

Multichannel electroencephalogram frequency domain feature extraction method for brain activity state transition detection

Stančin, Igor

Doctoral thesis / Disertacija

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Electrical Engineering and Computing / Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:168:341252>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-05-20**



Repository / Repozitorij:

[FER Repository - University of Zagreb Faculty of Electrical Engineering and Computing repository](#)





University of Zagreb
FACULTY OF ELECTRICAL ENGINEERING AND
COMPUTING

Igor Stančin

**MULTICHANNEL
ELECTROENCEPHALOGRAM
FREQUENCY DOMAIN FEATURE
EXTRACTION METHOD FOR BRAIN
ACTIVITY STATE TRANSITION DETECTION**

DOCTORAL THESIS

Zagreb, 2023



University of Zagreb
FACULTY OF ELECTRICAL ENGINEERING AND
COMPUTING

Igor Stančin

**MULTICHANNEL
ELECTROENCEPHALOGRAM
FREQUENCY DOMAIN FEATURE
EXTRACTION METHOD FOR BRAIN
ACTIVITY STATE TRANSITION DETECTION**

DOCTORAL THESIS

Supervisor: Associate Profesor Alan Jović, PhD

Zagreb, 2023



Sveučilište u Zagrebu
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Igor Stančin

**METODA IZLUČIVANJA ZNAČAJKI
VIŠEKANALNOGA
ELEKTROENCEFALOGRAMA U
FREKVENCIJSKOJ DOMENI ZA
DETEKCIJU PROMJENE STANJA
MOŽDANE AKTIVNOSTI**

DOKTORSKI RAD

Mentor: izv. prof. dr. sc. Alan Jović

Zagreb, 2023.

This doctoral thesis was completed at the University of Zagreb Faculty of Electrical Engineering and Computing, in the Department of Electronics, Microelectronics, Computer and Intelligent Systems.

It was supported in part by the European Regional Development Fund under Grant KK.01.2.1.01.0136 (DFDM)

Supervisor: Associate Profesor Alan Jović, Ph.D.

The thesis has 133 pages.

Thesis No.: _____

About the Supervisor

Alan Jović was born in Zagreb in 1982. He received his Dipl. Ing. and PhD degrees in computer science from the University of Zagreb Faculty of Electrical Engineering and Computing (FER), Zagreb, Croatia, in 2006 and 2012, respectively.

From September 2006 till March 2007, he worked as expert associate at the Ruđer Bošković Institute. Since April 2007, he has been working at the Department of Electronics, Microelectronics, Computer and Intelligent Systems at FER. In June 2016, he became assistant professor, and in April 2020, associate professor at FER.

He is the author of 24 published papers in A-category scientific journals, 3 scientific book chapters, 2 papers in other scientific journals, and 44 papers in proceedings of international conferences. His works were cited more than 800 times according to the Web of Science database (h-index 14). He is Editor-in-Chief of the scientific journal "CIT. Journal of Computing and Information Technology". He was a program committee member for about 15 international conferences and held several invited lectures. He is a program committee member of the international ICT convention MIPRO as well as conference chair of MIPRO's "Artificial Intelligence Systems" (AIS) conference. He received several national and international awards and acknowledgments for his scientific work.

He was principal investigator on four scientific and industrial national projects and collaborator on six national and international projects. His professional areas of interest include data mining, application of computing in medicine, and knowledge representation in computer systems. He is a member of IEEE, EMBS and MIPRO organizations.

O mentoru

Alan Jović rođen je u Zagrebu 1982. godine. Diplomirao je i doktorirao računarstvo na Sveučilištu u Zagrebu Fakultetu elektrotehnike i računarstva (FER), Zagreb, Hrvatska, 2006. odnosno 2012. godine.

Od rujna 2006. do ožujka 2007. radio je kao stručni suradnik na Institutu Ruđer Bošković. Od travnja 2007. radi na Zavodu za elektroniku, mikroelektroniku, računalne i inteligentne sustave na FER-u. U lipnju 2016. postaje docent, a u travnju 2020. izvanredni profesor na FER-u.

Autor je 24 rada objavljenih u znanstvenim časopisima A-kategorije, 3 poglavlja u znanstvenim knjigama, 2 rada u drugim znanstvenim časopisima i 44 rada u zbornicima međunarodnih konferencija. Njegovi radovi citirani su više od 800 puta prema bazi podataka Web of Science (h-indeks 14). Glavni je urednik znanstvenog časopisa "CIT. Journal of Computing and Information Technology". Bio je član programskog odbora na 15-ak međunarodnih konferencija i održao je nekoliko pozvanih predavanja. Član je programskog odbora međunarodne ICT konferencije MIPRO, kao i voditelj MIPRO-ovog savjetovanja "Sustavi umjetne inteligencije" (AIS). Za svoj znanstveni rad dobio je nekoliko nacionalnih i međunarodnih nagrada i priznanja.

Bio je glavni istraživač na četiri znanstvena i industrijska nacionalna projekta i suradnik na šest nacionalnih i međunarodnih projekata. Njegova profesionalna područja od interesa uključuju dubinsku analizu podataka, primjenu računarstva u medicini i predstavljanje znanja u računalnim sustavima. Član je organizacija IEEE, EMBS i MIPRO.

Preface

I am grateful to my mentor, Associate Professor Alan Jović, for the opportunity and invaluable guidance throughout my entire doctoral journey. I also extend my thanks to all the colleagues with whom I collaborated during my PhD studies.

I would like to thank my family and friends for their support. A special note of appreciation goes to my wife, Ana, for her support and understanding.

Abstract

Focusing on the improvements in detecting brain state transitions, this thesis specifically investigates the transition from a waking state to a drowsy state during driving. Driver drowsiness is a significant road safety concern that contributes to numerous accidents and fatalities worldwide. The conducted research focuses on the detection and prediction of driver drowsiness using multichannel electroencephalography (EEG) signals. The published review paper [Pub1] systematically examines EEG signal features, drowsiness detection systems, and discusses potential improvements. A key insight found during the review is that current methods focus on a single brain region, while frequency-domain features from different brain regions behave differently during transition to drowsiness state.

Based on this insight, a novel brain state transition detection method that extracts multichannel frequency-domain features for improved drowsiness identification was developed [Pub2]. Further research also revealed that male and female subjects have different brain states and brain state transitions when experiencing drowsiness during driving [Pub4]. Incorporating information on the driver's sex improved drowsiness detection accuracy, and a reliable sex classifier was constructed using EEG data [Pub3].

For future research, developing a unified definition of drowsiness and creating a publicly accessible database with diverse datasets would enable more reliable comparisons of different models. Additionally, enhancing the developed method by allowing operations other than adding in the construction of the ratio-based indices could yield even better results.

Prošireni sažetak

Naslov:

Metoda izlučivanja značajki višekanalnoga elektroencefalograma u frekvencijskoj domeni za detekciju promjene stanja moždane aktivnosti

Prošireni sažetak:

Ova disertacija usmjerena je na poboljšanje detekcije prijelaza između stanja moždane aktivnosti, s posebnim fokusom na prijelaz iz budnog stanja u stanje pospanosti tijekom vožnje. Pospanost vozača veliki je problem za sigurnost na cestama jer značajno pridonosi nesrećama i smrtnim slučajevima diljem svijeta. Smanjena sposobnost prosuđivanja i prostorne percepcije te odgođeno vrijeme reakcije uzrokovani pospanošću ometaju vozačevu sposobnost da učinkovito reagira na dinamične uvjete na cesti. Rješavanje pospanosti vozača važno je za spašavanje života i ublažavanje ekonomskih posljedica povezanih s prometnim nesrećama. Pospanost tijekom vožnje može se mjeriti ljestvicama za subjektivnu procjenu, matematičkim modelima dinamike spavanja, mjerama temeljenim na karakteristikama vožnje, mjerama temeljenim na ponašanju vozača u vožnji te mjerenjem elektrofizioloških signala vozača.

Subjektivne ljestvice pospanosti od ispitanika zahtijevaju periodičnu samoprocjenu pospanosti zbog čega nisu prikladne za mjerenje u realnim uvjetima već se pretežito koriste u laboratorijskom okruženju. Najpoznatija subjektivna ljestvica pospanosti naziva se Karolinska ljestvica pospanosti gdje se pospanost procjenjuje na ljestvici od jedan do devet. Svaka razina pospanosti ima pridijeljeni tekstni opis, a dodatni nedostatak primjene Karolinske ljestvice je mala razlika među opisima susjednih razina. Zbog toga je i korištenje ljestvice otežano. Matematički modeli dinamike spavanja također nisu prikladni za korištenje u realnim uvjetima jer se temelje na malom broju parametara i prvenstveno su zamišljeni za primjenu kod određivanja smjena radnika.

Modeli temeljeni na karakteristikama vožnje su trenutno najčešći u upotrebi u realnim uvjetima. Senzori automobila daju informacije o parametrima vožnje kao što su promjene kuta volana, frekvencija promjene kuta volana, jačina pritiska papučice gasa, pozicija automobila u traci i slično. Praćenjem navedenih parametara kroz vrijeme može se napraviti procjena pospanosti vozača. Takve procjene temelje se na istraživanjima koja su utvrdila određene korelacije s pospanošću. Glavni nedostatak ovih modela je njihova ovisnost o vozačevoj motivaciji (vozač svjesno može utjecati na parametre koji se mjere) i nemogućnost rane detekcije pospanosti. Modeli temeljeni na ponašanju vozača su obećavajuć pristup za koji se

pretpostavlja da će u budućnosti biti korišten u realnim uvjetima. U većini slučajeva takvi sustavi oslanjaju se na informacije s kamera u prostoru vozača automobila. Glavni problem je što tehnologija još uvijek nije dovoljno kvalitetna da bi se iz slika vozača kontinuirano i pouzdano izlučile ponašajne značajke vozača. Značajke koje se često koriste u ovim sustavima su stupanj zatvorenosti oka, broj treptaja, frekvencija treptaja, pozicija glave, detektirano zijevanje i slično.

Mjerenje elektrofizioloških signala vozača također predstavlja pouzdanu i obećavajuću metodu detekcije pospanosti. Glavni nedostatak ovog pristupa je njegova nametljivost, koja ujedno narušava slobodu kretanja vozača u vožnji. Zbog navedenog nedostatka malo je vjerojatno da će takav pristup biti korišten u realnim uvjetima ukoliko ne dođe do drastičnog smanjenja nametljivosti korištenih tehnologija. Međutim, zbog visoke pouzdanosti tih sustava, oni su često korišteni u laboratorijskom okruženju. Najčešće korišteni elektrofiziološki signali za detekciju pospanosti vozača su elektroencefalogram (EEG), elektrokardiogram i elektromiogram. U ovoj disertaciji pospanost se detektira korištenjem signala EEG-a.

U sklopu istraživanja proveden je opsežan pregled literature [Pub1], sustavno ispitujući značajke signala EEG-a i sustave za otkrivanje pospanosti. Pregledni rad dao je četiri ključna doprinosa: (1) temeljito ispitivanje i sistematizacija značajki izlučenih iz EEG signala, (2) opsežan pregled postojećih sustava temeljenih na EEG signalu za otkrivanje pospanosti vozača, (3) komparativnu analizu sličnih pregleda i (4) raspravu o mogućim poboljšanjima najsuvremenijih sustava za detekciju pospanosti kod vozača. Osim toga, pregledni rad je dao i nekoliko prijedloga i smjernica za buduća istraživanja, kao što je povećanje veličine uzorka, validacija sustava za otkrivanje pospanosti temeljenih na EEG-u s neviđenim ispitanicima i stavljanje javno dostupnih skupova podataka koji bi omogućili poštene usporedbe različitih modela.

Pregledane značajke EEG-a svrstane su u kategorije: značajke vremenske domene, frekvencijske domene, vremensko-frekvencijske domene, nelinearne značajke, entropije, prostorno-vremenske značajke i značajke kompleksnih mreža. U istraživanju su korišteni strogi kriteriji za uključivanje članaka kako bi se osiguralo da samo najkvalitetniji sustavi za detekciju pospanosti budu uključeni. Kriteriji su: 1) članak je objavljen u znanstvenom časopisu, 2) objavljen je 2010. godine ili kasnije, 3) članak ima minimalno 3 citata za svaku godinu od objave, 4) članci iz zadnje dvije godine koji nemaju dovoljan broj citata uključeni su ako su objavljeni u časopisu iz prvog kvartila prema bazi Web of Science, i 5) broj ispitanika u eksperimentu mora biti veći od deset. Koristeći navedene kriterije uključeno je 39 članaka. Najzastupljenije su značajke frekvencijske domene koje se nalaze u ukupno 24 članka.

Pregledani sustavi za detekciju pospanosti uglavnom rade binarnu klasifikaciju, uz svega tri sustava koji rade klasifikaciju u više od dvije klase. Točnosti su u rasponu od 67% do 99,5%, s medijanom od 93,5%. Prilikom označavanja ciljne varijable signala za periode pospanosti dominira pristup subjektivne ljestvice pospanosti.

Važan uvid stečen tijekom postupka pregleda bio je da se trenutne metode fokusiraju na jednu regiju mozga za izdvajanje značajki, unatoč činjenici da se značajke frekvencijske domene iz različitih regija mozga ponašaju različito prilikom prijelaza iz stanja budnosti u stanje pospanosti. Na temelju ovog zapažanja razvijena je nova metoda detekcije prijelaza stanja moždane aktivnosti koja izdvaja značajke frekvencijske domene iz višekanalnog EEG-a kako bi preciznije identificirala pospanost [Pub2]. Predložena metoda je pouzdana, a značajke frekvencijske domene temeljene na signalima iz više regija mozga nadmašuju one koje se temelje na jednoj regiji mozga u raznim metrikama. Također, metoda se može primijeniti u drugim domenama detekcije prijelaza stanja moždane aktivnosti.

Razvijena metoda koristi višekriterijsku optimizaciju. Optimiraju se parametri za svaku pojedinu frekvencijsku značajku kako bi se dobio višekanalni omjer (omjer frekvencijskih značajki iz različitih regija mozga). Vrijednostima parametara određuje se utjecaj dodijeljene mu frekvencijske značajke na višekanalni omjer. Tako, primjerice, postavljanjem parametra na nulu isključujemo utjecaj dodijeljene mu značajke u konačnom višekanalnom omjeru, dok povećanjem vrijednosti parametra povećavamo i utjecaj dodijeljene mu značajke na konačni višekanalni omjer.

Za višekriterijsku optimizaciju koristi se genetski algoritam *Non-dominated sorting genetic algorithm II* (NSGA-II). Konačni cilj je dobiti višekanalni omjer koji će izgledom biti što je moguće sličniji funkciji skoka. Zbog toga višekriterijska optimizacija koristi dvije suprotstavljene ciljne funkcije. Prva ciljna funkcija genetskog algoritma nagrađuje jedinke koje imaju veliku razliku između srednjih vrijednosti signala lijevo od skoka i desno od skoka. Ovakva funkcija sama po sebi nije dovoljna jer rezultira mnogim rješenjima u kojima nema vidljivog skoka, već se na jednoj od strana signala pojavljuje veliki šiljak koji onda stvara privid velike razlike srednjih vrijednosti lijeve i desne strane. Zbog tog problema uvedena je druga ciljna funkcija genetskog algoritma – minimizacija oscilacija signala višekanalnog omjera lijevo i desno od skoka. Takvom ciljnom funkcijom kažnjavaju se jedinke koje u sebi imaju velike oscilacije i velike šiljke.

Kako u stvarnosti nije moguće ostvariti stvarnu funkciju skoka s trenutnim prijelazom iz jedna razine u drugu razinu uvedena je dodatna prilagodba u vidu prijelaznog prozora. Pomični prijelazni prozor korišten je u blizini sredine signala, odnosno u području gdje se očekuje

promjena iz jednog stanja u drugo. Signal unutar prijelaznog prozora se zanemaruje te se na tom dijelu signala ne računaju vrijednosti ranije objašnjenih ciljnih funkcija. Takvim prijelaznim prozorom signalu se daje vremena da promijeni vrijednost bez da pritom bude kažnjen temeljem ciljnih funkcija.

Razvijena metoda primjenjiva je na bilo koji problem detekcije promjene stanja moždane aktivnosti, a u okviru ove disertacije njena primjena isprobana je na primjeru detekcije pospanosti [Pub2]. U tu svrhu korišteni su javno dostupni podatci, a kao područje pospanosti korištena je prva faza sna. Korištenjem ove metode dobivena su dva višekanalna omjera. Usporedbom sa sedam postojećih jednokanalnih frekvencijskih omjera pokazano je da dobiveni višekanalni frekvencijski omjeri imaju višu točnost i preciznost, bolje statističke pokazatelje, te smanjuju vrijeme potrebno za detekciju pospanosti.

Dodatna istraživanja otkrila su da muški i ženski ispitanici pokazuju različita stanja moždane aktivnosti i prijelaze među stanjima moždane aktivnosti prilikom pospanosti tijekom vožnje [Pub4]. U sklopu ovog istraživanja korištene su značajke frekvencijske domene i značajke kvantificirajuće analize ponavljanja. Statističkim testovima pokazano je kako se pojedine značajke bitno razlikuju kod muškaraca i žena tijekom vožnje u pospanom stanju. Također, istraživanjem je utvrđeno kako postoje i značajne razlike u korelaciji značajki u različitim regijama mozga. Muški ispitanici imaju značajno više korelacije u usporedbi sa ženskim ispitanicima što može biti indikator jače sinkronizacije moždanih aktivnosti kod muškaraca.

Razlike u značajkama EEG signala među muškarcima i ženama otprije su poznate, no većina istraživanja koja se bavila tom problematikom promatrala je ispitanike tijekom odmora sa zatvorenim očima. Ovim istraživanjem potvrđeno je da razlika u značajkama EEG signala postoji i tijekom kompleksne aktivnosti kao što je vožnja automobila.

Slično vrijedi i za istraživanja koja su razvijala modele klasifikacije spola ispitanika temeljem značajki EEG signala. Postojeći modeli razvijani su koristeći podatke ispitanika u sjedećem položaju sa zatvorenim očima. Takvi modeli imaju visoke točnosti, u prosjeku oko 98%. U sklopu ovog istraživanja razvijen je model za klasifikaciju spola ispitanika tijekom kompleksne aktivnosti – vožnje automobila [Pub3]. Razvijeni model na ispitanicima tijekom vožnje ima točnosti usporedive s onima iz prethodnih istraživanja.

Dobiveni model iskorišten je za ostvarivanje dodatnog poboljšanja detekcije prijelaza iz stanja budnosti u stanje pospanosti tijekom vožnje. Uključivanjem podataka o spolu vozača dobivenih razvijenim modelom poboljšana je točnost detekcije pospanosti [Pub3]. Odvajanjem skupa podataka na jedan s muškim ispitanicima i jedan sa ženskim ispitanicima postižu se

značajna poboljšanja detekcije pospanosti – 3% za muške ispitanike i 7% za ženske ispitanike. Takav pronalazak nije iznenađujuć s obzirom na značajne razlike u EEG značajkama muškaraca i žena.

U budućem radu u području detekcije pospanosti vozača temeljenog na EEG-u, ključno je omogućiti pouzdanije usporedbe različitih modela. Dva glavna aspekta mogu pridonijeti ovom napretku: (1) razvijanje jedinstvene, standardne definicije pospanosti koju prihvaća istraživačka zajednica, smanjujući subjektivnu pristranost i olakšavajući usporedbu među različitim studijama, i (2) stvaranje javno dostupne baze podataka s više od 100 ispitanika, što omogućuje pravednu usporedbu različitih modela.

Također, dodatna poboljšanja razvijene metode za dobivanje višekanalnih omjera u okviru istraživanja [Pub2] mogla bi dati još bolje rezultate. Trenutna metoda stvara višekanalne omjere na temelju omjera zbrojeva između značajki frekvencijske domene iz različitih regija mozga. Uključivanje ostalih matematičkih operacija u konstrukciju višekanalnih omjera bio bi obećavajući korak naprijed u istraživanju.

Ključni pojmovi: detekcija pospanosti, eeg značajke, značajke frekvencijske domene, višekanalne frekvencijske značajke

Contents

CHAPTER 1	INTRODUCTION	1
1.1	BACKGROUND AND MOTIVATION	1
1.2	SCOPE OF THE THESIS	2
1.3	SCIENTIFIC CONTRIBUTIONS.....	3
1.4	ORGANIZATION OF THE THESIS.....	4
CHAPTER 2	AN OVERVIEW OF DRIVER DROWSINESS DETECTION METHODS	5
2.1	SUBJECTIVE MEASURES OF DROWSINESS	7
2.2	MATHEMATICAL MODELS OF SLEEP-WAKE DYNAMICS	8
2.3	VEHICLE-BASED MEASUREMENTS OF DROWSINESS	8
2.4	DRIVER'S BEHAVIOR-BASED MEASUREMENT OF DROWSINESS	9
2.5	HUMAN PHYSIOLOGICAL SIGNAL-BASED MEASUREMENTS OF DROWSINESS	10
2.5.1	<i>Electrocardiogram</i>	11
2.5.2	<i>Electrooculogram</i>	12
2.5.3	<i>Other physiological measures</i>	12
2.6	HYBRID SYSTEMS	14
CHAPTER 3	THEORETICAL BACKGROUND OF METHODS USED IN PUBLICATIONS.....	15
3.1	EEG SIGNAL PROCESSING	15
3.1.1	<i>Signal preprocessing</i>	16
3.1.2	<i>Frequency domain analysis</i>	16
3.2	STATISTICAL METHODS	18
3.2.1	<i>Correlation</i>	18
3.2.2	<i>Hypothesis testing</i>	18
3.2.3	<i>Bonferroni's correction</i>	20
3.3	MACHINE LEARNING METHODS.....	21
3.3.1	<i>Naive Bayes</i>	21
3.3.2	<i>K-nearest neighbors</i>	22
3.3.3	<i>Logistic regression</i>	23
3.3.4	<i>Support vector machines</i>	24
3.3.5	<i>Decision tree</i>	24
3.3.6	<i>Random forest</i>	25
3.3.7	<i>Extreme gradient boosting</i>	26
3.4	FEATURE SELECTION	27
3.5	EVALUATION METRICS.....	28
3.6	EVOLUTIONARY MULTICRITERIA OPTIMIZATION	29
3.6.1	<i>Non-dominated sorting genetic algorithm</i>	30
CHAPTER 4	THE MAIN SCIENTIFIC CONTRIBUTION OF THE THESIS	31
4.1	MULTICHANNEL ELECTROENCEPHALOGRAM FREQUENCY DOMAIN FEATURES FOR BRAIN ACTIVITY STATE TRANSITION DETECTION, OBTAINED USING COMBINATION RATIOS OF POWER SPECTRAL DENSITIES IN INDIVIDUAL FREQUENCY BAND.....	31
4.2	COMPUTATIONAL METHOD FOR MULTICHANNEL ELECTROENCEPHALOGRAM FREQUENCY DOMAIN FEATURE EXTRACTION BASED ON MULTICRITERIA OPTIMIZATION	32

4.3	BRAIN ACTIVITY STATE TRANSITION DETECTION MODEL BASED ON MULTICHANNEL ELECTROENCEPHALOGRAPH FEATURES AND SUBJECT'S SEX INFORMATION	32
CHAPTER 5	CONCLUSIONS AND DIRECTIONS FOR FUTURE WORK	34
5.1	THE MAIN CONCLUSIONS OF THE THESIS	34
5.2	FURTHER RESEARCH DIRECTIONS	35
CHAPTER 6	LIST OF PUBLICATIONS	36
CHAPTER 7	AUTHOR'S CONTRIBUTION TO THE PUBLICATIONS.....	37
REFERENCES		39
PUBLICATIONS		52
PUBLICATION 1		53
PUBLICATION 2		84
PUBLICATION 3		107
PUBLICATION 4		121
LIST OF ABBREVIATIONS.....		128
BIOGRAPHY		130
ŽIVOTOPIS.....		133

Chapter 1

Introduction

Brain state transition occurs many times during a day. It can be caused by many diverging factors like sensory stimuli, stress, hormonal fluctuations, illness, medications, weather, etc. Brain state transition can be very simple and fast or complex and slow. Opening one's eyes is an example of a simple and fast action that causes a simple brain state transition [1], while changing from the preclinical stage of dementia to the late stage of dementia can be slow and complex in terms of brain changes [2]. Another example of a brain state transition is changes in sleep stages [3], which is closely related to the changes between wake and drowsy state [4]. Early and reliable detection of these and many other brain state transitions can lead to improved quality of life and/or saving lives. For this reason, methods for the detection of brain state transition recently emerged as an important research question and is the topic of this thesis [5]–[7]. The rest of this chapter is structured as follows: first, the background and the motivation for the thesis are further elaborated, followed by the detailed definition of the thesis scope, then the main scientific contributions of the thesis are introduced and the description of the thesis structure is provided.

1.1 Background and motivation

Early detection of certain changes in the brain states can have a positive impact on the observed subject and his surroundings. Monitoring the brain states of patients during surgery makes the anesthesiologist's job easier and reduces the risk of patient awareness during surgery [8]. Epileptic seizures can last from seconds to minutes and can cause impaired consciousness and convulsions which may lead to drowning, driving accidents and other risks for an observed subject and its surrounding. Early detection of the transition from a normal state to a seizure in epileptic patients can improve their quality of life and reduce the risk of severe injuries and

fatalities [9]. High exposure to stress can have a negative long-term effect on well-being, as well as a negative short-term effect on functioning and performance. All workers are prone to risks caused by stress, but occupations like healthcare workers or air-traffic control officers can be exposed to extreme levels of stressful situations at the workplace. At the same time, a decrease in their performance can be life-threatening, so monitoring their stress levels and changes in their emotional states can reduce the risk of life-threatening situations occurrence [10]. Early detection of the transition from a waking state to a drowsy state during driving reduces the risk of car accidents that can lead to severe injuries and fatalities [11].

There is various medical equipment that can measure brain activity or structure – electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), magnetic resonance imaging (MRI), computed tomography (CT), and others. The measuring device for early detection of brain state transitions must be able to measure in real time. The most commonly used devices capable of real-time measurement of brain activity are EEG, MEG and fMRI. Out of these, the EEG device is the most accessible and affordable one. Also, many research groups work on small and wearable EEG devices [12]. These facts make EEG the most probable device for application in real-world scenarios.

1.2 Scope of the thesis

The central point of this thesis is research on improvements in detection of brain state transitions. As mentioned above, each individual has many brain state transitions during each day that are caused by different sources. To ensure that the research is not too extensive but narrow and specific, it was constrained to one specific category of brain state transitions – the detection of a transition from a waking state to a drowsy state during driving.

Some researchers define the term drowsiness as a state between the wake state and sleep [13]–[15]. Tiredness and sleepiness are sometimes used as synonyms for drowsiness [16]–[18]. According to Gonçalves et al. [19], 17% of drivers encountered microsleep during driving, with 7% of them experiencing accidents due to drowsiness. The high frequency and prevalence of drowsiness-related accidents highlight the need for early detection systems. Researchers are working towards finding solutions for early detection of drowsiness in drivers [20]. While vehicle-based systems are most common in commercial use, they are mostly unreliable and depend heavily on the driver's motivation [20], [21]. Therefore, physiological signals are being seen as a promising alternative for reliable drowsiness detection [22].

According to a recent review on the EEG-based driver drowsiness detection systems, 61% of included papers used frequency-domain features [23]. There are also other groups of features that are less used than frequency-domain features, like entropies [24] or nonlinear features [25]. The prevalence of frequency-domain feature usage can be attributed to the fact that the alpha wave was the first EEG feature ever used, back in 1924 [26], but also to their interpretability. Many common brain states are described with the usage of frequency-domain features, e.g., recognition of the different sleep stages is primarily based on the frequency-domain features [27]. Usually, these features are based on one specific EEG electrode, i.e., from a single brain region. Many researchers concluded that different features have different trends in different brain regions during drowsiness-related tasks [28], [29]. These findings suggest that multichannel features could bring new insights into the field of drowsiness detection. Based on the aforementioned studies, the focus of this thesis was mostly on the frequency-domain features.

The brain activity of male and female subjects differs significantly [30]. The hypothesis is that male and female subjects have different brain states and brain states transitions. If the hypothesis is correct, it also suggests that it should be possible to classify subjects' sex based only on the EEG signal features, and the information about sex should also lead to the improvement of the brain state transitions. Research related to this hypothesis is also part of this thesis.

1.3 Scientific contributions

The main scientific contributions of this thesis are the following:

- Multichannel electroencephalogram frequency domain features for brain activity state transition detection, obtained using combination ratios of power spectral densities in individual frequency bands
- Computational method for multichannel electroencephalogram frequency domain feature extraction based on multicriteria optimization
- Brain activity state transition detection model based on multichannel electroencephalogram features and subject's sex information

The first contribution involves novel multichannel EEG frequency-domain features that improve driver drowsiness detection when compared with the single-channel features. These features are obtained using combination ratios of power spectral densities in individual frequency bands. The second contribution is the developed computational method for

multichannel electroencephalogram frequency domain feature extraction that applies to any brain state transition problem and its effectiveness is demonstrated in the driver drowsiness detection problem. As part of the third contribution, a predictive model that differentiates male and female subjects based on EEG signal features was developed, and the separation of a dataset based on the model's predictions leads to improvements in driver drowsiness detection.

1.4 Organization of the thesis

The thesis is structured as follows. Chapter 2 gives an overview of different approaches to measurement and detection of driver drowsiness. Chapter 3 provides theoretical background of methods used in publications. Chapter 4 presents the main scientific contributions of the thesis. Chapter 5 concludes the thesis and provides guidelines for future research in the field. All publications that are included in the thesis are listed in Chapter 6. Chapter 7 summarizes the author's contributions to the publications. That is followed by the bibliography and then publications which contain detailed descriptions of the thesis' scientific contributions.

Chapter 2

An overview of driver drowsiness detection methods

There is no unique and broadly accepted definition of drowsiness, and different researchers have different views of drowsiness and its definition. Per [14], drowsiness is “the transient state between wakefulness and sleep, during which a reduction in vigilance (alertness, attentiveness) and performance is often observed.” Such a definition is useful for a better understanding of drowsiness by humans, but it is too vague to be translated into a computer program for the automatic detection of drowsiness. In the definition, drowsiness is put in correlation with sleep. Sleep itself is relatively well defined, including the stage of wakefulness, three stages of non-rapid eye movement (NREM) and one stage of rapid eye movement (REM) [31]. NREM and REM stages may repeat several times during sleep. The majority of authors agree that drowsiness is equivalent to the first NREM stage [21], [32]–[35], but some authors dispute that claim [36]. In the context of this thesis, drowsiness is considered to be equivalent to the first stage of NREM sleep.

Terms like sleepiness, fatigue and tiredness have different meanings than drowsiness but are related. For example, mental fatigue is “a kind of subtle feeling that creates unwillingness toward performing any activity” [37]. Mental fatigue's underlying cause can be generated by tasks with high workloads and prolonged work hours or sleep deprivation [38]. Based on the Oxford learner's dictionaries [39], tiredness is a synonym for fatigue, and its definition is “the feeling that you would like to sleep or rest”, while the definition of sleepiness is “the feeling of needing sleep or being ready to go to sleep.” In the related literature, the terms sleepiness, fatigue and tiredness are often used as synonyms for drowsiness [16], [18], [21], [40], so the same applies to this thesis.

Driver drowsiness detection methods are separated into categories – subjective measures of drowsiness, a mathematical model of the sleep-wake cycle, vehicle-based measures of

drowsiness, driver's behavior-based measures of drowsiness, and physiological measures of drowsiness. Physiological measures of drowsiness are further separated based on the type of the signal used to acquire the measures – EEG, electrocardiogram (ECG), electrooculogram (EOG), electromyogram (EMG), respiratory inductance plethysmography (RIP), galvanic skin response (GSR), and skin temperature (ST). Survey of EEG-based drowsiness detection is excluded from this chapter because [Pub1] brings an extensive overview of the EEG-based driver drowsiness detection systems together with an extensive overview of EEG features.

In order to provide a more complete view of the topic, the rest of the chapter overviews other driver drowsiness detection methods. The overview of non-EEG-based methods of drowsiness detection is not as extensive as the one published in [Pub1] since these methods are not closely related to the field of this thesis. Yet, this overview is important for a better understanding the driver drowsiness detection and possibilities for future improvements. Since most authors conduct their research on datasets that are not publicly available, comparing the accuracies of different approaches is unreliable, which is why no such comparison is made in this chapter. Instead, this chapter emphasizes describing the most commonly used techniques in each category of driver drowsiness detection. Before proceeding to the specifics of each category, Table 2.1 brings the advantages and limitations of each category of drowsiness detection systems based on available literature [20], [41], [42].

Table 2.1 Advantages and limitations for each category of driver drowsiness detection systems. Seven important criteria are evaluated – A) ease of use, B) objective, C) reliable, D) is not dependent on the subject's motivation (the subject can not consciously and intentionally influence the measurement), E) non-intrusive, F) continuous monitoring, and G) early detection. Sign „+“ indicates an advantage for a given criterion, while sign „-“ indicates a limitation.

Category	A	B	C	D	E	F	G
Subjective measures	+	-	-	-	+	-	-
Mathematical model	+	+	-	-	+	+	-
Vehicle-based measures	+	+	-	-	+	+	-
Behavioral measures	+	+	-	-	+	+	-
Physiological measures	-	+	+	+	-	+	+

2.1 Subjective measures of drowsiness

Karolinska sleepiness scale (KSS) is the most commonly used subjective measure of drowsiness. KSS has nine levels with descriptions. The first four levels correspond to the different stages of alertness, the fifth level is neither alert nor sleepy, and the last four levels correspond to different drowsiness stages [43]. The descriptions for all levels of KSS are provided in Table 2.2. Other well-known subjective scales are the Stanford sleepiness scale (SSS) [44], with eight descriptive levels of drowsiness, and the Wierwille and Ellsworth drowsiness scale [45], with three levels of drowsiness – not drowsy, slightly drowsy and moderately drowsy.

Table 2.2 Karolinska sleepiness scale.

Scale	Verbal description
1	Extremely alert
2	Very alert
3	Alert
4	Fairly alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to keep alert
8	Sleepy, some effort to keep alert
9	Very sleepy, great effort to keep alert

The measurement procedure is simple, the driver subjectively rates their drowsiness level periodically, usually every five minutes [46], [47]. There are several disadvantages of the subjective measures of drowsiness – the periodicity of assessment makes it impossible to detect sudden changes in drowsiness level, subjective rates can be misjudged, and frequent self-assessment can influence driver's attention [48]. Because of these problems, subjective measurement of drowsiness is only used in the laboratory environment as an additional source of information about a driver's drowsiness level, which can be used to calibrate other drowsiness detection systems.

2.2 Mathematical models of sleep-wake dynamics

Mathematical models offer a quantitative analysis of the effect of sleep cycles on individual performance by incorporating inputs, such as circadian cycles, duration of sleep, duration of wakefulness, and sleep history, to predict the risk of fatigue and performance quality. The well-known model is the two process model [49], based on the interaction of the circadian and homeostatic processes. An upgrade to that model is the three process model [50], which considers the duration of sleep and wakefulness as inputs. There are also several more recent models, like the aircrew fatigue evaluation model (SAFE) [51], or the sleep, activity, fatigue, and task effectiveness model (SAFTE) [52]. SAFE presents the alertness levels of the flight crew and it is based on laboratory experiments. SAFTE has applications in the military and industrial fields, and it is similar to the SAFE model but is based on both real-life and simulated data.

Although mathematical models have been employed to predict drowsiness levels and cognitive performance in generic settings such as workplaces and factories, they have limitations. One key drawback is that these models cannot predict transient changes in drowsiness. Instead, they provide only one-time evaluations. Furthermore, recent advances in intelligent systems have shifted researchers' focus from mathematical models to more sophisticated approaches [53].

2.3 Vehicle-based measurements of drowsiness

Drowsiness reduces the driver's ability to perform at a high level. An increase in the number of steering wheel adjustments, changes in the load distribution on the driver's seat, gear changes, speed corrections, and line crossings are all indicators of deterioration in performance [54]. Although vehicle-based measurement of drowsiness is entirely non-intrusive, it is hard to create a general model, since it is highly dependent on many factors, like weather, road quality, driver experience, vehicle type, and traffic conditions [55]. Furthermore, alterations in driver's behavior are not always caused by drowsiness; they can be influenced by factors like the driver's personality, experience, and motivation [56]. Despite all the mentioned limitations, these systems are the most common ones in commercial use. Based on the analysis of the largest car companies, four out of five companies use the vehicle-based measurements of drowsiness in their cars [21]. A review of the commercially available drowsiness detection devices also reveals prevalence in the number of vehicle-based systems [57].

In their review paper on predicting driver drowsiness using vehicle measures, Liu et al. [54] conclude that the most frequent discovery was that drowsiness reduced driver performance by increasing the frequency of lane departures. Moreover, drowsiness increased the variability of the steering wheel movements and the vehicle's lane position, indicating that these factors may be helpful in the prediction of the beginning and progression of driver drowsiness. Figure 2.1 shows the difference between drowsy and awake drivers on the steering wheel movement frequency-amplitude graph. According to [58], most drowsiness-related crashes are run-off-road crashes; therefore, many vehicle-based systems for drowsiness detection try to predict lane departures caused by drowsiness. In a more recent review, Lenne and Jacobs [59] conclude that subjective ratings using the KSS, blink duration, and steering behavior are the most promising features for detecting drowsiness-related lane crossing.

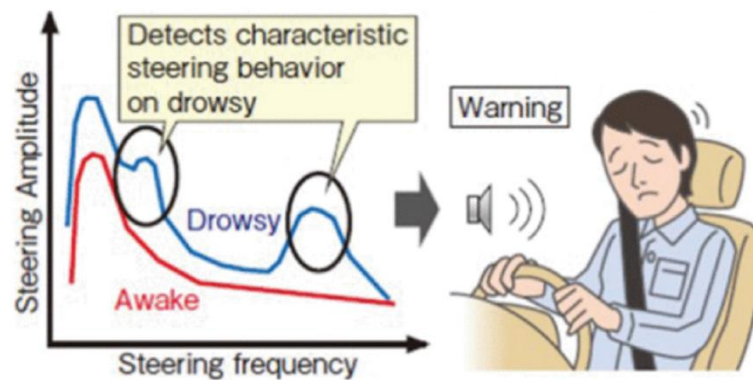


Figure 2.1 Steering wheel movement-based detection of drowsiness. Source: [60] © 2019 IEEE

2.4 Driver's behavior-based measurement of drowsiness

Measurement of driver's behavioral features is non-intrusive. It is done with an in-cockpit camera pointed at the driver. Despite recent advances in computer vision, there are still many challenges associated with extracting behavioral features from images. The obtained images vary significantly due to non-uniform illumination, out-of-plane orientations, glasses, movement, and facial deformations [61], [62]. These irregularities, in most cases, lead to inaccuracies in the measured behavioral features.

There are many developed algorithms for face detection, like eigenfaces [63], local binary patterns [64], or the Viola-Jones algorithm [65]. Current state-of-the-art algorithms are based on deep learning techniques, i.e., DeepFace [66] or FaceNet [67]. Face detection is just the first step in the process of extracting features. Besides face detection, there are many other

algorithms used for feature extraction from facial images, e.g., face tracking [68], face landmarks localization [69], blink detection [70], eye closures [71], mouth dropping and yawning recognition [71], and head pose and movement [70].

Driver's behavior-based drowsiness detection can be divided into three categories based on the types of features used: eye-related features, facial expression features, and head movement features [57]. The most commonly used behavioral feature for drowsiness detection is the percentage of eyelid closure (PERCLOS), which is part of an eye-related group of features [72]. Other features from this group that are commonly used are slow eye movement, rate of eye blinking, and average eye closure speed. Changes in the values of these features may indicate drowsiness, i.e., a decrease in blink rate or an increase in PERCLOS are indicators of drowsiness [73]. Facial expression features, like the presence of yawning, also indicate drowsiness [74]. Jaw drop, lip stretch, and raising of the inner and outer eyebrows are facial expression features that can indicate drowsiness [75]. Regarding the head position features, head nodding and scaling down usually occur when drowsiness is in a severe stage [76].

2.5 Human physiological signal-based measurements of drowsiness

The foundation of physiological approaches is the extraction of features from driver's physical state. Measurement methods like EEG, ECG, EOG, EMG, RIP, GSR and ST represent this kind of approach. The most significant advantage of the physiological measures of drowsiness is their capability for early detection and reliability. At the same time, the important limitation is the intrusiveness of the system. There have been significant improvements in the field of wearable devices for measuring physiological signals in recent years [77]. Although these small and wearable devices reduce the effect of the aforementioned limitation, it will never be completely removed as long as there is the need for electrodes to be placed on the driver. There are several non-contact devices for measuring physiological signals, like non-contact photoplethysmography (PPG) for real-time heart rate detection [78]. Despite significant recent advancements in the area of non-contact PPG, it is still challenging to obtain a trustworthy heart rate variability signal [79].

Figure 2.2 shows a typical block diagram for a physiological signal-based driver drowsiness detection system. Since the electrodes are used to acquire physiological signals, the skin-electrode contact can cause motion artifacts as electrodes have negligible internal resistance [80]. Besides the motion-related artifacts, the nearby powerline also adds noise to the measured

physiological signal [81]. The noise and artifacts should be removed before further processing of the acquired signal. After the physiological signals are cleared of noise, typically the feature extraction phase begins. At the end of the process, the extracted features are used together with machine learning algorithms to detect drowsiness.



Figure 2.2 Block diagram representation of a typical physiological signal-based drivers' drowsiness detection system using EEG sensors. Source: [21] © 2018 IEEE

More details about filtering the physiological signals are presented in subchapter 3.1. The feature extraction step is specific for each type of physiological signal. In the following subchapters, the most common features for each type of physiological signals are presented.

2.5.1 Electrocardiogram

ECG signal records the heart's electrical activity through repeated cardiac cycles. The three main components of an ECG are the P wave, which denotes depolarization of the atria, the QRS complex, which denotes depolarization of the ventricles; and the T wave, which denotes repolarization of the ventricles [82]. This process represents one heartbeat. The physiological phenomenon of variability in the time between heartbeats is known as heart rate variability (HRV). The variation in the normal beat-to-beat interval is used to measure HRV [83]. To extract the HRV signal from the ECG, the detection of the QRS complex is the first step. There are few well-established algorithms for the detection of QRS complex, like Pan-Tompkins [84] or Elgendi's [85].

Various features extracted from the HRV signal are used to detect driver drowsiness. Heart rate (HR) is the number of heartbeats per minute, and a decrease in HR indicates drowsiness [86]. The power spectral density of the HRV signal provides significant features in the analysis of drowsiness. The HRV signal is divided into three distinct frequency bands: very low frequency (VLF), which typically falls between 0.008 Hz and 0.04 Hz; low frequency (LF), which ranges from 0.04 Hz to 0.15 Hz; and high frequency (HF) that spans from 0.15 Hz to 0.5 Hz [87]. The LF to HF bands power ratio (LF/HF) is a common feature in the analysis of HRV signals. A decrease in the LF/HF feature indicates the process of falling asleep and losing alertness [87]. Also, a decrease in the VLF indicates drowsiness, according to the same study.

2.5.2 Electrooculogram

Electrooculography is a technique employed to evaluate the corneoretinal standing potential between the anterior and posterior sections of the human eye, producing a signal called the EOG [88]. Unlike electroretinogram, EOG does not assess responses to individual visual stimuli. To monitor eye movement, pairs of electrodes are positioned either above and below the eye or on its left and right sides. As the eye moves toward an electrode, it "detects" the positive aspect of the retina. In contrast, the opposing electrode "detects" the negative aspect, leading to a potential difference between the two. The placement of EOG electrodes is crucial for collecting accurate data, as the strength of the received EOG signal decreases when electrodes are farther from the eyes [89]. Eye activity, including eye blinks and movements, can change this potential difference, leading to variations in the EOG signal [90]. A blink occurs when the upper eyelid comes into contact with the lower eyelid, which typically lasts between 200 and 400 milliseconds [91].

EOG features are reliable indicators of drowsiness. Usually, indicators of drowsiness based on eyelid movements include blink rate, blink amplitude, blink duration, latency of lid reopening, and PERCLOS. Increases in blink duration and frequency were found to be drowsiness indicators [16], [92]. Also, longer lid reopening indicates sleep propensity [93]. As mentioned in subsection 2.4, PERCLOS is an important indicator of drowsiness in driver's behavior-based systems. The same applies to the PERCLOS extracted from the EOG signal [94]. Slow eye movements (SEM) have been identified as reliable indicators of drowsiness. Experiments have shown that SEM is common in drowsy drivers and is present in a significant portion of accident cases involving drowsy driving, with the "thalamic gating" phenomenon causes sleep onset while eyes remain open [89].

2.5.3 Other physiological measures

An EMG measures the electrical signals generated by muscle contractions [95], with surface EMG (sEMG) electrodes placed on the skin for non-invasive collection. Research has established a connection between muscle fatigue and EMG amplitude, as the amplitude represents muscle strength, which decreases with fatigue [96]. A decrease in the simple features like skewness and kurtosis of the EMG signal, as well as a decrease in the signal's frequency, were shown to indicate drowsiness [97], [98], [99].

Functional Near-Infrared Spectroscopy (fNIRS) is a non-invasive neuroimaging technique that measures brain activity by using near-infrared light to monitor changes in blood

oxygenation and hemodynamics in the cerebral cortex. fNIRS is based on the principle of near-infrared spectroscopy (NIRS), but it specifically focuses on the functional aspects of brain activity, i.e., it provides information about the neural activity in specific brain regions [100]. The oxygenated hemoglobin (HbO) level is a characteristic feature extracted from fNIRS. A notable increase in HbO levels during the transition from wakefulness to drowsiness can be used as an effective predictor of drowsiness [101].

GSR, also referred to as Electro Dermal Response (EDR), Psycho Galvanic Reflex (PGR), and Skin Conductance Response (SCR), measures skin conductance, which varies due to sweat gland secretion. The autonomic nervous system's sympathetic arousal controls sweat gland secretion. During drowsiness, the parasympathetic nervous system becomes more active, reducing sweating, increasing skin resistivity, and decreasing skin conductivity, or the opposite when arousal rises. However, a primary challenge with GSR is its high sensitivity to many internal and external factors like temperature, humidity, age, sex, time of the day, season, and emotions, which can affect the accuracy of the measurements [102].

RIP offers a non-invasive alternative for assessing breathing by tracking chest and abdominal wall movements, thus eliminating the need for cumbersome and invasive devices such as masks or mouthpieces [103]. This method utilizes recording belts placed around the thorax and abdomen, making it well-suited for continuous lung volume measurements. Respiration rate variability is a feature extracted from the RIP signal which can be used as an indicator of drowsiness [104].

ST measurement techniques assess the skin's surface temperature, which typically ranges between 32°C-35°C for healthy individuals [105]. The thermoregulation system in humans primarily maintains body temperature within a specific range, with ST being a direct outcome of this process. In contrast to ST, core body temperature reflects the internal operating temperature of body organs. Studies have demonstrated a significant decrease in forehead temperature values when transitioning from an alert state to increased drowsiness, with the forehead temperature continuing to decline as drowsiness intensifies [106].

There are many other techniques for measuring physiological signals, such as fMRI, MRI, NIRS, CT, or MEG. However, to the best of my knowledge when writing this thesis, these signals have not been applied in the domain of drowsiness detection in drivers but in some relatively related fields. For instance, fMRI has been employed for the study of intoxicated drivers [107], MEG for the detection of driver distraction [108], and MRI for the analysis of drivers with mental illness [109].

2.6 Hybrid systems

Hybrid drowsiness detection systems combine multiple methods, such as vehicle-, behavioral-, and physiological-based measures, to improve accuracy and reliability. These systems take into account the shortcomings of different approaches, leading to more robust drowsiness detection. The main idea behind the use of hybrid drowsiness detection systems is the improvement of drowsiness detection quality. However, sometimes the use of hybrid system adds additional limitations that initially were not there. For example, if EEG-based drowsiness detection is added to an existing vehicle-based system, the accuracy and reliability of the system would increase, but at the same time, the system would become more intrusive and harder to use. Research on drowsiness detection systems based on subjective, behavioral and physiological-based measures has resulted in significantly better detection compared to individual systems [110]. The same conclusion applies to the system based on behavioral and vehicle measures [111]. There are many other research studies with a hybrid approach in controlled environments [101], [112], [113], and it would be interesting to test the ability of these systems in a real-world environment [42].

Chapter 3

Theoretical background of methods used in publications

This chapter will present the theoretical background of methods used in publications [Pub2, Pub3, and Pub4]. These methods were mentioned in the papers but not elaborated because the main focus was on the scientific contributions of each paper. Here, a detailed explanation of the methods will be provided, making it easier for readers to understand the research. The article [Pub1] is a review article; therefore, no methods were used that need further explanation. The chapter is divided into four parts: EEG signal processing, statistical methods, machine learning methods, and evolutionary multicriteria optimization. By the end of this chapter, readers should have a good understanding of the techniques used in the papers, which will contribute to a better comprehension of the published works and their scientific contributions.

3.1 EEG signal processing

In most cases, EEG signals need to be preprocessed after they are acquired and before they are ready for analysis. Preprocessing involves transforming or reorganizing data before analyzing it. Different preprocessing steps can either reorganize data, remove bad data, or modify otherwise clean data. Preprocessing is needed because EEG signals contain both useful information (signal) and unwanted information (noise). Separating the two during the preprocessing stage can be challenging, and it is hard to eliminate all the noise without losing some of the useful signal. When the signal is preprocessed, usually feature extraction techniques are applied. For extraction of frequency-domain features, the signal first needs to be transformed from time domain into frequency domain. In continuation of this subchapter, first, filtering methods used in [Pub2], [Pub3] and [Pub4] are introduced, and second, methods for frequency domain analysis are explained.

3.1.1 Signal preprocessing

Filtering

The most important part of preprocessing is filtering. Filtering data can reduce the impact of high-frequency artifacts and low-frequency drifts, and notch filters weaken electrical line noise. There are two types of filters – finite impulse response (FIR) and infinite impulse response (IIR). These concepts refer to how a filter reacts to an impulse signal. FIR filters have a finite response, meaning their response ends at a certain point, while IIR filters have an infinite response, meaning their response never ends. FIR filters are known for their stability and tendency to produce fewer nonlinear phase distortions. While FIR filters may have a higher computational cost due to their increased filter order, this is typically not a significant drawback with modern computation capability [114].

Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is a method for separating multivariate signals into additive, statistically independent components. In the context of EEG data, the measured signal at each electrode is a mixture of different signal sources:

$$x = a_1s_1 + \dots + a_ns_n = \mathbf{A}\mathbf{s} \quad (3.1)$$

where s represents the original signal (e.g., brain activity, eye movement artifacts, etc.), and a_k are the weights. The goal is to estimate the matrix A and vector s by making as general assumptions as possible. The first assumption is that the original signals s are statistically independent, and the second assumption is that the original signals must have a distribution different from Gaussian. After estimating matrix A , its inverse W can be calculated. With the matrix W , the original signals can be easily computed using the equation 3.2 [115].

$$\mathbf{S} = \mathbf{W}\mathbf{x} \quad (3.2)$$

In terms of EEG preprocessing, ICA is used to extract different components from signal and remove unwanted components. In most cases ICA is used to remove artifacts like eye movements or presence of heart beats in the EEG signal.

3.1.2 Frequency domain analysis

Fast Fourier transform (FFT)

The Fast Fourier Transform (FFT) is an efficient algorithm used to compute a sequence's Discrete Fourier Transform (DFT). The DFT is a mathematical technique that converts a time-domain signal into its frequency-domain representation, allowing analysis of the frequencies present in the signal. The DFT of a time series is defined with equation 3.3.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j \frac{2\pi kn}{N}} \quad (3.3)$$

where x_n is the input signal, X_k is the output, N is the sequence length, and j is the imaginary unit. The FFT algorithm significantly reduces the computation time for the DFT by exploiting the symmetries in the calculation. The most common FFT algorithm is the Cooley-Tukey algorithm, which recursively breaks down the DFT calculation into smaller DFTs. It uses a divide-and-conquer approach that separates the input sequence into even and odd-indexed components. This process is repeated until the subproblems are small enough to be solved directly.

The FFT reduces the complexity of the DFT calculation from $O(N^2)$ in the case of the direct DFT computation to $O(N * \log(N))$ using the Cooley-Tukey algorithm [116], making it much faster for large sequences. It can be said that an FFT is any method to compute the same results as DFT in $O(N * \log(N))$ operations.

The resulting frequency-domain representation X_k can be squared to obtain power spectral density (PSD). This approach of squaring the FFT output to obtain PSD is often referred to as a periodogram. Once PSD is obtained, it is used for spectral analysis of the signal.

