID	b297f8e-2086-11ec-9621-0242a
Package name	com.whatsapp
Before text	Con
Current text	Cont
Is deleted	0
Is password	0

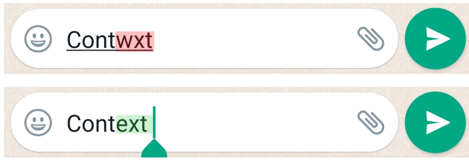
(a) Insertion example (single character)

Action	User removes “y" character
Timestamp	1595852197213
Device ID	b297f8e-2086-11ec-9621-0242a
Package name	com.whatsapp
Before text	Cony
Current text	Con
Is deleted	1
Is password	0

(b) Deletion example

Action	User completes to “Context"
Timestamp	1595852198123
Device ID	b297f8e-2086-11ec-9621-0242a
Package name	com.whatsapp
Before text	Con
Current text	Context
Is deleted	0
Is password	0

(c) Insertion example (multiple characters)



Action	Keyboard corrects to “Context"
Timestamp	15958521998712
Device ID	b297f8e-2086-11ec-9621-0242a
Package name	com.whatsapp
Before text	Contwxt
Current text	Context
Is deleted	0
Is password	0

(d) Substitution example

Figure 3.4: Sample actions and corresponding data model

The AWARE Framework masks the text entered into password fields; therefore, it does not collect the password phrases. We used this field to ignore such phrases in the overall process. Finally, we added a boolean field to indicate if the user is entering or deleting text. Using this data model, it is possible to generate the input stream and distinguish between the type of actions by comparing two consecutive transactions. If the simultaneous actions edit discontiguous parts of the text, they are considered as substitution [235].

3.3.2 Trial Identification and Tokenization

Evans and Wobbrock [18] and Nicolau et al. [17] refer to the set of keyboard interactions when participants complete a single task as trials, as in the laboratory experiments. To analyze data systematically, we have also applied several steps to segment overall keyboard data into trials. First, the overall text input stream was grouped by the participants, and then the timestamp values sorted the keyboard data list of each participant in ascending order. Then, iterating over the keyboard data lists, we compared two consecutive keyboard data to check if a new trial had started. If the participant switched to another app, we considered a new trial. Finally, we compared *before text* and *current text* values of two consecutive keyboard data. For instance, if a non-empty *current text* was followed by an *empty before* text, it indicated that the user either submitted or cleared the text just entered, and a new trial started. Unlike Evans and Wobbrock [18], and Nicolau et al. [17], we did not use screen events for starting a new trial since they may indicate an interception due to context. However, we adopted Evans and Wobbrock [18] and Nicolau et al. [17]'s approach to detect pauses. In summary, we calculated the mean time interval between keyboard events and added three standard deviations to obtain a threshold. Using the dataset obtained from 48 participants, we calculated that the mean difference between two successive non-backspaces or backspaces was 285 ms, a non-backspace following a backspace was 742 ms, and a backspace following a non-backspace was 899 ms. After adding three standard deviations to each mean value, we had 2,346 ms, 9,867 ms, and 23,189 ms, respectively, as the pause segmentation times. Overall, we segmented 938,431 keyboard interaction data into 42,018 trials.

3.3.2.1 Trial Validation

We excluded the trials if they only included a URL, a numeric value, a password, or a text with less than five characters. Moreover, since our participants entered text in any language they like, we also applied language criteria. We used Apache Tika

tika-langdetect package[4] and language-detector library[5] for language detection. We ignored the trials other than Turkish and English. The text language was used later to determine the proper resource to check if a token was correct or a typing error. Details are given below. Overall, we excluded 9,497 (22.6%) trials.

3.3.2.2 Tokenization and Token Selection

We tokenized the trials by using the whitespaces. We did not use punctuation characters in tokenization to detect typing errors caused by unintentional punctuation characters between the tokens. However, we removed punctuation characters and emojis at the end of the tokens. Some types of tokens serve a particular purpose in the text; however, they are out-of-vocabulary due to their structural appearance. They are very prevalent, especially in social media, and can be listed as follows [236]:

- URLs
- Email addresses
- Mentions (i.e., @mention)
- Hashtags (i.e., #hashtag)
- Emojis (i.e., :D)
- Vocatives (i.e., hahaha[6])

In addition to these, we recognized serial numbers (one or two upper case letters followed by a set of numeric characters), websites or domain names (i.e., metu.edu.tr), and file names with extensions (i.e., sample.pdf). We used regular expressions to check if a token matched with one of these cases. When our implementation found a match, it excluded the token from the dataset, similar to Han and Baldwin [237]. We excluded 4,574 tokens and had 135,254 valid tokens overall with 38,768 distinct tokens. Table 3.4 illustrates an overview of the dataset in terms of trials and actions after trial and token validations.

[4] `https://tika.apache.org/`, last access: 29.09.2021
[5] `https://github.com/optimaize/language-detector`, last access: 29.09.2021
[6] Turkish word equivalent to lol in English.

Table 3.4: Overview of the dataset in terms of trials and actions

	Size	Min	Max	Mean	Median	Std. Dev	Overall
Keystrokes in trials	32,301	5	325	23.56	18.00	20.40	760,980
Characters entered in trials	32,301	5	302	24.31	19.00	19.34	785,383
Participants' daily trials	384	1	1,220	84.12	38.00	121.05	32,301
Participants' overall trials	48	46	3,307	672.94	369.50	776.84	32,301
Insertions	48	800	68,018	14,508.79	9,305.00	16,548.20	696,422
Deletions	48	55	4,830	1,151.21	652.50	1,328.65	55,258
Substitutions	48	0	761	108.23	28.00	173.46	5,195
Auto completions	48	0	952	73.60	27.00	162.18	3,533

3.3.3 Error Detection

The previous section explained how we segmented the input stream into trials and tokenized each trial. The next step was to process these tokens to calculate the performance metrics. First, we needed to identify if a token has a typing error. Then, we had to distinguish between typing error corrections and edits when participants removed some characters and reentered new text. When participants realized typing errors and corrected them, the total interaction time corresponded to the final output. On the other hand, when the participants changed their minds and decided to write something else, the total interaction time corresponded to the overall text that the participants intentionally wrote. Therefore, to better measure the metrics related to typing speed, we had to distinguish between these two cases.

Algorithm 1 represents the pseudocode to validate a token. To check if a token has a typing error, we used several resources. First, we used Hunspell spellchecker [238] as it has been widely used in similar studies and supports multiple languages, including Turkish. Moreover, we checked if a token appeared in METU Turkish Corpus [239], a collection of 2 million words of Turkish text. Finally, we used the spellchecker implementation of the Zemberek project [240]. We only used METU Turkish Corpus and spellchecker of Zemberek if the participant's native language or text language was Turkish. If a token was identified as correct in one of these tools, it was accepted as a correct word without any typing error. In addition to these, we used several resources for lookup purposes. These resources include location names and coun-

Algorithm 1 Algorithm to validate a token

Require: $token \neq$ “”

procedure IsValid ($token, lang$)

  $hunspellInstance \leftarrow Hunspell.instance(lang)$

  if $hunspellInstance.isValid(token)$ **then return** TRUE

  end if

  if $lang =$ “tr” **then**

    if $zemberek.isValid(token)$ **then return** TRUE

    else if $metuCorpus.isValid(token)$ **then return** TRUE

    else if $addressLookup.contains(token)$ **then return** TRUE

    else if $abbreviationLookup.contains(token)$ **then return** TRUE

    else if $textSpeakLookup.contains(token)$ **then return** TRUE

    end if

  else if $lang =$ “en” **then**

    if $textSpeakLookup.contains(token)$ **then return** TRUE

    end if

  end if

  if $Bing.query(token, options =$ ”$spellcheck : true$”$)\neq$ EMPTY **then return** TRUE

  else if $Bing.query(token, options =$ ”$site : tureng.com$”$)\neq$ EMPTY **then return** TRUE

  else if $Bing.query(token, options =$ ”$site : urbandictionary.com$”$)\neq$ EMPTY **then return** TRUE

  end if

  return FALSE

end procedure

try codes [241], Turkish abbreviations[7], and a set of Turkish slang and text speak words [236]. Finally, we used Hunspell and Zemberek suggestions for vowel restoration.

Daily conversations or social media posts may also include some out-of-vocabulary but valid words, such as brand names, social media accounts, or technical terms. Even if a user intends to type such words, offline resources fail to identify these words as correct. Therefore, we used Bing Spell Check and Search APIs, similar to [18]. To reduce the number of calls to these APIs, we only sent requests for tokens that offline tools could not recognize. Moreover, we did not send the overall text content of the

[7] https://tdk.gov.tr/wp-content/uploads/2019/01/K%c4%b1saltmalar_Dizini.pdf, last access: 02.10.2021

trial for the spell checking; we only sent a single token at a time. Finally, we used additional query options such as filtering results for Urban Dictionary[8] and Tureng Multilingual Dictionary[9] sites to retrieve specific search results. Urban Dictionary is a crowdsourced resource and can be used to check the words in English slang and daily language [242]. Tureng Dictionary is a Turkish and English dictionary [243], and it makes use of resources in many different fields, such as engineering, law, and medicine.

Table 3.5: Token validation rules

Rule	Description	Algorithm	Examples
Case alternatives	Tokens that become valid after converting to lower, upper, and proper noun cases	**return** $isValid(toLower(token))$ **or** $isValid(toUpper(token))$ **or** $isValid(toProper(token))$	*en.* usa → USA *tr.* ankara → Ankara *en.* COME → come
Dialectical or accent use	Tokens that are written in informal forms in text speak and become valid after applying dialectical and accent transitions	$dSet \leftarrow \{dialectSet\}$ $tSet \leftarrow \{transitionSet\}$ **for** $i = 0; i < dSet.length; i{+}{+}$ **do** ... **if** $token.contains(dSet[i])$ **then** ...... $token.replace(dSet[i], tSet[i])$ ...... **if** $isValid(token)$ **then** .... **return** $true$ **return** $false$	*tr.* yapcaz → yapacağız (we will do) *tr.* yapıyom → yapıyorum (I am doing) *tr.* yapmicam → yapmayacağım (I won't do) *en.* goin → going
Repeating characters	Tokens that become valid after removing repetitive characters, that are generally used for expressing emotions	**for** $i = 1; i < token.length; i{+}{+}$ **do** ... **if** $token[i] = token[i-1]$ **then** ...... $n \leftarrow token.remove(i)$ ...... **if** $isValid(n)$ **then return** $true$ **return** $false$	*tr.* evettttt → evet (yes) *en.* hiiiii → hi
Removing vowels	Tokens constructed by removing vowels from a valid token	$suggests \leftarrow suggestions(token)$ **for** s **in** $suggests$ **do** ... **if** $token = s.removeVowels()$ ...... **then return** $true$ **return** $false$	*tr.* tmm → tamam (OK) *tr.* slm → selam (hi) *en.* msg → message

Continued on next page

[8] https://www.urbandictionary.com/, last access: 14.01.2022
[9] https://tureng.com/, last access: 14.01.2022

Table 3.5 – *Continued from previous page*

Rule	Description	Algorithm	Examples
English & French words	English and French words in a non-English or non-French text	**return** $isValid(token, "en")$ **or** $isValid(token, "fr")$	*en.* playlist *en.* data *fr.* voilà
Deascii-fication	Tokens that become valid after applying deascification, to detect use of "i", "o", "u", "c", "g", and "s" instead of "ı", "ö", "ü", "ç", "ğ", and "ş" characters	$ascii \leftarrow \{i, o, u, c, g, s\}$ $tr \leftarrow \{ı, ö, ü, ç, ğ, ş\}$ **for** $i = 0; i < ascii.length; i{+}{+}$ **do** ... $d \leftarrow token.replace(ascii[i], tr[i])$ ... **if** $isValid(d)$ **then return** $true$ **return** $false$	*tr.* Turkce → Türkçe *tr.* isik → ışık (light)
Proper nouns	Proper nouns with missing a-postrophes, generally ignored in text speak	**for** $i = 1; i < token.length; i{+}{+}$ **do** ... $n \leftarrow token.put(i, "'")$ ... **if** $isValid(n)$ **then return** $true$ **return** $false$	*tr.* Elginin → Elgin'in *tr.* Ankaraya → Ankara'ya *en.* Elgins → Elgin's
Phonetic substitution	Tokens that are intentionally corrupted by replacing some characters with phonetically similar forms or nonalphabetic characters	$pSet \leftarrow \{phoneticRuleSet\}$ $tSet \leftarrow \{transitionSet\}$ **for** $i = 0; i < pSet.length; i{+}{+}$ **do** ... **if** $token.contains(pSet[i])$ **then** ...... $token.replace(dSet[i], tSet[i])$ ...... **if** $isValid(token)$ **then** ....**return** $true$ **return** $false$	*tr.* kardeshim → kardeşim (my sister/brother) *tr.* qanqa → kanka (dude) *tr.* \$eker → Şeker (Sugar) *tr.* yawrum → yavrum (my little one) *en.* c@ → cat
Misspelled conjunction	Tokens that ends with a frequently misspelled conjunction	**for** c **in** $conjunctionSet$ **do** ... **if** $token.endsWith(c)$ **and** $isValid(token.remove(s), "tr")$ ...... **then return** $true$ **return** $false$	*tr.* tamammı → tamam mı (is it OK) *tr.* alırmısın? → alır mısın? (would you take?)
Frequents	Frequent spelling mistakes	$fmSet \leftarrow \{frequentMistakesSet\}$ **return** $fmSet.contains(token)$	*tr.* yalnış → yanlış (wrong) *en.* succesful → successful
Neologism	Non-Turkish words followed by a Turkish suffix	**for** s **in** $suffixSet$ **do** ... **if** $token.endsWith(s)$ **and** $isValid(token.remove(s), "en")$ ...... **then return** $true$ **return** $false$	*tr.* hack-lemek (hacking) *tr.* item-ler (items) *tr.* edit-lemek (editing) *tr.* pick-leyip (picking)

Overall, we combined the approaches of Evans and Wobbrock [18], Nicolau et al. [17], and Torunoğlu and Eryiğit [20]. To check if a token is valid in the corresponding language, we mainly followed the approaches of Evans and Wobbrock [18] and Nicolau et al. [17], except for the manual analysis. To identify text-speak words, we followed Torunoğlu and Eryiğit [20]. Moreover, we converted the words to lower, upper and proper noun cases and checked if they were valid. We applied a set of transition rules on the tokens. For instance, we removed repeating characters and checked if the resulting word was valid. Table 3.5 presents these rules with corresponding algorithms and examples. If the transformed word was valid, then it was accepted as correct.

We considered the following cases as typing errors:

- transposition errors (i.e., cont[xe]t),
- punctuation marks separating two words without any whitespace (i.e., context[.]factor),
- invalid tokens becoming valid after changing some characters with adjacent characters on the keyboard (i.e., cont[r]xt),
- tokens with missing or extra characters with respect to Hunspell and Zemberek suggestions (i.e., cont[]xt, conte[r]xt),
- one of the adjacent characters to spacebar separating two words (i.e., context[n]factor),
- two consecutive words as a token without any whitespace between them (i.e., context[]factor).

To distinguish between edits and error corrections, we first checked for adjacent character errors similar to Nicolau et al. [17]. For this purpose, we modified the minimum string distance calculation to accept two characters as equal if they are adjacent on the keyboard. If this new distance value is zero but removed and reentered texts are different, it is accepted as a correction of adjacent character error. In addition to the method of Nicolau et al. [17], we also checked for transposition, missing and extra character, bounce (repetition), and wrong character errors. We detected the difference between removed and reentered text. If the removed text segment is the reverse of the reentered text segment, it is considered a correction of a transposition error. If the removed text segment and reentered text segment have only one character, it is considered missing, extra, or wrong character error. If the removed text segment is a repetition of a single character, it is considered a bounce error correction. We observed that unintentional space characters and punctuation were commonly corrected. Finally, we applied suggestion checks similar to Nicolau et al. [17], and Evans and Wobbrock [18].

According to Zhang et al. [244], using auto-correction could help to prevent typos; while, it may also result in typos. Unfortunately, Nicolau et al. [17] and Evans and Wobbrock [18] did not consider the effect of auto-correction on typing errors. If the user removed an auto-corrected text, we considered it as an error correction. We also checked if the removed text is valid. If so, we considered it as an edit. Finally, we calculated the edit distance between removed and reentered text. If the edit distance is more than half of the lengths of both texts, we considered it an edit. Otherwise, it was classified as an error correction. According to Arif and Stuerzlinger [245]'s experiments, half of the users correct typing errors immediately (character-level), and the other half correct after a few keystrokes (word-level). The majority of the users that apply word-level correction correct after two to five characters. For this reason, we applied this method to both in-text and end-of-text replacements.

Algorithm 2 represents the pseudocode to distinguish between error corrections and edits. The corresponding procedure accepts non-empty removed and reentered text and a boolean value to indicate whether an auto-correction event occurred within the removed text's typing process. We compared the current text with the before text value to check this. If the minimum string distance between the current text and the before text is more than one, it indicates either an auto-complete or an auto-correction. If the current text starts with the before text, it is an auto-complete event (see Figure 3.4c); otherwise, it is an auto-correction event (see Figure 3.4d).

Algorithm 2 Algorithm to distinguish between error corrections and edits

Require: $removed \neq$ "", $reentered \neq$ ""

procedure ERRORCORRECTIONOREDIT ($removed, reentered, autoCorrection$)

$msd_{adj} \leftarrow MSD_{adj}(removed, reentered)$ ▷ Adjacent characters are accepted as equal in MSD

if $msd_{adj} = 0$ **then return** CORRECTION ▷ Adjacent character error

else if $startsWith(removed, \text{"}z\text{"})$ **and** $startsWithUppercase(reentered)$

and $startsWith(toLowerCase(reentered), toLowerCase(removeFirst(removed)))$ **then**

return CORRECTION ▷ Failing to switch to uppercase error

end if

$removed_{diff} \leftarrow getDifference_{removed}(removed, reentered)$ ▷ Consider only the difference

$reentered_{diff} \leftarrow getDifference_{reentered}(removed, reentered)$

if $removed_{diff} = reverse(reentered_{diff})$ **then**

return CORRECTION ▷ Transposition error

else if removeSpaces($removed_{diff}$) $= removeSpaces(reentered_{diff})$ **then**

return CORRECTION ▷ Missing space error

else if removeRepeatedChars($removed_{diff}$) $= removeRepeatedChars(reentered_{diff})$ **then**

return CORRECTION ▷ Bounce error

else if $length(removed_{diff}) = 0$ **and** $length(reentered_{diff}) = 1$ **then**

return CORRECTION ▷ Missing character error

else if $length(removed_{diff}) = 1$ **and** $length(reentered_{diff}) = 0$ **then**

return CORRECTION ▷ Extra character error

else if $length(removed_{diff}) = 1$ **and** $length(reentered_{diff}) = 1$ **then**

return CORRECTION ▷ Wrong character error

else if $getHunspellSuggestions(removed).contains(reentered)$ **then**

return CORRECTION

else if $getZemberekNormalizations(removed).contains(reentered)$ **then**

return CORRECTION

else if $autoCorrection$ **and** $startsWith(reentered, beforeCorrection(removed))$ **then**

return CORRECTION ▷ Auto-correction error

else if $MSD(removed, reentered) > (length(removed) + length(reentered))/4$ **then**

return EDIT ▷ Edit by distance

else if $!autoCorrection$ **and** $isValid(removed)$ **then**

return EDIT ▷ Edit by removing a correct word

end if

return CORRECTION

end procedure

Out of 21,683 text changes, we classified 18,192 (83.9%) error corrections and 3,491 (16.1%) edits. Participants corrected errors 379 times on average (standard deviation: 480.62, median: 204.5) and edited 72 times (standard deviation is 82.81, median is 46).

3.3.4 Evaluation

Before using our findings to investigate the effect of context on users' typing performance, we had to evaluate our error detection implementation. Due to our commitments in the consent form (see Section 3.2.5), we conducted a follow-up study with the same participants in our first user study and asked them to evaluate our implementation on the data they sent during the first experiment. As we have indicated in the consent form, we only automatically processed their data.

3.3.4.1 Procedure

In this follow-up study, we automatically prepared Excel files that included user data and our system's classification typing error and correction. These files are used to collect users' feedback such that we could compare users' feedback with the system's classification. This study mainly included the following three parts:

1. Invitation: We sent invitation emails to the participants who provided their email addresses (46 participants). We briefly explained the purpose of the study and asked the participants to reply if they agree to participate voluntarily in the follow-up study. We did not offer compensation for this follow-up study. For the analysis of this evaluation, the participants were asked to permit to process the responses manually. They were free to leave the study anytime they wanted. Moreover, we asked them to remove any text from the file without changing the row order if they feel uncomfortable sharing it.

2. Uncorrected Error Detection Task: For the first section, we randomly selected ten words that our system classified as correct and ten words that our system classified as typing errors. These were automatically chosen. Next, we listed

the overall text participant entered with the selected words and asked participants to enter '`F`' if they think they made a typo and '`T`' otherwise.

3. Edits & Error Correction Detection Task: In the second section, we selected ten cases classified as edit and ten cases classified as error correction. These were again automatically chosen. Next, we listed the overall text participant entered with the removed and reentered texts and asked participants to enter '`F`' if they think they corrected an error and '`T`' otherwise.

3.3.4.2 Material

For both sections, we only selected Turkish and English texts. To help participants to remember the context, we provided the overall text participant entered. Moreover, we selected the texts with at least three words. For the words and cases to be evaluated, we selected the words with at least three characters. The Excel files were created automatically and sent to participants without manual revision.

We did not provide the verdict of our system in the Excel file that we sent to participants. Moreover, the words and cases were randomly listed in the Excel file so that participants could not predict the system verdict. In a separate file, we saved the system verdict in the same order in the Excel file. We asked participants not to change the order of the words and cases to match the system verdict with the participant response. We provided the instructions with relevant examples to better guide the participants.

3.3.4.3 Study Duration and Participation

We sent the initial invitation on April 13rd, 2021. As of April 19th, 2021, we sent the Excel files to all participants who responded positively. The overall evaluation process was completed on May 15th, 2021. 30 of 46 participants agreed to participate in the follow-up study. We had to eliminate one participant since there was not enough text in Turkish and English. One participant changed their mind and decided not to participate due to their busy schedule. Two participants did not respond after we sent the Excel file. Overall, we received evaluation results from 26 participants.

Table 3.6: Evaluation results of the Follow-Up Study

	Proposed Approach		Nicolau et al. [17]		Evans and Wobbrock [18]	
	Error Detection	Error Corr. Detection	Error Detection	Error Corr. Detection	Error Detection	Error Corr. Detection
Accuracy	0.797	0.761	0.661	0.744	0.651	0.460
Sensitivity	0.818	0.726	0.979	0.671	0.839	0.234
Specificity	0.789	0.849	0.536	0.918	0.577	0.993
Precision	0.603	0.922	0.453	0.951	0.438	0.987
F1 Score	0.694	0.782	0.620	0.760	0.576	0.525

3.3.4.4 Results

We compared participants' responses to the system verdict and calculated the system accuracy. We also implemented the approaches proposed by Evans and Wobbrock [18] and Nicolau et al. [17] to compare our results with the literature. Table 3.6 presents the evaluation results. According to these results, our system has higher accuracy than Evans and Wobbrock [18], and Nicolau et al. [17]'s approaches. They are more sensitive since they classify the words that do not appear in offline lexicon and online query services. On the other hand, this results in lower specificity.

We created confusion matrixes to analyze the results of the evaluation. When deciding on the error rule set, one of our assumptions was that there must be a space character after punctuation characters. Our system classified 26 cases as typing errors in the evaluation data set. However, participants labelled 21 of these cases as correctly spelt text. Moreover, we used Hunspell suggestions and Tureng query results, which resulted in 10 and 13 incorrect classifications, respectively.

We compared removed text with the reentered text of the same length to detect edits and error corrections similar to Evans and Wobbrock [18], and Nicolau et al. [17]. However, this resulted in higher string distances in case of unintentional or missing characters. Moreover, we assumed that if the removed text consists of valid words, it was an edit. Unfortunately, participants labelled 29 of such cases as error correction. Finally, we observed that auto-completed text replacement might not indicate an error

Table 3.7: Evaluation results of the revised system

	Error Detection	Error Correction Detection
Accuracy	0.913	0.871
Sensitivity	0.923	0.881
Specificity	0.909	0.849
Precision	0.800	0.935
F1 Score	0.857	0.813

correction in all cases.

Based on these observations, we updated the above rules to increase the system's accuracy. For error detection, we accepted commonly made mistakes as correct since the participants may have written them intentionally (i.e., *tommorow-tomorrow* (English) or *lavobo-lavabo* (Turkish)). Moreover, we assumed that punctuations should follow the last word without a space character, and a space character must be inserted after the punctuation. However, some participants stated that they either put no space character after the punctuation or intentionally put a space character before the punctuation. Finally, we assumed that if a participant used any Turkish characters in a trial, any deasciified character corresponds to a typing error. However, we observed that some participants deasciified specific Turkish characters while using the others without deasciification. Therefore, we relaxed this assumption. To distinguish between error detection and edits, we calculated the edit distance between the removed text and the reentered text with the same length. However, this method failed with the missing character problems. We moved forward in the reentered text as long as the edit distance decreased. In some cases, the participants unintentionally tapped on adjacent characters while switching to uppercase mode. We implemented a rule for these cases. Finally, we assumed that it was an edit if both removed and reentered text were correct words. However, the participant responses showed that it was not a valid assumption. After these changes, we compared the new verdicts with the participant responses. Table 3.7 presents the evaluation results of the modified system.

3.4 Summary

This section began by describing our in-situ user study and how we collected text entry interactions and corresponding context factors in the wild. Then, it described the implemented mechanism to decide a given text contains typing errors automatically. It went on to describe the process of the follow-up study to evaluate these mechanisms and their results. The following chapter presents the statistical procedures and the results obtained from them to investigate the effect of context on user performance by using the data we collected in our main user study explained in Section 3.2.

CHAPTER 4

THE EFFECT OF THE CONTEXT ON USER PERFORMANCE

After we completed error and edit/error correction detection mechanisms and evaluated them, we investigated the effect of the context on user performance in text entry tasks. This chapter explains the procedure and results of this investigation.

4.1 Design and Procedure

In Section 3.1.1, the metrics for typing performance were identified. We used those four metrics in our investigation as dependent variables. We calculated the total error rate for ER metric by summing up the corrected and uncorrected error rates. Moreover, we accepted intentional errors due to text-speak as correct and did not include them in the ER calculation. For independent variables, we used participants' responses to context labels in five dimensions: environment, mobility, social, multitasking, and distraction. Table 4.1 shows the groups for the independent variables and corresponding context labels. We excluded the participants' data from the dataset of the contexts if the participant's context labels did not include samples for two groups. For instance, if participants did not provide samples for both indoor and outdoor groups, we excluded their data from all statistical calculations regarding environment context. We calculated the performance metrics for each sample and associated them with the contextual labels users have assigned.

Our study did not provide a predefined task. The participants have interacted with their smartphones as in their daily lives. Some participants have spent more time with their smartphones and entered text more than the others. Figure 4.1 illustrates the histogram for the number of trials each participant made during the study. Moreover,

Table 4.1: Independent variables and corresponding context labels

Context	Groups	Options
Environment	Indoor	Indoors, In vehicle
	Outdoor	Outdoors, Crosswalk
Mobility	Stable	Lying down, Sitting, Standing
	Mobile	Walking, Running
Social	Alone	Alone
	Not Alone	With 2-4 friends/ family members/colleagues, With a friend/ family member/colleague, With more than 4 friends/family members/colleagues, With strangers (crowded), With strangers (not crowded)
Multitasking	Nothing	Nothing
	Multitasking	I am carrying a box/bag/other, I am doing home-activities (cleaning, cooking, etc), I am having a conversation with someone around me, I am having breakfast/lunch/dinner, I am shopping, I am trying to avoid collision while walking, I am working, Multiple of these
Distractions	Nothing	Nothing
	Multitasking	I am in a hurry, I am interrupted by someone, I am interrupted by something unexpected, I need to check something from time to time, There are obstacles/people/cars on walking path, Multiple of these

a Kruskal-Wallis H test showed that there was a statistically significant difference between the participants' performances, in terms of WPM ($\chi^2(47) = 6923.066$, $p < 0.0001$), KSPS ($\chi^2(47) = 8563.796$, $p < 0.0001$), KSPC ($\chi^2(47) = 1630.444$, $p < 0.0001$), and ER ($\chi^2(47) = 1156.542$, $p < 0.0001$). Comparing the samples of each context group in such an unbalanced data could cause biases in the results. Therefore, we calculated mean and median values of each metric for all participants under different context groups.

Further statistical analysis showed that typing speed metrics (WPM and KSPS) were normally distributed for most participant and context group pairs. In contrast, error rate metrics (KSPC and ER) significantly deviated from a normal distribution for all

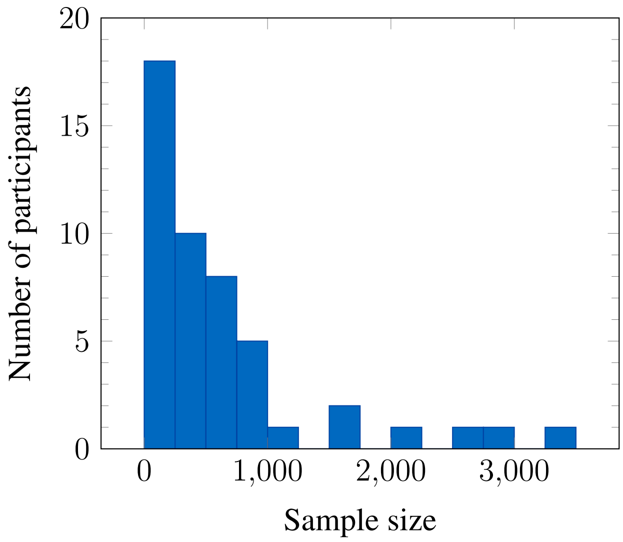


Figure 4.1: Histogram for participants' sample sizes

participant and context group pairs. The histograms for both KSPC and ER were in the long tail form. A KSPC value of 1 means no correction and corresponds to 56.1% of the cases in the data. Similarly, an ER value of 0 means no uncorrected typing error and corresponds to 59.5% of the cases. WPM and KSPS, on the other hand, had no such values that dominated the sample. Moreover, WPM and KSPS were measured based on the text length and duration of the corresponding trial. On the contrary, KSPC and ER were calculated based on our error detection implementation results. The mean for these metrics would result in a poor estimate of central tendency, while the median would yield more valid results [246]. As a result, we used the mean values of WPM and KSPS and median values of KSPC and ER to investigate context effects on user performance. In our statistical analysis, the value of each performance metric under one context group was compared to the other group in a pairwise manner on each context dimension. Therefore, the p-value was adjusted using the Bonferroni correction method to reduce Type-I errors [247] and divided by the number of pairwise comparisons (0.05/5 = 0.01).

Our statistical analysis first checked if the data were normally distributed for all groups for each context factor. We conducted Kolmogorov-Smirnov and Shapiro-Wilk tests. We used the Wilcoxon Signed-Rank Test when the test results showed that the data significantly deviated from a normal distribution. Otherwise, we used

Paired T-Test to compare the user performance under two context factors. All tests were conducted in 95% confidence intervals.

4.2 Research Questions

We addressed the following research questions in our investigation:

R1 – Environment: *Does being in an outdoor environment affect text entry performance in terms of typing speed (WPM and KSPS), and error rate (KSPC and ER) compared to being in an indoor environment?*

R2 – Mobility: *Does walking affect text entry performance in terms of typing speed (WPM and KSPS), and error rate (KSPC and ER) compared to being stable?*

R3 – Social context: *Does the presence of other people around affect text entry performance in terms of typing speed (WPM and KSPS), and error rate (KSPC and ER) compared to being alone?*

R4 – Multitasking: *Does multitasking affect text entry performance in terms of typing speed (WPM and KSPS), and error rate (KSPC and ER) compared to having no multitasking?*

R5 – Distractions *Does the presence of distractions affect text entry performance in terms of typing speed (WPM and KSPS), and error rate (KSPC and ER) compared to having no distractions?*

4.3 Results

Table 4.2 and Table 4.3 presents the results of our investigation. Table 4.4 summarizes these results for each context and performance metric.

4.3.1 R1 – Environment

Environment of the participant significantly affects user performance in terms of KSPC ($Z = 82.0$, $p = 0.001$) and ER ($Z = 42.0$, $p < 0.0001$). Participants in outdoor

Table 4.2: Paired T-Test results for the effect of context on users' mean WPM and KSPS values

Context	Metric	Group	N	Mean	Median	Std. dev.	Results
Environment	WPM	Indoor	31	41.832	41.975	6.834	t(30)=0.499, p=0.622
		Outdoor	31	41.307	41.155	7.467	
	KSPS	Indoor	31	3.670	3.688	0.604	t(30)=-0.079, p=0.938
		Outdoor	31	3.677	3.663	0.646	
Mobility	WPM	Stable	15	45.466	46.047	5.501	t(14)=-0.912, p=0.377
		Mobile	15	46.785	44.064	9.441	
	KSPS	Stable	15	4.011	4.032	0.460	t(14)=-0.517, p=0.613
		Mobile	15	4.069	3.894	0.763	
Social	WPM	Alone	38	42.281	42.203	6.868	t(37)=1.001, p=0.323
		Not Alone	38	41.500	41.225	6.636	
	KSPS	Alone	38	3.711	3.731	0.606	t(37)=0.614, p=0.543
		Not Alone	38	3.670	3.636	0.619	
Multitasking	WPM	Nothing	35	41.461	42.049	7.287	t(34)=-1.377, p=0.178
		Multitask	35	42.635	43.364	6.415	
	KSPS	Nothing	35	3.644	3.697	0.609	t(34)=-2.217, p=0.033
		Multitask	35	3.793	3.850	0.569	
Distractions	WPM	Nothing	35	41.987	41.959	6.836	t(34)=-0.169, p=0.867
		Multitask	35	42.145	41.941	8.290	
	KSPS	Nothing	35	3.704	3.726	0.597	t(34)=-0.364, p=0.718
		Multitask	35	3.732	3.712	0.727	

N Sample size, *p < 0.01, **p < 0.0001

condition had higher KSPC (1.052 ± 0.083) than participants in indoor condition (1.013 ± 0.028). Similarly, ER was higher for outdoor condition (1.217 ± 1.822) than indoor condition (0.311 ± 1.028). Environment does not significantly affect user performance in terms of WPM (t(30)=0.499, p=0.622) and KSPS (t(30)=-0.079, p=0.938).

4.3.2 R2 – Mobility

Mobility of the participant significantly affects user performance in terms of only ER (Z = 10.0, p = 0.003). ER was lower for stable condition (0.442 ± 1.256) than mobile

Table 4.3: Wilcoxon Signed-Rank Test results for the effect of context on users' median KSPC and ER values

Context	Metric	Group	N	Mean	Median	Std. dev.	Results
Environment	KSPC	Indoor	31	1.013	1.000	0.028	Z = 82.0, p = 0.001*
		Outdoor	31	1.052	1.024	0.083	
	ER	Indoor	31	0.311	0.000	1.028	Z = 42.0, p < 0.0001**
		Outdoor	31	1.217	0.000	1.822	
Mobility	KSPC	Stable	15	1.016	1.000	0.033	Z = 30.0, p = 0.095
		Mobile	15	1.040	1.031	0.045	
	ER	Stable	15	0.442	0.000	1.256	Z = 10.0, p = 0.003*
		Mobile	15	1.057	0.000	1.443	
Social	KSPC	Alone	38	1.013	1.000	0.031	Z = 115.0, p < 0.0001**
		Not Alone	38	1.028	1.008	0.038	
	ER	Alone	38	0.303	0.000	0.932	Z = 96.0, p < 0.0001**
		Not Alone	38	0.593	0.000	1.153	
Multitasking	KSPC	Nothing	35	1.027	1.000	0.063	Z = 157.0, p = 0.009*
		Multitask	35	1.042	1.014	0.059	
	ER	Nothing	35	0.494	0.000	1.410	Z = 129.0, p = 0.002*
		Multitask	35	1.223	0.000	2.160	
Distractions	KSPC	Nothing	35	1.017	1.000	0.034	Z = 81.0, p < 0.001*
		Multitask	35	1.052	1.008	0.081	
	ER	Nothing	35	0.467	0.000	1.136	Z = 99.0, p < 0.001*
		Multitask	35	1.460	0.000	2.122	

N Sample size, *p < 0.01, **p < 0.0001

condition (1.057 ± 1.443). Mobility does not significantly affect user performance in terms of WPM (t(14)=-0.912, p=0.377), KSPS (t(14)=-0.517, p=0.613) and KSPC (Z = 30.0, p = 0.095).

4.3.3 R3 – Social Context

Social context significantly affects user performance in terms of KSPC (Z = 115.0, p < 0.0001) and ER (Z = 96.0, p < 0.0001). The presence of other people resulted in higher KSPC (1.028 ± 0.038) than participants in alone condition (1.013 ± 0.031). Similarly, ER increased with the presence of other people (0.593 ± 1.153) compared

Table 4.4: The effect of context on user performance (↓: decreased, ↑: increased, ⊘: no significant effect)

Context	Factor	Typing Speed		Error Rate	
		WPM	KSPS	KSPC	ER
Environment (indoor/outdoor)	Being outdoors	⊘	⊘	↑	↑
Mobility (stable/mobile)	Being mobile	⊘	⊘	⊘	↑
Social (alone/not alone)	Presence of other people	⊘	⊘	↑	↑
Multitasking (with/without multitask)	Presence of multitasking	⊘	⊘	↑	↑
Distraction (with/without distraction)	Presence of distraction	⊘	⊘	↑	↑

to alone condition (0.303 ± 0.932). The presence of other people does not significantly affect user performance in terms of WPM (t(37)=1.001, p=0.323) and KSPS (t(37)=0.614, p=0.543).

4.3.4 R4 – Multitasking

Multitasking significantly affects user performance in terms of KSPC ($Z = 157.0$, $p = 0.009$), and ER ($Z = 129.0$, $p = 0.002$). Multitasking resulted in higher KSPC (1.042 ± 0.059) than no multitasking (1.027 ± 0.063). Similarly, ER was higher for multitasking conditions (1.223 ± 2.160) than no multitasking condition (0.494 ± 1.410). Multitasking does not significantly affect user performance in terms of WPM (t(34)=-1.377, p=0.178), and KSPS (t(34)=-2.217, p=0.033).

4.3.5 R5 – Distractions

Distractions significantly affect user performance in terms of KSPC ($Z = 81.0$, $p < 0.0001$) and ER ($Z = 99.0$, $p < 0.0001$). Presence of distraction resulted in higher KSPC (1.052 ± 0.081) than no distraction (1.017 ± 0.034). Similarly, ER was higher for distraction condition (1.460 ± 2.122) than no distraction condition (0.467 ± 1.136). Distractions do not significantly affect user performance in terms of WPM (t(34)=-0.169, p=0.867) and KSPS (t(34)=-0.364, p=0.718).

4.3.6 Task Context

Participants entered text in 231 different apps during our study, and 139 apps left after trial and token validations. The most frequently used apps include WhatsApp, Instagram, Messenger, Google Chrome, Tinder, and Telegram. We retrieved the category of each app by using the categories in Google Play Store[1]. Then, we grouped the apps by their categories and selected the most frequently used app categories: communication (i.e., Whatsapp), social (i.e., Instagram), tools (i.e., Google), and productivity (i.e., Notes). Finally, we investigated the effect of the category of the app used on user performance. A repeated-measures ANOVA with a Greenhouse-Geisser correction determined that mean WPM differed statistically significantly between different app types (F(1.878, 20.663) = 3.955, p = 0.037). Participants were fastest while using a communication app (44.857 $\pm$ 1.962), slowest while using a productivity app (36.083 $\pm$ 2.604), had 42.037 $\pm$ 1.955 WPM in social apps, and had 43.549 $\pm$ 3.861 WPM in tool apps. Post hoc analysis with a Bonferroni adjustment revealed that WPM was statistically significantly increased from productivity apps to communication apps (8.774 (95% CI, 1.034 to 16.514), p = 0.036), and from social apps to communication apps (2.820 (95% CI, 0.201 to 5.439), p = 0.048), but not from productivity apps to social apps (5.954 (95% CI, -2.128 to 14.036), p = 0.295) and not from tools to others. Similarly, a repeated-measures ANOVA with a Greenhouse-Geisser correction determined that mean KSPS differed statistically significantly between app types (F(1.970, 21.672) = 4.859, p = 0.018). Participants were fastest while using a communication app (3.964 $\pm$ 0.171), slowest while using a productivity app (3.179 $\pm$ 0.203), had 3.708 $\pm$ 0.172 KSPS in social apps, and had 3.765 $\pm$ 0.319 KSPS in tool apps. Post hoc analysis with a Bonferroni adjustment revealed that KSPS was statistically significantly increased from productivity apps to communication apps (0.784 (95% CI, 0.202 to 1.366), p = 0.012), and from social apps to communication apps (0.256 (95% CI, 0.043 to 0.469), p = 0.025), but not from productivity apps to social apps (0.198 (95% CI, -0.444 to 0.840), p = 1.000) and not from tools to others. On the other hand, the effect of using different app types on error rate was not statistically significant in terms of KSPC ($\chi^2(3)$ = 4.480, p = 0.214) and ER ($\chi^2(3)$ = 2.418, p = 0.490).

