JYU DISSERTATIONS 765

Xin Zuo

Automatic Detection of Driver's Abnormal State Based on Physiological Signals

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF INFORMATION TECHNOLOGY

JYU DISSERTATIONS 765

Xin Zuo

Automatic Detection of Driver's Abnormal State Based on Physiological Signals

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi Mattilanniemen auditoriossa MaA103
huhtikuun 5. päivänä 2024 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Information Technology of the University of Jyväskylä,
in building Mattilanniemi, auditorium MaA103, on April 5, 2024, at 12 o'clock.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2024

Editors
Marja-Leena Rantalainen
Faculty of Information Technology, University of Jyväskylä
Päivi Vuorio
Open Science Centre, University of Jyväskylä

ISBN 978-952-86-0099-2 (PDF)
URN:ISBN:978-952-86-0099-2
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-952-86-0099-2

ABSTRACT

Zuo, Xin
Automatic detection of driver's abnormal state based on physiological signals
Jyväskylä: University of Jyväskylä, 2024, 72 p. + included articles
(JYU Dissertations
ISSN 2489-9003; 765)
ISBN 978-952-86-0099-2 (PDF)

Abnormal driver state affecting environment perception, decision-making, and actions is one of the main traffic accidents causes. Physiological signals, reflecting drivers' actual internal state, have been used to detect abnormal status. Time and frequency domain features are commonly adopted to study driver state, while they are sensitive to residual noise and neglect signals' complexity. Besides, high temporal resolution is necessary to detect and analyze the subtle changes in driver status timely at a certain time but increases the sample rate and may decrease the computational efficiency. In long-term driving operations, the long-term temporal dependency is also vital for the automatic detection of the diver's abnormal state. However, difficulties exist in optimizing sample rate considering the time complexity of physiological signals and detecting abnormal driver status automatically with respect to long-term context information.

This dissertation focuses on the above challenges and proposes to detect the driver's abnormal state with multiscale entropy of physiological signals and bidirectional long short-term memory (BiLSTM) network. **Article I** explores the complexity of electroencephalography (EEG) with multiscale entropy on relative time scales (MSE) and the information compensation manner among features in detecting distraction. **Article II** investigates the fluctuation patterns of MSE and considers the long-term dependency on features with BiLSTM for distraction detection. **Article III** studies the appropriate sample rate by calculating multiscale entropy on absolute time scales (MSaE) and explores the distraction information in multiple physiological signals to detect distraction. **Articles II** and **III** also analyze the behavioral signals to validate the changes in driving performance due to distraction. In **Article IV**, a cross-subject emotion recognition framework based on fused entropy features and BiLSTM is proposed to integrate the merits of different features and learn the contextual information in EEG.

In summary, this dissertation investigates the fluctuation patterns of physiological signals with multiscale entropy of the optimized sample rate under different mental statuses to detect abnormal states with BiLSTM. The proposed framework indicates the potential of understanding and detecting a driver's abnormal state with multiple signals.

Keywords: driver state, distraction, emotion, multiscale entropy, long short-term memory, physiological signals, driving performance, multi-modality analysis

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Zuo, Xin
Automaattinen kuljettajan poikkeavan tilan havaitseminen fysiologisten signaalien perusteella
Jyväskylä: Jyväskylän yliopisto, 2024, 72 s. + artikkelit
(JYU Dissertations
ISSN 2489-9003; 765)
ISBN 978-952-86-0099-2 (PDF)