Welch's method

Welch's method [117] is a technique used for estimating the PSD of a signal. It is an extension of the periodogram method, as it reduces noise and variance in the PSD in exchange for reducing the frequency resolution. The method involves dividing the input signal into overlapping segments, applying a window function to each segment, and then calculating the periodogram for each windowed segment. The periodograms are computed using the FFT. The final PSD estimate is obtained by averaging the periodograms across all segments.

Thomson multitaper method

The Thomson multitaper method [118] is an advanced technique for estimating the PSD of a signal. It improves upon traditional methods, such as the Periodogram and Welch's method, by reducing variance while maintaining frequency resolution. The multitaper method uses a set of orthogonal tapering functions known as Slepian sequences. Each taper is multiplied with the signal; then a periodogram is applied to these signals resulting in multiple tapered PSDs. By averaging these tapered PSDs, the final PSD is obtained.

3.2 Statistical methods

3.2.1 Correlation

The concept of correlation was utilized in [Pub3] and [Pub4]. In [Pub3], a new approach was taken where correlations between two features of EEG signal were used as novel features. These novel features proved to be good predictors of the driver's sex. In [Pub4], the correlation strength between male and female drivers was compared, and the study concluded that male drivers had stronger correlations between certain EEG features compared to female drivers.

The calculation of the correlation between two signals results in the correlation coefficient. The correlation coefficient is a well-known basic measure of the relationship between variables and a starting point for various more complicated statistics. There are many types of correlation coefficients; the most common one is the Pearson correlation coefficient, which will be explained in the continuation of this subchapter [119].

In analyzing the relationship between two variables, researchers focus on two fundamental characteristics of correlation coefficients: direction and strength. The direction of a correlation coefficient can be either positive or negative. The values of the two variables move in the same direction when there is a positive correlation, whereas the opposite is true when there is a negative correlation. It is important to remember that a correlation does not necessarily indicate a causation relationship and that the found connection may not apply to every individual in the sample or population. The strength or magnitude of the relationship is the second essential characteristic of correlation coefficients. Ranging from -1.0 to +1.0, the coefficient quantifies the strength of the relationship between the variables. A coefficient of 0 indicates no relationship, while coefficients closer to -1.0 or +1.0 signify stronger relationships.

Equation 3.4 shows how the Pearson coefficient is calculated,

$$r = \frac{\sum(z_x z_y)}{N} \quad (3.4),$$

where r represents the Pearson correlation coefficient, z_x is the z score of variable X , z_y is the z score of variable Y and N is the number of pairs of X and Y scores. z score represents a number obtained after the standardization of each sample, i.e., after subtracting the population's mean and dividing by the population's standard deviation.

3.2.2 Hypothesis testing

Hypothesis testing is a method employed to answer questions about a population based on sample data, as access to the entire population is typically unavailable. A degree of confidence,

such as 95% or 99%, is defined, but reaching 100% confidence in the results is unfeasible. Both parametric and non-parametric hypothesis tests can be used, with the former assuming a certain probability distribution for the sample data. The null hypothesis assumes no difference in the population data, while the alternative hypothesis assumes a real difference in the population data. Hypothesis testing generally involves examining data, formulating hypotheses, calculating a test statistic, and comparing it to a critical value to decide if the null hypothesis can be rejected. There are three prevalent use cases of hypothesis tests depending on the type of data: one-sample, two-sample paired data, and two-sample unpaired data. The one-sample data type involves a single sample with multiple measurements of the observed variable. Two-sample paired data refers to cases where two measurements of the observed variable are collected on the same population before and after an experiment of interest. The two-sample unpaired data type arises when multiple measurements of the observed variable are gathered from two distinct sample groups [120].

Hypothesis testing was used in [Pub2] and [Pub3]. In [Pub2], two samples were paired, i.e., the same subjects were measured in the wake and drowsy states, so the used test was non-parametric Wilcoxon signed-rank test. In [Pub3], two samples were unpaired, i.e., two independent samples – male and female drivers, were measured during driving, so the used test was a non-parametric Mann-Whitney U test. In continuation, these two tests are presented in more detail.

Wilcoxon signed-rank test

Wilcoxon signed-rank test was developed in 1945 by Frank Wilcoxon. The test is carried out with the following steps for the two-sample paired data:

- Determining the null hypothesis – The null hypothesis states that the population's median consisting of paired data differences equals zero, while the alternative hypothesis suggests otherwise.
- Compute the test statistics – compute the differences between the paired data samples, rank each sample based on the absolute value of difference, sum up the ranks of positive and negative differences and use a smaller value as the test statistic.
- Compare the test statistic to a critical value – rejecting the null hypothesis occurs when the test statistic is smaller than the critical value.

Critical values depend on the three pieces of information: sample size, significance level, and single-tailed or two-tailed test. Critical values are precomputed and organized in a table for all the most commonly used sample sizes and significance levels. The only assumption of the Wilcoxon signed-rank test is that the distribution of differences between the two samples is

symmetric. This assumption allows the conversion of the non-parametric test into a parametric one. Suppose the assumption is correct and the null hypothesis is true (the median of the population of differences is zero). In that case, there is the same number of different values larger and smaller than zero (because of the definition of median). The central limit theorem allows us to use the normal distribution to make inferences about the population based on the distribution of the rank sums. This brings us to the parametric test with parameters of normal distribution calculated based on the properties of the arithmetic series (since ranks are arithmetic series) [121].

Mann-Whitney U test

The Mann-Whitney U test is similar to the Wilcoxon signed-rank test as it uses ranks. Since the data is unpaired, it can not rank the sample differences but ranks all values as the same distribution. The test is carried out with the following steps [121]:

- Determine the null hypothesis – both populations have the same distribution, while the alternative hypothesis suggests that the two distributions share identical properties except for differing medians.
- Compute the test statistics – pool both samples of sizes n_c and n_t into a single sample of size $n_c + n_t$, sort values and then rank them, calculate sums of ranks for control sample R_c and test sample R_t , calculate U statistic U_c (equation 3.5) and U_t (equation 3.6) and use the lower value as test statistics.

$$U_c = n_t n_c + 0.5 n_c (n_c + 1) - R_c \quad (3.5),$$

$$U_t = n_t n_c + 0.5 n_t (n_t + 1) - R_t \quad (3.6),$$

- Compare the test statistics to a critical value – rejecting the null hypothesis occurs when the test statistic is smaller than the critical value.

As with the Wilcoxon signed-rank test, common critical values are precomputed and organized in a table. The idea behind the Mann-Whitney U test is also to transform the non-parametric problem into the parametric one. This time it is done using the U statistic, which is normally distributed and a parametric test can be applied to it [122].

3.2.3 Bonferroni's correction

Multiple hypothesis tests were conducted in [Pub2] and [Pub4]. However, conducting numerous hypothesis tests can increase the likelihood of Type I errors, which occur when the null hypothesis is rejected even though there is no actual difference in the population variable. To address this issue, Bonferroni's correction was applied to adjust each test's critical value.

Bonferroni's correction involves dividing the significance level α by the number of tests m being performed. This adjusted critical value helps maintain the risk of getting at least one Type I error despite the multiple tests conducted. Bonferroni's correction usage ensures that the analysis remains robust and avoids an inflated probability of Type I errors while exploring the relationships between various factors and the outcome of interest.

3.3 Machine learning methods

Machine learning, an integral subfield of data science, enables computers to learn from data and make informed decisions. Three primary learning paradigms are supervised, unsupervised, and reinforcement learning. Supervised learning involves labeled data and focuses on classification or regression tasks, while unsupervised learning seeks to discover hidden patterns in unlabeled data through clustering or dimensionality reduction techniques. Reinforcement learning, on the other hand, revolves around training agents to interact with an environment and optimize their actions based on rewards or penalties. This thesis focuses on supervised classification, which aims to accurately predict discrete class labels for new, unseen instances based on historically labeled data. Seven different classifiers have been utilized for brain state transition detection in publications [Pub2] and [Pub3]. These seven classifiers are naive Bayes, K-nearest neighbors, logistic regression, support vector machines, decision tree, random forest and XGBoost. In the following subsections, the theoretical background of each classifier is briefly presented to provide a better understanding of their underlying principles and functionality.

3.3.1 Naive Bayes

Naive Bayes refers to a group of probabilistic algorithms based on Bayes' theorem (equation 3.7). In this subsection, the term naive Bayes refers to the Gaussian naive Bayes algorithm that assumes the conditional likelihood of the features to be Gaussian. The “naive” part comes from the assumption that the features are mutually independent given the class variable. Although this assumption is often violated in practice, the naive Bayes algorithm is still known to work very well for some problems, like spam filtering. Bayes' theorem is defined with the following equation 3.7:

$$P(y|\mathbf{x}) = \frac{P(y)P(\mathbf{x}|y)}{P(\mathbf{x})} \quad (3.7),$$

where:

- y denotes class variable and \mathbf{x} denotes feature vector
- $P(y|\mathbf{x})$ is the posterior probability – interpretation: given the observed feature vector \mathbf{x} , what is the chance it belongs to class y ?
- $P(y)$ is the prior probability (prior knowledge) – interpretation: the probability that any new outcome is equal to y .
- $P(\mathbf{x}|y)$ is the conditional probability (likelihood) – interpretation: How likely is it to observe feature vector \mathbf{x} given that it belongs to class y ?
- $P(\mathbf{x})$ is called evidence – interpretation: the probability of encountering feature vector \mathbf{x} .

Naive independence of features is defined with equation 3.8.

$$P(\mathbf{x}_i|y, \mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_n) = P(\mathbf{x}_i|y) \quad (3.8)$$

With equation 3.8 applied for each i , conditional probability in equation 3.7 can be redefined, so the equation for posterior probability would now be

$$P(y|\mathbf{x}) = \frac{P(y) \prod_{i=1}^n P(\mathbf{x}_i|y)}{P(\mathbf{x})} \quad (3.9),$$

which lead to the classification rule

$$\mathbf{y} = \underset{y}{\operatorname{argmax}} P(y) \prod_{i=1}^n P(\mathbf{x}_i|y) \quad (3.10).$$

The evidence $P(\mathbf{x})$ is removed from the classification rule as it is only a scaling factor. The conditional probability (likelihood) of the Gaussian naive Bayes algorithm is defined with equation 3.11,

$$P(\mathbf{x}_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(\mathbf{x}_i - \mu_y)^2}{2\sigma_y^2}\right) \quad (3.11),$$

where σ_y and μ_y are estimated using maximum likelihood. [123]

3.3.2 K-nearest neighbors

K-nearest neighbors (KNN) algorithm, in the process of decision-making about instance x , uses k nearest instances to x and makes a decision based on the majority voting among these instances [124]. KNN training is done by storing all the training instances in the memory. The optimal value for parameter k is data-dependent. Generally, higher k suppresses noise and makes classification boundaries more stable. In some implementations, *weights* can be used so that closer instances contribute more to the final decision. Since the KNN algorithm relies on

the distances between instances, the method is less effective in cases of highly dimensional data due to the “curse of dimensionality” [125].

3.3.3 Logistic regression

Logistic regression is a statistical method used for binary classification. It estimates the probability of an observation belonging to a particular class. The relationship between the input features and the predicted probability is modeled using the sigmoid function, also known as the logistic function, defined with equation 3.12 and shown in Figure 3.1.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (3.12)$$

Here, z represents the linear combination of input features X and their corresponding weights w , as expressed with equation 3.13.

$$z = \mathbf{w}^T \mathbf{x} = w_0 x_0 + w_1 x_1 + \dots + w_m x_m \quad (3.13)$$

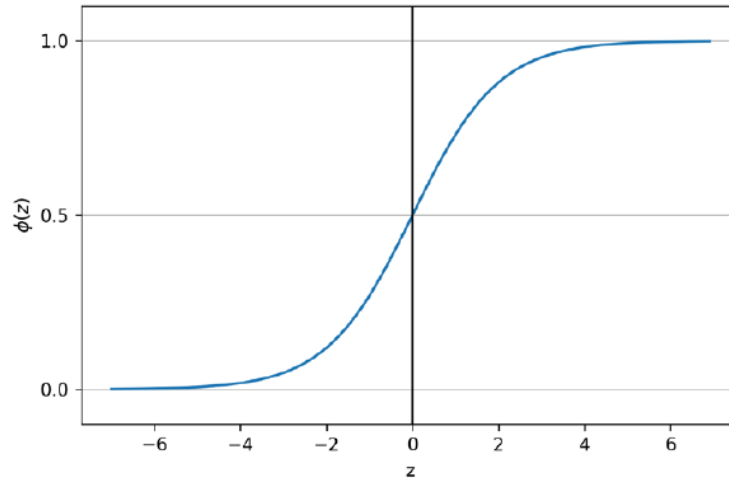


Figure 3.1 Sigmoid function.

During the training process, the algorithm adjusts the weights to minimize the differences between the predicted probabilities and the corresponding actual class labels. To find the best weights, the algorithm iteratively updates them based on the gradient of the log-likelihood function, as defined with equation 3.14.

$$\mathbf{w}_j = \mathbf{w}_j + \eta \sum_{i=1}^n \left(y^{(i)} - \sigma(z^{(i)}) \right) x_j^{(i)} \quad (3.14)$$

Here, w represents the weights, η is the learning rate, and the sum represents the gradient of the log-likelihood function. In the log-likelihood function, the cost tends to zero when a correct prediction is made. On the other hand, when the prediction is incorrect, the cost tends toward infinity. The central idea is that erroneous predictions incur a progressively higher cost, effectively penalizing such inaccuracies. [126]

3.3.4 Support vector machines

Support vector machines (SVM) are powerful supervised learning algorithms used for both classification and regression tasks. In the context of classification, SVMs aim to find the optimal hyperplane that separates the classes in the feature space with the largest margin possible. The margin represents the distance between the hyperplane and the closest data points from each class, known as support vectors. Maximizing the margin helps improve the generalization capabilities of the model, leading to better performance on unseen data.

In its simplest form, the linear SVM seeks to maximize the margin while ensuring that all samples are correctly classified. The margin is the distance between the positive and negative hyperplane and is defined with equation 3.15.

$$M = \frac{2}{\|w\|} \quad (3.15)$$

The margin M needs to be maximized with the constraint that all the examples are classified correctly. This problem can be solved with quadratic programming.

For the extension of SVM to non-linear classification problems, kernels play a crucial role. By employing kernel functions, it is possible to transform the input data into a higher-dimensional space, making the classes linearly separable in this new space. This technique is known as the "kernel trick." The kernel trick involves computing the inner product between two data points in the higher-dimensional space without explicitly transforming the data points. A kernel function κ is defined with equation 3.16.

$$\kappa(x^{(i)}, x^{(j)}) = \phi(x^{(i)})^T \phi(x^{(j)}) \quad (3.16)$$

Here, $x^{(i)}$ and $x^{(j)}$ are two input samples, and ϕ is the transformation function that maps the input samples to the higher-dimensional space. Popular kernel functions used in kernel SVMs include the linear, polynomial, Gaussian radial basis function, and sigmoid kernels. Each kernel function has a hyperparameter set that controls its behavior and decision boundary. [127]

3.3.5 Decision tree

Decision trees (DT) are non-parametric models whose main advantage is their interpretability. The primary goal is to recursively split the input data into subsets based on features, resulting in a tree-like structure with decision nodes and leaf nodes. Decision nodes represent splits based on feature values, while leaf nodes correspond to the predicted class labels. The process of constructing the decision tree involves selecting the most informative feature for each split. This is achieved by maximizing the information gain IG , which quantifies

the reduction in impurity achieved by splitting the dataset on a specific feature. Information gain is defined with equation 3.17.

$$IG(D_p, f) = I(D_p) - \frac{N_{left}}{N_p} I(D_{left}) - \frac{N_{right}}{N_p} I(D_{right}) \quad (3.17)$$

Here, f represents the feature used for splitting, while D_p denotes the dataset of the parent, and D_{left} and D_{right} denote the dataset of the left and right child nodes. I is the impurity measure, N_p refers to the total number of training examples at the parent node, and N_{left} and N_{right} correspond to the number of examples in the left and right child nodes. As evident, the information gain is essentially the difference between the parent node's impurity and the sum of the child nodes' impurities. The greater the reduction in child node impurities, the higher the information gain.

The most used impurity measures are entropy, Gini impurity and classification error. Entropy is defined with equation 3.18.

$$I_H(t) = - \sum_{i=1}^c p(i|t) \log_2 p(i|t) \quad (3.18)$$

In this case, $p(i|t)$ represents the proportion of examples belonging to class i for a specific node, t . Entropy is zero if all examples at a node are from the same class, and it reaches its maximum value when there is a uniform class distribution. Consequently, it can be stated that the entropy criterion seeks to maximize the mutual information within the tree. On the other hand, Gini impurity criterion can be interpreted as a criterion to minimize the possibility of misclassification. Gini impurity is defined with equation 3.19.

$$I_G(t) = - \sum_{i=1}^c p(i|t)(1 - p(i|t)) \quad (3.19)$$

Like entropy, the Gini impurity reaches its maximum value when the classes are perfectly mixed. In practice, however, Gini impurity and entropy generally produce similar results. Classification error can also be used as impurity measure, but it is more often used as a criterion for tree pruning. Classification error is defined with equation 3.20.

$$I_E(t) = 1 - \max\{p(i|t)\} \quad (3.20)$$

3.3.6 Random forest

The random forest (RF) [128] is an ensemble learning algorithm used for classification and regression tasks. It uses multiple decision trees during the training phase and combines their

predictions to produce a more accurate and robust result. By aggregating the outputs of multiple trees, RF reduces the chance of overfitting and improves generalization.

Here's an overview of how the random forest classifier is trained:

1. Select a random subset of the training data with replacement (bootstrap sample) for each tree in the forest.
2. For each tree, randomly select a subset of features from the complete feature set. Usually, \sqrt{M} is used as the size of the random subset, where M is a total number of features.
3. Find the best split based on an impurity measure while using only the selected features.
4. Continue growing the tree until a stopping criterion is met, such as a maximum depth or minimum number of samples per leaf.
5. Repeat steps 1-4 for each tree in the forest.
6. Once all trees are trained, predictions are made by aggregating the individual tree predictions. For classification, a majority vote is taken.

By using bootstrapped samples and random feature selection, the RF introduces diversity among the trees, which reduces overfitting and improves generalization. This ensemble approach results in a more accurate and robust model compared to a single decision tree.

3.3.7 Extreme gradient boosting

Extreme gradient boosting (XGBoost) is an open source implementation of the gradient boosting trees algorithm [129]. It is a powerful machine learning algorithm used for classification and regression tasks. It is an ensemble learning method based on gradient boosting, which builds multiple weak learners (typically decision trees) sequentially, improving the model by minimizing an objective function that measures the prediction error.

Here's an overview of how the XGBoost method is trained:

1. Initialize the model with a constant prediction value or a simple base learner, like a shallow decision tree.
2. Calculate the residuals, which are the differences between the actual values and the current ensemble predictions.
3. Fit a new decision tree to the residuals, learning to correct the mistakes made by the current ensemble.
4. Update the model by adding the new decision tree, scaled by a learning rate, to the current ensemble.

5. Repeat steps 2-4 for a specified number of boosting rounds or until a stopping criterion, such as early stopping based on validation performance, is met.
6. Once all rounds are complete, predictions are made by aggregating the outputs of all decision trees in the ensemble.

XGBoost introduces several improvements to the standard gradient boosting algorithm, such as regularization terms to control model complexity and prevent overfitting, parallelization and distributed computing support for faster training. These features make XGBoost a highly efficient and accurate algorithm for various machine learning tasks.

3.4 Feature selection

Feature selection is an important step in the machine learning pipeline, where the most relevant features or variables are selected from the original dataset to build a more efficient and accurate model. The primary goal of feature selection is to reduce the dimensionality of the dataset, thus minimizing noise, computational complexity, and the risk of overfitting. Common feature selection approaches include filter, wrapper and embedded methods [130].

Filter methods are a category of feature selection techniques that evaluate the importance of individual features based on their statistical properties, independently of the model. Some popular filter methods include the chi-squared test, mutual information, and ANOVA F-test. In the paper [Pub3], these three methods were used.

The chi-squared test measures the relationship between categorical features and categorical target variables. The chi-squared statistic is defined with equation 3.21.

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad (3.21)$$

Here, O_i is the observed frequency, E_i is the expected frequency, and the summation is over all samples. If two features are independent, the observed frequency is expected to be similar to the expected frequency, resulting in a lower chi-squared value. Therefore, a high chi-squared value suggests that the hypothesis of independence is not valid. Features with high chi-squared values are considered more relevant for the target variable.

The mutual information method quantifies the dependency between features and the target variable by measuring the amount of information one can obtain about the target variable by observing a specific feature. It is a similar concept to information gain that was defined in equation 3.17. Mutual information between feature X and the target variable Y is defined with equation 3.22.

$$MI(X; Y) = H(X) - H(X|Y) \quad (3.22)$$

Here, $MI(X; Y)$ is the mutual information for X and Y , $H(X)$ is the entropy for X , and $H(X|Y)$ is the conditional entropy for X given Y . Features with higher mutual information values indicate a stronger dependency between the feature and the target variable.

ANOVA F-test measures the relationship between each continuous predictive feature and a categorical target variable. It assesses whether the means of two or more groups are significantly different. The F-test statistic is defined with equation 3.23.

$$F = \frac{MS_{between}}{MS_{within}} \quad (3.23)$$

Here, $MS_{between}$ is the mean sum of squares between groups (i.e., the variance between predictive feature's and target variable's means), and MS_{within} is the mean sum of squares within groups (i.e., the variance within each group – variance of predictive feature and variance of target variable). Features with higher F-test values indicate a stronger association between the feature and the target variable.

3.5 Evaluation metrics

In the context of machine learning, proper evaluation of models is crucial for several reasons. First, it ensures the model's performance is accurately assessed, helping to avoid overfitting or underfitting by measuring generalization capabilities on unseen data. Second, evaluating models allows for effective comparison between different algorithms or configurations, enabling informed decisions in selecting the most suitable approach for a given task. In papers [Pub2] and [Pub3], two evaluation metrics were used – accuracy and precision. In continuation, they will be described, together with several other basic concepts important for better understanding.

An important concept for evaluation is the confusion matrix, shown in Table 3.1 for binary classification. TN denotes true negative, FP denotes false positive, FN denotes false negative and TP denotes true positive. They are all represented with natural numbers. Based on these metrics, several other useful metrics can be calculated, as shown with the following equations.

$$accuracy = \frac{TN + TP}{TN + FP + FN + TP} \quad (3.24)$$

$$precision = \frac{TP}{TP + FP} \quad (3.25)$$

$$recall = \frac{TP}{TP + FN} \quad (3.26)$$

Accuracy (equation 3.24) is the number of correctly predicted samples divided by the total number of samples. Precision (equation 3.25) is the number of true positive predictions divided by the total number of class 1 predictions. Recall (equation 3.26) is the number of true positive predictions divided by the number of true class 1 samples.

Table 3.1 Confusion matrix.

		Predicted	
		Class 0	Class 1
True	Class 0	TN	FP
	Class 1	FN	TP

3.6 Evolutionary multicriteria optimization

Computational optimization is a field that focuses on finding the best possible solutions to complex problems by minimizing or maximizing objective functions, often subject to various constraints. Within this field, evolutionary optimization is a subdomain that employs bio-inspired algorithms to mimic natural processes such as evolution, selection, and mutation to explore the solution space efficiently. Evolutionary multicriteria optimization (EMO) is a specialized area within evolutionary optimization which addresses problems with multiple conflicting objectives. When using EMO for non-trivial problems, no single solution minimizes all criteria simultaneously. Instead, there are Pareto optimal solutions – solutions that cannot be improved in any criterion without worsening another. Such solutions are also called non-dominated. All Pareto optimal solutions are equally good in a mathematical sense, and the final decision on choosing one solution over the others is usually made externally, based on expert knowledge.

Several widely-used algorithms have been developed for evolutionary multicriteria optimization, including the non-dominated sorting genetic algorithm II (NSGA-II), the strength Pareto evolutionary algorithm 2 (SPEA2), and the multi-objective particle swarm optimization (MOPSO). These algorithms employ different mechanisms for handling multiple objectives and maintaining diversity in the population of solutions, ensuring efficient exploration and convergence towards the Pareto front. The next subchapter brings a description of the NSGA-II algorithm as it was used in [Pub2].

3.6.1 Non-dominated sorting genetic algorithm

The genetic algorithm is the base for the NSGA-II algorithm [131]. Genetic algorithms use local search to find the best solution by exploring the neighborhood of the current solution. Crossover and mutation are used to modify the current solution and create new ones. The best solutions are selected for reproduction in each generation until a stopping criterion is met.

NSGA-II performs two steps before the operations characteristic for genetic algorithms. First, NSGA-II sorts the population of candidate solutions into multiple non-dominated Pareto fronts. A solution is said to dominate another if it is better or equal in all objectives and strictly better in at least one objective. The first front represents the best set of non-dominated solutions found so far. Second, the algorithm assigns a crowding distance to each solution within a front, which is a measure of the solution's proximity to its neighbors in the objective space. Higher crowding distances indicate sparser regions, and the algorithm prefers solutions with larger crowding distances to maintain diversity.

Finally, NSGA-II uses genetic operators, such as crossover and mutation, to create a new population of solutions for the next iteration. The algorithm employs a selection process based on both non-domination rank and crowding distance to choose solutions for reproduction. This ensures that the new population includes diverse and non-dominated solutions that gradually converge toward the Pareto front.

Chapter 4

The main scientific contribution of the thesis

The main scientific contributions of this thesis are as follows: first, multichannel electroencephalogram frequency domain features for brain activity state transition detection, obtained using combination ratios of power spectral densities in individual frequency bands, and second, computational method for multichannel electroencephalogram frequency domain feature extraction based on multicriteria optimization; these contributions are presented in [Pub2]. The third scientific contribution, brain activity state transition detection model based on multichannel electroencephalogram features and subject's sex informatio is presented in [Pub3] and [Pub4]. [Pub1] is a review of EEG signal features and their application in driver drowsiness detection systems which indirectly contributed to all three scientific contributions of the thesis.

4.1 Multichannel electroencephalogram frequency domain features for brain activity state transition detection, obtained using combination ratios of power spectral densities in individual frequency band

The first contribution is the result of the method described in the second contribution when applied to the drowsiness detection problem. The dataset consisted of 28 subjects, out of which 12 subjects were used exclusively as the test set to confirm the generalization ability of the obtained ratio index features. Two novel ratio index features are presented, one complex with 13 addends and their coefficients in the equation, and one simple with only 5 addends and no coefficients [Pub2]. Both proposed multichannel ratio indices performed better in terms of

statistical significance, accuracy, precision, and execution time than the seven most used single-channel ratio indices.

4.2 Computational method for multichannel electroencephalogram frequency domain feature extraction based on multicriteria optimization

The second contribution of the thesis is a multicriteria optimization method for the extraction of novel multichannel frequency-domain features. More specifically, the method extracts ratio index features that are based on individual frequency band features from different EEG channels. The method is constructed to preserve only the most informative frequency band features from each channel (i.e., brain region) that are combined together into novel ratio index features. It is applicable to any brain state transition problem and the only constraint for the method to work properly is that the dataset used for the extraction of novel features has the transition near the middle of the signal. The method is designed with the assumption that the transition is near the middle of the signal. For the application of the method on the datasets where that assumption is not satisfied, additional preprocessing and realigning of the signal is needed. The method tries to create a feature that will look like a step function. This is done with optimization based on the evolutionary metaheuristic algorithm called elitist non-dominated sorting genetic algorithm II (NSGA-II) [131] with two objective functions., The first one is to maximize the absolute difference between the left and right parts of the signal, and the second one is to minimize oscillations on each side of the signal. Since the transition of step function can not be instantaneous, a transition window of flexible size and position is used near the middle of the signal. The signal within the transition window is ignored by the objective functions. The used transition window enables the method to adjust to inaccuracies in the original transition labels, as shown in [Pub2].

4.3 Brain activity state transition detection model based on multichannel electroencephalogram features and subject's sex information

The third contribution is brain activity state transition detection model based on multichannel electroencephalogram features and subject's sex information [Pub3] and [Pub4]. Within the [Pub3] paper, two sub-contributions are presented. The first one is the use of EEG

feature correlations as new features in a model for sex classification, and the second one is the improvements of drowsiness detection with information on driver's sex. The model for sex classification is based on the EEG frequency-domain features, recurrence quantification analysis features and correlations between them used as features. [Pub3] is extended research based on the [Pub4]. In [Pub4], the smaller dataset was used and only a statistical analysis of the difference between male and female drivers in terms of extracted EEG features was presented.

Chapter 5

Conclusions and directions for future work

5.1 The main conclusions of the thesis

Driver drowsiness poses a significant risk to road safety, as it contributes to a substantial number of accidents and fatalities worldwide. The deceptive nature of drowsiness often results in impaired judgment, delayed reaction times, and reduced situational awareness, which collectively compromise a driver's ability to respond effectively to dynamic road conditions. It is important to address driver drowsiness as its mitigation saves lives and reduces economic consequences related to car accidents.

The review paper [Pub1] offers four key contributions: (1) a thorough examination and systematization of EEG signal features, (2) an extensive review of drowsiness detection systems, (3) a comparative analysis of similar reviews, and (4) a discussion on potential improvements to the state-of-the-art drowsiness detection systems. Besides these contributions, a few suggestions and guidelines are given. First, larger sample sizes, around 100 participants or more, are needed to ensure diverse datasets and model robustness while minimizing inter-individual differences. Second is the need for validation of the EEG-based drowsiness detection systems using data from unseen subjects. Third, the suggestion is to make datasets publicly available for fair comparisons.

While working on the review, a gap in the body of knowledge related to the prediction of driver drowsiness was found. Many studies show that frequency-domain features from different brain regions behave differently when it comes to drowsiness. However, current methods only focus on a single brain region for the extraction of these features. This insight led to a novel brain state transition detection method, described in [Pub2]. The method was used to extract multichannel frequency-domain features that help identify drowsiness. The proposed method

is reliable and the frequency-domain features based on the signals from multiple brain regions have better performance on multiple metrics compared to the same type of features that are based on the single brain region. An additional benefit of the method is that it can be used for any brain state transition problem.

Within the [Pub3] and [Pub4], the hypothesis that male and female subjects have different brain states and brain states transitions were investigated on the subjects that were experiencing drowsiness during driving. According to this part of the research, incorporating the information on driver's sex improves drowsiness detection accuracy. Moreover, a trustworthy driver's sex classifier based on the recorded EEG data was constructed.

5.2 Further research directions

In future work within the EEG-based driver drowsiness detection field, enabling a more reliable comparison of different models is of great importance. Two main aspects could contribute to this advancement. First, the development of a unified, standard definition of drowsiness should be developed and accepted by the research community, which would reduce subjective bias and facilitate easier comparison among various studies. Second, as mentioned in the previous subsection, creating a publicly accessible database with more than 100 subjects would significantly impact the field. This would allow for benchmarking different models and comparing their performance.

Additional improvements in the developed method within [Pub2] could yield even better results. By default, the method creates the ratio of sums between frequency-domain features from different brain regions. Allowing operations other than just adding in the construction of the ratio-based indices would be a promising research step forward.

Chapter 6

List of publications

- Pub 1 I. Stancin, M. Cifrek, and A. Jovic, “A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems,” *Sensors*, vol. 21, no. 11, p. 3786, May 2021, doi: 10.3390/s21113786.
- Pub 2 I. Stancin, N. Frid, M. Cifrek, and A. Jovic, “EEG Signal Multichannel Frequency-Domain Ratio Indices for Drowsiness Detection Based on Multicriteria Optimization,” *Sensors*, vol. 21, no. 20, p. 6932, Oct. 2021, doi: 10.3390/s21206932.
- Pub 3 I. Stancin, M. Z. Zeba, K. Friganovic, M. Cifrek, and A. Jovic, “Information on Drivers’ Sex Improves EEG-Based Drowsiness Detection Model,” *Applied Sciences*, vol. 12, no. 16, p. 8146, Aug. 2022, doi: 10.3390/app12168146.
- Pub 4 I. Stancin, K. Friganovic, M. Z. Zeba, M. Cifrek, and A. Jovic, “Gender differences in EEG features while driving,” *Proceedings of the 2nd International Conference on Advances in Signal Processing and Artificial Intelligence*, p. 127, Nov. 2020

Chapter 7

Author's contribution to the publications

This doctoral thesis presents findings derived from research conducted between 2018 and 2022 at the University of Zagreb Faculty of Electrical Engineering and Computing, located at Unska 3, HR-10000, Zagreb, Croatia. The research was part of the project supported in part by the European Regional Development Fund under Grant KK.01.2.1.01.0136 (DFDM). The thesis features four publications co-authored with several collaborators. The author contributed to each paper in several different categories. Categories of contribution evaluated for each contribution are conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, writing – original draft preparation, writing – review and editing, visualizations, supervision, project administration and funding acquisition. In continuation, when stating contributions for each paper, authors are presented with their initials, so the author of this thesis is presented with the initials I.S. After presenting contribution evaluation of all papers, short summary of the author's contribution to the publications is given.

[Pub1] The paper “**A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems**” have contributions among the authors as follows: conceptualization, I.S. and A.J.; methodology, I.S. and A.J.; validation, I.S., M.C. and A.J.; formal analysis, I.S. and A.J.; investigation, I.S.; resources, I.S.; data curation, I.S.; writing—original draft preparation, I.S.; writing—review and editing, M.C. and A.J.; visualization, I.S.; supervision, A.J.; project administration, M.C. and A.J.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript, including this categorization of contributions.

[Pub2] The paper “**EEG Signal Multichannel Frequency-Domain Ratio Indices for Drowsiness Detection Based on Multicriteria Optimization**” have contributions among the authors as follows: conceptualization, I.S., N.F. and A.J.; methodology, I.S., N.F. and A.J.;

software, I.S. and N.F.; validation, I.S., N.F., M.C. and A.J.; formal analysis, I.S., N.F. and A.J.; investigation, I.S.; resources, I.S. and N.F.; data curation, I.S.; writing—original draft preparation, I.S. and N.F.; writing—review and editing, I.S., N.F., M.C. and A.J.; visualization, I.S.; supervision, A.J.; project administration, M.C.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript, including this categorization of contributions.

[Pub3] The paper “**Information on Drivers’ Sex Improves EEG-Based Drowsiness Detection Model**” have contributions among the authors as follows: conceptualization, I.S., M.Z.Z., K.F., M.C. and A.J.; methodology, I.S., M.Z.Z., K.F. and A.J.; software, I.S.; validation, I.S., M.Z.Z. and A.J.; formal analysis, M.C. and A.J.; investigation, I.S., M.Z.Z. and K.F.; resources, M.C. and A.J.; data curation, I.S., M.Z.Z. and K.F.; writing—original draft preparation, I.S. and M.Z.Z.; writing—review and editing, I.S., M.Z.Z., K.F., M.C. and A.J.; visualization, I.S.; supervision, M.C. and A.J.; project administration, M.C. and A.J.; funding acquisition, M.C. and A.J. All authors have read and agreed to the published version of the manuscript, including this categorization of contributions.

[Pub4] The paper “**Gender differences in EEG features while driving**” have contributions among the authors as follows: conceptualization, I.S., M.Z.Z., K.F., M.C. and A.J.; methodology, I.S., M.Z.Z., K.F. and A.J.; software, I.S.; validation, I.S., M.Z.Z. and A.J.; formal analysis, M.C. and A.J.; investigation, I.S., M.Z.Z. and K.F.; resources, M.C. and A.J.; data curation, I.S., M.Z.Z. and K.F.; writing—original draft preparation, I.S. and M.Z.Z.; writing—review and editing, I.S., M.Z.Z., K.F., M.C. and A.J.; visualization, I.S.; supervision, M.C. and A.J.; project administration, M.C. and A.J.; funding acquisition, M.C. and A.J. All authors have read and agreed to the published version of the manuscript, including this categorization of contributions.

To summarize, the author of this thesis is the first author of all four publications, with the most contributions among the observed categories. The author was the main contributor in the most important categories – conceptualization and methodology. The author is the only author of the developed software in papers [Pub3] and [Pub4], while in [Pub2], the author is responsible for more than half of the developed software. The vast majority of writing the papers is also the author's merit. Finally, the core of the ideas behind the scientific contributions of this thesis and the developed methodologies for their realization is mostly the author's merit.

References

- [1] J. Maheshwari, S. D. Joshi, and T. K. Gandhi, “Tracking the Transitions of Brain States: An Analytical Approach Using EEG,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 8, pp. 1742–1749, Aug. 2020, doi: 10.1109/TNSRE.2020.3005950.
- [2] L. Pugnetti *et al.*, “EEG Evidence of Posterior Cortical Disconnection in PD and Related Dementias,” *Int. J. Neurosci.*, vol. 120, no. 2, pp. 88–98, Feb. 2010, doi: 10.3109/00207450903436346.
- [3] N. Giannakeas, “EEG-Based Automatic Sleep Stage Classification,” *Biomed. J. Sci. Tech. Res.*, vol. 7, no. 4, Aug. 2018, doi: 10.26717/BJSTR.2018.07.001535.
- [4] T. L. T. da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, “Automated drowsiness detection through wavelet packet analysis of a single EEG channel,” *Expert Syst. Appl.*, vol. 55, pp. 559–565, Aug. 2016, doi: 10.1016/j.eswa.2016.02.041.
- [5] Z. Huang, H. Cai, T. Dan, Y. Lin, P. Laurienti, and G. Wu, “Detecting Brain State Changes by Geometric Deep Learning of Functional Dynamics on Riemannian Manifold,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Cham, 2021, pp. 543–552. doi: 10.1007/978-3-030-87234-2_51.
- [6] J. Lu *et al.*, “fNIRS-based brain state transition features to signify functional degeneration after Parkinson’s disease,” *J. Neural Eng.*, vol. 19, no. 4, p. 046038, Aug. 2022, doi: 10.1088/1741-2552/ac861e.
- [7] Y.-B. Lee, K. Yoo, J. H. Roh, W.-J. Moon, and Y. Jeong, “Brain-State Extraction Algorithm Based on the State Transition (BEST): A Dynamic Functional Brain Network Analysis in fMRI Study,” *Brain Topogr.*, vol. 32, no. 5, pp. 897–913, Sep. 2019, doi: 10.1007/s10548-019-00719-7.
- [8] P. Myles, K. Leslie, J. McNeil, A. Forbes, and M. Chan, “Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial,” *The*

- Lancet*, vol. 363, no. 9423, pp. 1757–1763, May 2004, doi: 10.1016/S0140-6736(04)16300-9.
- [9] D. Sopic, A. Aminifar, and D. Atienza, “e-Glass: A Wearable System for Real-Time Detection of Epileptic Seizures,” in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, Florence, 2018, pp. 1–5. doi: 10.1109/ISCAS.2018.8351728.
- [10] P. Arico *et al.*, “Human Factors and Neurophysiological Metrics in Air Traffic Control: A Critical Review,” *IEEE Rev. Biomed. Eng.*, vol. 10, pp. 250–263, 2017, doi: 10.1109/RBME.2017.2694142.
- [11] S. L. Chellappa, C. J. Morris, and F. A. J. L. Scheer, “Effects of circadian misalignment on cognition in chronic shift workers,” *Sci. Rep.*, vol. 9, no. 1, p. 699, Jan. 2019, doi: 10.1038/s41598-018-36762-w.
- [12] A. J. Casson, “Wearable EEG and beyond,” *Biomed. Eng. Lett.*, vol. 9, no. 1, pp. 53–71, Feb. 2019, doi: 10.1007/s13534-018-00093-6.
- [13] M. Corsi-Cabrera, C. Arce, J. Ramos, I. Lorenzo, and M. A. Guevara, “Time course of reaction time and EEG while performing a vigilance task during total sleep deprivation,” *Sleep*, vol. 19, no. 7, pp. 563–569, Sep. 1996, doi: 10.1093/sleep/19.7.563.
- [14] M. L. Jackson *et al.*, “The utility of automated measures of ocular metrics for detecting driver drowsiness during extended wakefulness,” *Accid. Anal. Prev.*, vol. 87, pp. 127–133, Feb. 2016, doi: 10.1016/j.aap.2015.11.033.
- [15] B. S. Oken, M. C. Salinsky, and S. M. Elsas, “Vigilance, alertness, or sustained attention: physiological basis and measurement,” *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 117, no. 9, pp. 1885–1901, Sep. 2006, doi: 10.1016/j.clinph.2006.01.017.
- [16] C. Papadelis *et al.*, “Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents,” *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 118, no. 9, pp. 1906–1922, Sep. 2007, doi: 10.1016/j.clinph.2007.04.031.
- [17] S. K. Lal and A. Craig, “A critical review of the psychophysiology of driver fatigue,” *Biol. Psychol.*, vol. 55, no. 3, pp. 173–194, Feb. 2001, doi: 10.1016/s0301-0511(00)00085-5.
- [18] M. A. Kamran, M. M. N. Mannan, and M. Y. Jeong, “Drowsiness, Fatigue and Poor Sleep’s Causes and Detection: A Comprehensive Study,” *IEEE Access*, vol. 7, pp. 167172–167186, 2019, doi: 10.1109/ACCESS.2019.2951028.

- [19] M. Gonçalves *et al.*, “Sleepiness at the wheel across Europe: a survey of 19 countries,” *J. Sleep Res.*, vol. 24, no. 3, pp. 242–253, Jun. 2015, doi: 10.1111/jsr.12267.
- [20] R. P. Balandong, R. F. Ahmad, M. N. Mohamad Saad, and A. S. Malik, “A Review on EEG-Based Automatic Sleepiness Detection Systems for Driver,” *IEEE Access*, vol. 6, pp. 22908–22919, 2018, doi: 10.1109/ACCESS.2018.2811723.
- [21] A. Chowdhury, R. Shankaran, M. Kavakli, and Md. M. Haque, “Sensor Applications and Physiological Features in Drivers’ Drowsiness Detection: A Review,” *IEEE Sens. J.*, vol. 18, no. 8, pp. 3055–3067, Apr. 2018, doi: 10.1109/JSEN.2018.2807245.
- [22] T. Kundinger, N. Sofra, and A. Riener, “Assessment of the Potential of Wrist-Worn Wearable Sensors for Driver Drowsiness Detection,” *Sensors*, vol. 20, no. 4, Art. no. 4, Jan. 2020, doi: 10.3390/s20041029.
- [23] I. Stancin, M. Cifrek, and A. Jovic, “A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems,” *Sensors*, vol. 21, no. 11, p. 3786, May 2021, doi: 10.3390/s21113786.
- [24] U. R. Acharya *et al.*, “Characterization of focal EEG signals: A review,” *Future Gener. Comput. Syst.*, vol. 91, pp. 290–299, Feb. 2019, doi: 10.1016/j.future.2018.08.044.
- [25] C. J. Stam, “Nonlinear dynamical analysis of EEG and MEG: review of an emerging field,” *Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol.*, vol. 116, no. 10, pp. 2266–2301, Oct. 2005, doi: 10.1016/j.clinph.2005.06.011.
- [26] R. İnce, S. S. Adanır, and F. Sevmez, “The inventor of electroencephalography (EEG): Hans Berger (1873–1941),” *Childs Nerv. Syst.*, vol. 37, no. 9, pp. 2723–2724, Sep. 2021, doi: 10.1007/s00381-020-04564-z.
- [27] “AASM Scoring Manual - American Academy of Sleep Medicine,” *American Academy of Sleep Medicine – Association for Sleep Clinicians and Researchers*. <https://aasm.org/clinical-resources/scoring-manual/> (accessed Mar. 04, 2023).
- [28] M. Jagannath and V. Balasubramanian, “Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator,” *Appl. Ergon.*, vol. 45, no. 4, pp. 1140–1147, Jul. 2014, doi: 10.1016/j.apergo.2014.02.001.
- [29] C. Wang *et al.*, “Spectral Analysis of EEG During Microsleep Events Annotated via Driver Monitoring System to Characterize Drowsiness,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 2, pp. 1346–1356, Apr. 2020, doi: 10.1109/TAES.2019.2933960.
- [30] B. Kaur, D. Singh, and P. P. Roy, “Age and gender classification using brain–computer interface,” *Neural Comput. Appl.*, vol. 31, no. 10, pp. 5887–5900, Oct. 2019, doi: 10.1007/s00521-018-3397-1.

- [31] A. K. Patel, V. Reddy, K. R. Shumway, and J. F. Araujo, "Physiology, Sleep Stages," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2022. Accessed: Mar. 06, 2023. [Online]. Available: <http://www.ncbi.nlm.nih.gov/books/NBK526132/>
- [32] U. Budak, V. Bajaj, Y. Akbulut, O. Atila, and A. Sengur, "An Effective Hybrid Model for EEG-Based Drowsiness Detection," *IEEE Sens. J.*, vol. 19, no. 17, pp. 7624–7631, Sep. 2019, doi: 10.1109/JSEN.2019.2917850.
- [33] S. Majumder, B. Guragain, C. Wang, and N. Wilson, "On-board Drowsiness Detection using EEG: Current Status and Future Prospects," in *2019 IEEE International Conference on Electro Information Technology (EIT)*, Brookings, SD, USA, May 2019, pp. 483–490. doi: 10.1109/EIT.2019.8833866.
- [34] N. Sriraam, T. K. Padma Shri, and U. Maheshwari, "Recognition of wake-sleep stage 1 multichannel eeg patterns using spectral entropy features for drowsiness detection," *Australas. Phys. Eng. Sci. Med.*, vol. 39, no. 3, pp. 797–806, Sep. 2016, doi: 10.1007/s13246-016-0472-8.
- [35] T. Nakamura, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, "Automatic detection of drowsiness using in-ear EEG," in *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Jul. 2018, pp. 1–6. doi: 10.1109/IJCNN.2018.8489723.
- [36] M. W. Johns, "A new perspective on sleepiness," *Sleep Biol. Rhythms*, vol. 8, no. 3, pp. 170–179, Jul. 2010, doi: 10.1111/j.1479-8425.2010.00450.x.
- [37] E. Grandjean, "Fatigue in industry," *Br. J. Ind. Med.*, vol. 36, no. 3, pp. 175–186, Aug. 1979.
- [38] A. Williamson, R. Friswell, J. Olivier, and R. Grzebieta, "Are drivers aware of sleepiness and increasing crash risk while driving?," *Accid. Anal. Prev.*, vol. 70, pp. 225–234, Sep. 2014, doi: 10.1016/j.aap.2014.04.007.
- [39] "Oxford Learner's Dictionaries | Find definitions, translations, and grammar explanations at Oxford Learner's Dictionaries." <https://www.oxfordlearnersdictionaries.com/> (accessed Mar. 15, 2023).
- [40] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver Drowsiness Classification Using Fuzzy Wavelet-Packet-Based Feature-Extraction Algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–131, Jan. 2011, doi: 10.1109/TBME.2010.2077291.

- [41] M. R. Ullah, M. Aslam, M. I. Ullah, and M.-E. A. Maria, “Driver’s Drowsiness Detection Through Computer Vision: A Review,” in *Advances in Computational Intelligence*, Cham, 2018, pp. 272–281. doi: 10.1007/978-3-030-02840-4_22.
- [42] A. Sahayadhas, K. Sundaraj, and M. Murugappan, “Detecting Driver Drowsiness Based on Sensors: A Review,” *Sensors*, vol. 12, no. 12, Art. no. 12, Dec. 2012, doi: 10.3390/s121216937.
- [43] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, “Karolinska Sleepiness Scale (KSS),” in *STOP, THAT and One Hundred Other Sleep Scales*, A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, Eds. New York, NY: Springer, 2012, pp. 209–210. doi: 10.1007/978-1-4419-9893-4_47.
- [44] H. E, “The history and use of the Stanford sleepiness scale,” *Psychophysiology*, vol. 9, p. 150, 1971.
- [45] W. W. Wierwille and L. A. Ellsworth, “Evaluation of driver drowsiness by trained raters,” *Accid. Anal. Prev.*, vol. 26, no. 5, pp. 571–581, Oct. 1994, doi: 10.1016/0001-4575(94)90019-1.
- [46] A. Kosmadopoulos *et al.*, “The efficacy of objective and subjective predictors of driving performance during sleep restriction and circadian misalignment,” *Accid. Anal. Prev.*, vol. 99, no. Pt B, pp. 445–451, Feb. 2017, doi: 10.1016/j.aap.2015.10.014.
- [47] M. Poursadeghiyan, A. Mazloumi, G. N. Saraji, A. Niknezhad, A. Akbarzadeh, and M. H. Ebrahimi, “Determination the Levels of Subjective and Observer Rating of Drowsiness and Their Associations with Facial Dynamic Changes,” *Iran. J. Public Health*, vol. 46, no. 1, pp. 93–102, Jan. 2017.
- [48] T. Abe, D. Mollicone, M. Basner, and D. F. Dinges, “Sleepiness and Safety: Where Biology Needs Technology,” *Sleep Biol. Rhythms*, vol. 12, no. 2, pp. 74–84, Apr. 2014, doi: 10.1111/sbr.12067.
- [49] A. A. Borbely, “A Two Process Model of Sleep Regulation,” *Hum. Neurobiol.*, pp. 195–204, 1982.
- [50] T. Åkerstedt and S. Folkard, “Validation of the S and C Components of the Three-Process Model of Alertness Regulation,” *Sleep*, vol. 18, no. 1, pp. 1–6, Jan. 1995, doi: 10.1093/sleep/18.1.1.
- [51] A. J. Belyavin and M. B. Spencer, “Modeling Performance and Alertness: The QinetiQ Approach,” *Aviat. Space Environ. Med.*, vol. 75, no. 3, pp. A93–A103, Mar. 2004.

- [52] C. Fox, V. Hall, T. Schaefer, and R. Wolford, “407 EMF Identification of Emergency Medicine Fatigue At-Risk Periods Using Actigraphy and Computer Modeling: A Pilot Study,” *Ann. Emerg. Med.*, vol. 66, no. 4, pp. S146–S147, Oct. 2015, doi: 10.1016/j.annemergmed.2015.07.444.
- [53] G. Sikander and S. Anwar, “Driver Fatigue Detection Systems: A Review,” *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019, doi: 10.1109/TITS.2018.2868499.
- [54] C. C. Liu, S. G. Hosking, and M. G. Lenné, “Predicting driver drowsiness using vehicle measures: Recent insights and future challenges,” *J. Safety Res.*, vol. 40, no. 4, pp. 239–245, Aug. 2009, doi: 10.1016/j.jsr.2009.04.005.
- [55] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, “Uncorrelated fuzzy neighborhood preserving analysis based feature projection for driver drowsiness recognition,” *Fuzzy Sets Syst.*, vol. 221, pp. 90–111, Jun. 2013, doi: 10.1016/j.fss.2012.12.003.
- [56] P. Li, W. Jiang, and F. Su, “Single-channel EEG-based mental fatigue detection based on deep belief network,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug. 2016, pp. 367–370. doi: 10.1109/EMBC.2016.7590716.
- [57] M. Doudou, A. Bouabdallah, and V. Berge-Cherfaoui, “Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges,” *Int. J. Intell. Transp. Syst. Res.*, vol. 18, no. 2, pp. 297–319, May 2020, doi: 10.1007/s13177-019-00199-w.
- [58] A. J. Filtiness, K. A. Armstrong, A. Watson, and S. S. Smith, “Sleep-related crash characteristics: Implications for applying a fatigue definition to crash reports,” *Accid. Anal. Prev.*, vol. 99, no. Pt B, pp. 440–444, Feb. 2017, doi: 10.1016/j.aap.2015.11.024.
- [59] M. G. Lenné and E. E. Jacobs, “Predicting drowsiness-related driving events: a review of recent research methods and future opportunities,” *Theor. Issues Ergon. Sci.*, vol. 17, no. 5–6, pp. 533–553, Sep. 2016, doi: 10.1080/1463922X.2016.1155239.
- [60] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, “A Survey on State-of-the-Art Drowsiness Detection Techniques,” *IEEE Access*, vol. 7, pp. 61904–61919, 2019, doi: 10.1109/ACCESS.2019.2914373.
- [61] M. Flores, J. Armingol, and A. de la Escalera, “Driver Drowsiness Warning System Using Visual Information for Both Diurnal and Nocturnal Illumination Conditions,”

- EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, Art. no. 1, Dec. 2010, doi: 10.1155/2010/438205.
- [62] M. Tayab Khan *et al.*, “Smart Real-Time Video Surveillance Platform for Drowsiness Detection Based on Eyelid Closure,” *Wirel. Commun. Mob. Comput.*, vol. 2019, p. e2036818, Mar. 2019, doi: 10.1155/2019/2036818.
- [63] M. A. Turk and A. P. Pentland, “Face recognition using eigenfaces,” in *1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceedings*, Jun. 1991, pp. 586–591. doi: 10.1109/CVPR.1991.139758.
- [64] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006, doi: 10.1109/TPAMI.2006.244.
- [65] P. Viola and M. J. Jones, “Robust Real-Time Face Detection,” *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.
- [66] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708. Accessed: Mar. 09, 2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2014/html/Taigman_DeepFace_Closing_the_2014_CVPR_paper.html
- [67] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823. Accessed: Mar. 09, 2023. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Schroff_FaceNet_A_Unified_2015_CVPR_paper.html
- [68] W. Deng and R. Wu, “Real-Time Driver-Drowsiness Detection System Using Facial Features,” *IEEE Access*, vol. 7, pp. 118727–118738, 2019, doi: 10.1109/ACCESS.2019.2936663.
- [69] L. Zhao, Z. Wang, X. Wang, and Q. Liu, “Driver drowsiness detection using facial dynamic fusion information and a DBN,” *IET Intell. Transp. Syst.*, vol. 12, no. 2, pp. 127–133, 2018, doi: 10.1049/iet-its.2017.0183.
- [70] I.-H. Choi, C.-H. Jeong, and Y.-G. Kim, “Tracking a Driver’s Face against Extreme Head Poses and Inference of Drowsiness Using a Hidden Markov Model,” *Appl. Sci.*, vol. 6, no. 5, Art. no. 5, May 2016, doi: 10.3390/app6050137.