[1] https://play.google.com/store/apps/details?id=com.whatsapp&hl=en_US&gl=US, last access: 21.05.2022

4.3.7 Language Context

Before participating in our study, we asked about our participants' native language. The distribution of the participants' native languages is illustrated in Figure 3.2.e in Section 3.2.6. We also detected the language of the texts entered during the study (see Section 3.3.2.1 for details). Using participants' native language and the language of text they entered, we investigated the effect of using the native or a non-native language on the users' performance. Pairwise comparisons adjusted with Bonferroni showed statistically significant differences in terms of WPM ($t(31)=8.139$, $p<0.0001$), KSPS ($t(31)=7.641$, $p<0.0001$), KSPC ($Z = 67.0$, $p<0.01$), and ER ($Z = 19.0$, $p<0.0001$) between native and non-native language usage. The participants were faster while typing in their native languages (WPM: 43.005 ± 6.211, KSPS: 3.793 ± 0.551) than in a non-native language (WPM: 36.253 ± 7.704, KSPS: 3.241 ± 0.673). Moreover, the participants were more accurate in their native language (KSPC: 1.014 ± 0.029, ER: 0.277 ± 0.965) than in a non-native language (KSPC: 1.057 ± 0.071, ER: 1.565 ± 2.574).

4.3.8 Technical Context

We further investigated the effect of technical context on the participants' performance in smartphone brands, screen size, and keyboards used. One-way ANOVA tests showed that there were no statistically significant differences between the participants who used smartphones in different brands in terms of WPM ($F(7,40)= 0.740$, $p = 0.639$) and KSPS ($F(7,40)= 0.963$, $p = 0.471$). Similarly, Kruskal-Wallis H tests showed that the differences between smartphone brands in terms of KSPC ($\chi^2(7) = 10.853$, $p = 0.145$) and ER ($\chi^2(7) = 6.753$, $p = 0.455$) were not statistically significant.

We divided the participants into three based on the screen sizes of their smartphones: small, medium, and large screens. First, we calculated Q1 (5.5 inches) and Q3 (6.32 inches) based on the overall samples of screen sizes. Then, we classified the screen sizes smaller than Q1 as small screens, those larger than Q3 as large screens, and the others as medium screens. One-way ANOVA tests showed that there were no statistically significant differences between the participants who used smartphones in different screen sizes in terms of WPM (F(2,45)= 0.055, p = 0.947) and KSPS (F(2,45)= 0.061, p = 0.941). Similarly, Kruskal-Wallis H tests showed that the differences between screen sizes in terms of KSPC (χ^2(2) = 0.759, p = 0.684) and ER (χ^2(2) = 1.595, p = 0.450) were not statistically significant.

Our participants used four different types of soft keyboards during the study: Samsung, Microsoft SwiftKey, Gboard, and Fleksy keyboards (see Table D.1). We excluded the Fleksy keyboard from our statistical analysis since only one participant used this keyboard. One-way ANOVA tests showed that there were no statistically significant differences between different keyboard groups in terms of WPM (F(2,44)= 1.382, p = 0.262) and KSPS (F(2,44)= 1.686, p = 0.197). Similarly, Kruskal-Wallis H tests showed that the differences between three keyboard groups in terms of KSPC (χ^2(2) = 4.558, p = 0.102) and ER (χ^2(2) = 1.633, p = 0.442) were not statistically significant.

4.3.9 Demographics

Our statistical analysis could not find a main effect of demographic groups, including age, gender, education level, experience with a mobile device, experience with the current mobile device, daily screen time, and occupation on typing speed and error rate performance metrics.

4.3.10 Individual User Performances

To investigate the individual user performances, we repeated the same procedure in Section 4.1 on the dataset of each participant by using the same performance metrics. Figures E.1-E.5 in Appendix E illustrates the individual performance metrics for each participant under different contextual factors. Table 4.5 also illustrates the percent of the participants for each metric that have a higher value for each context factor. For example, 54.8% of participants had higher WPM in indoor conditions, while 45.2% had higher WPM in outdoor conditions. It is possible to observe individual differences in the effect of context. The effects of all context dimensions on each participant are available in our online repository (see “Online Repository” Sec-

Table 4.5: Percent of the participants that corresponding metric is higher for the context factor (Participants who had the same value under both conditions are excluded)

Context	Factor	WPM (%)	KSPS (%)	KSPC (%)	ER (%)
Environment	Indoor	54.8	48.4	16.1	6.5
	Outdoor	45.2	51.6	51.6	35.5
Mobility	Stable	53.3	46.7	26.7	6.7
	Mobile	46.7	53.3	46.7	33.3
Social	Alone	57.9	55.3	13.2	7.9
	Not Alone	42.1	44.7	42.1	23.7
Multitasking	Nothing	37.1	25.7	20.0	14.3
	Multitasking	62.9	74.3	40.0	28.6
Distractions	Nothing	45.7	42.9	11.4	11.4
	Distraction	54.3	57.1	42.9	37.1

tion on page 7)[2]. These results show that some participants' typing speed or error rate increase under certain context factors, while the same factor decreases the other participants' typing speed or error rate.

4.4 Conclusion

This chapter aimed to observe the effect of context on users' text entry performance in their daily settings and without any predefined task model. We calculated a set of performance metrics using this mechanism and associated them with the corresponding context labels. Finally, we investigated the effect of context on the participants' performance by using these metrics. Our findings show that contextual factors mainly affect participants' typing performance, and their effects on each participant are different.

[2] `https://iam.ncc.metu.edu.tr/cabas-individual-context-comparisons/`, last access: 21.01.2022

CHAPTER 5

MODELLING CONTEXT AND USER PERFORMANCE

The previous two chapters presented our user study and data collection mechanism. This chapter aims to implement a prediction mechanism based on the sensors and user performance. The chapter starts with a list of available sensors and widely used features in activity recognition studies. It continues with data cleaning and preprocessing mechanisms. Then, it investigates the sensors that can detect five context dimensions. Finally, it explains the mechanism to predict user performance problems due to the context.

Figure 5.1 illustrates the pipeline used for context recognition. The rest of the chapter details all steps in this pipeline.

5.1 Sensor Selection

This section presents the smartphone sensors collected during our user study. These sensors are available in the AWARE framework [233] and the data format presented for each sensor is the format supported by the AWARE Framework. Data collection process was explained in Section 3.2. The sensors are grouped by the category of data they measure, including motion, environment, position, and other sensors.

5.1.1 Motion Sensors

Motion sensors provide data about the smartphone's motion events caused by either user input or the environment. The motion sensors collected in this study include accelerometer, gravity, gyroscope, rotation, linear accelerometer, and significant mo-

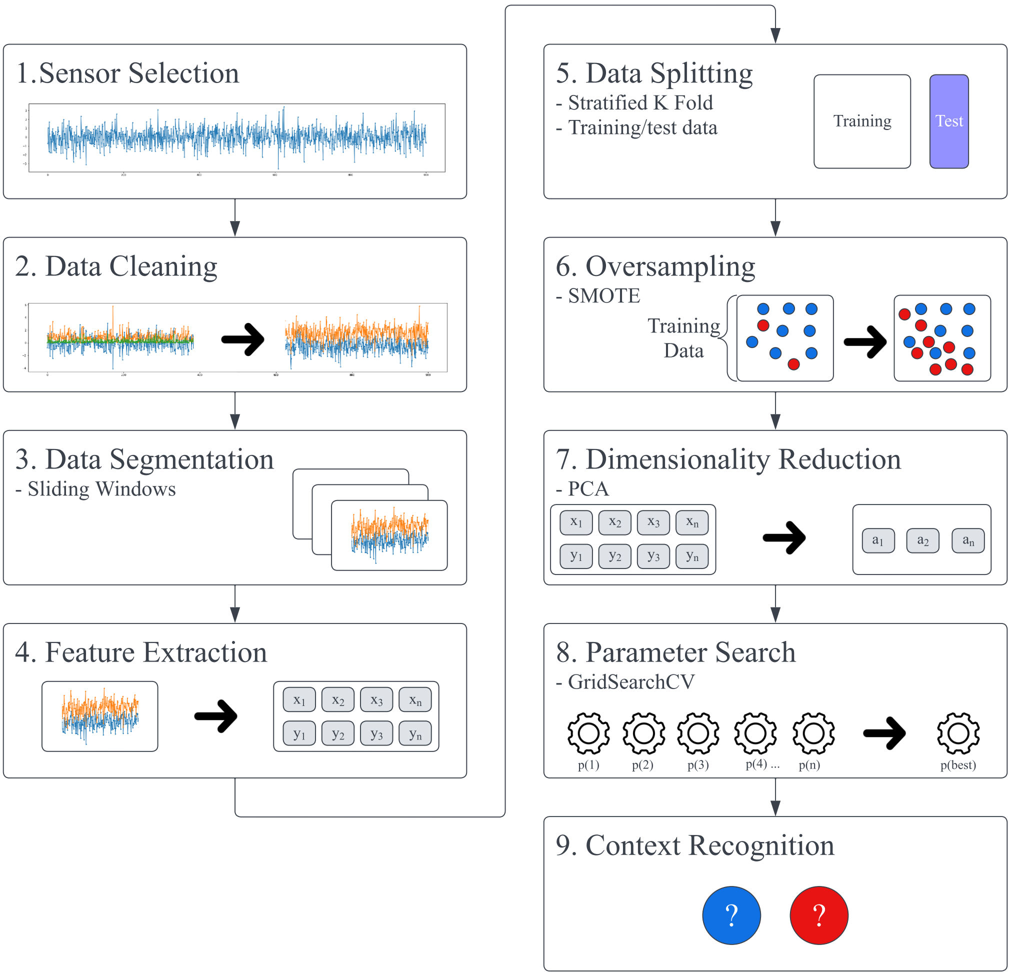


Figure 5.1: The context recognition pipeline

tion.

Listing 5.1 illustrates a sample motion sensor data. `double_values_0`, `double_values_1`, and `double_values_2` correspond to the values of X, Y, and Z axes, respectively. `accuracy` indicates the accuracy level of the recording whereas 3 means high level of accuracy and 0 means unreliable sensor measurements.

5.1.1.1 Accelerometer

The accelerometer sensor has been popular in the activity and context recognition domain that measures the physical acceleration in three axes. The measurement is

```
{
  "_id": "1164050",
  "timestamp": "1597662501397",
  "device_id": "4b11b28f-59b4-...",
  "double_values_0": "0.16641500592232",
  "double_values_1": "8.5006656646729",
  "double_values_2": "5.2661051750183",
  "accuracy": "3"
}
```


Listing 5.1: Motion sensor data example

relative to the free fall and is affected by the force of gravity [248]. The X-axis is horizontal, and the Y-axis is vertical relative to the phone's screen. The orientation changes do not swap the axis [249].

5.1.1.2 Gravity

The gravity sensor measures the gravity force applied to the smartphone corresponding to the direction and magnitude of gravity [249, 250]. It can be used to decide if the smartphone is moving [250]. The gravity force always applies to smartphones independent of orientation and mobility [251].

5.1.1.3 Gyroscope

The gyroscope sensor measures the rotation rate caused by the user exerting angular speed on the phone along three axis [249, 252]. These measurements can be used to detect the position or orientation [253].

5.1.1.4 Rotation

The rotation sensor is a synthetic sensor that measures the rotation of the global coordinate system to the device's coordinate system by using the accelerometer, the magnetometer, and the gyroscope sensors [249, 254]. It can be used to distinguish between walking and running [248], or monitoring orientation changes [249].

5.1.1.5 Linear Accelerometer

The linear accelerometer sensor is similar to the acceleration sensor, except that it excludes the force of gravity [249, 254].

5.1.1.6 Significant Motion

The significant motion sensor detects significant motion events that might cause the user's location to change, such as walking or sitting in a moving vehicle [255, 256]. Listing 5.2 illustrates a sample significant motion data. A significant motion is captured when `is_moving` value is 1.


```
{
    "_id": "2",
    "timestamp": "1595848255634",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "is_moving": "1"
}
```


Listing 5.2: Significant Motion data example

5.1.2 Environment Sensors

Android platform provides four sensors to sense environmental attributes: ambient temperature, humidity, light, and ambient air pressure. Although the AWARE Framework supports ambient temperature, none of the devices in our user study has this sensor. Moreover, the AWARE Framework does not provide integration with the ambient humidity sensor. Therefore, we only collected data for ambient air pressure (barometer) and light sensors.

5.1.2.1 Barometer

The barometer sensor measures ambient air pressure to predict weather changes [249] and determine the altitude when a GPS fix cannot be retrieved [254]. It can also be used for floor localization [257, 258]. Due to internal hardware smoothing, the barometer does not require a high sampling rate [259]. Sample barometer data is illustrated in Listing 5.3.


```
{
    "_id": "24",
    "timestamp": "1596455388308",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "double_values_0": "997.19091796875",
    "accuracy": "3"
}
```


Listing 5.3: Barometer data example, `double_values_0` value corresponds to the ambient air pressure in mbar/hPa units

5.1.2.2 Light

The light sensor is used to measure the ambient light. Commonly it is used to optimize screen brightness for environmental lighting conditions [260]. Although the light sensor can be helpful when determining where the user (indoor/outdoor) or smartphone (into a pocket or bag) is, additional information about the context is required [261]. The official light constants are as follows: [249]

- No moon: 0.001
- Full moon: 0.25
- Cloudy sky: 100.0
- Sunrise: 400.0
- Overcast: 10000.0

- Shade: 20000.0
- Sunlight: 110000.0
- Sunlight maximum: 120000.0

Sample light data is illustrated in Listing 5.4.


```
{
    "_id": "2",
    "timestamp": "1595848162029",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "double_light_lux": "579",
    "accuracy": "3"
}
```


Listing 5.4: Light data example, `double_light_lux` value corresponds to the ambient light in lux units

5.1.3 Position Sensors

The AWARE Framework supports two position sensors: magnetometer and proximity sensors.

5.1.3.1 Magnetometer

The magnetometer sensor measures the strength of the geomagnetic field around the device in three axis [248, 249]. It can be helpful when determining the device orientation [261]. The magnetometer data format is the same as motion sensors illustrated in Listing 5.1.

5.1.3.2 Proximity

The proximity sensor measures the distance between the smartphone and the object in front of it. It triggers an event when the object is closer than three cm [249, 261].

Sample proximity data is illustrated in Listing 5.5.


```
{
   "_id": "1",
   "timestamp": "1595848160998",
   "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
   "double_proximity": "5",
   "accuracy": "3"
}
```


Listing 5.5: Proximity data example

5.1.4 Other Sensors

Besides motion, environment, and position sensors, the AWARE Framework provides additional information that might be used to define the context. These data sources include currently used applications, battery status, incoming and outgoing calls, screen status, network and sim information, Wi-Fi connections, and GPS coordinates.

5.1.4.1 Applications

The applications sensor keeps track of currently used applications, including those running in the background [249]. Listing 5.6 illustrates a sample applications data. The framework adds a new record whenever a new application starts running (either in the foreground or background).

5.1.4.2 Battery

The battery sensor provides battery-related information such as battery level, charging, and discharging events [249]. Battery, charge and discharge data samples are illustrated in Listing 5.5.

```
{
  "_id": "1",
  "timestamp": "1595848606695",
  "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
  "package_name": "com.android.settings",
  "application_name": "Settings",
  "is_system_app": "1"
}
```

Listing 5.6: Applications data example

```
{
  "_id": "400",
  "timestamp": "1596001149614",
  "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
  "battery_status": "3",
  "battery_level": "55",
  "battery_scale": "100",
  "battery_voltage": "3820",
  "battery_temperature": "32",
  "battery_adaptor": "0",
  "battery_health": "2",
  "battery_technology": "Li-ion"
}
```

Listing 5.7: Battery data example

5.1.4.3 Calls

The communication sensor logs call events without any personal information [249]. Listing 5.8 illustrates a sample call data.


```
{
    "_id": "2",
    "timestamp": "1595863377177",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "call_type": "1",
    "call_duration": "597",
    "trace": "83f25dcc8d927ba24a1"
}
```


Listing 5.8: Call data example

5.1.4.4 Screen

The screen sensor logs the transitions between the screen on, screen off, screen locked, and screen unlocked states [249]. Listing 5.9 illustrates a sample screen data.


```
{
    "_id": "1177",
    "timestamp": "1596009727675",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "screen_status": "1"
}
```


Listing 5.9: Screen data example

5.1.4.5 Telephony

The telephony sensor provides network and sim information of the smartphone, as well as connected and neighboring cell towers [249]. Listing 5.10, Listing 5.11, and Listing 5.12 illustrate sample telephony data.

```
{
    "_id": "1514",
    "timestamp": "1596190348400",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "data_enabled": "2",
    "imei_meid_esn": "",
    "software_version": "02",
    "line_number": "b3357e446cbd95e1e...",
    "network_country_iso_mcc": "tr",
    "network_operator_code": "28601",
    "network_operator_name": "Turkcell",
    "network_type": "13",
    "phone_type": "1",
    "sim_state": "5",
    "sim_operator_code": "28601",
    "sim_operator_name": "Paycell | Turkcell",
    "sim_serial": "",
    "subscriber_id": ""
}
```

Listing 5.10: Telephony data example

```
{
    "_id": "1341",
    "timestamp": "1596224457369",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "cid": "10814012",
    "lac": "50901",
    "psc": "417",
    "signal_strength": "99",
    "bit_error_rate": "0"
}
```


Listing 5.11: GSM data example, `cid`, `lac`, and `psc` correspond to GSM tower's Cell ID, Location Area Code, and Primary Scrambling Code, respectively


```
{
    "_id": "39",
    "timestamp": "1596126307591",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "cid": "-1",
    "lac": "-1",
    "psc": "417",
    "signal_strength": "-89"
}
```


Listing 5.12: GSM neighbour data example, `cid`, `lac`, and `psc` correspond to GSM tower's Cell ID, Location Area Code, and Primary Scrambling Code, respectively

5.1.4.6 Wi-Fi

The Wi-Fi sensor logs the Wi-Fi scan results with detected devices [249]. The Wi-Fi sensor has been used to detect movement along with cellular signal strength [259]. Listing 5.13 illustrates a sample Wi-Fi data.


```
{
    "_id": "1265",
    "timestamp": "1595952010544",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "bssid": "04:d3:11:32:10:ab",
    "ssid": "Superbox_WiFi_1014",
    "security": "[WPA2-PSK-CCMP][ESS][WPS]",
    "frequency": "2417",
    "rssi": "-56",
    "label": ""
}
```


Listing 5.13: WiFi data example, `bssid`, `ssid`, and `rssi` correspond to detected device's MAC address, name, and RSSI dB, respectively

5.1.4.7 Locations

The locations sensor provides GPS data with the current coordinates of the user and the current speed [259]. Although GPS can be helpful to detect the user's activity, it has high power consumption and cannot cover indoor conditions [259].

5.2 Data Cleaning

We cleaned the sensor data in several ways:

- There were duplicate samples in all sensor datasets due to synchronization problems. The mobile app might have sent the data more than once, or the

```
{
    "_id": "27",
    "timestamp": "1595965218267",
    "device_id": "d6dfe3c6-bf0f-4375-b10b-bbe7a06045e3",
    "double_latitude": "38.2304098",
    "double_longitude": "26.33042396",
    "double_bearing": "149.10000610352",
    "double_speed": "0.18000000715256",
    "double_altitude": "144.74523925781",
    "provider": "gps",
    "accuracy": "17.152"
}
```


Listing 5.14: Locations data example

problem might have been related to the server. We grouped each sensor dataset by the corresponding participant and sorted it by timestamp in ascending order. We checked the timestamp values of consecutive rows and ignored the second one if their timestamp values were equal.

- Some participants' devices only had a few sensors; therefore, we excluded these participants from the dataset. Samsung SM-G610F (Galaxy J7 Prime, three participants) and Samsung SM-J710FQ (Galaxy J7, one participant) had only the accelerometer and proximity sensors among motion, position, and environment sensors. Similarly, Samsung SM-A710F (Galaxy A7, one participant) only had the accelerometer, light, magnetometer, and proximity sensors. Finally, Xiaomi Redmi 6 only had the accelerometer, gravity, light, magnetometer, and proximity sensors. Overall, we excluded six participants.
- Our data collection mechanism had two conditions to collect sensor data. First, it was designed to collect data only when the screen is on. Moreover, it synchronized the data only if participants entered text longer than five characters within the current session. However, if the data included rows when the screen was off or without text within the session, then we removed such samples from the dataset.

5.3 Data Segmentation

The data collected during our user study includes raw sensor recording in one or more axes based on the sensor type. Modeling on such kinds of raw sensors has challenges; therefore, the sensor data should be divided into portions of the same length for feature extraction. This process is called data segmentation, or windowing [262].

Data segmentation has been widely used in activity recognition studies, and different window sizes have been used in the literature. According to Ferrari et al. [263], a window consisting of two to five seconds of sensor data would be enough to recognize simple activities, such as sitting. The complex activities, on the other hand, required more oversized windows. Gao et al. [264] commented that even if longer windows produced more successful results than smaller windows, it depended on the classification methods. We used 2, 5, 10, and 20 seconds in our study as window sizes and compared their results.

Another issue to consider while segmenting the data is the overlap between windows. There are different approaches, including no overlap [265], 50% overlap [266], and overlap of one second [267]. Figo et al. [268] do not recommend using non-overlapping windows since it reduces the smoothness of the data.

5.4 Feature Extraction

Section 5.1 presented the set of sensors, and data for each sensor was collected in raw format. This raw data extracts critical features for exploring useful context information by being converted or transformed. This process is referred to as feature extraction [269]. Feature extraction helps to achieve more accurate classification results rather than using raw data [248, 270], and eliminates redundant features [248]. Moreover, feature extraction minimizes the noise in raw sensor data [248]. In this thesis work, we reviewed the literature to find commonly used features in the activity recognition domain. Table 5.1 presents these features with their short descriptions and formulas. We used tsfresh Python package [271] (version 0.19.0) to calculate these features.

5.5 Data Splitting

In order to cross-validate the models, we used the K-fold approach to split data into training and test sets. However, we wanted to keep the same distribution of labels among training and test sets. Therefore, in our experiments, we used the `StratifiedKFold` object in scikit-learn Python package [273] (version 1.1.1). This object extends `KFold` by preserving the distribution of the samples in each class [273]. We used five splits; therefore, 20% of the data was used as the test set in each fold.

5.6 Oversampling

Our user study collected data in participants' natural settings without a predefined task model. As a result, the study did not control the balance in the distribution of labels of different contextual factors. We used the synthetic minority over-sampling technique (SMOTE) [274] for oversampling to compare classification models on balanced data. We used imbalanced-learn Python package [275] (version 0.9.1) and only oversampled the training data in each fold. We compared the classification models with both imbalanced and oversampled data.

5.7 Dimensionality Reduction

Taking the sensors in Section 5.1 and features in Section 5.4 together, it is evident that we have a wide feature space. However, having high dimensionality increases the cost of modeling and classification. Moreover, it affects the training phase, which may reduce the accuracy [276]. For dimensionality reduction, we used the principal component analysis (PCA) [277] method.

Table 5.1: Features commonly used in the literature

Feature	Description	Formula
Mean	The average value of a sequence	$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
Median	The value that 50% of a sequence lies above [246]	$M = \left(\frac{\frac{n}{2} - cf}{f}\right)(w) + L_m$
Variance	Total sum of squared deviations from the mean of sequence divided by one minus the sequence size [246]	$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$
Standard Deviation	Square root value of the variance of a sequence [246]	$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2}$
Minimum	The smallest value in a sequence	$min = min_{j=1,...n}(x_j)$
Maximum	The largest value in a sequence	$max = max_{j=1,...n}(x_j)$
Range	Difference between max. and min. [246]	$range = max - min$
Skewness	Whether sequence is skewed to larger or smaller values [246].	$S = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \mu}{\sigma}\right]^3$
Kurtosis	Tendency of sequence to have extreme values [246]	$K = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \mu}{\sigma}\right]^4$
Mean absolute deviation	Mean of the absolute values of deviations from mean of the sequence [246]	$\frac{1}{n}\sum_{i=1}^{n}\lvert x_i - \mu\rvert$
Root mean square	Square root of normalized value of sum of squares of values in sequence	$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$
Total sum	Sum of values in a sequence	$sum = \sum_{i=1}^{n} x_i$
Zero crossing rate	Indexes where values change sign in a seq. [248]	$ZCR = \frac{1}{X-1}\sum_{x=1}^{X-1}$ $\{X_t X_{t-1} < 0\}$
Maximum latency	Index of max. value in a sequence [272]	$m_{x_{max}} = \{m \mid x_m = x_{max}\}$
Minimum latency	Index of min. value in a sequence [272]	$m_{x_{min}} = \{m \mid x_m = x_{min}\}$
First quartile	25^{th} percentile of a sequence [272]	$Q_1 = x_{\frac{n+1}{4}}$
Third quartile	75^{th} percentile of a sequence [272]	$Q_3 = x_{\frac{3(n+1)}{4}}$
Mean of signal gradient	Normalized mean of 1^{st} diff. of seq. [272]	$\bar{\bigtriangledown} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\lvert x_i - x_{i-1}\rvert}{x_{max}}\right)$
Mean of signal laplacian	Normalized mean of 2^{nd} diff. of seq. [272]	$\bar{\bigtriangleup} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\lvert x_{i+1} - 2x_i + x_{i-1}\rvert}{x_{max}}\right)$
Entropy	Rate of change in a sequence [272]	$H(x) = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i)$
Energy	Sum of squared values	$\sum_{i=1}^{n}\lvert x_i\rvert^2$

5.8 Parameter Search

In our pipeline, some components must be appropriately configured to produce better classification results. One of these components is the PCA. The number of dimensions that PCA reduces to is configurable. Moreover, the classification model K-Nearest Neighbors takes K-value as a parameter. The training set should be validated with different parameters and the parameters resulting in the best results should be used for the test set. Therefore, we used `GridSearchCV` object in scikit learn Python library to choose these parameters [273]. Given a classifier and a set of predefined parameters, `GridSearchCV` evaluates the model's performance with every parameter condition and results in the best parameters. Similar to oversampling, this process was only applied to the training data.

5.9 Context Classification Results

Context recognition has been a popular topic in the literature for the last two decades. These studies generally use different sets of sensors and features and compare the performances of several models with this data. These models can be listed as follows:

- K-Nearest Neighbors (KNN) [278]
- Decision Tree (DT) [278]
- Random Forest (RF) [265, 278–280]
- Multilayer Perceptron (MLP) [278, 280, 281]
- Support-vector Machine (SVM) [253, 282–285]
- AdaBoost [263]

This section aims to compare the performance of these models in recognizing five context dimensions, environment, mobility, social, multitasking, and distractions, and the sensor data collected in our user study. First, we investigated the relevant sensors in the literature for each context dimension. Then, using the pipeline presented in the previous sections, we investigated which contextual factors can be detected using smartphone sensors.

5.9.1 Environment

In our systematic review, we considered the environment as one of the physical context dimensions. Researchers attempted to evaluate the effect of the environment on users' performance in terms of being indoor/outdoor [8, 62], ambient noise [50], weather conditions [106], temperature [47], and vibration/noise level [123]. A considerable amount of literature has been published to detect indoor/outdoor conditions of the environment. These studies have used the following sensors:

- Accelerometer [248, 286–291]
- Wi-Fi [287, 289–294]
- Magnetometer [248, 286, 291, 292]
- Light [260, 291–293]
- Locations [289, 291, 293, 295]
- GSM [289, 292, 293]
- Barometer [290, 292, 296]
- Gyroscope [287, 288]
- Rotation [248]
- Microphone [291, 297]
- Bluetooth [287, 291]
- Ambient temperature [290, 293, 294]

Our user study did not collect microphone, Bluetooth, and temperature sensors; therefore, we ignored these sensors. However, as Sarsenbayeva et al. [49] suggested using battery temperature to predict ambient temperature, we used the battery sensor as well. In our user study, we asked participants to label their current environment as indoors, in a vehicle, outdoors, and crosswalk options. Then we grouped these labels as indoor and outdoor as presented in Table 4.1.

Table 5.2 presents the classification results by using different models and window sizes with 50% overlap for oversampled data. Figure 5.2 compares F1 scores. According to these results, the MLP classifier outperformed the other models and the random baseline. Moreover, the accuracy and F1-score values are close to each other for different window sizes.

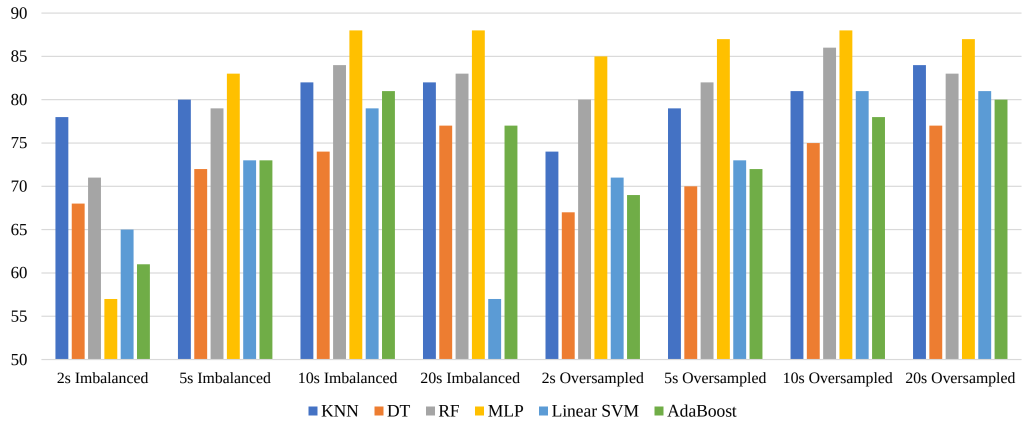


Figure 5.2: F1 Score comparison between different classification models and window sizes on environment context

Table 5.2: The results for environment context classification using different models and 50% percent overlapping window sizes

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
2s	KNN	0.78	0.66	0.71 ± 0.01	0.81 ± 0.01	0.74 ± 0.01	0.81 ± 0.01
	DT	2.64	0.05	0.66 ± 0.01	0.71 ± 0.01	0.67 ± 0.01	0.79 ± 0.01
	RF	13.79	0.12	0.85 ± 0.01	0.78 ± 0.02	0.80 ± 0.02	0.90 ± 0.01
	MLP	19.69	0.07	$\mathbf{0.85 \pm 0.02}$	$\mathbf{0.85 \pm 0.01}$	$\mathbf{0.85 \pm 0.01}$	$\mathbf{0.92 \pm 0.01}$
	Linear SVM	19.39	0.15	0.69 ± 0.00	0.79 ± 0.01	0.71 ± 0.01	0.79 ± 0.01
	AdaBoost	8.82	0.08	0.67 ± 0.02	0.74 ± 0.02	0.69 ± 0.02	0.78 ± 0.01
	Baseline	0.65	0.06	0.50 ± 0.01	0.51 ± 0.02	0.44 ± 0.01	0.51 ± 0.01
5s	KNN	0.56	0.25	0.77 ± 0.02	0.86 ± 0.02	0.79 ± 0.02	0.84 ± 0.02
	DT	1.02	0.03	0.69 ± 0.01	0.73 ± 0.01	0.70 ± 0.01	0.79 ± 0.01
	RF	7.30	0.06	0.85 ± 0.02	0.80 ± 0.02	0.82 ± 0.02	0.89 ± 0.01
	MLP	9.27	0.04	$\mathbf{0.87 \pm 0.01}$	$\mathbf{0.87 \pm 0.01}$	$\mathbf{0.87 \pm 0.01}$	$\mathbf{0.92 \pm 0.01}$
	Linear SVM	12.13	0.06	0.72 ± 0.01	0.79 ± 0.01	0.73 ± 0.01	0.80 ± 0.01
	AdaBoost	5.27	0.06	0.70 ± 0.02	0.76 ± 0.02	0.72 ± 0.02	0.79 ± 0.02
	Baseline	1.19	0.08	0.50 ± 0.01	0.50 ± 0.01	0.45 ± 0.01	0.50 ± 0.01
10s	KNN	0.16	0.04	0.79 ± 0.01	0.87 ± 0.01	0.81 ± 0.01	0.85 ± 0.01
	DT	0.31	0.01	0.73 ± 0.02	0.77 ± 0.03	0.75 ± 0.02	0.81 ± 0.01
	RF	1.77	0.03	0.88 ± 0.02	0.84 ± 0.03	0.86 ± 0.03	0.91 ± 0.01
	MLP	3.71	0.02	$\mathbf{0.89 \pm 0.03}$	$\mathbf{0.88 \pm 0.04}$	$\mathbf{0.88 \pm 0.03}$	$\mathbf{0.92 \pm 0.02}$
	Linear SVM	4.34	0.02	0.79 ± 0.01	0.84 ± 0.01	0.81 ± 0.01	0.85 ± 0.01
	AdaBoost	1.24	0.03	0.76 ± 0.03	0.81 ± 0.03	0.78 ± 0.03	0.83 ± 0.02
	Baseline	0.20	0.01	0.51 ± 0.02	0.51 ± 0.02	0.47 ± 0.02	0.51 ± 0.02
20s	KNN	0.10	0.02	0.82 ± 0.03	0.88 ± 0.02	0.84 ± 0.03	0.87 ± 0.03
	DT	0.14	0.01	0.77 ± 0.04	0.78 ± 0.04	0.77 ± 0.04	0.83 ± 0.03
	RF	0.75	0.02	0.88 ± 0.02	0.81 ± 0.03	0.83 ± 0.02	0.89 ± 0.01
	MLP	1.41	0.01	$\mathbf{0.88 \pm 0.05}$	$\mathbf{0.87 \pm 0.04}$	$\mathbf{0.87 \pm 0.04}$	$\mathbf{0.91 \pm 0.03}$
	Linear SVM	0.32	0.01	0.81 ± 0.02	0.83 ± 0.05	0.81 ± 0.03	0.86 ± 0.01
	AdaBoost	0.62	0.02	0.80 ± 0.05	0.81 ± 0.04	0.80 ± 0.05	0.85 ± 0.04
	Baseline	0.08	0.01	0.48 ± 0.03	0.47 ± 0.04	0.45 ± 0.04	0.49 ± 0.05

5.9.2 Mobility

Mobility is another physical context dimension in our systematic review. Activity recognition in terms of mobility has been a popular topic in the literature. Many researchers have attempted to detect mobility condition by using accelerometer and gyroscope sensors [253, 267, 282, 284, 285, 298–306]. The magnetometer sensor has also been widely used [284, 298, 300, 303–305, 307]. Other motion sensors such as gravity [307, 307], rotation [264], and linear acceleration [307] were also used to detect mobility condition. In this thesis work, we used all motion sensors.

Our user study asked participants to label their current mobility conditions as lying down, sitting, standing, walking, and running. Then we grouped these labels as stable and mobile as presented in Table 4.1.

Table 5.3 presents the classification results by using different models and window sizes with 50% overlap for oversampled data. Figure 5.3 compares F1 scores. According to these results, the MLP classifier outperformed the other models and the random baseline for windows of 5, 10, and 20 seconds. The MLP and Random Forest classifiers have relative values for windows of 2 seconds. While the accuracy and F1 score values are close to each other for windows of 5, 10, and 20 seconds, these values decrease for windows of 2 seconds. This outcome was expected as the literature suggested more oversized windows to capture complex activities.

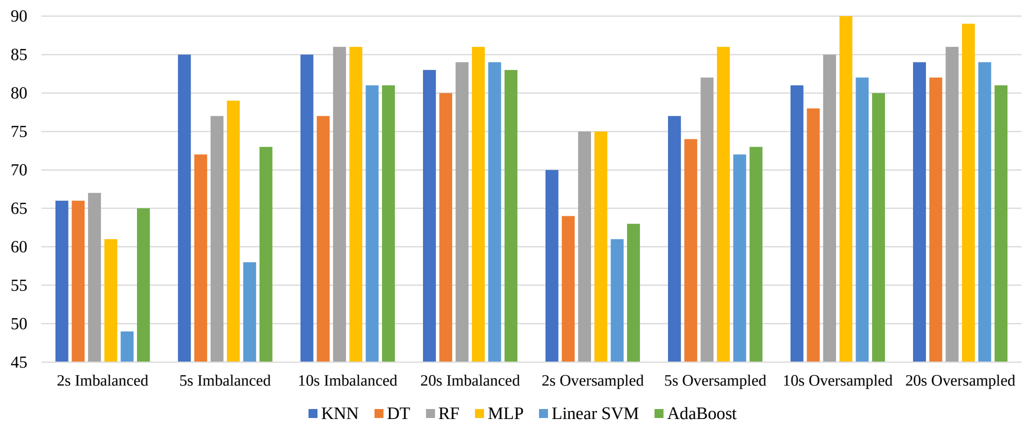


Figure 5.3: F1 Score comparison between different classification models and window sizes on mobility context

Table 5.3: The results for mobility context classification using different models and 50% percent overlapping window sizes

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Score	Accuracy
2s	KNN	1.39	0.72	0.65 ± 0.01	0.86 ± 0.01	0.70 ± 0.01	0.90 ± 0.01
	DT	2.12	0.06	0.61 ± 0.01	0.73 ± 0.01	0.64 ± 0.01	0.89 ± 0.01
	RF	20.72	0.11	$\mathbf{0.79 \pm 0.02}$	$\mathbf{0.72 \pm 0.02}$	$\mathbf{0.75 \pm 0.01}$	$\mathbf{0.96 \pm 0.00}$
	MLP	12.25	0.07	0.76 ± 0.02	0.75 ± 0.03	0.75 ± 0.03	0.95 ± 0.00
	Linear SVM	30.93	0.12	0.59 ± 0.01	0.79 ± 0.02	0.61 ± 0.01	0.84 ± 0.01
	AdaBoost	12.92	0.10	0.60 ± 0.00	0.77 ± 0.02	0.63 ± 0.01	0.86 ± 0.01
	Baseline	1.36	0.07	0.50 ± 0.00	0.50 ± 0.02	0.38 ± 0.01	0.50 ± 0.01
5s	KNN	0.67	0.22	0.72 ± 0.02	0.90 ± 0.01	0.77 ± 0.02	0.91 ± 0.01
	DT	0.87	0.04	0.70 ± 0.02	0.81 ± 0.01	0.74 ± 0.01	0.91 ± 0.01
	RF	9.32	0.06	0.83 ± 0.02	0.81 ± 0.03	0.82 ± 0.01	0.95 ± 0.00
	MLP	7.55	0.04	$\mathbf{0.85 \pm 0.03}$	$\mathbf{0.86 \pm 0.02}$	$\mathbf{0.86 \pm 0.02}$	$\mathbf{0.96 \pm 0.01}$
	Linear SVM	12.81	0.05	0.68 ± 0.01	0.87 ± 0.02	0.72 ± 0.02	0.88 ± 0.01
	AdaBoost	7.60	0.05	0.69 ± 0.01	0.83 ± 0.03	0.73 ± 0.02	0.90 ± 0.01
	Baseline	0.69	0.04	0.50 ± 0.01	0.50 ± 0.04	0.39 ± 0.01	0.51 ± 0.01
10s	KNN	0.20	0.05	0.76 ± 0.02	0.91 ± 0.01	0.81 ± 0.02	0.91 ± 0.01
	DT	0.29	0.02	0.74 ± 0.04	0.83 ± 0.04	0.78 ± 0.04	0.91 ± 0.02
	RF	1.41	0.03	0.84 ± 0.01	0.87 ± 0.05	0.85 ± 0.02	0.95 ± 0.01
	MLP	2.73	0.02	$\mathbf{0.90 \pm 0.04}$	$\mathbf{0.90 \pm 0.02}$	$\mathbf{0.90 \pm 0.03}$	$\mathbf{0.96 \pm 0.01}$
	Linear SVM	3.75	0.02	0.78 ± 0.03	0.88 ± 0.01	0.82 ± 0.03	0.93 ± 0.01
	AdaBoost	1.96	0.03	0.77 ± 0.02	0.83 ± 0.04	0.80 ± 0.03	0.92 ± 0.01
	Baseline	0.23	0.02	0.50 ± 0.01	0.51 ± 0.03	0.41 ± 0.02	0.51 ± 0.03
20s	KNN	0.08	0.02	0.80 ± 0.05	0.94 ± 0.02	0.84 ± 0.05	0.92 ± 0.03
	DT	0.10	0.01	0.79 ± 0.04	0.86 ± 0.06	0.82 ± 0.04	0.92 ± 0.02
	RF	0.51	0.02	0.90 ± 0.04	0.83 ± 0.05	0.86 ± 0.04	0.95 ± 0.01
	MLP	1.22	0.01	$\mathbf{0.88 \pm 0.03}$	$\mathbf{0.91 \pm 0.05}$	$\mathbf{0.89 \pm 0.03}$	$\mathbf{0.96 \pm 0.01}$
	Linear SVM	0.18	0.01	0.80 ± 0.05	0.90 ± 0.07	0.84 ± 0.06	0.93 ± 0.02
	AdaBoost	0.56	0.02	0.80 ± 0.09	0.83 ± 0.07	0.81 ± 0.07	0.92 ± 0.03
	Baseline	0.13	0.01	0.50 ± 0.04	0.50 ± 0.10	0.41 ± 0.05	0.51 ± 0.05

5.9.3 Social Context

A relatively small body of literature is concerned with detecting social context. Exler et al. [278] used accelerometer, gyroscope, and location sensors. Miluzzo et al. Chen et al. [308] used accelerometer, location, Bluetooth, and microphone. While [309] used only Bluetooth, Adams et al. [310] combined Bluetooth with location data. In this thesis work, we used the accelerometer, gyroscope, and location sensors to detect the social context. We also included the light and Wi-Fi sensors based on our experiments.