Kuljettajan epänormaali tila, joka vaikuttaa hänen ympäristönsä havainnointiin, päätöksentekoon ja toimintaan, on yksi suurimmista liikenneonnettomuuksien syistä. Fysiologisia signaaleja, jotka heijastavat kuljettajan sisäistä tilaa, käytetään poikkeavuuksien havaitsemiseen. Aika- ja taajuusalueen piirteitä käytetään yleisesti kuljettajan tilan tutkimiseen, mutta ne ovat herkkiä jälkiäänelle eivätkä erota signaalien monimutkaisuutta. Lisäksi korkea ajallinen resoluutio on tarpeen kuljettajan tilan muutoksen havaitsemiseksi ja analysoimiseksi tiettynä aikana, mutta se lisää näkörataa ja saattaa vähentää laskennallista tehokkuutta. Pitkäkestoisen ajon aikana pitkäaikainen ajallinen riippuvuus on myös elintärkeää kuljettajan poikkeavan tilan automaattiselle havaitsemiselle. Näköradan optimointiin liittyy kuitenkin vaikeuksia, kun otetaan huomioon fysiologisten signaalien aikavaativuus ja automaattinen kuljettajan poikkeavan tilan havaitseminen suhteessa pitkän aikavälin kontekstitietoon.

Tämä väitöskirja keskittyy edellä mainittuihin haasteisiin ja ehdottaa kuljettajan poikkeavan tilan havaitsemista fysiologisten signaalien moniskaalaentropian ja kaksisuuntaisen pitkäaikaisen lyhyen aikavälin muistin verkoston avulla. Artikkeli I tutkii elektroenkefalografian (EEG) moniskaalaentropian avulla suhteellisia aikaskaaloja käyttäen aivojen monimutkaisuutta ja informaatiokorvaustapaa ominaisuuksien välillä häiriön havaitsemiseksi. Artikkeli II tutkii moniskaalaentropian vaihtelumalleja ja ottaa huomioon pitkäaikaisen riippuvuuden ominaisuuksissa kaksisuuntaisen pitkäaikaisen lyhyen aikavälin muistin avulla häiriön havaitsemiseksi. Artikkeli III tutkii sopivaa näkörataa laskemalla moniskaalaentropiaa absoluuttisilla aikaskaaloilla ja häiriötietoa useissa fysiologisissa signaaleissa häiriön havaitsemiseksi. Artikkelit II ja III analysoivat myös käyttäytymissignaaleja varmistaakseen häiriön aiheuttamat muutokset ajosuorituksessa. Artikkelissa IV esitetään monitieteellinen tunteidentunnistuskehys, joka perustuu sulautettuihin entropioihin ja kaksisuuntaiseen pitkäaikaiseen lyhyen aikavälin muistiin, jotta voidaan integroida eri piirteiden edut ja oppia EEG:n kontekstitietoa.

Avainsanat: kuljettajan tila, häiriö, tunne, moniskaalaentropia, pitkäaikainen lyhyen aikavälin muisti, fysiologiset signaalit, ajosuoritus, monimuotoisuusanalyysi

Author Xin Zuo
Faculty of Information Technology
University of Jyväskylä
Finland
Email: zuoxin93@foxmail.com
ORCID: 0000-0002-0828-9748

Supervisors Timo Hämäläinen
Faculty of Information Technology
University of Jyväskylä
Finland

Fengyu Cong
Faculty of Information Technology
University of Jyväskylä
Finland

Jian Zhao
School of Mechanical Engineering
Dalian University of Technology
China

Chi Zhang
School of Biomedical Engineering
Dalian University of Technology
China

Reviewers Zhiguo Zhang
School of Computer Science Technology
Harbin Institute of Technology, Shenzhen
China

Yu Sun
College of Biomedical Engineering and Instrument Science
Zhejiang University
China

Opponent Jari Hyttinen
Faculty of Medicine and Health Technology
Tampere University
Finland

ACKNOWLEDGEMENTS

I embarked my doctoral study in Finland soon after the onset of COVID-19 pandemic, which was a time full of uncertainties and challenges. Nevertheless, I managed to fulfill this dissertation with the support and help of a village people. I would appreciate all the people who have been a part of this unforgettable and incredible journey with me.

First and foremost, I would like to express my sincere thanks to Prof. Tapani Ristaniemi. It is a pity that the pandemic did not allow me to meet him in person in 2020. But I will always remember his kindness and support for accepting me to his research group in University of Jyväskylä. May his soul rest in peace.