- [71] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, “Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network,” in *Computer Vision – ACCV 2016 Workshops*, Cham, 2017, pp. 117–133. doi: 10.1007/978-3-319-54526-4_9.
- [72] T. P. Nguyen, M. T. Chew, and S. Demidenko, “Eye tracking system to detect driver drowsiness,” in *2015 6th International Conference on Automation, Robotics and Applications (ICARA)*, Feb. 2015, pp. 472–477. doi: 10.1109/ICARA.2015.7081194.
- [73] T. Abe, “PERCLOS-based technologies for detecting drowsiness: Current evidence and future directions,” *SLEEP Adv.*, p. zpad006, Jan. 2023, doi: 10.1093/sleepadvances/zpad006.
- [74] T. Azim, M. A. Jaffar, M. Ramzan, and A. M. Mirza, “Automatic Fatigue Detection of Drivers through Yawning Analysis,” in *Signal Processing, Image Processing and Pattern Recognition*, Berlin, Heidelberg, 2009, pp. 125–132. doi: 10.1007/978-3-642-10546-3_16.
- [75] M. A. Assari and M. Rahmati, “Driver drowsiness detection using face expression recognition,” in *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Nov. 2011, pp. 337–341. doi: 10.1109/ICSIPA.2011.6144162.
- [76] I. Teyeb, O. Jemai, M. Zaied, and C. Ben Amar, “A novel approach for drowsy driver detection using head posture estimation and eyes recognition system based on wavelet network,” in *IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications*, Jul. 2014, pp. 379–384. doi: 10.1109/IISA.2014.6878809.
- [77] N. R. Adão Martins, S. Annaheim, C. M. Spengler, and R. M. Rossi, “Fatigue Monitoring Through Wearables: A State-of-the-Art Review,” *Front. Physiol.*, vol. 12, p. 790292, 2021, doi: 10.3389/fphys.2021.790292.
- [78] A. R. Kavsaoglu, K. Polat, and M. R. Bozkurt, “Real time heart rate detection using non-contact photoplethysmography signals,” in *2014 22nd Signal Processing and Communications Applications Conference (SIU)*, Apr. 2014, pp. 196–199. doi: 10.1109/SIU.2014.6830199.
- [79] V. A. A. van Es, R. G. P. Lopata, E. P. Scilingo, and M. Nardelli, “Contactless Cardiovascular Assessment by Imaging Photoplethysmography: A Comparison with Wearable Monitoring,” *Sensors*, vol. 23, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/s23031505.
- [80] D. Buxi *et al.*, “Correlation Between Electrode-Tissue Impedance and Motion Artifact in Biopotential Recordings,” *IEEE Sens. J.*, vol. 12, no. 12, pp. 3373–3383, Dec. 2012, doi: 10.1109/JSEN.2012.2221163.

- [81] A. Lay-Ekuakille, A. Scarano, and A. Trotta, “Power Line Interference Cancelling in EEG Inspection,” 2005. Accessed: Mar. 15, 2023. [Online]. Available: <https://www.semanticscholar.org/paper/Power-Line-Interference-Cancelling-in-EEG-Lay-Ekuakille-Scarano/b65b29c7fb34761691102d5bf6d04fc899dad2bb>
- [82] “30. Cardiology,” in *Kumar and Clark’s Clinical Medicine E-Book*, Elsevier Health Sciences, 2020.
- [83] “Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology,” *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996.
- [84] J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.
- [85] M. Elgendi, “Fast QRS Detection with an Optimized Knowledge-Based Method: Evaluation on 11 Standard ECG Databases,” *PLOS ONE*, vol. 8, no. 9, p. e73557, Sep. 2013, doi: 10.1371/journal.pone.0073557.
- [86] Y. Sun, X. Yu, J. Berilla, Z. Liu, and G. Wu, “An In-Vehicle Physiological Signal Monitoring System for Driver Fatigue Detection,” 2011. Accessed: Mar. 15, 2023. [Online]. Available: <https://www.semanticscholar.org/paper/An-In-Vehicle-Physiological-Signal-Monitoring-for-Sun-Yu/14663bde3553762e837a48fc9e4f5227eb141d50>
- [87] G. D. Furman, A. Baharav, C. Cahan, and S. Akselrod, “Early detection of falling asleep at the wheel: A Heart Rate Variability approach,” in *2008 Computers in Cardiology*, Sep. 2008, pp. 1109–1112. doi: 10.1109/CIC.2008.4749240.
- [88] G. B. Arden and J. H. Kelsey, “Some observations on the relationship between the standing potential of the human eye and the bleaching and regeneration of visual purple,” *J. Physiol.*, vol. 161, no. 2, pp. 205–226, May 1962.
- [89] D. Shin, H. Sakai, and Y. Uchiyama, “Slow eye movement detection can prevent sleep-related accidents effectively in a simulated driving task,” *J. Sleep Res.*, vol. 20, no. 3, pp. 416–424, Sep. 2011, doi: 10.1111/j.1365-2869.2010.00891.x.
- [90] B. Thorslund, *Electrooculogram analysis and development of a system for defining stages of drowsiness*. Statens väg- och transportforskningsinstitut, 2004. Accessed: Mar. 16, 2023. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:vti:diva-5130>

- [91] J. L. Andreassi, *Psychophysiology: Human Behavior and Physiological Response*, 5th edition. Mahwah, NJ: Psychology Press, 2006.
- [92] Hargutt and H. Kruger, "EYELID MOVEMENTS AND THEIR PREDICTIVE VALUE FOR FATIGUE STAGES," 2001. Accessed: Mar. 16, 2023. [Online]. Available: <https://www.semanticscholar.org/paper/EYELID-MOVEMENTS-AND-THEIR-PREDICTIVE-VALUE-FOR-Hargutt-Kruger/dae67f05513e4a884810c836739e12901456cdal>
- [93] R. Schleicher, N. Galley, S. Briest, and L. Galley, "Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?," *Ergonomics*, vol. 51, no. 7, pp. 982–1010, Jul. 2008, doi: 10.1080/00140130701817062.
- [94] X.-Q. Huo, W.-L. Zheng, and B.-L. Lu, "Driving fatigue detection with fusion of EEG and forehead EOG," in *2016 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2016, pp. 897–904. doi: 10.1109/IJCNN.2016.7727294.
- [95] M. Paoletti, A. Belli, L. Palma, M. Vallasciani, and P. Pierleoni, "A Wireless Body Sensor Network for Clinical Assessment of the Flexion-Relaxation Phenomenon," *Electronics*, vol. 9, no. 6, Art. no. 6, Jun. 2020, doi: 10.3390/electronics9061044.
- [96] M.-I. Lin, H.-W. Liang, K.-H. Lin, and Y.-H. Hwang, "Electromyographical assessment on muscular fatigue—an elaboration upon repetitive typing activity," *J. Electromyogr. Kinesiol.*, vol. 14, no. 6, pp. 661–669, Dec. 2004, doi: 10.1016/j.jelekin.2004.03.004.
- [97] M. K. Wali, "FFBPNN-BASED HIGH DROWSINESS CLASSIFICATION USING EMG AND WPT," *Biomed. Eng. Appl. Basis Commun.*, vol. 32, no. 03, p. 2050023, Jun. 2020, doi: 10.4015/S1016237220500234.
- [98] A. Luttmann, M. Jäger, J. Sökeland, and W. Laurig, "Electromyographical study on surgeons in urology. II. Determination of muscular fatigue," *Ergonomics*, vol. 39, no. 2, pp. 298–313, Feb. 1996, doi: 10.1080/00140139608964460.
- [99] V. Balasubramanian and K. Adalarasu, "EMG-based analysis of change in muscle activity during simulated driving," *J. Bodyw. Mov. Ther.*, vol. 11, no. 2, pp. 151–158, Apr. 2007, doi: 10.1016/j.jbmt.2006.12.005.
- [100] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *NeuroImage*, vol. 63, no. 2, pp. 921–935, Nov. 2012, doi: 10.1016/j.neuroimage.2012.03.049.

- [101] T. Nguyen, S. Ahn, H. Jang, S. C. Jun, and J. G. Kim, "Utilization of a combined EEG/NIRS system to predict driver drowsiness," *Sci. Rep.*, vol. 7, no. 1, p. 43933, Mar. 2017, doi: 10.1038/srep43933.
- [102] A. Widyanti, Hanna, K. Muslim, and I. Z. Satalaksana, "The sensitivity of Galvanic Skin Response for assessing mental workload in Indonesia," *Work*, vol. 56, no. 1, pp. 111–117, Feb. 2017, doi: 10.3233/WOR-162479.
- [103] D. J. Marlin and C. M. Deaton, "15 - Pulmonary Function Testing," in *Equine Respiratory Medicine and Surgery*, B. C. McGorum, P. M. Dixon, N. E. Robinson, and J. Schumacher, Eds. Edinburgh: W.B. Saunders, 2007, pp. 211–233. doi: 10.1016/B978-0-7020-2759-8.50020-9.
- [104] F. Guede-Fernández, M. Fernández-Chimeno, J. Ramos-Castro, and M. A. García-González, "Driver Drowsiness Detection Based on Respiratory Signal Analysis," *IEEE Access*, vol. 7, pp. 81826–81838, 2019, doi: 10.1109/ACCESS.2019.2924481.
- [105] S. Zhang, N. Zhu, and S. Lu, "Responses of human perception and skin temperature to directed thermal radiation in hot environments," *Build. Environ.*, vol. 197, p. 107857, Jun. 2021, doi: 10.1016/j.buildenv.2021.107857.
- [106] S. Bando, K. Oiwa, and A. Nozawa, "Evaluation of dynamics of forehead skin temperature under induced drowsiness," *IEEJ Trans. Electr. Electron. Eng.*, vol. 12, no. S1, pp. S104–S109, 2017, doi: 10.1002/tee.22423.
- [107] G. Battistella *et al.*, "Weed or Wheel! fMRI, Behavioural, and Toxicological Investigations of How Cannabis Smoking Affects Skills Necessary for Driving," *PLOS ONE*, vol. 8, no. 1, p. e52545, Jan. 2013, doi: 10.1371/journal.pone.0052545.
- [108] A. Fort *et al.*, "Attentional demand and processing of relevant visual information during simulated driving: A MEG study," *Brain Res.*, vol. 1363, pp. 117–127, Dec. 2010, doi: 10.1016/j.brainres.2010.09.094.
- [109] E. Papageorgiou *et al.*, "The neural correlates of impaired collision avoidance in hemianopic patients," *Acta Ophthalmol. (Copenh.)*, vol. 90, no. 3, pp. e198–e205, 2012, doi: 10.1111/j.1755-3768.2011.02315.x.
- [110] G. Yang, Y. Lin, and P. Bhattacharya, "A driver fatigue recognition model based on information fusion and dynamic Bayesian network," *Inf. Sci.*, vol. 180, no. 10, pp. 1942–1954, May 2010, doi: 10.1016/j.ins.2010.01.011.
- [111] B. Cheng, W. Zhang, Y. Lin, R. Feng, and X. Zhang, "Driver Drowsiness Detection Based on Multisource Information," *Hum. Factors Ergon. Manuf. Serv. Ind.*, vol. 22, Sep. 2012, doi: 10.1002/hfm.20395.

- [112] W.-L. Zheng and B.-L. Lu, “A multimodal approach to estimating vigilance using EEG and forehead EOG,” *J. Neural Eng.*, vol. 14, no. 2, p. 026017, Feb. 2017, doi: 10.1088/1741-2552/aa5a98.
- [113] M. Awais, N. Badruddin, and M. Drieberg, “A Hybrid Approach to Detect Driver Drowsiness Utilizing Physiological Signals to Improve System Performance and Wearability,” *Sensors*, vol. 17, no. 9, p. 1991, Aug. 2017, doi: 10.3390/s17091991.
- [114] “Bandpass Filtering and the Hilbert Transform,” in *Analyzing Neural Time Series Data: Theory and Practice*, 2014. doi: 10.7551/mitpress/9609.001.0001.
- [115] J. V. Stone, “Independent component analysis: an introduction,” *Trends Cogn. Sci.*, vol. 6, no. 2, pp. 59–64, Feb. 2002, doi: 10.1016/S1364-6613(00)01813-1.
- [116] J. W. Cooley and J. W. Tukey, “An algorithm for the machine calculation of complex Fourier series,” *Math. Comput.*, vol. 19, no. 90, pp. 297–301, 1965, doi: 10.1090/S0025-5718-1965-0178586-1.
- [117] P. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *IEEE Trans. Audio Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967, doi: 10.1109/TAU.1967.1161901.
- [118] D. J. Thomson, “Spectrum estimation and harmonic analysis,” *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sep. 1982, doi: 10.1109/PROC.1982.12433.
- [119] T. C. Urdan, *Statistics in Plain English*. Taylor & Francis, 2016.
- [120] A. P. King and R. J. Eckersley, “Chapter 5 - Inferential Statistics II: Parametric Hypothesis Testing,” in *Statistics for Biomedical Engineers and Scientists*, A. P. King and R. J. Eckersley, Eds. Academic Press, 2019, pp. 91–117. doi: 10.1016/B978-0-08-102939-8.00014-1.
- [121] A. P. King and R. J. Eckersley, “Chapter 6 - Inferential Statistics III: Nonparametric Hypothesis Testing,” in *Statistics for Biomedical Engineers and Scientists*, A. P. King and R. J. Eckersley, Eds. Academic Press, 2019, pp. 119–145. doi: 10.1016/B978-0-08-102939-8.00015-3.
- [122] H. B. Mann and D. R. Whitney, “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other,” *Ann. Math. Stat.*, vol. 18, no. 1, pp. 50–60, 1947.
- [123] S. Raschka, “Naive Bayes and Text Classification I - Introduction and Theory.” arXiv, Feb. 14, 2017. doi: 10.48550/arXiv.1410.5329.
- [124] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.

- [125] K. Chomboon, P. Chujai, P. Teerarassammee, K. Kerdprasop, and N. Kerdprasop, “An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm,” in *The Proceedings of the 2nd International Conference on Industrial Application Engineering 2015*, 2015, pp. 280–285. doi: 10.12792/iciae2015.051.
- [126] “Modeling class probabilities via logistic regression,” in *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, 3rd Edition*, 3rd edition., Birmingham Mumbai: Packt Publishing, 2019.
- [127] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data Min. Knowl. Discov.*, vol. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.
- [128] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [129] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [130] A. Jovic, K. Brkic, and N. Bogunovic, “A review of feature selection methods with applications,” *2015 38th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO*, pp. 1200–1205, May 2015, doi: 10.1109/MIPRO.2015.7160458.
- [131] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002, doi: 10.1109/4235.996017.

Publications

Publication 1

I. Stancin, M. Cifrek, and A. Jovic, "A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems," *Sensors*, vol. 21, no. 11, p. 3786, May 2021, doi: 10.3390/s21113786

Review

A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems

Igor Stancin, Mario Cifrek  and Alan Jovic * 

Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia; igor.stancin@fer.hr (I.S.); mario.cifrek@fer.hr (M.C.)

* Correspondence: alan.jovic@fer.hr

Abstract: Detecting drowsiness in drivers, especially multi-level drowsiness, is a difficult problem that is often approached using neurophysiological signals as the basis for building a reliable system. In this context, electroencephalogram (EEG) signals are the most important source of data to achieve successful detection. In this paper, we first review EEG signal features used in the literature for a variety of tasks, then we focus on reviewing the applications of EEG features and deep learning approaches in driver drowsiness detection, and finally we discuss the open challenges and opportunities in improving driver drowsiness detection based on EEG. We show that the number of studies on driver drowsiness detection systems has increased in recent years and that future systems need to consider the wide variety of EEG signal features and deep learning approaches to increase the accuracy of detection.

Keywords: drowsiness detection; EEG features; feature extraction; machine learning; drowsiness classification; fatigue detection; deep learning



Citation: Stancin, I.; Cifrek, M.; Jovic, A. A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems. *Sensors* **2021**, *21*, 3786. <https://doi.org/10.3390/s21113786>

Academic Editor: Giovanni Sparacino

Received: 29 March 2021

Accepted: 28 May 2021

Published: 30 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Many industries (manufacturing, logistics, transport, emergency ambulance, and similar) run their operations 24/7, meaning their workers work in shifts. Working in shifts causes misalignment with the internal biological circadian rhythm of many individuals, which can lead to sleeping disorders, drowsiness, fatigue, mood disturbances, and other long-term health problems [1–4]. Besides misalignment of the internal circadian rhythms with a work shift, sleep deprivation and prolonged physical or mental activity can also cause drowsiness [5–7]. Drowsiness increases the risk of accidents at the workplace [8–10], and it is one of the main risk factors in road and air traffic accidents per reports from NASA [11] and the US National Transportation Safety Board [12].

Drowsiness is the intermediate state between awareness and sleep [13–15]. Terms like tiredness and sleepiness are used as synonyms for drowsiness [16–18]. Some researchers also use fatigue as synonymous with drowsiness [19,20]. Definitions and differences between drowsiness and fatigue are addressed in many research papers [21–23]. The main difference between the two states is that short rest abates fatigue, while it aggravates drowsiness [24]. However, although the definitions are different, drowsiness and fatigue show similar behavior in terms of features measured from electroencephalogram (EEG) signal [25–28]. Because of this fact, in this review paper, we consider all the research papers whose topic was drowsiness, sleepiness, or fatigue, and we make no distinction among them.

The maximum number of hours that professional drivers are allowed to drive in a day is limited, yet drowsiness is still a major problem in traffic. A system for drowsiness detection with early warnings could address this problem. The most commonly used methods for drowsiness detection are self-assessment of drowsiness, driving events measures, eye-tracking measures, and EEG measures. Among these methods, drowsiness detection systems based on the EEG signal show the most promising results [18,29].

Brain neural network is a nonlinear dissipative system, i.e., it is a non-stationary system with a nonlinear relationship between causes and effects [30]. One way to analyze brain neural network is through feature extraction from the EEG signal. The most used techniques for feature extraction are linear, such as Fast Fourier Transform (FFT). Although it is a linear method, FFT also assumes that the amplitudes of all frequency components are constant over time, which is not the case with brain oscillations, since they are non-stationary. Because of the complexity of brain dynamics, there is a need for feature extraction methods that can properly take into account the nonlinearity and non-stationarity of brain dynamics. With an increase of computational power in recent years, many researchers work on improving the feature extraction methods, and there is a growing number of various features extracted from the EEG signal.

This paper aims to review the features extracted from the EEG signal and the applications of these features to the problem of driver drowsiness detection. We review the features since the large number of features described in the literature makes it difficult to understand their interrelationships, and also makes it difficult to choose the right ones for the given problem. To our knowledge, there is no similar review work that covers all the features discussed in this review. After the EEG features review, we continue with the review of driver drowsiness detection systems based on EEG. The main goal is to gain insight into the most commonly used EEG features and recent deep learning approaches for drowsiness detection, which would allow us to identify possibilities for further improvements of drowsiness detection systems. Finally, the main contributions of our work are the following: (1) Comprehensive review, systematization, and a brief introduction of the existing features of the EEG signal, (2) comprehensive review of the drowsiness detection systems based on the EEG signal, (3) comprehensive review of the existing similar reviews, and (4) discussion of various potential ways to improve the state of the art of drowsiness detection systems.

The paper is organized as follows: In Section 2, we present the overview of the existing review papers that are close to the topic of this paper, Section 3 provides the overview of the different features extracted from the EEG signal, Section 4 reviews the papers dealing with driver drowsiness detection systems, Section 5 provides a discussion about the features and drowsiness detection systems, and Section 6 brings the future directions of research and concludes the paper.

The search for the relevant papers included in our paper was done in the Web of Science Core Collection database. The search queries used were: (1) In Section 2.1—“{review, overview} {time, frequency, spectral, nonlinear, fractal, entropy, spatial, temporal, network, complex network} EEG features”, (2) in Section 2.2—“{review, overview} driver {drowsiness, sleepiness, fatigue} {detection, classification}”, (3) in Section 3—“<feature name> EEG feature”, (4) in Section 4—“EEG driver {“, deep learning, neural network} {drowsiness, sleepiness, fatigue} {detection, classification}”. Beyond the mentioned queries, when appropriate, we also reviewed the papers cited in the results obtained through the query. Additional constraints for papers in Section 4 were: (1) They had to be published in a scientific journal, (2) they had to be published in 2010 or later, (3) at least three citations per year since the paper was published, (4) papers from 2020 or 2021 were also considered with less than three citations per year, but published in Q1 journals, and (5) the number of participants in the study experiment had to be greater than 10. The goal of these constraints was to ensure that only high quality and relevant papers were included in our study.

2. Related Work

2.1. Reviews of the EEG Signal Features

Stam [30] in his seminal review paper about the nonlinear dynamical analysis of the EEG and magnetoencephalogram (MEG) signals included more than 20 nonlinear and spatiotemporal features (e.g., correlation dimension, Lyapunov exponent, phase synchronization). The theoretical background of these features and dynamical systems were also covered. The paper gave an overview of the other research works that include explanations

of the features from the fields of normal resting-state EEG, sleep, epilepsy, psychiatric diseases, normal cognition, distributed cognition, and dementia. The main drawback of the paper nowadays is that it is somewhat dated (from 2005) because additional approaches have been introduced in the meantime. Ma et al. [31] reviewed the most-used fractal-based features and entropies for the EEG signal analysis, and focused on the application of these features to sleep analysis. The authors concluded that using fractal or entropy methods may facilitate automatic sleep classification. Keshmiri [32], in a recent paper, provided a review on the usage of entropy in the fields of altered state of consciousness and brain aging. The author's work is mostly domain-specific, as it emphasizes incremental findings in each area of research rather than the specific types of entropies that were utilized in the reviewed research papers. Sun et al. [33] reviewed the complexity features in mild cognitive impairment and Alzheimer's disease. They described the usage of five time-domain entropies, three frequency-domain entropies, and four chaos theory-based complexity measures.

Motamedi-Fakhr et al. [34], in their review paper, provided an overview of more than 15 most-used features and methods (e.g., Hjorth parameters, coherence analysis, short-time Fourier transform, wavelet transform) for human sleep analysis. The features were classified into temporal, spectral, time-frequency, and nonlinear features. Besides these features, they also reviewed the research papers about sleep stages classification. Rashid et al. [35] reviewed the current status, challenges, and possible solutions for EEG-based brain-computer interface. Within their work, they also briefly discussed the most used features for brain-computer interfaces classified into time domain, frequency domain, time-frequency domain, and spatial domain.

Bastos and Schoffelen [36] provided a tutorial review of methods for functional connectivity analysis. The authors aimed to provide an intuitive explanation of how functional connectivity measures work and highlighted five interpretational caveats: The common reference problem, the signal-to-noise ratio, the volume conduction problem, the common input problem, and the sample size problem. Kida et al. [37], in their recent review paper, provided the definition, computation, short history, and pros and cons of the connectivity and complex network analysis applied to EEG/MEG signals. The authors briefly described the recent developments in the source reconstruction algorithms essential for the source-space connectivity and network analysis.

Khosla et al. [38], in their review, covered the applications of the EEG signals based on computer-aided technologies, ranging from the diagnosis of various neurological disorders such as epilepsy, major depressive disorder, alcohol use disorder, and dementia to the monitoring of other applications such as motor imagery, identity authentication, emotion recognition, sleep stage classification, eye state detection, and drowsiness monitoring. By reviewing these EEG signal-based applications, the authors listed features observed in these papers (without explanations), publicly available databases, preprocessing methods, feature selection methods, and used classification models. For the application of drowsiness monitoring, the authors reviewed only three papers, while other applications were better covered.

Ismail and Karwowski [39] overview paper dealt with a graph theory-based modeling of functional brain connectivity based on the EEG signal in the context of neuroergonomics. The authors concluded that the graph theory measures have attracted increasing attention in recent years, with the highest frequency of publications in 2018. They reviewed 20 graph theory-based measures and stated that the clustering coefficient and characteristic path length were mostly used in this domain.

Figure 1 shows the reviews presented in this section in chronological order of publication.

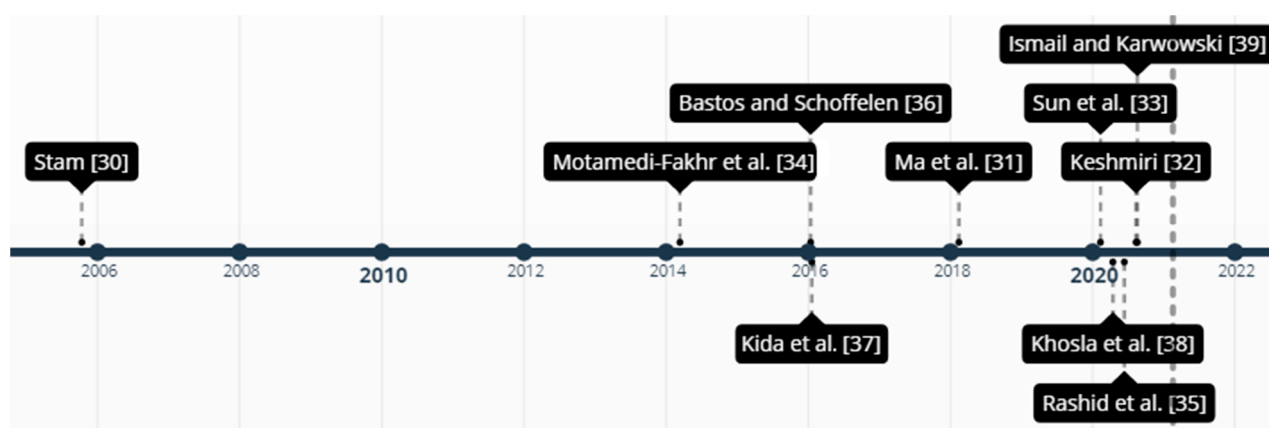


Figure 1. Chronologically ordered reviews of the EEG signal features.

2.2. Reviews of the Driver Drowsiness Detection

Lal and Craig [18], in their review of driver drowsiness systems, discussed the concept of fatigue and summarized the psychophysiological representation of driver fatigue. They concluded that most studies had found a correlation of theta and delta activity with the transition to fatigue.

Lenne and Jacobs [40], in their review paper, assessed the recent developments in the detection and prediction of drowsiness-related driving events. The driving events observed were the number of line crossings, the standard deviation of lateral position, the variability of lateral position, steering wheel variability, speed adjustments, and similar events. The authors concluded that these driving performance measures correlate with drowsiness in the experimental settings, although they stipulated that the new findings from on-road studies show a different impact on performance measures. Doudou et al. [41] reviewed the vehicle-based, video-based, and physiological signals-based techniques for drowsiness detection. They also reviewed the available commercial market solutions for drowsiness detection. When it comes to the EEG signal drowsiness detection, the authors included six papers that consider frequency-domain features in this field.

Sahayadhas et al. [42] reviewed vehicle-based measures, behavior-based measures, and physiological measures for driver drowsiness detection. The section on physiological measures included 12 papers with only frequency-domain features. Sikander and Anwar [43] reviewed drowsiness detection methods and categorized them into five groups—subjective reporting, driver biological features, driver physical features, vehicular features while driving, and hybrid features. When it comes to drowsiness detection using EEG signals, the authors focused more on explaining frequency-domain features used for drowsiness detection rather than presenting research that had already been done in this field.

Chowdhury et al. [44] reviewed different physiological sensors applied to driver drowsiness detection. Observed physiological methods for measuring drowsiness included electrocardiogram (ECG), respiratory belt, EEG, electrooculogram (EOG), electromyogram (EMG), galvanic skin response (GSR), skin temperature, and hybrid techniques. Related to EEG methods, the authors included papers based on the spectral power features, event-related potentials, and entropies. The authors also discussed different materials used for dry electrodes and the problem of measurement intrusiveness for the drivers.

Balandong et al. [45] split driver drowsiness detection systems into six categories based on the used technique—(1) subjective measures, (2) vehicle-based measures, (3) driver's behavior-based system, (4) mathematical models of sleep–wake dynamics, (5) human physiological signal-based systems, and (6) hybrid systems. The authors emphasized human physiological signal-based systems, but only the systems that rely on a limited number of EEG electrodes, as these kinds of systems are more practical for real-world applications. The authors concluded that the best results were obtained when the problem

was observed as a binary classification problem and that the fusion of the EEG features with other physiological signals should lead to improved accuracy.

Other review papers of driver drowsiness systems are specialized for a certain aspect of the field, e.g., Hu and Lodewijks [46] focused on differentiating the detection of passive fatigue, active fatigue, and sleepiness based on physiological signals, subjective assessment, driving behavior, and ocular metrics, Soares et al. [47] studied simulator experiments for drowsiness detection, Bier et al. [48] put focus on the monotony-related fatigue, and Philips et al. [49] studied operational actions (e.g., optimal staff, optimal schedule design) that reduce risk of drowsiness occurrence.

Figure 2 shows the reviews presented in this section in chronological order of publication.

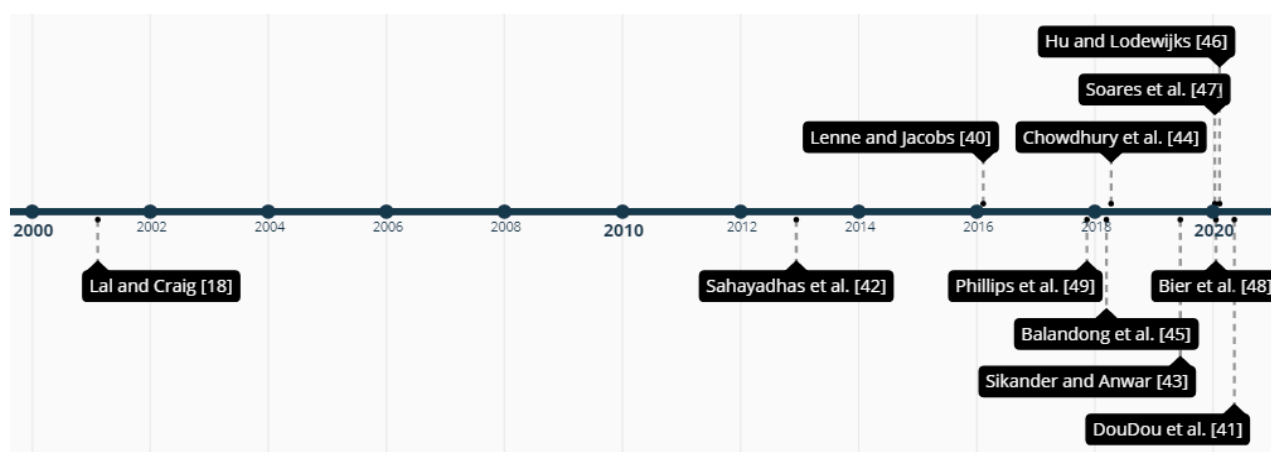


Figure 2. Chronologically ordered reviews of driver drowsiness detection methods.

3. EEG Features

The purpose of this section is to introduce features that researchers extract from the EEG signal. We will not go into the details of the computation for each feature. For the readers who are interested in the detailed computation for each feature, we suggest reading the cited papers. Instead, the main idea is to present, with a brief explanation, as many features as possible, which will later allow us to identify opportunities for further improvements in the area of driver drowsiness detection. Tables 1 and 2 show the list of all the features introduced in the following subsections. In the rest of this Section, we will use **bold** letters for the first occurrence of a particular feature name and *italic* letters for the first occurrence of a particular feature transformation or extraction method name.

3.1. Time, Frequency and Time-Frequency Domain Features

3.1.1. Time-Domain Features

The simplest features of the EEG signal are statistical features, like **mean**, **median**, **variance**, **standard deviation**, **skewness**, **kurtosis**, and similar [50]. **Zero-crossing rate (ZCR)** [51] is not a statistical feature, yet it is also a simple feature. It is the number of times that the signal crosses the *x*-axis. The *period-amplitude analysis* is based on the analysis of the half-waves, i.e., signals between two zero-crossings. With the period amplitude analysis, one can extract the **number of waves**, **wave duration**, **peak amplitude**, and **instantaneous frequency (IF)** (based only on the single observed half-wave) [52].

Hjorth parameters are features that are based on the variance of the derivatives of the EEG signal. **Mobility**, **activity**, and **complexity** [53] are the first three derivatives of the signal and the most-used Hjorth parameters. Mean absolute value of mobility, activity, and complexity can also be used as a features [54]. **K-complex** [55] is a characteristic waveform of the EEG signal that occurs in stage two of the non-rapid eye movement sleep phase. **Energy (E)** of the signal is the sum of the squares of amplitude.

3.1.2. Frequency-Domain Features

The power spectral density (PSD) of the signal, which is the base for calculation of the frequency domain features, can be calculated with several parametric and non-parametric methods. Non-parametric methods are used more often and include methods like Fourier transform (usually calculated with *Fast Fourier transform* algorithm, FFT [56]), *Welch's method* [57], or *Thompson multitaper method* [58]. Examples of parametric methods for the PSD estimation are the *autoregressive* (AR) models [59], *multivariate autoregressive* models [60], or the *autoregressive-moving average* (ARMA) models [61]. The non-parametric models have a more widespread usage, because there is no need for selecting parameters such as the model's order, which is the case for autoregressive models.

Statistical features like **mean**, **median**, **variance**, **standard deviation**, **skewness**, **kurtosis**, and similar are also used in the frequency domain. Relative powers of the certain frequency bands are the most used frequency-domain features in all fields of analysis of the EEG signals. The most commonly used frequency bands are **delta** (δ , 0.5–4 Hz), **theta** (θ , 4–8 Hz), **alpha** (α , 8–12 Hz), **beta** (β , 12–30 Hz), and **gamma** (γ , >30 Hz), band. There is also the **sigma band** (σ , 12–14 Hz) that is sometimes called **sleep spindles** [62]. Several ratios between frequency bands are widely used as features in the EEG signal analysis, i.e., θ/α [63], β/α [63], $(\theta + \alpha)/\beta$ [64], θ/β [64], $(\theta + \alpha)/(\alpha + \beta)$ [64], γ/δ [65] and $(\gamma + \beta)/(\delta + \alpha)$ [65].

Table 1. The list of time-domain, frequency domain and nonlinear features reviewed in this work.

Group	Feature Name	Abbr.	Group	Feature Name	Abbr.
Time-domain	Mean		Frequency-domain	θ/β	
	Median			$(\theta + \alpha)/(\alpha + \beta)$	
	Variance			γ/δ	
	Standard deviation			$(\gamma + \beta)/(\delta + \alpha)$	
	Skewness			Reflection coefficients	
	Kurtosis			Partial correlation coefficient	
	Zero-crossing rate	ZCR		Wavelet coefficients	
	Number of waves			Phase coupling	
	Wave duration		Nonlinear	Hurst exponent	H
	Peak amplitude			Renyi scaling exponent	
	Instantaneous frequency	IF		Renyi gener. dim. multifractals	
	Hjorth parameters			Capacity dimension D0	D0
	Mobility			Information dimension D1	D1
	Activity			Correlation dimension D2	D2
	Complexity			Katz fractal dimension	KFD
	K-complex			Petrosian fractal dimension	PFD
	Energy	E		Higuchi fractal dimension	HFD
Frequency-domain	Mean			Fractal spectrum	
	Median			Lyapunov exponents	LE
	Variance			Lempel-Ziv complexity	LZC
	Standard deviation			Central tendency measure	CTM
	Skewness			Auto-mutual information	AMI
	Kurtosis			Temporal irreversibility	
	Delta	δ		Recurrence rate	RR
	Theta	θ		Determinism	Det
	Alpha	α		Laminarity	Lam
	Beta	β		Average diagonal line length	L
	Gamma	γ		Maximum length of diagonal	Lmax
	Sigma	σ		Max. length of vertical lines	Vmax
	θ/α			Trapping time	TT
	β/α			Divergence	Div
	$(\theta + \alpha)/\beta$			Entropy of recurrence plot	ENTR

Table 2. The list of entropies, undirected and directed spatiotemporal (spt.), and complex network features reviewed in this work.

Group	Feature Name	Abbr.	Group	Feature Name	Abbr.
Entropies	Shannon entropy		Undirected spt.	Imaginary component of Coh	
	Renyi's entropy			Phase-lag index	PLI
	Tsallis entropy			Weighted phase lag index	wPLI
	Kraskov entropy	KE		Debiased weighted PLI	dwPLI
	Spectral entropy	SEN		Pairwise phase consistency	PPC
	Quadratic Renyi's SEN	QRSEN		Generalized synchronization	
	Response entropy	RE		Synchronization likelihood	SL
	State entropy	SE		Mutual information	MI
	Wavelet entropy	WE		Mutual information in freq.	MIF
	Tsallis wavelet entropy	TWE		Cross-RQA	
	Rényi's wavelet entropy	RWE		Correlation length	ξKLD
	Hilbert-Huang SEN	HHSE	Directed spt.	Granger causality	
	Log energy entropy	LogEn		Spectral Granger causality	
	Multiresolution entropy			Phase slope index	PSI
	Kolmogorov's entropy		Complex networks	Number of vertices	
	Nonlinear forecasting entropy			Number of edges	
	Maximum-likelihood entropy			Degree	D
	Coarse-grained entropy			Mean degree	
	Correntropy	CoE		Degree distribution	
	Approximate entropy	ApEn		Degree correlation	r
	Sample entropy	SampEn		Kappa	k
	Quadratic sample entropy	QSE		Clustering coefficient	
	Multiscale entropy	MSE		Transitivity	
	Modified multiscale entropy	MMSE		Motif	
	Composite multiscale entropy	CMSE		Characteristic path length	
	Permutation entropy	PE		Small worldness	
	Renyi's permutation entropy	RPE		Assortativity	
	Permutation Rényi entropy	PEr		Efficiency	
	Multivariate PE	MvPE		Local efficiency	
	Tsallis permutation entropy	TPE		Global efficiency	
	Dispersion entropy	DisE		Modularity	
	Amplitude-aware PE	AAPE		Centrality degree	
	Bubble entropy	BE		Closeness centrality	
	Differential entropy	DifE		Eigenvalue centrality	
	Fuzzy entropy	FuzzyEn		Betweenness centrality	
	Transfer entropy	TrEn		Diameter	d
Undirected spt.	Coherence			Eccentricity	Ecc
	Partial coherence			Hubs	
	Phase coherence			Rich club	
	Phase-locking value	PLV		Leaf fraction	
	Coherency	Coh		Hierarchy	Th

The frequency domain of the signal can also be obtained using *wavelet decomposition* [66,67] and *matching pursuit decomposition* [68,69] methods. Unlike Fourier transform, which decomposes a signal into sinusoids, wavelet decomposition uses an underlying mother wavelet function for decomposition, and matching pursuit decomposition uses the dictionaries of signals to find the best fit for the signal.

From autoregressive models, one can extract features such as **reflection coefficients** or **partial correlation coefficients**. **Wavelet coefficients** obtained after applying wavelet decomposition can also be used as features. PSD is usually used to obtain the second-order statistics of the EEG signal. However, one can also consider the higher-order spectrum. For

example, **phase coupling** [70] of different frequency components can be obtained with the higher-order spectral analysis.

3.1.3. Time-Frequency Features

The analysis of the EEG signal in the domains of time and frequency simultaneously is a powerful tool, since the EEG signal is a non-stationary signal [71,72]. The most important component of time-frequency domain analysis is the possibility to observe changes in the frequency over time. Short-time Fourier transform (STFT) is the simplest function that uses uniform separation of the observed signal and calculates its frequency components. A *spectrogram* [71] can be obtained with the application of STFT. Wavelet transform [73] is the usual alternative method to spectrogram that also provides coefficients as features from the time-frequency domain. The main advantage of wavelet transform compared to spectrogram is a variable window size, dependent on spectrum frequencies.

3.2. Nonlinear Features

Brain dynamics constitute a complex system. A system is complex when it is constructed from many nonlinear subsystems that cannot be separated into smaller subsystems without changing their dynamical properties. Fractal systems are often used for describing the brain dynamics measured with the EEG signal. To explain fractal systems, first, we need to introduce the scaling law. The scaling law is describing (asymptotically) a self-similar function F as a function of the scale parameter s , i.e., $F(s) \sim s^\alpha$. When applied to a self-affine signal, each axis should be scaled by a different power factor to obtain statistically equivalent changes in both directions. If s is used in the x -axis direction, then $s' = s^H$ should be used in the y -axis direction. The power factor H is called the **Hurst exponent** [74,75]. The Hurst exponent is a measure of long-term memory of the signal and is related to the fractal dimension with the equation $D_0 = 2 - H$ for self-similar time-series, where fractal dimension D_0 is defined in the next paragraph. Time-series q is monofractal if it is linearly interdependent with its **Renyi scaling exponent** $\tau(q)$, otherwise, it is multifractal. The **Renyi generalized dimension of multifractals** is defined as $D(q) = \tau(q)/(q - 1)$. For more detailed explanations about fractality and multifractality of the time-series, we refer the reader to [76–78].

In EEG signal analysis, all fractal dimensions are estimated based on the underlying attractor (a geometric structure towards which stationary dissipative system gravitates in its state space) of the signal [79]. In a strict mathematical sense, most time-series have the one-dimensional **support fractal dimension** D_0 (or capacity dimension or Hausdorff dimension) if there are no missing values. Regardless of the value of the D_0 , the **information dimension** D_1 and **correlation dimension** D_2 [79–81] can be calculated. The correlational dimension D_2 can be calculated with both monofractal and multifractal approaches. The **Katz fractal dimension** (KFD) [82], the **Petrosian fractal dimension** (PFD) [83], and the **Higuchi fractal dimension** (HFD) [84] are different approaches to the estimation of the fractal dimension. With multifractal time-series analysis, a **fractal spectrum** consisting of multiple fractal dimensions can be obtained [85,86].

Methods for fractal time-series analysis can be classified [76] into stationary analysis methods (such as *Fluctuation Analysis* [87], *Hurst's Rescaled-Range Analysis* [74], and similar), non-stationary analysis (such as *Detrended Fluctuation Analysis* [88], *Centered Moving Average Analysis* [89], *Triangle Total Areas* [90], and similar), and multifractal analysis (such as *Wavelet Transform Modulus Maxima* [91], *Multifractal Detrended Fluctuation Analysis* [92], and similar). Each of these methods provides its own estimation of fractal dimension or scaling exponent features.

Lyapunov exponents (LE) [93] are measures of the attractor's complexity. If a system has at least one positive Lyapunov exponent, then the system can be characterized as a chaotic dynamical system. A positive Lyapunov exponent points to exponential divergence of the two nearby trajectories in the attractor over time [94]. **Lempel-Ziv complexity** (LZC) [95] is a measure of complexity that binarizes time-series and then searches for the

occurrence of consecutive binary characters or “words” and counts the number of times a new “word” is encountered. The **Central tendency measure** (CTM) [96] is a measure of the variability of the observed time-series and represents the percentage of points on the scatter plot that fall into a given radius. **Auto-mutual information** (AMI) [97] is a mutual information measure applied to time-delayed versions of the same EEG time-series. **Temporal irreversibility** [98] of a time-series implies the influence of nonlinear dynamics, non-Gaussian noise, or both. It is a statistical property that differs based on the direction in which time proceeds, e.g., any sequence of measurements has a different probability of occurrence than its time reverse.

A *recurrence plot* [99] is a graphical method for the detection of reoccurring patterns in the time-series. *Recurrence quantification analysis* (RQA) [100] is a group of algorithms for the automatic quantification of recurrence plots. RQA is a noise resistant method, meaning it gives good results even when the signal-to-noise ratio of considered signals is unfavorable [101]. The **recurrence rate** (RR) is the probability that a specific state of a time-series will reoccur. **Determinism** (Det) is the percentage of points that form diagonal lines on the recurrence plot and **laminarity** (Lam) is the percentage of points forming vertical lines in the recurrence plot. The **average diagonal line length** (L), **maximum length of diagonal** (Lmax), and **maximum length of vertical lines** (Vmax) are also used as RQA-based features. **Trapping time** (TT) is the average vertical line length and it relates to the predictability of the time-series. **Divergence** (Div) is the reciprocal value of the maximal diagonal line length and it can have a trend similar to the positive Lyapunov exponents. **Entropy of the recurrence plot** (ENTR) reflects the complexity of the deterministic structure of the system.

3.3. Entropies

Entropy was first introduced to the field of information theory by Shannon in 1948 [102,103]. **Shannon’s information entropy** is calculated based on the expression $-\sum_j p_j \log(p_j)$, where p_j is the probability distribution of the observed data. It is used to measure uncertainty or randomness in the observed time-series. There are many derived variations of information entropy used in EEG analysis. The entropies may be considered as nonlinear features, but we describe them in a separate subsection due to their specific calculation.

Rényi’s entropy [104] is defined with the expression $-\frac{\alpha}{1-\alpha} \sum \log p_k^\alpha$, where $\alpha > 0$ and $\alpha \neq 1$. It is a generalization of Shannon’s entropy in the case of a limited value of $\alpha \rightarrow 1$. **Quadratic Rényi’s entropy** (or just Rényi’s entropy) is the case where $\alpha = 2$. **Tsallis entropy** (q-entropy) [105] is a generalization of the Boltzman–Gibbs entropy from statistical thermodynamics and is defined with the expression $\frac{k}{q-1} \left(1 - \sum_i p_i^q\right)$, where k is a positive constant and q is the non-extensivity parameter. For $q > 1$, the entropy has a more significant reaction to the events that occur often, whereas for $0 < q < 1$, the entropy has a more significant reaction to rare events.

The three aforementioned entropies can be calculated from the raw EEG signal. Besides that, they are a base for calculating several other entropies in the field of EEG analysis. **Kraskov entropy** (KE) [50] is an unbiased estimator of Shannon’s entropy for a d -dimensional random sample. **Spectral entropy** (SEN) [106] is calculated with the expression for Shannon’s entropy based on the normalized PSD of the EEG signal. **Quadratic Rényi’s spectral entropy** (QRSEN) [107] is calculated with the usage of Rényi’s entropy expression, and the difference compared to the spectral entropy is that it gives the higher weights to the lower frequencies. Commercial M-Entropy Module [108] uses two different components of spectral entropy—**response entropy** (RE) and **state entropy** (SE). State entropy includes the spectrum between 0.8 and 32 Hz, while response entropy includes the spectrum between 0.8 and 47 Hz.

Wavelet entropy (WE) [109,110] is somewhat similar to spectral entropy. The difference is that it is calculated based on the coefficients of the wavelet decomposition of the

given time-series. There are two generalizations of wavelet entropy—**Tsallis wavelet entropy** (TWE) and **Rényi's wavelet entropy** (RWE) [111]. **Hilbert–Huang spectral entropy** (HHSE) [112] applies Shannon's entropy to the Hilbert–Huang spectrum, which is obtained by the *Hilbert–Huang transform* [111,113]. **Log energy entropy** (LogEn) [114] is similar to the wavelet entropy, but only uses summation of logarithms of the probabilities. **Multiresolution entropy** [115] uses the combination of windowing and wavelet transform for the detection of changes in parameters that define the observed process (i.e., the parameters of brain dynamics).

Kolmogorov's entropy [116] is an embedding entropy and is defined as the sum of positive Lyapunov exponents. It represents the rate of information loss and a degree of predictability (regularity) of the attractor. Accurate computation of Kolmogorov's entropy is computationally expensive, so several entropies are used for the estimation of Kolmogorov's entropy based on the less computationally expensive methods. **Nonlinear forecasting entropy** [117] is the estimation of Kolmogorov's entropy for time-series with too few points. It is based on the forecasting of the time-series data, i.e., on the correlation coefficient of the forecasted points with actually observed points. The estimation method is independent of the forecasting method used. **Maximum-likelihood entropy** [118] is also the estimation of Kolmogorov entropy. It is derived with the application of maximum-likelihood to the correlation integral, which is treated as a probability distribution. **Coarse-grained entropy** [119] is an estimation of the attractor's entropy for cases where standardly used dimensions, Lyapunov exponents, and Kolmogorov's entropy are not suitable due to the high dimensionality of the observed process. **Correntropy** (CoE) [120] is an estimation of nonlinear autocorrelation.

Approximate entropy (ApEn) [121] is derived from Kolmogorov's entropy and its use in the analysis of the EEG signal (and other physiological signals) is widespread. It addresses the irregularity of a time-series. Predictable time-series, i.e., time-series with many repetitive patterns will have a small value of approximate entropy. **Sample entropy** (SampEn) [122] was introduced as an improvement to approximate entropy. It reduces the error of the approximate entropy by eliminating its two disadvantages—(1) self-matches and (2) dependence on the time-series length. Sample entropy is also an approximation of signal complexity. **Quadratic sample entropy** (QSE) [123] is SampEn insensitive to the data radius parameter r . It allows r to vary as needed to achieve confident estimates of the conditional probability. **Multiscale entropy** (MSE) [124] is a generalization of an entropy measure (such as sample entropy) to different time scales. **Modified multiscale entropy** (MMSE) [125] uses the same procedure as MSE, but replaces coarse-graining with a moving average procedure. **Composite multiscale entropy** (CMSE) [126] is a modification of the MSE that tackles the problem of increased variance and error estimation for short time-series.

Permutation entropy (PE) [127] estimates signal variability based on the repetition of the ordinal patterns. The algorithm requires parameter m (permutation order) to obtain ordinal patterns and their probabilities of occurrence. These probabilities are then applied in Shannon's entropy expression. Moreover, **Rényi's permutation entropy** (RPE) [128], **permutation Rényi entropy** (PER) [129], **multivariate permutation entropy** (MvPE) [130], and **Tsallis permutation entropy** (TPE) [111] can be calculated for the ordinal patterns. **Dispersion entropy** (DisE) [131] is a modification of permutation entropy that tackles the problem of amplitude information loss (since permutation entropy only considers the order of the amplitude values but not the values themselves). **Amplitude-aware permutation entropy** (AAPE) [132] is based on the similar idea of using the value of the signal with the permutation entropy. **Bubble entropy** (BE) [133] is similar to permutation entropy with the main difference in the method used for ranking vectors in the embedding space. Namely, permutation entropy uses repetition of the ordinal patterns and bubble entropy uses the number of steps needed to sort a vector with the bubble sort algorithm. **Differential entropy** (DifE) [134] calculation is based on Shannon's entropy expression and the estimation of the underlying probability density function of time-series. **Fuzzy entropy**

(FuzzyEn) [135] is based on the concept of fuzzy sets, first introduced by Zadeh [136]. It is similar to sample entropy, but instead of using the Heaviside function for distance calculation, it uses a fuzzy membership function. **Transfer entropy** (TrEn) [137] uses concepts similar to mutual information (see Section 3.4) with the ability to quantify the exchange of information between two systems. It is an asymmetric measure for information transfer from process X to process Y, which measures the effect of the past values of processes X and Y on the present value of process Y.

3.4. Spatiotemporal Features

Features that were introduced above are all calculated based on a single EEG channel. Since EEG recording devices can have hundreds of channels nowadays, features that describe the relationship between different channels bring further insight into the understanding of brain functions. This is the main idea behind the usage of the spatiotemporal features—to describe the relationship between different brain regions for particular states or events. Spatiotemporal features can be divided into two groups—directed and non-directed. The non-directed ones relate to the synchronization of two or more channels without any knowledge of the direction, while the directed ones include the causation between them, i.e., they measure functional connectivity.

3.4.1. Non-Directed Spatiotemporal Features

Coherence [138] is a cross-correlation equivalent in the frequency-domain, i.e., the cross-correlation of the PSD from two different channels. It reflects the synchronization of the changes of frequency components between the observed channels. **Partial coherence** [139] is an adjusted coherence with removed common signal's linear effect based on the third channel, which is not physically close to the two observed channels. **Phase coherence** [140] is the coherence of the phases of the signals. It was introduced to overcome the problem of detection of nonlinear dependencies between the two channels.