Our user study asked participants to label their current social context as alone, with 2-4 friends/family members/colleagues, with a friend/family member/colleague, with more than 4 friends/family members/colleagues, with strangers (crowded), with strangers (not crowded). Then we grouped these labels as alone and not alone as presented in Table 4.1.

Table 5.4 presents the classification results by using different models and window sizes with 50% overlap for oversampled data. Figure 5.4 compares F1 scores. According to these results, the MLP and Random Forest classifiers outperformed the other models and the random baseline for windows of 10 and 20 seconds, while only Random Forest has the most successful results for windows of 2 and 5 seconds. Both accuracy and F1 score values increase as the window size gets larger.

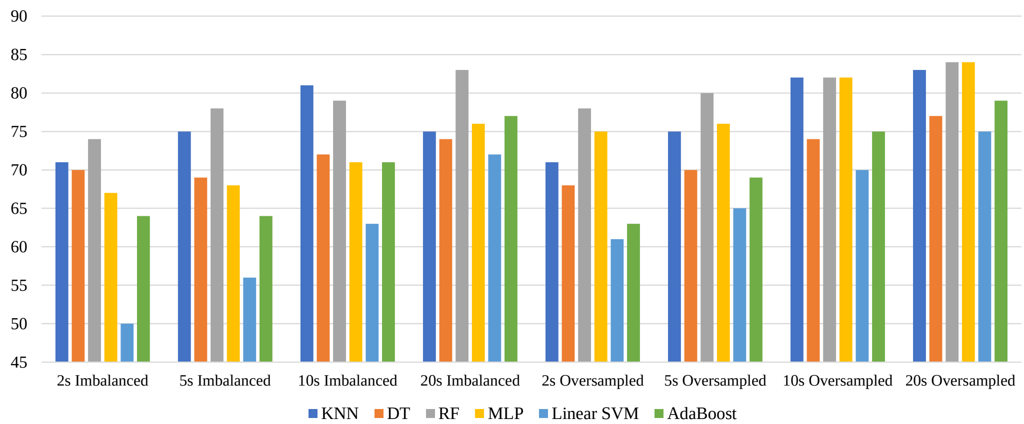


Figure 5.4: F1 Score comparison between different classification models and window sizes on social context

Table 5.4: The results for social context classification using different models and 50% percent overlapping window sizes

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
2s	KNN	0.94	0.22	0.71 ± 0.01	0.72 ± 0.01	0.71 ± 0.01	0.73 ± 0.01
	DT	2.22	0.06	0.67 ± 0.01	0.68 ± 0.01	0.68 ± 0.01	0.70 ± 0.01
	RF	11.64	0.11	$\mathbf{0.80 \pm 0.01}$	$\mathbf{0.77 \pm 0.01}$	$\mathbf{0.78 \pm 0.01}$	$\mathbf{0.82 \pm 0.01}$
	MLP	15.74	0.06	0.74 ± 0.01	0.76 ± 0.01	0.75 ± 0.01	0.77 ± 0.01
	Linear SVM	16.73	0.13	0.61 ± 0.01	0.61 ± 0.01	0.61 ± 0.01	0.65 ± 0.02
	AdaBoost	3.97	0.09	0.63 ± 0.03	0.65 ± 0.03	0.63 ± 0.03	0.65 ± 0.03
	Baseline	1.07	0.07	0.50 ± 0.01	0.50 ± 0.01	0.49 ± 0.01	0.50 ± 0.01
5s	KNN	0.64	0.14	0.74 ± 0.04	0.76 ± 0.04	0.75 ± 0.04	0.77 ± 0.04
	DT	0.88	0.04	0.70 ± 0.02	0.71 ± 0.03	0.70 ± 0.03	0.73 ± 0.02
	RF	4.85	0.08	$\mathbf{0.81 \pm 0.01}$	$\mathbf{0.79 \pm 0.01}$	$\mathbf{0.80 \pm 0.01}$	$\mathbf{0.83 \pm 0.01}$
	MLP	8.97	0.04	0.76 ± 0.06	0.77 ± 0.05	0.76 ± 0.05	0.79 ± 0.05
	Linear SVM	7.13	0.05	0.65 ± 0.04	0.65 ± 0.02	0.65 ± 0.03	0.68 ± 0.04
	AdaBoost	4.05	0.05	0.69 ± 0.01	0.70 ± 0.01	0.69 ± 0.01	0.71 ± 0.01
	Baseline	0.72	0.05	0.49 ± 0.01	0.49 ± 0.01	0.48 ± 0.01	0.49 ± 0.01
10s	KNN	0.20	0.04	0.81 ± 0.02	0.84 ± 0.02	0.82 ± 0.02	0.84 ± 0.02
	DT	0.29	0.02	0.73 ± 0.04	0.75 ± 0.04	0.74 ± 0.04	0.76 ± 0.04
	RF	1.64	0.03	$\mathbf{0.84 \pm 0.01}$	$\mathbf{0.81 \pm 0.02}$	$\mathbf{0.82 \pm 0.02}$	$\mathbf{0.85 \pm 0.01}$
	MLP	3.28	0.02	$\mathbf{0.83 \pm 0.01}$	$\mathbf{0.82 \pm 0.01}$	$\mathbf{0.82 \pm 0.00}$	$\mathbf{0.85 \pm 0.01}$
	Linear SVM	2.18	0.02	0.72 ± 0.05	0.70 ± 0.02	0.70 ± 0.03	0.74 ± 0.04
	AdaBoost	1.33	0.03	0.75 ± 0.02	0.75 ± 0.02	0.75 ± 0.02	0.78 ± 0.01
	Baseline	0.24	0.02	0.51 ± 0.02	0.51 ± 0.02	0.50 ± 0.02	0.51 ± 0.03
20s	KNN	0.10	0.02	0.83 ± 0.05	0.84 ± 0.06	0.83 ± 0.05	0.86 ± 0.04
	DT	0.17	0.01	0.77 ± 0.06	0.78 ± 0.05	0.77 ± 0.06	0.81 ± 0.05
	RF	0.56	0.02	$\mathbf{0.87 \pm 0.02}$	$\mathbf{0.82 \pm 0.05}$	$\mathbf{0.84 \pm 0.04}$	$\mathbf{0.87 \pm 0.02}$
	MLP	1.52	0.01	$\mathbf{0.85 \pm 0.05}$	$\mathbf{0.83 \pm 0.06}$	$\mathbf{0.84 \pm 0.05}$	$\mathbf{0.87 \pm 0.04}$
	Linear SVM	0.25	0.01	0.77 ± 0.05	0.75 ± 0.04	0.75 ± 0.03	0.79 ± 0.04
	AdaBoost	0.63	0.02	0.80 ± 0.03	0.79 ± 0.01	0.79 ± 0.02	0.83 ± 0.03
	Baseline	0.12	0.01	0.53 ± 0.04	0.53 ± 0.05	0.50 ± 0.05	0.53 ± 0.05

5.9.4 Multitasking

Several studies aim to detect the parallel task while using a smartphone. These studies have used the accelerometer [283, 299, 302, 311–313], gyroscope [299, 302, 312], locations [299, 312], Wi-Fi [299, 312], magnetometer [302], and temperature [302].

Our user study asked participants to label their current multitasking conditions as nothing, carrying a box/bag/other, doing home-activities (cleaning, cooking, etc), having a conversation with someone around me, having breakfast/lunch/dinner, shopping, trying to avoid collision while walking, working, and multiple of these. Then we grouped these labels as nothing and multitasking as presented in Table 4.1.

Table 5.5 presents the classification results by using different models and window sizes with 50% overlap for oversampled data. Figure 5.5 compares F1 scores. According to these results, the MLP classifier outperformed the other models and the random baseline for windows of 2, 5, and 20 seconds. The Random Forest classifier outperformed the other models and the random baseline for windows of 10 seconds. Both accuracy and F1 score values increase as the window size gets larger.

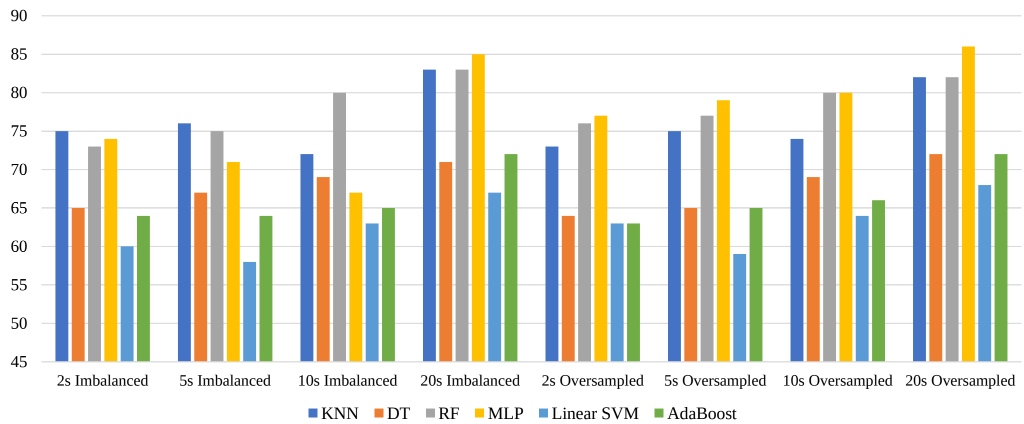


Figure 5.5: F1 Score comparison between different classification models and window sizes on multitasking context

Table 5.5: The results for multitasking context classification using different models and 50% percent overlapping window sizes

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Score	Accuracy
2s	KNN	0.81	0.40	0.73 ± 0.01	0.75 ± 0.01	0.73 ± 0.01	0.74 ± 0.01
	DT	1.75	0.05	0.64 ± 0.01	0.65 ± 0.01	0.64 ± 0.01	0.65 ± 0.01
	RF	10.65	0.11	0.78 ± 0.01	0.75 ± 0.01	0.76 ± 0.01	0.78 ± 0.01
	MLP	19.90	0.07	$\mathbf{0.77 \pm 0.02}$	$\mathbf{0.77 \pm 0.02}$	$\mathbf{0.77 \pm 0.02}$	$\mathbf{0.78 \pm 0.02}$
	Linear SVM	16.66	0.15	0.63 ± 0.01	0.63 ± 0.01	0.63 ± 0.01	0.64 ± 0.01
	AdaBoost	7.44	0.09	0.63 ± 0.01	0.64 ± 0.01	0.63 ± 0.01	0.64 ± 0.01
	Baseline	0.92	0.08	0.50 ± 0.01	0.50 ± 0.01	0.50 ± 0.01	0.50 ± 0.01
5s	KNN	0.61	0.25	0.75 ± 0.02	0.76 ± 0.02	0.75 ± 0.02	0.75 ± 0.02
	DT	0.94	0.03	0.65 ± 0.03	0.66 ± 0.03	0.65 ± 0.03	0.66 ± 0.03
	RF	4.66	0.07	0.79 ± 0.01	0.77 ± 0.01	0.77 ± 0.01	0.78 ± 0.01
	MLP	9.53	0.04	$\mathbf{0.79 \pm 0.02}$	$\mathbf{0.79 \pm 0.02}$	$\mathbf{0.79 \pm 0.02}$	$\mathbf{0.80 \pm 0.02}$
	Linear SVM	5.77	0.06	0.62 ± 0.02	0.59 ± 0.02	0.59 ± 0.02	0.63 ± 0.02
	AdaBoost	3.26	0.06	0.65 ± 0.01	0.65 ± 0.01	0.65 ± 0.01	0.65 ± 0.01
	Baseline	0.51	0.04	0.50 ± 0.01	0.50 ± 0.01	0.50 ± 0.01	0.50 ± 0.01
10s	KNN	0.14	0.04	0.74 ± 0.03	0.74 ± 0.03	0.74 ± 0.03	0.74 ± 0.02
	DT	0.20	0.02	0.69 ± 0.01	0.69 ± 0.01	0.69 ± 0.01	0.69 ± 0.01
	RF	1.28	0.03	$\mathbf{0.81 \pm 0.04}$	$\mathbf{0.79 \pm 0.03}$	$\mathbf{0.80 \pm 0.03}$	$\mathbf{0.81 \pm 0.03}$
	MLP	2.91	0.02	0.80 ± 0.06	0.79 ± 0.06	0.80 ± 0.06	0.80 ± 0.06
	Linear SVM	0.92	0.02	0.66 ± 0.03	0.64 ± 0.03	0.64 ± 0.03	0.66 ± 0.03
	AdaBoost	0.50	0.02	0.66 ± 0.01	0.66 ± 0.01	0.66 ± 0.01	0.66 ± 0.02
	Baseline	0.17	0.02	0.47 ± 0.02	0.47 ± 0.02	0.47 ± 0.02	0.47 ± 0.02
20s	KNN	0.06	0.02	0.82 ± 0.03	0.82 ± 0.03	0.82 ± 0.03	0.82 ± 0.03
	DT	0.08	0.01	0.73 ± 0.05	0.73 ± 0.05	0.72 ± 0.05	0.73 ± 0.05
	RF	0.44	0.02	0.83 ± 0.04	0.82 ± 0.04	0.82 ± 0.04	0.83 ± 0.04
	MLP	1.16	0.01	$\mathbf{0.86 \pm 0.02}$	$\mathbf{0.86 \pm 0.02}$	$\mathbf{0.86 \pm 0.02}$	$\mathbf{0.86 \pm 0.02}$
	Linear SVM	0.15	0.01	0.70 ± 0.04	0.68 ± 0.04	0.68 ± 0.04	0.70 ± 0.04
	AdaBoost	0.28	0.02	0.72 ± 0.05	0.72 ± 0.05	0.72 ± 0.05	0.72 ± 0.05
	Baseline	0.08	0.01	0.48 ± 0.05	0.48 ± 0.06	0.48 ± 0.05	0.48 ± 0.05

5.9.5 Distractions

The literature on detecting distraction factors around smartphone users using sensors is limited. We related the studies that aimed to detect stress factors with this context. Can et al. [280] used the accelerometer and temperature sensors. Sano and Picard [314] used the accelerometer sensor, location sensor, and calls. Finally, Tigwell et al. [315] used the calls, locations, and microphone.

Our user study asked participants to label their current distraction conditions as nothing, being in a hurry, being interrupted by someone, being interrupted by something unexpected, needing to check something from time to time, obstacles/people/cars on walking path, and multiple of these. Then we grouped these labels as nothing and distraction as presented in Table 4.1.

Table 5.6 presents the classification results by using different models and window sizes with 50% overlap for oversampled data. Figure 5.6 compares F1 scores. According to these results, the MLP classifier outperformed the other models and the random baseline for windows of 5, 10, and 20 seconds. The Random Forest classifier outperformed the other models and the random baseline for windows of 5 seconds. Both accuracy and F1 score values increase as the window size gets larger.

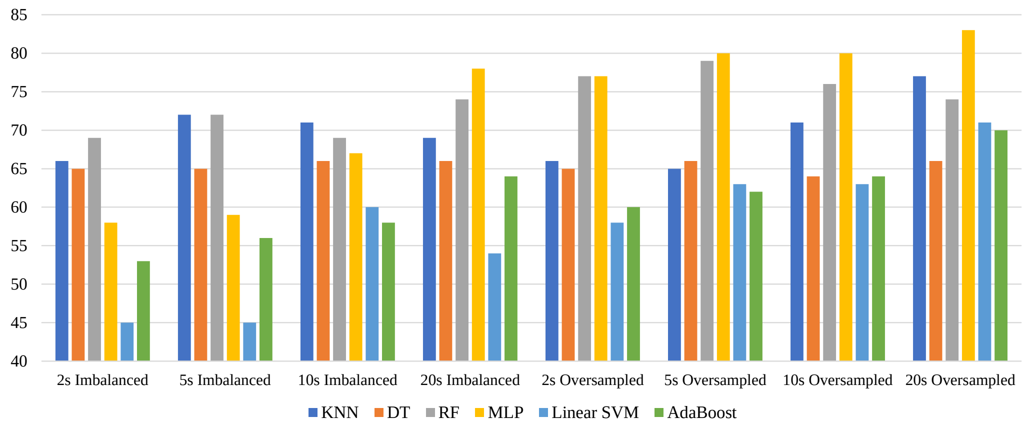


Figure 5.6: F1 Score comparison between different classification models and window sizes on distraction context

Table 5.6: The results for distraction context classification using different models and 50% percent overlapping window sizes

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
2s	KNN	0.78	0.24	0.65 ± 0.02	0.72 ± 0.02	0.66 ± 0.02	0.73 ± 0.02
	DT	2.88	0.06	0.64 ± 0.02	0.67 ± 0.02	0.65 ± 0.02	0.74 ± 0.02
	RF	15.46	0.11	$\mathbf{0.83 \pm 0.02}$	$\mathbf{0.73 \pm 0.02}$	$\mathbf{0.77 \pm 0.02}$	$\mathbf{0.87 \pm 0.01}$
	MLP	26.84	0.07	0.77 ± 0.02	0.77 ± 0.01	0.77 ± 0.01	0.86 ± 0.01
	Linear SVM	34.52	0.19	0.59 ± 0.01	0.63 ± 0.01	0.58 ± 0.01	0.65 ± 0.01
	AdaBoost	8.94	0.10	0.60 ± 0.01	0.65 ± 0.01	0.60 ± 0.02	0.67 ± 0.02
	Baseline	0.76	0.07	0.50 ± 0.01	0.50 ± 0.01	0.45 ± 0.01	0.50 ± 0.01
5s	KNN	0.55	0.15	0.65 ± 0.03	0.71 ± 0.04	0.65 ± 0.03	0.73 ± 0.03
	DT	1.53	0.04	0.65 ± 0.01	0.69 ± 0.01	0.66 ± 0.01	0.75 ± 0.01
	RF	7.37	0.07	0.85 ± 0.01	0.76 ± 0.02	0.79 ± 0.01	0.89 ± 0.00
	MLP	14.66	0.04	$\mathbf{0.80 \pm 0.02}$	$\mathbf{0.79 \pm 0.01}$	$\mathbf{0.80 \pm 0.01}$	$\mathbf{0.88 \pm 0.01}$
	Linear SVM	14.70	0.08	0.63 ± 0.01	0.69 ± 0.01	0.63 ± 0.01	0.70 ± 0.01
	AdaBoost	6.09	0.06	0.62 ± 0.01	0.67 ± 0.01	0.62 ± 0.01	0.70 ± 0.01
	Baseline	0.60	0.05	0.50 ± 0.01	0.50 ± 0.02	0.45 ± 0.01	0.51 ± 0.01
10s	KNN	0.23	0.04	0.70 ± 0.03	0.79 ± 0.04	0.71 ± 0.03	0.78 ± 0.03
	DT	0.42	0.02	0.63 ± 0.03	0.67 ± 0.04	0.64 ± 0.03	0.74 ± 0.03
	RF	2.05	0.03	0.82 ± 0.02	0.73 ± 0.04	0.76 ± 0.04	0.88 ± 0.01
	MLP	4.67	0.02	$\mathbf{0.81 \pm 0.02}$	$\mathbf{0.80 \pm 0.03}$	$\mathbf{0.80 \pm 0.02}$	$\mathbf{0.88 \pm 0.01}$
	Linear SVM	3.24	0.02	0.63 ± 0.01	0.70 ± 0.02	0.63 ± 0.02	0.71 ± 0.02
	AdaBoost	1.83	0.03	0.63 ± 0.03	0.67 ± 0.04	0.64 ± 0.03	0.74 ± 0.01
	Baseline	0.21	0.02	0.50 ± 0.01	0.50 ± 0.02	0.45 ± 0.01	0.50 ± 0.01
20s	KNN	0.07	0.02	0.75 ± 0.04	0.82 ± 0.04	0.77 ± 0.04	0.85 ± 0.03
	DT	0.12	0.01	0.65 ± 0.02	0.67 ± 0.04	0.66 ± 0.03	0.79 ± 0.01
	RF	0.71	0.02	0.88 ± 0.06	0.69 ± 0.06	0.74 ± 0.06	0.89 ± 0.02
	MLP	1.48	0.01	$\mathbf{0.85 \pm 0.03}$	$\mathbf{0.83 \pm 0.04}$	$\mathbf{0.83 \pm 0.03}$	$\mathbf{0.91 \pm 0.02}$
	Linear SVM	0.24	0.01	0.69 ± 0.04	0.74 ± 0.05	0.71 ± 0.04	0.81 ± 0.03
	AdaBoost	0.67	0.02	0.71 ± 0.06	0.70 ± 0.05	0.70 ± 0.05	0.83 ± 0.02
	Baseline	0.07	0.01	0.51 ± 0.04	0.53 ± 0.07	0.44 ± 0.05	0.49 ± 0.04

5.9.6 Summary and Discussion

In this section, we compared the performances of different classification models in recognizing five context dimensions. In general, MLP was the most successful model. Moreover, larger window sizes resulted in better results, as Ferrari et al. [263] suggested. We achieved the best classification results in the environment (0.92 accuracy and 0.88 F1-score) and mobility (0.96 accuracy and 0.90 F1-score) contexts. These two contexts have relatively a larger body of literature, and many different approaches have been tried to recognize these contexts. Available motion, environment, and position sensors have been widely used in relevant studies.

On the other hand, social context, multitasking, and distraction had relatively more minor interest in the literature. This minority might be due to the challenges of detecting these contexts with available sensors. Further studies should be conducted to investigate these contexts.

5.10 Predicting Errors

The previous section investigated the sensors that detect five context dimensions: environment, mobility, social, multitasking, and distractions. This section continues this investigation by predicting participants' errors using these available sensors.

In Chapter 4, we showed that context affects each individual differently. While it may positively impact a participant's performance, it may cause performance problems for another participant. The interaction between a user and his/her smartphone is unique and should be considered accordingly [316]. We focus on ability-based adaptations in this study; therefore, we aim to identify and predict the users' abilities. Moreover, the number of samples for each participant has a wide range (see Figure 4.1). As a result, we decided to apply regression models to each participant separately.

We applied the same pipeline in Figure 5.1 except that we used regressor implementation of each model instead of classifiers [273]. Moreover, we calculated the mean squared errors instead of accuracy, precision, recall, and F1 score for the regression tasks. We could not apply deep learning methods such as RNN-LSTM due to the in-

Table 5.7: The regression results for error detection (mean squared error)

P	Sample Size	KNN	RF	SVR	AB	Baseline
P03	99	0.26 ± 0.06	0.28 ± 0.04	$\mathbf{0.25 \pm 0.05}$	0.31 ± 0.08	0.36 ± 0.02
P05	928	0.23 ± 0.02	$\mathbf{0.21 \pm 0.01}$	$\mathbf{0.21 \pm 0.01}$	0.25 ± 0.00	0.64 ± 0.00
P08	76	0.25 ± 0.08	0.29 ± 0.08	$\mathbf{0.23 \pm 0.07}$	0.35 ± 0.10	0.55 ± 0.03
P09	1006	0.25 ± 0.01	$\mathbf{0.24 \pm 0.01}$	0.32 ± 0.04	0.25 ± 0.00	0.39 ± 0.00
P12	82	0.26 ± 0.09	$\mathbf{0.23 \pm 0.04}$	$\mathbf{0.23 \pm 0.06}$	0.24 ± 0.05	0.29 ± 0.03
P14	322	0.26 ± 0.03	$\mathbf{0.23 \pm 0.02}$	0.39 ± 0.34	0.24 ± 0.00	0.44 ± 0.01
P15	3332	0.24 ± 0.01	$\mathbf{0.23 \pm 0.00}$	0.32 ± 0.06	0.25 ± 0.00	0.59 ± 0.00
P16	673	0.26 ± 0.02	$\mathbf{0.25 \pm 0.01}$	0.28 ± 0.03	$\mathbf{0.25 \pm 0.00}$	0.55 ± 0.00
P17	360	0.26 ± 0.02	$\mathbf{0.21 \pm 0.02}$	0.22 ± 0.02	0.24 ± 0.01	0.29 ± 0.01
P19	281	0.26 ± 0.02	$\mathbf{0.23 \pm 0.01}$	0.40 ± 0.33	0.25 ± 0.01	0.52 ± 0.00
P22	109	$\mathbf{0.28 \pm 0.06}$	0.29 ± 0.10	0.29 ± 0.07	0.29 ± 0.08	0.39 ± 0.02
P23	120	0.25 ± 0.05	0.24 ± 0.07	0.30 ± 0.13	$\mathbf{0.21 \pm 0.03}$	0.75 ± 0.00
P25	677	0.28 ± 0.01	$\mathbf{0.23 \pm 0.01}$	0.26 ± 0.04	0.26 ± 0.01	0.36 ± 0.00
P26	98	0.27 ± 0.02	0.25 ± 0.04	$\mathbf{0.22 \pm 0.05}$	0.23 ± 0.05	0.51 ± 0.01
P27	817	0.26 ± 0.02	$\mathbf{0.21 \pm 0.00}$	0.25 ± 0.02	0.25 ± 0.00	0.35 ± 0.00
P28	219	0.26 ± 0.02	$\mathbf{0.19 \pm 0.02}$	0.21 ± 0.04	0.22 ± 0.02	0.26 ± 0.01
P29	313	0.25 ± 0.03	$\mathbf{0.23 \pm 0.02}$	0.26 ± 0.02	0.24 ± 0.02	0.43 ± 0.01
P31	1198	0.26 ± 0.02	$\mathbf{0.24 \pm 0.02}$	0.25 ± 0.02	0.25 ± 0.01	0.54 ± 0.00
P32	199	0.24 ± 0.04	$\mathbf{0.23 \pm 0.02}$	$\mathbf{0.23 \pm 0.09}$	$\mathbf{0.23 \pm 0.01}$	0.53 ± 0.01
P33	110	$\mathbf{0.20 \pm 0.06}$	0.26 ± 0.04	0.24 ± 0.04	0.25 ± 0.03	0.39 ± 0.02
P35	1104	0.25 ± 0.01	$\mathbf{0.23 \pm 0.01}$	0.25 ± 0.02	0.25 ± 0.00	0.38 ± 0.00
P38	154	0.26 ± 0.06	$\mathbf{0.21 \pm 0.02}$	0.54 ± 0.42	0.23 ± 0.02	0.35 ± 0.01
P43	299	0.26 ± 0.03	$\mathbf{0.19 \pm 0.02}$	$\mathbf{0.19 \pm 0.02}$	0.23 ± 0.01	0.28 ± 0.01
P45	505	0.27 ± 0.04	$\mathbf{0.25 \pm 0.01}$	$\mathbf{0.25 \pm 0.03}$	$\mathbf{0.25 \pm 0.01}$	0.50 ± 0.00
P47	335	0.28 ± 0.01	$\mathbf{0.24 \pm 0.02}$	0.25 ± 0.04	0.25 ± 0.01	0.38 ± 0.01

sufficient number of samples [317]. As there are no other studies or publicly available datasets that we could compare our results, we compared the results of our system to a random baseline. We used `DummyClassifier` [273] model of scikit-learn library for random baseline calculations.

Table 5.7 shows the mean squared error values of the different regression models for each participants. Some participants were ignored due to insufficient data. Among 42 participants, seven participants had a non-Turkish native language, and most of their interactions were eliminated due to language restrictions in this study. Another ten participants had less than 10 error cases; therefore, we also ignored them.

According to these results, Random Forest regression resulted in the most minor mean squared errors among different models. Moreover, the Random Forest regression model outperformed the baseline for all participants.

5.11 Summary and Conclusion

This section started with a description of a pipeline to recognize context with available sensors. The steps in the pipeline included sensor selection, data cleaning, data segmentation to windows, feature extraction, splitting data to training and test sets, oversampling to balance labels, dimensionality reduction, parameter search, and classification/regression applications.

We first investigated the sensors to recognize context. The literature has many studies for mobility and environment contexts; therefore, we based our sensor selections on existing approaches. Even though our sampling rate was less than these approaches, classification models performed promising results for mobility and environment. On the other hand, the research on social context, multitasking, and distractions is relatively limited. Therefore, we used different combinations of sensors and compared their results. However, the classification results for these contexts are not as good as mobility and environment. This might be explained by the fact that environmental and motion sensors provide sufficient information about both mobility and environment contexts. On the other hand, available sensors are limited for the other contexts. For instance, a smartphone user might put his/her phone on a table and have lunch or talk to someone else while also texting. Using camera recordings or audio might only be effective in such cases. However, it would be costly and obtrusive.

We continued our investigation with error prediction. We automatically calculated participants' typing errors in transcription streams and associated these errors with sensor data. Since our previous findings showed that context affects each individual differently, and the number of samples is unbalanced, we decided to apply regression models to individual datasets. The success of these models varies from individual to individual, as we already expected. In overall, the Random Forrest regressor performed the best for the majority of the participants and overall dataset. It also outperformed the random baseline.

CHAPTER 6

ADAPTATION AND DISCUSSION

The previous chapters investigated the methods to measure users' performance without a task model, investigated the effect of context on users' performance, and compared different methods to classify context and predict users' performance. This chapter discusses different adaptation techniques to overcome SIIDs and the general findings of this thesis work. Section 6.1 starts with explaining the ability-based design and possible adaptations. Section 6.2 explores the adaptations for smartphones. Section 6.3 continues with possible approaches that might use the findings of this thesis work to overcome SIIDs. Finally, Section 6.4 discusses the general findings of this thesis work.

6.1 Ability-Based Design and Examples

Chapter 4 investigated the effect of context on users' typing performance and found that it affects every individual differently. This diversity shows that every user needs different strategies to overcome performance problems. One strategy that improves the performance of some users might not be effective for others [318]. The ability-based design considers the users' abilities under different conditions instead of considering their disabilities [14]. The ability-based design has seven principles [14]:

1. Ability: An ability-based system must focus on the users' abilities.
2. Accountability: In case of performance problems, an ability-based system must change itself without expecting users to change.
3. Adaptation: An ability-based system must provide mechanisms for adaptations,

either by automatically or users' preference.

4. Transparency: An ability-based system should notify users about adaptations and provide mechanisms to revert these adaptations.

5. Performance: An ability-based system should monitor and predict users' performance.

6. Context: An ability-based system should be aware of the context and its effects on users' performance.

7. Commodity: An ability-based system should rely on available hardware and software resources.

One of the examples of the ability-based design is SUPPLE++ [319]. SUPPLE++ models users' pointing performance in desktop settings and adapts itself for motor or vision abilities. In case of poor performance, it changes the orientations and sizes of the widgets on the graphical user interface. Adaptive Click-and-Cross [320] is another example. This approach provides an alternative representation of the interaction elements (such as small targets) when users click on or near these elements. Smart Touch [321] aims to increase the touch accuracy of people with motor impairments by overcoming multiple or unintentional touch problems. It predicts users' intended touch points with a personalized template matching algorithm. Finally, PointAssist [322] aims to improve the pointing performance of older adults by configuring cursor speed in case of a decrease in performance. If the speed and length of submovement are under predefined thresholds, the precision mode was turned on, reducing the cursor speed to its half.

The possible adaptations can be summarized as follows:

- The system may change the visual elements' colors, sizes, or orientations (e.g., SUPPLE++).
- The system may suggest alternative representations (e.g., Adaptive Click-and-Cross).
- The system may predict user input and correct it in case of performance problems (e.g., Smart Touch).

- The system may change its configurations, such as cursor speed or touch accuracy (e.g., PointAssist).

6.2 Adaptations for Smartphones

Using a smartphone itself can cause performance problems similar to those experienced by users with motor impairments [13]. Contextual factors such as the environment, the current position, or the accessories used can cause additional problems. Users generally adopt different strategies to overcome these problems. For instance, smartphone users change their current locations, remove the accessories that may introduce problems, adjust the smartphone's brightness, use their hands for shadow, or postpone their task for a better condition to prevent situational visual impairments [323]. On the other hand, these performance problems can be addressed by applying adaptations similar to those for users with physical impairments [324]. For example, to maintain the same performance with a stable condition while walking, target sizes might be increased [78]. Moreover, modern mobile operating systems support some adaptations. For example, when the light condition is not suitable enough to see the contents on the screen, the operating system controls the screen's brightness [325]. However, Yu et al. [325] showed that this adaptation is ineffective.

The adaptation process can be automated by using different available data sources. Goel et al. [36] showed that the accelerometer sensor can be used to overcome the performance problems experienced while walking. According to Goel et al. [326], detecting hand posture can prevent situations like carrying something with the dominant hand or grabbing a handle in public transportation from causing performance problems. Furthermore, Sarsenbayeva et al. [49] suggested that using the smartphone's battery temperature can help adapt to interaction in cold environments. This study collected text entry data in the wild and processed this data offline to measure the users' performance. The same approach in Chapter 5 can be applied to measure the performance online and support the adaptation of the user interfaces to the users' abilities and context.

Section 3.2.7 presents the participants' self-evaluations on whether they made a typing error or the reason for the typing error as they perceived. In the majority of the cases, participants did not correlate their editing/correcting behavior with a typing error or associate the typing problem with a particular reason. Our error correction/edit classification implementation, on the other hand, classified most of the cases as error corrections. This finding also supports the necessity of an automated adaptation mechanism so that even if a user does not perceive the SIIDs, this mechanism can prevent users from the possible negative impacts of the SIIDs.

Our results also showed individual differences between the effects of different contextual factors on participants. According to our findings in Chapter 4, the same context factor has different effects on each participant. It may reduce the typing speed or increase the error rate for one participant, while it increases the typing speed or reduces the error rate for another participant. A possible explanation for these results may be the different strategies employed to overcome SIIDs. For instance, a user who needs to send a text message while walking in a public area may wish to complete the typing tasks as soon as possible, increasing the typing speed and possibly increasing the error rate. Another user, on the other hand, may prioritize paying attention to the surroundings and decrease his/her typing speed. Therefore, each strategy users intentionally or unintentionally employ under different scenarios may affect user performance differently. For this reason, user-specific ability-based interfaces that adapt themselves based on the users' abilities should be considered [106]. As software libraries to sense the context become available, mobile app developers can use them to create adaptive interfaces [327]. For instance, a background process can send broadcasts whenever a user is in a situation that may affect his/her performance, and the apps receiving these broadcasts can apply different adaptations based on the requirements of the user interface.

An index of performance can be calculated using Fitts' law [328,329], which uses the distance to a target and the size of the target. According to Fitts' law, the difficulty of a task decreases when movement distance decreases or the target size increases. A possible implication of Fitts' law might be that increasing the key sizes in a soft keyboard balances the task difficulty under conditions that users may have interaction problems due to contextual factors.

6.3 Adaptations to Context

There are two approaches that can be used to overcome SIIDs based on our findings in this thesis work. One possible approach is using the data from available smartphone sensors and user keyboard interaction data. The keyboard interaction data is converted to the user performance metrics using the methods presented in Chapter 3. In this thesis work, we combined three approaches in the literature to detect typing errors and distinguish between edits and error corrections. According to the results presented in Section 3.3.4.4, our implementation outperformed the existing approaches in the literature. This approach can be used to measure users' typing performance without a predefined task model. Then, using regression models compared in Chapter 5, performance is predicted based on current sensor recordings and previous user interactions. If the corresponding regression model predicts a performance problem, the smartphone notifies the actively running app of possible adaptations. This approach is illustrated in Figure 6.1.

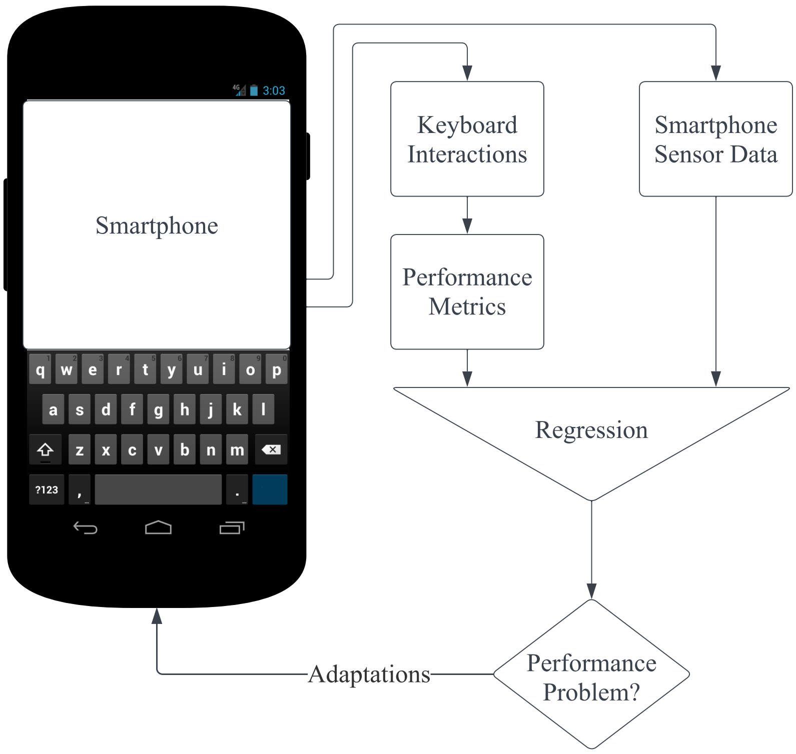


Figure 6.1: Adaptations based on sensor data and keyboard interactions

Another possible approach is adapting based on the context types that affect users' performance. The keyboard interaction data is converted to the user performance metrics, similar to the other approach. Moreover, the system classifies the current context using the available sensor data and context labels provided by the user, as explained in Chapter 5. Then, statistical methods presented in Chapter 4 are used to find correlations between the current context and the user performance. The results of our statistical analysis presented in Table 4.4 in Section 4.3 showed that environment, mobility, social context, multitasking, and distraction conditions have the main effects on users' typing errors. Moreover, our investigations revealed individual user typing speed differences under different contextual factors. Suppose the user's performance is negatively affected by the current context. In that case, the smartphone again notifies the actively running app of possible adaptations.

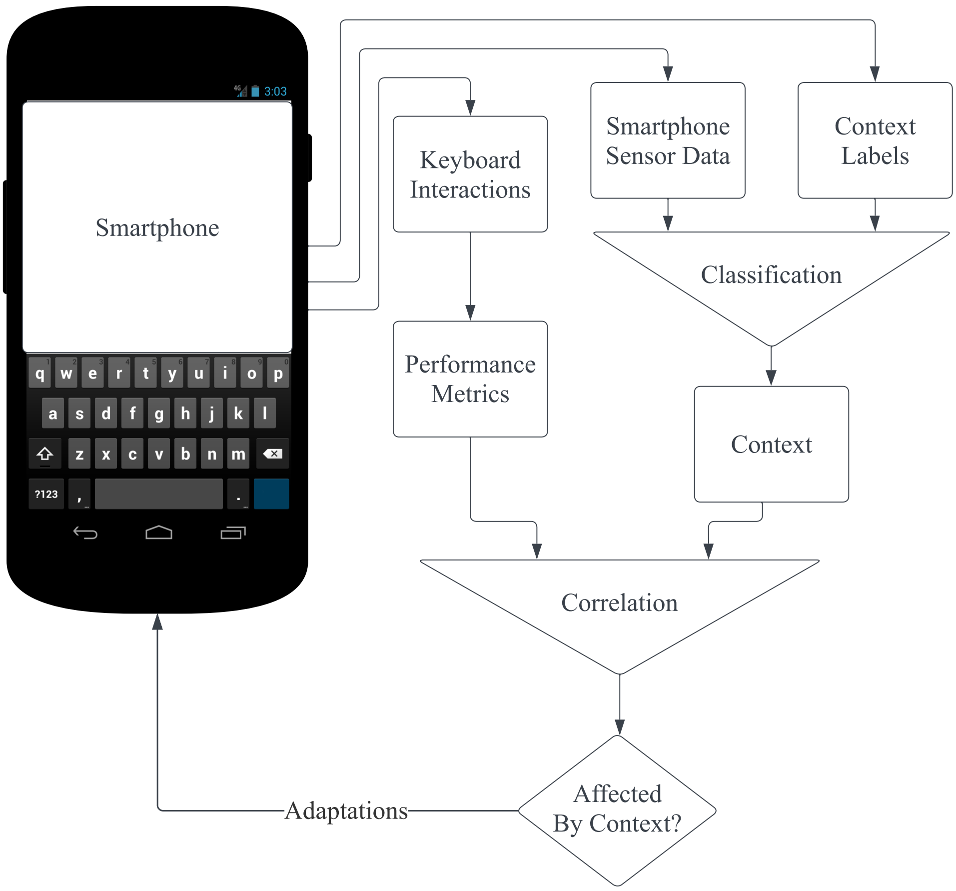


Figure 6.2: The context recognition pipeline

The computational cost of the first approach would be higher since it must monitor performance changes all the time while detecting only context changes in the second approach would be sufficient. However, requesting users to provide context labels might be frustrating. Moreover, it is possible that the user may not have previously labeled a new context. The second approach applies the adaptations whenever a user is in a specific context if that context was correlated with performance problems. On the other hand, the first approach applies adaptations based on situations. Both approaches must update themselves as users' abilities and disabilities change over time.

6.4 General Discussion

In this study, we investigated the effect of context on smartphone users' text entry performance in real-world settings. We conducted a user study in the wild and collected participants' text entry data, sensor data, and context labels. We identified a set of performance metrics to measure users' typing performance systematically. We combined several existing approaches to detect typing errors and distinguish between edits and corrections to better measure typing speed. Finally, we investigated the effect of context on user performance by combining the performance metrics and context labels in five dimensions: environment, mobility, social, multitasking and distraction.

In reviewing the literature, the text entry studies investigating the effect of context on users' performance have been conducted in controlled settings. Our study, on the other hand, was conducted in the wild. For this purpose, we extended an existing framework and captured the participants' keyboard interactions, a set of sensor data, and context labels submitted by the participants. In our user experiment, the participants interacted with their smartphones as they do in their daily lives without a predefined task model. This approach helped us to collect user data in more realistic settings.