I would like to thank my supervisors Prof. Fengyu Cong and Prof. Jian Zhao. Their firm support, encouragement and guidance are invaluable during the entire journey. I would not get to know the chance of pursuing my doctoral degree in Finland without the recommendation of Prof. Zhao. It is also him who suggested me to further my study in this prospective research field at the initial stage of planning my research proposal. I am very grateful for Prof. Cong who helped me successfully apply for the financial support of China Scholarship Council. He also warmly welcomed me to join his professional research team to pursue my doctoral study with overwhelming academic guidance, and I am so proud to have been a part of our ASAP family. I am also gratitude for their careful consideration about my future and for being so supportive and understanding. Their passion and precise attitudes to research as well as their extensive academic knowledge have impressed me greatly and will inspire me in my future research work.

I would like to express my heartfelt appreciation to my supervisor Prof. Timo Hämäläinen, who took over from Tapani as my doctoral supervisor in Finland. Timo is a wise, responsible, and kind supervisor, which makes me feel that he is my colleague rather than a supervisor. Every time I email or meet him to ask for advice or help, he responses almost immediately to my puzzles even outside the working hours. Thanks for his every "good work" that encourages me to be more relaxed and work patiently. I also appreciate his kind help in revising my manuscripts and this dissertation as well as translating the Finnish sections. Thanks for all the help and support of him since my arrival in Finland.

I would like to express my sincere gratitude to my supervisor Associate Professor Chi Zhang for his constant guidance and encouragement in this journey. I feel so lucky to have him as my mentor. From the initial beginning of my research to the final submission of this dissertation, his wisdom and support have been instrumental in building my academic knowledge. His patience, expertise and dedication to my research have always inspired me to complete my doctoral study without worrying about any difficulties. Thanks for our weekly group meetings which are the pivotal checkpoints to keep me on track. Thanks for his sparking ideas and constructive comments on my research, manuscripts, and dissertation. I am very grateful for his contributions to my research.

I would like to sincerely thank Prof. Zhiguo Zhang (School of Computer Science Technology, Harbin Institute of Technology, Shenzhen) and Prof. Yu Sun (College of Biomedical Engineering and Instrument Science, Zhejiang University) for reviewing this dissertation. I would also like to sincerely thank Prof. Jari Hyttinen from Tampere University to be my opponent.

I would like to thank the China Scholarship Council for providing me with financial support to pursue my doctoral degree in Finland. Without this funding, the completion of this dissertation would not be as smooth.

I am profoundly grateful for all my friends who have accompanied me throughout the whole journey and created numerous wonderful memories. I would like to thank Dr. Huashuai Xu, Dong Tang, Liting Song, Xiangyu Rong and Lina Sun for their help and company. The various get-together activities organized by them have lightened and enriched my humdrum daily life. Special thanks go to Dr. Jia Liu, Dr. Wenya Liu, Dr. Lili Tian, Dr. Guanghui Zhang, Dr. Xiaoshuang Wang, Dr. Dongdong Zhou, Jiaqi Zheng, Xinying Chen, Dr. Zhonghua Chen, Dr. Chaoxiong Ye and Yalin Sun, their pleasant presence makes my life more interesting. I would also extend my thanks to other members in the EEG Group and ASAP family at Dalian University of Technology as well as the colleagues in the Faculty of Information Technology at University of Jyväskylä.

Finally, I would like to express my deepest gratitude to my family. Thanks to my dear father who always believe in my abilities and support my decisions. His unconditional love and tireless work have a profound influence on my life. For our family, 2023 has been a tough year, and I am sorry I did not go home to accompany them. Mom, you are the most strong-willed and bravest woman in my heart, and I am so proud of you. My best wishes are for your new lease on life and good luck. I would like to thank my little sister and little brother for being responsible enough to look after our parents and take care of their affairs. Their thoughtfulness and consideration have taken my mind off worrying about our family. I also appreciate my uncle who emphasizes the importance of education and gives me invaluable support and help throughout these year studies. I am fortunate to have my beloved grandparents accompanied since I was born. They have made immense sacrifices for me and my family and given us selfless dedication which continues to encourage me in my future life. Lastly, I want to express my special appreciation to my boyfriend, Dr. Changyi Peng. His unwavering love, company, and belief in me have been one of the greatest sources of motivation and strength in this long academic journey in Finland. Best wishes to our future journey!