The **phase-locking value** (PLV) [141] represents the measure of the transient phase locking that is completely independent of the signal's amplitude, which is not the case for the coherence measure. **Coherency** [142] is calculated similar to coherence, but without applying the magnitude operator to the cross-spectral density of two channels. The obtained complex-valued quantity is called coherency. The **imaginary component of coherency** (iCoh) [143] reflects the nonlinear interaction between the two underlying time-series. **Phase-lag index** (PLI) [144] is a measure of the asymmetry of the distribution of phase differences between two signals. It brings improvement compared to the imaginary component of coherency by removing the effect of amplitude information. The **weighted phase lag index** (wPLI) [145] uses weights to reduce a phase lag index's sensitivity to noise, while the **debiased weighted phase lag index** (dwPLI) [145] additionally reduces a sample-size bias. **Pairwise phase consistency** (PPC) [146] is a measure similar to PLV, but it quantifies the distribution of all pairwise phase differences across observations.

Generalized synchronization [147] incorporates the nonlinear property of the dynamical systems into its calculation. The idea is to observe two dynamical systems, a response system and a driving system, where the response system is a function of the driving system. Authors propose a numerical method called *mutual false nearest neighbors* for distinguishing between synchronized and unsynchronized behavior of the systems. *Arnhold's measure* [148] is another algorithm for measuring such interdependence between two dynamical systems. **Synchronization likelihood** (SL) [149] brings several improvements into these methods—it is sensitive to linear and nonlinear brain dynamics and is suitable for an analysis of the non-stationary systems. It is calculated based on the similarity of the time-delayed embeddings in the state space.

Mutual information (MI) [150] quantifies the amount of information obtained about one time-series through observing the other time-series. It is a commonly used measure in the information theory and is calculated based on Shannon's entropy. **Mutual information**

in frequency (MIF) [151] is a recently developed measure that calculates the mutual information between the PSDs of two time-series. Its interpretation is similar to coherence.

Cross-recurrence quantification analysis [101] is similar to RQA, but instead of observing the self-similarity of a single signal, the similarity of two different channels is observed. The features extracted are the same as in the case of single-channel RQA (see Section 3.2). The **correlation length** (ξ_{KLD}) [152] is a measure of the spatio-temporal disorder based on the Karhunen–Loeve decomposition.

3.4.2. Directed Spatiotemporal Features

Granger causality [153] is a well-known statistical test, which tests whether one time-series forecasts (causes) the other time-series, and vice-versa. It is based on the autoregressive forecast models of the two time-series. **Spectral Granger causality** [154] can also be calculated and it is based on the estimation of the spectral transfer matrix and the covariance of the autoregressive model's residuals. The **phase slope index** (PSI) [155] is a robust estimation of the information flow direction. It is insensitive to the mixtures of the independent sources, which is the main problem for Granger causality. Transfer entropy, which is explained in Section 3.3, can also be considered a directed spatiotemporal feature.

3.5. Complex Networks

The features introduced in Section 3.1, Section 3.2, and Section 3.3 were based only on a single channel of the EEG signal. Section 3.4 introduced features calculated based on the pairwise interactions between the two channels. In this section, the main goal is to introduce the features that observe the interactions between more than two channels. Complex networks are a graph-theory-based approach to EEG signal analysis. A connectivity matrix obtained by observing all pairwise connections between channels is used to obtain a graph. Any method explained in Section 3.4 can be used to determine connectivity matrix, and popular choices are correlation, PLI, or MI. Graphs can be weighted based on the values of the connectivity matrix or unweighted by applying thresholding to the connectivity matrix. A minimum spanning tree can also be used as a method for obtaining an acyclic graph with all vertices included. For more details about graph construction and complex networks, we refer the reader to papers [156,157]. In continuation of this section, we introduce features that are calculated based on the obtained graph. These features are functional connectivity features.

Once the graph is obtained, the **number of vertices** and the **number of edges** can be used as features. The **degree** (D) [158] of a vertex is the number of edges connected to the vertex. The **mean degree** of the network is a metric of density. The **degree distribution** is a probability distribution of the degrees and it provides information about the structure of the graph. **Degree correlation** (r) [159] is the correlation coefficient of degrees of pairs of neighbors in a graph. **Kappa** (k) [159] is a measure of the degree diversity and it measures the broadness of the degree distribution. The **clustering coefficient** [160] is a measure of the vertices connectedness in a graph and it can be local (for a sub-graph) or global. If the local clustering coefficient is equal to one, it means that the corresponding local sub-graph is fully connected. The global clustering coefficient is sometimes called **transitivity** [161]. A **motif** [162] is a generalized version of the clustering coefficient and a pattern of local connectivity. The average of all pairwise shortest path lengths is called **characteristic path length** [160]. **Small worldness** [163] is a second-order graph statistic and its calculation is based on the trade-off between high local clustering and short path length. **Assortativity** [164] is the measure of vertex tendency to link with other vertices with a similar number of edges.

Efficiency [165] is a measure of the efficiency of the information exchange in the graph. **Local efficiency** [165] is the inverse of the shortest path lengths between vertices on the observed sub-graph, where the sub-graph consists of all neighbors of the observed vertex. **Global efficiency** [165] is the average efficiency of the graph divided by the average

efficiency of a fully connected graph. **Modularity** [166] describes the structure of the graph and represents the degree to which a graph is subdivided into non-overlapping clusters.

Each vertex in the graph has a measure of **centrality degree** [167], which represents the number of shortest paths in the graph that the observed vertex is involved in. Similarly, each vertex in the graph has a measure of **closeness centrality** [168], which represents the average distance of the observed vertex from all other vertices in the graph. **Eigenvalue centrality** [169] is a measure of the ease of accessibility of a vertex to other vertices. It is computed based on the relative vertex scores, with the basic idea that the high-scoring connections should contribute more to vertex influence than the low-scoring vertices. **Betweenness centrality** [170] is a measure of the importance of the vertex in a graph. It is computed based on the number of times a vertex occurs along the shortest path between two other vertices.

Diameter (d) [159] is the longest shortest path of a graph. **Eccentricity** (Ecc) [159] is the longest shortest path from a referenced vertex to any other vertex in the graph. **Hubs** [171] are vertices with high centrality. Hubs tend to be connected and this property is called assortativity. **Rich club** [172] is a sub-graph of highly interconnected hubs. **Leaf fraction** [159] of a graph is the number of vertices with exactly one edge. **Hierarchy** (T_H) [159] captures the ratio between a small diameter on one hand and overloading of the hub nodes on the other hand.

4. Driver Drowsiness Detection Systems

The aim of this Section is to review the work on drowsiness detection focusing on the features used. The inclusion criteria for the papers are stated in Section 1. Tables 3 and 4 show a summary of the reviewed work on driver drowsiness detection, and the rest of the Section briefly presents each work.

Balam et al. [173] used a convolutional neural network (CNN) for the classification based on the raw EEG signal from the Cz-Oz channel. They used data from the Sleep-EDF Expanded Database and their ground truth for drowsiness was the S1 sleep stage. Since the authors used a publicly available database, they compared their deep learning (DL) approach with the other feature-based approaches, and they concluded that this approach resulted in at least 3% better results. Chaabene et al. [174] used frequency-domain features for defining the ground truth. They used CNN with raw EEG signal from seven electrodes as input and achieved 90% drowsiness detection accuracy.

Yingying et al. [175] used a Long Short-Term Memory (LSTM) network to classify sleepiness in two classes and their final classification accuracy achieved was 98%. Their ground truth labels for classification were based on the alpha-blocking phenomenon and the alpha wave attenuation-disappearance phenomenon. The authors claimed that these two phenomena represent two different sleepiness levels, relaxed wakefulness and sleep onset, respectively. The authors used only the O2 channel of the EEG signal and performed a continuous wavelet transform to obtain the PSD. Zou et al. [176] used multiscale PE, multiscale SampEn, and multiscale FuzzyEn. Their ground truth labels were based on Li's subjective fatigue scale and the accuracy achieved was 88.74%. Chaudhuri and Routray [177] used only three entropies as features—ApEn, SampEn, and modified SampEn. Their experiment was designed to slowly increase the fatigue level of the participants because of the effects of physical and mental workload, along with the effects of sleep deprivation. The experiment was divided into 11 stages and stages 7 and later were labeled as the fatigue state. The authors used SVM and achieved 86% accuracy.

Budak et al. [178] used MIT/BIH Polysomnographic EEG database in their study. Their ground truth for binary classification was based on sleep stages labeled by an expert. The awake stage was labeled the awake state and stage I of sleep was labeled the drowsy state. The authors used ZCR, E, IF, and SEN as traditional features, and also used AlexNet on the spectrogram images to obtain additional 4096 features (layers fc6 and fc7 of AlexNet). The accuracy of the binary classification was 94.31%, which is the best result achieved on this dataset, according to the authors. Mehreen et al. [179] used δ , δ/α , θ , θ/φ , $\delta/\alpha+\beta+\gamma$,

and δ/θ EEG features, along with blink features and head movement features and achieved 92% accuracy of drowsiness detection. Based on EEG features only, the accuracy was 76%. The authors used subjective evaluation with Karolinska Sleepiness Scale (KSS) as the ground truth. It is unclear how the authors converted nine levels of KSS into a two-level ground truth. Chen et al. [180] used the clustering coefficient and characteristic path length of the graph obtained for δ , θ , α , and β frequency bands. The graph was obtained using the phase lag index. The ground truth labels were binary. The first three minutes of participants' driving were labeled as alert state and the last three minutes as fatigue state. SVM was selected for classification and achieved 94.4% accuracy. The authors conclude that the functional connectivity of the brain differs significantly between the alert and fatigue state, particularly in the α and β bands.

Table 3. The summary of metadata of the reviewed driver drowsiness detection papers.

Author	Year	Participants	Electrodes
Chaabene et al. [174]	2021	12	14 channels
Balam et al. [173]	2021	23	Pz-Oz
Yingying et al. [175]	2020	12	O1 and O2
Zou et al. [176]	2020	16	32 channels
Chaudhuri and Routray [177]	2020	12	19 Channels
Budak et al. [178]	2019	16	C3-O1, C4-A1, and O2-A1
Chen et al. [179]	2019	14	14 channels
Mehreen et al. [180]	2019	50	AF7, AF8, TP9 and TP10
Martensson et al. [181]	2019	86	Fz-A1, Cz-A2 and Oz-Pz
Barua et al. [182]	2019	30	30 channels
Ogino and Mitsukura [183]	2018	29	Fp1
Chen et al. [184]	2018	15	30 channels
Chen et al. [185]	2018	15	30 channels
Chen et al. [186]	2018	12	40 channels
Hu and Min [187]	2018	22	30 channels
Dimitrakopoulos et al. [188]	2018	40	64 channels
Hong et al. [189]	2018	16	Ear channel
Li and Chung [190]	2018	17	O1 and O2
Min et al. [191]	2017	12	32 channels
Awais et al. [192]	2017	22	19 channels
Nguyen et al. [193]	2017	11	64 channels
Hu [194]	2017	28	32 channels
Chai et al. [195]	2017	43	32 channels
Chai et al. [196]	2017	43	32 channels
Mu et al. [197]	2017	11	27 channels
Fu et al. [198]	2016	12	O1 and O2
Ahn et al. [199]	2016	11	64 channels
Huang et al. [200]	2016	12	30 channels
Li et al. [201]	2015	20	O1 and O2
Chen et al. [202]	2015	16	9 channels
Sauvet et al. [203]	2014	14	C3-M2 and O1-M2
Lee et al. [204]	2014	20	Fpz-Cz and Pz-Oz
Garces Correa et al. [205]	2014	18	C3-O1, C4-A1 and O2-A1
Zhang et al. [110]	2014	20	O1 and O2
Hu et al. [206]	2013	40	Fz-A1, Cz-A2 and Oz-Pz
Picot et al. [207]	2012	20	F3, C3, P3 and O1
Zhao et al. [208]	2011	13	32 channels
Khushaba et al. [20]	2011	31	Fz, T8 and Oz
Liu et al. [209]	2010	50	13 channels

Table 4. The summary of reviewed driver drowsiness detection papers. The meanings of the abbreviations are: TD—time-domain, FD—frequency-domain, N—nonlinear, EN—entropies, CN—complex networks, SIG—signal-based labeling, Li’s—Li’s subjective fatigue scale, SD—sleep deprivation, NREM1—labels based on the sleep stages, BE3—first and last three minutes as two labels, BE5—first and last five minutes as two labels, BIH—behavior-based labeling, WIE—Wierwille scale, RT—reaction time based labeling, EXP—expert labeling, LSTM—long-short term memory, KNN—k nearest neighbor, SVM—support vector machine, RF—random forest, ELM—extreme learning machine, GBDT—gradient boosting decision tree, NN—neural network, FLDA—Fisher linear discriminant analysis, SDBN—sparse deep belief network, HMM—hidden Markov model, and Thres.—thresholding-based algorithm.

Author	Features	Target	Algorithm	No. Classes	Acc.
Chaabene et al. [174]	Raw	SIG	CNN	2	90.14
Balam et al. [173]	Raw	NREM1	CNN	2	94.00
Yingying et al. [175]	FD	SIG	LSTM	2	98.14
Zou et al. [176]	EN	Li’s	KNN		88.74
Chaudhuri and Routray [177]	EN	SD	SVM	2	86.00
Budak et al. [178]	TD, FD, EN and special	NREM1	LSTM	2	94.31
Chen et al. [179]	CN	BE3	SVM	2	94.40
Mehreen et al. [180]	FD	KSS	SVM	2	92.00
Martensson et al. [181]	FD, N and EN	KSS	RF	2	93.50
Barua et al. [182]	TD, FD and EN	KSS	SVM	2 and 3	93.00 and 79.00
Ogino and Mitsukura [183]	FD and EN	KSS	SVM	2	67.00
Chen et al. [184]	CN	KSS			
Chen et al. [185]	CN	KSS	KNN	2	98.60
Chen et al. [186]	CN	BE3	ELM	2	95.00
Hu and Min [187]	EN	BE5	GBDT	2	94.00
Dimitrakopoulos et al. [188]	CN	BE5	SVM	2	92.10
Hong et al. [189]	FD, N and EN	EBE	SVM	5	99.50
Li and Chung [190]	FD	WIE	SVM	5	93.87
Min et al. [191]	FD and EN	BE5	NN	2	98.30
Awais et al. [192]	TD, FD and EN	BIH	SVM	2	80.00
Nguyen et al. [193]	FD	SIG	FLDA	2	79.20
Hu [194]	EN	BE5	AdaBoost	2	97.50
Chai et al. [195]	FD	BE5	SDBN	2	90.60
Chai et al. [196]	FD	BE5	NN	2	88.20
Mu et al. [197]	EN	Li’s	SVM	2	97.00
Fu et al. [198]	FD	KSS	HMM	3	AUC 0.841
Ahn et al. [199]	FD	SD	FLDA	2	75.90
Huang et al. [200]	FD	RT			
Li et al. [201]	FD	BIH	SVM	2	93.16
Chen et al. [202]	FD, N and EN	SIG	ELM	2	95.60
Sauvet et al. [203]	FD	EXP	Threshold	2	98.30
Lee et al. [204]	TD and FD	NREM1	SVM	4	98.50
Garces Correa et al. [205]	TD and FD	NREM1	NN	2	87.40
Zhang et al. [110]	N and EN	SIG	NN	4	96.50
Hu et al. [206]	FD	KSS	SVM	2	75.00
Picot et al. [207]	FD	SIG	Threshold	5	80.60
Zhao et al. [208]	FD	Li’s	SVM	3	81.60
Khushaba et al. [20]	FD	WIE	LDA	5	95.00
Liu et al. [209]	EN	KSS and Li’s	HMM	2	84.00

Martensson et al. [181] used θ , α , $\theta/(\theta + \alpha)$, $\alpha/(\theta + \alpha)$, $(\theta + \alpha)/\beta$, α/β , $(\theta + \alpha)/(\theta + \beta)$, θ/β , SampEn, and HFD from three EEG channels together with features from EOG and ECG signals. The authors performed a sequential forward floating feature selection method for dimensionality reduction and six EEG features were selected—HFD, θ , $\alpha/(\theta + \alpha)$, θ/β , $\theta/(\theta + \alpha)$ and α . Random forest was selected as the best model and achieved 93.5% accuracy on the test set and 84% on the leave-one-subject-out validation scheme. The ground truth was obtained with the KSS. The severely sleepy class was for a KSS score greater than seven and the sufficiently alert class was for a KSS score of less than seven.

KSS scores equal to seven were discarded as outlined. Barua et al. [182] used δ , θ , α , β , γ , $(\theta + \alpha)/\beta$, α/β , $(\theta + \alpha)/(\alpha + \beta)$, and θ/β from 30 EEG channels along with features from EOG and contextual information (e.g., time awake, duration of last sleep, and the like). The authors achieved the best accuracy of 93% for binary classification and 79% for classification into three classes. Self-evaluation with KSS score was used as ground truth and KSS score was classified into three classes—alert class for KSS scores below six, somewhat sleepy class for KSS scores below eight, and sleepy for KSS scores equal to eight or nine. In the binary classification, the authors used two methods (fuzzy centroid redistribution and SVM predicted redistribution) for redistribution of somewhat sleepy classes into the alert and sleepy classes. Ogino and Mitsukura [183] used δ , θ , α , β , and γ as frequency domain features, and parameters of the autoregressive model and MSE were also added to the feature set. Only the Fp1 channel was used and the authors achieved 67% accuracy by using SVM. The ground truth labels were based on the KSS score, where the alert class was for a KSS score less than four and the drowsy class was for a KSS score greater than six.

Chen et al. [184] analyzed the difference in complex network features for each frequency band (δ , θ , α , and β) between alert and drowsy states. The authors used the features: Number of vertices, number of edges, D, leaf fraction, d, Ecc, betweenness centrality, k, Th, and r. Their ground truth was based on the KSS score. A significant difference was found in the four features of the δ -band and five features of the θ -band. In addition, the authors suggested a more linear graph configuration in alert states and a more star-shaped graph configuration in drowsy states. Chen et al. [185] used the same experiment for drowsiness classification in a related study. Three complex network features (degree, degree correlation, and kappa) were extracted for each frequency band (δ , θ , α , and β). The ground truth was based on the KSS score and they performed binary classification. The highest accuracy of 98.6% was achieved using the k nearest neighbor (KNN) algorithm. Chen et al. [186] used phase synchronization, phase coherence, k, betweenness centrality, and Th as features. The first three minutes of participants' driving were labeled as an alert state and the last three minutes as a fatigue state. The highest accuracy achieved was 95% using the extreme learning machine (ELM) algorithm. Dimitrakopoulos et al. [187] used 64 channels and computed three complex network features—clustering coefficient, characteristic path length, and small-worldness. The authors achieved 92.1% accuracy for drowsiness classification. The network values of the first and the last 5-min windows were used to indicate the states of maximum alertness and maximum fatigue, respectively.

Hong et al. [188] used δ , θ , α , β , ratio indices, frequency domain statistics, the generalized Hurst exponent, HFD, SEN, and PE from the ear channel together with photoplethysmography (PPG) and ECG. The highest accuracy achieved was 99.5%. The ground truth labels were divided into five levels and were labeled by experts based on behavioral expressions. The authors ranked the features using four different methods, and in each method, at least four of the seven best-ranked features were nonlinear features. Hu and Min [189] used 30 channels and four entropies from each channel—SEN, ApEn, SampEn, and FuzzyEn. The authors achieved 94% accuracy in drowsiness classification. They used a ground truth based on self-reported fatigue. If the measurement lasted longer than 30 min before the participant self-reported fatigue, the signals from the 5th to 10th minute were used as the normal state and the signals from the last five minutes before the end of the experiment were used as the fatigued state. Li and Chung [190] used θ , α , and β features from O1 and O2 channels along with gyroscope-based head movement measurement. The subjective Wierwille scale was used to obtain five-level ground truth. The achieved accuracy for five-level classification was 93% and for binary classification it was 96%. Awais et al. [191] used mean, variance, minimum, maximum, E, SampEn, δ , θ , α , β , and γ from 19 channels along with ECG signal. SVM was used for classification and they achieved 80% accuracy for binary classification. When only EEG features were used, the accuracy was 76%. The authors used video-based facial features including eye blink duration, facial expressions, facial tone, eye blinking rate, and movements such as

head-nodding and yawning for establishing ground truth. When a drowsy event began, five minutes before it were marked as the alert state and five minutes after it were marked as the drowsy state.

Min et al. [192] used SEN, ApEn, SampEn, and FuzzyEn for fatigue detection. These four entropies gave better results than AR coefficients. An experiment was terminated based on the subjective report of fatigue. To confirm these fatigue reports, the authors utilized the Chalder Fatigue Scale and Li's Subjective Fatigue Scale before and after the experiment. The first five minutes of the recording were labeled as the normal state and the last five minutes of the recording were labeled as the fatigued state. The authors achieved an accuracy of 98.3%. Nguyen et al. [193] used δ , θ , α , β , and γ features from 64 channels along with near-infrared spectroscopy (NIRS). EOG and ECG signals were also measured, but they were only used to establish the ground truth labels. Fisher linear discriminant analysis (FLDA) was used for binary classification with 79.2% accuracy when EEG and NIRS were used. The accuracy when only EEG features were used was 70.5%. The authors introduced the drowsiness detection index, a variable derived for drowsiness detection, and they reported that it predicts the onset of drowsiness on average 3.6 s earlier. Hu [194] used SEN, ApEn, SampEn, and FuzzyEn features from 32 channels of the EEG signal. An experiment was terminated based on the EOG parameter associated with fatigue and self-reported fatigue. The first five minutes were labeled as the normal state and the last five minutes were labeled as the fatigue state. The AdaBoost classification algorithm was used and achieved 97.5% accuracy. Chai et al. [195] used AR coefficients as features. The ground truth labels were binary, with the first five minutes of driving labeled as the alert state and the last five minutes of driving labeled as the fatigued state. An experiment was terminated when the participant drove off the road for 15 s or when consistent signs of fatigue (such as head nodding and prolonged eye closure) were detected. The authors used NN for classification and achieved 88.2% accuracy. Chai et al. [196] used AR features from 32 channels. The first five minutes of data were used as an alert state and the last five minutes as a drowsy state. The authors used a sparse deep belief network as a classification algorithm and achieved 90% accuracy. Mu et al. [197] used FuzzyEn from Fp1 and Fp2 channels and achieved 97% accuracy using the SVM algorithm. The ground truth labels were binary with the first 10 min labeled as the normal state and the last 10 min labeled as the fatigued state. The stopping criteria of the experiment were based on Li's subjective fatigue scale and Borg's CR-10 scale.

Fu et al. [198] used θ , α , and β features from O1 and O2 channels along with EMG and respiration. The ground truth was set based on the KSS score, where level one was KSS score equal to one or two, level two was KSS score equal to three or four, and level three was KSS score equal to five or six. The reported average area under the curve (AUC) was 0.841. When only EEG features were used, the average AUC was 0.644. Ahn et al. [199] used δ , θ , α , β , and γ along with EOG, ECG, and fNIRS. FLDA was used for binary classification with 79.2% accuracy using all the available sensors. The accuracy based only on the EEG signal features was 59.7%. Binary ground truth was used with the well-rested group and the sleep-deprived group. Huang et al. [200] used only the α feature. The system developed in this study did not use a classification algorithm. It was based on measuring the response times of the subjects. Drowsiness was labeled for the moments when the response time was 2.5 times greater than the mean response time, which helped the authors to determine a threshold for α feature value indicating drowsiness. An auditory warning system was developed to help subjects to remain alert.

Li et al. [201] used θ , α , and β features from O1 and O2 channels. The ground truth alert and drowsy data were labeled based on the percentage of eyelid closure (PERCLOS) and the number of adjustments on the steering wheel. The best accuracy of 93.16% was achieved using the SVM classifier and only θ and β features. The authors used the probability of prediction instead of the discrete class label to develop an early warning system with a probability threshold of 0.424. Chen et al. [202] used δ , θ , α , β , γ , ApEn, SampEn, Rényi's entropy, and RQA features, along with the EOG. Two neurologists manually labeled binary

ground truth values based on the EOG features and frequency domain features. ELM was used for classification based on the nonlinear features only and achieved 95.6% accuracy. Sauvet et al. [203] used θ , α , β , $(\theta + \alpha)/\beta$, and fuzzy fusion of these features. Feature thresholding was applied for classification and an accuracy of 98.3% was achieved. The ground truth was based on expert scoring, but it is unclear how this scoring was performed.

Lee et al. [204] used δ , θ , α , β , time-domain statistics, ZCR, and several ratio indices from Fpz-Cz and Pz-Oz EEG channels. The ground truth was classified into four classes: Awake, slightly drowsy, moderately drowsy, and extremely drowsy. These classes were determined by experienced physicians, with the first three classes being derived from the awake-sleep stage and the extremely drowsy class corresponding to the N1 sleep stage. SVM was used for classification and the best accuracy achieved was 98.5%. Garces Correa et al. [205] used MIT-BIH Polysomnographic Database in their research. Eighteen subjects were selected and δ , θ , α , β , γ , time-domain statistics, and frequency domain statistics features were extracted. The ground truth alert and drowsy labels were determined based on the awake and S1 sleep stages, respectively. A neural network was used for classification and it achieved 87.4% accuracy. Zhang et al. [110] used LZC and peak-to-peak versions of ApEn and SampEn. Peak-to-peak means that instead of using all the data points of the features, the authors used only the difference between the maximum and minimum values in the sliding window. Four levels of ground truth labels were used, referred to as normal state, mild fatigue, mood swing, and excessive fatigue. These labels were determined based on the various entropy patterns used in the paper, but it is unclear exactly how the labels were determined. A neural network was used for classification and it achieved 96.5% accuracy.

Hu et al. [206] used δ , θ , α , β , and frequency domain statistics along with EOG signal features. The authors achieved a final drowsiness detection accuracy of 75%. Binary ground truth labels were used. The alert state was defined with a KSS score less than 8 and Karolinska drowsiness score (KDS) equal to 0, while drowsiness was defined with a KSS score greater than 7 and a KDS score equal to or greater than 50. The KDS is an EEG/EOG-based drowsiness scoring experiment where the final score is between 0% (alert) and 100% (drowsy) [210]. Picot et al. [207] used only α and β features from the P3 channel together with the EOG signal. The ground truth was labeled by experts based on the EEG and EOG signal. Five levels were used in labeling the ground truth, but three levels were used to evaluate the drowsiness detection system. The drowsiness detection system was based on the statistical test to compare the two populations and thresholding, and achieved an accuracy of 80.6%. Zhao et al. [208] used multivariate autoregressive coefficients as features along with the EOG signal. The accuracy achieved with the SVM classifier was 81.6%. The ground truth labels were based on Li's subjective fatigue scale. Khushaba et al. [20] introduced a hybrid type of EEG features called fuzzy mutual information-based wavelet-packet features, and achieved a drowsiness detection accuracy of 95%. Their ground truth had five levels and was based on Wierwille and Ellsworth criteria. Wierwille and Ellsworth criteria [211] is a textual description of the drowsiness continuum based on behavioral and facial signs that should prepare raters to rate participants' drowsiness based on observations of the video while driving. Liu et al. [209] used ApEn and Kolmogorov entropy of the δ , θ , α , and β frequency bands. The ground truth was binary with pre-task time as the alert state and post-task time as the fatigue state. The authors confirmed a statistically significant increase in fatigue level based on the five different subjective scales—KSS, Stanford sleepiness scale, Samn–Perelli checklist, Li's subjective fatigue scale, and Borg's CR-10 scale. A hidden Markov model was used for classification and achieved 84% accuracy.

5. Discussion

Section 3 presented 147 features that were classified into 7 categories, as shown in Tables 1 and 2. As mentioned, Tables 3 and 4 show a summary of 39 reviewed papers on drowsiness detection. The year with the most papers meeting the inclusion criteria is 2018

with eight included papers. Figure 3 shows the number of included papers and the number of papers as a result of the search query: “EEG driver drowsiness detection”. Based on both trends, it can be seen that the number of papers on this topic is increasing.

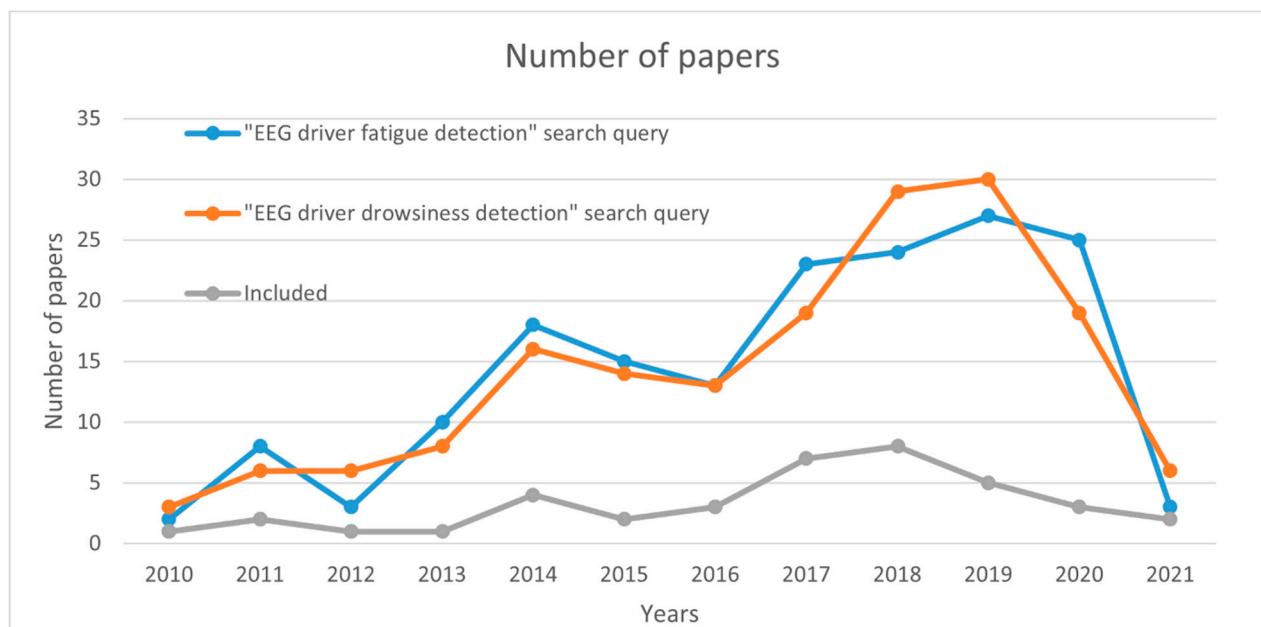


Figure 3. The number of papers included in the study and the number of papers obtained as a result of the “EEG driver drowsiness detection” and “EEG driver fatigue detection” search query, data until April 2021.

From 2013 to 2016, there were only two papers that used entropies and eight papers that used only frequency domain statistics. Although there is a higher number of published papers in recent years, there are fewer papers that rely only on frequency-domain features. Nonlinear features, entropies, and complex network features have been increasingly used in recent years. Reported drowsiness detection accuracies have remained more or less the same over the years and are usually between 80% and 99%. There is an increasing body of work that has been done with higher numbers of participants (30 or more), and it is reasonable to assume that the accuracies from these works are the most reliable.

Although we often refer to accuracy as a quality measure for the developed system, it must be noted that it is not possible to fairly compare the accuracy of different works because most of the works have been performed with a private dataset based on different experimental designs.

Besides the different datasets used, we observe that the methodology used for validation of the drowsiness detection systems is also a common problem. As mentioned earlier, EEG signal is a non-stationary and nonlinear signal with high inter-individual differences. Because of these properties, the only proper way for model validation is the validation on the signals from an unseen subject. Empirical tests show that there is a large difference in the accuracies between validation on the unseen subjects and validation on the unseen parts of the signal [212]. Reporting of validation with improper methodology can create overexpectation of the model performance, bad generalization on the unseen subjects, and can lead other researchers in the wrong direction. This effect is visible through the examples of papers that use validation on the unseen subjects, but also report about validation on the unseen parts of the signal in order to be comparable with existing research [173]. The fourth inclusion constraint defined in Section 1 was used to eliminate the papers that have a low probability of achieving good generalization due to a low number of participants.

The highest accuracy achieved was 99.5% in the work of Hong et al. [188]. It is interesting to note that the authors included features from three different categories. The authors used standard frequency bands and ratio indices, the nonlinear generalized Hurst

exponent and HFD, and the entropies SEN and PE. Although this is not a large number of features, it is reasonable to assume that their diversity leads to the high accuracy of drowsiness detection. It is difficult to say how reliable the given accuracy is because only 16 participants took part in the experiment and there may be a high sampling bias in the data. The study by Martensson et al. [181] also used features from three different categories. The features used were standard frequency domain features and ratio indices, entropy SampEn, and nonlinear HFD. This study had the largest number of participants (86), and the accuracy achieved was 93.5%. These two studies suggest that using different types of features should result in high accuracy of drowsiness detection.

Complex network features for EEG signal analysis have become very popular in recent years, and this is also true for drowsiness detection systems. There are four papers [179,184–186] that include complex network features. One of them only provides analysis without classification and the remaining three have high accuracy—93%, 94%, and 98%. Complex networks are a promising approach, but confirming the reliability of such a system, especially when combined with features from other categories, requires studies with a large number of participants.

There is also a growing body of research on drowsiness detection using deep learning models. Deep learning models are known for their high ability to learn hidden structures in the data, but they often require a large amount of data for proper training. They can be used with the raw data as input, but also with features, or both. There are five papers using deep learning that met our inclusion criteria. In the first one, the authors used the LSTM network with raw data and different types of features and achieved 94.4% accuracy [178]. Their research was based on only 16 participants. The second one also used LSTM, but for prediction of the underlying alpha phenomena that is the base for determining drowsiness level [175]. The other three papers used CNN as a classification method. The highest accuracy achieved was 94% and the model used only raw data, without any pre-computed EEG signal features [173].

The reported accuracies for these deep learning models are in line with the accuracies of other models but, as we stated earlier, a direct comparison of the accuracies may lead to the wrong conclusions. Balam et al. [173] provided a proper comparison of different approaches. The authors used a publicly available dataset, so they were able to provide a fair comparison of different approaches. Their CNN approach was compared with one research based on the LSTM network and seven feature-based research studies. The best accuracy was obtained with their proposed method, while the LSTM method had a slightly lower accuracy. All seven feature-based approaches had more than 5% lower accuracy on average. A similar comparison was provided in Budak et al. [178] on a different publicly available dataset. Furthermore, the difference was that the authors used features and raw data for their LSTM model. The comparison was made with one deep learning approach and six other feature-based approaches. Again, the feature-based approaches had a lower performance by about 7%, on average.

These two pieces of research suggest that the deep learning approach is more appropriate and has higher performance for drowsiness detection than the feature-based approach. Nevertheless, it must be noted that all of the feature-based approaches that had lower accuracy used only time-domain and/or frequency-domain features. As shown and discussed earlier, the addition of different types of features could lead to an improvement of these models. From the inspected literature, it is currently unclear whether the inclusion of additional features would outperform deep learning models. In addition, it would be interesting to examine what effect would the addition of the features that are a measure of signal's memory (like Hurst exponent) have, since the LSTM model also relies on the previous values of the signal. However, we can speculate that the addition of the memory-based features would increase the accuracy of these feature-based models, but probably not enough to outperform LSTM models. The reason for this is because deep models have a higher capacity for learning hidden structures than the memory-based features, but additional research should be made to support the speculation.

A larger amount of data is needed for proper training of deep learning models compared to non-deep learning models. Acquiring the data is often a problem when it comes to EEG-based drowsiness detection. Authors of research studies that use deep learning approaches often employ generative adversarial networks for the augmentation of the dataset [175]. This process often leads to an improved performance of the model. Regardless of the possibilities for augmentation of the dataset, researchers should strive to gather as much as possible real EEG signals. The larger number of participants would ensure greater diversity of the dataset, reduce the influence of inter-individual differences in EEG signals, make models more robust, and allow enough data for proper validation of models.

As we discussed earlier, there is evidence that different types of features improve drowsiness detection models. In the papers that met our inclusion criteria, about 50 different features were used, while we introduced 147 EEG-based features in our review. Approximately 100 unused features provide much room for further research. In particular, spatiotemporal features were only used to obtain a graph for complex network features [184].

Another way to improve such systems is to set better ground truth labels. Currently, many works use subjective self-evaluation as ground truth. The KSS is used most often for this purpose. The KSS is a nine-level scale, with the first four levels describing alertness, the 5th neutral level, and the last four levels describing sleepiness. The four levels for alertness and sleepiness have detailed descriptions, and they are very similar. It is also hard to tell if the scale is linear with the same distances between adjacent levels. Since it is a subjective scale with small differences between adjacent levels, it may lead to subjectivity bias and inconsistencies in the ground truth labels, which was confirmed in [191], where the authors state after the statistical test results: “Subjective measures were not reliable for detecting drowsiness alone, and that solely relying on self-reported measures may not provide a meaningful measure of a person’s actual physiological state.” Future research on how to provide a unified definition and description of drowsiness is needed to combat this subjectivity bias.

For future research, we recommend the development of a drowsiness detection system that consider raw data, features from all seven categories, and deep learning models. Ground truth labels should be based on the unified, standard definition and description of drowsiness. If there is not yet research providing such a unified definition of drowsiness, then ground truth should be confirmed with multiple independent sources to reduce subjectivity bias (even expert labels are prone to subjectivity). Because electrophysiological signals have high interindividual differences, a large number of participants (about 100 or more [181]) is needed to reduce sample bias and increase the chances of a model to have good generalization.

6. Conclusions

With this review paper, we bring four contributions: (1) Comprehensive review, systematization, and a brief description of the existing features of the EEG signal, (2) comprehensive review of the drowsiness detection systems, (3) comprehensive review of the existing similar reviews, and (4) discussion of various potential ways to improve the state of the art of drowsiness detection systems. In continuation, we summarize our suggestions for the general improvement of the field of drowsiness detection systems.

A higher number of participants in the experiments (about 100 or more) is needed to ensure diversity of a dataset, reduce the influence of inter-individual differences of EEG signals, make models more robust, and allow enough data for proper validation of models. Validation of EEG-based driver drowsiness detection should always be done based on the data from unseen subjects (for example, using leave-one-subject-out cross-validation). Whenever possible, datasets should be published publicly to allow fair comparison of different approaches. Based only on the papers from this review, without additional research, we were not able to identify a single feature or a feature category that guarantees the best performance of the drowsiness detection system. What we can conclude is that

a higher number of features from at least four different categories should lead to more reliable drowsiness detection systems with lower sampling bias and higher generalization ability. Deep learning models exhibit higher performance for drowsiness detection than the considered non-deep learning models based on time and frequency-domain features. Nevertheless, the use of pre-computed EEG signal features together with deep learning models should always be considered (in addition to raw EEG data modeling), since in some cases, the addition of pre-computed features to deep learning models additionally boosted performance.

For future research that would have a strong impact on the field of drowsiness detection systems, we suggest the development of a unified, standard definition and description of drowsiness, which would lead to a reduction in subjective bias and easier comparison of different studies.

Author Contributions: Conceptualization, I.S. and A.J.; methodology, I.S. and A.J.; validation, I.S., M.C. and A.J.; formal analysis, I.S. and A.J.; investigation, I.S.; resources, I.S.; data curation, I.S.; writing—original draft preparation, I.S.; writing—review and editing, M.C. and A.J.; visualization, I.S.; supervision, A.J.; project administration, M.C. and A.J.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been carried out within the project “Research and development of the system for driver drowsiness and distraction identification—DFDM”, funded by the European Regional Development Fund in the Republic of Croatia under the Operational Programme Competitiveness and Cohesion 2014–2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Eastman, C.I.; Martin, S.K. How to use light and dark to produce circadian adaptation to night shift work. *Ann. Med.* **1999**, *31*, 87–98. [\[CrossRef\]](#)
- Chellappa, S.L. Circadian misalignment: A biological basis for mood vulnerability in shift work. *Eur. J. Neurosci.* **2020**, *52*, 3846–3850. [\[CrossRef\]](#)
- Kang, J.-H.; Miao, N.-F.; Tseng, I.-J.; Sithole, T.; Chung, M.-H. Circadian Activity Rhythms and Sleep in Nurses Working Fixed 8-hr Shifts. *Biol. Res. Nurs.* **2015**, *17*, 348–355. [\[CrossRef\]](#)
- Arendt, J. Shift work: Coping with the biological clock. *Occup. Med.* **2010**, *60*, 10–20. [\[CrossRef\]](#)
- Brown, I.D. Driver Fatigue. *Hum. Factors J. Hum. Factors Ergon. Soc.* **1994**, *36*, 298–314. [\[CrossRef\]](#)
- Soleimanloo, S.S.; Wilkinson, V.E.; Cori, J.M.; Westlake, J.; Stevens, B.; Downey, L.; Shiferaw, B.A.; Rajaratnam, S.M.W.; Howard, M.E. Eye-Blink Parameters Detect On-Road Track-Driving Impairment Following Severe Sleep Deprivation. *J. Clin. Sleep Med.* **2019**, *15*, 1271–1284. [\[CrossRef\]](#) [\[PubMed\]](#)
- Vicente, J.; Laguna, P.; Bartra, A.; Bailón, R. Drowsiness detection using heart rate variability. *Med. Biol. Eng. Comput.* **2016**, *54*, 927–937. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chellappa, S.L.; Morris, C.J.; Scheer, F.A.J.L. Effects of circadian misalignment on cognition in chronic shift workers. *Sci. Rep.* **2019**, *9*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
- Folkard, S. Shift work, safety and productivity. *Occup. Med.* **2003**, *53*, 95–101. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ross, J.K. Offshore industry shift work—health and social considerations. *Occup. Med.* **2009**, *59*, 310–315. [\[CrossRef\]](#)
- Orasanu, J.; Parke, B.; Kraft, N.; Tada, Y.; Hobbs, A.; Anderson, B.; Dulchinos, V. *Evaluating the Effectiveness of Schedule Changes for Air Traffic Service (ATS) Providers: Controller Alertness and Fatigue Monitoring Study*; Technical Report; Federal Aviation Administration, National Technical Information Service (NTIS): Alexandria, VA, USA, 2012.
- National Transportation Safety Board. *Most Wanted List of Transportation Safety Improvements: Reduce Fatigue-Related Accidents*; National Transportation Safety Board: Washington, DC, USA, 2018.
- Corsi-Cabrera, M.; Arce, C.; Ramos, J.; Lorenzo, I.; Guevara, M.A. Time Course of Reaction Time and EEG While Performing a Vigilance Task During Total Sleep Deprivation. *Sleep* **1996**, *19*, 563–569. [\[CrossRef\]](#)
- Jackson, M.; Kennedy, G.A.; Clarke, C.; Gullo, M.; Swann, P.; Downey, L.; Hayley, A.C.; Pierce, R.J.; Howard, M.E. The utility of automated measures of ocular metrics for detecting driver drowsiness during extended wakefulness. *Accid. Anal. Prev.* **2016**, *87*, 127–133. [\[CrossRef\]](#) [\[PubMed\]](#)

15. Oken, B.S.; Salinsky, M.C.; Elsas, S.M. Vigilance, alertness, or sustained attention: Physiological basis and measurement. *Clin. Neurophysiol.* **2006**, *117*, 1885–1901. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Kamran, M.A.; Mannan, M.M.N.; Jeong, M.Y. Drowsiness, Fatigue and Poor Sleep's Causes and Detection: A Comprehensive Study. *IEEE Access* **2019**, *7*, 167172–167186. [\[CrossRef\]](#)
17. Papadelis, C.; Chen, Z.; Kourtidou-Papadeli, C.; Bamidis, P.D.; Chouvarda, I.; Bekiaris, E.; Maglaveras, N. Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clin. Neurophysiol.* **2007**, *118*, 1906–1922. [\[CrossRef\]](#)
18. Lal, S.K.; Craig, A. A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* **2001**, *55*, 173–194. [\[CrossRef\]](#)
19. Lin, C.-T.; Wu, R.-C.; Liang, S.-F.; Chao, W.-H.; Chen, Y.-J.; Jung, T.-P. EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2005**, *52*, 2726–2738. [\[CrossRef\]](#)
20. Khushaba, R.N.; Kodagoda, S.; Lal, S.; Dissanayake, G. Driver Drowsiness Classification Using Fuzzy Wavelet-Packet-Based Feature-Extraction Algorithm. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 121–131. [\[CrossRef\]](#)
21. Neu, D.; Linkowski, P.; Le Bon, O. Clinical complaints of daytime sleepiness and fatigue: How to distinguish and treat them, especially when they become 'excessive' or 'chronic'? *Acta Neurol. Belg.* **2010**, *110*, 15–25. [\[PubMed\]](#)
22. Neu, D.; Mairesse, O.; Verbanck, P.; Linkowski, P.; Le Bon, O. Non-REM sleep EEG power distribution in fatigue and sleepiness. *J. Psychosom. Res.* **2014**, *76*, 286–291. [\[CrossRef\]](#)
23. Phillips, R.O. A review of definitions of fatigue—And a step towards a whole definition. *Transp. Res. Part F Traffic Psychol. Behav.* **2015**, *29*, 48–56. [\[CrossRef\]](#)
24. Johns, M.W.; Chapman, R.J.; Crowley, K.E.; Tucker, A. A new method for assessing the risks of drowsiness while driving. *Somnologie Schlafforschung Und Schlafmed.* **2008**, *12*, 66–74. [\[CrossRef\]](#)
25. Simon, M.; Schmidt, E.A.; Kincses, W.E.; Fritzsche, M.; Bruns, A.; Aufmuth, C.; Bogdan, M.; Rosenstiel, W.; Schrauf, M. EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clin. Neurophysiol.* **2011**, *122*, 1168–1178. [\[CrossRef\]](#)
26. Craig, A.; Tran, Y.; Wijesuriya, N.; Nguyen, H. Regional brain wave activity changes associated with fatigue. *Psychophysiology* **2012**, *49*, 574–582. [\[CrossRef\]](#)
27. Aeschbach, D.; Matthews, J.R.; Postolache, T.T.; Jackson, M.A.; Giesen, H.A.; Wehr, T.A. Dynamics of the human EEG during prolonged wakefulness: Evidence for frequency-specific circadian and homeostatic influences. *Neurosci. Lett.* **1997**, *239*, 121–124. [\[CrossRef\]](#)
28. Borghini, G.; Astolfi, L.; Vecchiato, G.; Mattia, D.; Babiloni, F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* **2014**, *44*, 58–75. [\[CrossRef\]](#)
29. Kundinger, T.; Sofra, N.; Riener, A. Assessment of the Potential of Wrist-Worn Wearable Sensors for Driver Drowsiness Detection. *Sensors* **2020**, *20*, 1029. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Stam, C. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clin. Neurophysiol.* **2005**, *116*, 2266–2301. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Ma, Y.; Shi, W.; Peng, C.-K.; Yang, A.C. Nonlinear dynamical analysis of sleep electroencephalography using fractal and entropy approaches. *Sleep Med. Rev.* **2018**, *37*, 85–93. [\[CrossRef\]](#)
32. Keshmiri, S. Entropy and the Brain: An Overview. *Entropy* **2020**, *22*, 917. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Sun, J.; Wang, B.; Niu, Y.; Tan, Y.; Fan, C.; Zhang, N.; Xue, J.; Wei, J.; Xiang, J. Complexity Analysis of EEG, MEG, and fMRI in Mild Cognitive Impairment and Alzheimer's Disease: A Review. *Entropy* **2020**, *22*, 239. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Motamedi-Fakhr, S.; Moshrefi-Torbati, M.; Hill, M.; Hill, C.M.; White, P.R. Signal processing techniques applied to human sleep EEG signals—A review. *Biomed. Signal Process. Control.* **2014**, *10*, 21–33. [\[CrossRef\]](#)
35. Rashid, M.; Sulaiman, N.; Majeed, A.P.P.A.; Musa, R.M.; Nasir, A.F.A.; Bari, B.S.; Khatun, S. Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. *Front. Neurobot.* **2020**, *14*. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Bastos, A.M.; Schoffelen, J.-M. A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Front. Syst. Neurosci.* **2016**, *9*, 175. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Kida, T.; Tanaka, E.; Kakigi, R. Multi-Dimensional Dynamics of Human Electromagnetic Brain Activity. *Front. Hum. Neurosci.* **2016**, *9*. [\[CrossRef\]](#)
38. Khosla, A.; Khandnor, P.; Chand, T. A comparative analysis of signal processing and classification methods for different applications based on EEG signals. *Biocybern. Biomed. Eng.* **2020**, *40*, 649–690. [\[CrossRef\]](#)
39. Ismail, L.E.; Karwowski, W. A Graph Theory-Based Modeling of Functional Brain Connectivity Based on EEG: A Systematic Review in the Context of Neuroergonomics. *IEEE Access* **2020**, *8*, 155103–155135. [\[CrossRef\]](#)
40. Lenné, M.G.; Jacobs, E.E. Predicting drowsiness-related driving events: A review of recent research methods and future opportunities. *Theor. Issues Ergon. Sci.* **2016**, *17*, 533–553. [\[CrossRef\]](#)
41. Doudou, M.; Bouabdallah, A.; Berge-Cherfaoui, V. Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges. *Int. J. Intell. Transp. Syst. Res.* **2019**, *18*, 297–319. [\[CrossRef\]](#)
42. Sahayadhas, A.; Sundaraj, K.; Murugappan, M. Detecting Driver Drowsiness Based on Sensors: A Review. *Sensors* **2012**, *12*, 16937–16953. [\[CrossRef\]](#)
43. Sikander, G.; Anwar, S. Driver Fatigue Detection Systems: A Review. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2339–2352. [\[CrossRef\]](#)

44. Chowdhury, A.; Shankaran, R.; Kavakli, M.; Haque, M. Sensor Applications and Physiological Features in Drivers' Drowsiness Detection: A Review. *IEEE Sens. J.* **2018**, *18*, 3055–3067. [\[CrossRef\]](#)
45. Balandong, R.P.; Ahmad, R.F.; Saad, M.N.M.; Malik, A.S. A Review on EEG-Based Automatic Sleepiness Detection Systems for Driver. *IEEE Access* **2018**, *6*, 22908–22919. [\[CrossRef\]](#)
46. Hu, X.; Lodewijks, G. Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue. *J. Saf. Res.* **2020**, *72*, 173–187. [\[CrossRef\]](#)
47. Soares, S.; Ferreira, S.; Couto, A. Driving simulator experiments to study drowsiness: A systematic review. *Traffic Inj. Prev.* **2020**, *21*, 29–37. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Bier, L.; Wolf, P.; Hilsenbek, H.; Abendroth, B. How to measure monotony-related fatigue? A systematic review of fatigue measurement methods for use on driving tests. *Theor. Issues Ergon. Sci.* **2018**, *21*, 22–55. [\[CrossRef\]](#)
49. Phillips, R.O.; Kecklund, G.; Anund, A.; Sallinen, M. Fatigue in transport: A review of exposure, risks, checks and controls. *Transp. Rev.* **2017**, *37*, 742–766. [\[CrossRef\]](#)
50. Jiang, D.; Lu, Y.-N.; Ma, Y.; Wang, Y. Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement. *Expert Syst. Appl.* **2019**, *121*, 188–203. [\[CrossRef\]](#)
51. Michielli, N.; Acharya, U.R.; Molinari, F. Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals. *Comput. Biol. Med.* **2019**, *106*, 71–81. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Geering, B.A.; Achermann, P.; Eggimann, F.; Borbély, A.A. Period-amplitude analysis and power spectral analysis: A comparison based on all-night sleep EEG recordings. *J. Sleep Res.* **1993**, *2*, 121–129. [\[CrossRef\]](#)
53. Hjorth, B. EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* **1970**, *29*, 306–310. [\[CrossRef\]](#)
54. Wang, X.-W.; Nie, D.; Lu, B.-L. EEG-Based Emotion Recognition Using Frequency Domain Features and Support Vector Machines. In *International Conference on Neural Information Processing, ICONIP 2011*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 734–743. [\[CrossRef\]](#)
55. Oliveira, G.H.; Coutinho, L.R.; da Silva, J.C.; Pinto, I.J.; Ferreira, J.M.; Silva, F.J.; Santos, D.V.; Teles, A.S. Multitaper-based method for automatic k-complex detection in human sleep EEG. *Expert Syst. Appl.* **2020**, *151*, 113331. [\[CrossRef\]](#)
56. Nussbaumer, H.J. *The Fast Fourier Transform*; Springer: Berlin/Heidelberg, Germany, 1981; pp. 80–111. [\[CrossRef\]](#)
57. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73. [\[CrossRef\]](#)
58. Thomson, D. Spectrum estimation and harmonic analysis. *Proc. IEEE* **1982**, *70*, 1055–1096. [\[CrossRef\]](#)
59. Akaike, H. Fitting autoregressive models for prediction. *Ann. Inst. Stat. Math.* **1969**, *21*, 243–247. [\[CrossRef\]](#)
60. Neumaier, A.; Schneider, T. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw.* **2001**, *27*, 27–57. [\[CrossRef\]](#)
61. Said, S.E.; Dickey, D.A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **1984**, *71*, 599–607. [\[CrossRef\]](#)
62. De Gennaro, L.; Ferrara, M. Sleep spindles: An overview. *Sleep Med. Rev.* **2003**, *7*, 423–440. [\[CrossRef\]](#)
63. Eoh, H.J.; Chung, M.K.; Kim, S.-H. Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *Int. J. Ind. Ergon.* **2005**, *35*, 307–320. [\[CrossRef\]](#)
64. Jap, B.T.; Lal, S.; Fischer, P.; Bekiaris, E. Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Syst. Appl.* **2009**, *36*, 2352–2359. [\[CrossRef\]](#)
65. da Silva, T.L.; Kozakevicius, A.J.; Rodrigues, C.R. Automated drowsiness detection through wavelet packet analysis of a single EEG channel. *Expert Syst. Appl.* **2016**, *55*, 559–565. [\[CrossRef\]](#)
66. Adeli, H.; Zhou, Z.; Dadmehr, N. Analysis of EEG records in an epileptic patient using wavelet transform. *J. Neurosci. Methods* **2003**, *123*, 69–87. [\[CrossRef\]](#)
67. Stépiane, M. *A Wavelet Tour of Signal Processing*; Elsevier: Amsterdam, The Netherlands, 2009. [\[CrossRef\]](#)
68. Franaszczuk, P.; Bergey, G.K.; Durka, P.; Eisenberg, H.M. Time-frequency analysis using the matching pursuit algorithm applied to seizures originating from the mesial temporal lobe. *Electroencephalogr. Clin. Neurophysiol.* **1998**, *106*, 513–521. [\[CrossRef\]](#)
69. Durka, P.; Ircha, D.; Blinowska, K. Stochastic time-frequency dictionaries for matching pursuit. *IEEE Trans. Signal Process.* **2001**, *49*, 507–510. [\[CrossRef\]](#)
70. Eiselt, M.; Schelenz, C.; Witte, H.; Schwab, K. Time-variant Parametric Estimation of Transient Quadratic Phase Couplings during Electroencephalographic Burst Activity. *Methods Inf. Med.* **2005**, *44*, 374–383. [\[CrossRef\]](#)
71. Cohen, L. *Time-Frequency Analysis*; Prentice Hall: Hoboken, NJ, USA, 1995; Volume 778.
72. Abbate, A.; DeCusatis, C.M.; Das, P.K. Time-Frequency Analysis of Signals. In *Wavelets and Subbands*; Birkhäuser Boston: Boston, MA, USA, 2002; pp. 103–187. [\[CrossRef\]](#)
73. Meyer, Y. *Wavelets and Operators*; Cambridge University Press: Cambridge, UK, 1993. [\[CrossRef\]](#)
74. Hurst, H.E. THE PROBLEM OF LONG-TERM STORAGE IN RESERVOIRS. *Int. Assoc. Sci. Hydrol. Bull.* **1956**, *1*, 13–27. [\[CrossRef\]](#)
75. Lloyd, E.H.; Hurst, H.E.; Black, R.P.; Simaika, Y.M. Long-Term Storage: An Experimental Study. *J. R. Stat. Soc. Ser. A (Gen.)* **1966**, *129*, 591. [\[CrossRef\]](#)
76. Kantelhardt, J.W. Fractal and Multifractal Time Series. In *Encyclopedia of Complexity and Systems Science*; Springer: New York, NY, USA, 2009; pp. 3754–3779. [\[CrossRef\]](#)
77. Barnsley, M.F. *Fractals Everywhere*; Elsevier: Amsterdam, The Netherlands, 1993. [\[CrossRef\]](#)