Measuring user performance without a task model is challenging. There are several approaches to detect typing errors and measure typing speed; however, these lookup-based approaches handle daily texting language manually or treat them as typing errors. Daily texting language is too common that considering them as typing errors since they are out-of-vocabulary would yield incorrect interpretations about the effect of context on users' performance. On the other hand, manual analysis introduces privacy issues and is not applicable for possible applications of error detection mechanisms. We combined several existing approaches to cover daily texting language and detect typing errors in English and Turkish. Our evaluation showed that our implementation improved the error detection accuracy compared to the literature. However, even though we applied the text speak rules in the literature, some participants' verdicts for error detection introduced new text speak uses that we did not cover initially. Therefore, an error detection mechanism should learn common usage patterns and adapt itself to users.

The majority of the text entry studies investigating the effect of context on users' performance have primarily focused on different mobility conditions. It may be the case that different mobility conditions can be easily replicated during a study. Moreover, there was contradicting evidence in the literature regarding the effect of mobility. This study considered the context in a broader perspective in five dimensions: environment, mobility, social, multitasking, and distraction. The results of our experiment yielded that being in an outdoor environment, being mobile, presence of other people and having distractions increased error rate, while they did not affect typing speed. Multitasking increased the number of keystrokes in a second and error rate. These are the first empirical evidence on the effect of context on users' typing performance in a study conducted in the wild.

6.4.1 The Effect of Context on Users' Typing Performance

In this thesis work, we focused on five context dimensions. For the environment, the error rate was significantly lower for the indoor group than for the outdoor group in terms of KSPC and ER. However, no significant difference between the two groups was evident for WPM and KSPS. Generally, we are exposed to more external factors in the outdoor environments. Therefore, people likely pay more attention to these external factors than typing, or some factors make it difficult to type, resulting in higher error rates. Prior studies have focused on a single aspect of the environment. Sarsenbayeva et al. [50] considered ambient noise and Sarsenbayeva et al. [226] investigated the effect of ambient light. The present study was designed to consider all aspects of the environment.

Mobility had a significant effect on the uncorrected error rate. Participants' error rate was higher when mobile than when they were stable in terms of ER. No significant difference between the two groups was evident for WPM, KSPS, and KSPC. We mainly focus on our surroundings to avoid hazards when we are walking. In general, therefore, it seems that this causes more typing errors. It also seems possible that users do not correct their typing errors in mobile conditions. Comparison of the findings with those of other studies confirms the increase in error rate in the case of mobility. In contrast to earlier findings, however, no evidence of the effect of mobility on typing speed was detected.

For the social context, the presence of other people increased the error rate in terms of KSPC and ER. The participants made fewer typing errors alone than when there were people around. Similar to the environment and mobility, it did not significantly affect the typing speed. The presence of other people and social interaction with them may have shifted the focus from the text entry task to the interaction. Therefore, this resulted in more typing errors. In reviewing the literature, no data was found on the effect of social context on users' typing performance.

Multitasking affected the participants' error rate, increasing both KSPC and ER. Multitasking did not have a significant effect on typing speed. Like the social context, focusing on other tasks may have increased the error rate. This finding was also reported by Crease et al. [76]. However, the findings of the current study do not support Sarsenbayeva et al. [52] who reported no significant effect of multitasking on error rate. Chen et al. [53] also found the main effect on texting time in case of a dual task while crossing a street. Our results are also partially supported by the literature on the target selection task domain with encumbrance conditions. Ng et al. [41,111,113,159] showed that accuracy of target selection was significantly affected by encumbrance condition. On the other hand, while they showed a significant effect on task completion time in two studies [41, 159], they could not find a main effect in another study [111].

The presence of distractions increased the error rate in terms of KSPC and ER; however, it did not affect typing speed. It may be that distraction factors took the participants' primary focus similar to the environment and mobility, and the participants made more typing errors when they were interrupted. This outcome is contrary to that of Jain and Balakrishnan [153] who found an increase in typing speed and a decrease in error rate when participants were distracted.

It is interesting to see individual differences in the effect of context on different participants. A context factor may affect a participant negatively by reducing the typing speed or increasing the error rate, while the same factor may improve another participant's performance by increasing the typing speed or reducing the error rate. Ability-based design is an approach in which users do not adapt themselves to a system; instead, it measures the user performance and adapts itself. For instance, if a user has problems tapping on a key on the keyboard, the system may increase the size of the keys to prevent the error. The ability-based design identifies and exploits users' abilities rather than their disabilities to enhance interaction using available resources [14]. Overall, these results show that ability-based design could be an approach to better consider users' context. Further research is needed to show the actual effect of ability-based designed applications on the users' performance.

6.4.2 Challenges of conducting studies in the wild

There are several issues related to conducting a user study remotely in the wild. Since our study was remote, participants were asked to install an application on their smartphones and share their data during the study. The app running as a background service consumed battery and bandwidth with data collection and participants' attentional resources with questionnaires. Finding voluntary participants that would install such an app and keep it for at least three days was challenging even if we offered a small amount of compensation. Overall, we collected data from 48 participants. Another significant issue is privacy. When asked to share daily data with strangers, people could have privacy concerns. We clearly explained how and why we processed the data to address the participants' concerns. Moreover, we provided a mechanism to pause and resume the experiment so that participants could stop sharing data when they felt uncomfortable. Still, we could find more participants if we did not transfer keyboard data to our server and process them on the participants' devices. However, we needed keyboard data to work on a typing error detection mechanism. Data security and anonymity are essential in such studies, and researchers should pay attention to these issues.

Conducting a study in the wild enabled us to collect real-world data from the users while doing their daily tasks in their everyday context. However, controlling the samples to maintain a balance between independent groups is challenging in such studies. This balance is typically ensured in controlled studies. The researchers can specify the number of observations required for each context factor and continue the experiment until the expected number of samples is collected. In this study, on the other hand, we collected data labels during participants' daily life. We did not ask participants to change their normal behaviour and use their smartphones under conditions they would not normally do. Some participants may prefer not to use their smartphones under specific conditions, such as while walking. Moreover, some participants may not have encountered certain context factors during the experiment. The imbalance of the contextual factors is a tradeoff between controlled and in-situ studies.

The study was conducted during the Covid19 pandemic. During this pandemic, people were encouraged to isolate themselves from each other and stay at home. To not risk researchers and participants, we have conducted this study as a completely remote study. Since the participants would download, install and configure the app independently, we had to set clear instructions for this process. When a participant failed to complete this process, we tried to assist him/her remotely. Some participants abandoned early due to some technical problems, and we could not investigate the problem effectively since we did not have access to the smartphones. Moreover, since people were at home most of the time, this might have limited the coverage of contextual factors in submitted questionnaire answers. On the other hand, we could reach participants with a broad range of demographic profiles by conducting a remote experiment.

6.4.3 Context Classification

A considerable amount of literature has been published on human activity recognition. This thesis work compares various classification models with our sensor dataset and context labels. These comparisons showed that each classification model outperformed the random baseline. The data collection framework collected sensor data at the standard sampling rate. Moreover, our user study did not force participants to interact with their smartphones under predefined contexts, resulting imbalanced distribution of labels for some contexts. The accuracies and F1 scores of classification models can be further improved by increasing the sampling rate or collecting more data to balance contextual factors.

There is a large volume of published studies classifying environment and mobility contexts. On the other hand, a relatively small body of literature is concerned with social context, multitasking, and distractions. We followed the existing approaches to selecting relevant environmental and mobility sensors. However, we had to compare the different sets of sensors for the other contexts. Furthermore, available environment, motion, and position sensors in smartphones provide information about the current environment and mobility conditions.

Nevertheless, these sensors might be limited in some cases, such as when the phone is on a stable platform. There are existing approaches; however, they propose obtrusive methods. For instance, ObstacleWatch [330] aims to detect obstacles by using the microphones in smartphones; however, it requires users to hold their smartphone in a specific position. Putze et al. [331] aim to detect auto-correction errors using brain activity and eye gaze. However, capturing brain activity requires additional hardware, and the user has to use the smartphone in a specific position to capture eye gaze.

6.4.4 Error Prediction

According to Suchman [332] "interaction between people and machines implies mutual intelligibility, or shared understanding" (p. 6). This thesis work aims to take this implication a step forward and provide a mechanism to predict users' intentions and possible interaction problems. For this purpose, the current study compared several regression methods for individuals and overall users as a single dataset. This comparison showed that the Random Forest regressor produced the most successful results and outperformed the random baseline.

Similar to the context classification task, sensing some contextual factors might be challenging. For instance, a user's emotional condition or something that confuses the user's mind might also cause performance problems. In these cases, the sensors in a smartphone might be limited. Moreover, predicting performance problems and applying corresponding adaptations would take a certain amount of time. An immediate distraction that prevents a user from interacting with the smartphone for only a moment might decrease the user's performance. However, the system might not capture this immediate distraction or respond in a reasonable time to adapt. Therefore, any adaptation technique is limited by smartphones' computational and sensing capacities.

6.4.5 Best Practices for Adaptation

People use their smartphones in various contexts and are exposed to the adverse effects of SIIDs. Monitoring user performance and context can help systems adapt to overcome these adverse effects of SIIDs. Modern smartphones can support these adaptations with available sensors. The system can monitor the user performance, context, and different usage patterns to provide a personalized user interface for different users. These personalized user interfaces can make users focus on the task and handle task complexity by recognizing their intents [333]. However, the benefits of the adaptations must outweigh the side effects caused by design problems [334]. This section covers the usability problems highlighted in the previous studies and provides suggestions to overcome these problems.

6.4.5.1 Smartphone Capabilities

Smartphones with touchscreens have advantages and disadvantages over the desktop computers. The soft keyboards support dynamic resizing of the keys and different orientations [335]. On the other hand, smartphones are limited in screen size, processing capacities, and bandwidth.

The soft keyboard can be resized in the text entry domain to overcome performance problems. The adaptive behavior can provide bigger keys when users have difficulty tapping on a specific key. Moreover, tactile feedback can inform users when a touch is successful [335].

The two approaches presented in Section 6.3 require real-time user performance and sensor data monitoring. However, these adaptations should be designed considering smartphones' limited capacities and resources. One possible design choice is to train the models in an external server and use the best model in the smartphone. However, this would require synchronizing the performance and sensor data regularly. It might consume the bandwidth of the smartphone. In any case, the performance and sensor data for training should be kept at an optimal level.

6.4.5.2 Learning the Adaptive Behavior

The adaptive systems change the layout or modality to overcome performance problems. Novice users may have difficulty learning these adaptations and reverting some adaptations. Easy-to-use mechanisms are mostly insufficient; therefore, the system should support users in the learning phase [336]. Alvarez-Cortes et al. [333] and Jameson [337] suggested providing help mechanisms so that users can learn the adaptive behavior with explanations of each function. Moreover, users should be informed about the consequences of their actions, and the system should make recommendations on different use cases [337]. Finally, the models with a hard-to-explain process might prevent users from learning why the system changes in such a way [333].

6.4.5.3 Users' Acceptance

People are more willing to use adaptive systems on smartphones due to their limited capabilities [338]. However, usability, trust, and acceptance issues are associated with adaptive systems [339]. People may have concerns about giving control to an automated system even if it empowers them [340]. Peissner and Edlin-White [339] suggested that using implicit and explicit feedback provides a transparent and controllable interaction and increases user acceptance. According to this approach, the system should request permissions before applying an adaptation or show a notification after it. Jameson [337] commented that users should be provided the system's benefits so that they continue to use it. The following two sections continue with other issues that may jeopardize user acceptance: predictability and accuracy.

6.4.5.4 Predictability of Adaptations

According to Norman [341], a system should be consistent in structure and design so that users can memorize and predict the operations with minimum memory problems. People build mental models to predict system behavior. Inconsistent adaptations might cause unpredictable behavior, resulting in user frustration [339]. On the other hand, a predictable system satisfies users more [342]. Therefore, the adaptations should be predictable [337, 343] and reliable [344] for user acceptance. Explaining the actions to the user might increase predictability [337]. Following Nielsen's usability heuristic on consistency [345], adaptation conventions and standards should be established. These conventions may help users to build common mental models across different applications.

6.4.5.5 Accuracy of the System

An adaptive system can not be considered without users, and it can be effective only if it improves users' interaction [346]. Even if an adaptive system is designed to be consistent and predictable, it can have accuracy problems due to the underlying model. The accuracy of an adaptive system has a significant impact on users' performance [347]. For instance, Gajos et al. [342] showed that increasing the system's accuracy improved the task required to complete a task and the utilization of the system. This positive impact on user performance is more significant on small screens than on large screens [338]. Inaccurate system behavior, on the other hand, causes confusion and trust issues [337]. Therefore, the system's accuracy should be considered one of the most critical factors affecting user acceptance.

When user control is involved in the interaction, both accurate and inaccurate systems produce successful results. Moreover, user control increases the user performance even if the system has an inaccurate behavior [348]. Therefore, users should be provided with a control mechanism if the corresponding models are not sufficiently accurate.

6.4.5.6 Collecting Sufficient Data

Different user characteristics including education level, personality, cognitive skills, preferences, current mood, goals [349], abilities, interests [350], culture, native language, and religion [351] might affect the effectiveness of an adaptive system. Moreover, difficulty and motor or cognitive demands of the corresponding task are also important [349]. Jameson [337] suggested involving users by combining the adaptation decision process with user feedback.

An adaptive system's data collection process must be considered an online process. The data must be updated after each interaction in terms of user performance, and preference [346]. An adaptive system should learn rapidly, and the number of training samples is often limited. The developers should focus on the models that can be able to train with small sample sizes [346].

6.4.5.7 Privacy Concerns

The developers of an adaptive system should consider the trade-off between usability and privacy. The users of such a system may have privacy concerns [337] as data collection mechanisms might introduce privacy vulnerabilities, and possible threats [352]. Jameson [337] suggested that developers should limit the storage of sensitive data to prevent these issues. Moreover, the users should be provided with descriptive information about how the collected data will be used, along with mechanisms to check and modify the user model [337]. Moreover, the captured keyboard interactions must not be transferred to a server or stored in the smartphone for security and privacy reasons [6]. Therefore, the system should only store the performance metrics, not the textual content.

6.4.5.8 Other Concerns

The other usability concerns can be listed as follows:

1. A fully adaptive system might impair users' skills over time. This negative effect can be overcome by involving users with a mixed-initiative [349].
2. The changes in the underlying layout might result in undesirable user interfaces. Therefore, adaptations should be a part of the early design process [337].
3. The frequency of adaptations can play a significant role in the acceptance of an adaptive system [347]. Dynamic resizing user interface elements unpredictably might frustrate users [335].
4. The frequency of use of the feature associated with the adaptive behavior is also a concern. The adaptations are beneficial in routine tasks. However, intermediate levels of adaptations should be preferred for non-routine tasks [349].
5. An adaptive system might be unresponsive if the underlying model is slow [333]. On the other hand, users are concerned about timely adaptations [340].

6.4.6 Design Guidelines

The previous section presented the concerns related to adaptive interfaces. This section continues with a discussion of our findings in this thesis work related to these concerns and proposals for design guidelines. These guidelines are suggested for both mobile operating system developers and mobile app developers to consider SIID-related issues and adaptations.

- **Smartphone Capabilities:** In Section 3.2.6 we mentioned the diversity of smartphone brands and models participants used during our user study. In Section 5.2 we stated that six of these devices had only a few sensors; therefore, the corresponding data was eliminated in our investigations. The absence of available sensors might also be the case in a real-world scenario. An adaptive system should be able to cover such situations.

Data synchronization might be interrupted due to low battery or offline status of the smartphone if the underlying model is trained on an external server. Moreover, the external server might be temporarily unavailable. In data loss, the system might fail to predict user performance problems. In such cases, the system might remain responsive in order not to affect users' interactions.

Some participants commented on the excessive battery use and the system locking down during data collection. Although these cases have a cost on users, they might be ignored if the system proves itself to improve user interaction.

- **Learning the Adaptive Behavior:** Our user study was remote, and participants had to install and configure the application. Some participants had difficulty following the steps even if we provided a detailed description. When the application stopped unexpectedly by a technical problem or the operating system, the participants could not notice it or needed guidance to recover the application. This learning issue might be the case for an adaptive system. For users to properly install, configure and understand an adaptive system, sufficient help mechanisms should be provided.

- **Users' Acceptance:** Our literature review in Chapter 2 revealed that environment, mobility, social context, multitasking, and distractions significantly affect users' performance. The results of our statistical analysis presented in Table 4.4 in Section 4.3 also supported these findings. The literature also suggested that these performance problems due to contextual factors can be reduced with adaptations, such as reducing the sizes of user interface elements, automatically changing modality or configurations, or providing alternative representations. However, the users' perception of successful interaction might not match the theoretical assumptions [347]. Therefore, the adaptation techniques proposed in this thesis work should be evaluated with a user experiment.

- **Predictability of Adaptations:** The approach in Figure 6.1 in Section 6.3 aims to detect user performance problems by using current sensor data. Although this approach is more situation-specific than the other approach, it also might make users unable to predict the adaptations. On the other hand, the approach in Figure 6.2 associates the context with the user performance and might provide more predictable actions.

- **Accuracy of the System:** According to the results presented in Section 3.3.4.4, even if our error detection implementation outperformed the existing methods, it still has an 87.1% of accuracy. This result shows that our error detection mechanism might fail to identify some typing errors or classify some correctly typed words as errors. This might affect the accuracy of the user performance predictions and prevent users from trusting the system.

  Moreover, although our error prediction mechanism outperformed the baseline according to the results in Section 5.10, there is still room for improvement. Inaccurate adaptation behavior might cause users to lose trust in the adaptive system.

  As the literature suggests, an adaptive system should explain its actions to the user. Moreover, it should collect feedback after specific actions, which should be used to improve the system's performance.

- **Collecting Sufficient Data:** Some participants have entirely or partially ignored the context label questions during our user study. This result might be due to the operating system preventing notifications of our application from appearing in the notification panel. Moreover, the users might not have preferred to pay attention to these questions under certain contextual factors. If the approach in Figure 6.2 in Section 6.3, insufficient number of context labels might prevent system to find correlations between the context and user performance.

  According to Figure 4.1 in Section 4.1, some participants have spent less time with their smartphones and entered text less than others. It may take longer to collect a sufficient size of samples for such users, especially with a model that requires large samples. This longer period might result in giving up the benefits of the system and uninstallation of the application. Moreover, the users must interact with the system under different conditions for a certain period to collect training data. However, it might cause another frustration once the user expects the system to adapt but fail due to insufficient data size for training.

- **Privacy Concerns:** We collected participants' keyboard interactions during our user study. This experimental decision raised many questions among our participants, especially on how we would use, store and process their data. An adaptive system should only store and synchronize the user performance data.

Moreover, the users should be well explained about how their data is used for their benefit.

Some adaptation techniques, such as changing modality, might also cause privacy issues. For instance, if the system changes its modality to audio interaction in a public environment due to occupied hands, it might result in other people witnessing the user's potentially private interaction. Therefore, the system should ask users' permission before taking action.

CHAPTER 7

CONCLUSION AND FUTURE WORK

Smartphones are an essential part of our daily lives, and we can take them wherever we go. As a result, we use them under various contextual factors. Some of these contextual factors can cause SIIDs, which are temporary reductions in user performance due to exposure to a set of contextual factors. On the other hand, the ability-based design approach aims to reduce the impact of SIIDs by adapting the system to the user.

This thesis work aims to understand how context affects user performance and develop a mechanism to detect performance changes due to the context. Moreover, this thesis work aims to predict performance problems unobtrusively using available smartphone sensors and propose adaptations for the system. For this purpose, we systematically reviewed the literature on smartphone use under different contextual factors. We investigated the effect of context on user performance in five dimensions: physical context, temporal context, social context, task context, and technical context. Then, we conducted a user study to collect data for user performance measurement and sensing the context. We implemented an error detection mechanism by combining several approaches in the literature. We applied statistical analysis to see the effect of context on user performance. Finally, we compared different machine learning models to predict user performance problems due to context.

Although context has many dimensions, the research on the effect of context on users' performance has mainly focused on different mobility conditions. Our review also revealed that only a few studies had been conducted in the natural settings of the users. Although reproducibility and performance measurement is based on more reliable methods, studies in the wild reflect more realistic use case scenarios. Therefore, we designed our empirical study to be conducted in the wild without a predefined task model.

A user study to measure user performance without a task model requires a mechanism to predict users' intentions. There are two main issues to consider in a text entry task. First, the mechanism must identify intentional and unintentional typing errors. Then, the mechanism must determine when a user corrected a typing error or changed his/her mind to type something else. The existing approaches in the literature were limited as they did not consider daily texting language. Therefore, we combined the approaches of Nicolau et al. [17], Evans and Wobbrock [18], and Torunoğlu and Eryiğit [20] to cover daily texting language and detect typing errors in both English and Turkish. Our evaluation showed that the combined approach improved error detection and correction/editing detection.

We asked our participants to label the context in five dimensions: environment, mobility, social, multitasking, and distractions. These five dimensions were identified based on our systematic review. We investigated the effect of these context dimensions by using performance metrics we calculated from the participants' transcription streams. Statical analysis showed that five context dimensions did not have a main effect on typing speed; however, they significantly affected the error rate. Moreover, our analysis showed that each participant was differently affected by each context dimension. While some participants performed better under a particular contextual factor, others had performance problems under the same factor.

We also collected sensor data in our user study. First, we used this data to explore relevant sensors to classify different contexts. We compared different classification models with different parameters. These models outperformed the random baseline. Then, we applied regression models to individual user data and overall user data as a single dataset. We combined the sensors to classify different contexts and associated them with users' performance data. Random Forrest regressor produced the best results and outperformed the random baseline.

Finally, we reviewed the existing adaptive systems and identified possible adaptation approaches. Then, we discussed how to use the sensing, error detection, context classification, and performance prediction methods in an adaptable system.

7.1 Limitations

Our study is not without limitations. Even though we combine many different techniques, our accuracy is still not 100% for calculating performance metrics, therefore there is always a risk of not assessing the users' typing errors fully. Furthermore, the following cases were a true challenge for our automated assessment. First, some of the text-speak uses are identical to typing errors. For instance, character repetitions may both indicate emotions and a typing error. Therefore, some out-of-vocabulary words that we identified as intentional errors or text speak may correspond to unintentional typing errors.

A typing error may result in another valid word. Moreover, the spelling of a word may be correct; however, it may not be grammatically correct in the sentence. Our implementation does not detect these errors. This problem could be addressed by checking the occurrence frequencies of the tokens with surrounding words. However, further studies are needed to explore such Natural Language Processing (NLP) techniques.

This thesis work has collected participants' data in their daily settings without a predefined task model. As a result, the distribution of the context labels was imbalanced. A more comprehensive study might enable more data collection in different contexts and applying deep learning mechanisms to predict users' performance.

7.2 Future Work

The typing error detection mechanism accepts a token as a correctly spelled word if a spellchecker or a corpus validates the token or its transformed forms. However, a typing error, such as a transposition error, may result in another valid token. In such a case, the word's spelling may be correct; however, it may not be grammatically correct in the sentence. In our thesis work, we ignored these cases. However, applying NLP techniques can detect such kind of typing errors.

Our study focused on typing performance. However, our literature review showed that smartphone interactions also have different task domains, such as target acquisition or navigation. Measuring user performance would require different metrics and approaches for these task domains. Detecting user performance problems for these task domains in the wild would introduce new challenges. Further research can investigate different task domains and possible adaptations to overcome SIIDs.

Our analysis investigated the effect of context on users' performance in five context dimensions. It showed that each contextual factor affected the individuals differently. However, a user is exposed to the contextual factors in all five dimensions simultaneously. Further research can also be conducted to investigate which context has the highest performance impact on a particular user or how composite context combinations negatively affect users' performance.

This thesis work reviewed and discussed possible adaptations to overcome performance problems due to SIIDs. The adaptation approaches discussed in this thesis work can be implemented, and the effectiveness of these adaptations can be explored in terms of user performance and perception. The literature states that users might perceive a different understanding of a successful interaction than theoretical assumptions. Moreover, Section 6.4.5 listed the concerns on adaptive systems, especially user acceptance. Therefore, further user studies should assess and compare different automatic system adaptations regarding user acceptance and effectiveness.

One of the major challenges of investigating the effect of context on users' performance and predicting user performance with context data is to find or collect data. In this thesis work, we collected user performance measurements, context labels, and sensor data from available smartphone sensors. We published our dataset in our public repository[1].

Finally, our user study collected data from each participant for at most one week. Considering the pandemic situation and the general nature of the experiment in the wild, the distribution of different contexts was imbalanced. A user study with a longer duration might ensure the collection of sufficient data for different contexts.

[1] `https://github.com/melgin/cabas-dataset`

REFERENCES

[1] M. S. Ahmed, R. Jahangir Rony, and N. Ahmed, "Identifying high and low academic result holders through smartphone usage data," in *Asian CHI Symposium 2021*, Asian CHI Symposium 2021, (New York, NY, USA), p. 114–121, ACM, 2021.

[2] E. S. Troll, M. Friese, and D. D. Loschelder, "How students' self-control and smartphone-use explain their academic performance," *Computers in Human Behavior*, vol. 117, p. 106624, 2021.

[3] K. Lukoff, C. Yu, J. Kientz, and A. Hiniker, "What makes smartphone use meaningful or meaningless?," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, mar 2018.

[4] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer, "Falling asleep with angry birds, facebook and kindle: A large scale study on mobile application usage," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, (New York, NY, USA), pp. 47–56, ACM, 2011.

[5] A. M. Roffarello and L. De Russis, "Understanding, discovering, and mitigating habitual smartphone use in young adults," *ACM Trans. Interact. Intell. Syst.*, vol. 11, jul 2021.

[6] A. Rodrigues, H. Nicolau, A. Santos, D. Branco, J. Rainey, D. Verweij, J. D. Smeddinck, K. Montague, and T. Guerreiro, "Investigating the tradeoffs of everyday text-entry collection methods," in *CHI Conference on Human Factors in Computing Systems*, CHI '22, (New York, NY, USA), Association for Computing Machinery, 2022.

[7] K. Palin, A. M. Feit, S. Kim, P. O. Kristensson, and A. Oulasvirta, "How do people type on mobile devices? observations from a study with 37,000 vol-

unteers," in *Proc. of the 21st Int. Conf. on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '19, (NY, USA), ACM, 2019.

[8] A. Oulasvirta, S. Tamminen, V. Roto, and J. Kuorelahti, "Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile hci," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, (New York, NY, USA), pp. 919–928, ACM, 2005.

[9] G. W. Tigwell, R. Menzies, and D. R. Flatla, "Designing for situational visual impairments: Supporting early-career designers of mobile content," in *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, (New York, NY, USA), pp. 387–399, ACM, 2018.

[10] J. O. Wobbrock, "Situationally-induced impairments and disabilities," in *Web Accessibility: A Foundation for Research* (Y. Yesilada and S. Harper, eds.), pp. 59–92, London: Springer London, 2019.

[11] A. K. Dey, G. D. Abowd, and D. Salber, "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications," *Hum.-Comput. Interact.*, vol. 16, pp. 97–166, Dec. 2001.

[12] S. Greenberg, "Context as a dynamic construct," *Human–Computer Interaction*, vol. 16, no. 2-4, pp. 257–268, 2001.

[13] Y. Yesilada, S. Harper, T. Chen, and S. Trewin, "Small-device users situationally impaired by input," *Computers in Human Behavior*, vol. 26, no. 3, pp. 427 – 435, 2010.

[14] J. O. Wobbrock, S. K. Kane, K. Z. Gajos, S. Harada, and J. Froehlich, "Ability-based design: Concept, principles and examples," *ACM Trans. Access. Comput.*, vol. 3, pp. 9:1–9:27, Apr. 2011.

[15] A. Sears and M. Young, "Physical disabilities and computing technologies: An analysis of impairments," in *The Human-computer Interaction Handbook* (J. A. Jacko and A. Sears, eds.), pp. 482–503, Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 2003.

[16] A. Sears, M. Lin, J. Jacko, and Y. Xiao, "When computers fade … pervasive computing and situationally-induced impairments and disabilities," in *Proceedings of HCII 2003*, (Mahwah, NJ), pp. 1298–1302, Lawrence Erlbaum, 2003.

[17] H. Nicolau, K. Montague, T. Guerreiro, A. Rodrigues, and V. L. Hanson, "Investigating laboratory and everyday typing performance of blind users," *ACM Trans. Access. Comput.*, vol. 10, Mar. 2017.

[18] A. Evans and J. Wobbrock, "Taming wild behavior: The input observer for obtaining text entry and mouse pointing measures from everyday computer use," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), p. 1947–1956, ACM, 2012.

[19] A. Chamberlain and A. Crabtree, "Research 'in the wild'," in *Into the Wild: Beyond the Design Research Lab* (A. Chamberlain and A. Crabtree, eds.), (Cham), pp. 1–6, Springer International Publishing, 2020.

[20] D. Torunoğlu and G. Eryiğit, "A cascaded approach for social media text normalization of Turkish," in *Proc. of the 5th Workshop on Language Analysis for Social Media (LASM)*, (Sweden), pp. 62–70, ACL, Apr. 2014.

[21] Sufatrio, D. J. J. Tan, T.-W. Chua, and V. L. L. Thing, "Securing android: A survey, taxonomy, and challenges," *ACM Comput. Surv.*, vol. 47, pp. 58:1–58:45, May 2015.

[22] M. Pierer, "Mobile device types," in *Mobile Device Management: Mobility Evaluation in Small and Medium-Sized Enterprises*, pp. 31–36, Wiesbaden: Springer Fachmedien Wiesbaden, 2016.

[23] Y. Yesilada, G. Brajnik, and S. Harper, "Barriers common to mobile and disabled web users," *Interacting With Computers*, vol. 23, no. 5, pp. 525–542, 2011.

[24] S. Harper, Y. Yesilada, and T. Chen, "Mobile device impairment … similar problems, similar solutions?," *Behaviour & Information Technology*, vol. 30, no. 5, pp. 673–690, 2011.

[25] S. Jumisko-Pyykkö and T. Vainio, "Framing the context of use for mobile hci," *Int. J. Mob. Hum. Comput. Interact.*, vol. 2, pp. 1–28, Oct. 2010.

[26] C. K. Coursaris and D. J. Kim, "A meta-analytical review of empirical mobile usability studies," *J. Usability Studies*, vol. 6, pp. 11:117–11:171, May 2011.

[27] L. G. Motti, N. Vigouroux, and P. Gorce, "Interaction techniques for older adults using touchscreen devices: A literature review," in *Proceedings of the 25th Conference on L'Interaction Homme-Machine*, IHM '13, (New York, NY, USA), pp. 125:125–125:134, ACM, 2013.

[28] Y. Liu, J. Goncalves, D. Ferreira, S. Hosio, and V. Kostakos, "Identity crisis of ubicomp?: Mapping 15 years of the field's development and paradigm change," in *Proc. of the 2014 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing*, UbiComp '14, (New York, NY, USA), pp. 75–86, ACM, 2014.

[29] Z. Sarsenbayeva, N. van Berkel, C. Luo, V. Kostakos, and J. Goncalves, "Challenges of situational impairments during interaction with mobile devices," in *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, OZCHI '17, (New York, NY, USA), pp. 477–481, ACM, 2017.

[30] K. Ghezzi-Kopel, "Libguides: A guide to conducting systematic reviews: Steps in a systematic review," Feb 2016. (accessed: 06.09.2018).

[31] J. Clawson, K. Lyons, T. Starner, and E. Clarkson, "The impacts of limited visual feedback on mobile text entry for the twiddler and mini-qwerty keyboards," in *Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, ISWC '05, (Washington, DC, USA), pp. 170–177, IEEE Computer Society, 2005.

[32] M. Pielot, N. Henze, and S. Boll, "Supporting map-based wayfinding with tactile cues," in *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, (New York, NY, USA), pp. 23:1–23:10, ACM, 2009.

[33] B. Schildbach and E. Rukzio, "Investigating selection and reading performance on a mobile phone while walking," in *Proceedings of the 12th International*

Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '10, (New York, NY, USA), pp. 93–102, ACM, 2010.

[34] A. B. Brush, A. K. Karlson, J. Scott, R. Sarin, A. Jacobs, B. Bond, O. Murillo, G. Hunt, M. Sinclair, K. Hammil, and S. Levi, "User experiences with activity-based navigation on mobile devices," in *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '10, (New York, NY, USA), pp. 73–82, ACM, 2010.

[35] M. Pielot and S. Boll, "Tactile wayfinder: Comparison of tactile waypoint navigation with commercial pedestrian navigation systems," in *Proceedings of the 8th International Conference on Pervasive Computing*, Pervasive'10, (Berlin, Heidelberg), pp. 76–93, Springer-Verlag, 2010.

[36] M. Goel, L. Findlater, and J. Wobbrock, "Walktype: Using accelerometer data to accomodate situational impairments in mobile touch screen text entry," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 2687–2696, ACM, 2012.

[37] N. Banovic, K. Yatani, and K. N. Truong, "Escape-keyboard: A sight-free one-handed text entry method for mobile touch-screen devices," *Int. J. Mob. Hum. Comput. Interact.*, vol. 5, pp. 42–61, July 2013.

[38] J. Wither, C. E. Au, R. Rischpater, and R. Grzeszczuk, "Moving beyond the map: Automated landmark based pedestrian guidance using street level panoramas," in *Proc. of the 15th Int. Conf. on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, (New York, NY, USA), pp. 203–212, ACM, 2013.

[39] M. Böhmer, C. Lander, S. Gehring, D. P. Brumby, and A. Krüger, "Interrupted by a phone call: Exploring designs for lowering the impact of call notifications for smartphone users," in *Proc. of the SIGCHI Conf. on Human Factors in Comp. Systems*, CHI '14, (New York, NY, USA), pp. 3045–3054, ACM, 2014.

[40] J. R. Marone, P. B. Patel, C. P. Hurt, and M. D. Grabiner, "Frontal plane margin of stability is increased during texting while walking," *Gait & Posture*, vol. 40, no. 1, pp. 243 – 246, 2014.

[41] A. Ng, J. H. Williamson, and S. A. Brewster, “Comparing evaluation methods for encumbrance and walking on interaction with touchscreen mobile devices,” in *Proc. of the 16th Int. Conf. on Human-computer Interaction with Mobile Devices & Services*, MobileHCI ’14, (New York, NY, USA), pp. 23–32, ACM, 2014.

[42] Q. Xu, L. Li, J. H. Lim, C. Y. C. Tan, M. Mukawa, and G. Wang, “A wearable virtual guide for context-aware cognitive indoor navigation,” in *Proceedings of the 16th Int. Conf. on Human-computer Interaction with Mobile Devices & Services*, MobileHCI ’14, (New York, NY, USA), pp. 111–120, ACM, 2014.

[43] D. Wenig, J. Schöning, B. Hecht, and R. Malaka, “Stripemaps: Improving map-based pedestrian navigation for smartwatches,” in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI ’15, (New York, NY, USA), pp. 52–62, ACM, 2015.

[44] Y. Zhang and P.-L. P. Rau, “Playing with multiple wearable devices,” *Comput. Hum. Behav.*, vol. 50, pp. 148–158, Sept. 2015.

[45] A. Colley, L. Virtanen, T. Ojala, and J. Häkkilä, “Guided touch screen: Enhanced eyes-free interaction,” in *Proc. of the 5th ACM Int. Symp. on Pervasive Displays*, PerDis ’16, (NY, USA), pp. 80–86, ACM, 2016.

[46] J. Hong, L. Stearns, J. Froehlich, D. Ross, and L. Findlater, “Evaluating angular accuracy of wrist-based haptic directional guidance for hand movement,” in *Proceedings of the 42Nd Graphics Interface Conference*, GI ’16, (School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada), pp. 195–200, Canadian Human-Computer Communications Society, 2016.

[47] Z. Sarsenbayeva, J. Goncalves, J. García, S. Klakegg, S. Rissanen, H. Rintamäki, J. Hannu, and V. Kostakos, “Situational impairments to mobile interaction in cold environments,” in *Proc. of the 2016 ACM Int. Joint Conf. on Pervasive and Ubiq. Comp.*, UbiComp ’16, (New York, NY, USA), pp. 85–96, ACM, 2016.

[48] J. Goncalves, Z. Sarsenbayeva, N. van Berkel, C. Luo, S. Hosio, S. Risanen,

H. Rintamäki, and V. Kostakos, "Tapping task performance on smartphones in cold temperature," *Interacting with Computers*, vol. 29, no. 3, pp. 355–367, 2017.

[49] Z. Sarsenbayeva, N. Van Berkel, A. Visuri, S. Rissanen, H. Rintamaki, V. Kostakos, and J. Goncalves, "Sensing cold-induced situational impairments in mobile interaction using battery temperature," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, pp. 98:1–98:9, Sept. 2017.

[50] Z. Sarsenbayeva, N. van Berkel, E. Velloso, V. Kostakos, and J. Goncalves, "Effect of distinct ambient noise types on mobile interaction," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, pp. 82:1–82:23, July 2018.

[51] S.-H. Chen, O.-Y. Lo, T. Kay, and L.-S. Chou, "Concurrent phone texting alters crossing behavior and induces gait imbalance during obstacle crossing," *Gait & Posture*, vol. 62, pp. 422 – 425, 2018.

[52] Z. Sarsenbayeva, N. van Berkel, D. Hettiachchi, W. Jiang, T. Dingler, E. Velloso, V. Kostakos, and J. Goncalves, "Measuring the effects of stress on mobile interaction," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, pp. 24:1–24:18, Mar. 2019.

[53] P.-L. Chen, W. Saleh, and C.-W. Pai, "Texting and walking: a controlled field study of crossing behaviours and inattentional blindness in taiwan," *Behaviour & Information Technology*, vol. 36, no. 4, pp. 435–445, 2017.

[54] M. Pielot, B. Poppinga, W. Heuten, and S. Boll, "Pocketnavigator: Studying tactile navigation systems in-situ," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 3131–3140, ACM, 2012.

[55] I. E. Hyman, S. M. Boss, B. M. Wise, K. E. McKenzie, and J. M. Caggiano, "Did you see the unicycling clown? inattentional blindness while walking and talking on a cell phone," *Applied Cognitive Psychology*, vol. 24, no. 5, pp. 597–607, 2010.

[56] R. Alsaleh, T. Sayed, and M. H. Zaki, "Assessing the effect of pedestrians' use

of cell phones on their walking behavior: A study based on automated video analysis," *Transportation Research Record*, vol. 2672, no. 35, pp. 46–57, 2018.

[57] S. Harper, T. Chen, and Y. Yesilada, "How do people use their mobile phones?: A field study of small device users," *Int. J. Mob. Hum. Comput. Interact.*, vol. 3, pp. 37–54, Jan. 2011.

[58] P.-L. Chen and C.-W. Pai, "Pedestrian smartphone overuse and inattentional blindness: an observational study in taipei, taiwan," *BMC Public Health*, vol. 18, 2018.

[59] T. Horberry, R. Osborne, and K. Young, "Pedestrian smartphone distraction: Prevalence and potential severity," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 60, pp. 515 – 523, 2019.

[60] A. Hiniker, K. Sobel, H. Suh, Y.-C. Sung, C. P. Lee, and J. A. Kientz, "Texting while parenting: How adults use mobile phones while caring for children at the playground," in *Proc. of the 33rd Annual ACM Conf. on Human Factors in Comp. Systems*, CHI '15, (New York, NY, USA), pp. 727–736, ACM, 2015.

[61] D. Fitton, I. S. MacKenzie, J. C. Read, and M. Horton, "Exploring tilt-based text input for mobile devices with teenagers," in *Proceedings of the 27th International BCS Human Computer Interaction Conference*, BCS-HCI '13, (Swinton, UK, UK), pp. 25:1–25:6, British Computer Society, 2013.

[62] S. Brewster, "Overcoming the lack of screen space on mobile computers," *Personal Ubiquitous Comput.*, vol. 6, pp. 188–205, Jan. 2002.

[63] K. Vertanen and P. O. Kristensson, "Parakeet: A continuous speech recognition system for mobile touch-screen devices," in *Proceedings of the 14th International Conference on Intelligent User Interfaces*, IUI '09, (New York, NY, USA), pp. 237–246, ACM, 2009.

[64] K. Vertanen and P. O. Kristensson, "Parakeet: A demonstration of speech recognition on a mobile touch-screen device," in *Proceedings of the 14th International Conference on Intelligent User Interfaces*, IUI '09, (New York, NY, USA), pp. 483–484, ACM, 2009.

[65] J. Conradi and T. Alexander, “Analysis of visual performance during the use of mobile devices while walking,” in *11th International Conference on Engineering Psychology and Cognitive Ergonomics - Volume 8532*, (New York, NY, USA), pp. 133–142, Springer-Verlag New York, Inc., 2014.

[66] P. Plummer, S. Apple, C. Dowd, and E. Keith, “Texting and walking: Effect of environmental setting and task prioritization on dual-task interference in healthy young adults,” *Gait & Posture*, vol. 41, no. 1, pp. 46 – 51, 2015.

[67] D. McMillan, B. Brown, A. Lampinen, M. McGregor, E. Hoggan, and S. Pizza, “Situating wearables: Smartwatch use in context,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI ’17, (New York, NY, USA), pp. 3582–3594, ACM, 2017.

[68] J. Kjeldskov and J. Stage, “New techniques for usability evaluation of mobile systems,” *International Journal of Human-Computer Studies*, vol. 60, no. 5, pp. 599 – 620, 2004. HCI Issues in Mobile Computing.