Jyväskylä 15.03.2024
Xin Zuo

ACRONYMS

AE	Approximate entropy
ANN	Artificial neural networks
BiLSTM	Bidirectional long short-term memory network
BVP	Blood volume pulse
CNN	Convolutional neural network
DE	Differential entropy
DL	Deep learning
ECG	Electrocardiogram
EEG	Electroencephalography
EMG	Electromyography
FE	Fuzzy entropy
GSR	Galvanic skin response
HRV	Heart rate variability
KNN	K-nearest neighbor
LPV	Lane position variability
LSTM	Long short-term memory
MI	Mutual information
ML	Machine learning
MSaE	Multiscale entropy on absolute time scales
MSE	Multiscale entropy on relative time scales
NCA	Neighborhood component analysis
NMF	Non-negative matrix factorization
OSA	Obstructive sleep apnea
PSD	Power spectral density
RE	Rényi entropy
RF	Random forest
RNN	Recurrent neural network
SE	Sample entropy
SFS	Sequential forward selection
SVM	Support vector machine
WT	Wavelet transform

FIGURES

TABLES

CONTENTS

LIST OF INCLUDED ARTICLES

I **Xin Zuo**, Chi Zhang, Jian Zhao, Timo Hämäläinen, and Fengyu Cong (2024). Driver distraction detection based on EEG feature fusion using Random Forest. In *ICBSP '23: Proceedings of the 2023 8th International Conference on Biomedical Imaging, Signal Processing* (pp. 104-109). ACM, https://doi.org/10.1145/3634875.3634890

II **Xin Zuo**, Chi Zhang, Fengyu Cong, Jian Zhao, and Timo Hämäläinen (2022). Driver distraction detection using bidirectional long short-term network based on multiscale entropy of EEG. *IEEE Transactions on Intelligent Transportation Systems*, 23(10), 19309-19322, http://doi.org/10.1109/TITS.2022.3159602

III **Xin Zuo**, Chi Zhang, Fengyu Cong, Jian Zhao, and Timo Hämäläinen (2023). Driver distraction detection based on MSaE of multi-modality physiological signals. Submitted to *IEEE Transactions on Intelligent Transportation Systems*.

IV **Xin Zuo**, Chi Zhang, Timo Hämäläinen, Hanbing Gao, Yu Fu, and Fengyu Cong (2022). Cross-subject emotion recognition using fused entropy features of EEG. *Entropy*. 24(9), 1281, https://doi.org/10.3390/e24091281

1 INTRODUCTION

Nowadays, the number of cars on road is growing rapidly with the increasing of modern economic level, which is usually followed by the more frequent traffic accidents. It will result in not only financial loss but also more injuries and deaths. According to the U. S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA), an estimated about 2.7 million police-reported motor vehicle accidents happen in 2021 (NHTSA, 2022). A more detailed crash statistics reported by the Fatality Analysis Reporting System (FARS) in the same year shows that there are more than 61 thousand fatal vehicle crashes leading to approximately 43 thousand people died nationwide (NHTSA, 2021). As for the crash investigation all over the world, around 1.3 million fatalities occur on account of road traffic crashes in the survey released by the World Health Organization (WHO) in 2022 (WHO, 2023). The number of people suffering from different levels of injuries is even larger that between 20 and 50 million more injuries are caused by car accidents.

It has been reported that the human mistakes and violations induced by driver's abnormal mental status are the main causes of most accidents and have apparent effects on road safety (Shahverdy et al., 2020). Since driving is an activity of driver-vehicle-environment interaction, drivers play important roles in this process. It affects not only making decisions but also taking actions based on the driving environment if the driver's state changes to be deviant with time, thus leading to risks of incidents. Therefore, driver's states have significant influence on driving safety so long as they are involved in controlling the vehicle. In this case, it is important to detect driver's abnormal state in time and remind the driver to pay attention to driving safely to avoid crashes.