78. Kantz, H.; Schreiber, T. *Nonlinear Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2003.
79. Pritchard, W.; Duke, D. Measuring Chaos in the Brain—A Tutorial Review of EEG Dimension Estimation. *Brain Cogn.* **1995**, *27*, 353–397. [\[CrossRef\]](#)
80. Grassberger, P.; Procaccia, I. Estimation of the Kolmogorov entropy from a chaotic signal. *Phys. Rev. A* **1983**, *28*, 2591–2593. [\[CrossRef\]](#)
81. Grassberger, P.; Procaccia, I. Measuring the strangeness of strange attractors. *Phys. D Nonlinear Phenom.* **1983**, *9*, 189–208. [\[CrossRef\]](#)
82. Dash, D.P.; Kolekar, M.H.; Jha, K. Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and Hidden Markov Model. *Comput. Biol. Med.* **2020**, *116*, 103571. [\[CrossRef\]](#)
83. Moctezuma, L.A.; Molinas, M. Towards a minimal EEG channel array for a biometric system using resting-state and a genetic algorithm for channel selection. *Sci. Rep.* **2020**, *10*, 1–14. [\[CrossRef\]](#)
84. Ferenets, R.; Lipping, T.; Anier, A.; Jantti, V.; Melto, S.; Hovilehto, S. Comparison of Entropy and Complexity Measures for the Assessment of Depth of Sedation. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1067–1077. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Zorick, T.; Mandelkern, M.A. Multifractal Detrended Fluctuation Analysis of Human EEG: Preliminary Investigation and Comparison with the Wavelet Transform Modulus Maxima Technique. *PLoS ONE* **2013**, *8*, e68360. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Muzy, J.F.; Bacry, E.; Arneodo, A. Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. *Phys. Rev. E* **1993**, *47*, 875–884. [\[CrossRef\]](#)
87. Bunde, A.; Havlin, S. (Eds.) *Fractals in Science*; Springer: Berlin/Heidelberg, Germany, 1994. [\[CrossRef\]](#)
88. Peng, C.-K.; Buldyrev, S.; Havlin, S.; Simons, M.; Stanley, H.E.; Goldberger, A.L. Mosaic organization of DNA nucleotides. *Phys. Rev. E* **1994**, *49*, 1685–1689. [\[CrossRef\]](#)
89. Álvarez-Ramírez, J.; Rodríguez, E.; Echeverría, J.C. Detrending fluctuation analysis based on moving average filtering. *Phys. A Stat. Mech. Its Appl.* **2005**, *354*, 199–219. [\[CrossRef\]](#)
90. Lotfalinezhad, H.; Maleki, A. TTA, a new approach to estimate Hurst exponent with less estimation error and computational time. *Phys. A Stat. Mech. Its Appl.* **2020**, *553*, 124093. [\[CrossRef\]](#)
91. Muzy, J.F.; Bacry, E.; Arneodo, A. Wavelets and multifractal formalism for singular signals: Application to turbulence data. *Phys. Rev. Lett.* **1991**, *67*, 3515–3518. [\[CrossRef\]](#) [\[PubMed\]](#)
92. Kantelhardt, J.W.; Zschiegner, S.A.; Koscielny-Bunde, E.; Havlin, S.; Bunde, A.; Stanley, H. Multifractal detrended fluctuation analysis of nonstationary time series. *Phys. A Stat. Mech. Its Appl.* **2002**, *316*, 87–114. [\[CrossRef\]](#)
93. Shevchenko, I.I. Lyapunov exponents in resonance multiplets. *Phys. Lett. A* **2014**, *378*, 34–42. [\[CrossRef\]](#)
94. Shen, C.; Yu, S.; Lu, J.; Chen, G. Designing Hyperchaotic Systems With Any Desired Number of Positive Lyapunov Exponents via A Simple Model. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2014**, *61*, 2380–2389. [\[CrossRef\]](#)
95. Lempel, A.; Ziv, J. On the Complexity of Finite Sequences. *IEEE Trans. Inf. Theory* **1976**, *22*, 75–81. [\[CrossRef\]](#)
96. Cohen, M.; Hudson, D.; Deedwania, P. Applying continuous chaotic modeling to cardiac signal analysis. *IEEE Eng. Med. Biol. Mag.* **1996**, *15*, 97–102. [\[CrossRef\]](#)
97. Abásolo, D.; Escudero, J.; Hornero, R.; Gómez, C.; Espino, P. Approximate entropy and auto mutual information analysis of the electroencephalogram in Alzheimer’s disease patients. *Med. Biol. Eng. Comput.* **2008**, *46*, 1019–1028. [\[CrossRef\]](#)
98. Daw, C.S.; Finney, C.E.A.; Kennel, M.B. Symbolic approach for measuring temporal “irreversibility”. *Phys. Rev. E* **2000**, *62*, 1912–1921. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Eckmann, J.-P.; Kamphorst, S.O.; Ruelle, D. Recurrence Plots of Dynamical Systems. *Europhys. Lett. (EPL)* **1987**, *4*, 973–977. [\[CrossRef\]](#)
100. Webber, C.L.; Zbilut, J.P. Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **1994**, *76*, 965–973. [\[CrossRef\]](#)
101. Zbilut, J.P.; Giuliani, A.; Webber, C.L. Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Phys. Lett. A* **1998**, *246*, 122–128. [\[CrossRef\]](#)
102. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [\[CrossRef\]](#)
103. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [\[CrossRef\]](#)
104. Rényi, A. *Probability Theory*; North-Holland: Amsterdam, The Netherlands, 1970.
105. Tsallis, C.; Mendes, R.; Plastino, A. The role of constraints within generalized nonextensive statistics. *Phys. A Stat. Mech. Its Appl.* **1998**, *261*, 534–554. [\[CrossRef\]](#)
106. Fell, J.; Röschke, J.; Mann, K.; Schäffner, C. Discrimination of sleep stages: A comparison between spectral and nonlinear EEG measures. *Electroencephalogr. Clin. Neurophysiol.* **1996**, *98*, 401–410. [\[CrossRef\]](#)
107. Kannathal, N.; Choo, M.L.; Acharya, U.R.; Sadasivan, P. Entropies for detection of epilepsy in EEG. *Comput. Methods Programs Biomed.* **2005**, *80*, 187–194. [\[CrossRef\]](#)
108. Viertio-Oja, H.; Maja, V.; Sarkela, M.; Talja, P.; Tenkanen, N.; Tolvanenlaakso, H.; Paloheimo, M.P.J.; Vakkuri, A.P.; Ylihankala, A.M.; Meriläinen, P. Description of the Entropytm algorithm as applied in the Datex-Ohmeda S/5tm Entropy Module. *Acta Anaesthesiol. Scand.* **2004**, *48*, 154–161. [\[CrossRef\]](#)
109. Särkelä, M.O.K.; Ermes, M.J.; van Gils, M.J.; Yli-Hankala, A.M.; Jäntti, V.H.; Vakkuri, A.P. Quantification of Epileptiform Electroencephalographic Activity during Sevoflurane Mask Induction. *Anesthesiology* **2007**, *107*, 928–938. [\[CrossRef\]](#)

110. Zhang, C.; Wang, H.; Fu, R. Automated Detection of Driver Fatigue Based on Entropy and Complexity Measures. *IEEE Trans. Intell. Transp. Syst.* **2013**, *15*, 168–177. [\[CrossRef\]](#)
111. Eliang, Z.; Ewang, Y.; Esun, X.; Eli, D.; Voss, L.J.; Sleight, J.W.; Ehagihira, S.; Eli, X. EEG entropy measures in anesthesia. *Front. Comput. Neurosci.* **2015**, *9*, 16. [\[CrossRef\]](#)
112. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [\[CrossRef\]](#)
113. Li, X.; Li, D.; Liang, Z.; Voss, L.J.; Sleight, J.W. Analysis of depth of anesthesia with Hilbert–Huang spectral entropy. *Clin. Neurophysiol.* **2008**, *119*, 2465–2475. [\[CrossRef\]](#) [\[PubMed\]](#)
114. Aydın, S.; Saraoğlu, H.M.; Kara, S. Log Energy Entropy-Based EEG Classification with Multilayer Neural Networks in Seizure. *Ann. Biomed. Eng.* **2009**, *37*, 2626–2630. [\[CrossRef\]](#)
115. Torres, M.E.; Gamero, L.G.; Flandrin, P.; Abry, P. Multiresolution entropy measure. In *Wavelet Applications in Signal and Image Processing V*; Aldroubi, A., Laine, A.F., Unser, M.A., Eds.; Optical Science, Engineering and Instrumentation: San Diego, CA, USA, 1997; Volume 3169, pp. 400–407. [\[CrossRef\]](#)
116. Schuster, H.G.; Just, W. *Deterministic Chaos*; Wiley: Hoboken, NJ, USA, 2005. [\[CrossRef\]](#)
117. Wales, D.J. Calculating the rate of loss of information from chaotic time series by forecasting. *Nat. Cell Biol.* **1991**, *350*, 485–488. [\[CrossRef\]](#)
118. Schouten, J.C.; Takens, F.; Bleek, C.M.V.D. Maximum-likelihood estimation of the entropy of an attractor. *Phys. Rev. E* **1994**, *49*, 126–129. [\[CrossRef\]](#)
119. Paluš, M. Coarse-grained entropy rates for characterization of complex time series. *Phys. D Nonlinear Phenom.* **1996**, *93*, 64–77. [\[CrossRef\]](#)
120. Santamaria, I.; Pokharell, P.; Principe, J. Generalized correlation function: Definition, properties, and application to blind equalization. *IEEE Trans. Signal Process.* **2006**, *54*, 2187–2197. [\[CrossRef\]](#)
121. Pincus, S.M.; Gladstone, I.M.; Ehrenkranz, R.A. A regularity statistic for medical data analysis. *J. Clin. Monit.* **1991**, *7*, 335–345. [\[CrossRef\]](#) [\[PubMed\]](#)
122. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Circ. Physiol.* **2000**, *278*, H2039–H2049. [\[CrossRef\]](#) [\[PubMed\]](#)
123. García-Martínez, B.; Martínez-Rodrigo, A.; Cantabrana, R.Z.; García, J.M.P.; Alcaraz, R. Application of Entropy-Based Metrics to Identify Emotional Distress from Electroencephalographic Recordings. *Entropy* **2016**, *18*, 221. [\[CrossRef\]](#)
124. Courtiol, J.; Perdikis, D.; Petkoski, S.; Müller, V.; Huys, R.; Sleimen-Malkoun, R.; Jirsa, V.K. The multiscale entropy: Guidelines for use and interpretation in brain signal analysis. *J. Neurosci. Methods* **2016**, *273*, 175–190. [\[CrossRef\]](#)
125. Wu, S.-D.; Wu, C.-W.; Lee, K.-Y.; Lin, S.-G. Modified multiscale entropy for short-term time series analysis. *Phys. A Stat. Mech. Its Appl.* **2013**, *392*, 5865–5873. [\[CrossRef\]](#)
126. Wu, S.-D.; Wu, C.-W.; Lin, S.-G.; Wang, C.-C.; Lee, K.-Y. Time Series Analysis Using Composite Multiscale Entropy. *Entropy* **2013**, *15*, 1069–1084. [\[CrossRef\]](#)
127. Cuesta-Frau, D.; Murillo-Escobar, J.P.; Orrego, D.A.; Delgado-Trejos, E. Embedded Dimension and Time Series Length. Practical Influence on Permutation Entropy and Its Applications. *Entropy* **2019**, *21*, 385. [\[CrossRef\]](#) [\[PubMed\]](#)
128. Zhao, X.; Shang, P.; Huang, J. Permutation complexity and dependence measures of time series. *EPL (Europhys. Lett.)* **2013**, *102*, 40005. [\[CrossRef\]](#)
129. Mammone, N.; Duun-Henriksen, J.; Kjaer, T.W.; Morabito, F.C. Differentiating Interictal and Ictal States in Childhood Absence Epilepsy through Permutation Rényi Entropy. *Entropy* **2015**, *17*, 4627–4643. [\[CrossRef\]](#)
130. He, S.; Sun, K.; Wang, H. Multivariate permutation entropy and its application for complexity analysis of chaotic systems. *Phys. A Stat. Mech. Its Appl.* **2016**, *461*, 812–823. [\[CrossRef\]](#)
131. Rostaghi, M.; Azami, H. Dispersion Entropy: A Measure for Time-Series Analysis. *IEEE Signal Process. Lett.* **2016**, *23*, 610–614. [\[CrossRef\]](#)
132. Azami, H.; Escudero, J. Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation. *Comput. Methods Programs Biomed.* **2016**, *128*, 40–51. [\[CrossRef\]](#)
133. Manis, G.; Aktaruzzaman; Sassi, R. Bubble Entropy: An Entropy Almost Free of Parameters. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2711–2718. [\[CrossRef\]](#)
134. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley: Hoboken, NJ, USA, 2005.
135. Chen, W.; Wang, Z.; Xie, H.; Yu, W. Characterization of Surface EMG Signal Based on Fuzzy Entropy. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2007**, *15*, 266–272. [\[CrossRef\]](#)
136. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [\[CrossRef\]](#)
137. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [\[CrossRef\]](#)
138. Randolph, P.H. Spectral Analysis and Its Applications. *Technometrics* **1970**, *12*, 174. [\[CrossRef\]](#)
139. Kuś, R.; Kamiński, M.; Blinowska, K.J. Determination of EEG Activity Propagation: Pair-Wise Versus Multichannel Estimate. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1501–1510. [\[CrossRef\]](#)
140. Mormann, F.; Lehnertz, K.; David, P.; Elger, C.E. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Phys. D Nonlinear Phenom.* **2000**, *144*, 358–369. [\[CrossRef\]](#)

141. Lachaux, J.-P.; Rodriguez, E.; Martinerie, J.; Varela, F.J. Measuring phase synchrony in brain signals. *Human Brain Mapp.* **1999**, *8*, 194–208. [\[CrossRef\]](#)
142. Nunez, P.L.; Srinivasan, R.; Westdorp, A.F.; Wijesinghe, R.S.; Tucker, D.M.; Silberstein, R.B.; Cadusch, P.J. EEG coherency. *Electroencephalogr. Clin. Neurophysiol.* **1997**, *103*, 499–515. [\[CrossRef\]](#)
143. Nolte, G.; Bai, O.; Wheaton, L.; Mari, Z.; Vorbach, S.; Hallett, M. Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin. Neurophysiol.* **2004**, *115*, 2292–2307. [\[CrossRef\]](#) [\[PubMed\]](#)
144. Stam, C.J.; Nolte, G.; Daffertshofer, A. Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum. Brain Mapp.* **2007**, *28*, 1178–1193. [\[CrossRef\]](#) [\[PubMed\]](#)
145. Vinck, M.; Oostenveld, R.; van Wingerden, M.; Battaglia, F.; Pennartz, C.M.A. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage* **2011**, *55*, 1548–1565. [\[CrossRef\]](#)
146. Vinck, M.; van Wingerden, M.; Womelsdorf, T.; Fries, P.; Pennartz, C. The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization. *NeuroImage* **2010**, *51*, 112–122. [\[CrossRef\]](#)
147. Rulkov, N.F.; Sushchik, M.M.; Tsimring, L.S.; Abarbanel, H.D.I. Generalized synchronization of chaos in directionally coupled chaotic systems. *Phys. Rev. E* **1995**, *51*, 980–994. [\[CrossRef\]](#)
148. Arnhold, J.; Grassberger, P.; Lehnertz, K.; Elger, C.E. A robust method for detecting interdependences: Application to intracranially recorded EEG. *Phys. D Nonlinear Phenom.* **1999**, *134*, 419–430. [\[CrossRef\]](#)
149. Stam, C.; Van Dijk, B. Synchronization likelihood: An unbiased measure of generalized synchronization in multivariate data sets. *Phys. D Nonlinear Phenom.* **2002**, *163*, 236–251. [\[CrossRef\]](#)
150. Pereda, E.; Quiroga, R.Q.; Bhattacharya, J. Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobiol.* **2005**, *77*, 1–37. [\[CrossRef\]](#) [\[PubMed\]](#)
151. Young, J.; Dragoi, V.; Aazhang, B. Precise measurement of correlations between frequency coupling and visual task performance. *Sci. Rep.* **2020**, *10*, 1–14. [\[CrossRef\]](#) [\[PubMed\]](#)
152. Zoldi, S.M.; Greenside, H.S. Karhunen-Loève Decomposition of Extensive Chaos. *Phys. Rev. Lett.* **1997**, *78*, 1687–1690. [\[CrossRef\]](#)
153. Granger, C.W.J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **1969**, *37*, 424–438. [\[CrossRef\]](#)
154. Geweke, J. Measurement of Linear Dependence and Feedback Between Multiple Time Series. *J. Am. Stat. Assoc.* **1982**, *77*, 304. [\[CrossRef\]](#)
155. Nolte, G.; Ziehe, A.; Nikulin, V.V.; Schlögl, A.; Krämer, N.; Brismar, T.; Müller, K.-R. Robustly Estimating the Flow Direction of Information in Complex Physical Systems. *Phys. Rev. Lett.* **2008**, *100*, 234101. [\[CrossRef\]](#)
156. Stam, C.J. Modern network science of neurological disorders. *Nat. Rev. Neurosci.* **2014**, *15*, 683–695. [\[CrossRef\]](#)
157. Zou, Y.; Donner, R.V.; Marwan, N.; Donges, J.F.; Kurths, J. Complex network approaches to nonlinear time series analysis. *Phys. Rep.* **2019**, *787*, 1–97. [\[CrossRef\]](#)
158. Rubinov, M.; Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **2010**, *52*, 1059–1069. [\[CrossRef\]](#)
159. Stam, C.J.; van Straaten, E.C.W. The organization of physiological brain networks. *Clin. Neurophysiol.* **2012**, *123*, 1067–1087. [\[CrossRef\]](#) [\[PubMed\]](#)
160. Watts, D.J.; Strogatz, S. Collective dynamics of ‘small-world’ networks. *Nat. Cell Biol.* **1998**, *393*, 440–442. [\[CrossRef\]](#)
161. Newman, M.E.J. The Structure and Function of Complex Networks. *SIAM Rev.* **2003**, *45*, 167–256. [\[CrossRef\]](#)
162. Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; Alon, U. Network Motifs: Simple Building Blocks of Complex Networks. *Science* **2002**, *298*, 824–827. [\[CrossRef\]](#)
163. Humphries, M.D.; Gurney, K. Network ‘Small-World-Ness’: A Quantitative Method for Determining Canonical Network Equivalence. *PLoS ONE* **2008**, *3*, e0002051. [\[CrossRef\]](#)
164. Fallani, F.D.V.; Astolfi, L.; Cincotti, F.; Mattia, D.; Tocci, A.; Salinari, S.; Marciani, M.; Witte, H.; Colosimo, A.; Babiloni, F. Brain Network Analysis From High-Resolution EEG Recordings by the Application of Theoretical Graph Indexes. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2008**, *16*, 442–452. [\[CrossRef\]](#) [\[PubMed\]](#)
165. Latora, V.; Marchiori, M. Efficient Behavior of Small-World Networks. *Phys. Rev. Lett.* **2001**, *87*, 198701. [\[CrossRef\]](#) [\[PubMed\]](#)
166. Leicht, E.; Newman, M.E.J. Community Structure in Directed Networks. *Phys. Rev. Lett.* **2008**, *100*, 118703. [\[CrossRef\]](#)
167. Newman, M. *Networks*; Oxford University Press: Oxford, UK, 2010.
168. Sabidussi, G. The centrality index of a graph. *Psychometrika* **1966**, *31*, 581–603. [\[CrossRef\]](#) [\[PubMed\]](#)
169. Batool, K.; Niazi, M.A. Towards a Methodology for Validation of Centrality Measures in Complex Networks. *PLoS ONE* **2014**, *9*, e90283. [\[CrossRef\]](#) [\[PubMed\]](#)
170. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239. [\[CrossRef\]](#)
171. Iakovidou, N.D. Graph Theory at the Service of Electroencephalograms. *Brain Connect.* **2017**, *7*, 137–151. [\[CrossRef\]](#) [\[PubMed\]](#)
172. Heuvel, M.V.D.; Sporns, O. Rich-Club Organization of the Human Connectome. *J. Neurosci.* **2011**, *31*, 15775–15786. [\[CrossRef\]](#)
173. Balam, V.P.; Sameer, V.U.; Chinara, S. Automated classification system for drowsiness detection using convolutional neural network and electroencephalogram. *IET Intell. Transp. Syst.* **2021**, *15*, 514–524. [\[CrossRef\]](#)
174. Chaabene, S.; Bouaziz, B.; Boudaya, A.; Hökelmann, A.; Ammar, A.; Chaari, L. Convolutional Neural Network for Drowsiness Detection Using EEG Signals. *Sensors* **2021**, *21*, 1734. [\[CrossRef\]](#)

175. Jiao, Y.; Deng, Y.; Luo, Y.; Lu, B.-L. Driver sleepiness detection from EEG and EOG signals using GAN and LSTM networks. *Neurocomputing* **2020**, *408*, 100–111. [\[CrossRef\]](#)
176. Zou, S.; Qiu, T.; Huang, P.; Bai, X.; Liu, C. Constructing Multi-scale Entropy Based on the Empirical Mode Decomposition(EMD) and its Application in Recognizing Driving Fatigue. *J. Neurosci. Methods* **2020**, *341*, 108691. [\[CrossRef\]](#)
177. Chaudhuri, A.; Routray, A. Driver Fatigue Detection Through Chaotic Entropy Analysis of Cortical Sources Obtained From Scalp EEG Signals. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 185–198. [\[CrossRef\]](#)
178. Budak, U.; Bajaj, V.; Akbulut, Y.; Atilla, O.; Sengur, A. An Effective Hybrid Model for EEG-Based Drowsiness Detection. *IEEE Sens. J.* **2019**, *19*, 7624–7631. [\[CrossRef\]](#)
179. Chen, J.; Wang, H.; Wang, Q.; Hua, C. Exploring the fatigue affecting electroencephalography based functional brain networks during real driving in young males. *Neuropsychologia* **2019**, *129*, 200–211. [\[CrossRef\]](#) [\[PubMed\]](#)
180. Mehreen, A.; Anwar, S.M.; Haseeb, M.; Majid, M.; Ullah, M.O. A Hybrid Scheme for Drowsiness Detection Using Wearable Sensors. *IEEE Sens. J.* **2019**, *19*, 5119–5126. [\[CrossRef\]](#)
181. Martensson, H.; Keelan, O.; Ahlstrom, C. Driver Sleepiness Classification Based on Physiological Data and Driving Performance From Real Road Driving. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 421–430. [\[CrossRef\]](#)
182. Barua, S.; Ahmed, M.U.; Ahlström, C.; Begum, S. Automatic driver sleepiness detection using EEG, EOG and contextual information. *Expert Syst. Appl.* **2019**, *115*, 121–135. [\[CrossRef\]](#)
183. Ogino, M.; Mitsukura, Y. Portable Drowsiness Detection through Use of a Prefrontal Single-Channel Electroencephalogram. *Sensors* **2018**, *18*, 4477. [\[CrossRef\]](#)
184. Chen, J.; Wang, H.; Hua, C.; Wang, Q.; Liu, C. Graph analysis of functional brain network topology using minimum spanning tree in driver drowsiness. *Cogn. Neurodyn.* **2018**, *12*, 569–581. [\[CrossRef\]](#)
185. Chen, J.; Wang, H.; Hua, C. Assessment of driver drowsiness using electroencephalogram signals based on multiple functional brain networks. *Int. J. Psychophysiol.* **2018**, *133*, 120–130. [\[CrossRef\]](#) [\[PubMed\]](#)
186. Chen, J.; Wang, H.; Hua, C. Electroencephalography based fatigue detection using a novel feature fusion and extreme learning machine. *Cogn. Syst. Res.* **2018**, *52*, 715–728. [\[CrossRef\]](#)
187. Hu, J.; Min, J. Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model. *Cogn. Neurodyn.* **2018**, *12*, 431–440. [\[CrossRef\]](#)
188. Dimitrakopoulos, G.N.; Kakkos, I.; Dai, Z.; Wang, H.; Sgarbas, K.; Thakor, N.; Bezerianos, A.; Sun, Y. Functional Connectivity Analysis of Mental Fatigue Reveals Different Network Topological Alterations Between Driving and Vigilance Tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 740–749. [\[CrossRef\]](#)
189. Hong, S.; Kwon, H.; Choi, S.H.; Park, K.S. Intelligent system for drowsiness recognition based on ear canal electroencephalography with photoplethysmography and electrocardiography. *Inf. Sci.* **2018**, *453*, 302–322. [\[CrossRef\]](#)
190. Li, G.; Chung, W.-Y. Combined EEG-Gyroscope-tDCS Brain Machine Interface System for Early Management of Driver Drowsiness. *IEEE Trans. Hum. Mach. Syst.* **2017**, *48*, 50–62. [\[CrossRef\]](#)
191. Min, J.; Wang, P.; Hu, J. Driver fatigue detection through multiple entropy fusion analysis in an EEG-based system. *PLoS ONE* **2017**, *12*, e0188756. [\[CrossRef\]](#)
192. Awais, M.; Badruddin, N.; Driberg, M. A Hybrid Approach to Detect Driver Drowsiness Utilizing Physiological Signals to Improve System Performance and Wearability. *Sensors* **2017**, *17*, 1991. [\[CrossRef\]](#)
193. Nguyen, T.; Ahn, S.; Jang, H.; Jun, S.C.; Kim, J.G. Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Sci. Rep.* **2017**, *7*, 43933. [\[CrossRef\]](#)
194. Hu, J. Automated Detection of Driver Fatigue Based on AdaBoost Classifier with EEG Signals. *Front. Comput. Neurosci.* **2017**, *11*, 72. [\[CrossRef\]](#) [\[PubMed\]](#)
195. Chai, R.; Ling, S.H.; San, P.P.; Naik, G.R.; Nguyen, T.N.; Tran, Y.; Craig, A.; Nguyen, H. Improving EEG-Based Driver Fatigue Classification Using Sparse-Deep Belief Networks. *Front. Neurosci.* **2017**, *11*, 103. [\[CrossRef\]](#) [\[PubMed\]](#)
196. Chai, R.; Naik, G.R.; Nguyen, T.N.; Ling, S.H.; Tran, Y.; Craig, A.; Nguyen, H. Driver Fatigue Classification With Independent Component by Entropy Rate Bound Minimization Analysis in an EEG-Based System. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 715–724. [\[CrossRef\]](#)
197. Mu, Z.; Hu, J.; Yin, J. Driving Fatigue Detecting Based on EEG Signals of Forehead Area. *Int. J. Pattern Recognit. Artif. Intell.* **2017**, *31*, 1750011. [\[CrossRef\]](#)
198. Fu, R.; Wang, H.; Zhao, W. Dynamic driver fatigue detection using hidden Markov model in real driving condition. *Expert Syst. Appl.* **2016**, *63*, 397–411. [\[CrossRef\]](#)
199. Ahn, S.; Nguyen, T.; Jang, H.; Kim, J.G.; Jun, S.C. Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data. *Front. Hum. Neurosci.* **2016**, *10*, 219. [\[CrossRef\]](#) [\[PubMed\]](#)
200. Huang, K.-C.; Huang, T.-Y.; Chuang, C.-H.; King, J.-T.; Wang, Y.-K.; Lin, C.-T.; Jung, T.-P. An EEG-Based Fatigue Detection and Mitigation System. *Int. J. Neural Syst.* **2016**, *26*, 1650018. [\[CrossRef\]](#) [\[PubMed\]](#)
201. Li, G.; Lee, B.-L.; Chung, W.-Y. Smartwatch-Based Wearable EEG System for Driver Drowsiness Detection. *IEEE Sens. J.* **2015**, *15*, 7169–7180. [\[CrossRef\]](#)
202. Chen, L.-L.; Zhao, Y.; Zhang, J.; Zou, J.-Z. Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Syst. Appl.* **2015**, *42*, 7344–7355. [\[CrossRef\]](#)

-
203. Sauvet, F.; Bougard, C.; Coroenne, M.; Lely, L.; Van Beers, P.; Elbaz, M.; Guillard, M.; Leger, D.; Chennaoui, M. In-Flight Automatic Detection of Vigilance States Using a Single EEG Channel. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 2840–2847. [[CrossRef](#)] [[PubMed](#)]
 204. Lee, B.G.; Lee, B.L.; Chung, W.-Y. Mobile Healthcare for Automatic Driving Sleep-Onset Detection Using Wavelet-Based EEG and Respiration Signals. *Sensors* **2014**, *14*, 17915–17936. [[CrossRef](#)] [[PubMed](#)]
 205. Correa, A.G.; Orosco, L.; Laciari, E. Automatic detection of drowsiness in EEG records based on multimodal analysis. *Med. Eng. Phys.* **2014**, *36*, 244–249. [[CrossRef](#)] [[PubMed](#)]
 206. Hu, S.; Zheng, G.; Peters, B. Driver fatigue detection from electroencephalogram spectrum after electrooculography artefact removal. *IET Intell. Transp. Syst.* **2013**, *7*, 105–113. [[CrossRef](#)]
 207. Picot, A.; Charbonnier, S.; Caplier, A. On-Line Detection of Drowsiness Using Brain and Visual Information. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2011**, *42*, 764–775. [[CrossRef](#)]
 208. Zhao, C.; Zheng, C.; Zhao, M.; Tu, Y.; Liu, J. Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic. *Expert Syst. Appl.* **2011**, *38*, 1859–1865. [[CrossRef](#)]
 209. Liu, J.; Zhang, C.; Zheng, C. EEG-based estimation of mental fatigue by using KPCA–HMM and complexity parameters. *Biomed. Signal Process. Control.* **2010**, *5*, 124–130. [[CrossRef](#)]
 210. Gillberg, M.; Kecklund, G.; Åkerstedt, T. Sleepiness and performance of professional drivers in a truck simulator—comparisons between day and night driving. *J. Sleep Res.* **1996**, *5*, 12–15. [[CrossRef](#)] [[PubMed](#)]
 211. Wierwille, W.W.; Ellsworth, L.A. Evaluation of driver drowsiness by trained raters. *Accid. Anal. Prev.* **1994**, *26*, 571–581. [[CrossRef](#)]
 212. Kamrud, A.; Borghetti, B.; Kabban, C.S. The Effects of Individual Differences, Non-Stationarity, and the Importance of Data Partitioning Decisions for Training and Testing of EEG Cross-Participant Models. *Sensors* **2021**, *21*, 3225. [[CrossRef](#)]

Publication 2

I. Stancin, N. Frid, M. Cifrek, and A. Jovic, “EEG Signal Multichannel Frequency-Domain Ratio Indices for Drowsiness Detection Based on Multicriteria Optimization,” *Sensors*, vol. 21, no. 20, p. 6932, Oct. 2021, doi: 10.3390/s21206932.

Article

EEG Signal Multichannel Frequency-Domain Ratio Indices for Drowsiness Detection Based on Multicriteria Optimization

Igor Stancin, Nikolina Frid, Mario Cifrek  and Alan Jovic * 

Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia; igor.stancin@fer.hr (I.S.); nikolina.frid@fer.hr (N.F.); mario.cifrek@fer.hr (M.C.)

* Correspondence: alan.jovic@fer.hr

Abstract: Drowsiness is a risk to human lives in many occupations and activities where full awareness is essential for the safe operation of systems and vehicles, such as driving a car or flying an airplane. Although it is one of the main causes of many road accidents, there is still no reliable definition of drowsiness or a system to reliably detect it. Many researchers have observed correlations between frequency-domain features of the EEG signal and drowsiness, such as an increase in the spectral power of the theta band or a decrease in the spectral power of the beta band. In addition, features calculated as ratio indices between these frequency-domain features show further improvements in detecting drowsiness compared to frequency-domain features alone. This work aims to develop novel multichannel ratio indices that take advantage of the diversity of frequency-domain features from different brain regions. In contrast to the state-of-the-art, we use an evolutionary metaheuristic algorithm to find the nearly optimal set of features and channels from which the indices are calculated. Our results show that drowsiness is best described by the powers in delta and alpha bands. Compared to seven existing single-channel ratio indices, our two novel six-channel indices show improvements in (1) statistically significant differences observed between wakefulness and drowsiness segments, (2) precision of drowsiness detection and classification accuracy of the XGBoost algorithm and (3) model performance by saving time and memory during classification. Our work suggests that a more precise definition of drowsiness is needed, and that accurate early detection of drowsiness should be based on multichannel frequency-domain features.

Keywords: drowsiness detection; EEG; frequency-domain features; multicriteria optimization; machine learning



Citation: Stancin, I.; Frid, N.; Cifrek, M.; Jovic, A. EEG Signal Multichannel Frequency-Domain Ratio Indices for Drowsiness Detection Based on Multicriteria Optimization. *Sensors* **2021**, *21*, 6932. <https://doi.org/10.3390/s21206932>

Academic Editors: Maysam Abbod and Jiann-Shing Shieh

Received: 30 August 2021

Accepted: 17 October 2021

Published: 19 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Drowsiness is the intermediate state between wakefulness and sleep [1]. Terms such as sleepiness or tiredness are used synonymously with drowsiness in related studies [2–4]. Although it is intuitively clear what drowsiness is, it is not so easy to determine exactly whether a person is in a drowsy state or not. The reason for this is the unclear definition of drowsiness. Some researchers define drowsiness as stage 1 sleep (S1) [5–9], which is also known as non-rapid eye movement 1 (NREM 1) sleep. Da Silveira et al. [10] used S1 sleep stage data in their research of drowsiness. Johns [11] claims that the S1 sleep stage is equivalent to microsleep (episodes of psychomotor insensitivity due to sleep-related wakefulness loss [12]), while drowsiness is stated to occur before S1 sleep, but it is not stated when it begins and what characterizes it. Researchers who do not use any of the aforementioned definitions of drowsiness typically use a subjective assessment of drowsiness, e.g., the Karolinska sleepiness scale [13]. In this paper, the term drowsiness is used as a synonym for the S1 sleep stage.

In a drowsy state, people are not able to function at the level required to safely perform an activity [14], due to the progressive loss of cortical processing efficiency [15]. Drowsiness is, therefore, a significant risk factor for human lives in many occupations, e.g., for air traffic

controllers, pilots and regular car drivers [16]. According to the reports from NASA [17] and the National Transportation Safety Board [18], one of the main factors in road and air accidents is drowsiness. Gonçalves et al. [19] conducted a study across 19 European countries and concluded that in the last two years, 17% of drivers fell asleep while driving, while 7% of them had an accident due to drowsiness. The high frequency and prevalence of drowsiness-related accidents speak in favor of the development of early drowsiness detection systems, which is the subject of this paper.

Many researchers are trying to solve the problem of early detection of drowsiness in drivers. Balandong et al. [20], in their recent review, divided the techniques for detecting driver drowsiness into six categories: (1) subjective measures, (2) vehicle-based systems, (3) driver's behavior-based systems, (4) mathematical models of sleep–wake dynamics, (5) human physiological signal-based systems and (6) hybrids of one or more of these techniques. Currently, the most common techniques used in practice are vehicle-based systems [5], but these systems are mostly unreliable and depend largely on the driver's motivation to drive as well as possible [20].

Physiological signals are the promising alternative for reliable drowsiness detection [21]. The main problem with this approach is that these systems are often not easy to use and are intrusive to drivers [20]. Nevertheless, many researchers are working on small, automated and wearable devices [21–24], or on steering wheel devices [25,26] in order to overcome these obstacles. Techniques for detecting drowsiness based on physiological signals can be further subdivided according to the type of signal used, such as electroencephalogram (EEG) [27], electrooculogram (EOG) [28] or electrocardiogram (ECG) [29].

The most studied and applied physiological signal to detect drowsiness is the EEG. In this paper, frequency-domain features of the EEG signal are analyzed and two novel multichannel ratio indices for the detection of drowsiness are proposed. Besides the frequency-domain features, there are also other types of features: (1) nonlinear features [30], (2) spatiotemporal (functional connectivity) features [31] and (3) entropies [32]. These three groups of features have a lower frequency of use compared to the frequency-domain features, so in this paper, we focus only on frequency-domain features. Based on the recent review [33] of EEG-based drowsiness detection systems, 61% of the included papers used frequency-domain features, 38% used entropies, 10% used nonlinear features and 10% used spatiotemporal features (some papers used multiple groups of features, so the sum of the percentages is greater than 100%). This shows the difference in the use of drowsiness detection systems, and the difference is even greater in the general field of neurophysiological scientific papers. Although the three feature groups mentioned above are used less frequently, there are still a certain number of papers that include them, especially entropies.

Frequency-domain features estimate the power spectral density in a given frequency band. The bands typically used in the analysis of EEG signals are delta (δ , 0.5–4 Hz), theta (θ , 4–8 Hz), alpha (α , 8–12 Hz), beta (β , 12–30 Hz) and gamma (γ , >30 Hz). An increase in theta activity [34] and an increase in alpha activity [35] indicate drowsiness. An increase in the beta activity, however, is a sign of wakefulness and alertness [36]. There are several widely used frequency-domain ratio indices for detecting drowsiness. Eoh et al. [36] proposed the θ/α and β/α ratio indices, Jap et al. [37] proposed the $(\theta + \alpha)/\beta$, θ/β and $(\theta + \alpha)/(\alpha + \beta)$ ratio indices and da Silva et al. [10] proposed the γ/δ and $(\gamma + \beta)/(\delta + \alpha)$ ratio indices. These ratio indices provide improvement in the detection of drowsiness compared to the frequency-domain features alone and are shown to correlate with drowsiness.

All these frequency-domain features and ratio indices are calculated from a single EEG channel, i.e., from a single brain region. In recent research, Wang et al. [38] showed that the significance of a decrease in delta and an increase in $(\theta + \alpha)/\beta$ indices depends on the brain region. This significant diversity of the correlation of features with drowsiness in different brain regions is the motivation for this research. Since all currently used frequency-domain features and ratio indices are based on a single channel (single brain region), this work aims

to use the best distinguishing features of each brain region for the detection of drowsiness and to combine them into a single multichannel ratio index feature.

In our work, we use a computational method based on multicriteria optimization to extract the multichannel EEG-based frequency-domain ratio index features. This method allows us to discover new multichannel ratio indices that show improvements in the detection of drowsiness compared to single-channel ratio indices. Finally, with the use of machine learning models, we prove that multichannel indices detect drowsiness with higher accuracy, higher precision, reduced memory and faster computation compared to single-channel features.

In the Materials and Methods Section, we show the methodology of our work, including a description of the dataset, preprocessing and feature extraction methods used. Novel multichannel ratio indices and the multi-objective optimization method are also described there. In the Results Section, we present the results of our work, including statistical analysis, drowsiness prediction and computational properties of the proposed indices. In the Discussion Section, we discuss in more detail the topics covered in this paper. Finally, in the last section, we conclude the paper.

2. Materials and Methods

2.1. Dataset, Preprocessing and Feature Extraction

The data used in this paper were obtained from the PhysioNet portal [39], in particular from the 2018 PhysioNet computing in cardiology challenge [40]. The original dataset contains data records from 1985 subjects, and each recording includes a six-channel EEG, an electrooculogram, an electromyogram, a respiration signal from the abdomen and chest, airflow and oxygen saturation signals and a single-channel electrocardiogram during the all-night sleep. The records were divided into training and test sets of equal size. The sleep stages [41] of all subjects were annotated by clinical staff based on the American Academy of Sleep Medicine (AASM) manual for the scoring of sleep [42]. There are six types of annotations for different stages: wakefulness (W), stage 1 (S1), stage 2 (S2), stage 3 (S3), rapid eye movement (REM) and undefined.

In this research, we wanted to use a training set (992 subjects) to detect drowsiness. The officially provided way of acquiring the data is through torrent download, but we managed to download only 393 subjects completely, due to a lack of seeders. Of these 393 subjects, EEG signal recordings from 28 subjects were selected, based on the condition that each recording had at least 300 s of the W stage and, immediately after that, at least 300 s of the S1 stage. From each recording, a fragment of 600 s (300 s of W stage and 300 s of S1 stage) was used for analysis. In the original dataset, each EEG signal recording consists of six channels (F3, F4, C3, C4, O1 and O2, based on the International 10/20 System), with a sampling frequency of 200 Hz. Table 1 shows the identification numbers of all the selected subjects. The subjects were divided into two groups, one group used for training of the model (16 subjects) and the other one for the test of the obtained models (12 subjects). The training set was used to obtain novel ratio indices (with the method described below) and the test set was used to check these novel indices on the unseen data.

Table 1. The identification numbers of all the selected subjects. The training set is in the upper part and the test set is in the lower part of the table.

tr03-0092	tr03-0256	tr03-0876	tr03-1389
tr04-0649	tr04-0726	tr05-1434	Tr05-1675
tr07-0168	tr07-0458	tr07-0861	tr08-0021
tr08-0111	tr09-0175	tr10-0872	tr13-0204
tr04-0653	tr07-0127	tr09-0453	tr13-0170
tr05-0028	tr08-0157	tr12-0255	tr13-0508
tr05-0332	tr09-0328	tr12-0441	tr13-0653

Before feature extraction, the EEG signal must be filtered. For this purpose, the DC component was removed from the signal and the signal was filtered with a Butterworth filter to remove high-frequency artifacts and low-frequency drifts. We used the sixth-order Butterworth filter, the low-cut frequency of 1 Hz and the high-cut frequency of 40 Hz. In the selected fragments of the recordings, there was an insignificant number of eye-related artifacts, so we decided not to use the independent component analysis for their removal in order to prevent potential information loss due to component removal.

The signals were divided into epochs to calculate features. The epochs were five seconds long with a 50% overlap between them. Frequency-domain features are often used in EEG signal analysis. These features were extracted from the power spectral density (PSD) of the signal. To obtain the PSD of the signal, Welch's method [43] was used. Welch's method is used more often than Fast Fourier transform in the field of EEG signal analysis since it produces PSD with lower variance. The standard frequency-domain features were calculated, i.e., delta (δ , 0.5–4 Hz), theta (θ , 4–8 Hz), alpha (α , 8–12 Hz) and beta (β , 12–30 Hz) bands. We also calculated the less frequently used frequency-domain features, i.e., gamma (γ , >30 Hz), sigma (σ , 12–14 Hz), low alpha (α_1 , 8–10 Hz) and high alpha (α_2 , 10–12 Hz) bands [44].

2.2. Novel Multichannel Ratio Indices

Ratios between frequency-domain features have often been used as new features in different areas of EEG signal analysis [10,36]. All these features have a simple mathematical formulation but often lead to an improvement in detection and reduction of dimensionality for drowsiness. Moreover, they are calculated based on a single channel only. The idea behind the novel indices we present in this work is to design the feature formulation in such a way that frequency-domain features from different channels can be combined. Figure 1 illustrates the difference between these two approaches. For simplicity of visualization, only four epochs, two channels (F3 and F4) and three features per channel are shown in Figure 1.

	F3			F4		
	θ	α	β	θ	α	β
Epoch 1						
Epoch 2						
Epoch 3						
Epoch 4						

	F3			F4		
	θ	α	β	θ	α	β
Epoch 1						
Epoch 2						
Epoch 3						
Epoch 4						

Figure 1. A visualization of tables with features. The green color represents the possibilities for creating a ratio index, the first table (**top**) are the possibilities reported in the related work to create a single-channel ratio index, while the second table (**bottom**) are the possibilities explored in our novel multichannel approach.

We define a new index, I , for each epoch, e , which is calculated as a ratio of the feature values, $F(e)$, for all six channels in the epoch, e . In both the nominator and denominator,

the feature value of each channel, j , is multiplied with a dedicated coefficient, C_{ij} or K_{ij} respectively, as indicated in the Equation (1):

$$I(e) = \frac{\sum_{i=\text{features}} \sum_{j=\text{channels}} C_{ij} F_{ij}(e)}{\sum_{i=\text{features}} \sum_{j=\text{channels}} K_{ij} F_{ij}(e)} \quad (1)$$

The purpose of the coefficients is to reduce or even eliminate the influence of certain channels of frequency-domain features, by setting the value in the range $[0, 1)$, or increase the influence of certain channels of the frequency-domain features by setting the corresponding coefficient to a value in the range $[1, \infty)$. There are 48 (6 channels and 8 features per channel) C coefficients and 48 K coefficients.

The ideal output of $I(e)$ should look like a step function (or an inverse step function), which would indicate a clear difference between the two stages: W and S1. Figure 2 illustrates the main features of the output. The output can be divided into two parts: the left one corresponds to stage W and the right one to S1. While the output in each part should be as smooth as possible, i.e., with minimal oscillations, it is expected that there will be a transition period between the phases, which may have significant oscillations. This transition period would ideally be the step function, but in realistic settings, it is expected that the transition between phases of brain activity will probably last several epochs and would not be considered as either stage W or S1.

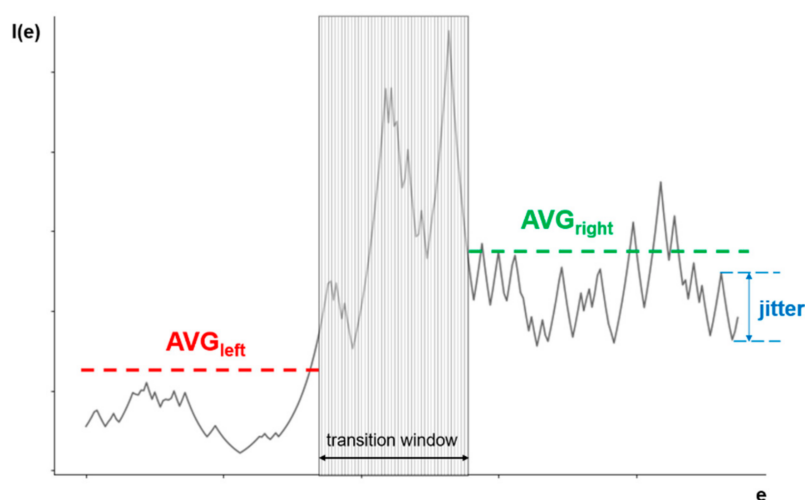


Figure 2. An illustration of all the elements needed for an evaluation of solutions of the multi-objective optimization in drowsiness detection.

In order to determine the appropriate value of the coefficients that would provide the output as close as possible to the ideal, at least two criteria must be taken into account: the absolute difference between the mean values left and right of the transition window and the quantification of the oscillations in each part. This can be defined as a multi-objective optimization problem that we want to solve using a metaheuristic multi-objective evolutionary optimization method, as described in the next section. To the best of our knowledge, this state transition problem has never been approached with evolutionary computation.

2.3. Multi-Objective Optimization

The optimization of a step function that is representative of the problem of flat surfaces is generally a challenge for any optimization algorithm because it does not provide information about which direction is favorable and an algorithm can get stuck on one of the flat plateaus [45]. To overcome this challenge, instead of optimizing the function according to one criterion, we define two objectives that we optimize simultaneously: (1) to maximize the absolute difference between the mean value of $I(e)$ output for the W

and S1 stages, and (2) to minimize the oscillations of the output value around the mean value in each stage. According to Figure 2, the left part of the $I(e)$ output occurring before the transition phase corresponds to the W stage, and the right part, occurring after the transition phase, corresponds to the S1 stage. Since optimization problems are usually expressed as minimization problems, where the first objective function, O1, is defined as the inverse absolute difference between the mean value of $I(e)$ of the left part (avg_{left}) and the right part (avg_{right}), Equation (2) is established:

$$O1 = \frac{1}{|avg_{right} - avg_{left}|} \quad (2)$$

The second objective function, O2, expresses the oscillations in the function and is defined as the number of times the difference between the output values of $I(e)$ for two adjacent epochs was greater than a given limit. The exact value of this limit will be discussed later in this section as it is closely related to the specifics of the optimization method used. The main goal of the objective function O2 is to minimize the influence of the biggest flaw in the way that the objective O1 is calculated, i.e., to use the averaging function. For example, if a possible solution is a completely straight line, except for a large negative spike in the left part and a large positive spike in the right part, based only on the objective function O1, this would be a good solution, while the objective function O2 would penalize this solution.

As mentioned above, the transition between two stages will probably take several epochs and show significant oscillations of the function output values. According to the annotation made by clinical personnel, the transition phase should be approximately in the middle of the $I(e)$ output, but it cannot be determined exactly how long it will last. In our work, which is based on expert knowledge of human behavior in the case of drowsiness, we assume that it lasts about one minute, which corresponds to about 30 epochs. Within the transition window, neither one of the two objective functions is calculated, since it is assumed to belong neither to the W nor to the S1 stages. We also allow it to move around the center, shifting left and right, due to a possible error of the human observer who marked the data.

The multi-objective optimization problem can now be expressed as $\min\{O1, O2\}$, where O1 and O2 are the conflicting objective functions, as defined above. The evolutionary meta-heuristic algorithm NSGA-II [46] was applied to solve this multi-objective optimization problem. The genetic algorithms (GAs) are normally used to solve complex optimization and search problems [47]. NSGA-II is one of the most popular evolutionary multicriteria optimization methods due to its versatility and ability to easily adapt to different types of optimization problems. The strong points of this MO algorithm are: (1) the fast non-dominated sorting ranking selection method used to emphasize Pareto-optimal solutions, (2) maintaining the population diversity by using the crowding distance and (3) the elitism approach, which ensures the preservation of best candidates through generations without the setting of any new parameters other than the normal genetic algorithm parameters, such as population size, termination parameter, crossover and mutation probabilities. Additionally, it was often used for the elimination of EEG channels with the similar purpose as in our case-dimensionality reduction [48]. This paper uses the implementation of NSGA-II provided by the MOEA framework [49] and is based on the guidelines defined in [46,50].