[69] J. Lumsden and A. Gammell, “Mobile note taking: Investigating the efficacy of mobile text entry,” in *Mobile Human-Computer Interaction - MobileHCI 2004* (S. Brewster and M. Dunlop, eds.), (Berlin, Heidelberg), pp. 156–167, Springer Berlin Heidelberg, 2004.

[70] M. Lin, K. J. Price, R. Goldman, A. Sears, and J. A. Jacko, “Tapping on the move-fitts’ law under mobile conditions,” in *Proceedings of Information Resources Management Assoc. International Conference*, (Hershey), pp. 132–135, Idea Group, 2005.

[71] B. MacKay, D. Dearman, K. Inkpen, and C. Watters, “Walk ’n scroll: A comparison of software-based navigation techniques for different levels of mobility,” in *Proc. of the 7th Int. Conf. on Human Computer Interaction with Mobile Devices & Services*, MobileHCI ’05, (New York, NY, USA), pp. 183–190, ACM, 2005.

[72] J. Price, M. Lin, J. Feng, R. Goldman, A. Sears, and A. Jacko, “Motion does matter: An examination of speech-based text entry on the move,” *Univers. Access Inf. Soc.*, vol. 4, pp. 246–257, Feb. 2006.

[73] M. J. Schedlbauer, R. L. Pastel, and J. M. Heines, “Effect of posture on target acquisition with a trackball and touch screen,” in *28th International Conference on Information Technology Interfaces, 2006.*, (Los Alamitos, CA), pp. 257–262, IEEE Computer Society, 2006.

[74] K. Vadas, N. Patel, K. Lyons, T. Starner, and J. Jacko, “Reading on-the-go: A comparison of audio and hand-held displays,” in *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI ’06, (New York, NY, USA), pp. 219–226, ACM, 2006.

[75] L. Barnard, J. S. Yi, J. A. Jacko, and A. Sears, “Capturing the effects of context on human performance in mobile computing systems,” *Personal Ubiquitous Comput.*, vol. 11, pp. 81–96, Jan. 2007.

[76] M. Crease, J. Lumsden, and B. Longworth, “A technique for incorporating dynamic paths in lab-based mobile evaluations,” in *Proc. of the 21st British HCI Group Annual Conf. on People and Computers: HCI...But Not As We Know It - Volume 1*, BCS-HCI ’07, (Swinton, UK, UK), pp. 99–108, British Computer Society, 2007.

[77] E. Hoggan and S. Brewster, “Designing audio and tactile crossmodal icons for mobile devices,” in *Proc. of the 9th Int. Conf. on Multimodal Interfaces*, ICMI ’07, (New York, NY, USA), pp. 162–169, ACM, 2007.

[78] M. Lin, R. Goldman, K. J. Price, A. Sears, and J. Jacko, “How do people tap when walking? an empirical investigation of nomadic data entry,” *Int. J. Hum.-Comput. Stud.*, vol. 65, pp. 759–769, Sept. 2007.

[79] Y. J. Choi, “How changes in context affect human behaviours and performance in mobile computing systems,” *WSEAS Trans. Sys. Ctrl.*, vol. 3, pp. 664–675, Aug. 2008.

[80] Y. J. Choi, “Investigation of the effects of context on users’ behaviors and performance in mobile computing systems,” in *Proceedings of the 12th WSEAS International Conference on Systems*, ICS’08, (Stevens Point, Wisconsin, USA), pp. 675–680, World Scientific and Engineering Academy and Society (WSEAS), 2008.

[81] A. Crossan, J. Williamson, S. Brewster, and R. Murray-Smith, "Wrist rotation for interaction in mobile contexts," in *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '08, (New York, NY, USA), pp. 435–438, ACM, 2008.

[82] L. Dai, A. Sears, and R. Goldman, "Shifting the focus from accuracy to recallability: A study of informal note-taking on mobile information technologies," *ACM Trans. Comp.-Hum. Interact.*, vol. 16, pp. 4:1–4:46, Apr. 2009.

[83] H. Lü and Y. Li, "Gesture avatar: A technique for operating mobile user interfaces using gestures," in *Proc. of the SIGCHI Conf. on Human Factors in Comp. Systems*, CHI '11, (New York, NY, USA), pp. 207–216, ACM, 2011.

[84] G. Wilson, S. A. Brewster, M. Halvey, A. Crossan, and C. Stewart, "The effects of walking, feedback and control method on pressure-based interaction," in *Proc. of the 13th Int. Conf. on Hum. Comp. Interact. with Mobile Devices and Services*, MobileHCI '11, (New York, NY, USA), pp. 147–156, ACM, 2011.

[85] B. van Tonder and J. Wesson, "The impact of sensor fusion on tilt interaction in a mobile map-based application," in *Proceedings of the South African Institute of Computer Scientists and Information Technologists Conference on Knowledge, Innovation and Leadership in a Diverse, Multidisciplinary Environment*, SAICSIT '11, (New York, NY, USA), pp. 249–258, ACM, 2011.

[86] H. Nicolau and J. Jorge, "Touch typing using thumbs: Understanding the effect of mobility and hand posture," in *Proc. of the SIGCHI Conf. on Human Factors in Comp. Systems*, CHI '12, (New York, NY, USA), pp. 2683–2686, ACM, 2012.

[87] L. Chan, R.-H. Liang, M.-C. Tsai, K.-Y. Cheng, C.-H. Su, M. Y. Chen, W.-H. Cheng, and B.-Y. Chen, "Fingerpad: Private and subtle interaction using fingertips," in *Proc. of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, (New York, NY, USA), pp. 255–260, ACM, 2013.

[88] Y. Fukatsu, B. Shizuki, and J. Tanaka, "No-look flick: Single-handed and eyes-free japanese text input system on touch screens of mobile devices," in *Proc. of*

the 15th Int. Conf. on Human-computer Interaction with Mobile Devices and Services, MobileHCI '13, (New York, NY, USA), pp. 161–170, ACM, 2013.

[89] T. Mustonen, M. Berg, J. Kaistinen, T. Kawai, and J. Häkkinen, "Visual task performance using a monocular see-through head-mounted display (hmd) while walking," *Journal of Experimental Psychology: Applied*, vol. 19, no. 4, pp. 333–344, 2013.

[90] J. Clawson, T. Starner, D. Kohlsdorf, D. P. Quigley, and S. Gilliland, "Texting while walking: An evaluation of mini-qwerty text input while on-the-go," in *Proc. of the 16th Int. Conf. on Human-computer Interaction with Mobile Devices & Services*, MobileHCI '14, (New York, NY, USA), pp. 339–348, ACM, 2014.

[91] F. Heller, S. Ivanov, C. Wacharamanotham, and J. Borchers, "Fabritouch: Exploring flexible touch input on textiles," in *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, ISWC '14, (New York, NY, USA), pp. 59–62, ACM, 2014.

[92] H. Nicolau, T. Guerreiro, D. Lucas, and J. Jorge, "Mobile text-entry and visual demands: reusing and optimizing current solutions," *Universal Access in the Information Society*, vol. 13, pp. 291–301, Aug 2014.

[93] D. Dobbelstein, C. Winkler, G. Haas, and E. Rukzio, "Pocketthumb: A wearable dual-sided touch interface for cursor-based control of smart-eyewear," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, pp. 9:1–9:17, June 2017.

[94] M. Harvey and M. Pointon, "Perceptions of the effect of fragmented attention on mobile web search tasks," in *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, CHIIR '17, (New York, NY, USA), pp. 293–296, ACM, 2017.

[95] M. Harvey and M. Pointon, "Searching on the go: The effects of fragmented attention on mobile web search tasks," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, (New York, NY, USA), pp. 155–164, ACM, 2017.

[96] F. Mo, J. Zhou, and S. Yi, "Adapting the navigation interface of smart watches to user movements," *International Journal of Human–Computer Interaction*, vol. 33, no. 6, pp. 460–474, 2017.

[97] E. Hoggan, S. A. Brewster, and J. Johnston, "Investigating the effectiveness of tactile feedback for mobile touchscreens," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, (New York, NY, USA), pp. 1573–1582, ACM, 2008.

[98] S. Fitchett and A. Cockburn, "Multiscroll: Using multitouch input to disambiguate relative and absolute mobile scroll modes," in *Proceedings of the 24th BCS Interaction Specialist Group Conference*, BCS '10, (Swinton, UK, UK), pp. 393–402, British Computer Society, 2010.

[99] M. Harvey and M. Pointon, "Understanding in-context interaction: An investigation into on-the-go mobile search," *Information Processing & Management*, vol. 56, no. 6, p. 102089, 2019.

[100] R. Rzayev, P. W. Woźniak, T. Dingler, and N. Henze, "Reading on smart glasses: The effect of text position, presentation type and walking," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, (New York, NY, USA), pp. 45:1–45:9, ACM, 2018.

[101] A. Chamberlain and R. Kalawsky, "A comparative investigation into two pointing systems for use with wearable computers while mobile," in *Proceedings of the Eighth International Symposium on Wearable Computers*, ISWC '04, (Washington, DC, USA), pp. 110–117, IEEE Computer Society, 2004.

[102] S. Mizobuchi, M. Chignell, and D. Newton, "Mobile text entry: Relationship between walking speed and text input task difficulty," in *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services*, MobileHCI '05, (New York, NY, USA), pp. 122–128, ACM, 2005.

[103] G. N. Marentakis and S. A. Brewster, "Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane,"

in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, (New York, NY, USA), pp. 359–368, ACM, 2006.

[104] M. Schedelbauer and J. Heines, "Selecting while walking: An investigation of aiming performance in a mobile work context," in *Proceedings of the 13th Americas Conference on Information Systems*, AMCIS '07, (Atlanta, GA), p. 433, Association for Information Systems, 2007.

[105] D. L. Ashbrook, J. R. Clawson, K. Lyons, T. E. Starner, and N. Patel, "Quickdraw: The impact of mobility and on-body placement on device access time," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, (New York, NY, USA), pp. 219–222, ACM, 2008.

[106] S. K. Kane, J. O. Wobbrock, and I. E. Smith, "Getting off the treadmill: Evaluating walking user interfaces for mobile devices in public spaces," in *Proc. of the 10th Int. Conf. on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '08, (New York, NY, USA), pp. 109–118, ACM, 2008.

[107] A. K. Karlson and B. B. Bederson, "One-handed touchscreen input for legacy applications," in *Proc. of the SIGCHI Conf. on Human Factors in Comp. Systems*, CHI '08, (New York, NY, USA), pp. 1399–1408, ACM, 2008.

[108] K. B. Perry and J. P. Hourcade, "Evaluating one handed thumb tapping on mobile touchscreen devices," in *Proceedings of Graphics Interface 2008*, GI '08, (Toronto, Ont., Canada, Canada), pp. 57–64, Canadian Information Processing Society, 2008.

[109] A. Crossan, M. McGill, S. Brewster, and R. Murray-Smith, "Head tilting for interaction in mobile contexts," in *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, (New York, NY, USA), pp. 6:1–6:10, ACM, 2009.

[110] J. Bergstrom-Lehtovirta, A. Oulasvirta, and S. Brewster, "The effects of walking speed on target acquisition on a touchscreen interface," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, (New York, NY, USA), pp. 143–146, ACM, 2011.

[111] A. Ng, S. Brewster, and A. Crossan, "The effects of encumbrance on mobile gesture interactions," 2011.

[112] J. J. Lim and C. Feria, "Visual search on a mobile device while walking," in *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '12, (New York, NY, USA), pp. 295–304, ACM, 2012.

[113] A. Ng, S. A. Brewster, and J. Williamson, "The impact of encumbrance on mobile interactions," in *Human-Computer Interaction – INTERACT 2013* (P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, eds.), (Berlin, Heidelberg), pp. 92–109, Springer Berlin Heidelberg, 2013.

[114] J. Conradi, O. Busch, and T. Alexander, "Optimal touch button size for the use of mobile devices while walking," *Procedia Manufacturing*, vol. 3, pp. 387 – 394, 2015. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.

[115] D. Dobbelstein, G. Haas, and E. Rukzio, "The effects of mobility, encumbrance, and (non-)dominant hand on interaction with smartwatches," in *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, ISWC '17, (New York, NY, USA), pp. 90–93, ACM, 2017.

[116] K. Sun, Y. Wang, C. Yu, Y. Yan, H. Wen, and Y. Shi, "Float: One-handed and touch-free target selection on smartwatches," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, (New York, NY, USA), pp. 692–704, ACM, 2017.

[117] T. Yamabe and K. Takahashi, "Experiments in mobile user interface adaptation for walking users," in *Proc. of the The 2007 Int. Conf. on Intel. Pervasive Comp.*, IPC '07, (Washington, DC, USA), pp. 280–284, IEEE Comp. Society, 2007.

[118] G. Singh, W. Delamare, and P. Irani, "D-swime: A design space for smartwatch interaction techniques supporting mobility and encumbrance," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, (New York, NY, USA), pp. 634:1–634:13, ACM, 2018.

[119] L. Barnard, J. S. Yi, J. A. Jacko, and A. Sears, “An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices,” *Int. J. Hum.-Comput. Stud.*, vol. 62, pp. 487–520, Apr. 2005.

[120] J. A. Feld and P. Plummer, “Visual scanning behavior during distracted walking in healthy young adults,” *Gait & Posture*, vol. 67, pp. 219 – 223, 2019.

[121] A. Mariakakis, M. Goel, M. T. I. Aumi, S. N. Patel, and J. O. Wobbrock, “Switchback: Using focus and saccade tracking to guide users’ attention for mobile task resumption,” in *Proc. of the 33rd Annual ACM Conf. on Human Factors in Comp. Systems*, CHI ’15, (New York, NY, USA), pp. 2953–2962, ACM, 2015.

[122] T. Roumen, S. T. Perrault, and S. Zhao, “Notiring: A comparative study of notification channels for wearable interactive rings,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15, (New York, NY, USA), pp. 2497–2500, ACM, 2015.

[123] E. Hoggan, A. Crossan, S. A. Brewster, and T. Kaaresoja, “Audio or tactile feedback: Which modality when?,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’09, (New York, NY, USA), pp. 2253–2256, ACM, 2009.

[124] A. Pirhonen, S. Brewster, and C. Holguin, “Gestural and audio metaphors as a means of control for mobile devices,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’02, (New York, NY, USA), pp. 291–298, ACM, 2002.

[125] B. P. Van Tonder and J. L. Wesson, “Improving the controllability of tilt interaction for mobile map-based applications,” *Int. J. Hum.-Comput. Stud.*, vol. 70, pp. 920–935, Dec. 2012.

[126] S. Robinson, P. Eslambolchilar, and M. Jones, “Evaluating haptics for information discovery while walking,” in *Proc. of the 23rd British HCI Group Annual Conf. on People and Computers: Celebrating People and Technology*, BCS-HCI ’09, (Swinton, UK, UK), pp. 93–102, British Computer Society, 2009.

[127] K. Yatani and K. N. Truong, "An evaluation of stylus-based text entry methods on handheld devices studied in different user mobility states," *Pervasive and Mobile Computing*, vol. 5, no. 5, pp. 496 – 508, 2009.

[128] K. Yatani and K. N. Truong, "An evaluation of stylus-based text entry methods on handheld devices in stationary and mobile settings," in *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '07, (New York, NY, USA), pp. 487–494, ACM, 2007.

[129] J. D. Hincapié-Ramos and P. Irani, "Crashalert: Enhancing peripheral alertness for eyes-busy mobile interaction while walking," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, (New York, NY, USA), pp. 3385–3388, ACM, 2013.

[130] J. E. Zucco, B. H. Thomas, and K. Grimmer, "Evaluation of four wearable computer pointing devices for drag and drop tasks when stationary and walking," in *2006 10th IEEE International Symposium on Wearable Computers*, (Washington, DC, USA), pp. 29–36, IEEE Computer Society, Oct 2006.

[131] E. R. Hooten, S. T. Hayes, and J. A. Adams, "Communicative modalities for mobile device interaction," *Int. J. Hum.-Comput. Stud.*, vol. 71, pp. 988–1002, Oct. 2013.

[132] S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker, "Multimodal 'eyes-free' interaction techniques for wearable devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, (New York, NY, USA), pp. 473–480, ACM, 2003.

[133] M. Pielot, A. Kazakova, T. Hesselmann, W. Heuten, and S. Boll, "Pocketmenu: Non-visual menus for touch screen devices," in *Proc. of the 14th Int. Conf. on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '12, (New York, NY, USA), pp. 327–330, ACM, 2012.

[134] M. Pielot, B. Poppinga, W. Heuten, and S. Boll, "6th senses for everyone!: The value of multimodal feedback in handheld navigation aids," in *Proceedings of*

the 13th International Conference on Multimodal Interfaces, ICMI '11, (New York, NY, USA), pp. 65–72, ACM, 2011.

[135] A. Lucero and A. Vetek, "Notifeye: Using interactive glasses to deal with notifications while walking in public," in *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, ACE '14, (New York, NY, USA), pp. 17:1–17:10, ACM, 2014.

[136] M. Pielot, B. Poppinga, W. Heuten, and S. Boll, "A tactile compass for eyes-free pedestrian navigation," in *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II*, INTERACT'11, (Berlin, Heidelberg), pp. 640–656, Springer-Verlag, 2011.

[137] S. Schneegass and A. Voit, "Gesturesleeve: Using touch sensitive fabrics for gestural input on the forearm for controlling smartwatches," in *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, ISWC '16, (New York, NY, USA), pp. 108–115, ACM, 2016.

[138] T. Yamabe, K. Takahashi, and T. Nakajima, "Design issues and an empirical study in mobility oriented service development," in *Proceedings of the 1st Workshop on Mobile Middleware: Embracing the Personal Communication Device*, MobMid '08, (New York, NY, USA), pp. 1:1–1:6, ACM, 2008.

[139] T. Yamabe, K. Takahashi, and T. Nakajima, "Towards mobility oriented interaction design: Experiments in pedestrian navigation on mobile devices," in *Proceedings of the 5th Annual International Conf. on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, Mobiquitous '08, (ICST, Brussels, Belgium), pp. 46:1–46:10, ICST, 2008.

[140] M. Iwata, H. Miyamoto, T. Hara, D. Komaki, K. Shimatani, T. Mashita, K. Kiyokawa, T. Uemukai, G. Hattori, S. Nishio, and H. Takemura, "A content search system considering the activity and context of a mobile user," *Personal Ubiquitous Comput.*, vol. 17, pp. 1035–1050, June 2013.

[141] D. C. Schwebel, D. Stavrinos, K. W. Byington, T. Davis, E. E. O'Neal, and D. de Jong, "Distraction and pedestrian safety: How talking on the phone,

texting, and listening to music impact crossing the street," *Accident Analysis & Prevention*, vol. 45, pp. 266 – 271, 2012.

[142] D. Stavrinos, K. W. Byington, and D. C. Schwebel, "Distracted walking: Cell phones increase injury risk for college pedestrians," *Journal of Safety Research*, vol. 42, no. 2, pp. 101 – 107, 2011.

[143] J. R. Williamson, S. Brewster, and R. Vennelakanti, "Mo!games: Evaluating mobile gestures in the wild," in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI '13, (New York, NY, USA), pp. 173–180, ACM, 2013.

[144] S. Reyal, S. Zhai, and P. O. Kristensson, "Performance and user experience of touchscreen and gesture keyboards in a lab setting and in the wild," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, (New York, NY, USA), pp. 679–688, ACM, 2015.

[145] J. R. Wiliamson, A. Crossan, and S. Brewster, "Multimodal mobile interactions: Usability studies in real world settings," in *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI '11, (New York, NY, USA), pp. 361–368, ACM, 2011.

[146] M. Aliannejadi, M. Harvey, L. Costa, M. Pointon, and F. Crestani, "Understanding mobile search task relevance and user behaviour in context," in *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, CHIIR '19, (New York, NY, USA), pp. 143–151, ACM, 2019.

[147] M. A. Timmis, H. Bijl, K. Turner, I. Basevitch, M. J. D. Taylor, and K. N. van Paridon, "The impact of mobile phone use on where we look and how we walk when negotiating floor based obstacles," *PLOS ONE*, vol. 12, pp. 1–20, 06 2017.

[148] A. Bragdon, E. Nelson, Y. Li, and K. Hinckley, "Experimental analysis of touch-screen gesture designs in mobile environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, (New York, NY, USA), pp. 403–412, ACM, 2011.

[149] E. M. Lamberg and L. M. Muratori, "Cell phones change the way we walk," *Gait & Posture*, vol. 35, no. 4, pp. 688 – 690, 2012.

[150] A. S. Arif, B. Iltisberger, and W. Stuerzlinger, "Extending mobile user ambient awareness for nomadic text entry," in *Proceedings of the 23rd Australian Computer-Human Interaction Conference*, OzCHI '11, (New York, NY, USA), pp. 21–30, ACM, 2011.

[151] A. Ng, J. Williamson, and S. Brewster, "The effects of encumbrance and mobility on touch-based gesture interactions for mobile phones," in *Proceedings of the 17th Int. Conf. on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, (New York, NY, USA), pp. 536–546, ACM, 2015.

[152] V. Agostini, F. Lo Fermo, G. Massazza, and M. Knaflitz, "Does texting while walking really affect gait in young adults?," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, p. 86, Sep 2015.

[153] M. Jain and R. Balakrishnan, "User learning and performance with bezel menus," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 2221–2230, ACM, 2012.

[154] L. Jongil, C. S. Ho, L. Jihyun, and K. Kijeong, "Effects of smartphone texting on the visual perception and dynamic walking stability," *J Exerc Rehabil*, vol. 13, no. 1, pp. 48–54, 2017.

[155] S. Azenkot, C. L. Bennett, and R. E. Ladner, "Digitaps: Eyes-free number entry on touchscreens with minimal audio feedback," in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, (New York, NY, USA), pp. 85–90, ACM, 2013.

[156] Y. Wang, C. Yu, J. Liu, and Y. Shi, "Understanding performance of eyes-free, absolute position control on touchable mobile phones," in *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, (New York, NY, USA), pp. 79–88, ACM, 2013.

[157] J. Conradi, "Influence of letter size on word reading performance during walking," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, (New York, NY, USA), pp. 16:1–16:9, ACM, 2017.

[158] P. Crowley, P. Madeleine, and N. Vuillerme, "The effects of mobile phone use on walking: a dual task study," *BMC Research Notes*, vol. 12, no. 1, pp. 352–357, 2019.

[159] A. Ng, S. A. Brewster, and J. H. Williamson, "Investigating the effects of encumbrance on one- and two- handed interactions with mobile devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, (New York, NY, USA), pp. 1981–1990, ACM, 2014.

[160] K. A. Li, P. Baudisch, and K. Hinckley, "Blindsight: Eyes-free access to mobile phones," in *Proc. of the SIGCHI Conf. on Human Factors in Comp. Systems*, CHI '08, (New York, NY, USA), pp. 1389–1398, ACM, 2008.

[161] S. Fitchett and A. Cockburn, "Evaluating reading and analysis tasks on mobile devices: A case study of tilt and flick scrolling," in *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, OZCHI '09, (New York, NY, USA), pp. 225–232, ACM, 2009.

[162] P. Chopra, D. Castelli, and J. Dingwell, "Cognitively demanding object negotiation while walking and texting," *Scientific Reports*, vol. 8, 12 2018.

[163] Y. Tian, Y. Huang, J. He, and K. Wei, "What affects gait performance during walking while texting? a comparison of motor, visual and cognitive factors," *Ergonomics*, vol. 61, no. 11, pp. 1507–1518, 2018.

[164] S. M. Lopresti-Goodman, A. Rivera, and C. Dressel, "Practicing safe text: the impact of texting on walking behavior," *Applied Cognitive Psychology*, vol. 26, no. 4, pp. 644–648, 2012.

[165] S. M. Schabrun, W. van den Hoorn, A. Moorcroft, C. Greenland, and P. W. Hodges, "Texting and walking: Strategies for postural control and implications for safety," *PLOS ONE*, vol. 9, pp. 1–8, 01 2014.

[166] S. G. Gustafson, B. Rabe, and P. M. Baudisch, "Understanding palm-based imaginary interfaces: The role of visual and tactile cues when browsing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, (New York, NY, USA), pp. 889–898, ACM, 2013.

[167] T. Davies and A. Beeharee, "The case of the missed icon: Change blindness on mobile devices," in *Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 1451–1460, ACM, 2012.

[168] A. Oulasvirta and J. Bergstrom-Lehtovirta, "Ease of juggling: Studying the effects of manual multitasking," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, (New York, NY, USA), pp. 3103–3112, ACM, 2011.

[169] J. Lumsden and P. Drost, "A comparison of the impact of avoidance cues in hazard avoidance during evaluation of text entry," in *Proc. of the 22Nd British HCI Group Annual Conf. on People and Computers: Culture, Creativity, Interaction - Vol. 2*, BCS-HCI '08, (Swindon, UK), pp. 57–60, BCS Learning & Development Ltd., 2008.

[170] T. Yang, M. Ferati, Y. Liu, R. Rohani Ghahari, and D. Bolchini, "Aural browsing on-the-go: Listening-based back navigation in large web architectures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 277–286, ACM, 2012.

[171] H. Tinwala and I. S. MacKenzie, "Eyes-free text entry on a touchscreen phone," in *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, (New York), pp. 83–88, IEEE, Sept 2009.

[172] K. Yatani and K. N. Truong, "Semfeel: A user interface with semantic tactile feedback for mobile touch-screen devices," in *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, (New York, NY, USA), pp. 111–120, ACM, 2009.

[173] A. Oulasvirta and J. Bergstrom-Lehtovirta, "A simple index for multimodal

flexibility," in *Proc. of the SIGCHI Conf. on Human Factors in Comp. Systems*, CHI '10, (New York, NY, USA), pp. 1475–1484, ACM, 2010.

[174] T. Pakkanen, R. Raisamo, K. Salminen, and V. Surakka, "Haptic numbers: Three haptic representation models for numbers on a touch screen phone," in *Int. Conf. on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10, (New York, NY, USA), pp. 35:1–35:4, ACM, 2010.

[175] H. Tinwala and I. S. MacKenzie, "Eyes-free text entry with error correction on touchscreen mobile devices," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, (New York, NY, USA), pp. 511–520, ACM, 2010.

[176] M. Negulescu, J. Ruiz, Y. Li, and E. Lank, "Tap, swipe, or move: Attentional demands for distracted smartphone input," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, (New York, NY, USA), pp. 173–180, ACM, 2012.

[177] O. Metatla, F. Martin, T. Stockman, and N. Bryan-Kinns, "Non-visual menu navigation: The effect of an audio-tactile display," in *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - Holiday HCI*, BCS-HCI '14, (UK), pp. 213–217, BCS, 2014.

[178] Y. Kubo, B. Shizuki, and J. Tanaka, "B2b-swipe: Swipe gesture for rectangular smartwatches from a bezel to a bezel," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, (New York, NY, USA), pp. 3852–3856, ACM, 2016.

[179] M. Seuter, M. Pfeiffer, G. Bauer, K. Zentgraf, and C. Kray, "Running with technology: Evaluating the impact of interacting with wearable devices on running movement," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, pp. 101:1–101:17, Sept. 2017.

[180] J. Alexander, T. Han, W. Judd, P. Irani, and S. Subramanian, "Putting your best foot forward: Investigating real-world mappings for foot-based gestures,"

in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 1229–1238, ACM, 2012.

[181] T. Han, J. Alexander, A. Karnik, P. Irani, and S. Subramanian, "Kick: Investigating the use of kick gestures for mobile interactions," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, (New York, NY, USA), pp. 29–32, ACM, 2011.

[182] H. W. Bristow, C. Baber, J. Cross, J. F. Knight, and S. I. Woolley, "Defining and evaluating context for wearable computing," *International Journal of Human-Computer Studies*, vol. 60, no. 5, pp. 798 – 819, 2004. HCI Issues in Mobile Computing.

[183] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload* (P. A. Hancock and N. Meshkati, eds.), vol. 52 of *Advances in Psychology*, pp. 139 – 183, Amsterdam: North-Holland, 1988.

[184] B. van Tonder and J. Wesson, "Intellitilt: An enhanced tilt interaction technique for mobile map-based applications," in *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II*, INTERACT'11, (Berlin, Heidelberg), pp. 505–523, Springer-Verlag, 2011.

[185] Y. Zhou, T. Xu, B. David, and R. Chalon, "Interaction on-the-go: A fine-grained exploration on wearable procam interfaces and gestures in mobile situations," *Univers. Access Inf. Soc.*, vol. 15, pp. 643–657, Nov. 2016.

[186] H. Tu, X. Ren, and S. Zhai, "Differences and similarities between finger and pen stroke gestures on stationary and mobile devices," *ACM Trans. Comput.-Hum. Interact.*, vol. 22, pp. 22:1–22:39, Aug. 2015.

[187] S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch, "Earpod: Eyes-free menu selection using touch input and reactive audio feedback," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, (New York, NY, USA), pp. 1395–1404, ACM, 2007.

[188] Q. Roy, S. T. Perrault, S. Zhao, R. C. Davis, A. Pattena Vaniyar, V. Vechev, Y. Lee, and A. Misra, “Follow-my-lead: Intuitive indoor path creation and navigation using interactive videos,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI ’17, (New York, NY, USA), pp. 5703–5715, ACM, 2017.

[189] D. Wenig, A. Steenbergen, J. Schöning, B. Hecht, and R. Malaka, “Scrollinghome: Bringing image-based indoor navigation to smartwatches,” in *Proceedings of the 18th Int. Conf. on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI ’16, (New York, NY, USA), pp. 400–406, ACM, 2016.

[190] A. Mulloni, H. Seichter, and D. Schmalstieg, “Handheld augmented reality indoor navigation with activity-based instructions,” in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI ’11, (New York, NY, USA), pp. 211–220, ACM, 2011.

[191] S. Robinson, M. Jones, P. Eslambolchilar, R. Murray-Smith, and M. Lindborg, “"i did it my way": Moving away from the tyranny of turn-by-turn pedestrian navigation,” in *Proc. of the 12th Int. Conf. on Human Comp. Interact. with Mob. Devices and Services*, MobileHCI ’10, (New York, NY, USA), pp. 341–344, ACM, 2010.

[192] S. Burigat and L. Chittaro, “Pedestrian navigation with degraded gps signal: Investigating the effects of visualizing position uncertainty,” in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI ’11, (New York, NY, USA), pp. 221–230, ACM, 2011.

[193] S. T. Perrault, E. Lecolinet, J. Eagan, and Y. Guiard, “Watchit: Simple gestures and eyes-free interaction for wristwatches and bracelets,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’13, (New York, NY, USA), pp. 1451–1460, ACM, 2013.

[194] J. Lumsden and S. Brewster, “A paradigm shift: Alternative interaction techniques for use with mobile & wearable devices,” in *Proceedings of the 2003*

Conf. of the Centre for Advanced Studies on Collaborative Research, CASCON '03, (Indianapolis, IN), pp. 197–210, IBM Press, 2003.

[195] D. Ghosh, P. S. Foong, S. Zhao, D. Chen, and M. Fjeld, "Editalk: Towards designing eyes-free interactions for mobile word processing," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, (New York, NY, USA), pp. 403:1–403:10, ACM, 2018.

[196] S. Heo and G. Lee, "Forcetap: Extending the input vocabulary of mobile touch screens by adding tap gestures," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, (New York, NY, USA), pp. 113–122, ACM, 2011.

[197] M. Ciman and K. Wac, "Individuals' stress assessment using human - smartphone interaction analysis," *IEEE Transactions on Affective Computing*, vol. 9, pp. 51–65, Jan 2018.

[198] T. Reis, L. Carriço, and C. Duarte, "Mobile interaction: Automatically adapting audio output to users and contexts on communication and media control scenarios," in *Universal Access in Human-Computer Interaction. Intelligent and Ubiq. Interaction Environments* (C. Stephanidis, ed.), (Berlin, Heidelberg), pp. 384–393, Springer, 2009.

[199] J. Mäntyjärvi and T. Seppänen, "Adapting applications in handheld devices using fuzzy context information," *Interacting with Computers*, vol. 15, no. 4, pp. 521 – 538, 2003.

[200] A. Möller, M. Kranz, S. Diewald, L. Roalter, R. Huitl, T. Stockinger, M. Koelle, and P. A. Lindemann, "Experimental evaluation of user interfaces for visual indoor navigation," in *Proc. of the SIGCHI Conf. on Human Factors in Comp. Systems*, CHI '14, (New York, NY, USA), pp. 3607–3616, ACM, 2014.

[201] S. Bertel, T. Dressel, T. Kohlberg, and V. von Jan, "Spatial knowledge acquired from pedestrian urban navigation systems," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, (New York, NY, USA), pp. 32:1–32:6, ACM, 2017.

[202] S. Mayer, L. Lischke, P. W. Woźniak, and N. Henze, "Evaluating the disruptiveness of mobile interactions: A mixed-method approach," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, (New York, NY, USA), pp. 406:1–406:14, ACM, 2018.

[203] R. Murray-Smith, A. Ramsay, S. Garrod, M. Jackson, and B. Musizza, "Gait alignment in mobile phone conversations," in *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '07, (New York, NY, USA), pp. 214–221, ACM, 2007.

[204] S. Wakamiya, H. Kawasaki, Y. Kawai, A. Jatowt, E. Aramaki, and T. Akiyama, "Lets not stare at smartphones while walking: Memorable route recommendation by detecting effective landmarks," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, (New York, NY, USA), pp. 1136–1146, ACM, 2016.

[205] A. Roudaut, A. Rau, C. Sterz, M. Plauth, P. Lopes, and P. Baudisch, "Gesture output: Eyes-free output using a force feedback touch surface," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, (New York, NY, USA), pp. 2547–2556, ACM, 2013.

[206] T. Han, X. Xiao, L. Shi, J. Canny, and J. Wang, "Balancing accuracy and fun: Designing camera based mobile games for implicit heart rate monitoring," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, (New York, NY, USA), pp. 847–856, ACM, 2015.

[207] M. Ciman, K. Wac, and O. Gaggi, "isensestress: Assessing stress through human-smartphone interaction analysis," in *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '15, (ICST, Brussels, Belgium, Belgium), pp. 84–91, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2015.

[208] F. Ioannidou, F. Hermens, and T. L. Hodgson, "Mind your step: the effects of mobile phone use on gaze behavior in stair climbing," *Journal of Technology in Behavioral Science*, vol. 2, pp. 109–120, Dec 2017.

[209] H. Han and G. Shin, “Head flexion angle when web-browsing and texting using a smartphone while walking,” *Applied Ergonomics*, vol. 81, p. 102884, 2019.

[210] C. Ranasinghe, S. Heitmann, A. Hamzin, M. Pfeiffer, and C. Kray, “Pedestrian navigation and gps deteriorations: User behavior and adaptation strategies,” in *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, OzCHI ’18, (New York, NY, USA), pp. 266–277, ACM, 2018.

[211] A. J. Piazza, A. P. Knowlden, E. Hibberd, J. Leeper, A. M. Paschal, and S. Usdan, “Mobile device use while crossing the street: Utilizing the theory of planned behavior,” *Accident Analysis & Prevention*, vol. 127, pp. 9 – 18, 2019.

[212] J. Wobbrock, “Measures of text entry performance,” in *Text Entry Systems* (K. Tanaka-Ishii and I. S. MacKenzie, eds.), ch. 3, pp. 47–74, San Francisco: Morgan Kaufmann, 2007.

[213] R. W. Soukoreff and I. S. MacKenzie, “Metrics for text entry research: An evaluation of msd and kspc, and a new unified error metric,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’03, (New York, NY, USA), p. 113–120, ACM, 2003.

[214] I. Durham, D. A. Lamb, and J. B. Saxe, “Spelling correction in user interfaces,” *Commun. ACM*, vol. 26, p. 764–773, Oct. 1983.

[215] T. Chen, Y. Yesilada, and S. Harper, “What input errors do you experience? typing and pointing errors of mobile web users,” *International Journal of Human-Computer Studies*, vol. 68, no. 3, pp. 138–157, 2010.

[216] K. K. Greene, M. A. Gallagher, B. C. Stanton, and P. Y. Lee, “I can’t type that! p@$$w0rd entry on mobile devices,” in *Proceedings of the Second International Conference on Human Aspects of Information Security, Privacy, and Trust - Volume 8533*, (Berlin, Heidelberg), p. 160–171, Springer-Verlag, 2014.

[217] S. K. Chan, B. He, and I. Ounis, “An in-depth study of the automatic detection and correction of spelling mistakes,” in *Proc. of the 5th Dutch-Belgian Information Retrieval Workshop (DIR)*, (Utrecht, the Netherlands), pp. 71–81, Utrecht University, 2005.

[218] J. Head, P. N. Russell, M. J. Dorahy, E. Neumann, and W. S. Helton, "Text-speak processing and the sustained attention to response task," *Experimental Brain Research*, vol. 216, pp. 103–111, Jan 2012.

[219] L. V. Subramaniam, S. Roy, T. A. Faruquie, and S. Negi, "A survey of types of text noise and techniques to handle noisy text," in *Proc. of the Third Workshop on Analytics for Noisy Unstructured Text Data*, AND '09, (NY, USA), p. 115–122, ACM, 2009.

[220] F.-J. Forbes and E. Buchanan, ""textisms": The comfort of the recipient.," *Psychology of popular media culture*, vol. 8, pp. 358–364, 2019.

[221] S. H. Hsieh and T. H. Tseng, "Playfulness in mobile instant messaging: Examining the influence of emoticons and text messaging on social interaction," *Computers in Human Behavior*, vol. 69, pp. 405 – 414, 2017.

[222] M. Choudhury, R. Saraf, V. Jain, A. Mukherjee, S. Sarkar, and A. Basu, "Investigation and modeling of the structure of texting language," *Int. Journal of Document Analysis and Recognition*, vol. 10, pp. 157–174, 2007.

[223] L. D. Rosen, J. Chang, L. Erwin, L. M. Carrier, and N. A. Cheever, "The relationship between "textisms" and formal and informal writing among young adults," *Communication Research*, vol. 37, no. 3, pp. 420–440, 2010.

[224] C. Varnhagen, G. P. McFall, N. Pugh, L. Zederayko(Routledge), H. Sumida-MacDonald, and T. Kwong, "Lol: New language and spelling in instant messaging," *Reading and Writing*, vol. 23, pp. 719–733, 07 2010.

[225] J. O. Wobbrock and B. A. Myers, "Analyzing the input stream for character-level errors in unconstrained text entry evaluations," *ACM Trans. Comput.-Hum. Interact.*, vol. 13, p. 458–489, dec 2006.

[226] Z. Sarsenbayeva, N. van Berkel, W. Jiang, D. Hettiachchi, V. Kostakos, and J. Goncalves, "Effect of ambient light on mobile interaction," in *Human-Computer Interaction – INTERACT 2019* (D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, and P. Zaphiris, eds.), (Cham), pp. 465–475, Springer International Publishing, 2019.

[227] R. Schlögl, C. Wimmer, and T. Grechenig, "Hyper typer: A serious game for measuring mobile text entry performance in the wild," in *Extended Abstracts of the 2019 CHI Conf. on Human Factors in Computing Systems*, CHI EA '19, (NY, USA), p. 1–6, ACM, 2019.

[228] C. Wimmer, R. Schlögl, K. Kappel, and T. Grechenig, "Measuring mobile text entry performance and behaviour in the wild with a serious game," in *Proc. of the 18th Int. Conf. on Mobile and Ubiquitous Multimedia*, MUM '19, (NY, USA), ACM, 2019.

[229] A. Komninos, M. Dunlop, K. Katsaris, and J. Garofalakis, "A glimpse of mobile text entry errors and corrective behaviour in the wild," in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '18, (New York, NY, USA), p. 221–228, ACM, 2018.

[230] W. Wong, W. Liu, and M. Bennamoun, "Integrated scoring for spelling error correction, abbreviation expansion and case restoration in dirty text," in *Proc. of the Fifth Australasian Conf. on Data Mining and Analystics - Volume 61*, AusDM '06, (AUS), p. 83–89, Australian Computer Society, Inc., 2006.

[231] A. Battestini, V. Setlur, and T. Sohn, "A large scale study of text-messaging use," in *Proc. of the 12th Int. Conf. on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '10, (NY, USA), p. 229–238, ACM, 2010.

[232] R. Larson and M. Csikszentmihalyi, "The experience sampling method," in *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*, pp. 21–34, Dordrecht: Springer Netherlands, 2014.

[233] D. Ferreira, V. Kostakos, and A. K. Dey, "Aware: Mobile context instrumentation framework," *Frontiers in ICT*, vol. 2, p. 6, 2015.

[234] N. V. Berkel, D. Ferreira, and V. Kostakos, "The experience sampling method on mobile devices," *ACM Comput. Surv.*, vol. 50, pp. 93:1–93:40, Dec. 2017.

[235] M. R. Zhang and J. O. Wobbrock, "Beyond the input stream: Making text entry evaluations more flexible with transcription sequences," in *Proceedings of the*

32nd Annual ACM Symposium on User Interface Software and Technology, UIST '19, (New York, NY, USA), p. 831–842, Association for Computing Machinery, 2019.

[236] G. Eryiğit and D. Torunoğlu-Selamet, "Social media text normalization for turkish," *Natural Language Engineering*, vol. 23, no. 6, p. 835–875, 2017.

[237] B. Han and T. Baldwin, "Lexical normalisation of short text messages: Makn sens a #twitter," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, (USA), p. 368–378, Association for Computational Linguistics, 2011.