NSGA-II was used with the following configuration. The chromosome was divided into two parts: in the first part, genes represented the nominator coefficient values (C_{ij}), and in the second part, genes represented the denominator coefficient values (K_{ij}). In each part, the genes were grouped by frequency-domain features and channels, as illustrated in Figure 3. The genes were encoded as real values in the range [0.0, 10.0], and standard NSGA-II crossover and mutation operators were used to support operation on real values.

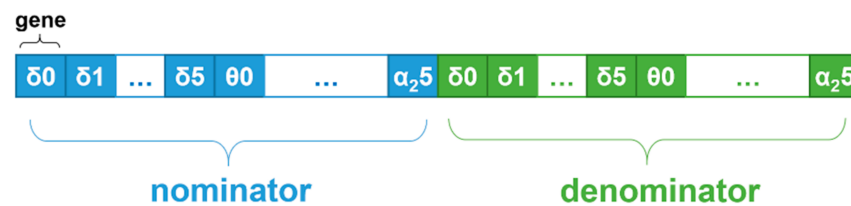


Figure 3. An illustration of a chromosome structure in the proposed optimization problem solution.

Each solution is evaluated based on the values of objectives O1 and O2, as described in the pseudocode in Algorithm 1. First, the chromosome is decoded (line 1). Then, for each test fragment, two values are calculated: (1) the inverse absolute difference (IAD) between the mean index value, $I(e)$, of the left part and the right part, represented by the `invAbsDiff` variable in the pseudocode, and (2) the oscillations in the function, represented by the `oscillation` variable in the pseudocode (lines 3–5). Finally, the value of each objective O1 for the given solution is defined as the average value of `invAbsDiff` for all test fragments, and the value of objective O2 is defined as the average value of `oscillation` for all test fragments (lines 7–8).

Algorithm 1. Evaluation.

```

1: decode chromosome to get coefficient values
2: for each fragment do
3:   indexVals[] = calculate index value for each epoch
4:   invAbsDiff += IADCalc(indexVals[], windowStart)
5:   oscillation += OscillationCalc(indexVals[], windowStart, winSize)
6: end for
7: objective1 = invAbsDiff/number_of_fragments
8: objective2 = oscillation/number_of_fragments

```

The algorithm for the IAD calculation is provided in the pseudocode in Algorithm 2. The calculation of the IAD for each fragment was slightly modified compared to Equation (1) to allow a faster convergence of the search algorithm. The transition phase was not in the same position in each fragment but allowed to move more loosely away from the center because the annotation in the original dataset was performed manually and there was a possibility of human error in case the observer would register a transition from W to S1 a little too early or too late. The algorithm allows the transition phase to begin no earlier than 30 epochs from the fragment start, and end no later than 60 epochs before the fragment end (line 2). The algorithm assumes the transition phase by looking for a window of 30 epochs which has the maximum difference of index, $I(e)$, values between the left and the right part (lines 9–13).

The gradation of the absolute difference between the mean value of the left and the right parts is also introduced (lines 19–22) to allow easier and faster convergence of the algorithm. The optimization of the objective O1 can be considered as an optimization problem with soft constraints that are related to how much O1 deviates from the optimal value. However, it is quite difficult to determine the optimal value precisely a priori. As indicated in [51,52], constraints are often treated with penalties in optimization techniques. The basic idea is to transform a constrained optimization problem into an unconstrained one by introducing a penalty into the original objective function to penalize violations of constraints. According to a comprehensive overview in [51], the penalty should be based on the degree of constraint violation of an individual. In [53], it is also recommended that instead of having just one fixed penalty coefficient, the penalty coefficient should increase when higher levels of constraint violation are reached. The greatest challenge, however, is to determine the exact penalty values. If the penalty is too high or too low, evolutionary algorithms spend either too much or too little time exploring the infeasible region, so it is

necessary to find the right trade-off between the objective function and the penalty function so that the search moves towards the optimum in the feasible space. As the authors have shown in [54], the choice of penalty boundaries is problem-dependent and difficult to generalize. Since we cannot strictly determine the optimal value of $O1$ in our case, we have chosen several thresholds for the absolute difference value, with the penalty increasing by a factor of 10 for each new threshold. The exact thresholds were selected based on the experience gained from the first few trial runs of the algorithm. Based on the observations from the trial runs, a third modification was also introduced: the difference is calculated with a relative, instead of absolute, value of $I(e)$. The relative value of $I(e)$ is calculated by using the lowest $I(e)$ value as a reference point, instead of zero, i.e., the zero is “moved”, as shown in code lines 16–18 in Algorithm 2.

Algorithm 2. IAD Calculation.

```

1: function IADCALC(indexVals[], windowStart)
2:   for j between 30 and (indexVals.size-60) do
3:     maxAbsDiff = 0
4:     left = 0
5:     right = 0
6:     avgLeft = average value of all Index values before j
7:     avgRight = average value of all Index values after j+30
8:     diff = ABS(avgRight-avgLeft)
9:     if diff  $\geq$  maxAbsDiff then
10:       maxAbsDiff = diff
11:       left = avgLeft
12:       right = avgRight
13:       windowStart = j
14:     end if
15:   end for
16:   lowestVal = GETLOWESTVAL(indexVals)
17:   movedZero = lowestVal-0.01*lowestVal
18:   absDiff = ABS(right-left)/MIN(left-movedZero, right-movedZero)
19:   if absDiff  $\geq$  5.0 then invAbsDiff = 1/absDiff
20:   else if absDiff  $\geq$  1.0 then invAbsDiff = 10/absDiff
21:   else if absDiff  $\geq$  0.5 then invAbsDiff = 100/absDiff
22:   else invAbsDiff = 1000
23:   end if
24:   return invAbsDiff
25: end function

```

The pseudocode for calculating the oscillations in the function as the second objective, $O2$, is provided in Algorithm 3. Again, the optimization of the oscillations can be considered a constrained optimization problem, so that, in the same way as in the case of the IAD calculation discussed previously, a gradation of the difference between the output values of $I(e)$ for two adjacent epochs is used to penalize the larger differences more severely (lines 7–10 and 15–18). The exact thresholds were chosen based on the experience gained from the first few trial runs of the algorithm. In order to make the algorithm converge more easily and quickly, the concept of “moved zero” was used again (lines 2, 3, 6 and 14).

Algorithm 3. Oscillation Calculation.

```

1: function OSCILLATIONCALC(indexVals[], windowStart, winSize)
2:   lowestVal = GETLOWESTVAL(indexVals)
3:   movedZero = lowestVal-0.01*lowestVal
4:   oscillation = 0
5:   for i between 1 and windowStart-1 do
6:     absDiff = ABS((indexVals[i]-indexVals[i-1])/(indexVals[i-1]-movedZero))
7:     if absDiff ≥ 5.0 then oscillation += 1000
8:     else if absDiff ≥ 1.0 then oscillation += 100
9:     else if absDiff ≥ 0.5 then oscillation += 10
10:    else if absDiff ≥ 0.25 then oscillation += 1
11:    end if
12:  end for
13:  for i between windowStart+winSize and indexVals.size()-1 do
14:    absDiff = ABS((indexVals[i]-indexVals[i-1])/(indexVals[i-1]-movedZero))
15:    if absDiff ≥ 5.0 then oscillation += 1000
16:    else if absDiff ≥ 1.0 then oscillation += 100
17:    else if absDiff ≥ 0.5 then oscillation += 10
18:    else if absDiff ≥ 0.25 then oscillation += 1
19:    end if
20:  end for
21:  return oscillation
22: end function

```

Finally, to further minimize the oscillations, and help the search algorithm converge more quickly, the maximum change in the $I(e)$ value between two adjacent epochs is set to 10% of the first of the two epochs. The mathematical formulation of this limit is provided in Equation (3):

$$Index(e) = \begin{cases} 1.1 * I(e-1), & \text{if } I(e) > 1.1 * I(e-1) \\ 0.9 * I(e-1), & \text{if } I(e) < 0.9 * I(e-1) \\ I(e), & \text{else} \end{cases} \quad (3)$$

3. Results

The optimization algorithm was executed over 107 generations, using 100 randomly selected chromosomes as a starting point. Ideally, the optimization algorithm would have many C and K coefficients equal to zero and only a few non-zero coefficients in order to obtain a simple and easily understandable mathematical formulation of a novel multichannel ratio index. Unfortunately, even the best solutions of the optimization algorithm had only up to 20 C and K coefficients equal to zero. Although such a novel multichannel ratio index showed good behavior in detecting drowsiness, it is impractical to use a formula with 76 coefficients. We consider anything above 15 coefficients to be impractical.

In order to reduce the number of coefficients and to simplify the formulation of the novel multichannel ratio index, some coefficients were manually set to zero. In order to decide which coefficients have the least influence on the final solution, we counted how often a large value of the coefficient is fixed to a certain frequency-domain feature. By analyzing the coefficients of all solutions in the final population of the optimization algorithm, we concluded that the most frequently selected features were δ , α , $\alpha1$ and $\alpha2$. After manually fixing the coefficients of all other frequency-domain features to zero, the search range for the optimization algorithm was reduced to half.

Although 48 C and K coefficients remained in the solution at that time, the algorithm provided equally good results in terms of drowsiness detection, but with a much simpler mathematical formulation. In addition to the 48 coefficients that were manually set to zero, the algorithm often set many more coefficients to zero. A decision on the best solution in the final population was made based on the $O1$ and $O2$ values of the optimization algorithm

in combination with the number of coefficients set to zero after using the floor operator on the coefficients. The floor operator was used to simplify the equation by removing the decimal numbers. Preferred solutions are those with a higher number of coefficients set to zero. Our choice was the solution with 13 non-zero coefficients, as shown in Equation (4):

$$I1(e) = \frac{\alpha_{F3} + 4\alpha_{O2} + 9\alpha_{1F3} + 3\alpha_{1C3} + 9\alpha_{1C4} + \alpha_{1O2} + 4\alpha_{2O1} + 8\alpha_{2O2}}{\delta_{F3} + 3\delta_{F4} + 3\delta_{C3} + 2\delta_{C4} + 9\delta_{O2}} \quad (4)$$

All C and K coefficients were rounded to a lower value (floor operator). Here, e represents the current epoch and all the features on the right side were from that same epoch.

The goal of the second condition of the optimization algorithm was to minimize the oscillations of the $I(e)$ function. The results were much better with this condition than without it, but the resulting function still oscillated strongly. In order to additionally minimize the oscillations, a limitation was performed. The maximum change between any two adjacent samples was set to 10% of the value of the first sample. Equation (5) shows the mathematical formulation of this limitation of the maximum change:

$$Index1(e) = \begin{cases} 1.1 * I1(e-1), & \text{if } I1(e) > 1.1 * I1(e-1) \\ 0.9 * I1(e-1), & \text{if } I1(e) < 0.9 * I1(e-1) \\ I1(e), & \text{else} \end{cases} \quad (5)$$

where $I1(e)$ is defined by Equation (4) and e is the current epoch. Limiting the maximum change of adjacent samples further improves the detection model, and therefore Equation (5) presents the first novel multichannel ratio index.

We have tried to further simplify the formulation of the multichannel ratio index. This time, brute force search for the best solution was applied with the following constraints: (1) encoding of all C and K coefficients was set to integer values of zero or one for the sake of simplicity, and (2) a maximum of five addends in the equation was allowed. With these constraints, we obtained Equation (6):

$$I2(e) = \frac{\delta_{F3} + \delta_{F4} + \delta_{O2}}{\alpha_{C3} + \alpha_{2O2}} \quad (6)$$

Again, similar to the first index, the maximum change was limited, so that the final equation for the second ratio index was obtained as:

$$Index2(e) = \begin{cases} 1.1 * I2(e-1), & \text{if } I2(e) > 1.1 * I2(e-1) \\ 0.9 * I2(e-1), & \text{if } I2(e) < 0.9 * I2(e-1) \\ I2(e), & \text{else} \end{cases} \quad (7)$$

where $I2(e)$ is defined by Equation (6) and e is the current epoch. After obtaining the two novel indices, they were normalized to the range $[0, 1]$ for each subject to eliminate interindividual differences between the subjects.

The two novel multichannel ratio indices defined by Equations (5) and (7) were compared with the seven existing indices θ/α and β/α [36], $(\theta + \alpha)/\beta$, θ/β and $(\theta + \alpha)/(\alpha + \beta)$ [37], and γ/δ and $(\gamma + \beta)/(\delta + \alpha)$ [10]. The indices γ/δ and $(\gamma + \beta)/(\delta + \alpha)$ were calculated based on the wavelet transform, i.e., in the same way as in the original paper. Figure 4 shows a comparison of our novel indices with the best and the worst channel for θ/α and $(\theta + \alpha)/\beta$ single-channel indices for subject tr08-0111. These two single-channel indices were selected because they are the best predictors of drowsiness for a given subject among all single-channel indices.

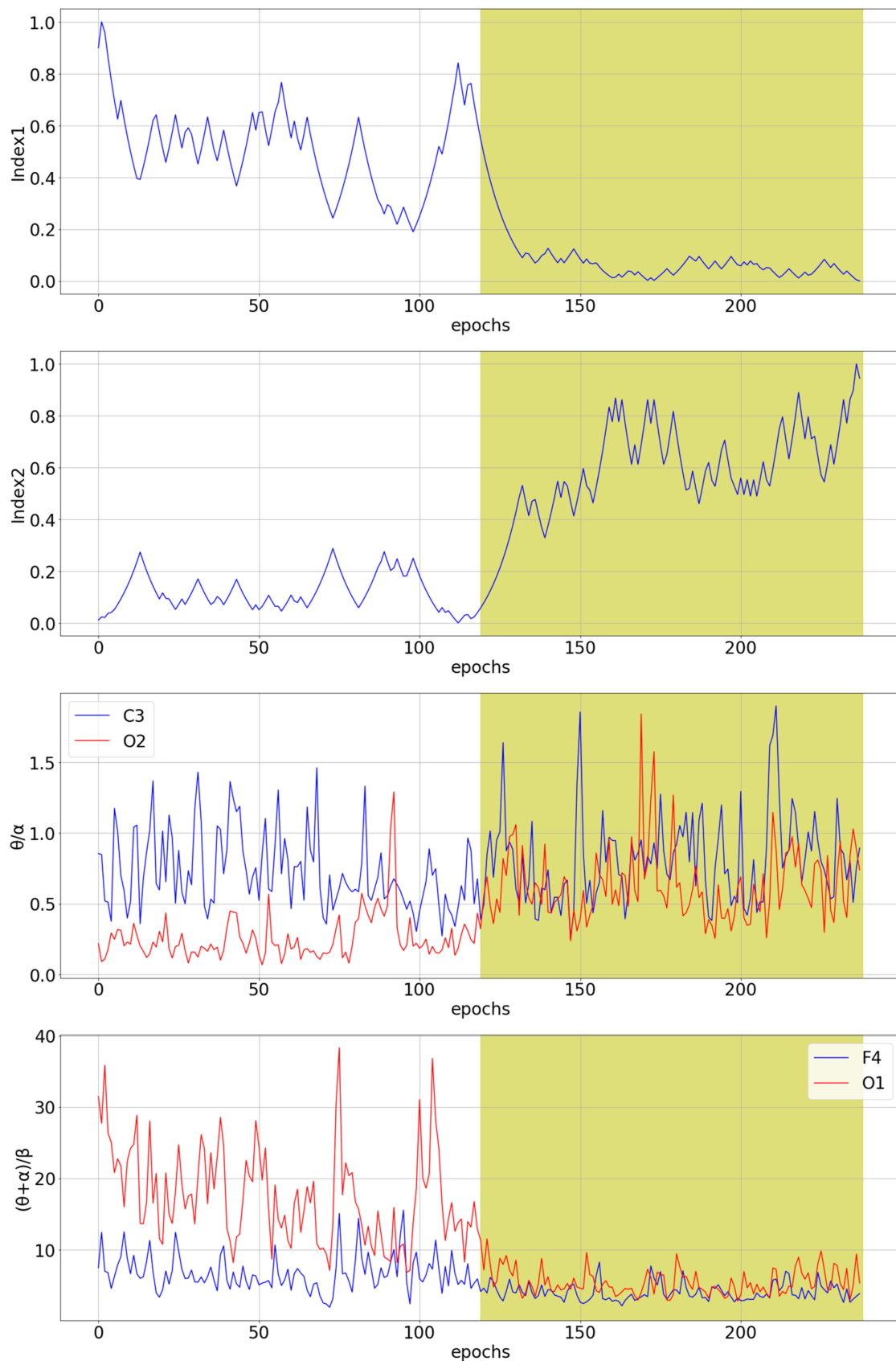


Figure 4. The comparison of the two novel multichannel indices with the best and the worst channel for θ/α and $(\theta + \alpha)/\beta$ single-channel indices for subject tr08-0111. The white part of the diagram represents an awake state, while the yellow part of the diagram represents stage 1 of sleep, i.e., a drowsiness state.

3.1. Statistical Analysis

The Wilcoxon signed-rank test [55] was used to analyze the statistical differences between the awake state and the S1 state. This test was chosen because it refers to data that do not necessarily follow the normal distribution. Table 2 shows p -values for each subject in the training set and each index. The significance level $\alpha_0 = 0.01$ was used together with the Bonferroni correction [55] to reduce the probability of false-positive results, as the test was repeated 144 times (16 subjects and 9 indices), giving us the final $\alpha_p = 6.9 \times 10^{-5}$. For the existing indices, the p -value was calculated for each channel, but only the p -values of the best channel (the lowest average of p -value for all subjects) are shown in Table 2.

The two novel indices show p -values lower than α_p for most subjects. From this, we can conclude that, for Index1, 14 of 16 subjects show two different distributions for the W stage and the S1 stage, while 13 of 16 subjects show significantly different distributions of the W stage and the S1 stage for Index2. There are only two existing indices where the p -value is lower than α_p in more than ten cases. These are θ/β and $(\theta + \alpha)/(\alpha + \beta)$, both by Jap et al. [37].

Table 3 shows p -values for each subject in the test set and each index. Again, the two novel indices, together with the $(\gamma + \beta)/(\delta + \alpha)$ [10] index, show p -values lower than α_p for most subjects.

3.2. Drowsiness Prediction Analysis

An additional comparison of ratio indices was performed by analyzing the drowsiness detection accuracy and precision, as obtained with the XGBoost algorithm [56]. Default parameters were applied: learning rate eta equal to 0.3, gamma equal to 0 and a maximum depth of a tree equal to 6. For a detailed comparison of the indices, classification accuracy and precision were calculated for each subject. Namely, each subject has 238 epochs of the measured signal, with the first half representing the W state and the second half the S1 state. The algorithm classified the subject's state for each epoch (238 classifications per subject), and the accuracy for each subject was calculated based on these classifications. The leave-one-subject-out cross-validation method was applied on the training set, i.e., the algorithm was trained on the data of 15 subjects and tested on the subject excluded from the training set, and this was repeated 16 times to evaluate drowsiness detection on each subject from the training set. Table 4 shows the classification accuracy achieved on the training set.

Table 2. Statistical significance p -values were obtained by the Wilcoxon signed-rank test for distinguishing the awake state from the S1 state. The shaded green cells with bold text represent the lowest p -value for each subject in the training set. At the bottom, the index having p -values lower than α_p for most subjects is marked in the same way.

Subject	Index1	Index2	θ/α	β/α	$(\theta + \alpha)/\beta$	θ/β	$(\theta + \alpha)/(\alpha + \beta)$	γ/δ	$(\gamma + \beta)/(\delta + \alpha)$
	This Work		Eoh et al. [36]			Jap et al. [37]	da Silveira et al. [10]		
tr03-0092	1.12×10^{-6}	4.82×10^{-3}	1.23×10^{-4}	7.22×10^{-1}	4.97×10^{-2}	1.62×10^{-08}	1.41×10^{-9}	1.06×10^{-6}	3.09×10^{-4}
tr03-0256	3.92×10^{-8}	1.43×10^{-13}	2.42×10^{-6}	6.88×10^{-4}	1.11×10^{-3}	4.04×10^{-2}	3.60×10^{-1}	9.08×10^{-3}	1.51×10^{-9}
tr03-0876	1.37×10^{-20}	3.87×10^{-17}	5.06×10^{-3}	1.63×10^{-7}	9.80×10^{-14}	2.94×10^{-6}	1.58×10^{-5}	1.09×10^{-1}	2.09×10^{-1}
tr03-1389	9.77×10^{-11}	1.35×10^{-8}	9.64×10^{-2}	2.06×10^{-1}	2.86×10^{-4}	1.27×10^{-1}	1.91×10^{-1}	1.82×10^{-1}	3.01×10^{-1}
tr04-0649	5.85×10^{-21}	4.11×10^{-21}	5.71×10^{-12}	2.38×10^{-9}	8.12×10^{-7}	1.89×10^{-1}	2.18×10^{-5}	3.33×10^{-7}	6.13×10^{-3}
tr04-0726	2.96×10^{-20}	3.19×10^{-20}	5.90×10^{-20}	2.31×10^{-16}	9.40×10^{-9}	2.78×10^{-14}	2.62×10^{-15}	2.36×10^{-19}	2.19×10^{-20}
tr05-1434	7.90×10^{-10}	9.79×10^{-13}	6.76×10^{-9}	3.70×10^{-10}	1.62×10^{-1}	3.96×10^{-17}	2.00×10^{-19}	5.42×10^{-17}	1.08×10^{-11}
tr05-1675	1.71×10^{-13}	1.85×10^{-11}	1.24×10^{-9}	7.82×10^{-3}	1.48×10^{-1}	1.10×10^{-14}	4.47×10^{-16}	1.58×10^{-2}	4.62×10^{-10}
tr07-0168	2.88×10^{-21}	5.15×10^{-21}	1.75×10^{-13}	4.73×10^{-11}	8.08×10^{-6}	1.05×10^{-8}	4.49×10^{-11}	8.87×10^{-16}	1.01×10^{-1}
tr07-0458	8.34×10^{-11}	1.77×10^{-16}	1.66×10^{-4}	1.77×10^{-4}	5.51×10^{-1}	1.32×10^{-2}	6.62×10^{-3}	3.68×10^{-4}	4.17×10^{-1}
tr07-0861	2.88×10^{-21}	3.11×10^{-21}	3.14×10^{-3}	3.55×10^{-7}	3.64×10^{-2}	9.96×10^{-8}	1.11×10^{-6}	1.49×10^{-17}	1.50×10^{-12}
tr08-0021	2.88×10^{-21}	2.88×10^{-21}	4.09×10^{-2}	2.54×10^{-9}	4.55×10^{-8}	3.34×10^{-10}	4.91×10^{-5}	4.19×10^{-13}	2.10×10^{-6}
tr08-0111	2.88×10^{-21}	2.88×10^{-21}	4.41×10^{-2}	7.54×10^{-5}	1.94×10^{-20}	2.04×10^{-20}	3.92×10^{-4}	4.50×10^{-15}	3.41×10^{-3}
tr09-0175	7.78×10^{-5}	7.92×10^{-2}	4.64×10^{-2}	3.10×10^{-4}	5.35×10^{-14}	2.23×10^{-5}	2.68×10^{-6}	7.18×10^{-2}	1.30×10^{-5}
tr × 10-0872	2.62×10^{-15}	1.96×10^{-14}	1.76×10^{-2}	3.89×10^{-2}	5.91×10^{-3}	5.09×10^{-6}	7.52×10^{-5}	2.14×10^{-5}	2.33×10^{-5}
tr13-0204	1.71×10^{-3}	6.62×10^{-1}	6.30×10^{-4}	4.91×10^{-5}	6.36×10^{-5}	2.91×10^{-10}	2.63×10^{-10}	5.59×10^{-1}	2.16×10^{-2}
No. subjects with $p < 6.9 \times 10^{-5}$	14	13	6	9	8	12	11	9	8

Table 3. Statistical significance p -values were obtained by the Wilcoxon signed-rank test for distinguishing the awake state from the S1 state. The shaded green cells with bold text represent the lowest p -value for each subject in the test set. At the bottom, the index having p -values lower than α_p for most subjects is marked in the same way.

Subject	Index1	Index2	θ/α	β/α	$(\theta + \alpha)/\beta$	θ/β	$(\theta + \alpha)/(\alpha + \beta)$	γ/δ	$(\gamma + \beta)/(\delta + \alpha)$
	This Work		Eoh et al. [36]		Jap et al. [37]		da Silveira et al. [10]		
tr04-0653	3.19×10^{-9}	2.71×10^{-9}	2.61×10^{-7}	1.80×10^{-5}	6.83×10^{-1}	3.63×10^{-9}	2.17×10^{-8}	2.69×10^{-6}	3.33×10^{-10}
tr05-0028	1.26×10^{-9}	2.36×10^{-7}	3.78×10^{-3}	2.99×10^{-1}	3.45×10^{-1}	3.56×10^{-1}	5.16×10^{-2}	1.70×10^{-1}	4.36×10^{-2}
tr05-0332	2.88×10^{-21}	2.88×10^{-21}	8.04×10^{-16}	2.10×10^{-5}	2.66×10^{-3}	7.39×10^{-14}	8.17×10^{-18}	1.84×10^{-16}	3.00×10^{-16}
tr07-0127	3.71×10^{-18}	1.27×10^{-15}	8.92×10^{-2}	1.44×10^{-2}	3.44×10^{-5}	1.38×10^{-2}	2.51×10^{-5}	9.07×10^{-16}	3.36×10^{-16}
tr08-0157	2.88×10^{-21}	2.88×10^{-21}	6.51×10^{-5}	1.26×10^{-7}	9.66×10^{-1}	6.67×10^{-3}	1.85×10^{-3}	2.82×10^{-11}	5.92×10^{-11}
tr09-0328	1.80×10^{-10}	1.14×10^{-2}	7.90×10^{-10}	9.56×10^{-7}	1.70×10^{-7}	2.01×10^{-1}	1.92×10^{-3}	3.54×10^{-2}	7.01×10^{-5}
tr09-0453	2.73×10^{-1}	7.80×10^{-4}	1.63×10^{-1}	2.96×10^{-1}	1.03×10^{-7}	3.89×10^{-2}	3.17×10^{-2}	8.30×10^{-16}	6.45×10^{-9}
tr12-0255	1.37×10^{-2}	5.55×10^{-10}	4.31×10^{-19}	2.17×10^{-18}	1.89×10^{-16}	2.20×10^{-19}	8.22×10^{-19}	1.11×10^{-7}	2.33×10^{-7}
tr12-0441	2.76×10^{-11}	9.26×10^{-9}	6.95×10^{-4}	6.89×10^{-1}	2.46×10^{-3}	3.63×10^{-13}	1.53×10^{-4}	2.13×10^{-3}	5.68×10^{-8}
tr13-0170	7.59×10^{-7}	3.60×10^{-5}	3.74×10^{-2}	4.73×10^{-2}	1.91×10^{-4}	9.71×10^{-4}	2.16×10^{-2}	6.07×10^{-17}	5.23×10^{-16}
tr13-0508	2.69×10^{-1}	7.61×10^{-2}	1.87×10^{-5}	1.99×10^{-2}	1.10×10^{-9}	9.17×10^{-2}	6.22×10^{-5}	7.26×10^{-5}	7.26×10^{-2}
tr13-0653	1.09×10^{-16}	7.36×10^{-20}	4.90×10^{-8}	4.16×10^{-4}	8.05×10^{-2}	1.66×10^{-2}	1.03×10^{-9}	3.47×10^{-2}	1.40×10^{-5}
No. subjects with $p < 6.9 \times 10^{-5}$	9	9	7	5	5	4	6	7	9

Table 4. The classification accuracy was obtained with the XGBoost algorithm for each subject in the training set. The shaded green cells with bold text show the highest accuracy obtained for each subject. At the bottom, the best mean accuracy for each ratio index is marked in the same way.

Subject	Index1	Index2	θ/α	β/α	$(\theta + \alpha)/\beta$	θ/β	$(\theta + \alpha)/(\alpha + \beta)$	γ/δ	$(\gamma + \beta)/(\delta + \alpha)$
	This Work		Eoh et al. [36]		Jap et al. [37]		da Silveira et al. [10]		
tr03-0092	0.5420	0.5252	0.6387	0.5168	0.5504	0.6092	0.6050	0.4684	0.5527
tr03-0256	0.5924	0.6303	0.5840	0.5504	0.5672	0.6345	0.6345	0.4473	0.4346
tr03-0876	0.6387	0.5672	0.6008	0.6218	0.4748	0.6471	0.6303	0.5781	0.5992
tr03-1389	0.3487	0.3908	0.4874	0.5588	0.5042	0.5378	0.4664	0.5696	0.5148
tr04-0649	0.6975	0.8025	0.5462	0.5462	0.5630	0.4832	0.5210	0.6118	0.5527
tr04-0726	0.7983	0.7605	0.6681	0.6176	0.5966	0.6134	0.6765	0.7637	0.7511
tr05-1434	0.3739	0.3697	0.4160	0.6218	0.7059	0.5882	0.6933	0.5485	0.5063
tr05-1675	0.6849	0.6513	0.6933	0.5546	0.6218	0.5630	0.7227	0.5781	0.5781
tr07-0168	0.7773	0.8193	0.6008	0.5504	0.5630	0.6092	0.6303	0.4473	0.5105
tr07-0458	0.3109	0.2689	0.5378	0.4748	0.5084	0.5378	0.5504	0.4684	0.4979
tr07-0861	0.6933	0.7269	0.5378	0.5420	0.5504	0.5168	0.5714	0.6540	0.6160
tr08-0021	0.7857	0.6387	0.5630	0.4874	0.4664	0.3866	0.4748	0.6329	0.6245
tr08-0111	0.6891	0.8025	0.7143	0.7101	0.7227	0.5462	0.4748	0.6287	0.6118
tr09-0175	0.6050	0.5252	0.5966	0.4748	0.5420	0.6218	0.6008	0.5401	0.6498
tr10-0872	0.6134	0.6008	0.5084	0.5168	0.4874	0.5252	0.4832	0.5274	0.4810
tr13-0204	0.5042	0.4538	0.6597	0.4790	0.6387	0.6471	0.6303	0.5570	0.5570
Average	0.6035	0.5959	0.5846	0.5515	0.5664	0.5667	0.5853	0.5638	0.5649

Index1 has the highest average accuracy and the highest classification accuracy for 3 of 16 subjects. Index2 has the second-highest average accuracy and the highest classification accuracy for 4 of 16 subjects, which is the most of all indices. θ/α [36] and $(\theta + \alpha)/(\alpha + \beta)$ [37] are the only other indices with an average classification accuracy above 0.58, while θ/α [36] and θ/β [37] are the only other indices with the highest accuracy for 3 of 16 subjects. The β/α [36] index has the lowest average classification accuracy on the training set (0.5515).

Table5 shows the classification accuracy on the test set. Index1 has the highest average accuracy and the highest classification accuracy for 3 of 12 subjects. Index2 has the third-highest average accuracy and the highest classification accuracy for 4 of 12 subjects, which is the most of all indices. The only other index with comparable accuracy is θ/α [36], with the second-highest average accuracy. All other indices have at least 2.5% lower accuracy than the two novel indices.

Table6 shows the degree of precision of drowsiness detection on the training set. Index2 has the highest average precision of drowsiness detection and the highest precision

of drowsiness detection for five subjects, which is the highest of all indices. Index1 has the second-best average precision of drowsiness detection. $(\theta + \alpha)/(\alpha + \beta)$ [37] and γ/δ [10] have a precision of drowsiness detection comparable to Index1 and Index2, while all other ratio indices have lower precision.

Table 5. The classification accuracy was obtained with the XGBoost algorithm for each subject in the test set. The shaded green cells with bold text show the highest accuracy obtained for each subject. At the bottom, the best mean accuracy for each ratio index is marked in the same way.

Subject	Index1	Index2	θ/α	β/α	$(\theta + \alpha)/\beta$	θ/β	$(\theta + \alpha)/(\alpha + \beta)$	γ/δ	$(\gamma + \beta)/(\delta + \alpha)$
	This Work		Eoh et al. [36]		Jap et al. [37]		da Silva et al. [10]		
tr04-0653	0.5672	0.6345	0.6303	0.5042	0.5168	0.5840	0.5630	0.5612	0.5738
tr05-0028	0.4454	0.4202	0.5294	0.5588	0.4664	0.4076	0.5462	0.5021	0.5443
tr05-0332	0.8067	0.8277	0.7563	0.5630	0.5294	0.6555	0.6134	0.5654	0.6118
tr07-0127	0.5462	0.6092	0.4916	0.5294	0.5252	0.5000	0.4118	0.4304	0.4051
tr08-0157	0.6303	0.6681	0.5294	0.5294	0.5000	0.5084	0.5294	0.5232	0.4810
tr09-0328	0.6050	0.5084	0.5966	0.5168	0.5000	0.5042	0.5672	0.5738	0.5401
tr09-0453	0.5588	0.5252	0.5714	0.5420	0.5294	0.5504	0.5840	0.5105	0.4557
tr12-0255	0.5420	0.5546	0.6639	0.5462	0.4748	0.5630	0.5924	0.6329	0.5654
tr12-0441	0.6891	0.5756	0.5840	0.5168	0.5378	0.5630	0.5798	0.6498	0.5274
tr13-0170	0.6008	0.5630	0.5924	0.5546	0.6261	0.6471	0.5336	0.5612	0.4430
tr13-0508	0.4538	0.5084	0.6555	0.6008	0.5588	0.6261	0.6092	0.5654	0.5443
tr13-0653	0.6807	0.6303	0.5210	0.5462	0.5630	0.5336	0.6008	0.5359	0.5063
Average	0.5938	0.5854	0.5935	0.5424	0.5273	0.5536	0.5609	0.5510	0.5165

Table 6. The precision of drowsiness detection was obtained with the XGBoost algorithm for each subject in the training set. The shaded green cells with bold text show the highest precision obtained for each subject. At the bottom, the best mean precision for each ratio index is marked in the same way.

Subject	Index1	Index2	θ/α	β/α	$(\theta + \alpha)/\beta$	θ/β	$(\theta + \alpha)/(\alpha + \beta)$	γ/δ	$(\gamma + \beta)/(\delta + \alpha)$
	This Work		Eoh et al. [36]		Jap et al. [37]		da Silva et al. [10]		
tr03-0092	0.5439	0.5439	0.5439	0.5439	0.5439	0.5439	0.5439	0.5439	0.5439
tr03-0256	0.6222	0.6348	0.5676	0.5349	0.5571	0.6127	0.6096	0.4270	0.3974
tr03-0876	0.6514	0.5800	0.6765	0.5973	0.4808	0.7397	0.7123	0.5804	0.6264
tr03-1389	0.2273	0.3774	0.4906	0.5455	0.5034	0.5391	0.4556	0.5519	0.5120
tr04-0649	0.7582	0.8273	0.5733	0.7895	0.7778	0.4878	0.5294	0.9063	0.6000
tr04-0726	0.8318	1.0000	0.7128	0.6373	0.6055	0.6627	0.7333	0.8370	0.7706
tr05-1434	0.4051	0.4083	0.1429	0.6028	0.7168	0.7692	0.7805	0.6571	0.5027
tr05-1675	0.6642	0.6011	0.6885	0.5607	0.6559	0.5862	0.7265	0.5425	0.5433
tr07-0168	0.7500	0.7923	0.5759	0.5375	0.5397	0.5730	0.5963	0.4488	0.5156
tr07-0458	0.2816	0.0492	0.5437	0.4789	0.5088	0.5446	0.5732	0.4500	0.4930
tr07-0861	0.6264	0.6688	0.5338	0.5342	0.5349	0.5105	0.5521	0.6216	0.5780
tr08-0021	0.7464	0.6854	0.6119	0.4717	0.4535	0.4000	0.4688	0.6325	0.6355
tr08-0111	0.6692	0.8214	0.7297	0.8205	0.7912	0.5314	0.4840	0.6500	0.5985
tr09-0175	0.6404	0.5263	0.5742	0.4762	0.5379	0.6142	0.6053	0.5273	0.6207
tr10-0872	0.6174	0.5909	0.5045	0.5130	0.4892	0.5254	0.4882	0.5349	0.4627
tr13-0204	0.5054	0.4545	0.6357	0.4820	0.6170	0.6636	0.6348	0.6032	0.5970
Average	0.5963	0.5976	0.5691	0.5704	0.5821	0.5815	0.5934	0.5946	0.5623

Table 7 shows the degree of precision of drowsiness detection achieved on the test set. Index1 has the highest average precision. θ/α [36] and γ/δ [10] have 1% lower precision than Index1, while all other indices have at least 4% lower precision. Index2 has the second-highest average precision.

Table 7. The precision of drowsiness detection was obtained with the XGBoost algorithm for each subject in the test set. The shaded green cells with bold text show the highest precision obtained for each subject. At the bottom, the best mean precision for each ratio index is marked in the same way.

Subject	Index1	Index2	θ/α	β/α	$(\theta + \alpha)/\beta$	θ/β	$(\theta + \alpha)/(\alpha + \beta)$	γ/δ	$(\gamma + \beta)/(\delta + \alpha)$
This Work			Eoh et al. [36]		Jap et al. [37]		da Silveira et al. [10]		
tr04-0653	0.5769	0.6569	0.6742	0.5030	0.5133	0.6000	0.5862	0.5795	0.5914
tr05-0028	0.3818	0.4296	0.5178	0.5385	0.4762	0.4214	0.5364	0.5000	0.5862
tr05-0332	0.7483	0.9333	0.7905	0.5547	0.5321	0.7846	0.6709	0.8571	0.8824
tr07-0127	0.5401	0.6512	0.4722	0.5294	0.5231	0.5000	0.3600	0.4309	0.4207
tr08-0157	0.5812	0.6087	0.5158	0.5574	0.5000	0.5048	0.5153	0.5122	0.4886
tr09-0328	0.6147	0.5078	0.5650	0.5137	0.5000	0.5041	0.5678	0.5620	0.5372
tr09-0453	0.5398	0.5176	0.5436	0.5301	0.5248	0.5306	0.5538	0.5049	0.4721
tr12-0255	0.5316	0.5351	0.6054	0.5342	0.4826	0.5393	0.5545	0.5886	0.5389
tr12-0441	0.8000	0.5789	0.5633	0.5093	0.5249	0.5478	0.5785	0.6496	0.5231
tr13-0170	0.6333	0.5466	0.5724	0.5355	0.6056	0.6296	0.5313	0.6944	0.3478
tr13-0508	0.4337	0.5091	0.6331	0.5674	0.5443	0.6154	0.6140	0.5478	0.5352
tr13-0653	0.7048	0.6000	0.5177	0.5314	0.5419	0.5213	0.5845	0.5392	0.5037
Average	0.5905	0.5896	0.5809	0.5337	0.5224	0.5582	0.5544	0.5805	0.5356

3.3. Computational Analysis

With regard to the classification and the use of machine learning algorithms, an advantage of using the novel multichannel indices compared to the existing single-channel indices is also the saving of memory and time, due to the reduction of dimensionality. The accuracies of Index1 and Index2 from Table 4 were achieved with the model constructed from the single feature only, while all other indices had six features since the dataset contains six EEG channels. For this reason, storing the novel indices consumes six times less memory. The time consumption was measured as an average of 100 executions. The measured time included classifier initialization, classifier training, classifications on the test subject and calculation of classification accuracy. Table 8 shows the results of time consumption measurements. The use of the novel multichannel indices saves about 30% of time compared to all other traditionally used single-channel ratio indices.

Table 8. The average time of 100 executions of the XGBoost classifier's initialization, training, classifications on the test subject and calculation of classification accuracy, expressed in milliseconds. The shaded green cells with bold text represent the best values for each subject and the best average value.

Subject	Index1	Index2	θ/α	β/α	$(\theta + \alpha)/\beta$	θ/β	$(\theta + \alpha)/(\alpha + \beta)$	γ/δ	$(\gamma + \beta)/(\delta + \alpha)$
This Work			Eoh et al. [36]		Jap et al. [37]		da Silveira et al. [10]		
tr03-0092	86.3772	86.9689	122.7764	124.0313	129.9221	129.5543	130.6956	128.6490	128.9000
tr03-0256	86.7446	87.1034	123.3161	123.2911	130.3844	130.2867	130.6147	128.6415	128.7518
tr03-0876	86.3508	87.0188	122.6970	123.5249	130.2485	130.7789	130.6456	128.7160	128.3419
tr03-1389	85.9344	86.8811	122.1586	123.8243	130.4414	130.1170	131.3382	129.2281	128.8390
tr04-0649	86.9833	87.5527	123.6650	124.0832	130.2565	130.0574	130.2316	129.2234	128.8357
tr04-0726	86.5498	87.5921	123.7690	123.6945	129.6285	129.6750	130.6256	128.9002	128.6363
tr05-1434	86.5450	86.6549	123.0505	130.6853	131.9267	130.6205	131.3138	130.6215	131.3438
tr05-1675	87.0534	87.4627	123.3135	130.0399	130.4660	129.1552	129.5943	130.5365	128.9256
tr07-0168	86.9143	87.2251	122.9559	129.6185	130.5158	130.1070	129.7780	129.6381	128.7795
tr07-0458	86.5651	86.8690	122.6074	129.9533	130.1468	130.2667	130.4915	128.8906	129.5788
tr07-0861	86.8634	87.2801	122.7545	130.2319	130.0104	128.1135	129.9124	129.3949	128.7760
tr08-0021	86.9239	88.9566	123.0910	130.0868	129.1948	130.0221	130.1670	129.3241	129.1652
tr08-0111	86.6697	87.4626	122.7216	130.3879	130.5019	130.4011	129.8419	128.8413	129.1940
tr09-0175	87.3240	87.3827	123.4803	129.6729	130.9953	130.1271	131.1785	128.7062	128.9786
tr10-0872	86.9690	87.5381	124.0918	130.5091	129.4638	130.6490	130.2964	129.4843	129.3599
tr13-0204	87.2509	87.1135	123.2010	131.7928	130.4062	131.4087	130.3568	128.9199	128.1530
Average	86.7512	87.3164	123.1031	127.8392	130.2818	130.0838	130.4426	129.2322	129.0350

4. Discussion

The main idea of our research was to combine frequency-domain features from different brain regions into a multichannel ratio index to improve frequency-domain features for the detection of drowsiness and to gain new insights into drowsiness. The results in Tables 2–8 suggest that two novel multichannel ratio indices improve the detection of drowsiness based on the frequency-domain features and reduce the time required for detection.

We must note that the main idea of this research was not to create the best possible model for drowsiness detection but only to bring improvement into frequency-domain features that are often used for drowsiness detection. Our focus was on developing the method for obtaining these novel indices, which is explained in Section 2.3 “Multi-Objective Optimization”. In order to confirm that our conclusions also hold for other classifiers besides XGBoost, Table 9 shows the average accuracy on the test set obtained with Naïve Bayes, k nearest neighbors, logistic regression, decision tree, random forest and support vector machine classifiers (using the scikit-learn library at default settings). The average accuracies of two novel indices vary from 56% to 65% among the algorithms. All the algorithms show that our novel multichannel indices are better than existing single-channel indices.

Table 9. The average accuracy was obtained on the test set with different classification algorithms. Each row is colored with a pallet of colors ranging from dark green for the highest number in the row to dark red for the lowest number in the row. The algorithms are: NB—Naïve Bayes, KNN—k nearest neighbors, Logistic—logistic regression, DT—decision tree, RF—random forest and SVM—support vector machine.

Algorithm	Index1	Index2	θ/α	β/α	$(\theta + \alpha)/\beta$	θ/β	$(\theta + \alpha)/(\alpha + \beta)$	γ/δ	$(\gamma + \beta)/(\delta + \alpha)$
	This Work		Eoh et al. [36]			Jap et al. [37]		da Silveira et al. [10]	
NB	0.6399	0.6535	0.5947	0.5462	0.5432	0.5308	0.5663	0.5316	0.5277
KNN	0.5785	0.5840	0.5588	0.5387	0.5399	0.5452	0.5525	0.5378	0.5277
Logistic	0.6396	0.6543	0.6029	0.5131	0.5383	0.5626	0.5735	0.5793	0.5613
DT	0.5717	0.5629	0.5456	0.5050	0.5074	0.5420	0.5267	0.5356	0.5223
RF	0.5719	0.5659	0.5762	0.5380	0.5360	0.5549	0.5501	0.5321	0.5222
SVM	0.6325	0.6526	0.6200	0.5695	0.5714	0.5731	0.5801	0.5541	0.5478

Our results were compared with the seven existing single-channel ratio indices that are currently state-of-the-art frequency-domain features. The newest one was introduced in 2016 [10], but all of these single-channel ratio indices are often used in the more recent drowsiness detection papers [57–59].

The authors in the aforementioned research report 92% accuracy as the best-obtained accuracy [57]. This accuracy was obtained based on the epoch-level validation. Epoch-level validation is a cross-validation procedure on the epoch level, which means that there is a very high probability that all subjects will have epochs in the training set and in the test set at the same time. On the other hand, subject-level validation is validation where it is ensured that subjects in the test set are not contained in the training set. An example of a subject-level validation is the leave-one-subject-out cross-validation that we used in this research. The only proper way for model validation is subject-level validation, as it represents the real-life setting in which the data from a new subject are used only for testing the model. Empirical tests conducted in related research showed a large difference in the accuracies between epoch-level validation and subject-level validation [60].

In a study from Mehreen et al. [57], the authors also provide subject-level validation, and the accuracy achieved was 71.5% based on 15 frequency-domain features. The highest accuracy achieved in our research is shown in Table 9, and it was 65.45%, achieved by logistic regression. This 65.45% accuracy is relatively close to 71.5%, and it must be noted

that it was obtained based only on the Index2 feature, with a simple algorithm and without any parameter optimization. Due to this, we are confident that the addition of our two multichannel ratio indices would lead to an improvement in all state-of-the-art drowsiness detection systems that use frequency-domain features. Again, our aim was not to create the best possible drowsiness detection model but to prove that the novel multichannel indices are better than the existing single-channel frequency-domain features.

The Equations (4) and (6) for these multichannel ratio indices, obtained after optimizing the parameters with the optimization algorithm, suggest that alpha and delta are two of the most important frequency power bands for drowsiness detection. Equation (6) suggests that delta power in the frontal region describes drowsiness better than in the central region, while alpha power in the occipital and central regions describes drowsiness better than in the frontal region.

These results are consistent with several previous research papers on drowsiness detection that reported the importance of increasing alpha power [22,35,61,62]. Delta power is usually only present in deep sleep stages [36], so some researchers studying drowsiness do not include delta in their research [63]. However, there is still much research that includes delta power. The increase in delta power is considered to be an indicator of drowsiness [4]. Our research found that theta and beta powers are not as good drowsiness indicators as alpha and delta powers, while many other research studies disagree. A decrease in beta power was found to be an indicator of drowsiness in [4,36,64,65] and an increase in theta power was found to be an indicator of drowsiness in [27,34,61,62,65]. Wang et al. [38], in their study of microsleep events, found that alpha and delta rhythms characterize microsleep events. As mentioned earlier, there is an inconsistency in terminology, and some researchers consider sleep stage S1 as drowsiness [5–9], while Johns [11] considers it equivalent to microsleep events in the driving scenario. We used the data from sleep stage S1 and referred to it in this research as drowsiness. Since our results suggest that delta and alpha are the most significant for the detection of drowsiness, as in the work of Wang et al. [38] on microsleep events, our work suggests that sleep stage S1 may be more similar to microsleep events than to drowsiness, but further research is needed to support this as a fact.

Apart from the indication that drowsiness is closely related to microsleep events, it may also be closely linked to driver fatigue. Some researchers even use the term fatigue as a synonym for drowsiness [66]. Fatigue is a consequence of prolonged physical or mental activity [67] and can lead to drowsiness [68]. Normally, rest and inactivity relieve fatigue, however, they exacerbate drowsiness [69]. Lal and Craig [70] found that delta and theta band activities increase significantly during fatigue. Craig et al. [71] reported significant changes in the alpha 1, alpha 2, theta and beta bands, while they did not find any significant changes in the delta band when observing driver fatigue. Simon et al. [68] report that alpha band power and alpha spindles correlate with fatigue.

These three research papers [68,70,71] all use visual inspection to define the ground truth of fatigue. This approach to defining the ground truth is prone to subjectivity. A similar problem occurs when drowsiness is defined by using subjective drowsiness ratings, such as the Karolinska sleepiness scale [72].

Driver drowsiness, driver fatigue and microsleep events are defined as different internal states of the brain, but show similar behavior when observing the features obtained from the EEG. Possible explanations could be that fatigue, drowsiness and microsleep have a similar effect on brain functions and cause the driver's inability to function at the desired level. Most researchers of these three driver states only use frequency-domain features, while there are a number of other features (nonlinear features [30], spatiotemporal features [31] and entropies [32]) that could be used. Further studies with these features could find some features of the EEG signal that distinguish drowsiness, fatigue and microsleep. Distinguishing features of these three brain states could lead to the exact definitions of these terms. Precise and standardized definitions of fatigue, drowsiness and microsleep would help researchers to compare their work more easily.

Figure 4 shows that the proposed procedure for creating the novel ratio indices has succeeded in creating step-like indices for a given subject. In addition to Index1 and Index2, which show desirable behavior, the indices $(\theta + \alpha)/\beta$ and θ/β show similar, favorable behavior for a few channels. Figure 5 shows a comparison of novel ratio indices with the best and the worst channel for γ/δ and $(\gamma + \beta)/(\delta + \alpha)$ single-channel indices for subject tr04-0726. Index1, index2, θ/α [36], $(\theta + \alpha)/\beta$, $(\theta + \alpha)/(\alpha + \beta)$ [37] and γ/δ [10] show similar behavior. These indices seem to detect drowsiness well, but with about a 50 epochs delay. Since several different single-channel indices that were previously shown to correlate with drowsiness together with two novel multichannel indices show the same delay in detecting drowsiness, this suggests that there may be shortcomings in the labeling of the initial signals. The manual for scoring sleep [42] provides guidelines for labeling, and it may be possible that the professionals who labeled the sleep signals labeled an approximate time of transition from the W state to the S1 state, as it is known that labeling any kind of several-hour-long EEG signal is a very tedious, hard and time-consuming job [73]. For this reason, the loose transition window is applied in the optimization algorithm, as described in Section 2.3.

The main shortcoming in applying our approach is the need to place six EEG electrodes on the driver's scalp while driving. Apart from being intrusive, there is also a problem with noise in real-world applications that cannot be neutralized with the current state-of-the-art filter technology. All electrophysiological signals measured with wearable devices have a similar problem with intrusiveness and noise. ECG measurements, for example, are somewhat less susceptible to noise than EEG. Several recent works have shown that ECG can be used as a good predictor of sleep stages based on deep learning classifiers. Sun et al. [74] combined ECG with abdominal respiration and obtained a kappa value of 0.585, while Sridhar et al. [75] obtained a kappa value of 0.66. Combining EEG and ECG measurements has also been proposed in the context of driver drowsiness detection under simulator-based laboratory conditions [76]. Despite the problems of intrusiveness and noise susceptibility, research based on the electrophysiological signals brings a shift towards a precise definition of drowsiness. Once there is an exact definition of drowsiness or at least guidelines and manuals that accurately describe drowsiness (similar to the manuals for evaluating sleep stages), a big step will be taken to solve the problem of early detection of drowsiness [77]. It is doubtful that a wearable system based on electrophysiological signals will ever be widely used in real-world driving, but they still need to be developed. In our opinion, such wearable electrophysiological devices are more likely to be used for calibration/validation of non-intrusive systems (such as the driving performance-based or video-based systems) in controlled/simulated driving scenarios. In such scenarios, it is possible to control ambient noise, leading to a reduction in the effects of noise sensitivity.

An additional limitation of this work is that we were able to download data from 393 of 992 subjects completely, and only 28 of these 393 subjects were included in our study due to the inclusion condition that we described in Section 2.1 "Dataset, Preprocessing and Feature Extraction". Although it is a small subset of data, with the use of 12 subjects as a test set, we showed that the dataset is large enough to provide a good generalization (as seen in Tables 3, 5 and 7). In a recent review paper about state-of-the-art drowsiness detection [33], the authors reviewed 39 papers, and the average number of subjects in the included works is 23.5, which also indicates that our number of subjects included in the current study (28) is acceptable.