[238] L. Németh, "Hunspell." `http://hunspell.github.io/`, 2016. Last accessed 02.10.2021.

[239] B. Say, D. Zeyrek, K. Oflazer, and U. Özge, "Development of a corpus and a treebank for present-day written turkish, (proceedings of the eleventh international conference of turkish linguistics, august, 2002)," in *Current Research in Turkish Lingustics* (K. İmer and G. Doğan, eds.), (Famagusta, North Cyprus), pp. 183–192, Eastern Mediterranean University Press, 2004.

[240] A. A. Akın and M. D. Akın, "Zemberek, an open source nlp framework for turkic languages," *Structure*, vol. 10, pp. 1–5, 2007.

[241] G. Şeker and G. Eryiğit, "Extending a crf-based named entity recognition model for turkish well formed text and user generated content1," *Semantic Web*, vol. 8, pp. 1–18, 01 2017.

[242] S. R. Wilson, W. Magdy, B. McGillivray, and G. Tyson, "Analyzing temporal relationships between trending terms on twitter and urban dictionary activity," in *12th ACM Conference on Web Science*, WebSci '20, (New York, NY, USA), p. 155–163, ACM, 2020.

[243] A. Ucan, B. Naderalvojoud, E. A. Sezer, and H. Sever, "Sentiwordnet for new language: Automatic translation approach," in *2016 12th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, (Washington, DC, USA), pp. 308–315, IEEE, 2016.

[244] M. R. Zhang, H. Wen, and J. O. Wobbrock, “Type, then correct: Intelligent text correction techniques for mobile text entry using neural networks,” in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST ’19, (New York, NY, USA), p. 843–855, ACM, 2019.

[245] A. S. Arif and W. Stuerzlinger, “Predicting the cost of error correction in character-based text entry technologies,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’10, (New York, NY, USA), p. 5–14, ACM, 2010.

[246] P. Bruce, A. Bruce, and P. Gedeck, *Practical statistics for data scientists: 50+ essential concepts using r and python*. Sebastopol, CA: O’Reilly Media, Incorporated, 2020.

[247] M. A. Napierala, “What is the bonferroni correction?.” AAOS Now - American Academy of Orthopaedic Surgeons, Apr 2012.

[248] A. M. Otebolaku and M. T. Andrade, “User context recognition using smartphone sensors and classification models,” *J. Netw. Comput. Appl.*, vol. 66, p. 33–51, May 2016.

[249] D. Ferreira, “Sensors.” `https://awareframework.com/sensors/`, Nov 2016. Last access: 28.06.2022.

[250] A. P. J. Dwiyantoro, I. G. D. Nugraha, and D. Choi, “A simple hierarchical activity recognition system using a gravity sensor and accelerometer on a smartphone,” *International Journal of Technology*, vol. 7, no. 5, 2016.

[251] Y. Li, F. Xue, X. Fan, Z. Qu, and G. Zhou, “Pedestrian walking safety system based on smartphone built-in sensors,” *IET Communications*, vol. 12, no. 6, p. 751 – 758, 2018. Cited by: 11; All Open Access, Bronze Open Access.

[252] A. Jain and V. Kanhangad, “Gender classification in smartphones using gait information,” *Expert Systems with Applications*, vol. 93, pp. 257–266, 2018.

[253] E. Bulbul, A. Cetin, and I. A. Dogru, “Human activity recognition using smartphones,” in *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1–6, 2018.

[254] G. Milette and A. Stroud, *Professional Android Sensor Programming*. GBR: Wrox Press Ltd., 1st ed., 2012.

[255] Google, "Motion sensors | Android developers." `https://developer.android.com/guide/topics/sensors/sensors_motion`, 2016. Last access: 28.06.2022.

[256] M. F. Muthohar, I. G. D. Nugraha, and D. Choi, "Exploring significant motion sensor for energy-efficient continuous motion and location sampling in mobile sensing application," *International Journal of Technology*, vol. 7, no. 1, p. 38 – 49, 2016. Cited by: 1.

[257] H. Ye, T. Gu, X. Tao, and J. Lu, "Scalable floor localization using barometer on smartphone," *Wirel. Commun. Mob. Comput.*, vol. 16, p. 2557–2571, nov 2016.

[258] K. Muralidharan, A. J. Khan, A. Misra, R. K. Balan, and S. Agarwal, "Barometric phone sensors: More hype than hope!," in *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*, HotMobile '14, (New York, NY, USA), Association for Computing Machinery, 2014.

[259] K. Sankaran, M. Zhu, X. F. Guo, A. L. Ananda, M. C. Chan, and L.-S. Peh, "Using mobile phone barometer for low-power transportation context detection," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, SenSys '14, (New York, NY, USA), p. 191–205, ACM, 2014.

[260] L. Yang, Z. Wang, W. Wang, and Q. Zhang, "Naloc: Nonlinear ambient-light-sensor-based localization system," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, Dec. 2018.

[261] H. Martín, A. M. Bernardos, J. Iglesias, and J. R. Casar, "Activity logging using lightweight classification techniques in mobile devices," *Personal Ubiquitous Comput.*, vol. 17, p. 675–695, Apr. 2013.

[262] G. Yuan, Z. Wang, F. Meng, Q. Yan, and S. Xia, "An overview of human activity recognition based on smartphone," *Sensor Review*, vol. 39, 10 2018.

[263] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "On the personalization of classification models for human activity recognition," *IEEE Access*, vol. 8, pp. 32066–32079, 2020.

[264] Z. Gao, D. Liu, K. Huang, and Y. Huang, "Context-aware human activity and smartphone position-mining with motion sensors," *Remote Sensing*, vol. 11, no. 21, 2019.

[265] M. E. ul Haq and M. A. Azam, "Opportunistic sensing for inferring in-the-wild human contexts based on activity pattern recognition using smart computing," *Future Generation Computer Systems*, vol. 106, pp. 374–392, 2020.

[266] Z. Chen, C. Jiang, S. Xiang, J. Ding, M. Wu, and X. Li, "Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 3992–4001, 2020.

[267] J. V. Jeyakumar, L. Lai, N. Suda, and M. Srivastava, "Sensehar: A robust virtual activity sensor for smartphones and wearables," in *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, SenSys '19, (New York, NY, USA), p. 15–28, Association for Computing Machinery, 2019.

[268] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4566–4578, 2016.

[269] D. Figo, P. C. Diniz, D. R. Ferreira, and J. a. M. Cardoso, "Preprocessing techniques for context recognition from accelerometer data," *Personal Ubiquitous Comput.*, vol. 14, p. 645–662, Oct. 2010.

[270] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, "Hand-crafted features vs residual networks for human activities recognition using accelerometer," in *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*, pp. 153–156, 2019.

[271] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package)," *Neurocomputing*, vol. 307, pp. 72–77, 2018.

[272] M. Ehatisham-Ul-Haq, M. A. Azam, Y. Amin, and U. Naeem, "C2fhar: Coarse-to-fine human activity recognition with behavioral context modeling using smart inertial sensors," *IEEE Access*, vol. 8, pp. 7731–7747, 2020.

[273] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[274] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int. Res.*, vol. 16, p. 321–357, jun 2002.

[275] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017.

[276] W. Sousa Lima, E. Souto, K. El-Khatib, R. Jalali, and J. Gama, "Human activity recognition using inertial sensors in a smartphone: An overview," *Sensors*, vol. 19, no. 14, 2019.

[277] I. Jolliffe, "Principal component analysis," in *International Encyclopedia of Statistical Science* (M. Lovric, ed.), pp. 1094–1096, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[278] A. Exler, M. Braith, K. Mincheva, A. Schankin, and M. Beigl, "Smartphone-based estimation of a user being in company or alone based on place, time, and activity," in *Mobile Computing, Applications, and Services* (K. Murao, R. Ohmura, S. Inoue, and Y. Gotoh, eds.), (Cham), pp. 74–89, Springer International Publishing, 2018.

[279] M. E. ul Haq, M. Awais Azam, Y. Asim, Y. Amin, U. Naeem, and A. Khalid, "Using smartphone accelerometer for human physical activity and context recognition in-the-wild," *Procedia Computer Science*, vol. 177, pp. 24–31, 2020. The 11th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2020) / The 10th International Conference

on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2020) / Affiliated Workshops.

[280] Y. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, "Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study," *Sensors*, vol. 19, 04 2019.

[281] A. R. Javed, M. U. Sarwar, S. Khan, C. Iwendi, M. Mittal, and N. Kumar, "Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition," *Sensors*, vol. 20, no. 8, 2020.

[282] G. Şengül, E. Ozcelik, S. Misra, R. Damaševičius, and R. Maskeliūnas, "Fusion of smartphone sensor data for classification of daily user activities," *Multimedia Tools and Applications*, vol. 80, pp. 33527–33546, Oct 2021.

[283] T. Ikuta, K. Tsubouchi, and N. Nishio, "Improving the accuracy of walking context recognition using step rate," in *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, 2021.

[284] D. Garcia-Gonzalez, D. Rivero, E. Fernandez-Blanco, and M. R. Luaces, "A public domain dataset for real-life human activity recognition using smartphone sensors," *Sensors*, vol. 20, no. 8, 2020.

[285] S. K. Bashar, A. Al Fahim, and K. H. Chon, "Smartphone based human activity recognition with feature selection and dense neural network," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pp. 5888–5891, 2020.

[286] K. A. Nguyen, R. N. Akram, K. Markantonakis, Z. Luo, and C. Watkins, "Location tracking using smartphone accelerometer and magnetometer traces," in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, ARES '19, (New York, NY, USA), Association for Computing Machinery, 2019.

[287] M. S. Mashuk, J. Pinchin, P.-O. Siebers, and T. Moore, "A smart phone based multi-floor indoor positioning system for occupancy detection," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp. 216–227, 2018.

[288] A. Esmaeili Kelishomi, A. Garmabaki, M. Bahaghighat, and J. Dong, "Mobile user indoor-outdoor detection through physical daily activities," *Sensors*, vol. 19, no. 3, 2019.

[289] A. Ofstad, E. Nicholas, R. Szcodronski, and R. R. Choudhury, "Aampl: Accelerometer augmented mobile phone localization," in *Proceedings of the First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-Less Environments*, MELT '08, (New York, NY, USA), p. 13–18, Association for Computing Machinery, 2008.

[290] E. Welbourne, J. Lester, A. LaMarca, and G. Borriello, "Mobile context inference using low-cost sensors," in *Proceedings of the First International Conference on Location- and Context-Awareness*, LoCA'05, (Berlin, Heidelberg), p. 254–263, Springer-Verlag, 2005.

[291] L. Zavala, R. Dharurkar, P. Jagtap, T. Finin, and A. Joshi, "Mobile, Collaborative, Context-Aware Systems," in *Proceedings of the AAAI Workshop on Activity Context Representation: Techniques and Languages*, AAAI, AAAI Press, August 2011.

[292] Y. Zhu, H. Luo, F. Zhao, and R. Chen, "Indoor/outdoor switching detection using multisensor densenet and lstm," *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1544–1556, 2021.

[293] Y. Du, V. Issarny, and F. Sailhan, "User-centric context inference for mobile crowdsensing," in *Proceedings of the International Conference on Internet of Things Design and Implementation*, IoTDI '19, (New York, NY, USA), p. 261–266, Association for Computing Machinery, 2019.

[294] J.-H. Park, D.-K. Kim, D. Baek, and J. Lee, "Low-power sensing model considering context transition for location-based services," *Soft Computing*, vol. 21, pp. 5223–5233, 2017.

[295] Y. Xia, S. Pan, W. Gao, B. Yu, X. Gan, Y. Zhao, and Q. Zhao, "Recurrent neural network based scenario recognition with multi-constellation gnss measurements on a smartphone," *Measurement*, vol. 153, p. 107420, 2020.

[296] M. Yu, F. Xue, C. Ruan, and H. Guo, "Floor positioning method indoors with smartphone's barometer," *Geo-spatial Information Science*, vol. 22, no. 2, pp. 138–148, 2019.

[297] F. Li, H. Chen, X. Song, Q. Zhang, Y. Li, and Y. Wang, "Condiosense: High-quality context-aware service for audio sensing system via active sonar," *Personal Ubiquitous Comput.*, vol. 21, p. 17–29, Feb. 2017.

[298] J. Sena, J. Barreto, C. Caetano, G. Cramer, and W. R. Schwartz, "Human activity recognition based on smartphone and wearable sensors using multiscale dcnn ensemble," *Neurocomputing*, vol. 444, pp. 226–243, 2021.

[299] M. G. Campana and F. Delmastro, "Compass: Unsupervised and online clustering of complex human activities from smartphone sensors," *Expert Systems with Applications*, vol. 181, p. 115124, 2021.

[300] W. Ge and E. Agu, "Cruft: Context recognition under uncertainty using fusion and temporal learning," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 747–752, 2020.

[301] Khimraj, P. K. Shukla, A. Vijayvargiya, and R. Kumar, "Human activity recognition using accelerometer and gyroscope data from smartphones," in *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*, pp. 1–6, 2020.

[302] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Networks and Applications*, vol. 25, pp. 743–755, Apr 2020.

[303] H. Gao, "Behaviour-aided environment detection for context adaptive navigation," in *Proceedings of the 31st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2018)*, (Miami, Florida), pp. 3283–3296, 2018.

[304] Y. Gu, D. Li, Y. Kamiya, and S. Kamijo, "Integration of positioning and activity context information for lifelog in urban city area," *NAVIGATION: Journal of the Institute of Navigation*, vol. 67, no. 1, pp. 163–179, 2020.

[305] W. Qi, H. Su, and A. Aliverti, "A smartphone-based adaptive recognition and real-time monitoring system for human activities," *IEEE Transactions on Human-Machine Systems*, vol. 50, no. 5, pp. 414–423, 2020.

[306] N. Ahmed, J. I. Rafiq, and M. R. Islam, "Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model," *Sensors*, vol. 20, no. 1, 2020.

[307] R.-A. Voicu, C. Dobre, L. Bajenaru, and R.-I. Ciobanu, "Human physical activity recognition using smartphone sensors," *Sensors*, vol. 19, no. 3, 2019.

[308] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, "Sensing meets mobile social networks: The design, implementation and evaluation of the cenceme application," in *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, SenSys '08, (New York, NY, USA), p. 337–350, Association for Computing Machinery, 2008.

[309] Z. Chen, Y. Chen, X. Gao, S. Wang, L. Hu, C. C. Yan, N. D. Lane, and C. Miao, "Unobtrusive sensing incremental social contexts using fuzzy class incremental learning," in *Proceedings of the 2015 IEEE International Conference on Data Mining (ICDM)*, ICDM '15, (USA), p. 71–80, IEEE Computer Society, 2015.

[310] B. Adams, D. Phung, and S. Venkatesh, "Sensing and using social context," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 5, Nov. 2008.

[311] M. Ehatisham-ul Haq, F. Murtaza, M. A. Azam, and Y. Amin, "Daily living activity recognition in-the-wild: Modeling and inferring activity-aware human contexts," *Electronics*, vol. 11, no. 2, 2022.

[312] A. Bolat, D. Kim, and K.-J. Li, "Discovering user-context in indoor space," in *Proceedings of the 9th ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, ISA'18, (New York, NY, USA), p. 1–6, Association for Computing Machinery, 2018.

[313] Y. Asim, M. A. Azam, M. Ehatisham-ul Haq, U. Naeem, and A. Khalid, "Context-aware human activity recognition (cahar) in-the-wild using smart-

phone accelerometer," *IEEE Sensors Journal*, vol. 20, no. 8, pp. 4361–4371, 2020.

[314] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, ACII '13, (USA), p. 671–676, IEEE Computer Society, 2013.

[315] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements," *J. of Biomedical Informatics*, vol. 59, p. 49–75, Feb. 2016.

[316] A. Reiss, *Personalized Mobile Physical Activity Monitoring for Everyday Life*. PhD thesis, Technical University of Kaiserslautern, 2014.

[317] F. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with lstm," in *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, vol. 2, pp. 850–855 vol.2, 1999.

[318] K. Z. Gajos, A. Hurst, and L. Findlater, "Personalized dynamic accessibility," *Interactions*, vol. 19, p. 69–73, mar 2012.

[319] K. Z. Gajos, J. O. Wobbrock, and D. S. Weld, "Automatically generating user interfaces adapted to users' motor and vision capabilities," in *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, (New York, NY, USA), p. 231–240, Association for Computing Machinery, 2007.

[320] L. Li and K. Z. Gajos, "Adaptive click-and-cross: Adapting to both abilities and task improves performance of users with impaired dexterity," in *Proceedings of the 19th International Conference on Intelligent User Interfaces*, IUI '14, (New York, NY, USA), p. 299–304, Association for Computing Machinery, 2014.

[321] M. E. Mott, R.-D. Vatavu, S. K. Kane, and J. O. Wobbrock, "Smart touch: Improving touch accuracy for people with motor impairments with template matching," in *Proceedings of the 2016 CHI Conference on Human Factors in*

Computing Systems, CHI '16, (New York, NY, USA), p. 1934–1946, Association for Computing Machinery, 2016.

[322] J. P. Hourcade, C. M. Nguyen, K. B. Perry, and N. L. Denburg, "Pointassist for older adults: Analyzing sub-movement characteristics to aid in pointing tasks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, (New York, NY, USA), p. 1115–1124, Association for Computing Machinery, 2010.

[323] G. W. Tigwell, D. R. Flatla, and R. Menzies, "It's not just the light: Understanding the factors causing situational visual impairments during mobile interaction," in *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, NordiCHI '18, (New York, NY, USA), p. 338–351, Association for Computing Machinery, 2018.

[324] P. Biswas, P. M. Langdon, J. Umadikar, S. Kittusami, and S. Prashant, "How interface adaptation for physical impairment can help able bodied users in situational impairment," in *Inclusive Designing* (P. M. Langdon, J. Lazar, A. Heylighen, and H. Dong, eds.), (Cham), pp. 49–58, Springer International Publishing, 2014.

[325] J. Yu, J. Zhao, Y. Chen, and J. Yang, "Sensing ambient light for user experience-oriented color scheme adaptation on smartphone displays," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, SenSys '15, (New York, NY, USA), p. 309–321, Association for Computing Machinery, 2015.

[326] M. Goel, A. Jansen, T. Mandel, S. N. Patel, and J. O. Wobbrock, "Contexttype: Using hand posture information to improve mobile touch screen text entry," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, (New York, NY, USA), pp. 2795–2798, ACM, 2013.

[327] J. O. Wobbrock, "Situationally aware mobile devices for overcoming situational impairments," in *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '19, (New York, NY, USA), pp. 1:1–1:18, ACM, 2019.

[328] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimental Psychology*, vol. 47, no. 6, pp. 381–391, 1954.

[329] I. S. MacKenzie, "Fitts' law as a research and design tool in human-computer interaction," *Hum.-Comput. Interact.*, vol. 7, p. 91–139, mar 1992.

[330] Z. Wang, S. Tan, L. Zhang, and J. Yang, "Obstaclewatch: Acoustic-based obstacle collision detection for pedestrian using smartphone," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, dec 2018.

[331] F. Putze, T. Ihrig, T. Schultz, and W. Stuerzlinger, *Platform for Studying Self-Repairing Auto-Corrections in Mobile Text Entry Based on Brain Activity, Gaze, and Context*, p. 1–13. New York, NY, USA: Association for Computing Machinery, 2020.

[332] L. Suchman, "Interactive artifacts," in *Human-Machine Reconfigurations: Plans and Situated Actions*, Learning in Doing: Social, Cognitive and Computational Perspectives, p. 33–50, Cambridge University Press, 2 ed., 2006.

[333] V. Alvarez-Cortes, B. E. Zayas-Perez, V. H. Zarate-Silva, and J. A. Ramirez Uresti, "Current trends in adaptive user interfaces: Challenges and applications," in *Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007)*, pp. 312–317, 2007.

[334] L. Findlater and K. Z. Gajos, "Design space and evaluation challenges of adaptive graphical user interfaces," *AI Magazine*, vol. 30, p. 68, Sep. 2009.

[335] A. Gunawardana, T. Paek, and C. Meek, "Usability guided key-target resizing for soft keyboards," in *Proceedings of the 15th International Conference on Intelligent User Interfaces*, IUI '10, (New York, NY, USA), p. 111–118, Association for Computing Machinery, 2010.

[336] L. Findlater and J. McGrenere, "A comparison of static, adaptive, and adaptable menus," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, (New York, NY, USA), p. 89–96, Association for Computing Machinery, 2004.

[337] A. Jameson, "Understanding and dealing with usability side effects of intelligent processing," *AI Mag.*, vol. 30, pp. 23–40, 2009.

[338] L. Findlater and J. McGrenere, "Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, (New York, NY, USA), p. 1247–1256, Association for Computing Machinery, 2008.

[339] M. Peissner and R. Edlin-White, "User control in adaptive user interfaces for accessibility," in *Human-Computer Interaction – INTERACT 2013* (P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, eds.), (Berlin, Heidelberg), pp. 623–640, Springer Berlin Heidelberg, 2013.

[340] A. Martin-Hammond, A. Ali, C. Hornback, and A. K. Hurst, "Understanding design considerations for adaptive user interfaces for accessible pointing with older and younger adults," in *Proceedings of the 12th International Web for All Conference*, W4A '15, (New York, NY, USA), Association for Computing Machinery, 2015.

[341] D. A. Norman, "Design rules based on analyses of human error," *Commun. ACM*, vol. 26, p. 254–258, apr 1983.

[342] K. Z. Gajos, K. Everitt, D. S. Tan, M. Czerwinski, and D. S. Weld, "Predictability and accuracy in adaptive user interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, (New York, NY, USA), p. 1271–1274, Association for Computing Machinery, 2008.

[343] T. F. Paymans, J. Lindenberg, and M. Neerincx, "Usability trade-offs for adaptive user interfaces: Ease of use and learnability," in *Proceedings of the 9th International Conference on Intelligent User Interfaces*, IUI '04, (New York, NY, USA), p. 301–303, Association for Computing Machinery, 2004.

[344] T. Tsandilas and m. c. schraefel, "An empirical assessment of adaptation techniques," in *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, (New York, NY, USA), p. 2009–2012, Association for Computing Machinery, 2005.

[345] N. Jakob, "Usability heuristics for user interface design," *NN/g Nielsen Norman Group. Retrived from: https://www. nngroup. com/articles/ten-usability-heuristics*, 1994.

[346] P. Langley, "User modeling in adaptive interface," in *UM99 User Modeling* (J. Kay, ed.), (Vienna), pp. 357–370, Springer Vienna, 1999.

[347] K. Z. Gajos, M. Czerwinski, D. S. Tan, and D. S. Weld, "Exploring the design space for adaptive graphical user interfaces," in *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '06, (New York, NY, USA), p. 201–208, Association for Computing Machinery, 2006.

[348] S. Bouzit, D. Chêne, and G. Calvary, "Evanescent adaptation on small screens," in *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, (New York, NY, USA), p. 62–68, Association for Computing Machinery, 2015.

[349] T. Lavie and J. Meyer, "Benefits and costs of adaptive user interfaces," *International Journal of Human-Computer Studies*, vol. 68, no. 8, pp. 508–524, 2010. Measuring the Impact of Personalization and Recommendation on User Behaviour.

[350] K. Montague, V. L. Hanson, and A. Cobley, "Designing for individuals: Usable touch-screen interaction through shared user models," in *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, (New York, NY, USA), pp. 151–158, ACM, 2012.

[351] K. Reinecke and A. Bernstein, "Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces," *ACM Trans. Comput.-Hum. Interact.*, vol. 18, jul 2011.

[352] F. Hamidi, K. Poneres, A. Massey, and A. Hurst, "Who should have access to my pointing data? privacy tradeoffs of adaptive assistive technologies," in *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '18, (New York, NY, USA), p. 203–216, Association for Computing Machinery, 2018.

[353] A. Mariakakis, S. Parsi, S. N. Patel, and J. O. Wobbrock, "Drunk user interfaces: Determining blood alcohol level through everyday smartphone tasks," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, (New York, NY, USA), pp. 234:1–234:13, ACM, 2018.

[354] F. Wolf and R. Kuber, "Developing a head-mounted tactile prototype to support situational awareness," *Int. J. Hum.-Comput. Stud.*, vol. 109, pp. 54–67, Jan. 2018.

[355] D. Kim, K. Han, J. S. Sim, and Y. Noh, "Smombie guardian: We watch for potential obstacles while you are walking and conducting smartphone activities," *PLOS ONE*, vol. 13, pp. 1–21, 06 2018.

[356] Z. Wang, S. Tan, L. Zhang, and J. Yang, "Obstaclewatch: Acoustic-based obstacle collision detection for pedestrian using smartphone," *Proc. ACM Interact. Mob. Wearable Ubiquitous Tech.*, vol. 2, no. 4, pp. 194:1–194:22, 2018.

[357] W. Zamora, C. T. Calafate, J.-C. Cano, and P. Manzoni, "Smartphone tuning for accurate ambient noise assessment," in *Proceedings of the 15th International Conference on Advances in Mobile Computing & Multimedia*, MoMM2017, (New York, NY, USA), pp. 115–122, ACM, 2017.

[358] F. Lyddy, F. Farina, J. Hanney, L. Farrell, and N. Kelly O'Neill, "An analysis of language in university students' text messages," *Journal of Computer-Mediated Communication*, vol. 19, no. 3, pp. 546–561, 2014.

[359] A. Arslan, "Deasciification approach to handle diacritics in turkish information retrieval," *Inf. Process. Manage.*, vol. 52, p. 326–339, Mar. 2016.

[360] A. T. Koksal, O. Bozal, E. Yürekli, and G. Gezici, "#turki$hTweets: A benchmark dataset for Turkish text correction," in *Findings of the ACL: EMNLP 2020*, (Online), pp. 4190–4198, ACL, Nov. 2020.

[361] S. Yildirim and T. Yildiz, "An unsupervised text normalization architecture for turkish language," *Res. Comput. Sci.*, vol. 90, pp. 183–194, 2015.

[362] GSMArena, "Gsmarena.com - mobile phone reviews, news, specifications and more...." `https://www.gsmarena.com/`, 2000. Last accessed: 29.06.2022.

Appendix A

SUMMARY OF THE CONTEXTUAL FACTORS

This appendix summarizes the contextual factors studied in the literature.

A.1 Physical Context Summary

This section summarizes the literature related to physical context.

Table A.1: Overall physical contexts in the literature

Physical Cont.	Types	Papers
Location	Lab environment	[8, 31, 37, 39–41, 45, 46, 50, 52, 61–66, 68–70, 72–84, 87, 88, 90–92, 94–105, 107, 109–114, 116–119, 121, 124, 125, 127, 128, 144, 147, 148, 150–157, 159–162, 164, 166, 168–181, 184–187, 193, 195, 202, 205, 206, 353, 354]
	Indoor environment	[8, 34, 36, 42, 43, 47–49, 66, 71, 85, 86, 89, 93, 102, 106, 108, 115, 120, 122, 124–126, 129–132, 158, 163, 167, 188–190, 194, 196, 199, 200, 208]
	Outdoor environment	[32, 33, 38, 43, 44, 51, 53, 55, 56, 58, 59, 62, 65, 106, 136–140, 182, 191, 192, 199, 201, 204, 209, 210, 355, 356]
	Pedestrian street or public area	[8, 35, 57, 63, 64, 68, 133–135]
	Virtual environment	[114, 141, 142, 162]
	Stairs in a building	[127, 128]
	Public transportation	[97, 123]
	In the wild	[39, 54, 67, 143–146, 197, 198, 207]

Continued on next page

Table A.1 – *Continued from previous page*

Physical Cont.	Types	Papers
Mobility	Sitting	[8,31,36,37,45,46,52,61–64,68–100,123,125,127,128, 146, 148, 150, 153, 155, 161, 166, 171–173, 175, 176, 182, 184–187, 197, 202, 206, 207, 354]
	Standing	[8, 33, 47–50, 65, 66, 81, 88, 90, 91, 93, 101–118, 130, 157, 163, 168, 178, 182, 185]
	Walking on a route	[8, 32–35, 38, 41–44, 61, 63–65, 68, 69, 74–76, 78–81, 84, 86, 89–95, 98–109, 111–113, 115, 117, 119, 120, 124, 126–132, 134–139, 147, 149–151, 157, 159, 161, 163, 169, 170, 176, 184–186, 188–192, 194, 196, 200, 201, 204, 209, 210, 355, 356]
	Walking on a treadmill	[40,41,65,68,70,72,77,78,82,83,87,94–96,99,110,114, 116, 118, 119, 121, 122, 148, 154, 162, 179, 354]
	Walking on a mini-stepper	[124]
	Walking after a researcher	[36, 88]
	Walking on a straight path	[51, 62, 92, 152, 158, 165]
	Walking through a street or public area	[53, 55–59, 71]
	Going up or down stairs	[127, 128, 208]
	Public transportation	[8, 97, 146]
	Jogging or running	[44, 199]
	Walking	[54, 66, 73, 85, 125, 133, 141, 142, 146, 164, 182, 195, 199, 203]
Artifacts	Physical obstacles	[35,51,57,69,78,92,101,106,126,129,132,134,135,138, 139, 147, 194, 355, 356]
	Furniture	[71, 74, 75, 79, 80, 94, 95, 99, 112, 119, 150, 161]
	Pedestrians	[35, 53, 56, 57, 66, 106, 120, 129, 134, 135, 188, 356]
	Vehicles	[56, 356]
	Virtual vehicles or obstacles	[141, 142, 162]
	Unicycling clown	[55]
Sensed Environm. Attributes	Lighting levels (low/high)	[75, 79, 80, 119, 199]
	Weather (cloudy, partly cloudy, sunny)	[106]
	Vibration	[123]
	Environmental noise	[50, 123, 198, 199, 357]
	Temperature (cold, warm)	[47–49]
	Acceleration sensors	[34, 36, 85, 117, 146, 182, 184, 203]

Continued on next page

Table A.1 – *Continued from previous page*

Physical Cont.	Types	Papers
	Gyroscope	[85, 146, 203]
	Magnetic field sensors	[87, 203]
	Motion sensors	[90, 116]
	Heart rate	[116, 206]
	GPS	[182, 192]

A.2 Temporal Context Summary

This section summarizes the literature related to temporal context.

Table A.2: Overall temporal contexts in the literature

Temporal Context	Types	Papers
Duration	Single session for less than 10 minutes	[69, 99, 138, 139, 152, 154, 176]
	Single session for 10 - 30 minutes	[45, 60, 61, 89, 126, 191, 206, 355]
	Single session for 30 - 60 minutes	[36, 39, 40, 74, 87, 93–96, 108, 115, 121, 129, 136, 163, 170, 175, 189, 190, 197, 203, 207]
	Single session for 70 - 90 minutes	[8, 50, 52, 100, 134, 135, 151, 178, 187, 192]
	Single session for more than 90 minutes	[34, 63, 64, 122, 179, 210]
	Single session with unknown duration	[32, 33, 35, 38, 42, 44, 46, 51, 66, 70–73, 75, 78–81, 83, 85, 86, 92, 98, 101–105, 109–113, 117, 119, 120, 123, 125, 127, 128, 131, 133, 137, 140, 141, 147–149, 157–159, 161, 162, 164, 165, 169, 171, 173, 174, 177, 184, 185, 195, 202, 208, 209]
	Multiple sessions within a single day	[47–49, 144, 150]
	Multiple sessions on different days	[31, 37, 41, 43, 65, 84, 88, 90, 114, 116, 153, 155, 186, 188, 201, 204, 211, 353]

Continued on next page

Table A.2 – *Continued from previous page*

Temporal Context	**Types**	**Papers**
	Multiple experiments	[55,62,68,76,77,82,91,97,106,107,118,124, 130, 132, 142, 156, 160, 166–168, 172, 180–182, 193, 194, 196, 200, 205, 354]
	Longitudinal study	[39, 54, 143, 145, 146, 197, 198]
Before, during, after	ESM	[144, 146, 197, 207]
Synchronous/ Asynchronous	Synchronous	[44, 55, 141, 142, 147, 160]
	Asynchronous	[44, 141, 147]
Actions' relation to time	Hurrying, normal and waiting	[8]
	Presentation time	[157]
	Walking speed	[41, 65, 78, 157–159]
Time of day, week and year	Afternoon	[106]
	Busiest time of the day	[53, 71]
	Different hours of a day	[53, 57, 67]
	Spring	[35, 56, 60, 355]
	Summer	[43, 58, 60, 134]

A.3 Task Context Summary

This section summarizes the literature related to task context.

Table A.3: Overall task contexts in the literature

Task Context	Types	Papers
Multitasking	Walking	[8, 32–36, 38, 40–44, 53–55, 57, 61–66, 68–96, 98, 101–117, 119, 121, 122, 124–139, 141, 142, 147–152, 154, 157, 159, 161, 164, 165, 169, 170, 176, 179, 184–186, 188–192, 196, 200, 201, 203, 204]
	Encumbrance	[41, 111, 113, 115, 118, 151, 159]
	Exercising, jogging or running	[44, 67, 96, 122]
	Collision/hazard avoidance	[76, 129, 142, 162, 169]
	Playing a game	[68, 158, 160, 196, 354]
	Distraction or cognitive tasks	[40, 52, 82, 148, 201]
	Conversation or social interaction	[67, 202]
	Monitoring environment	[114, 154]
	Others: parenting [60], talking on the phone [160], holding objects with different sizes [168], working, eating, relax. and traveling [67]	
Interruptions	Obstacles	[35, 53, 56, 57, 66, 69, 71, 71, 74, 75, 78–80, 92, 94, 95, 101, 106, 112, 119, 126, 129, 132, 134, 135, 138, 139, 147, 150, 161, 188, 356]
	Eyes-free interaction	[31, 37, 45, 46, 88, 136, 145, 155, 156, 160, 166, 171–178, 187, 193, 195, 205]
	Hands-free interaction	[180, 181, 195]
	Stressor tasks	[52, 197, 207]
	Hazard checks	[76, 169]
	Distraction tasks	[121, 153]
	Stairs	[127, 128]
	Virtual objects or vehicles	[141, 142]
	Others: stop signs [170], visual disruptions [167], alcohol usage [353], interruptions from children [60], incoming phone calls [39]	
Action orien. task domain	Texting or talk. on the phone	[53, 198]
	Playing a game	[143, 206]
	Media control	[198]

Continued on next page

Table A.3 – *Continued from previous page*

Task Context	Types	Papers
Goal orien. task domain	Navigation	[32,34,35,38,42,43,54,85,131,134,136,138,139, 170,184,188–192,200,201,204,210]
	Target selection	[33,41,46–50,52,62,70,71,73,78,81,83,84,87,91, 93,101,104,106–108,110,111,113–116,124,126, 156,159,166,167,193,196]
	Text entry	[31,36,37,50–52,57,61,63,64,66,68,69,76,82,86, 88,90,92,97,102,120,123,127,128,144,147,150, 153–155,158,160,165,168,171–175,197,207–209, 353]
	Gesture based interaction	[45,118,125,130,148,151,176,178,185,186,194]
	Reading	[33, 44, 74, 75, 79, 80, 98, 100, 117, 119, 121, 145, 147,157,161,165,355]
	Searching	[8,50,52,75,94–96,99,112,117,119,140,146,197, 207,209]
	Talking on the phone	[44,55,141,142,147,149,158,203]
	Texting	[40,44,141,149,152,164]
	Playing a game	[129,162,167]
	Menu related tasks	[132,133,177,187]
	Other goal oriented tasks	[39,44,65,72,77,89,103,105,109,122,135,137, 141,160,169,179–182,195,199,202,205,353]

Other goal oriented tasks include audio target acquisition [103], cognitive tasks [89], cross-modal icon identification [77], dealing with incoming notifications or alerts [105, 122, 135], foot gesture based interaction [180, 181], gesture recognition [205], head gesture based target selection [109], recording phone number, checking calendar [160], question answering [39], remembering symbols shown [179], speech based text entry [72, 169, 195], sports tracking [44, 137], listening to music [141], information retrieval [182], browsing bus timetable [199], declining incoming calls [202], heart rate balancing, simple and choice reaction [353], and visual acuity [65].

A.4 Social Context Summary

This section summarizes the literature related to social context.

Table A.4: Overall social contexts in the literature

Social Context	Types	Papers
Persons present	Self	[8,31–55,57–106,108–122,124–182,184–201,204–210, 353–356]
	Other pedestrians	[35, 53, 56, 57, 66, 106, 120, 129, 134, 135, 188, 356]
	Accompanied	[53, 55, 57, 123, 202, 203]
	At least one child	[60]
	Experimenter	[160]
Interpersonal interaction	One to one	[44, 53, 141, 142, 147, 152, 160, 202, 203]
Culture	Users from UK and India	[143]

A.5 Technical Context Summary

This section summarizes the literature related to technical context.

Table A.5: Overall technical contexts in the literature

Technical Cont.	Types	Papers
Device	Smartphone with touchscreen	[33–39, 41, 44, 47–50, 52, 53, 66, 79, 80, 85, 86, 88, 92, 94–99, 105, 110, 113, 114, 120, 121, 125, 133–136, 140, 143, 144, 150, 151, 153–157, 159, 162, 163, 166, 167, 170–177, 179, 186, 190, 192, 195–197, 200–202, 206, 207, 209, 210, 353, 354, 356]
	PDA	[31, 62, 68, 70–72, 75, 78, 82, 102, 107, 108, 117, 119, 123, 124, 127, 128, 132, 194, 203]
	Tablet	[42, 45, 46, 61, 63, 64, 73, 94, 95, 104, 129, 131, 138, 139, 169, 181, 205]
	Wearable device	[32, 42, 46, 69, 77, 87, 91, 93, 101, 122, 132, 135, 137, 160, 166, 182, 185, 194]
	Smartwatch	[43, 44, 67, 96, 115, 116, 118, 137, 178, 179, 188, 189, 193]
	Participants' own devices	[54, 56, 58, 59, 141, 144, 146, 147, 152, 164, 165, 208]
	Smartphone with physical keyboard	[8, 76, 81, 97, 109, 111, 149, 160, 173, 191]
	Smartphone with touch. & phys. keyboard	[83, 90, 107, 112, 148, 168, 184]
	UMPC	[74, 84, 106, 126, 145]
	Media device with touchscreen	[65, 98, 161, 170, 187]
	Mobile phone with physical keyboard	[68, 142, 203]
	Smart glasses	[100, 188]
	Smart bracelet/Wristband	[44, 193]
	Others: Twiddler [31, 130], Trackball, gyroscopic mouse and touchpad [130], hand-held device [199], Protractor3D recognizer with acceler. [180], actuators on shoe, eye tracker [201], head mounted display [74, 89], MT-9B orientation trackers [103], e-book reader [74], Microsoft Xbox Kinect [180, 181]	
Informational artefacts	Laptop computer	[87, 88, 168, 171–173, 175]
	Projection screen or floor	[69, 72, 76, 154, 154, 169]

Continued on next page

Table A.5 – *Continued from previous page*

Technical Cont.	Types	Papers
	Large monitor or wall display	[40, 91, 148]
	External screen or LCD displays	[37, 89, 166]
	Desktop computer	[31, 91]
	Tablet	[45]
Interoperability	Smartphone - Smartphone	[200]
	Google Glass - PocketThumb	[93]
	PDA - wearable device	[132]
	PDA/Twiddler - Desktop computer	[31]
	Smartphone - Smart glasses & smartwatch	[188]
	Smartphone - Headphone	[173, 195]
	Smartphone - Laptop computer	[88, 171, 172]
	Wearable device - Desktop computer	[91]
	Laptop computer - PDA	[203]
	Wearable device - Laptop computer	[87]
	Wear. device - Smartphone - Ext. monitor	[166]
	Actuators on shoe, smartphone, eye tracker	[201]
Mixed reality systems	Virtual reality	[114, 141, 142]

Appendix B

TEXT-SPEAK EXAMPLES

Table B.1 illustrates common text-speak techniques in daily texting use and examples for these techniques.

Table B.1: Common text-speak techniques and their examples

Text-speak technique	Example	References
Deletion of vowels	“msg" for “message"	[222, 358]
Deletion of repeated characters	“tomorow" for “tomorrow"	[222]
Shortening of words	“lab" for “laboratory"	[222, 224, 358]
Deletion of punctuation	“dont" for “don’t"	[224, 358]
Deletion the ’g’ at the end in words ending “ing"	“goin" for “going"	[358]
Deletion of the final characters	“hav" for “have"	[358]
Phonetic substitution	“2" for “too" or “c" for “see"	[222, 224, 358]
Abbreviation	“lol” for “laughs out loud”	[222, 224, 358]
Dialectal and informal usage	“gonna” for “going to”	[222, 224, 358]
Deletion of function words and pronouns	“readin bk" for	[222]
Missed capitalization	“i’d" for “I’d"	[224, 358]
Spelling as pronunciation	“fone" for “phone"	[20, 224, 358]
Onomatopoeic/ exclamatory	“ha", “yay"	[224, 358]
Repeating characters for expression	“whaaaat" to express surprise	[224]
Using upper case/extra punctuation for emotion	“WHAT?????"	[224]
Using insider words	“hottie", “fugly"	[224]
Prevention of using Turkish characters	“kacmis" for “kaçmış"	[20, 359, 360]
Separation errors	“birşey" for “bir şey" “hiç biri" for “hiçbiri"	[360]
Use of English words in Turkish text		[360]
Neologisms	“hack-lemek"	[360]
Incorrect use of some suffixes	“kitapda" for “kitap da"	[361]

Appendix C

OVERALL ESM QUESTIONS

This chapter presents the ESM questions used in the study.

C.1 ESM Questions for Labelling Context

If the participants entered text longer than five characters, the app sent notifications to ask them to answer a set of questions related to their current context. The overall questions for context labelling and provided options are as follows:

1. Which one of these best describes your current location?
   - Indoors
   - Outdoors
   - Stairs
   - In vehicle
   - Crosswalk
   - Other
2. Which one of these best describes your mobility condition?
   - Lying down
   - Sitting
   - Standing
   - Walking
   - Running

- Other

3. Which one of these best describes people around you?

    - Alone
    - With a friend/family member/colleague
    - With 2-4 friends/family members/colleagues
    - With more than 4 friends/family members/colleagues
    - With strangers (not crowded)
    - With strangers (crowded)
    - Other

4. Did you handle any other task along with text entry?

    - Nothing
    - I am carrying a box/bag/other
    - I am trying to avoid collision while walking
    - I am having a conversation with someone around me
    - I am working
    - I am shopping
    - I am doing home-activities (cleaning, cooking, etc)
    - I am having breakfast/lunch/dinner
    - Multiple of these
    - Other

5. Is there anything that interrupted/distracted your interaction with mobile device?

    - Nothing
    - There are obstacles/people/cars on walking path
    - I am in a hurry
    - I need to check something from time to time (i.e. a child or cook)
    - I am interrupted by someone

- I am interrupted by something unexpected
- Multiple of these
- Other

C.2 ESM Questions for Participants' Self Evaluation on Typing Errors

If the participants deleted any characters during a session, we asked participants if they made a typing error after context questions. If the participants selected yes or maybe options, we asked them to specify the cause of the typing problem. The overall questions for self-evaluation and provided options are as follows:

1. Did you just make a typing error?
   - Yes
   - No
   - Maybe
2. What do you think caused this typing error?
   - My current location
   - My current mobility situation
   - People around me
   - Other task I am busy with
   - Something that interrupts me
   - Multiple of these
   - Other

Appendix D

PARTICIPANTS' DEVICE SUMMARY

Table D.1 provides a summary of participants' devices. The brand and model names and Android SDK versions were retrieved from participants' devices. The screen sizes were collected from product specifications [362].

Table D.1: Participants' smartphone brands, models, Android SDK versions and screen sizes

Brand	Model	SDK	Size (inches)	Keyboard	#
Asus	ASUS_X00QD (Zenfone 5)	28	6.2	Gboard	1
Google	Pixel 3	29	5.5	Gboard	1
Huawei	ANE-LX1 (P20 lite)	28	5.84	Microsoft SwiftKey	2
Huawei	BLA-L09 (Mate 10 Pro)	29	6.0	Gboard	1
Huawei	ELE-L29 (P30)	29	6.1	Microsoft SwiftKey	1
Huawei	FIG-LX1 (P smart)	28	5.65	Microsoft SwiftKey	1
Huawei	RNE-L21 (Mate 10 Lite)	26	5.9	Microsoft SwiftKey	1
Huawei	SNE-LX1 (Mate 20 lite)	29	6.3	Microsoft SwiftKey	1
Huawei	VTR-L09 (P10)	28	5.1	Microsoft SwiftKey	1
Lenovo	Lenovo P2a42 (P2)	24	5.5	Gboard	1
Nokia	Nokia 6.1	29	5.5	Gboard	1
Nokia	Nokia 7.2	29	6.3	Gboard	1
OnePlus	ONEPLUS A6000	29	6.28	Microsoft SwiftKey	1
Samsung	SM-A305F (Galaxy A30)	29	6.4	Samsung	1
Samsung	SM-A307FN (Galaxy A30s)	29	6.4	Samsung	1
Samsung	SM-A505F (Galaxy A50)	29	6.4	Samsung	2
Samsung	SM-A520F (Galaxy A5)	26	5.2	Samsung	2

Continued on next page

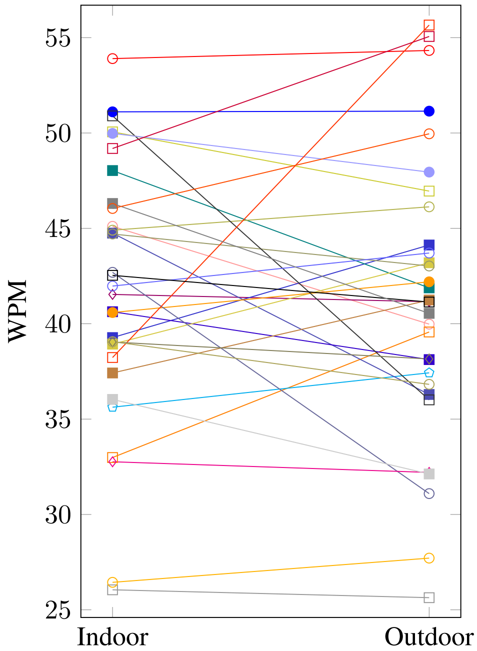
Table D.1 – *Continued from previous page*

Brand	Model	SDK	Size (inches)	Keyboard	#
Samsung	SM-A710F (Galaxy A7)	24	5.5	Microsoft SwiftKey	1
Samsung	SM-G610F (Galaxy J7 Prime)	24	5.5	Samsung	1
Samsung	SM-G610F (Galaxy J7 Prime)	27	5.5	Samsung	2
Samsung	SM-G930F (Galaxy S7)	26	5.1	Microsoft SwiftKey	1
Samsung	SM-G935F (Galaxy S7 Edge)	26	5.5	Fleksy	1
Samsung	SM-G935F (Galaxy S7 Edge)	26	5.5	Samsung	1
Samsung	SM-G950F (Galaxy S8)	28	5.8	Samsung	1
Samsung	SM-G950U (Galaxy S8)	28	5.8	Samsung	1
Samsung	SM-G965F (Galaxy S9+)	26	6.2	Samsung	1
Samsung	SM-G965U1 (Galaxy S9+)	29	6.2	Samsung	1
Samsung	SM-J710FQ (Galaxy J7)	27	5.5	Samsung	1
Samsung	SM-N950F (Galaxy Note8)	28	6.3	Samsung	1
Samsung	SM-N960F (Galaxy Note9)	29	6.4	Samsung	2
Xiaomi	MI 6	28	5.15	Gboard	1
Xiaomi	MI 6	28	5.15	Microsoft SwiftKey	1
Xiaomi	MI 8 Lite	29	6.26	Gboard	1
Xiaomi	MI CC 9e	28	6.01	Gboard	1
Xiaomi	Mi 9T	28	6.39	Gboard	1
Xiaomi	Redmi 6	28	5.45	Microsoft SwiftKey	1
Xiaomi	Redmi Note 5 Pro	28	5.99	Gboard	1
Xiaomi	Redmi Note 8	28	6.3	Gboard	1
Xiaomi	Redmi Note 8 Pro	28	6.53	Gboard	1
Xiaomi	Redmi Note 8 Pro	29	6.53	Gboard	3
Xiaomi	Redmi Note 9 Pro	29	6.67	Gboard	1

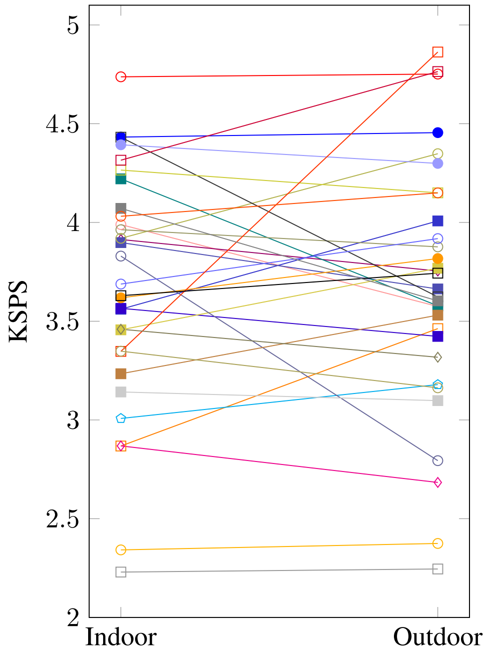
Appendix E

THE EFFECT OF CONTEXT ON PARTICIPANTS' INDIVIDUAL PERFORMANCE

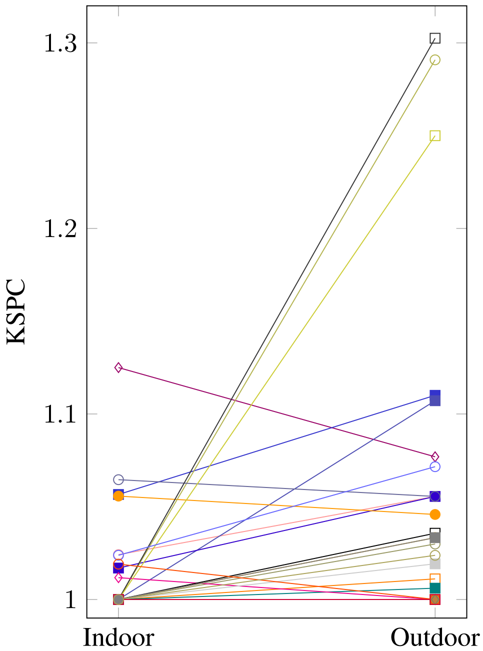
Figure E.1, Figure E.2, Figure E.3, Figure E.4, and Figure E.5 illustrate how individual performances change under different context groups of environment, mobility, social context, multitasking and distractions, respectively. The shape of the marker indicates the age of the participants (circle (○): 18 - 24, square (□): 25 - 34, diamond (◇): 35 - 54, triangle (△): 55+). Female participants are represented as filled (●), and male participants are represented as unfilled (○). The participants are illustrated with the same colors in all figures.



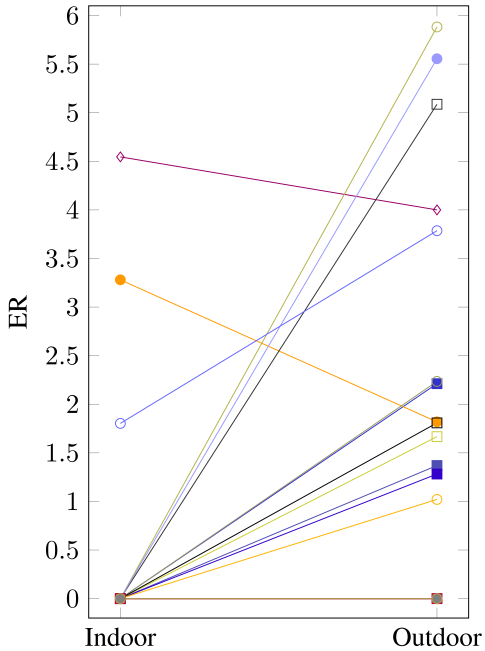
(a) The effect of environment on WPM



(b) The effect of environment on KSPS

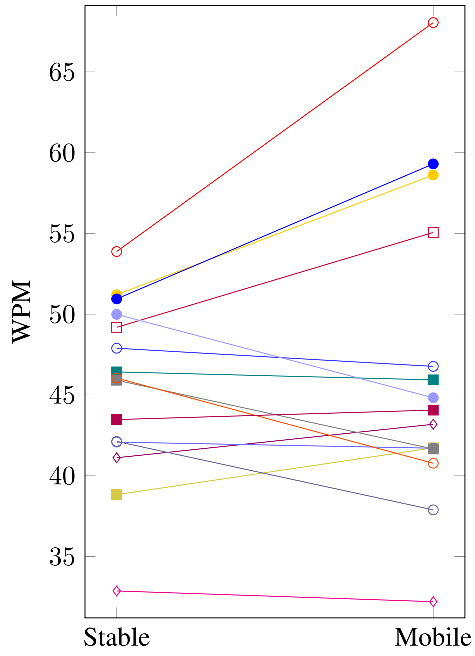


(c) The effect of environment on KSPC

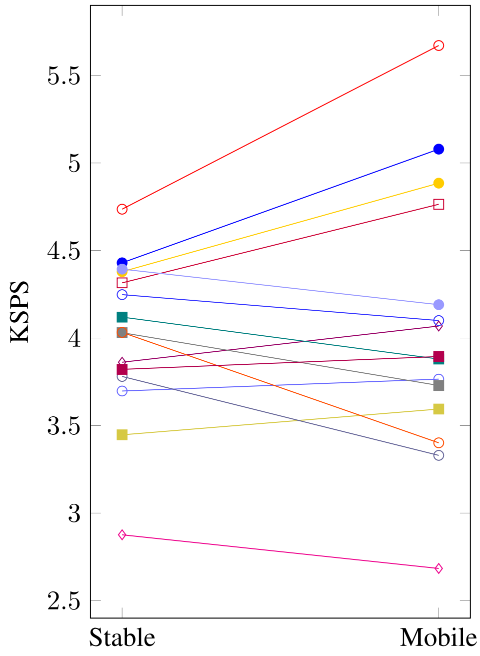


(d) The effect of environment on ER

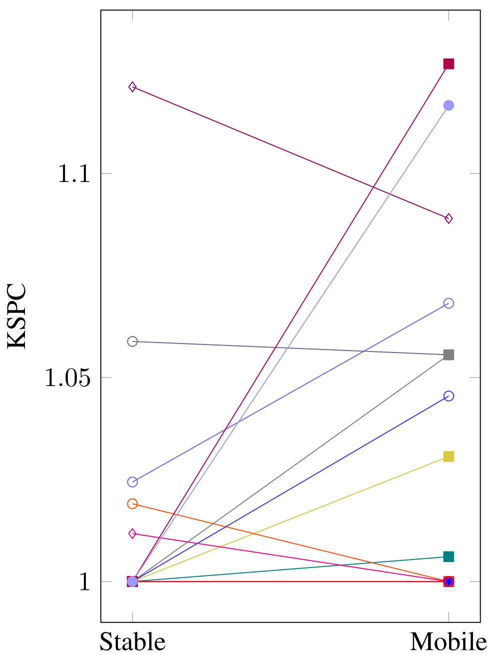
Figure E.1: The effect of environment on individual performances



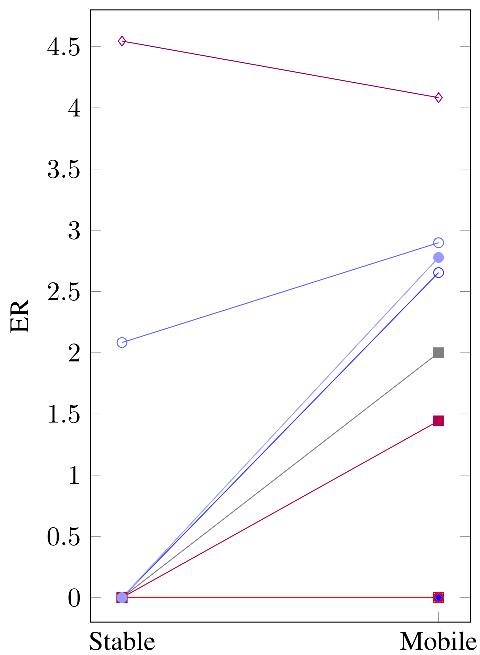
(a) The effect of mobility on WPM



(b) The effect of mobility on KSPS

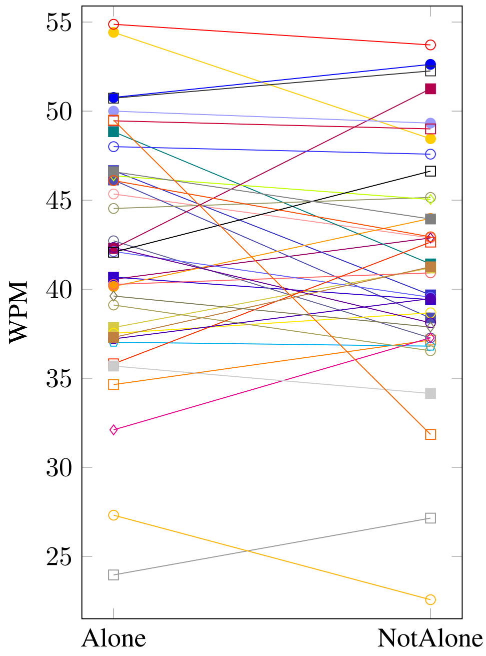


(c) The effect of mobility on KSPC

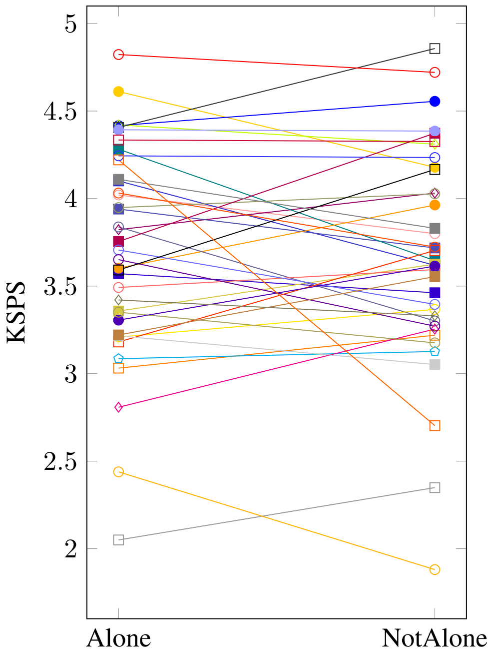


(d) The effect of mobility on ER

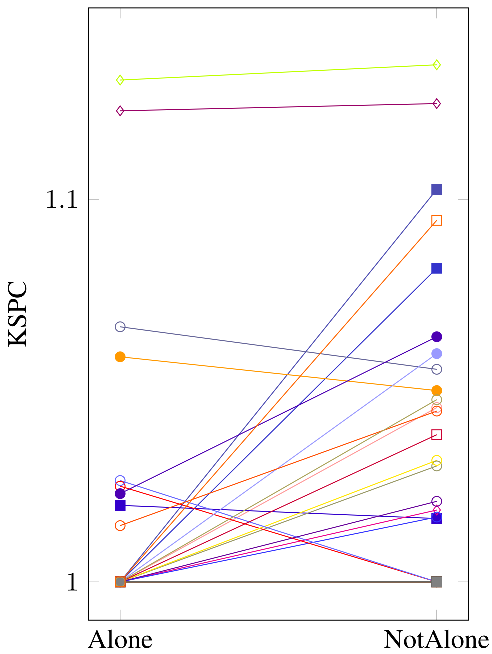
Figure E.2: The effect of mobility on individual performances



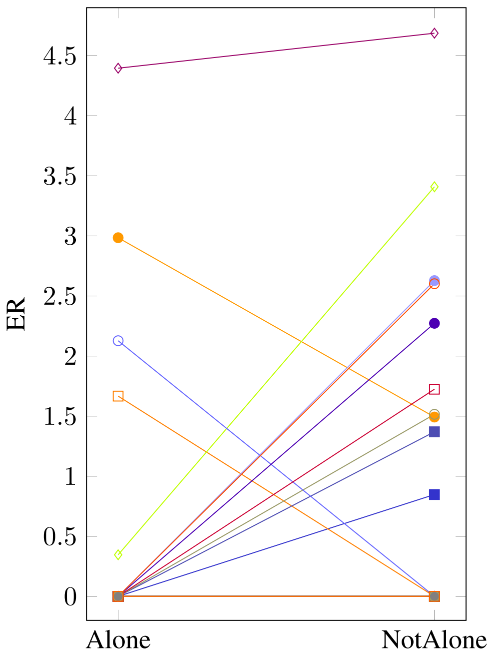
(a) The effect of social context on WPM



(b) The effect of social context on KSPS

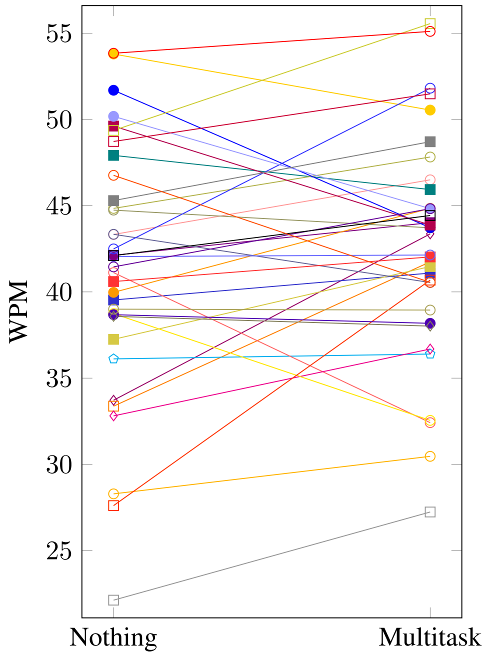


(c) The effect of social context on KSPC

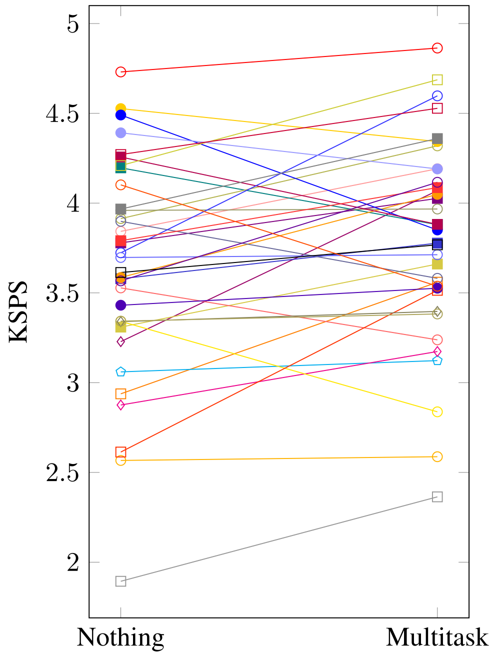


(d) The effect of social context on ER

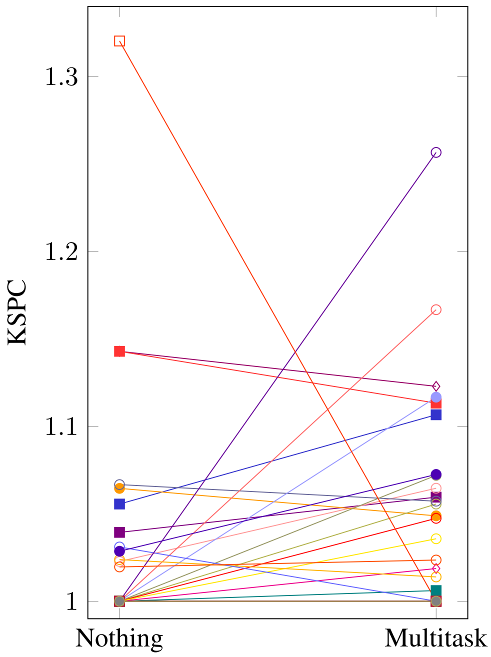
Figure E.3: The effect of social on individual performances



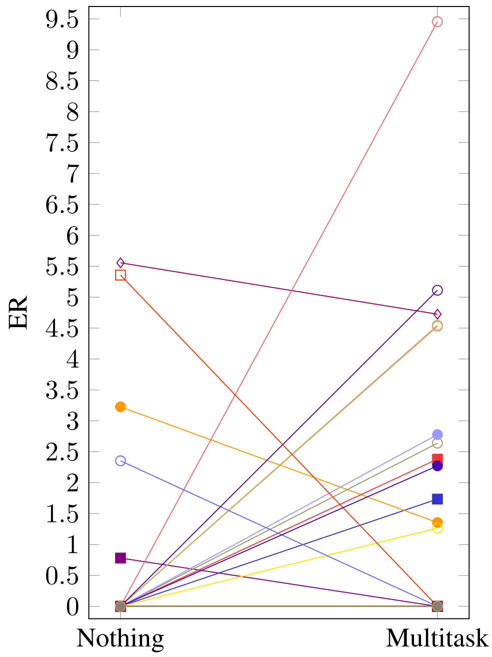
(a) The effect of multitasking on WPM



(b) The effect of multitasking on KSPS

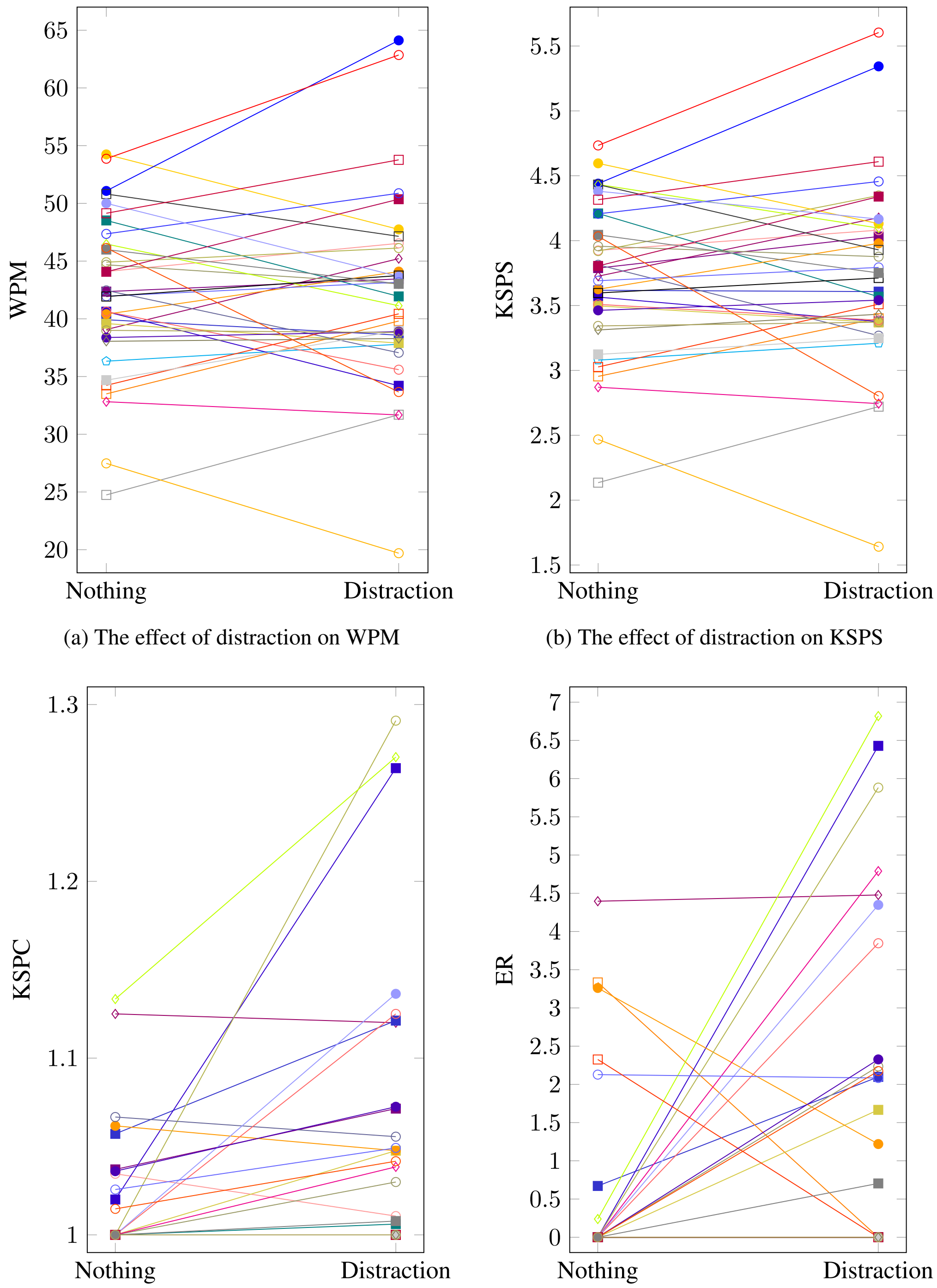


(c) The effect of multitasking on KSPC



(d) The effect of multitasking on ER

Figure E.4: The effect of multitasking on individual performances



(a) The effect of distraction on WPM

(b) The effect of distraction on KSPS

(c) The effect of distraction on KSPC

(d) The effect of distraction on ER

Figure E.5: The effect of distraction on individual performances

Appendix F

CONTEXT RECOGNITION RESULTS

F.1 Environment

Table F.1 illustrates environment context classification results of KNN, DT, RF, MLP, Linear SVM, and AdaBoost with 50% and 1 second sliding windows of 2, 5, 10 and 20 seconds on imbalanced and oversampled data.

Table F.1: The results for environment classification using different models

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
Imbalanced Data, 50% Overlap							
2s	KNN	0.43	0.44	0.86 ± 0.01	0.73 ± 0.02	0.78 ± 0.02	0.90 ± 0.01
	DT	1.12	0.06	0.68 ± 0.02	0.69 ± 0.02	0.68 ± 0.02	0.82 ± 0.01
	RF	5.86	0.10	0.90 ± 0.02	0.67 ± 0.02	0.71 ± 0.02	0.88 ± 0.01
	MLP	9.10	0.06	0.76 ± 0.05	0.56 ± 0.01	0.57 ± 0.01	0.84 ± 0.01
	Linear SVM	8.48	0.08	0.83 ± 0.02	0.62 ± 0.04	0.65 ± 0.04	0.86 ± 0.01
	AdaBoost	2.97	0.08	0.72 ± 0.06	0.60 ± 0.08	0.61 ± 0.11	0.85 ± 0.02
5s	KNN	0.23	0.18	0.85 ± 0.02	0.77 ± 0.03	0.80 ± 0.03	0.89 ± 0.01
	DT	0.42	0.03	0.72 ± 0.03	0.72 ± 0.02	0.72 ± 0.02	0.82 ± 0.02
	RF	2.07	0.06	0.90 ± 0.01	0.74 ± 0.02	0.79 ± 0.02	0.89 ± 0.01
	MLP	6.26	0.04	0.86 ± 0.02	0.80 ± 0.03	0.83 ± 0.02	0.90 ± 0.01
	Linear SVM	7.28	0.04	0.83 ± 0.02	0.70 ± 0.03	0.73 ± 0.03	0.86 ± 0.01
	AdaBoost	2.53	0.06	0.78 ± 0.03	0.70 ± 0.01	0.73 ± 0.02	0.85 ± 0.01
10s	KNN	0.10	0.04	0.86 ± 0.03	0.80 ± 0.02	0.82 ± 0.02	0.89 ± 0.01
	DT	0.12	0.02	0.73 ± 0.02	0.74 ± 0.03	0.74 ± 0.02	0.82 ± 0.01
	RF	0.77	0.03	0.92 ± 0.01	0.80 ± 0.02	0.84 ± 0.02	0.90 ± 0.01

Continued on next page

Table F.1 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
	MLP	2.48	0.02	**0.89 ± 0.03**	**0.88 ± 0.04**	**0.88 ± 0.03**	**0.92 ± 0.02**
	Linear SVM	1.14	0.02	0.82 ± 0.02	0.78 ± 0.01	0.79 ± 0.01	0.87 ± 0.01
	AdaBoost	0.70	0.03	0.83 ± 0.01	0.79 ± 0.03	0.81 ± 0.02	0.87 ± 0.01
20s	KNN	0.03	0.02	0.84 ± 0.02	0.81 ± 0.03	0.82 ± 0.02	0.88 ± 0.01
	DT	0.03	0.01	0.77 ± 0.02	0.78 ± 0.02	0.77 ± 0.02	0.83 ± 0.02
	RF	0.31	0.02	0.88 ± 0.03	0.80 ± 0.01	0.83 ± 0.01	0.89 ± 0.01
	MLP	0.89	0.01	0.89 ± 0.05	0.87 ± 0.05	0.88 ± 0.05	0.91 ± 0.03
	Linear SVM	0.08	0.02	0.54 ± 0.22	0.61 ± 0.16	0.57 ± 0.19	0.79 ± 0.05
	AdaBoost	0.23	0.02	0.78 ± 0.03	0.77 ± 0.01	0.77 ± 0.02	0.84 ± 0.02
Imbalanced Data, 1s Overlap							
2s	KNN	0.43	0.44	0.86 ± 0.01	0.73 ± 0.02	0.78 ± 0.02	0.90 ± 0.01
	DT	1.12	0.06	0.68 ± 0.02	0.69 ± 0.02	0.68 ± 0.02	0.82 ± 0.01
	RF	5.86	0.10	0.90 ± 0.02	0.67 ± 0.02	0.71 ± 0.02	0.88 ± 0.01
	MLP	9.10	0.06	0.76 ± 0.05	0.56 ± 0.01	0.57 ± 0.01	0.84 ± 0.01
	Linear SVM	8.48	0.08	0.83 ± 0.02	0.62 ± 0.04	0.65 ± 0.04	0.86 ± 0.01
	AdaBoost	2.97	0.08	0.72 ± 0.06	0.60 ± 0.08	0.61 ± 0.11	0.85 ± 0.02
5s	KNN	0.61	0.71	0.91 ± 0.01	0.84 ± 0.02	0.87 ± 0.02	0.92 ± 0.01
	DT	1.61	0.08	0.78 ± 0.01	0.78 ± 0.01	0.78 ± 0.01	0.86 ± 0.01
	RF	8.98	0.13	0.94 ± 0.01	0.80 ± 0.02	0.85 ± 0.02	0.92 ± 0.01
	MLP	13.54	0.09	0.95 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.96 ± 0.00
	Linear SVM	2.68	0.11	0.40 ± 0.00	0.50 ± 0.00	0.44 ± 0.00	0.80 ± 0.00
	AdaBoost	6.53	0.12	0.79 ± 0.01	0.72 ± 0.01	0.75 ± 0.01	0.86 ± 0.01
10s	KNN	0.55	0.37	0.95 ± 0.01	0.93 ± 0.01	0.94 ± 0.01	0.96 ± 0.00
	DT	0.76	0.05	0.87 ± 0.02	0.87 ± 0.02	0.87 ± 0.02	0.91 ± 0.01
	RF	5.68	0.09	0.97 ± 0.00	0.91 ± 0.01	0.94 ± 0.01	0.96 ± 0.01
	MLP	6.40	0.06	0.99 ± 0.00	0.98 ± 0.01	0.98 ± 0.01	0.99 ± 0.00
	Linear SVM	11.84	0.09	0.82 ± 0.04	0.80 ± 0.03	0.81 ± 0.03	0.87 ± 0.02
	AdaBoost	5.18	0.08	0.84 ± 0.01	0.79 ± 0.02	0.81 ± 0.02	0.88 ± 0.01
20s	KNN	0.61	0.35	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01
	DT	0.79	0.09	0.96 ± 0.01	0.95 ± 0.01	0.96 ± 0.01	0.97 ± 0.00
	RF	2.01	0.07	0.99 ± 0.00	0.98 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	MLP	4.38	0.04	**1.00 ± 0.00**	**1.00 ± 0.00**	**1.00 ± 0.00**	**1.00 ± 0.00**
	Linear SVM	8.65	0.05	0.85 ± 0.05	0.84 ± 0.03	0.84 ± 0.04	0.89 ± 0.03
	AdaBoost	2.19	0.06	0.88 ± 0.02	0.85 ± 0.01	0.86 ± 0.01	0.90 ± 0.01

Continued on next page

Table F.1 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
Oversampled Data, 1s Overlap							
2s	KNN	0.78	0.66	0.71 ± 0.01	0.81 ± 0.01	0.74 ± 0.01	0.81 ± 0.01
	DT	2.64	0.05	0.66 ± 0.01	0.71 ± 0.01	0.67 ± 0.01	0.79 ± 0.01
	RF	13.79	0.12	0.85 ± 0.01	0.78 ± 0.02	0.80 ± 0.02	0.90 ± 0.01
	MLP	19.69	0.07	0.85 ± 0.02	0.85 ± 0.01	0.85 ± 0.01	0.92 ± 0.01
	Linear SVM	19.39	0.15	0.69 ± 0.00	0.79 ± 0.01	0.71 ± 0.01	0.79 ± 0.01
	AdaBoost	8.82	0.08	0.67 ± 0.02	0.74 ± 0.02	0.69 ± 0.02	0.78 ± 0.01
5s	KNN	1.62	1.22	0.85 ± 0.01	0.92 ± 0.01	0.88 ± 0.01	0.91 ± 0.01
	DT	3.89	0.08	0.75 ± 0.01	0.79 ± 0.02	0.77 ± 0.01	0.84 ± 0.01
	RF	19.13	0.14	0.93 ± 0.01	0.88 ± 0.01	0.90 ± 0.01	0.94 ± 0.00
	MLP	14.87	0.09	0.95 ± 0.00	0.95 ± 0.01	0.95 ± 0.00	0.97 ± 0.00
	Linear SVM	44.61	0.20	0.75 ± 0.01	0.82 ± 0.01	0.77 ± 0.01	0.83 ± 0.01
	AdaBoost	14.06	0.12	0.71 ± 0.01	0.78 ± 0.01	0.73 ± 0.01	0.80 ± 0.01
10s	KNN	0.95	0.34	0.92 ± 0.02	0.96 ± 0.01	0.94 ± 0.02	0.96 ± 0.01
	DT	1.57	0.05	0.87 ± 0.01	0.89 ± 0.01	0.88 ± 0.01	0.91 ± 0.01
	RF	11.25	0.10	0.98 ± 0.00	0.96 ± 0.01	0.97 ± 0.01	0.98 ± 0.00
	MLP	7.74	0.07	0.98 ± 0.00	0.98 ± 0.00	0.98 ± 0.00	0.99 ± 0.00
	Linear SVM	27.89	0.10	0.80 ± 0.02	0.87 ± 0.01	0.83 ± 0.02	0.86 ± 0.02
	AdaBoost	8.87	0.08	0.79 ± 0.01	0.85 ± 0.01	0.81 ± 0.01	0.86 ± 0.00
20s	KNN	1.10	0.35	0.98 ± 0.01	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	DT	2.27	0.10	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.96 ± 0.01
	RF	6.80	0.12	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	MLP	5.55	0.06	$\mathbf{1.00 \pm 0.00}$	$\mathbf{0.99 \pm 0.00}$	$\mathbf{1.00 \pm 0.00}$	$\mathbf{1.00 \pm 0.00}$
	Linear SVM	15.26	0.06	0.82 ± 0.03	0.89 ± 0.02	0.84 ± 0.03	0.87 ± 0.02
	AdaBoost	5.06	0.08	0.85 ± 0.02	0.89 ± 0.01	0.87 ± 0.02	0.90 ± 0.01

F.2 Mobility

Table F.2 illustrates mobility context classification results of KNN, DT, RF, MLP, Linear SVM, and AdaBoost with 50% and 1 second sliding windows of 2, 5, 10 and 20 seconds on imbalanced and oversampled data.