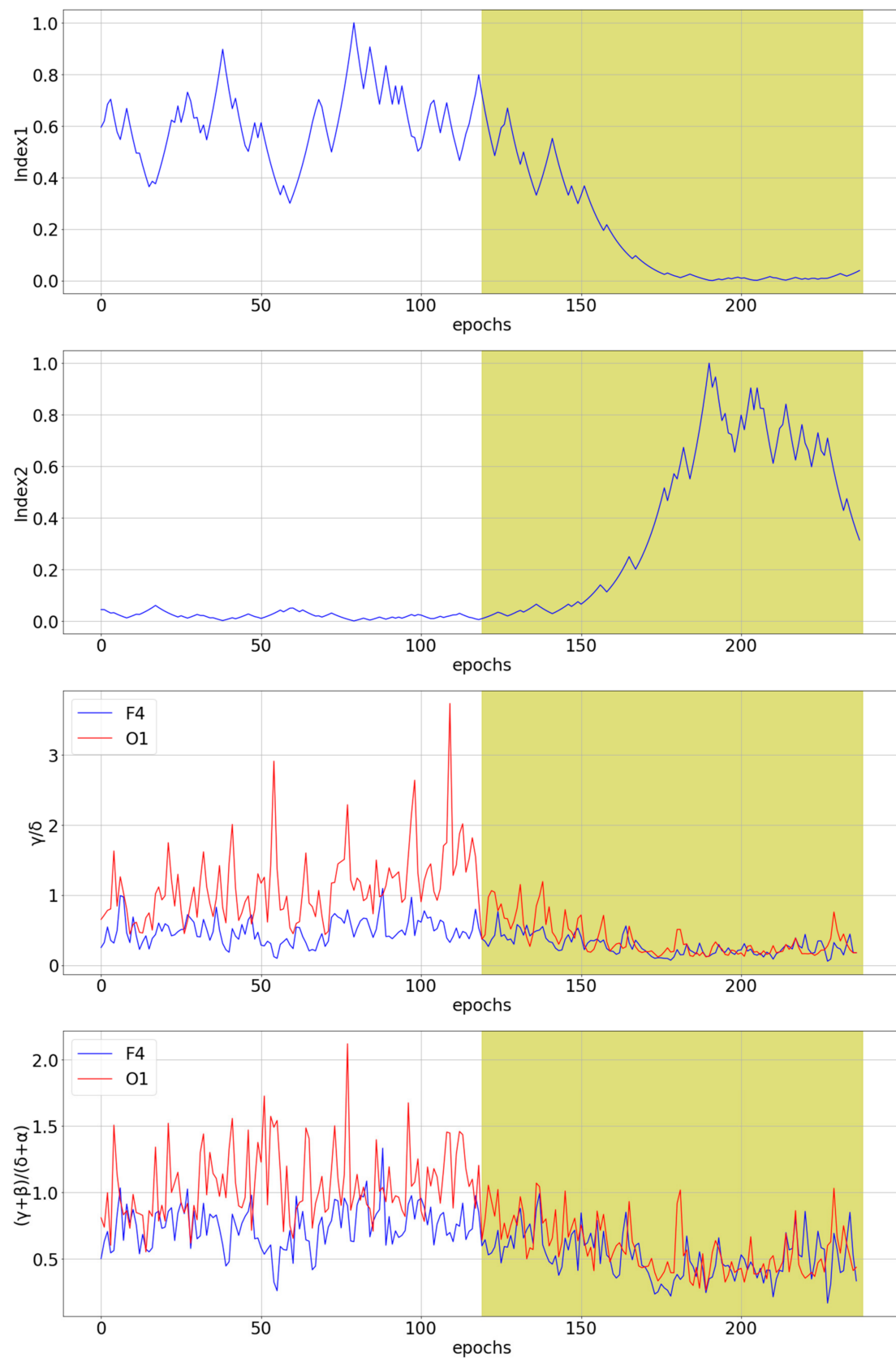


Figure 5. The comparison of the two novel multichannel indices with the best and the worst channel for γ/δ and $(\gamma + \beta)/(\delta + \alpha)$ single-channel indices for subject tr04-0726. The white part of the diagram represents the awake state, while the yellow part of the diagram represents the stage 1 of sleep, i.e., the drowsiness state.

5. Conclusions

This paper presented two novel multichannel ratio indices for the detection of drowsiness obtained by multi-objective optimization based on evolutionary computation. The results suggested that alpha and delta powers are good drowsiness indicators. The novel multichannel ratio indices were compared with seven existing single-channel ratio indices and showed better results in detecting drowsiness measured with precision and in the overall classification accuracy of both states using several machine learning algorithms. Our work suggests that a more precise definition of drowsiness is needed, and that accurate early detection of drowsiness should be based on multichannel frequency-domain ratio indices. The multichannel features also reduced the time needed for classification. The process of obtaining these indices by using a multi-objective optimization algorithm can also be applied to other areas of EEG signal analysis.

Research such as this, together with research on small hardware for physiology-based drowsiness detection, can eventually lead to an easy-to-use, non-intrusive device that reliably detects drowsiness. In addition, research on a reliable and standardized definition of drowsiness is needed and it would lead to improvements in the field of drowsiness detection.

Author Contributions: Conceptualization, I.S., N.F. and A.J.; methodology, I.S., N.F. and A.J.; software, I.S. and N.F.; validation, I.S., N.F., M.C. and A.J.; formal analysis, I.S., N.F. and A.J.; investigation, I.S.; resources, I.S. and N.F.; data curation, I.S.; writing—original draft preparation, I.S. and N.F.; writing—review and editing, I.S., N.F., M.C. and A.J.; visualization, I.S.; supervision, A.J.; project administration, M.C.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been carried out within the project “Research and development of the system for driver drowsiness and distraction identification—DFDM” (KK.01.2.1.01.0136), funded by the European Regional Development Fund in the Republic of Croatia under the Operational Programme Competitiveness and Cohesion 2014–2020.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the University of Zagreb, Faculty of Electrical Engineering and Computing (on 26 September 2020).

Informed Consent Statement: All the available information regarding patients is available from the PhysioNet portal [47], from the 2018 PhysioNet computing in cardiology challenge [48], at: <https://physionet.org/content/challenge-2018/1.0.0/> (accessed on 24 September 2021).

Data Availability Statement: The data used in this paper were obtained from the PhysioNet portal [47], from the 2018 PhysioNet computing in cardiology challenge [48], at: <https://physionet.org/content/challenge-2018/1.0.0/> (accessed on 24 September 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jackson, M.L.; Kennedy, G.A.; Clarke, C.; Gullo, M.; Swann, P.; Downey, L.A.; Hayley, A.C.; Pierce, R.J.; Howard, M.E. The utility of automated measures of ocular metrics for detecting driver drowsiness during extended wakefulness. *Accid. Anal. Prev.* **2016**, *127*, 127–133. [\[CrossRef\]](#)
2. Kamran, M.A.; Mannan, M.M.N.; Jeong, M.Y. Drowsiness, Fatigue and Poor Sleep’s Causes and Detection: A Comprehensive Study. *IEEE Access* **2019**, *7*, 167172–167186. [\[CrossRef\]](#)
3. Lal, S.K.L.; Craig, A. A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* **2001**, *173*, 173–194. [\[CrossRef\]](#)
4. Papadelis, C.; Chen, Z.; Kourtidou-Papadeli, C.; Bamidis, P.; Chouvarda, I.; Bekiaris, E.; Maglaveras, N. Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clin. Neurophysiol.* **2007**, *1906*, 1906–1922. [\[CrossRef\]](#)
5. Chowdhury, A.; Shankaran, R.; Kavakli, M.; Haque, M.M. Sensor Applications and Physiological Features in Drivers’ Drowsiness Detection: A Review. *IEEE Sens. J.* **2018**, *3055*, 3055–3067. [\[CrossRef\]](#)
6. Oken, B.S.; Salinsky, M.C.; Elsas, S.M. Vigilance, alertness, or sustained attention: Physiological basis and measurement. *Clin. Neurophysiol.* **2006**, *1885*, 1885–1901. [\[CrossRef\]](#)

7. Majumder, S.; Guragain, B.; Wang, C.; Wilson, N. On-board Drowsiness Detection using EEG: Current Status and Future Prospects. In Proceedings of the 2019 IEEE International Conference on Electro Information Technology (EIT), Brookings, SD, USA, 16–18 May 2019; pp. 483–490.
8. Sriraam, N.; Shri, T.P.; Maheshwari, U. Recognition of wake-sleep stage 1 multichannel eeg patterns using spectral entropy features for drowsiness detection. *Australas. Phys. Eng. Sci. Med.* **2016**, *797*, 797–806. [\[CrossRef\]](#)
9. Budak, U.; Bajaj, V.; Akbulut, Y.; Atila, O.; Sengur, A. An Effective Hybrid Model for EEG-Based Drowsiness Detection. *IEEE Sens. J.* **2019**, *7624*, 7624–7631. [\[CrossRef\]](#)
10. da Silveira, T.L.; Kozakevicius, A.J.; Rodrigues, C.R. Automated drowsiness detection through wavelet packet analysis of a single EEG channel. *Expert Syst. Appl.* **2016**, *559*, 559–565. [\[CrossRef\]](#)
11. Johns, M.W. A new perspective on sleepiness. *Sleep Biol. Rhythm.* **2010**, *170*, 170–179. [\[CrossRef\]](#)
12. Moller, H.J.; Kayumov, L.; Bulmash, E.L.; Nhan, J.; Shapiro, C.M. Simulator performance, microsleep episodes, and subjective sleepiness: Normative data using convergent methodologies to assess driver drowsiness. *Psychosom. J. Res.* **2006**, *335*, 335–342. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Martensson, H.; Keelan, O.; Ahlstrom, C. Driver Sleepiness Classification Based on Physiological Data and Driving Performance From Real Road Driving. *IEEE Trans. Intell. Transp. Syst.* **2019**, *421*, 421–430. [\[CrossRef\]](#)
14. Phillips, R.O. A review of definitions of fatigue—And a step towards a whole definition. *Transp. Res. Part F Traffic Psychol. Behav.* **2015**, *48*, 48–56. [\[CrossRef\]](#)
15. Slater, J.D. A definition of drowsiness: One purpose for sleep? *Med. Hypotheses* **2008**, *641*, 641–644. [\[CrossRef\]](#)
16. Subasi, A. Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert Syst. Appl.* **2005**, *701*, 701–711. [\[CrossRef\]](#)
17. Orasanu, J.; Parke, B.; Kraft, N.; Tada, Y.; Hobbs, A.; Anderson, B.; Dulchinos, V. *Evaluating the Effectiveness of Schedule Changes for Air Traffic Service (ATS) Providers: Controller Alertness and Fatigue Monitoring Study*; Technical Report; Federal Aviation Administration, Human Factors Division: Washington, DC, USA, 2012.
18. Hart, C.A.; Dinh-Zarr, T.B.; Sumwalt, R.; Weener, E. *Most Wanted List of Transportation Safety Improvements: Reduce Fatigue-Related Accidents*; National Transportation Safety Board: Washington, DC, USA, 2018.
19. Gonçalves, M.; Amici, R.; Lucas, R.; Åkerstedt, T.; Cirignotta, F.; Horne, J.; Léger, D.; McNicholas, W.T.; Partinen, M.; Téran-Santos, J.; et al. Sleepiness at the wheel across Europe: A survey of 19 countries. *Sleep J. Res.* **2015**, *24*, 242–253. [\[CrossRef\]](#)
20. Balandong, R.P.; Ahmad, R.F.; Saad, M.N.M.; Malik, A.S. A Review on EEG-Based Automatic Sleepiness Detection Systems for Driver. *IEEE Access* **2018**, *6*, 22908–22919. [\[CrossRef\]](#)
21. Kunding, T.; Sofra, N.; Riener, A. Assessment of the Potential of Wrist-Worn Wearable Sensors for Driver Drowsiness Detection. *Sensors* **2020**, *20*, 1029. [\[CrossRef\]](#)
22. Zheng, W.-L.; Gao, K.; Li, G.; Liu, W.; Liu, C.; Liu, J.-Q.; Wang, G.; Lu, B.-L. Vigilance Estimation Using a Wearable EOG Device in Real Driving Environment. *IEEE Trans. Intell. Transp. Syst.* **2020**, *170*, 170–184. [\[CrossRef\]](#)
23. Fu, R.; Wang, H.; Zhao, W. Dynamic driver fatigue detection using hidden Markov model in real driving condition. *Expert Syst. Appl.* **2016**, *397*, 397–411. [\[CrossRef\]](#)
24. Lin, C.-T.; Chuang, C.-H.; Tsai, S.-F.; Lu, S.-W.; Chen, Y.-H.; Ko, L.-W. Wireless and Wearable EEG System for Evaluating Driver Vigilance. *IEEE Trans. Biomed. Circuits Syst.* **2014**, *165*, 165–176.
25. Cassani, R.; Falk, T.H.; Horai, A.; Gheorghe, L.A. Evaluating the Measurement of Driver Heart and Breathing Rates from a Sensor-Equipped Steering Wheel using Spectrotemporal Signal Processing. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 2843–2847.
26. Li, Z.; Li, S.; Li, R.; Cheng, B.; Shi, J. Online Detection of Driver Fatigue Using Steering Wheel Angles for Real Driving Conditions. *Sensors* **2017**, *17*, 495. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Aeschbach, D.; Matthews, J.R.; Postolache, T.T.; Jackson, M.A.; Giesen, H.A.; Wehr, T.A. Dynamics, of the human EEG during prolonged wakefulness: Evidence for frequency-specific circadian and homeostatic influences. *Neurosci. Lett.* **1997**, *239*, 121–124. [\[CrossRef\]](#)
28. Barua, S.; Ahmed, M.U.; Ahlström, C.; Begum, S. Automatic driver sleepiness detection using EEG, EOG and contextual information. *Expert Syst. Appl.* **2019**, *121*, 121–135. [\[CrossRef\]](#)
29. Wang, L.; Li, J.; Wang, Y. Modeling and Recognition of Driving Fatigue State Based on R-R Intervals of ECG Data. *IEEE Access* **2019**, *7*, 175584–175593. [\[CrossRef\]](#)
30. Stam, C.J. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clin. Neurophysiol.* **2005**, *2266*, 2266–2301. [\[CrossRef\]](#)
31. Bastos, A.M.; Schoffelen, M.J. A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Front. Syst. Neurosci.* **2016**, *9*, 175. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Acharya, U.R.; Hagiwara, Y.; Deshpande, S.N.; Suren, S.; Koh, J.E.W.; Oh, S.L.; Arunkumar, N.; Ciaccio, E.J.; Lim, C.M. Characterization of focal EEG signals: A review. *Future Gener. Comput. Syst.* **2019**, *290*, 290–299. [\[CrossRef\]](#)
33. Stancin, I.; Cifrek, M.; Jovic, A. A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems. *Sensors* **2021**, *21*, 3786. [\[CrossRef\]](#)
34. Cajochen, C.; Brunner, D.P.; Krauchi, K.; Graw, P.; Wirz-Justice, A. Power Density in Theta/Alpha Frequencies of the Waking EEG Progressively Increases During Sustained Wakefulness. *Sleep* **1995**, *890*, 890–894. [\[CrossRef\]](#)

35. Astolfi, L.; Mattia, D.; Vecchiato, G.; Babiloni, F.; Borghini, G. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* **2012**, *58*, 58–75.
36. Eoh, H.J.; Chung, M.K.; Kim, S.H. Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *Int. Ind. J. Ergon.* **2005**, *307*, 307–320. [\[CrossRef\]](#)
37. Jap, B.T.; Lal, S.; Fischer, P.; Bekiaris, E. Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Syst. Appl.* **2009**, *36*, 2352–2359. [\[CrossRef\]](#)
38. Wang, C.; Guragain, B.; Verma, A.K.; Archer, L.; Majumder, S.; Mohamud, A.; Flaherty-Woods, E.; Shapiro, G.; Almashor, M.; Lenné, M.; et al. Spectral Analysis of EEG During Microsleep Events Annotated via Driver Monitoring System to Characterize Drowsiness. *IEEE Trans. Aerosp. Electron. Syst.* **2020**, *1346*, 1346–1356. [\[CrossRef\]](#)
39. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.-K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Ghassemi, M.; Moody, B.; Lehman, L.-W.; Song, C.; Li, Q.; Sun, H.; Westover, B.; Clifford, G. You Snooze, You Win: The PhysioNet/Computing in Cardiology Challenge 2018. In Proceedings of the 2018 Computing in Cardiology Conference (CinC), Maastricht, The Netherlands, 23–26 September 2018.
41. Institute of Medicine (US) Committee on Sleep Medicine and Research; Colten, H.; Altevogt, B. Sleep Physiology. In *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*; National Academies Press (US): Washington, DC, USA, 2006.
42. Berry, R.B.; Quan, S.F.; Abreu, A.R.; Bibbs, M.L.; Del Rosso, L.; Harding, S.M.; Mao, M.; Plante, D.T.; Pressman, M.R.; Troester, M.M.; et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*; Version 2.6; American Academy of Sleep Medicine: Darien, IL, USA, 2020.
43. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **1967**, *15*, 70–73. [\[CrossRef\]](#)
44. Basha, A.J.; Balaji, B.S.; Poornima, S.; Prathilothamai, M.; Venkatachalam, K. Support vector machine and simple recurrent network based automatic sleep stage classification of fuzzy kernel. *J. Ambient Intell. Humaniz. Comput.* **2020**, *12*, 6189–6197. [\[CrossRef\]](#)
45. Feoktistov, V. *Differential Evolution 5*; Springer US: Boston, MA, USA, 2006.
46. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *182*, 182–197. [\[CrossRef\]](#)
47. Chugh, T.; Sindhya, K.; Hakanen, J.; Miettinen, K. A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms. *Soft Comput.* **2019**, *23*, 3137–3166. [\[CrossRef\]](#)
48. Moctezuma, L.A.; Molinas, M. Towards a minimal EEG channel array for a biometric system using resting-state and a genetic algorithm for channel selection. *Sci. Rep.* **2020**, *10*, 14917. [\[CrossRef\]](#)
49. Hadka, D. MOEA Framework. Available online: <http://moeaframework.org/> (accessed on 20 September 2021).
50. Coello Coello, C.; Lamont, G.B.; van Veldhuizen, D.A. *Evolutionary Algorithms for Solving Multi-Objective Problems*; Springer: Boston, MA, USA, 2007.
51. Coello Coello, C.A. Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: A survey of the state of the art. *Comput. Methods Appl. Mech. Eng.* **2002**, *191*, 1245–1287. [\[CrossRef\]](#)
52. Wang, Y.; Cai, Z.; Zhou, Y.; Zeng, W. An Adaptive Tradeoff Model for Constrained Evolutionary Optimization. *IEEE Trans. Evol. Comput.* **2008**, *80*, 80–92. [\[CrossRef\]](#)
53. Homaifar, A.; Qi, C.X.; Lai, S.H. Constrained Optimization Via Genetic Algorithms. *Simulation* **1994**, *242*, 242–253. [\[CrossRef\]](#)
54. Runarsson, T.P.; Yao, X. Stochastic ranking for constrained evolutionary optimization. *IEEE Trans. Evol. Comput.* **2000**, *284*, 284–294. [\[CrossRef\]](#)
55. McDonald, J.H. *Handbook of Biological Statistics*, 3rd ed.; Sparky House Publishing: Baltimore, MD, USA, 2014.
56. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
57. Mehreen, A.; Anwar, S.M.; Haseeb, M.; Majid, M.; Ullah, M.O. A Hybrid Scheme for Drowsiness Detection Using Wearable Sensors. *IEEE Sens. J.* **2019**, *5119*, 5119–5126. [\[CrossRef\]](#)
58. Seok, W.; Yeo, M.; You, J.; Lee, H.; Cho, T.; Hwang, B.; Park, C. Optimal Feature Search for Vigilance Estimation Using Deep Reinforcement Learning. *Electronics* **2020**, *9*, 142. [\[CrossRef\]](#)
59. Wu, E.Q.; Peng, X.Y.; Zhang, C.Z.; Lin, J.X.; Sheng, R.S.F. Pilots' Fatigue Status Recognition Using Deep Contractive Autoencoder Network. *IEEE Trans. Instrum. Meas.* **2019**, *3907*, 3907–3919. [\[CrossRef\]](#)
60. Kamrud, A.; Borghetti, B.; Schubert Kabban, C. The Effects of Individual Differences, Non-Stationarity, and the Importance of Data Partitioning Decisions for Training and Testing of EEG Cross-Participant Models. *Sensors* **2021**, *21*, 3225. [\[CrossRef\]](#)
61. Lin, F.-C.; Ko, L.-W.; Chuang, C.-H.; Su, T.-P.; Lin, T.-C. Generalized EEG-Based Drowsiness Prediction System by Using a Self-Organizing Neural Fuzzy System. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2012**, *2044*, 2044–2055. [\[CrossRef\]](#)
62. Jap, B.T.; Lal, S.; Fischer, P. Comparing combinations of EEG activity in train drivers during monotonous driving. *Expert Syst. Appl.* **2011**, *996*, 996–1003. [\[CrossRef\]](#)




63. Akbar, I.A.; Rumagit, A.M.; Utsunomiya, M.; Morie, T.; Igasaki, T. Three drowsiness categories assessment by electroencephalogram in driving simulator environment. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11–15 July 2017; pp. 2904–2907.
64. Keckluno, G.; Åkersteot, T. Sleepiness in long distance truck driving: An ambulatory EEG study of night driving. *Ergonomics* **1993**, *1007*, 1007–1017. [[CrossRef](#)]
65. Dussault, C.; Jouanin, J.-C.; Philippe, M.; Guezennec, Y.C. EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviat. Space Environ. Med.* **2005**, *344*, 344–351.
66. Khushaba, R.N.; Kodagoda, S.; Lal, S.; Dissanayake, G. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *IEEE Trans. Biomed. Eng.* **2011**, *121*, 121–131. [[CrossRef](#)]
67. Brown, I.D. Driver Fatigue. *Hum. Factors* **1994**, *36*, 298–314. [[CrossRef](#)] [[PubMed](#)]
68. Simon, M.; Schmidt, E.A.; Kincses, W.E.; Fritzsche, M.; Bruns, A.; Aufmuth, C.; Bogdan, M.; Rosenstiel, W.; Schrauf, M. EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clin. Neurophysiol.* **2011**, *122*, 1168–1178. [[CrossRef](#)]
69. Johns, M.W.; Chapman, R.; Crowley, K.; Tucker, A. A new method for assessing the risks of drowsiness while driving. *Somnologie—Schlafforsch. Schlafmed.* **2008**, *66*, 66–74. [[CrossRef](#)]
70. Lal, S.K.L.; Craig, A. Driver fatigue: Electroencephalography and psychological assessment. *Psychophysiology* **2002**, *313*, 313–321. [[CrossRef](#)]
71. Craig, A.; Tran, Y.; Wijesuriya, N.; Nguyen, H. Regional brain wave activity changes associated with fatigue. *Psychophysiology* **2012**, *574*, 574–582. [[CrossRef](#)]
72. Kaida, K.; Takahashi, M.; Åkerstedt, T.; Nakata, A.; Otsuka, Y.; Haratani, T.; Fukasawa, K. Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clin. Neurophysiol.* **2006**, *1574*, 1574–1581. [[CrossRef](#)]
73. Gajic, D.; Djurovic, Z.; Di Gennaro, S.; Gustafsson, F. Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition. *Biomed. Eng. Appl. Basis Commun.* **2014**, *26*, 1450021. [[CrossRef](#)]
74. Sun, H.; Ganglberger, W.; Panneerselvam, E.; Leone, M.J.; Quadri, S.A.; Goparaju, B.; Tesh, R.A.; Akeju, O.; Thomas, R.J.; Westover, M.B. Sleep staging from electrocardiography and respiration with deep learning. *Sleep* **2020**, *43*, zsz306. [[CrossRef](#)] [[PubMed](#)]
75. Sridhar, N.; Shoeb, A.; Stephens, P.; Kharbouch, A.; Shimol, D.B.; Burkart, J.; Ghoreyshi, A.; Myers, L. Deep learning for automated sleep staging using instantaneous heart rate. *NPJ Digit. Med.* **2020**, *3*, 106. [[CrossRef](#)]
76. Awais, M.; Badruddin, N.; Drieberg, M. A Hybrid Approach to Detect Driver Drowsiness Utilizing Physiological Signals to Improve System Performance and Wearability. *Sensors* **2017**, *17*, 1991. [[CrossRef](#)] [[PubMed](#)]
77. Dong, Y.; Hu, Z.; Uchimura, K.; Murayama, N. Driver Inattention Monitoring System for Intelligent Vehicles: A Review. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 596–614. [[CrossRef](#)]

Publication 3

I. Stancin, M. Z. Zeba, K. Friganovic, M. Cifrek, and A. Jovic, “Information on Drivers’ Sex Improves EEG-Based Drowsiness Detection Model,” *Applied Sciences*, vol. 12, no. 16, p. 8146, Aug. 2022, doi: 10.3390/app12168146.

Article

Information on Drivers' Sex Improves EEG-Based Drowsiness Detection Model

Igor Stancin ¹, Mirta Zelenika Zeba ², Kresimir Friganovic ^{2,3} , Mario Cifrek ²  and Alan Jovic ^{1,*} ¹ Department of Electronics, Microelectronics, Computer and Intelligent Systems, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia² Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia³ Genome Institute of Singapore, Agency for Science Technology and Research, 1 Fusionopolis Way, #20-10 Connexis, Singapore 138632, Singapore

* Correspondence: alan.jovic@fer.hr

Abstract: Objective detection of a driver's drowsiness is important for improving driving safety, and the most prominent indicator of drowsiness is changes in electroencephalographic (EEG) activity. Despite extensively documented behavioral differences between male and female drivers, previous studies have not differentiated drowsiness detection models based on drivers' sex. Therefore, the overall aim of this study is to demonstrate that drowsiness detection can be improved with the use of drivers' sex information, either as a feature or as separate sex-dependent datasets. Additionally, we aim to provide a reliable EEG-based sex classification model. The used dataset consists of 17 male and 17 female drivers which were evaluated during alert and drowsy sessions. Frequency-domain and recurrence quantification analysis EEG features were used. Four classification algorithms and three feature selection methods were applied to build the models. The accuracy of drowsiness detection based on sex-dependent datasets is 84% for male drivers and 88% for female drivers, which is 3% and 7% better, respectively, than the classification without information about driver's sex (81%). The model for sex classification based on EEG achieved high accuracy: 97% correctly identified participants in alert sessions and 96% in drowsy sessions. All participants were correctly classified after the application of majority voting on five algorithm runs. The results suggest that sex-dependent datasets improve the accuracy of drowsiness models, which may be relevant to a variety of drowsiness detection systems currently being developed in the field.

Keywords: drowsiness detection; EEG features; machine learning; recurrence quantification analysis; sex classification; sex differences



Citation: Stancin, I.; Zeba, M.Z.; Friganovic, K.; Cifrek, M.; Jovic, A. Information on Drivers' Sex Improves EEG-Based Drowsiness Detection Model. *Appl. Sci.* **2022**, *12*, 8146. <https://doi.org/10.3390/app12168146>

Academic Editor: Alexander E. Hramov

Received: 19 July 2022

Accepted: 12 August 2022

Published: 15 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Driving requires various cognitive skills such as visual perception, attention, memory, executive functions, and motor skills [1]. Performance in these cognitive domains is related to driving performance and driving ability. In addition, psychophysiological aspects such as drowsiness and fatigue can affect cognitive processes during driving and lead to traffic accidents [2]. Moreover, fatigue and drowsy driving are some of the major causes of traffic accidents [3,4]. Therefore, objective detection of a driver's drowsiness is an important factor that could improve driving safety.

However, drowsiness in drivers is not easy to detect. The most used measures are self-assessments of drowsiness. The European Union regulation for the type-approval of motor vehicles with regard to their drivers' drowsiness and attention warning systems is based on the self-assessment of drowsiness with the Karolinska Sleepiness Scale [5]. Other methods for inferring correlates of drowsiness are image-based methods, vehicle-based methods, physiological-based methods, and hybrid methods [6]. Each of these categories can be further split into subcategories, e.g., physiological-based methods can be based

on electroencephalogram (EEG), electrocardiogram (ECG), respiration, electrooculogram, electromyogram, galvanic skin response, or skin temperature [7].

ECG records the heart's electrical activity (heartbeats), while heart rate variability (HRV) refers to the variations in times between two adjacent heartbeats. HRV is considered a good drowsiness predictor and accuracies of drowsiness detection based on HRV range from 56.6% to 95% among studies [8]. Respiration can be derived from HRV but is also measured using several other techniques. Siddiqui et al. [9], in their recent research, used radar for non-invasive respiration measurement and obtained 87% drowsiness classification accuracy. Measuring skin temperature can also be used for the assessment of drowsiness. Gielen and Aerts [10] measured temperature on the nose and wrist, and obtained accuracies of 68.4% and 88.9%, respectively. Electrooculogram measures eye movement characteristics. It is often used for drowsiness detection, achieving accuracies that range from 64% to 99% among studies [11].

EEG provides a very accurate assessment of the driver's mental state [12], achieving accuracies of drowsiness detection that range from 67% to 99% among studies [13]. Furthermore, changes in EEG activity are considered biomarkers of mental fatigue and drowsiness [14]. The most prominent indicator associated with mental fatigue and drowsiness is increased theta activity in frontal, central, and posterior cortical sites. In addition, increased alpha activity is associated with individual variability in cortical changes related to mental fatigue and drowsiness. Many studies also use ratio indices between these frequency bands as indicators of drowsiness [15], and a recent study suggests that multichannel ratio indices could bring an additional increase in drowsiness detection accuracy [16]. However, the results obtained are from studies with male participants or males and females as a group, without separating them by sex.

Differences in driving behavior between men and women have been extensively documented using experimental driving tasks, attitudes toward driving, behavioral analysis, risk perception, and the number of accidents.

Drowsiness also has differential effects on driving behavior between men and women. For example, self-rated levels of drowsiness while driving tend to be higher in women than in men [17]. Women also tend to report longer ideal sleep durations than men [18]. As a result, it appears that women's greater need for sleep leads to higher ratings of drowsiness.

Studies also report sex differences in brain organization that influence the regulation of brain activity during awake state and sleep [19]. In addition, sleep deprivation has a differential effect on brain activity between females and males. Therefore, it is reasonable to assume that changes in EEG activity related to drowsiness differ between men and women.

There are various methods of EEG signal analysis to detect drowsiness and alertness in drivers. They are mostly based on the different types of features extracted from the signal (i.e., frequency-domain features, recurrence quantification analysis features, entropies, etc.) [20–25]. To the best of our knowledge, there is no research that includes drivers' sex information in drowsiness detection systems.

A preliminary version of this work has been reported [26] and it shows the statistical difference between EEG features of alert male and female drivers. The main goal of this substantially extended study is to improve drowsiness detection by including the information about drivers' sex in the classifier. This is done in two ways: (1) by considering sex as a feature, and (2) by separating the datasets into male and female. In addition, the study aims to develop a reliable EEG-based sex classification model, where correlations between features are introduced as a novel feature that differentiates between male and female drivers.

2. Methodology

2.1. Experimental Design

In our study, 34 healthy participants were recorded during two sessions in a driving simulation. All participants had a valid driver's license and drove a car regularly but are not professional drivers. The EEG was recorded with a 32-channel actiCHamp EEG

amplifier (Brain Products, Munich, Germany) with passive sintered Ag/AgCl electrodes. The electrodes were located at prefrontal (Fp1, Fp2), frontal (F3, F4), central frontal (FC5, FC1, FC2, FC6), inferior frontal (F7, F8), midline frontal (Fz), central (C3, C4), midline central (Cz), midline parietal (Pz), central parietal (CP5, CP1, CP2, CP6), midline occipital (Oz), inferior temporal (T7, T8), posterior temporal (TP9, TP10), parietal (P3, P4), inferior parietal (P7, P8), posterior occipital (PO9, PO10), and occipital scalp sites (O1, O2). The electrodes were positioned according to the International 10–20 system guidelines. The electrode at the FCz position was used as a reference. A ground electrode was placed on the forehead. BrainVision Recorder software was used for impedance check and on-line monitoring. The EEG signals were recorded with a 1000 Hz sampling rate. A simulation scenario was shown on three wide LCD screens. The car was controlled with a professional steering wheel joystick and pedals. All participants were instructed to follow the traffic rules. Male participants averaged 30.24 of age with a standard deviation of 6.86 years and female participants averaged 30.12 years of age with a standard deviation of 6.98 years. The driving scenario was the same for all participants and consisted of driving on state roads and highways, and in an urban environment. The study was approved by the Ethics Committee of the University of Zagreb, Faculty of Electrical Engineering and Computing.

Each participant had two recording sessions. The first recording session (alert session) began at 2:00 p.m. and the second one (drowsy session) began at midnight. The sessions were recorded on different days. The driving scenario was the same for both sessions: after 15 min of adaptation driving (simple acceleration, braking, and turning exercises), all participants were instructed to drive on a highway for 90 min, with adjusted day/night lighting according to the time of the session. Highway driving was monotonic, with very few other cars in traffic, and did not include any unexpected events (such as road crossings of animals or pedestrians). Room lighting, temperature, and humidity were controlled and were the same for both sessions.

Each EEG recording was divided into 10 s epochs with five seconds' overlap between epochs. The first 250 epochs (~21 min) of highway driving in the alert session were labeled as periods with the highest alertness level for all participants. The last 250 epochs of the drowsy session (approximately after 70 min of highway driving) were labeled as periods with the highest drowsiness level for all participants. This labeling, where the beginning of the session is labeled as alert, and the end is labeled as drowsy, is commonly used in practice [27–30].

These classifications of two-phase classes were additionally confirmed by an expert in psychophysiological behavior based on their self-assessment of drowsiness (Karolinska Sleepiness Scale) before and after the session, and visual inspection of participants. Participants indicated their drowsiness level on the Karolinska Sleepiness Scale for the alert session as extremely alert, very alert, alert, or fairly alert. Additionally, their total score on the Fatigue Assessment Scale indicated no fatigue before the alert session. Table 1 shows the number of male and female drivers in each class.

Although we used a 32-channel EEG, we wanted to describe dependencies and differences between specific brain regions—front left (FL), front right (FR), occipital left (OL), and occipital right (OR) regions. We calculated the mean value for each feature from five channels in each region (see Table 2). In addition to the channels from these four regions, we also included Oz, Pz, and Cz channels in our analysis.

Table 1. The number of participants in each class.

	Male	Female
Alertness	17	17
Drowsiness	17	17

Table 2. Channels in brain regions.

Region	Channels
Front left (FL)	F7, F3, FC5, FC1, T7
Front right (FR)	F8, F4, FC6, FC2, T8
Occipital left (OL)	O1, P7, P3, CP5, CP1
Occipital right (OR)	O2, P4, P8, CP6, CP2

2.2. Preprocessing and Feature Extraction

The first step of preprocessing was to filter the raw EEG signals to remove unwanted artifacts from the signal. For filtering, we used a Butterworth bandpass filter from 0.5 Hz to 40 Hz.

In this way, we filtered out the line noise (50 Hz). Independent component analysis (ICA) was also used to remove artifacts such as eye movements [31]. The Infomax ICA algorithm was used for separating the original signal into independent components [32]. If there were waveforms between these components that were characteristic of eye movements, we removed these eye movements from the signal. Blinks were removed for all participants and left–right eye movements were removed whenever present (in approximately 50% of participants).

EEG features were calculated based on 10 s epochs with a five-second overlap between epochs. We computed the basic frequency-domain features: the relative power of the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–50 Hz) frequencies. They were calculated using the Thomson multitaper method [33] to obtain the power spectral density. We also used recurrence quantification analysis (RQA) [34–36] features: determinism (Det), laminarity (Lam), recurrence rate (RR), trapped time (TT), determinism divided by recurrence rate (Det/RR), longest diagonal (Lmax), longest vertical line (Vmax), average diagonal line length (Adll), divergence (Div), and entropy (Ent). The RQA features were calculated from the recurrence plot (RP) of the signal. RP is a 2D representation of the phase space trajectory of the signal [37]. It is a matrix with dimensions $N \times N$, where N is the length of the signal. The position (i, j) in the matrix is marked as one if the i -th and j -th points in the signal are close to each other.

The total number of extracted features was 15 per channel/region. We used four regions (with averaged feature value from 5 channels) and three additional channels (see above), which gave us a total of 105 features.

2.3. Drowsiness Detection

The 105 features described in Section 2.2 served as the initial feature set for the first drowsiness detection model. The aim of this analysis was to investigate whether sex as a feature could improve drowsiness detection.

The best combination of algorithms and hyperparameters was optimized using a grid search approach. The experiment was conducted using four classification algorithms (XGBoost, naïve Bayes, random forest, and support vector machines) and three feature selection algorithms (chi2, information gain, and ANOVA F-test). Table 3 shows the hyperparameters that were explored for classification algorithm selection on the training set. In addition, we also searched for the optimal number of features to include in the model (selection of 10, 20, 30, 40, 50, 60, and 100). Grid search was applied to all these hyperparameters, classification algorithms, and feature selection methods simultaneously. The decision was made based on maximizing the accuracy of a model without including sex as a feature.

Table 3. Table of optimization hyperparameters with a grid search approach. Naïve Bayes had no hyperparameters to optimize.

XGBoost	
Eta—{0.1, 0.3, 0.4, 0.9}	reg_alpha—{0, 0.5}
gamma—{0, 1, 5}	reg_lambda—{0.5, 1}
learning_rate—{0.05, 0.1, 0.5, 1}	
Random Forest	
n_estimators—{30, 100, 200, 500}	
min_samples_split—{2, 4, 6}	
Support Vector Machine	
C—{0.5, 1, 10}	kernel—{linear, rbf}

Once the best algorithm and feature set were obtained, the evaluation metrics for the drowsiness detection model without sex as a feature were calculated. The classification was performed for each epoch of each participant, resulting in a total of 17,000 epochs (both classes with 34 recording sessions, each session with 250 epochs) for drowsiness detection. A random 66% of the dataset was used for training the classifier and the remaining 33% was used for testing the classifier.

Then, sex as a feature was added to the dataset, and classification was performed in the same way. In the second analysis, to further investigate the influence of driver's sex on drowsiness detection, the dataset was divided into two subsets containing only male and female drivers, respectively. The classification was performed in the same way for these two subsets of data.

2.4. Sex Classification Model

A reliable EEG-based sex classification model was developed to make previous findings applicable in cases where the dataset does not contain information about drivers' sex. For the sex classification model, we divided each channel/region of each participant into nine segments (the first eight segments with 27 epochs and the ninth segment with the remaining 34 epochs). The values of each feature in each of these segments were averaged, as shown in Figure 1. We chose nine segments to increase our dataset for classification to 306 records (34 participants with 9 segments each), which was sufficient to properly split the dataset into training and test subsets. Moreover, each of the nine segments had enough epochs to compute correlations between features. The final feature set consisted of 105 features, explained in Section 2.2, and the correlations between each pair of features as new features. In total, this amounted to 5565 features. The training dataset consisted of each participant's six randomly selected segments and the test dataset consisted of each participant's remaining three segments. We also used leave-one-subject-out evaluation (LOSOVC). Hyperparameter optimization and algorithm selection were performed on the training dataset using 5-fold cross-validation. The same grid search approach as in Section 2.3 was used for hyperparameter optimization.

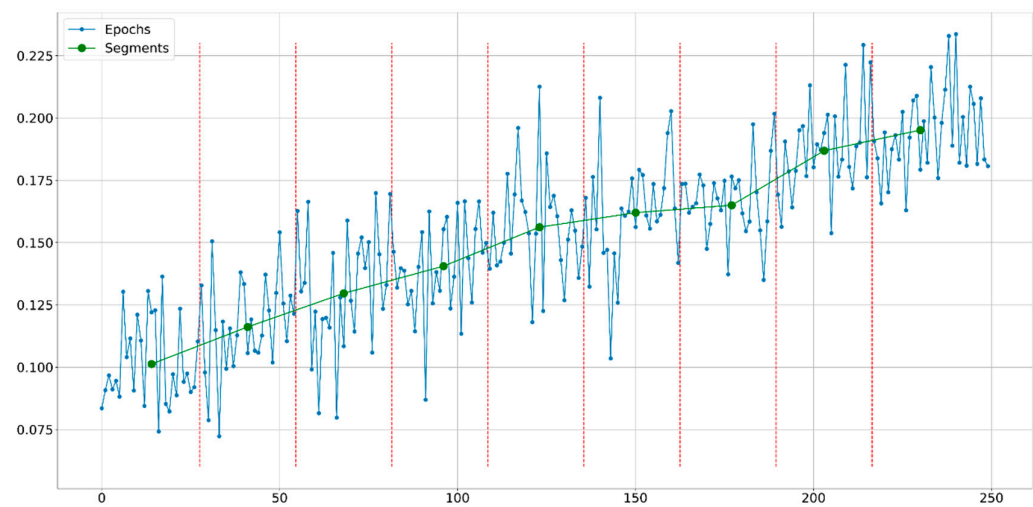


Figure 1. Illustration of segment averaging.

3. Results

3.1. Improvement of Drowsiness Detection with Addition of Sex as a Feature

Based on the optimization of grid search hyperparameters, XGBoost with default parameters, along with 50 features selected by the chi2 method, was selected for our drowsiness detection model. The decision was made based on maximizing the accuracy of a model without including sex as a feature.

Table 4 shows the evaluation metrics' scores of the drowsiness detection model with and without sex as a feature. Including sex as an additional feature in the feature set resulted in only an incremental improvement in the evaluation metrics. However, when the dataset was split into two datasets based on the sex information, it led to a significant improvement in the drowsiness detection evaluation metrics. Table 4 shows that the accuracy of the drowsiness detection is 84% for male drivers and 88% for female drivers, which is 3% and 7% better, respectively, than the classification without information about driver sex. Additionally, precision and recall are higher in both groups for both male and female drivers.

Table 4. Results of the drowsiness classification model.

	Without Sex Information		
	Precision	Recall	Accuracy
Alertness	0.81	0.82	0.81
Drowsiness	0.82	0.80	
	With sex information		
	Precision	Recall	Accuracy
Alertness	0.81	0.83	0.82
Drowsiness	0.82	0.81	
	Only male drivers		
	Precision	Recall	Accuracy
Alertness	0.82	0.86	0.84
Drowsiness	0.86	0.82	
	Only female drivers		
	Precision	Recall	Accuracy
Alertness	0.87	0.89	0.88
Drowsiness	0.89	0.87	

3.2. Model for Sex Classification

The best classification model for drivers' sex classification was XGBoost. The following hyperparameters were selected as the best ones and were the same for both groups: eta equals 0.1, gamma equals 0, max_depth equals 10, reg_alpha equals 0, reg_lambda equals 1, and learning_rate parameter equals 0.1. The chi2 method was selected as the feature selection method, and the 40 best-ranked features were selected for our final classification model.

Interestingly, among the selected features, there were no frequency-domain features. There were 36 selected RQA features and four selected correlation features. The same features were selected for all seven channels/regions—Det, Lam, Det/RR, Lmax, and Vmax. Besides these 35 features, the model also selected the RR feature from the Cz channel, the correlation between Cz Theta and Cz Delta, the correlation between FL Theta and Cz Gamma, the correlation between FR Theta and Cz Gamma, and the correlation between OL Theta and OL Delta. All the selected correlations for the drowsy group were between frequency-domain features. Among the selected features, the distribution of channels/regions was almost uniform, 5 ± 1 of all channels/regions were selected.

In general, the final accuracy of any classification model may vary with a different selection of the test set. Since we randomly selected our test set as described in Section 2.4, we applied the methodology described there five times (as seen in Tables 5 and 6) to verify the stability and robustness of the classification model.

Table 5. Classification accuracies on the test data for alert participants.

Participant	Accuracy						Target Class	Correctly Classified					
	Run1	Run2	Run3	Run4	Run5	Avg		Run1	Run2	Run3	Run4	Run5	Avg
1	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
2	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
3	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
4	0.67	1.00	1.00	0.67	1.00	0.87	M	1	1	1	1	1	1
5	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
6	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
7	0.67	1.00	1.00	1.00	1.00	0.93	M	1	1	1	1	1	1
8	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
9	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
10	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
11	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
12	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
13	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
14	1.00	1.00	0.67	1.00	1.00	0.93	M	1	1	1	1	1	1
15	1.00	1.00	1.00	1.00	0.67	0.93	M	1	1	1	1	1	1
16	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
17	1.00	1.00	0.67	0.67	0.67	0.80	F	1	1	1	1	1	1
18	1.00	1.00	0.67	0.67	1.00	0.87	M	1	1	1	1	1	1
19	1.00	1.00	1.00	0.67	1.00	0.93	F	1	1	1	1	1	1
20	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
21	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
22	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
23	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
24	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
25	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
26	0.67	0.67	1.00	1.00	0.67	0.80	M	1	1	1	1	1	1
27	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
28	1.00	0.67	0.67	1.00	1.00	0.87	M	1	1	1	1	1	1
29	0.33	0.67	0.67	1.00	0.67	0.67	M	0	1	1	1	1	0.8
30	0.33	0.33	1.00	0.67	0.67	0.60	F	0	0	1	1	1	0.6
31	1.00	0.67	0.67	0.67	0.67	0.73	F	1	1	1	1	1	1
32	1.00	1.00	0.33	0.33	1.00	0.73	F	1	1	0	0	1	0.6

Table 5. Cont.

Participant	Accuracy						Target Class	Correctly Classified					
	Run1	Run2	Run3	Run4	Run5	Avg		Run1	Run2	Run3	Run4	Run5	Avg
33	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
34	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
Average:	0.93	0.94	0.92	0.92	0.94	0.93	Sum:	32	33	33	33	34	33
							Accuracy:	0.94	0.97	0.97	0.97	1.00	0.97

Table 6. Classification accuracies on the test data for drowsy participants.

Participant	Accuracy						Target Class	Correctly Classified					
	Run1	Run2	Run3	Run4	Run5	Avg		Run1	Run2	Run3	Run4	Run5	Avg
1	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
2	0.67	1.00	0.33	1.00	0.67	0.73	F	1	1	0	1	1	0.8
3	1.00	1.00	0.67	1.00	1.00	0.93	F	1	1	1	1	1	1
4	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
5	1.00	0.67	1.00	0.67	0.67	0.80	M	1	1	1	1	1	1
6	1.00	1.00	1.00	1.00	0.67	0.93	M	1	1	1	1	1	1
7	1.00	0.67	1.00	1.00	0.67	0.87	F	1	1	1	1	1	1
8	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
9	1.00	1.00	0.67	1.00	1.00	0.93	F	1	1	1	1	1	1
10	0.67	1.00	1.00	0.67	0.67	0.80	M	1	1	1	1	1	1
11	0.33	1.00	0.33	1.00	0.67	0.67	M	0	1	0	1	1	0.6
12	1.00	1.00	0.67	1.00	0.67	0.87	M	1	1	1	1	1	1
13	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
14	1.00	1.00	1.00	0.67	1.00	0.93	F	1	1	1	1	1	1
15	0.67	1.00	1.00	0.67	0.67	0.80	M	1	1	1	1	1	1
16	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
17	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
18	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
19	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
20	0.67	0.67	0.67	1.00	0.33	0.67	F	1	1	1	1	0	0.8
21	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
22	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
23	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
24	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
25	0.67	0.67	1.00	0.33	1.00	0.73	M	1	1	1	0	1	0.8
26	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
27	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
28	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
29	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
30	1.00	0.67	0.33	1.00	0.67	0.73	F	1	1	0	1	1	0.8
31	1.00	0.67	1.00	1.00	1.00	0.93	F	1	1	1	1	1	1
32	1.00	1.00	1.00	1.00	1.00	1.00	M	1	1	1	1	1	1
33	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
34	1.00	1.00	1.00	1.00	1.00	1.00	F	1	1	1	1	1	1
Average:	0.93	0.94	0.90	0.94	0.89	0.92	Sum:	33	34	31	33	33	32.8
							Accuracy:	0.97	1.00	0.91	0.97	0.97	0.96

Tables 5 and 6 show the accuracy of the classification models for alert and drowsy participants, respectively. Each participant had three randomly selected segments in the test set. The classification accuracy of these segments is marked with blue. If at least two out of three of these segments have a correctly classified sex (accuracy 0.67 or 1.00), the participant is marked as correctly classified in the right, green part of the table. The average classification accuracy was calculated for each participant and each run of the methodology. The average classification accuracy of the segments (marked in blue) was 93% (alert participants) and 92% (drowsy participants), while the average classification

accuracy of the participants (marked in green) was 97% (alert participants) and 96% (drowsy participants). The classification accuracy with LOSOCV was 82%.

4. Discussion

Splitting the dataset into two subsets, male and female, resulted in a significant improvement in drowsiness detection (confirmed with Mann–Whitney U test). After adding sex as a feature to the dataset, drowsiness detection was only incrementally improved, but splitting into two sex-dependent datasets yielded a 3% and 7% improvement for male and female drivers, respectively. Precision and recall were higher for alert and drowsy states for both male and female subsets. Interestingly, precision and recall were higher for female drivers than for male drivers. This increase in prediction accuracy, precision, and recall for female drivers suggests differential changes in EEG activity associated with drowsiness compared to male drivers.

Theoretically, approaches using sex as a feature and manual split of the dataset could have the same accuracy. The reason why this is not the case could be the relatively high dimensionality of the dataset and the relatively small number of examples in the dataset. For this reason, the algorithm was unable to optimize the hyperplane to artificially separate the data based on the sex feature.

A recent review of EEG signal features and their application in driver drowsiness detection systems summarized 39 papers, none of which used participant's sex as a feature [13]. Based on our results, it is reasonable to assume that the model accuracy presented in many of these papers could be further increased by using sex as a feature or by splitting the dataset into male and female subsets.

For these results to be applicable to all datasets, a high-accuracy sex classifier is needed. The average classification accuracy of the segment's sex (blue markers in Tables 5 and 6) is 93% and 92% for alert and drowsy drivers, respectively. On average, the classifiers correctly classify 97% of the alert participants and 96% of the drowsy participants (green markers in Tables 5 and 6). When majority voting is applied to each participant across all five runs, the accuracy of both classifiers is 100%.

Other studies with high accuracy of sex classification models based on EEG were those of Kaur et al. [38] and Kaushik et al. [39]. Their classification accuracy was 96.7% and 97.5%, respectively. The experimental design was the same in both papers, participants were measured in a relaxed resting position with their eyes closed. In both works, the discrete wavelet transform was used to obtain the frequency-domain features. Their final classifiers were based only on these frequency-domain features. In comparison, our participants were measured while driving, which is a complex mental activity. Kaur et al. achieved their accuracy with the usage of LOSOCV on the 60 participants. Their high accuracy suggests that a higher number of participants could increase our LOSOCV closer to the segments split based accuracy.

Our previous work showed that frequency-domain features and RQA features differ significantly between alert male and female drivers [26]. Similar statistical results were found when brain activity was analyzed between male and female drowsy drivers. Female drowsy drivers showed significantly higher relative beta power in all regions and significantly higher relative alpha power in all regions except the Cz electrode. Male drowsy drivers showed significantly higher relative delta power in all regions and relative theta power in the occipital region. On the other hand, relative gamma power showed a different pattern between male and female drowsy drivers. In female drowsy drivers, relative gamma power was significantly higher in the occipital region, whereas in male drowsy drivers, relative gamma power was higher in the frontal region. Furthermore, the results of this paper showed that the features in the frequency domain are more correlated in males than in females during both alert and drowsy sessions.

In the current work, Figure 2 shows the difference between the average feature value for all male and all female drowsy drivers for two RQA features with the smallest *p*-values (*V*_{max} in the OR region and *L*_{max} in the Pz channel) and two frequency-domain features

(beta in the OR region and the OL region). Figure 3 shows a high correlation (0.79) between the Det feature from the FR region and the Lam feature from the Cz channel for male drowsy drivers. The right part of Figure 3 shows the same features for female drowsy drivers, which are only weakly correlated (0.37). These differences in the correlation of the features are the main reason for their use as features.

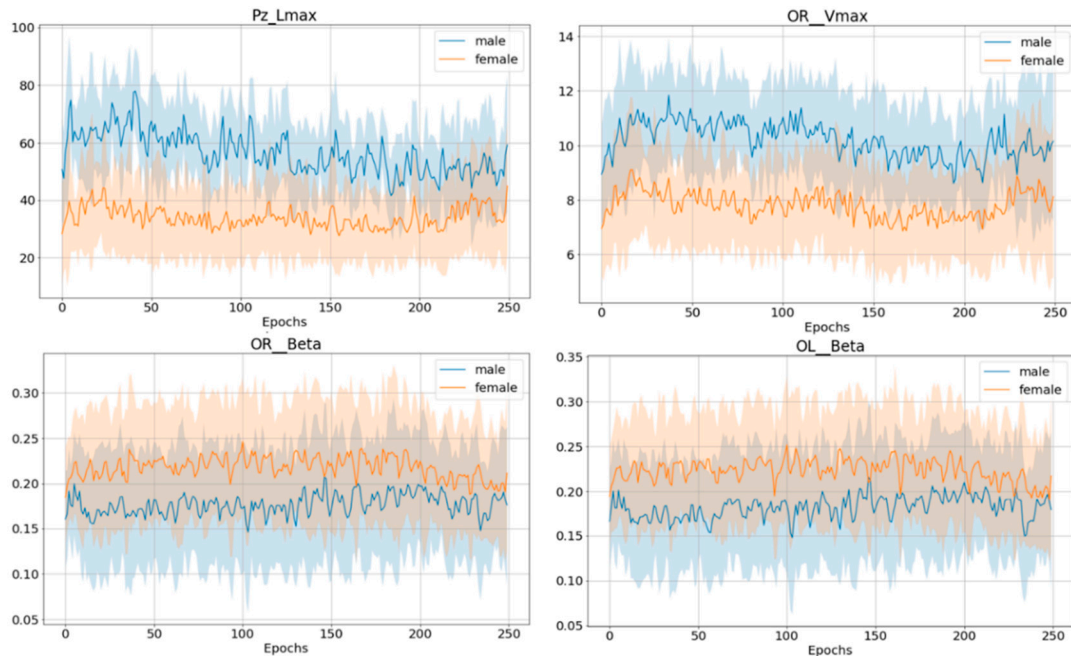


Figure 2. The difference between the average features of all male and all female drowsy drivers: Vmax from the OR region, Lmax from the Pz channel, relative beta power from the OR region, and relative beta power from the OL region. The Y-axis shows the values of observed features. The bold line is the mean of the feature, and the filled area represents the standard deviation.

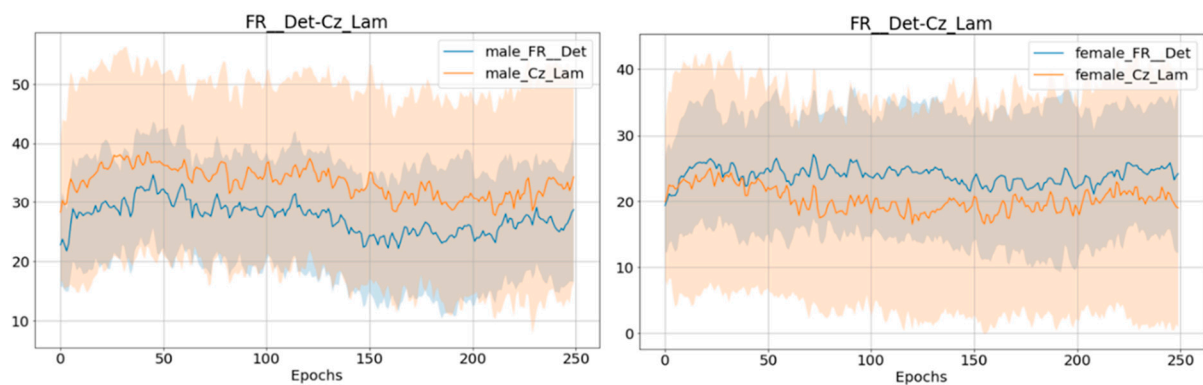


Figure 3. Left: highly correlated (0.79) determinism (Det) from the FR region and laminarity (Lam) from the Cz channel for male drivers; right: weakly correlated (0.37) determinism (Det) from the FR region and laminarity (Lam) from the Cz channel for female drivers. The Y-axis shows the values of the observed features. The bold line is the mean of the feature, and the filled area represents the standard deviation.

It should be noted that, although our dataset consisted of frequency-domain features, our feature selection method filtered out all the frequency-domain features for sex classification. The final feature sets for the sex classification model thus consisted of only RQA features and correlation features. Since RQA features discriminate male and female drivers better than frequency-domain features in our work, it is reasonable to assume that the studies (Kaur et al. [38] and Kaushik et al. [39]) reporting high accuracy of the sex

classification models would report even higher results with the inclusion of RQA features, but further research should be conducted to confirm this assumption.

Our system is based on the 10 s epochs with 5 s overlaps between them so the application in the real-world scenario could make a decision every 5 s. Classifiers make their decisions in less than a second, which means that if epochs would have a step of one second instead of five, the system would still be able to make decisions on time.

There are also some drawbacks to this work. One is the not very large dataset with 34 participants in 68 recording sessions. This results in a lower LOSOCV (82%) and the increase in the number of participants would increase accuracy. Another drawback is the exclusive use of binary drowsiness classification. In real-world applications, at least three levels of drowsiness are usually targeted [39,40].

For future work, the number of participants and the number of features considered should be increased. For example, an interesting point would be to observe driving performance data (e.g., line crossings, distance from an ideal position on the road) based on participants' sex and EEG features. Since our data suggest functional differences between males and females during drowsy driving, the next step would be to investigate whether these differences are related to driving performance. Investigating the relation between driving performance and EEG features could provide insight into how people drive and explain potential differences. The underlying mechanisms related to driving would therefore provide a more accurate model for driving safety systems. Another interesting topic for future work is to investigate the influence of driver's sex as a feature on deep learning models that have lately been used extensively and are showing promising results [40]. In addition, combining our findings (sex as a feature and RQA features as good drowsiness predictors) with decreasing the number of electrodes used [41] could lead to a reliable system that is also easy to implement in practice. Such a system's accuracy could also benefit from the usage of blink-related features derived from EEG [42].

5. Conclusions

This research has shown that including the information about driver's sex increases the accuracy of drowsiness detection. Furthermore, a reliable sex classifier based on EEG signals was developed. Although it is hard to implement exactly the same system in a real-time environment, due to the high number of electrodes, these important findings may benefit all other systems that are less intrusive simply by including sex as a feature in the existing systems.

The drowsiness detection model for drivers is usually based on the EEG features and without sex as a feature. After adding sex as an additional feature in the dataset, only incremental improvements in drowsiness detection accuracy were achieved. With the further step of manually splitting the dataset into male and female subsets, the drowsiness detection model accuracy increased by 3% and 7% for male and female datasets, respectively. We consider these results relevant to a variety of drowsiness detection systems currently being developed in the field.

The sex classification model based on EEG features achieved high accuracy. All participants were correctly classified after applying majority voting to the results of all five runs. Correlations between features used as features scored high on the feature selection list, suggesting that correlations between features from different brain regions/channels should be used more frequently as features.

Author Contributions: Conceptualization, I.S., M.Z.Z., K.F., M.C. and A.J.; Methodology, I.S., M.Z.Z., K.F. and A.J.; Software, I.S.; Validation, I.S., M.Z.Z. and A.J.; Formal Analysis, M.C. and A.J.; Investigation, I.S., M.Z.Z. and K.F.; Resources, M.C. and A.J.; Data Curation, I.S., M.Z.Z. and K.F.; Writing—Original Draft Preparation, I.S. and M.Z.Z.; Writing—Review and Editing, I.S., M.Z.Z., K.F., M.C. and A.J.; Visualization, I.S.; Supervision, M.C. and A.J.; Project Administration, M.C. and A.J.; Funding Acquisition, M.C. and A.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been carried out within the project “Research and development of the system for driver drowsiness and distraction identification—DFDM” (KK.01.2.1.01.0136), funded by the European Regional Development Fund in the Republic of Croatia under the Operational Programme Competitiveness and Cohesion 2014–2020.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the University of Zagreb, Faculty of Electrical Engineering and Computing (on 26 September 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Restrictions apply to the availability of these data. The data used in this work were obtained from Xylon d.o.o., Croatia, to which any request for data should be addressed (info@logicBRICKS.com).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ledger, S.; Bennett, J.M.; Chekaluk, E.; Batchelor, J.; Di Meco, A. Cognitive function and driving in middle adulthood: Does age matter? *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *66*, 471–484. [\[CrossRef\]](#)
2. Chuang, C.-H.; Cao, Z.; King, J.-T.; Wu, B.-S.; Wang, Y.; Lin, C.-T. Brain Electrodynamics and Hemodynamic Signatures against Fatigue during Driving. *Front. Neurosci.* **2018**, *12*, 181. [\[CrossRef\]](#) [\[PubMed\]](#)
3. National Highway Traffic Safety Administration. *Drowsy Driving 2015*; National Highway Traffic Safety Administration: Washington, DC, USA, 2017.
4. Owens, J.M.; Dingus, T.A.; Guo, F.; Fang, Y.; Perez, M.; McClafferty, J.; Tefft, B. *Prevalence of Drowsy Driving Crashes: Estimates from a Large-Scale Naturalistic Driving Study*; Foundation for Traffic Safety: Washington, DC, USA, 2018.
5. European Commission. *Commission Delegated Regulation (EU) 2021/1341*; Official Journal of the European Union; European Commission: Brussels, Belgium, 2021.
6. Albadawi, Y.; Takruri, M.; Awad, M. A Review of Recent Developments in Driver Drowsiness Detection Systems. *Sensors* **2022**, *22*, 2069. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Chowdhury, A.; Shankaran, R.; Kavakli, M.; Haque, M. Sensor Applications and Physiological Features in Drivers’ Drowsiness Detection: A Review. *IEEE Sens. J.* **2018**, *18*, 3055–3067. [\[CrossRef\]](#)
8. Burlacu, A.; Brinza, C.; Brezulanu, A.; Covic, A. Accurate and early detection of sleepiness, fatigue and stress levels in drivers through Heart Rate Variability parameters: A systematic review. *Rev. Cardiovasc. Med.* **2021**, *22*, 845–852. [\[CrossRef\]](#)
9. Siddiqui, H.; Saleem, A.; Brown, R.; Bademci, B.; Lee, E.; Rustam, F.; Dudley, S. Non-Invasive Driver Drowsiness Detection System. *Sensors* **2021**, *21*, 4833. [\[CrossRef\]](#)
10. Gielen, J.; Aerts, J.-M. Feature Extraction and Evaluation for Driver Drowsiness Detection Based on Thermoregulation. *Appl. Sci.* **2019**, *9*, 3555. [\[CrossRef\]](#)
11. Tian, Y.; Cao, J. Fatigue driving detection based on electrooculography: A review. *EURASIP J. Image Video Process.* **2021**, *2021*, 1–17. [\[CrossRef\]](#)
12. Johnson, R.R.; Popovic, D.P.; Olmstead, R.E.; Stikic, M.; Levendowski, D.J.; Berka, C. Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model. *Biol. Psychol.* **2011**, *87*, 241–250. [\[CrossRef\]](#)
13. Stancin, I.; Cifrek, M.; Jovic, A. A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems. *Sensors* **2021**, *21*, 3786. [\[CrossRef\]](#)
14. Tran, Y.; Craig, A.; Craig, R.; Chai, R.; Nguyen, H. The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analyses. *Psychophysiology* **2020**, *57*, e13554. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Lal, S.K.; Craig, A. A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* **2001**, *55*, 173–194. [\[CrossRef\]](#)
16. Stancin, I.; Frid, N.; Cifrek, M.; Jovic, A. EEG Signal Multichannel Frequency-Domain Ratio Indices for Drowsiness Detection Based on Multicriteria Optimization. *Sensors* **2021**, *21*, 6932. [\[CrossRef\]](#)
17. Soares, S.; Monteiro, T.; Lobo, A.; Couto, A.; Cunha, L.; Ferreira, S. Analyzing Driver Drowsiness: From Causes to Effects. *Sustainability* **2020**, *12*, 1971. [\[CrossRef\]](#)

18. Tonetti, L.; Fabbri, M.; Natale, V. Sex Difference in Sleep-Time Preference and Sleep Need: A Cross-Sectional Survey among Italian Pre-Adolescents, Adolescents, and Adults. *Chronobiol. Int.* **2008**, *25*, 745–759. [\[CrossRef\]](#)
19. Corsi-Cabrera, M.; Sánchez, A.I.; Del-Río-Portilla, Y.; Villanueva, Y.; Pérez-García, E. Effect of 38 h of total sleep deprivation on the waking EEG in women: Sex differences. *Int. J. Psychophysiol.* **2003**, *50*, 213–224. [\[CrossRef\]](#)
20. Borghini, G.; Astolfi, L.; Vecchiato, G.; Mattia, D.; Babiloni, F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* **2012**, *44*, 58–75. [\[CrossRef\]](#)
21. Chen, L.-L.; Zhao, Y.; Zhang, J.; Zou, J.-Z. Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Syst. Appl.* **2015**, *42*, 7344–7355. [\[CrossRef\]](#)
22. Cai, Q.; Gao, Z.-K.; Yang, Y.-X.; Dang, W.-D.; Grebogi, C. Multiplex Limited Penetrable Horizontal Visibility Graph from EEG Signals for Driver Fatigue Detection. *Int. J. Neural Syst.* **2019**, *29*, 1850057. [\[CrossRef\]](#)
23. Min, J.; Wang, P.; Hu, J. Driver fatigue detection through multiple entropy fusion analysis in an EEG-based system. *PLoS ONE* **2017**, *12*, e0188756. [\[CrossRef\]](#)
24. Nguyen, T.; Ahn, S.; Jang, H.; Jun, S.C.; Kim, J.G. Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Sci. Rep.* **2017**, *7*, 43933. [\[CrossRef\]](#)
25. Lin, C.-T.; King, J.-T.; Chuang, C.-H.; Ding, W.; Chuang, W.-Y.; Liao, L.-D.; Wang, Y. Exploring the Brain Responses to Driving Fatigue through Simultaneous EEG and fNIRS Measurements. *Int. J. Neural Syst.* **2019**, *30*, 1950018. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Stancin, I.; Friganovic, K.; Zeba, M.Z.; Jovic, A.; Cifrek, M. Gender Differences in EEG Features While Driving. In Proceedings of the 2nd International Conference on Advances in Signal Processing and Artificial Intelligence, Berlin, Germany, 30 June–2 July 2020; pp. 125–130.
27. Chen, J.; Wang, H.; Wang, Q.; Hua, C. Exploring the fatigue affecting electroencephalography based functional brain networks during real driving in young males. *Neuropsychologia* **2019**, *129*, 200–211. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Chen, J.; Wang, H.; Hua, C. Electroencephalography based fatigue detection using a novel feature fusion and extreme learning machine. *Cogn. Syst. Res.* **2018**, *52*, 715–728. [\[CrossRef\]](#)
29. Hu, J.; Min, J. Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model. *Cogn. Neurodynamics* **2018**, *12*, 431–440. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Dimitrakopoulos, G.N.; Kakkos, I.; Dai, Z.; Wang, H.; Sgarbas, K.; Thakor, N.; Bezerianos, A.; Sun, Y. Functional Connectivity Analysis of Mental Fatigue Reveals Different Network Topological Alterations between Driving and Vigilance Tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 740–749. [\[CrossRef\]](#)
31. Makeig, S.; Bell, A.J.; Jung, T.-P.; Sejnowski, T.J. Independent Component Analysis of Electroencephalographic Data. In Proceedings of the 8th International Conference on Neural Information Processing Systems, Denver, CO, USA, 27–30 November 1995; pp. 145–151.
32. Bell, A.J.; Sejnowski, T.J. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Comput.* **1995**, *7*, 1129–1159. [\[CrossRef\]](#)
33. Thomson, D.J. Spectrum estimation and harmonic analysis. *Proc. IEEE* **1982**, *70*, 1055–1096. [\[CrossRef\]](#)
34. Shabani, H.; Mikaili, M.; Noori, S.M.R. Assessment of recurrence quantification analysis (RQA) of EEG for development of a novel drowsiness detection system. *Biomed. Eng. Lett.* **2016**, *6*, 196–204. [\[CrossRef\]](#)
35. Webber, C.; Zbilut, J. Recurrence Quantification Analysis of Nonlinear Dynamical Systems. In *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*; National Science Foundation: Alexandria, VA, USA, 2005.
36. Acharya, U.R.; Sree, S.V.; Chattopadhyay, S.; Yu, W.; Ang, P.C.A. Application of recurrence quantification analysis for the automated identification of epileptic eeg signals. *Int. J. Neural Syst.* **2011**, *21*, 199–211. [\[CrossRef\]](#)
37. Eckmann, J.-P.; Kamphorst, S.O.; Ruelle, D. Recurrence Plots of Dynamical Systems. *Eur. Lett.* **1987**, *4*, 973–977. [\[CrossRef\]](#)
38. Kaur, B.; Singh, D.; Roy, P.P. Age and gender classification using brain–computer interface. *Neural Comput. Appl.* **2018**, *31*, 5887–5900. [\[CrossRef\]](#)
39. Kaushik, P.; Gupta, A.; Roy, P.P.; Dogra, D.P. EEG-Based Age and Gender Prediction Using Deep BLSTM-LSTM Network Model. *IEEE Sens. J.* **2018**, *19*, 2634–2641. [\[CrossRef\]](#)
40. Chaabene, S.; Bouaziz, B.; Boudaya, A.; Hökelmann, A.; Ammar, A.; Chaari, L. Convolutional Neural Network for Drowsiness Detection Using EEG Signals. *Sensors* **2021**, *21*, 1734. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Ko, L.-W.; Komarov, O.; Lai, W.-K.; Liang, W.-G.; Jung, T.-P. Eyeblink recognition improves fatigue prediction from single-channel forehead EEG in a realistic sustained attention task. *J. Neural Eng.* **2020**, *17*, 036015. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Shahbakhti, M.; Beiramvand, M.; Rejer, I.; Augustyniak, P.; Broniec-Wojcik, A.; Wierzbach, M.; Marozas, V. Simultaneous Eye Blink Characterization and Elimination from Low-Channel Prefrontal EEG Signals Enhances Driver Drowsiness Detection. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 1001–1012. [\[CrossRef\]](#)

Publication 4

I. Stancin, K. Friganovic, M. Z. Zeba, M. Cifrek, and A. Jovic, “Gender differences in EEG features while driving,” Proceedings of the 2nd International Conference on Advances in Signal Processing and Artificial Intelligence, p. 127, Nov. 2020

Gender differences in EEG features while driving

I. Stančin¹, K. Friganović¹, M. Zelenika Zeba², A. Jović¹ and M. Cifrek¹

¹ University of Zagreb, Faculty of Electrical Engineering and Computing., Unska 3, Zagreb, Croatia

² Polyclinic SUVAG, Ul. kneza Ljudevita Posavskog 10, Zagreb, Croatia

Tel.: + 38516129-554

E-mail: igor.stancin@fer.hr

Abstract: Gender differences in traffic are generally analyzed through the number of accidents reported to the police. Our research aims to observe gender differences in electroencephalographical signals (EEG) while participants are driving. Time-frequency domain features and recurrence quantification analysis (RQA) features are calculated in order to analyze the differences. To compare male and female drivers, we used Mann–Whitney U test and compared correlations between features and brain regions. Female drivers showed significantly higher beta relative power in the occipital right region and significantly higher alpha relative power in the frontal regions, while male drivers showed significantly higher theta relative power in all regions except in the front right region. Most RQA features show a significant difference between male and female drivers. Also, male drivers showed significantly higher correlations between the RQA features, especially between different brain regions. These results could reflect the differences in the information processing strategies of male and female drivers, e.g. they tend to focus on different information when performing the task. That could account for reported gender differences in the number of traffic accidents and traffic behavior.

Keywords: gender difference, EEG features, driving, recurrence quantification analysis

1. Introduction

Gender differences in traffic are generally analyzed through the number of accidents reported to the police [1]. In Ireland, in 2001, it was ten times more likely for a male driver to be killed in an accident than for a female driver [2]. Although the number of female drivers is still increasing, this difference in the number of accidents is still large [3]. The argument used against gender difference in driving performance is that male drivers drive more often and more kilometers [4], while other research concludes that even with taking that into account, there are still more accidents involving male drivers [5].

Behavioral and neurophysiological differences between genders should also be examined. International Transportation Forum reports that, in 2018, male drivers still run higher risks while driving [6]. Another indirect reason for this kind of gender difference is that spatial orientation tasks are harder for women [7], which makes them more careful in traffic. Additionally, other researchers concluded that female drivers are involved in accidents due to errors of perceptual nature and judgment error [8].

Gender differences based on the traffic statistics [2], [3], [5], [6] and from experiments while participants performing simple tasks [7], [9]–[12] are obvious. With our research, we aim to observe gender differences in electroencephalographical signals (EEG) while participants perform a complex task – drive a car, in order to provide a deeper understanding of differences between genders while driving, and in general.

2. Methodology

2.1. Experiment design

In our study, we used a research-grade EEG recorder with 32 channels for recording signals from 14 healthy participants (seven males and seven females) during a driving simulation. Although we used 32-channel EEG, we wanted to describe dependencies and differences between certain brain regions – front left (FL), front right (FR), occipital left (OL), and occipital right (OR) regions. We calculated the mean value for each feature calculated from five electrodes in that region (shown in Table 1). In addition to channels from these four regions, we also used Oz, Pz and Cz channels in our analysis.

Simulation scenario was shown on the screen. Steering the car was done using a professional steering wheel joystick. All participants were instructed to drive following traffic regulations. Male participants were 27.14 years old on average with a standard deviation of 2.94 years, while female participants were 26.0 years old on average with a standard deviation of 3.66 years. The driving scenario was the same for all

Table 1. Electrodes in brain region

Region	Electrodes
Front left (FL)	F7, F3, FC5, FC1, T7
Front right (FR)	F8, F4, FC6, FC2, T8
Occipital left (OL)	O1, P7, P3, CP5, CP1
Occipital right (OR)	O2, P4, P8, CP6, CP2

participants and consisted of driving on both state and highway roads, and in an urban city environment, for around 40 minutes. The recording sessions were held in the afternoon.

EEG features were calculated based on 10 seconds epochs with five seconds of overlap between epochs. To represent differences in EEG features between male and female drivers, we averaged the values of each feature in each epoch for all male and female drivers, respectively, thus constructing time-series of each feature for both genders.

For each feature, we calculated the Mann–Whitney U test between the time-series of both genders, with the usage of $\alpha_0 = 0.001$. Since we repeated our test 105 times, we also used Bonferroni correction to reduce the chances for type I error, which gave us $\alpha = 9.26 \times 10^{-6}$. Also, the correlation for the time-series of each feature between the two genders was calculated and compared.

2.2. Filtering and EEG features

The first step of preprocessing was to filter raw EEG signals in order to remove unwanted artifacts from the signal. For filtering, we used a bandpass filter from 0.5 Hz to 40 Hz. In this way, we filtered out line noise (50 Hz). Independent component analysis (ICA) was also used for removing artifacts like eye movements.

We calculated basic time-frequency domain features: relative power of alpha (8-12 Hz), beta (12-30 Hz), theta (4-8 Hz), delta (0.5-4 Hz) and gamma (30-50 Hz) frequencies. They were calculated with the usage of the Thomson multitaper method [13] for obtaining power spectral density. We also used recurrence quantification analysis (RQA) features: determinism (Det), laminarity (Lam), recurrence rate (RR), trapped time (TT), determinism divided by recurrence rate (Det/RR), longest diagonal (Lmax), longest vertical line (Vmax), average diagonal line length (Adll), divergence (Div) and entropy (Ent). RQA features were calculated from the recurrence plot (RP) of the signal. RP is a 2D representation of the phase space trajectory of the signal [14]. It is a matrix of dimensions $N \times N$, where N is the length of the signal. Position (i, j) in the matrix is marked with one if i -th and j -th point in the signal are close to each other.

The total number of features extracted was 15 per channel/region. We used three channels and four regions, which gave us a total of 105 features.

3. Results

Although we used the conservative $\alpha_0 = 0.001$ with Bonferroni correction, we still got 93 features with a p value smaller than α . Table 2 shows p values for all features. Among 12 features that have p value larger than α , eight are from the occipital regions (six from the Oz channel); 10 of them are from Oz, Pz or Cz channels; and only one is from the frontal regions of the brain. Among 50 features with the smallest p values, there are mostly RQA features, with a few exceptions – five relative beta powers, one relative

alpha power, and one relative theta power. Among the first nine features with equal (and smallest) p value, four of them are Lam and three of them are RR. Figure 1 shows two RQA features with the smallest p values and two time-frequency domain features with the smallest p values. We can see that relative beta power is higher for female drivers than for male drivers on the Pz channel and in the OR region, which is also true for all other channels and regions. Female drivers also have significantly larger values for relative alpha power in the FR region, FL region and on the Cz channel, while all other channels and regions have similar values. Male drivers have significantly larger values for relative theta power in all regions, except on the Oz channel and in the FR region. Male drivers have higher values for the RR on all channels and in all regions, except for Cz, where female drivers have higher values (RR on the Pz channel shown in Figure 1).

Figure 2 shows all correlations for male and female drivers. Female drivers have a lot more white color present in Figure 2, which represents weak correlations (correlation between -0.4 and 0.4). Male drivers, in general, show much stronger correlations between different features than female drivers, especially between different regions of the brain. Even in cases where female drivers show some correlations between different channels and regions of the brain, male drivers, in general, have stronger correlations on these same channels and in the same regions. Figure 3 shows an example of a strong correlation between Det in the FR region and Lam in the FL region for male drivers, together with a weak correlation of the same features for female drivers.

4. Discussion

Brain electrical activity, as assessed by relative power and RQA features, was significantly different between male and female drivers. Female drivers showed significantly higher beta relative power in the OR region and significantly higher alpha relative power in the frontal regions, while male drivers showed significantly higher theta relative power in all regions, except in the FR region. In addition, male drivers showed significantly higher correlations between the RQA features.

The higher alpha and beta relative power reported for women in other studies [10], [15] was confirmed by the present results. In addition, observed relative alpha power in our sample was the largest for the FR and FL regions, and Cz channel. Obtained results are in line with an fMRI study which found that females showed activation in frontal and parietal regions as they performed the spatial-cognition performance task, while men showed distinct activation of the left hippocampus [16]. Furthermore, the absence of a

Table 2. *p* values for all features

Feature	<i>p</i> value	Feature	<i>p</i> value	Feature	<i>p</i> value	Feature	<i>p</i> value
Pz_RR	1.15E-89	Pz_Det/RR	3.42E-89	Oz_Lam	3.04E-72	Cz_Adll	8.09E-25
Pz_Beta	1.15E-89	OR_Det	5.46E-89	OR_Theta	6.71E-72	Pz_Delta	1.67E-23
FL_Lam	1.15E-89	FL_Det/RR	1.42E-88	Cz_Theta	2.52E-70	Oz_Det	2.21E-23
Pz_Lam	1.15E-89	OL_Adll	1.90E-88	Cz_Alpha	1.22E-69	Oz_Alpha	5.45E-23
OL_Lmax	1.15E-89	Pz_Div	3.64E-88	FR_Beta	8.22E-69	Oz_Vmax	2.41E-21
OL_RR	1.15E-89	OR_Div	4.44E-88	Pz_Ent	1.33E-68	FL_Gamma	2.30E-17
OL_Lam	1.15E-89	OL_Ent	5.70E-88	Cz_Div	8.81E-68	FR_Div	8.22E-17
OR_RR	1.15E-89	FR_Vmax	5.87E-88	Cz_Delta	2.57E-65	FR_Gamma	4.93E-16
OR_Lam	1.15E-89	OL_Det/RR	8.22E-88	FR_Det	3.62E-64	FL_Delta	7.96E-15
FL_Vmax	1.16E-89	OR_Det/RR	8.49E-88	Pz_TT	9.11E-61	Oz_Theta	3.51E-08
OL_Div	1.16E-89	OL_TT	2.76E-87	FR_Adll	6.37E-59	Oz_Det/RR	6.42E-08
OR_Beta	1.17E-89	OR_Adll	4.95E-87	OR_Delta	1.56E-58	Oz_Delta	4.91E-06
FL_RR	1.18E-89	FL_Adll	1.37E-86	OL_Delta	1.15E-51	OL_Gamma	2.12E-04
FL_TT	1.19E-89	FL_Div	3.14E-86	FR_TT	2.93E-51	Oz_Div	1.03E-03
FL_Lmax	1.23E-89	Cz_Det/RR	6.43E-86	Cz_Lam	1.45E-50	OR_Gamma	2.21E-03
OR_Lmax	1.28E-89	OL_TT	1.42E-85	Oz_Lmax	4.43E-50	Oz_Beta	2.48E-03
OL_Det	1.29E-89	FL_Ent	6.69E-85	Cz_Lmax	7.31E-45	Oz_Ent	3.33E-03
OL_Beta	1.30E-89	OR_Ent	1.02E-84	FR_Delta	3.57E-42	Cz_Vmax	4.79E-03
Cz_Beta	1.33E-89	FL_Beta	1.49E-84	FL_Theta	5.45E-41	Oz_Gamma	4.92E-03
OL_Vmax	1.33E-89	Cz_RR	1.31E-82	Cz_Gamma	3.93E-38	Oz_Adll	1.00E-02
OR_Vmax	1.52E-89	Oz_RR	1.57E-82	FR_Det/RR	5.04E-34	Cz_TT	1.23E-02
FR_Lam	1.65E-89	FL_Alpha	8.39E-81	OR_Alpha	7.68E-34	Pz_Alpha	2.65E-01
FL_Det	1.74E-89	Pz_Theta	6.07E-80	OL_Alpha	2.72E-33	FR_Theta	5.86E-01
Pz_Det	2.02E-89	Pz_Vmax	1.27E-77	Oz_TT	8.58E-33	Cz_Ent	6.76E-01
FR_RR	2.27E-89	FR_Alpha	3.41E-76	Pz_Gamma	3.99E-28		
FR_Lmax	2.37E-89	OL_Theta	4.12E-74	FR_Ent	4.71E-28		
Pz_Lmax	3.15E-89	Pz_Adll	6.64E-73	Cz_Det	5.81E-27		

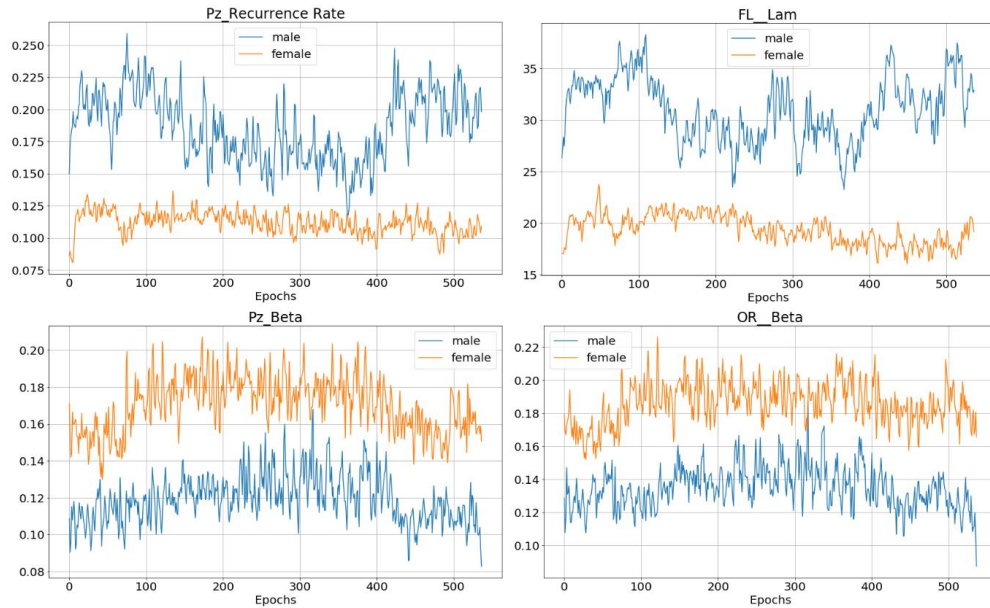


Fig. 1. Difference between male and female drivers: recurrence rate from Pz channel, laminarity (Lam) from FL region, relative alpha power from FR region and relative beta power from the Pz channel. Y-axis shows values of observed features.

reduction in relative alpha power has been related to information processing differences between men and women in spatial tasks [10].

Our results showed a higher contribution of theta relative power only in men. Increased relative theta power has been related to more active task-related processing [17]–[21]. Observed theta power was the largest for the FL region and Cz channel, which is in line with the previous results [11].

Neuroanatomical studies have shown that the hippocampus, the parietal lobes, and the right prefrontal areas are activated by visuospatial working memory tasks and are implicated in complex navigation [22], [23]. Also, allocentric or world-centered representation of the environment has been linked to the functioning of the hippocampus, while the posterior parietal cortex is involved in egocentric or body-centered spatial cognition [24]. Therefore,

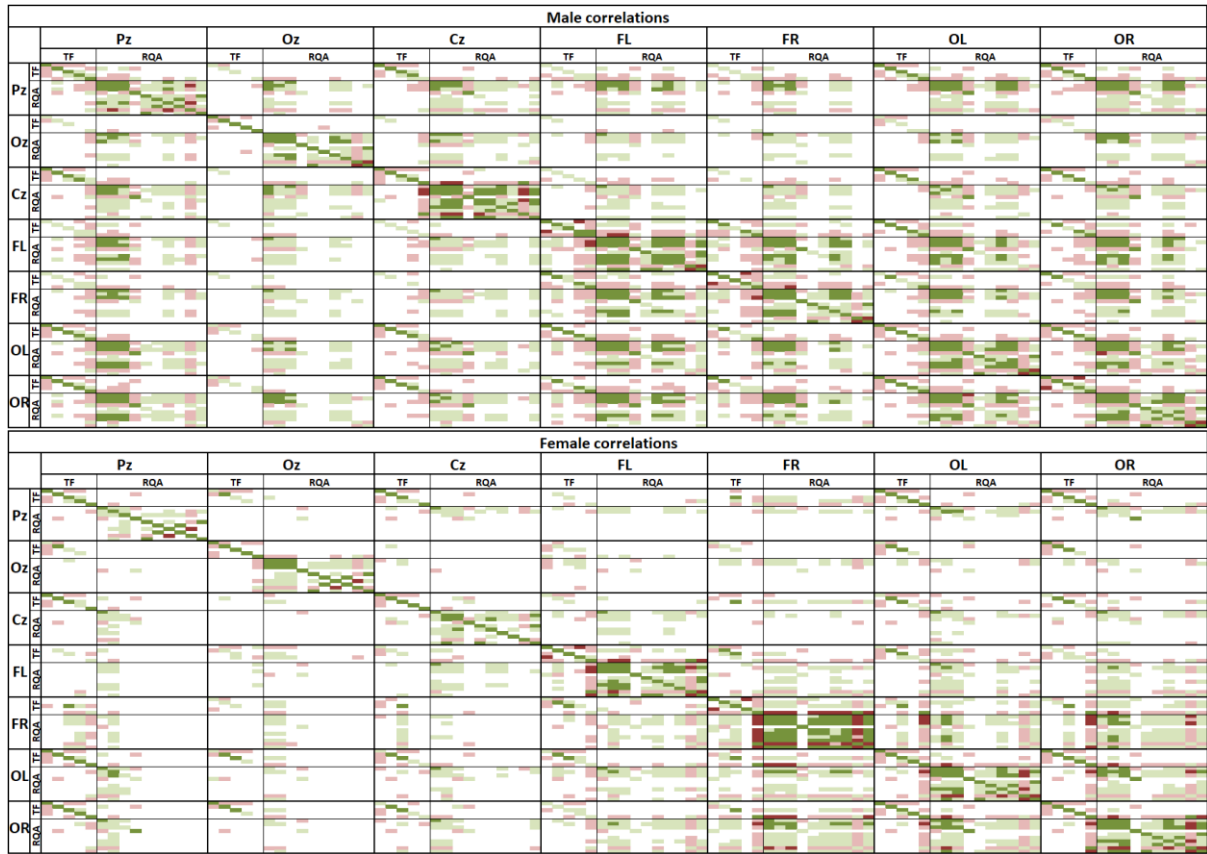


Fig. 2. Correlations for males and females with color-coded values. White – correlation between -0.4 and 0.4, light green – correlation between 0.4 and 0.7, dark green – correlation between 0.7 and 1, light red – correlation between -0.4 and -0.7 and dark red – correlation between -0.7 and -1. TF represents time-frequency domain features and RQA represents features from recurrence quantification analysis.

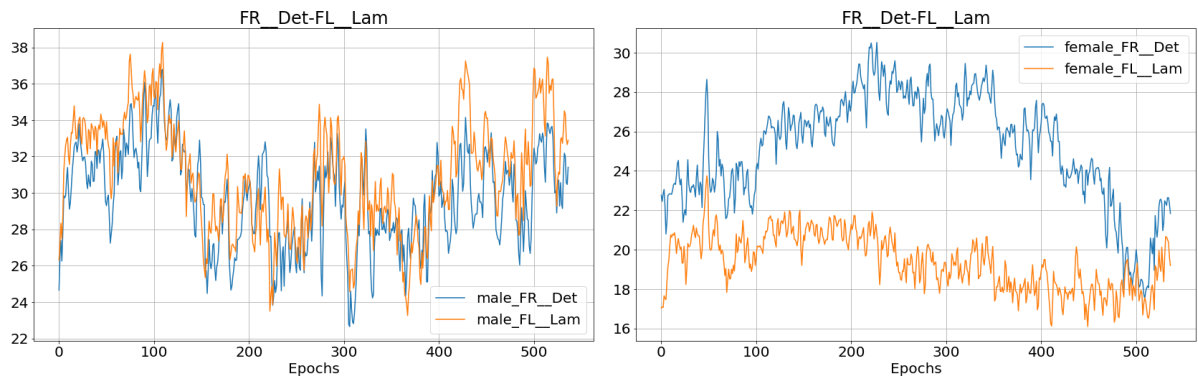


Fig. 3. On the left: highly correlated determinism (Det) from the FR region and laminarity (Lam) from the FL region for male drivers; on the right: weakly correlated determinism from the FR region and laminarity from the FL region for female drivers. Y-axis shows values of observed features.

observed differences between men and women in our sample could be related to different strategies used to perform the task, where men tend to focus more on the global-spatial information from the environment when performing the task, while women tend to use analytic information processing when performing the task [9].

Analysis of correlations between features points to the similar behavior of two signals. Therefore, higher correlations between different channels and regions could indicate shared neuronal processes. In our results, male drivers showed significantly higher correlations between the RQA features with respect to

female drivers. This could indicate stronger synchronization of neuronal activity for male drivers that are in line with the active task-related processing which has been reflected by increased relative theta power in men.

RQA features are used to analyze non-linear relation between the recurrence of states $x(i)$ in phase space [25]. They do not need assumptions about stationarity or length of signal and are not too sensitive to noise in the signal. These are all favorable characteristics for describing EEG signals, which are non-stationary and are often susceptible to noise [14].

The flaw of using RQA features is that they are harder to interpret.

RR represents the density of recurrence points on the RP. Recurrence plot of white noise consists of many single dots, while RP of deterministic signal consists of many long diagonals. Det is the ratio of points forming diagonal lines with all points and can be interpreted as predictability of signal. Like Det, Lam is the ratio of points forming vertical lines and represents the number of laminar states in the dynamical system. Long vertical lines mean that the dynamical system has slow changes. TT is the mean length of vertical lines and indicates that changes of states are slow. Ent measures the complexity of the signal.

Unfortunately, there are not many reliable studies including RQA features, especially related to gender differences. Therefore, it is hard to compare these features and results based on them directly with the existing studies about gender differences. Despite that, studies like ours are necessary for further expansion and acceptance of different kinds of features that show significant differences between the inspected groups.

5. Conclusion

The present results show that male and female drivers differ significantly in many ways in EEG features. Observed differences of the time-frequency domain features and RQA features, together with the strength of correlations between different features and brain regions, for male and female drivers, could reflect the differences in their information processing strategies during driving. Furthermore, these results could suggest that men and women tend to focus on different informations when driving. Therefore, observed results could account for reported gender differences in the number of traffic accidents and traffic behavior.

This conclusion, with further research, may lead to improvements in a better understanding of gender differences in brain functions during different cognitive tasks and to a better understanding of brain functions in general. There is a large number of applications that could benefit from a better understanding of gender differences in brain functions (better understanding of inter-gender behavior in traffic; improved quality of disease detection; understanding of learning, memory, and emotions; and similar).

In our further work, we plan to record more participants and implement more features, the ones that have already been used in studies related to gender differences in EEG signals, as well as the ones that have not yet been used in this field. With a larger number of features that can be related to existing studies, we hope that we will be able to provide better interpretations and explanations of differences between male and female drivers. At the same time, we would strive to lay the foundations for future studies that would analyze features which have not been inspected in the field of gender differences.

References

- [1] B. A. Jonah, "Accident risk and risk-taking behaviour among young drivers," *Accid. Anal. Prev.*, vol. 18, no. 4, pp. 255–271, Aug. 1986.
- [2] National Roads Authority, "Young driver accidents," 2000.
- [3] S. Laapotti, E. Keskinen, and S. Rajalin, "Comparison of young male and female drivers' attitude and self-reported traffic behaviour in Finland in 1978 and 2001," *J. Safety Res.*, vol. 34, no. 5, pp. 579–587, Jan. 2003.
- [4] P. Cordellieri, F. Baralla, F. Ferlazzo, R. Sgalla, L. Piccardi, and A. M. Giannini, "Gender Effects in Young Road Users on Road Safety Attitudes, Behaviors and Risk Perception," *Front. Psychol.*, vol. 7, p. 1412, Sep. 2016.
- [5] S. J. Wylie, "Young female drivers in New Zealand," *Accid. Anal. Prev.*, vol. 27, no. 6, pp. 797–805, Dec. 1995.
- [6] International Transportation Forum, "Road safety: annual report 2018," 2018.
- [7] D. Kimura, *Sex and Cognition*. The MIT Press, 1999.
- [8] V. J. Storie, "Male and Female Car Drivers: Differences Observed in Accidents," *Transp. Res. Int. Doc.*, vol. 761, 1977.
- [9] J. Ramos-Loyo and L. M. Sanchez-Loyo, "Gender differences in EEG coherent activity before and after training navigation skills in virtual environments," *Hum. Physiol.*, vol. 37, no. 6, pp. 700–707, Nov. 2011.
- [10] M. Corsi-Cabrera, J. Ramos, M. A. Guevara, C. Arce, and S. Gutierrez, "Gender Differences in the Eeg During Cognitive Activity," *Int. J. Neurosci.*, vol. 72, no. 3–4, pp. 257–264, Jan. 1993.
- [11] S. E. Kober, J. L. Reichert, C. Neuper, and G. Wood, "Interactive effects of age and gender on EEG power and coherence during a short-term memory task in middle-aged adults," *Neurobiol. Aging*, vol. 40, pp. 127–137, Apr. 2016.
- [12] B. Rescher and P. Rappelsberger, "Gender dependent EEG-changes during a mental rotation task," *Int. J. Psychophysiol.*, vol. 33, no. 3, pp. 209–222, Sep. 1999.
- [13] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.
- [14] M. Niknazar, S. R. Mousavi, B. Vosoughi Vahdat, and M. Sayyah, "A New Framework Based on Recurrence Quantification Analysis for Epileptic Seizure Detection," *IEEE J. Biomed. Heal. Informatics*, vol. 17, no. 3, pp. 572–578, May 2013.
- [15] S. Georgiev, Z. Minchev, C. G. Christova, and D. Philipova, "Gender Event-related Brain Oscillatory Differences in Normal Elderly Population EEG," *Int. J. Bioautomation*, vol. 15, no. 1, pp. 33–48, 2011.
- [16] G. Grön, A. P. Wunderlich, M. Spitzer, R. Tomczak, and M. W. Riepe, "Brain activation during human navigation: gender-different neural networks as substrate of performance," *Nat. Neurosci.*, vol. 3, no. 4, pp. 404–408, Apr. 2000.
- [17] C. Kamarajan *et al.*, "Theta oscillations during the processing of monetary loss and gain: A perspective on gender and impulsivity," *Brain Res.*, vol. 1235, pp. 45–62, Oct. 2008.
- [18] W. Klimesch, "EEG alpha and theta oscillations

- reflect cognitive and memory performance: a review and analysis," *Brain Res. Rev.*, vol. 29, no. 2–3, pp. 169–195, Apr. 1999.
- [19] S. E. Kober and C. Neuper, "Sex differences in human EEG theta oscillations during spatial navigation in virtual reality," *Int. J. Psychophysiol.*, vol. 79, no. 3, pp. 347–355, Mar. 2011.
- [20] M. Thordstein, N. L??fgren, A. Flisberg, K. Lindecrantz, and I. Kjellmer, "Sex differences in electrocortical activity in human neonates," *Neuroreport*, vol. 17, no. 11, pp. 1165–1168, Jul. 2006.
- [21] J. B. Caplan, J. R. Madsen, A. Schulze-Bonhage, R. Aschenbrenner-Scheibe, E. L. Newman, and M. J. Kahana, "Human theta oscillations related to sensorimotor integration and spatial learning," *J. Neurosci.*, vol. 23, no. 11, pp. 4726–36, Jun. 2003.
- [22] A. M. Owen, A. C. Evans, and M. Petrides, "Evidence for a Two-Stage Model of Spatial Working Memory Processing within the Lateral Frontal Cortex: A Positron Emission Tomography Study," *Cereb. Cortex*, vol. 6, no. 1, pp. 31–38, Jan. 1996.
- [23] E. Salmon *et al.*, "Regional brain activity during working memory tasks," *Brain*, vol. 119, no. 5, pp. 1617–1625, 1996.
- [24] J. O'Keefe and L. Nadel, *The Hippocampus as a Cognitive Map*. Oxford University Press, 1978.
- [25] U. R. ACHARYA, S. V. SREE, S. CHATTOPADHYAY, W. YU, and P. C. A. ANG, "APPLICATION OF RECURRENCE QUANTIFICATION ANALYSIS FOR THE AUTOMATED IDENTIFICATION OF EPILEPTIC EEG SIGNALS," *Int. J. Neural Syst.*, vol. 21, no. 03, pp. 199–211, Jun. 2011.

List of Abbreviations

EEG	Electroencephalogram
MEG	Magnetoencephalogram
fMRI	Functional magnetic resonance imaging
MRI	Magnetic resonance imaging
CT	Computational tomography
NREM	Non-rapid eye movement
REM	Rapid eye movement
ECG	Electrocardiogram
EOG	Electrooculogram
EMG	Electromyogram
RIP	Respiratory inductance plethysmography
GSR	Galvanic skin respons
ST	Skin temperature
KSS	Karolinska sleepiness scale
SSS	Stanford sleepiness scale
SAFE	Aircrew fatigue evaluation model
SAFTE	Sleep, activity, fatigue, and task effectiveness model
PERCLOS	Percentage of eyelid closure
PPG	Photoplethysmography
HRV	Heart rate variability
HR	Heart rate
LF/HF	Low frequency to high frequency bands power ratio
SEM	Slow eye movements
sEMG	surface EMG
NIRS	Near-infrared spectroscopy
fNIRS	Functional NIRS
HbO	Oxygenated hemoglobin
EDR	Electro dermal response
PGR	Psycho galvanic reflex
SCR	Skin conductance response

sEMG	surface EMG
FIR	Finite impulse response
IIR	Infinite impulse response
ICA	Independent component analysis
FFT	Fast Fourier transform
DFT	Discrete Fourier transform
PSD	Power spectral density
KNN	K-nearest neighbors
SVM	Support vector machine
DT	Decision tree
RF	Random forest
XGBoost	Extreme gradient boosting
EMO	Evolutionary multicriteria optimization
NSGA-II	Non-dominated sorting genetic algorithm II

Biography

Igor Stančin was born on September 18, 1992. in Zagreb. He finished his elementary and high school education in Zagreb, and in 2011 he enrolled in undergraduate studies in electrical engineering, majoring in automation, at the Faculty of Electrical Engineering and Computing (FER) of the University of Zagreb. After completing his undergraduate studies, he enrolled in a graduate study in computer science, at the Faculty of Electrical Engineering and Computing (FER) of the University of Zagreb. In September 2018, he enrolled in the post-graduate doctoral study in computer science at the same faculty, and since then he has been employed at the Faculty of Electrical Engineering and Computer Science as an assistant financed by the European Fund for Regional Development.

As an assistant employed at the Department of Electronics, Microelectronics, Computer and Intelligent Systems, he actively participates in the work on the industrial IRI project "Research and development of systems for recognizing driver fatigue and distraction" under the leadership of prof. Ph.D. Marija Cifreka from September 2018 to May 2022. From May 2022, he actively participates in the work on the industrial IRI project "Platform for intelligent and energy-efficient management of industrial IoT devices" under the leadership of associate professor Ph.D. Alan Jovic. He participates in classes as an assistant in the courses Software Engineering, Introduction to Data Science, Deep Data Analysis, Formal Verification of Program Support and Expert Systems. As an associate, he participated in: the industrial project "Intelligent analysis of the data of policyholders of the basic retail portfolio" of the company Croatia osiguranje d.d. and FER; and the industrial project "Prediction of telecommunication cell failures from time series of measured calls" of HashCode d.o.o. and FER.

His area of scientific interest includes the processing and analysis of time series and the application of artificial intelligence methods. He is a member of the Croatian section of IEEE.

List of works:

Papers in A category journals

1. Stancin, M. Cifrek, and A. Jovic, "A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems," *Sensors*, vol. 21, no. 11, p. 3786, May 2021, doi: 10.3390/s21113786 (international review, article, scientific)
2. Stancin, N. Frid, M. Cifrek, and A. Jovic, "EEG Signal Multichannel Frequency-Domain Ratio Indices for Drowsiness Detection Based on Multicriteria Optimization," *Sensors*, vol. 21, no. 20, p. 6932, Oct. 2021, doi: 10.3390/s21206932 (international review, article, scientific)
3. Stancin, M. Z. Zeba, K. Friganovic, M. Cifrek, and A. Jovic, "Information on Drivers' Sex Improves EEG-Based Drowsiness Detection Model," *Applied Sciences*, vol. 12, no. 16, p. 8146, Aug. 2022, doi: 10.3390/app12168146 (international review, article, scientific)

Papers in conference proceedings

1. Stančin, Igor; Friganović, Krešimir; Zelenika Zeba, Mirta; Jović, Alan; Cifrek, Mario, Gender Differences in EEG Features While Driving. // Proceedings of the 2nd International Conference on Advances in Signal Processing and Artificial Intelligence / Yurish, Sergey Y. (ur.). Barcelona, Spain: International Frequency Sensor Association (IFSA) Publishing, S. L., 2020. str. 125-130 (lecture, international review, full paper (in extenso), scientific)
2. Jović, Alan; Stančin, Igor; Friganović, Krešimir; Cifrek, Mario, Clinical Decision Support Systems in Practice: Current Status and Challenges. // Proceedings 43rd International Convention MIPRO 2020 / Skala, K. (ur.). Rijeka: Croatian Society for Information, Communication and Electronic Technology – MIPRO, 2020. str. 373-378. (<https://www.bib.irb.hr/1083692>) (lecture, international review, full paper (in extenso), scientific)
3. Stančin, Igor; Jović, Alan „An overview and comparison of free Python libraries for data mining and big data analysis“ // MIPRO 2019 Proceedings / Skala, Karolj (ur.). Rijeka: Croatian Society for Information and Communication Technology, Electronics and Microelectronics - MIPRO, 2019. str. 1161-1166

doi:10.23919/MIPRO.2019.8757088 (lecture, international review, full paper (in extenso), scientific)

4. Stančin, Igor; Jović, Alan „Analyzing the Influence of Player Tracking Statistics on Winning Basketball Teams“ // Proceedings of MIPRO 2018 / Skala, Karolj (ur.). Rijeka: Croatian Society for Information and Communication Technology, Electronics and Microelectronics - MIPRO, 2018. str. 1779-1784 doi:10.23919/MIPRO.2018.8400276 (lecture, international review, full paper (in extenso), scientific)

Životopis

Igor Stančin rođen je 18.9.1992. u Zagrebu. Osnovno i srednjoškolsko obrazovanje završava u Zagrebu, a 2011. upisuje preddiplomski studij elektrotehnike, smjer automatika, na Fakultetu elektrotehnike i računarstva (FER) Sveučilišta u Zagrebu. Po završetku preddiplomskog studija, upisuje diplomski studij računarstva, smjer računarska znanost, na Fakultetu elektrotehnike i računarstva (FER) Sveučilišta u Zagrebu. Poslijediplomski doktorski studij računarstva na istom fakultetu upisuje u rujnu 2018. godine, od kraja je i zaposlen na Fakultetu elektrotehnike i računarstva na mjesto asistenta financiranog od strane Europskog fonda za regionalni razvoj.

Kao asistent zaposlen na Zavodu za elektroniku, mikroelektroniku, računalne i inteligentne sustave aktivno sudjeluje u radu na industrijskom IRI projektu „Istraživanje i razvoj sustava za prepoznavanje umora i distrakcije vozača“ pod vodstvom prof. dr. sc. Marija Cifreka od rujna 2018. do svibnja 2022. Od svibnja 2022 aktivno sudjeluje u radu na industrijskom IRI projektu „Platforma za inteligentno i energetske efikasno upravljanje industrijskim IoT uređajima“ pod vodstvom izv. prof. dr. sc. Alana Jovića. U nastavi sudjeluje kao asistent na predmetima Programsko inženjerstvo, Uvod u znanost o podacima, Dubinska analiza podataka, Formalne verifikacija programske potpore i Ekspertni sustavi. Kao suradnik sudjelovao je na: industrijskom projektu "Inteligentna analiza podataka osiguranika osnovnog maloprodajnog portfelja" tvrtke Croatia osiguranje d.d. i FER-a; i industrijskom projektu " Predikcija kvarova telekomunikacijskih ćelija iz vremenskih nizova mjerenih poziva" tvrtke HashCode d.o.o. i FER-a.

Područje njegovog znanstvenog interesa obuhvaća obradbu i analizu vremenskih nizova te primjenu metoda umjetne inteligencije. Član je hrvatske sekcije IEEE-a.