Table F.2: The results for mobility classification using different models

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
Imbalanced Data, 50% Overlap							
2s	KNN	0.36	0.21	0.80 ± 0.02	0.61 ± 0.04	0.66 ± 0.05	0.95 ± 0.00
	DT	0.84	0.06	0.66 ± 0.03	0.67 ± 0.03	0.66 ± 0.03	0.93 ± 0.01
	RF	4.91	0.09	0.91 ± 0.04	0.61 ± 0.04	0.67 ± 0.05	0.96 ± 0.00
	MLP	7.83	0.06	0.77 ± 0.04	0.58 ± 0.02	0.61 ± 0.02	0.95 ± 0.00
	Linear SVM	0.96	0.07	0.47 ± 0.00	0.50 ± 0.00	0.49 ± 0.00	0.95 ± 0.00
	AdaBoost	2.05	0.08	0.77 ± 0.03	0.61 ± 0.03	0.65 ± 0.03	0.95 ± 0.00
5s	KNN	0.23	0.16	0.90 ± 0.02	0.81 ± 0.03	0.85 ± 0.02	0.96 ± 0.00
	DT	0.22	0.04	0.72 ± 0.02	0.72 ± 0.03	0.72 ± 0.02	0.92 ± 0.01
	RF	2.13	0.06	0.93 ± 0.02	0.71 ± 0.05	0.77 ± 0.05	0.95 ± 0.01
	MLP	5.71	0.04	0.83 ± 0.04	0.76 ± 0.02	0.79 ± 0.03	0.95 ± 0.01
	Linear SVM	1.76	0.04	0.61 ± 0.21	0.57 ± 0.10	0.58 ± 0.14	0.93 ± 0.01
	AdaBoost	1.03	0.05	0.80 ± 0.02	0.69 ± 0.03	0.73 ± 0.03	0.94 ± 0.01
10s	KNN	0.10	0.05	0.89 ± 0.01	0.82 ± 0.04	0.85 ± 0.02	0.95 ± 0.00
	DT	0.08	0.02	0.76 ± 0.05	0.77 ± 0.07	0.77 ± 0.06	0.92 ± 0.02
	RF	0.49	0.03	$\mathbf{0.94 \pm 0.02}$	$\mathbf{0.81 \pm 0.08}$	$\mathbf{0.86 \pm 0.07}$	$\mathbf{0.96 \pm 0.01}$
	MLP	2.02	0.02	0.88 ± 0.02	0.84 ± 0.03	0.86 ± 0.02	0.95 ± 0.01
	Linear SVM	0.55	0.01	0.84 ± 0.04	0.78 ± 0.04	0.81 ± 0.04	0.94 ± 0.01
	AdaBoost	0.56	0.02	0.83 ± 0.06	0.79 ± 0.04	0.81 ± 0.04	0.94 ± 0.02
20s	KNN	0.02	0.02	0.86 ± 0.07	0.81 ± 0.09	0.83 ± 0.08	0.94 ± 0.03
	DT	0.04	0.01	0.82 ± 0.07	0.79 ± 0.03	0.80 ± 0.04	0.93 ± 0.02
	RF	0.27	0.02	0.95 ± 0.03	0.78 ± 0.04	0.84 ± 0.04	0.95 ± 0.01
	MLP	0.62	0.01	0.89 ± 0.04	0.84 ± 0.06	0.86 ± 0.05	0.95 ± 0.02
	Linear SVM	0.09	0.01	0.83 ± 0.02	0.86 ± 0.08	0.84 ± 0.04	0.94 ± 0.01
	AdaBoost	0.19	0.02	0.85 ± 0.08	0.82 ± 0.11	0.83 ± 0.09	0.94 ± 0.03
Imbalanced Data, 1s Overlap							
2s	KNN	0.36	0.21	0.80 ± 0.02	0.61 ± 0.04	0.66 ± 0.05	0.95 ± 0.00
	DT	0.84	0.06	0.66 ± 0.03	0.67 ± 0.03	0.66 ± 0.03	0.93 ± 0.01
	RF	4.91	0.09	0.91 ± 0.04	0.61 ± 0.04	0.67 ± 0.05	0.96 ± 0.00
	MLP	7.83	0.06	0.77 ± 0.04	0.58 ± 0.02	0.61 ± 0.02	0.95 ± 0.00
	Linear SVM	0.96	0.07	0.47 ± 0.00	0.50 ± 0.00	0.49 ± 0.00	0.95 ± 0.00
	AdaBoost	2.05	0.08	0.77 ± 0.03	0.61 ± 0.03	0.65 ± 0.03	0.95 ± 0.00

Continued on next page

Table F.2 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
5s	KNN	0.90	0.69	0.92 ± 0.01	0.87 ± 0.01	0.89 ± 0.01	0.97 ± 0.00
	DT	0.56	0.09	0.77 ± 0.02	0.78 ± 0.02	0.77 ± 0.02	0.94 ± 0.00
	RF	5.72	0.13	0.96 ± 0.01	0.79 ± 0.03	0.86 ± 0.02	0.97 ± 0.00
	MLP	12.35	0.09	0.86 ± 0.04	0.78 ± 0.08	0.81 ± 0.07	0.95 ± 0.01
	Linear SVM	4.43	0.10	0.53 ± 0.15	0.54 ± 0.09	0.53 ± 0.11	0.93 ± 0.01
	AdaBoost	2.42	0.12	0.78 ± 0.03	0.66 ± 0.05	0.69 ± 0.05	0.94 ± 0.01
10s	KNN	0.57	0.35	0.97 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.99 ± 0.00
	DT	0.60	0.07	0.88 ± 0.03	0.89 ± 0.04	0.88 ± 0.03	0.96 ± 0.01
	RF	4.75	0.09	0.99 ± 0.01	0.89 ± 0.01	0.93 ± 0.01	0.98 ± 0.00
	MLP	5.75	0.06	0.95 ± 0.07	0.94 ± 0.05	0.94 ± 0.06	0.98 ± 0.02
	Linear SVM	5.25	0.06	0.79 ± 0.06	0.77 ± 0.05	0.77 ± 0.05	0.92 ± 0.02
	AdaBoost	2.58	0.08	0.84 ± 0.04	0.78 ± 0.03	0.80 ± 0.03	0.94 ± 0.01
20s	KNN	0.36	0.18	0.99 ± 0.01	0.98 ± 0.01	0.99 ± 0.01	0.99 ± 0.00
	DT	0.47	0.04	0.96 ± 0.01	0.95 ± 0.02	0.95 ± 0.01	0.98 ± 0.00
	RF	1.98	0.07	1.00 ± 0.00	0.97 ± 0.02	0.98 ± 0.01	0.99 ± 0.00
	MLP	6.31	0.05	0.94 ± 0.01	0.92 ± 0.02	0.93 ± 0.01	0.97 ± 0.00
	Linear SVM	4.09	0.04	0.87 ± 0.06	0.87 ± 0.04	0.87 ± 0.05	0.95 ± 0.02
	AdaBoost	2.09	0.07	0.93 ± 0.04	0.90 ± 0.04	0.91 ± 0.04	0.97 ± 0.01
Oversampled Data, 1s Overlap							
2s	KNN	1.39	0.72	0.65 ± 0.01	0.86 ± 0.01	0.70 ± 0.01	0.90 ± 0.01
	DT	2.12	0.06	0.61 ± 0.01	0.73 ± 0.01	0.64 ± 0.01	0.89 ± 0.01
	RF	20.72	0.11	0.79 ± 0.02	0.72 ± 0.02	0.75 ± 0.01	0.96 ± 0.00
	MLP	12.25	0.07	0.76 ± 0.02	0.75 ± 0.03	0.75 ± 0.03	0.95 ± 0.00
	Linear SVM	30.93	0.12	0.59 ± 0.01	0.79 ± 0.02	0.61 ± 0.01	0.84 ± 0.01
	AdaBoost	12.92	0.10	0.60 ± 0.00	0.77 ± 0.02	0.63 ± 0.01	0.86 ± 0.01
5s	KNN	2.04	0.89	0.80 ± 0.01	0.95 ± 0.00	0.85 ± 0.01	0.95 ± 0.01
	DT	4.05	0.08	0.75 ± 0.01	0.84 ± 0.02	0.78 ± 0.02	0.93 ± 0.00
	RF	29.94	0.15	0.89 ± 0.02	0.88 ± 0.02	0.88 ± 0.02	0.97 ± 0.00
	MLP	14.19	0.11	0.93 ± 0.01	0.92 ± 0.01	0.92 ± 0.01	0.98 ± 0.00
	Linear SVM	47.36	0.18	0.67 ± 0.01	0.87 ± 0.02	0.72 ± 0.02	0.87 ± 0.01
	AdaBoost	20.52	0.14	0.67 ± 0.01	0.84 ± 0.02	0.71 ± 0.02	0.88 ± 0.01
10s	KNN	1.36	0.33	0.92 ± 0.04	0.98 ± 0.01	0.95 ± 0.02	0.98 ± 0.01
	DT	2.97	0.06	0.87 ± 0.01	0.92 ± 0.01	0.89 ± 0.01	0.96 ± 0.00
	RF	13.86	0.10	0.96 ± 0.01	0.95 ± 0.02	0.95 ± 0.01	0.98 ± 0.00
	MLP	6.70	0.07	0.98 ± 0.01	0.99 ± 0.01	0.98 ± 0.01	0.99 ± 0.00

Continued on next page

Table F.2 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
	Linear SVM	25.18	0.10	0.73 ± 0.02	0.90 ± 0.02	0.78 ± 0.02	0.89 ± 0.02
	AdaBoost	8.87	0.09	0.76 ± 0.02	0.91 ± 0.01	0.81 ± 0.01	0.92 ± 0.01
20s	KNN	0.89	0.11	0.95 ± 0.01	0.99 ± 0.01	0.97 ± 0.01	0.99 ± 0.00
	DT	1.02	0.05	0.93 ± 0.02	0.96 ± 0.01	0.95 ± 0.01	0.98 ± 0.00
	RF	7.11	0.07	0.99 ± 0.00	0.99 ± 0.01	0.99 ± 0.00	1.00 ± 0.00
	MLP	4.24	0.05	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	Linear SVM	8.14	0.06	0.83 ± 0.01	0.95 ± 0.01	0.88 ± 0.00	0.94 ± 0.00
	AdaBoost	6.80	0.06	0.88 ± 0.03	0.95 ± 0.01	0.91 ± 0.02	0.96 ± 0.01

F.3 Social Context

Table F.3 illustrates social context classification results of KNN, DT, RF, MLP, Linear SVM, and AdaBoost with 50% and 1 second sliding windows of 2, 5, 10 and 20 seconds on imbalanced and oversampled data.

Table F.3: The results for social context classification using different models

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
Imbalanced Data, 50% Overlap							
2s	KNN	0.26	0.20	0.75 ± 0.01	0.70 ± 0.01	0.71 ± 0.01	0.77 ± 0.00
	DT	1.01	0.06	0.70 ± 0.01	0.70 ± 0.01	0.70 ± 0.01	0.74 ± 0.01
	RF	8.26	0.11	0.85 ± 0.01	0.71 ± 0.01	0.74 ± 0.01	0.80 ± 0.00
	MLP	10.05	0.06	0.74 ± 0.01	0.66 ± 0.01	0.67 ± 0.01	0.75 ± 0.00
	Linear SVM	36.11	0.09	0.81 ± 0.01	0.55 ± 0.00	0.50 ± 0.01	0.70 ± 0.00
	AdaBoost	2.95	0.09	0.69 ± 0.01	0.63 ± 0.01	0.64 ± 0.01	0.72 ± 0.00
5s	KNN	0.29	0.22	0.80 ± 0.02	0.74 ± 0.02	0.75 ± 0.02	0.80 ± 0.02
	DT	0.26	0.03	0.69 ± 0.02	0.69 ± 0.03	0.69 ± 0.03	0.73 ± 0.02
	RF	3.12	0.07	0.85 ± 0.01	0.76 ± 0.01	0.78 ± 0.01	0.83 ± 0.01
	MLP	4.76	0.04	0.77 ± 0.02	0.67 ± 0.01	0.68 ± 0.01	0.76 ± 0.01

Continued on next page

Table F.3 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Score	Accuracy
	Linear SVM	13.68	0.05	0.83 ± 0.02	0.59 ± 0.01	0.56 ± 0.02	0.72 ± 0.01
	AdaBoost	0.73	0.05	0.75 ± 0.01	0.63 ± 0.01	0.64 ± 0.02	0.74 ± 0.01
10s	KNN	0.08	0.04	0.85 ± 0.01	0.79 ± 0.02	0.81 ± 0.02	0.85 ± 0.01
	DT	0.16	0.02	0.73 ± 0.03	0.73 ± 0.04	0.72 ± 0.03	0.76 ± 0.02
	RF	1.18	0.03	0.86 ± 0.03	0.77 ± 0.03	0.79 ± 0.03	0.84 ± 0.02
	MLP	1.90	0.02	0.80 ± 0.04	0.70 ± 0.03	0.71 ± 0.04	0.79 ± 0.02
	Linear SVM	1.92	0.02	0.85 ± 0.02	0.63 ± 0.02	0.63 ± 0.03	0.76 ± 0.02
	AdaBoost	0.28	0.02	0.80 ± 0.02	0.69 ± 0.03	0.71 ± 0.03	0.78 ± 0.02
20s	KNN	0.03	0.02	0.83 ± 0.05	0.73 ± 0.04	0.75 ± 0.03	0.82 ± 0.02
	DT	0.04	0.01	0.75 ± 0.02	0.74 ± 0.02	0.74 ± 0.01	0.79 ± 0.01
	RF	0.44	0.02	$\mathbf{0.88 \pm 0.04}$	$\mathbf{0.80 \pm 0.04}$	$\mathbf{0.83 \pm 0.04}$	$\mathbf{0.87 \pm 0.02}$
	MLP	0.65	0.01	0.83 ± 0.02	0.74 ± 0.02	0.76 ± 0.02	0.83 ± 0.01
	Linear SVM	0.43	0.01	0.89 ± 0.03	0.69 ± 0.04	0.72 ± 0.05	0.82 ± 0.03
	AdaBoost	0.24	0.02	0.79 ± 0.04	0.75 ± 0.05	0.77 ± 0.05	0.82 ± 0.03
Imbalanced Data, 1s Overlap							
2s	KNN	0.26	0.20	0.75 ± 0.01	0.70 ± 0.01	0.71 ± 0.01	0.77 ± 0.00
	DT	1.01	0.06	0.70 ± 0.01	0.70 ± 0.01	0.70 ± 0.01	0.74 ± 0.01
	RF	8.26	0.11	0.85 ± 0.01	0.71 ± 0.01	0.74 ± 0.01	0.80 ± 0.00
	MLP	10.05	0.06	0.74 ± 0.01	0.66 ± 0.01	0.67 ± 0.01	0.75 ± 0.00
	Linear SVM	36.11	0.09	0.81 ± 0.01	0.55 ± 0.00	0.50 ± 0.01	0.70 ± 0.00
	AdaBoost	2.95	0.09	0.69 ± 0.01	0.63 ± 0.01	0.64 ± 0.01	0.72 ± 0.00
5s	KNN	0.72	0.73	0.87 ± 0.01	0.82 ± 0.01	0.84 ± 0.01	0.87 ± 0.01
	DT	1.29	0.09	0.78 ± 0.01	0.78 ± 0.01	0.78 ± 0.01	0.80 ± 0.01
	RF	11.01	0.15	0.90 ± 0.01	0.81 ± 0.02	0.83 ± 0.01	0.87 ± 0.01
	MLP	14.04	0.09	0.78 ± 0.01	0.73 ± 0.01	0.74 ± 0.01	0.79 ± 0.01
	Linear SVM	33.41	0.13	0.84 ± 0.01	0.60 ± 0.01	0.59 ± 0.01	0.74 ± 0.01
	AdaBoost	2.05	0.12	0.68 ± 0.02	0.63 ± 0.02	0.63 ± 0.02	0.72 ± 0.01
10s	KNN	0.38	0.40	0.92 ± 0.01	0.89 ± 0.01	0.90 ± 0.01	0.92 ± 0.01
	DT	0.76	0.05	0.87 ± 0.02	0.87 ± 0.02	0.87 ± 0.02	0.89 ± 0.02
	RF	6.80	0.11	0.95 ± 0.00	0.91 ± 0.01	0.92 ± 0.01	0.94 ± 0.01
	MLP	10.31	0.06	0.94 ± 0.01	0.92 ± 0.01	0.93 ± 0.01	0.94 ± 0.01
	Linear SVM	12.46	0.08	0.84 ± 0.01	0.67 ± 0.01	0.69 ± 0.01	0.78 ± 0.00
	AdaBoost	2.28	0.08	0.79 ± 0.01	0.74 ± 0.00	0.76 ± 0.01	0.80 ± 0.01

Continued on next page

Table F.3 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
20s	KNN	0.38	0.23	0.97 ± 0.00	0.96 ± 0.01	0.96 ± 0.00	0.97 ± 0.00
	DT	0.62	0.04	0.94 ± 0.02	0.94 ± 0.01	0.94 ± 0.01	0.95 ± 0.01
	RF	4.61	0.07	0.99 ± 0.00	0.98 ± 0.01	0.98 ± 0.00	0.99 ± 0.00
	MLP	6.93	0.05	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	Linear SVM	7.42	0.05	0.79 ± 0.08	0.71 ± 0.03	0.73 ± 0.04	0.80 ± 0.03
	AdaBoost	2.60	0.06	0.85 ± 0.02	0.81 ± 0.02	0.83 ± 0.02	0.86 ± 0.01
Oversampled Data, 1s Overlap							
2s	KNN	0.94	0.22	0.71 ± 0.01	0.72 ± 0.01	0.71 ± 0.01	0.73 ± 0.01
	DT	2.22	0.06	0.67 ± 0.01	0.68 ± 0.01	0.68 ± 0.01	0.70 ± 0.01
	RF	11.64	0.11	0.80 ± 0.01	0.77 ± 0.01	0.78 ± 0.01	0.82 ± 0.01
	MLP	15.74	0.06	0.74 ± 0.01	0.76 ± 0.01	0.75 ± 0.01	0.77 ± 0.01
	Linear SVM	16.73	0.13	0.61 ± 0.01	0.61 ± 0.01	0.61 ± 0.01	0.65 ± 0.02
	AdaBoost	3.97	0.09	0.63 ± 0.03	0.65 ± 0.03	0.63 ± 0.03	0.65 ± 0.03
5s	KNN	1.77	0.74	0.84 ± 0.03	0.86 ± 0.04	0.85 ± 0.03	0.86 ± 0.03
	DT	2.10	0.09	0.77 ± 0.02	0.79 ± 0.02	0.78 ± 0.02	0.80 ± 0.02
	RF	17.60	0.15	0.91 ± 0.01	0.87 ± 0.01	0.89 ± 0.01	0.90 ± 0.01
	MLP	25.37	0.09	0.88 ± 0.03	0.89 ± 0.02	0.88 ± 0.03	0.90 ± 0.02
	Linear SVM	20.45	0.17	0.65 ± 0.01	0.65 ± 0.01	0.65 ± 0.01	0.68 ± 0.01
	AdaBoost	12.69	0.13	0.70 ± 0.01	0.72 ± 0.01	0.70 ± 0.01	0.72 ± 0.01
10s	KNN	1.16	0.40	0.93 ± 0.01	0.95 ± 0.01	0.94 ± 0.01	0.95 ± 0.01
	DT	1.46	0.06	0.87 ± 0.01	0.88 ± 0.02	0.87 ± 0.02	0.89 ± 0.01
	RF	9.28	0.10	0.97 ± 0.00	0.94 ± 0.01	0.95 ± 0.01	0.96 ± 0.00
	MLP	15.91	0.06	0.95 ± 0.01	0.95 ± 0.01	0.95 ± 0.01	0.96 ± 0.01
	Linear SVM	14.30	0.11	0.68 ± 0.02	0.68 ± 0.01	0.68 ± 0.02	0.72 ± 0.02
	AdaBoost	3.63	0.08	0.73 ± 0.02	0.75 ± 0.02	0.74 ± 0.02	0.76 ± 0.02
20s	KNN	0.77	0.19	0.99 ± 0.01	0.99 ± 0.00	0.99 ± 0.01	0.99 ± 0.00
	DT	1.33	0.04	0.94 ± 0.01	0.95 ± 0.01	0.94 ± 0.01	0.95 ± 0.01
	RF	6.47	0.08	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	MLP	6.77	0.04	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	Linear SVM	15.18	0.07	0.78 ± 0.06	0.82 ± 0.06	0.79 ± 0.06	0.82 ± 0.05
	AdaBoost	4.86	0.06	0.83 ± 0.02	0.85 ± 0.01	0.84 ± 0.02	0.86 ± 0.01

F.4 Multitasking

Table F.4 illustrates multitasking context classification results of KNN, DT, RF, MLP, Linear SVM, and AdaBoost with 50% and 1 second sliding windows of 2, 5, 10 and 20 seconds on imbalanced and oversampled data.

Table F.4: The results for multitasking classification using different models

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
Imbalanced Data, 50% Overlap							
2s	KNN	0.29	0.38	0.76 ± 0.01	0.74 ± 0.01	0.75 ± 0.01	0.77 ± 0.01
	DT	0.80	0.06	0.65 ± 0.01	0.65 ± 0.01	0.65 ± 0.01	0.66 ± 0.01
	RF	8.42	0.11	0.80 ± 0.01	0.72 ± 0.01	0.73 ± 0.01	0.77 ± 0.01
	MLP	14.24	0.07	0.75 ± 0.05	0.74 ± 0.06	0.74 ± 0.06	0.76 ± 0.05
	Linear SVM	12.25	0.14	0.70 ± 0.00	0.61 ± 0.00	0.60 ± 0.01	0.68 ± 0.00
	AdaBoost	4.98	0.08	0.66 ± 0.01	0.63 ± 0.01	0.64 ± 0.01	0.67 ± 0.01
5s	KNN	0.25	0.17	0.78 ± 0.01	0.76 ± 0.02	0.76 ± 0.02	0.78 ± 0.02
	DT	0.56	0.03	0.67 ± 0.02	0.67 ± 0.01	0.67 ± 0.02	0.68 ± 0.02
	RF	4.41	0.07	0.80 ± 0.01	0.74 ± 0.01	0.75 ± 0.01	0.77 ± 0.01
	MLP	5.91	0.04	0.74 ± 0.05	0.71 ± 0.07	0.71 ± 0.07	0.74 ± 0.06
	Linear SVM	9.50	0.05	0.64 ± 0.02	0.59 ± 0.02	0.58 ± 0.02	0.64 ± 0.01
	AdaBoost	0.88	0.05	0.67 ± 0.01	0.64 ± 0.02	0.64 ± 0.02	0.68 ± 0.01
10s	KNN	0.05	0.04	0.73 ± 0.03	0.71 ± 0.03	0.72 ± 0.03	0.73 ± 0.02
	DT	0.11	0.01	0.69 ± 0.01	0.69 ± 0.01	0.69 ± 0.01	0.70 ± 0.01
	RF	1.03	0.03	0.82 ± 0.01	0.79 ± 0.01	0.80 ± 0.01	0.81 ± 0.01
	MLP	1.83	0.01	0.71 ± 0.02	0.67 ± 0.02	0.67 ± 0.03	0.70 ± 0.02
	Linear SVM	0.65	0.02	0.67 ± 0.03	0.63 ± 0.02	0.63 ± 0.02	0.67 ± 0.02
	AdaBoost	0.26	0.02	0.67 ± 0.03	0.65 ± 0.02	0.65 ± 0.02	0.67 ± 0.02
20s	KNN	0.02	0.02	0.84 ± 0.04	0.83 ± 0.04	0.83 ± 0.04	0.83 ± 0.04
	DT	0.03	0.01	0.71 ± 0.01	0.71 ± 0.02	0.71 ± 0.01	0.71 ± 0.01
	RF	0.37	0.02	0.84 ± 0.02	0.83 ± 0.02	0.83 ± 0.02	0.84 ± 0.02
	MLP	0.94	0.01	$\mathbf{0.85 \pm 0.02}$	$\mathbf{0.85 \pm 0.02}$	$\mathbf{0.85 \pm 0.02}$	$\mathbf{0.86 \pm 0.02}$
	Linear SVM	0.18	0.01	0.69 ± 0.04	0.67 ± 0.03	0.67 ± 0.04	0.69 ± 0.03
	AdaBoost	0.31	0.02	0.72 ± 0.05	0.72 ± 0.05	0.72 ± 0.05	0.72 ± 0.04

Continued on next page

Table F.4 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
Imbalanced Data, 1s Overlap							
2s	KNN	0.29	0.38	0.76 ± 0.01	0.74 ± 0.01	0.75 ± 0.01	0.77 ± 0.01
	DT	0.80	0.06	0.65 ± 0.01	0.65 ± 0.01	0.65 ± 0.01	0.66 ± 0.01
	RF	8.42	0.11	0.80 ± 0.01	0.72 ± 0.01	0.73 ± 0.01	0.77 ± 0.01
	MLP	14.24	0.07	0.75 ± 0.05	0.74 ± 0.06	0.74 ± 0.06	0.76 ± 0.05
	Linear SVM	12.25	0.14	0.70 ± 0.00	0.61 ± 0.00	0.60 ± 0.01	0.68 ± 0.00
	AdaBoost	4.98	0.08	0.66 ± 0.01	0.63 ± 0.01	0.64 ± 0.01	0.67 ± 0.01
5s	KNN	0.92	0.90	0.86 ± 0.00	0.85 ± 0.00	0.85 ± 0.00	0.86 ± 0.00
	DT	0.94	0.08	0.74 ± 0.01	0.74 ± 0.01	0.74 ± 0.01	0.75 ± 0.01
	RF	11.62	0.16	0.86 ± 0.01	0.81 ± 0.01	0.82 ± 0.01	0.84 ± 0.01
	MLP	21.05	0.10	0.87 ± 0.01	0.87 ± 0.00	0.87 ± 0.00	0.87 ± 0.00
	Linear SVM	32.17	0.21	0.71 ± 0.01	0.67 ± 0.01	0.67 ± 0.02	0.70 ± 0.01
	AdaBoost	4.82	0.11	0.68 ± 0.01	0.65 ± 0.01	0.65 ± 0.01	0.68 ± 0.01
10s	KNN	1.45	1.05	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	0.95 ± 0.02
	DT	1.20	0.06	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01	0.84 ± 0.01
	RF	9.74	0.18	0.94 ± 0.00	0.93 ± 0.01	0.94 ± 0.01	0.94 ± 0.01
	MLP	12.93	0.08	0.97 ± 0.01	0.97 ± 0.01	0.97 ± 0.01	0.97 ± 0.01
	Linear SVM	17.90	0.09	0.65 ± 0.02	0.62 ± 0.01	0.61 ± 0.01	0.65 ± 0.01
	AdaBoost	1.04	0.08	0.72 ± 0.02	0.69 ± 0.01	0.70 ± 0.01	0.72 ± 0.01
20s	KNN	0.39	0.22	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.01	0.96 ± 0.01
	DT	0.51	0.04	0.92 ± 0.01	0.92 ± 0.01	0.92 ± 0.01	0.92 ± 0.01
	RF	3.06	0.07	0.98 ± 0.02	0.97 ± 0.02	0.98 ± 0.02	0.98 ± 0.02
	MLP	6.57	0.04	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	Linear SVM	5.97	0.06	0.64 ± 0.01	0.63 ± 0.01	0.63 ± 0.01	0.65 ± 0.01
	AdaBoost	1.88	0.06	0.83 ± 0.01	0.82 ± 0.01	0.82 ± 0.01	0.83 ± 0.01
Oversampled Data, 1s Overlap							
2s	KNN	0.81	0.40	0.73 ± 0.01	0.75 ± 0.01	0.73 ± 0.01	0.74 ± 0.01
	DT	1.75	0.05	0.64 ± 0.01	0.65 ± 0.01	0.64 ± 0.01	0.65 ± 0.01
	RF	10.65	0.11	0.78 ± 0.01	0.75 ± 0.01	0.76 ± 0.01	0.78 ± 0.01
	MLP	19.90	0.07	0.77 ± 0.02	0.77 ± 0.02	0.77 ± 0.02	0.78 ± 0.02
	Linear SVM	16.66	0.15	0.63 ± 0.01	0.63 ± 0.01	0.63 ± 0.01	0.64 ± 0.01
	AdaBoost	7.44	0.09	0.63 ± 0.01	0.64 ± 0.01	0.63 ± 0.01	0.64 ± 0.01

Continued on next page

Table F.4 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
5s	KNN	1.89	1.22	0.83 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.83 ± 0.01
	DT	2.93	0.08	0.72 ± 0.01	0.73 ± 0.01	0.72 ± 0.01	0.73 ± 0.01
	RF	15.08	0.15	0.85 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.85 ± 0.01
	MLP	27.33	0.10	0.87 ± 0.01	0.87 ± 0.01	0.87 ± 0.01	0.87 ± 0.01
	Linear SVM	34.05	0.25	0.66 ± 0.03	0.65 ± 0.03	0.65 ± 0.04	0.67 ± 0.03
	AdaBoost	8.24	0.12	0.66 ± 0.02	0.66 ± 0.02	0.66 ± 0.02	0.67 ± 0.01
10s	KNN	1.44	0.44	0.94 ± 0.02	0.95 ± 0.02	0.94 ± 0.02	0.95 ± 0.02
	DT	1.61	0.06	0.83 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.83 ± 0.01
	RF	8.77	0.10	0.95 ± 0.00	0.94 ± 0.00	0.94 ± 0.00	0.94 ± 0.00
	MLP	11.77	0.08	0.97 ± 0.01	0.97 ± 0.01	0.97 ± 0.01	0.97 ± 0.01
	Linear SVM	9.52	0.10	0.61 ± 0.02	0.61 ± 0.02	0.61 ± 0.02	0.63 ± 0.02
	AdaBoost	1.96	0.07	0.70 ± 0.02	0.69 ± 0.01	0.70 ± 0.01	0.71 ± 0.01
20s	KNN	0.85	0.22	0.97 ± 0.01	0.97 ± 0.01	0.97 ± 0.01	0.97 ± 0.01
	DT	1.01	0.04	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01	0.91 ± 0.01
	RF	5.52	0.07	0.99 ± 0.00	0.98 ± 0.01	0.98 ± 0.00	0.98 ± 0.00
	MLP	7.47	0.05	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06
	Linear SVM	8.65	0.05	0.69 ± 0.07	0.69 ± 0.07	0.69 ± 0.07	0.69 ± 0.07
	AdaBoost	3.04	0.06	0.82 ± 0.02	0.82 ± 0.02	0.82 ± 0.02	0.83 ± 0.02

F.5 Distractions

Table F.5 illustrates distraction context classification results of KNN, DT, RF, MLP, Linear SVM, and AdaBoost with 50% and 1 second sliding windows of 2, 5, 10 and 20 seconds on imbalanced and oversampled data.

Table F.5: The results for distraction classification using different models

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Score	Accuracy
Imbalanced Data, 50% Overlap							
2s	KNN	0.25	0.24	0.78 ± 0.02	0.63 ± 0.02	0.66 ± 0.02	0.84 ± 0.01
	DT	1.07	0.06	0.65 ± 0.01	0.65 ± 0.01	0.65 ± 0.01	0.77 ± 0.01
	RF	6.70	0.10	0.91 ± 0.01	0.65 ± 0.01	0.69 ± 0.02	0.86 ± 0.01
	MLP	8.24	0.06	0.76 ± 0.04	0.57 ± 0.01	0.58 ± 0.01	0.82 ± 0.01
	Linear SVM	3.86	0.08	0.40 ± 0.00	0.50 ± 0.00	0.45 ± 0.00	0.80 ± 0.00
	AdaBoost	1.24	0.08	0.73 ± 0.03	0.54 ± 0.01	0.53 ± 0.02	0.81 ± 0.00
5s	KNN	0.24	0.24	0.82 ± 0.02	0.68 ± 0.02	0.72 ± 0.02	0.86 ± 0.01
	DT	0.30	0.04	0.65 ± 0.01	0.66 ± 0.02	0.65 ± 0.02	0.78 ± 0.01
	RF	3.89	0.06	0.91 ± 0.01	0.67 ± 0.01	0.72 ± 0.01	0.87 ± 0.01
	MLP	4.58	0.03	0.75 ± 0.03	0.58 ± 0.02	0.59 ± 0.03	0.82 ± 0.01
	Linear SVM	1.73	0.04	0.40 ± 0.00	0.50 ± 0.00	0.45 ± 0.00	0.81 ± 0.00
	AdaBoost	0.68	0.05	0.78 ± 0.03	0.56 ± 0.02	0.56 ± 0.04	0.82 ± 0.01
10s	KNN	0.11	0.04	0.81 ± 0.04	0.68 ± 0.01	0.71 ± 0.02	0.86 ± 0.01
	DT	0.16	0.02	0.65 ± 0.02	0.66 ± 0.02	0.66 ± 0.02	0.78 ± 0.02
	RF	1.09	0.03	0.87 ± 0.03	0.65 ± 0.02	0.69 ± 0.03	0.86 ± 0.01
	MLP	1.85	0.02	0.82 ± 0.05	0.65 ± 0.08	0.67 ± 0.08	0.85 ± 0.02
	Linear SVM	3.89	0.02	0.84 ± 0.02	0.59 ± 0.03	0.60 ± 0.04	0.84 ± 0.01
	AdaBoost	0.24	0.03	0.70 ± 0.07	0.57 ± 0.02	0.58 ± 0.03	0.82 ± 0.01
20s	KNN	0.02	0.02	0.86 ± 0.05	0.65 ± 0.03	0.69 ± 0.03	0.87 ± 0.01
	DT	0.04	0.01	0.66 ± 0.02	0.66 ± 0.04	0.66 ± 0.04	0.81 ± 0.02
	RF	0.36	0.02	0.90 ± 0.04	0.70 ± 0.06	0.74 ± 0.06	0.89 ± 0.02
	MLP	0.77	0.01	$\mathbf{0.87 \pm 0.03}$	$\mathbf{0.75 \pm 0.08}$	$\mathbf{0.78 \pm 0.08}$	$\mathbf{0.90 \pm 0.03}$
	Linear SVM	0.12	0.01	0.57 ± 0.21	0.55 ± 0.07	0.54 ± 0.12	0.84 ± 0.02
	AdaBoost	0.17	0.02	0.71 ± 0.09	0.62 ± 0.07	0.64 ± 0.09	0.84 ± 0.03
Imbalanced Data, 1s Overlap							
2s	KNN	0.25	0.24	0.78 ± 0.02	0.63 ± 0.02	0.66 ± 0.02	0.84 ± 0.01
	DT	1.07	0.06	0.65 ± 0.01	0.65 ± 0.01	0.65 ± 0.01	0.77 ± 0.01
	RF	6.70	0.10	0.91 ± 0.01	0.65 ± 0.01	0.69 ± 0.02	0.86 ± 0.01
	MLP	8.24	0.06	0.76 ± 0.04	0.57 ± 0.01	0.58 ± 0.01	0.82 ± 0.01
	Linear SVM	3.86	0.08	0.40 ± 0.00	0.50 ± 0.00	0.45 ± 0.00	0.80 ± 0.00
	AdaBoost	1.24	0.08	0.73 ± 0.03	0.54 ± 0.01	0.53 ± 0.02	0.81 ± 0.00

Continued on next page

Table F.5 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
5s	KNN	0.38	0.29	0.84 ± 0.01	0.73 ± 0.01	0.76 ± 0.01	0.88 ± 0.00
	DT	0.86	0.07	0.75 ± 0.02	0.75 ± 0.01	0.75 ± 0.01	0.84 ± 0.01
	RF	11.18	0.14	0.94 ± 0.00	0.73 ± 0.01	0.78 ± 0.01	0.89 ± 0.01
	MLP	13.87	0.08	0.81 ± 0.04	0.65 ± 0.00	0.68 ± 0.01	0.85 ± 0.01
	Linear SVM	3.34	0.14	0.40 ± 0.00	0.50 ± 0.00	0.45 ± 0.00	0.81 ± 0.00
	AdaBoost	1.54	0.10	0.74 ± 0.05	0.53 ± 0.01	0.51 ± 0.02	0.81 ± 0.01
10s	KNN	0.50	0.37	0.93 ± 0.01	0.84 ± 0.02	0.88 ± 0.02	0.93 ± 0.01
	DT	0.93	0.05	0.82 ± 0.02	0.83 ± 0.02	0.82 ± 0.02	0.89 ± 0.01
	RF	7.64	0.10	0.96 ± 0.00	0.82 ± 0.01	0.87 ± 0.01	0.93 ± 0.01
	MLP	9.67	0.05	0.95 ± 0.06	0.91 ± 0.13	0.92 ± 0.11	0.96 ± 0.05
	Linear SVM	1.47	0.07	0.41 ± 0.00	0.50 ± 0.00	0.45 ± 0.00	0.82 ± 0.00
	AdaBoost	1.07	0.07	0.75 ± 0.03	0.57 ± 0.01	0.59 ± 0.02	0.83 ± 0.00
20s	KNN	0.35	0.15	0.93 ± 0.03	0.90 ± 0.02	0.91 ± 0.03	0.95 ± 0.01
	DT	0.47	0.04	0.92 ± 0.01	0.92 ± 0.02	0.92 ± 0.02	0.95 ± 0.01
	RF	4.27	0.07	0.99 ± 0.00	0.95 ± 0.01	0.97 ± 0.00	0.98 ± 0.00
	MLP	4.62	0.04	0.98 ± 0.04	0.95 ± 0.09	0.96 ± 0.07	0.98 ± 0.04
	Linear SVM	1.21	0.04	0.42 ± 0.00	0.50 ± 0.00	0.45 ± 0.00	0.83 ± 0.00
	AdaBoost	0.73	0.05	0.78 ± 0.03	0.65 ± 0.01	0.69 ± 0.01	0.86 ± 0.01
Oversampled Data, 1s Overlap							
2s	KNN	0.78	0.24	0.65 ± 0.02	0.72 ± 0.02	0.66 ± 0.02	0.73 ± 0.02
	DT	2.88	0.06	0.64 ± 0.02	0.67 ± 0.02	0.65 ± 0.02	0.74 ± 0.02
	RF	15.46	0.11	0.83 ± 0.02	0.73 ± 0.02	0.77 ± 0.02	0.87 ± 0.01
	MLP	26.84	0.07	0.77 ± 0.02	0.77 ± 0.01	0.77 ± 0.01	0.86 ± 0.01
	Linear SVM	34.52	0.19	0.59 ± 0.01	0.63 ± 0.01	0.58 ± 0.01	0.65 ± 0.01
	AdaBoost	8.94	0.10	0.60 ± 0.01	0.65 ± 0.01	0.60 ± 0.02	0.67 ± 0.02
5s	KNN	1.50	1.21	0.80 ± 0.03	0.88 ± 0.03	0.83 ± 0.03	0.88 ± 0.03
	DT	3.78	0.07	0.72 ± 0.03	0.77 ± 0.03	0.74 ± 0.03	0.82 ± 0.02
	RF	22.63	0.14	0.93 ± 0.01	0.84 ± 0.00	0.88 ± 0.00	0.93 ± 0.00
	MLP	29.96	0.10	0.90 ± 0.01	0.89 ± 0.01	0.89 ± 0.01	0.93 ± 0.00
	Linear SVM	60.43	0.35	0.63 ± 0.01	0.70 ± 0.01	0.63 ± 0.01	0.70 ± 0.01
	AdaBoost	17.33	0.13	0.62 ± 0.00	0.68 ± 0.01	0.63 ± 0.01	0.71 ± 0.01

Continued on next page

Table F.5 – *Continued from previous page*

Win.	Method	Fit Time	Score Time	Precision	Recall	F1 Scrore	Accuracy
10s	KNN	1.02	0.49	0.90 ± 0.01	0.96 ± 0.01	0.93 ± 0.01	0.95 ± 0.01
	DT	2.84	0.05	0.78 ± 0.02	0.82 ± 0.02	0.80 ± 0.02	0.87 ± 0.01
	RF	13.24	0.09	0.97 ± 0.00	0.92 ± 0.01	0.94 ± 0.01	0.97 ± 0.00
	MLP	10.62	0.06	0.97 ± 0.01	0.96 ± 0.00	0.96 ± 0.00	0.98 ± 0.00
	Linear SVM	31.60	0.15	0.67 ± 0.01	0.75 ± 0.02	0.68 ± 0.01	0.75 ± 0.01
	AdaBoost	11.34	0.08	0.66 ± 0.01	0.74 ± 0.01	0.68 ± 0.01	0.76 ± 0.01
20s	KNN	0.79	0.22	0.98 ± 0.01	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00
	DT	1.19	0.04	0.92 ± 0.01	0.94 ± 0.01	0.93 ± 0.01	0.96 ± 0.01
	RF	7.06	0.07	1.00 ± 0.00	0.98 ± 0.01	0.99 ± 0.00	0.99 ± 0.00
	MLP	6.09	0.04	0.99 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	1.00 ± 0.00
	Linear SVM	14.30	0.07	0.74 ± 0.03	0.84 ± 0.02	0.77 ± 0.03	0.83 ± 0.03
	AdaBoost	7.22	0.06	0.78 ± 0.01	0.87 ± 0.02	0.81 ± 0.02	0.88 ± 0.01

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name:	Akpınar, Mehmet Elgin
Nationality:	Turkish (TC)
Date of Birth:	06.06.1988
Place of Birth:	Muğla
Marital Status:	Married
Phone:	+90 555 564 54 27
Email:	elgin.akpinar@metu.edu.tr

EDUCATION

Degree	**Institution**	**Year of Graduation**
M.S.	METU Computer Engineering	2014
B.S.	METU Computer Engineering	2011
High School	Muğla 75. Yıl Fen Lisesi	2006

PROFESSIONAL EXPERIENCE

Year	**Place**	**Enrollment**
2019 – Present	Mobiliz	Software Team Manager
2016 – 2019	Mobiliz	Software Design Leader
2015 – 2016	Comodo	Software Team Leader
2011 – 2015	TechNarts	Software Engineer
2010 – 2011	MikroBeta	Software Engineer, Part-time
2010	Minder IT Solutions	Software Engineer, Intern
2009	YD Software Co. Inc	Software Engineer, Intern

PUBLICATIONS

1. Elgin Akpinar, Yeliz Yeşilada, and Pınar Karagöz. 2022. Effect of Context on Smartphone Users' Typing Performance in the Wild *ACM Transactions on Computer-Human Interaction*, 43 pages. In submission (minor revision received)

2. Elgin Akpinar, Yeliz Yeşilada, and Selim Temizer, "The Effect of Context on Small Screen and Wearable Device Users' Performance - A Systematic Review." *ACM Computing Surveys* 53, 3, Article 52, 44 pages, ACM, 2020.

3. Elgin Akpinar, Yeliz Yeşilada, and Selim Temizer, "Ability and Context Based Adaptive System: A Proposal for Machine Learning Approach", in *Proceedings of the CHI'19 Workshop: Addressing the Challenges of Situationally-Induced Impairments and Disabilities in Mobile Interaction*, 8 pages, 2019.

4. Carlos Duarte, Ana Salvado, M. Elgin Akpinar, Yeliz Yeşilada, and Luís Carriço, "Automatic Role Detection of Visual Elements of Web Pages for Automatic Accessibility Evaluation", in *Proceedings of the 15th Int. Web for All Conference (W4A '18)*, Article 21, pp. 1–4, ACM, 2018.

5. M. Elgin Akpinar, Yeliz Yeşilada, "Discovering Visual Elements of Web Pages and Their Roles: Users' Perception", *Interacting with Computers*, Volume 29, Issue 6, Pages 845–867, 2017.

6. Elgin Akpınar and Yeliz Yeşilada, "'Old habits die hard!': eyetracking based experiential transcoding: a study with mobile users", in *Proceedings of the 12th Int. Web for All Conference (W4A '15)*, Article 12, pp. 1–5, ACM, 2015.

7. M. Elgin Akpınar, Yeliz Yeşilada, "Heuristic Role Detection of Visual Elements of Web Pages", in Daniel, F., Dolog, P., Li, Q. (eds) *Web Engineering. ICWE 2013. Lecture Notes in Computer Science*, vol 7977, Springer, 2013.

8. M. Elgin Akpınar, Yeliz Yeşilada, "Vision Based Page Segmentation Algorithm: Extended and Perceived Success", in Sheng, Q.Z., Kjeldskov, J. (eds) *Current Trends in Web Engineering. ICWE 2013. Lecture Notes in Computer Science*, vol 8295, Springer, 2013.

9. Serap. M. Altunkan, Alper Yasemin, İ. Taha Aykaç and Elgin Akpinar, “Turkish pharmaceuticals track & trace system”, *7th International Symposium on Health Informatics and Bioinformatics*, pp. 24-30, IEEE, 2012.