Driver distraction, one kind of abnormal mental status, is reported as the major contributor of car collisions making up 27% of all serious injuries in 2022 (Pandurov, 2023). Driving requires drivers to concentrate their attention to the surroundings and take actions immediately to any unexpected events. Risky driving behaviors (i.e., operating electronic devices, eating, talking etc.) can divert their attention to activities irrelevant to driving and increase the mental workload, thus, bringing about distraction (Hossain et al., 2022). It has been reported that using a cell phone while driving contributes the most for distracted driving (W. Zhang & Zhang, 2021). These abnormal behaviors weaken the driver's abilities to perceive information from environment as well as to control the vehicle. In addition, the response time can also be augmented when encountering upcoming events. Hence, more accidents are generated.

Besides, the probability of traffic crashes can be greatly increased by about 10 times when drivers experiencing strong emotions (such as sadness, happiness, and anger) while driving (Guettas et al., 2020). It is also reported that road rage has been increased several times over the past few years, which can obviously increase the risks of accidents (Ramzan et al., 2019). These feelings increase the cognitive load and decrease the attention resource in processing information related to safe driving and may therefore result in involuntary driving errors (Precht et al., 2017). Nowadays, the traffic environment is becoming more and more complex with a large number of cars on roads. The pressure of life and working is boosted with the hustle and bustle of city life as well. Under the circumstances, drivers may suffer from various emotions that bring about dangerous driving behaviors. As a result, the possibility of exposing drivers and other traffic participants like pedestrians to injuries increases drastically.

Since driver distraction and emotion have significant influence on driving safety, it is necessary to monitor and detect distraction and abnormal emotions accurately and timely in order to alert the driver to make appropriate adjustments, thereby reducing the potential for car accidents.

Next, the motivations of the conducted research are introduced. Then, the aims and overview of the entire dissertation are described. Finally, the structure of the dissertation is illustrated.

1.1 Research motivation

Visual signals and vehicle behavioral signals are most widely used to study driver state as they are the intuitive responses of driver status and easy to collect without interfere with drivers. But as driver's behaviors usually take place after the mental status alteration, there is usually a delay in visual and vehicle behavioral signals. Besides, it has been validated that visual signals can be affected by personal driving habits, illumination, wearing glasses, individual purposes and so on (B. Zhang et al., 2023; J. Zhang et al., 2020) and that vehicle behavioral signals are dependent on subjects and sensitive to the weather and road conditions (Healey & Picard, 2000). Thus, the validity and availability of these signals need to be carefully considered. Nowadays, with the development of less intrusive equipment, physiological signals are more and more popular in driver state detection. Generated by the real-time electrical potentials of human body, physiological signals can reflect the actual internal status alterations of drivers in real time and cannot be hidden on purpose (Y. Dong et al., 2011).

Physiological signals are usually analyzed from the perspectives of time domain and frequency domain. However, the stability and rhythmicity of these signals change along with the dynamic alterations of physiological status (Z. Zhang et al., 2018). So, there is also extensive complexity information. Entropy features (like SE, AE, and FE) have been utilized to manifest the complexity of physiological signals. But the residual noise will be enrolled in the calculation of entropy features, which will increase the computational cost and decrease the robustness of the results (J. Urigüen & Zapirain, 2015). Besides, to analyze and detect the subtle changes in driver state timely, it is necessary to use signals in high temporal resolution. Yet, the data size increases with resolution and may decrease the computational efficiency. Under the circumstances, it is important to explore the valuable driver status information using the complexity of physiological signals under appropriate sample rate.

Moreover, it is commonly agreed that different modalities signals can provide more useful information for studying driver state (He et al., 2019). So, it has been a promising area to detect driver state with multi-modality signals. What's more, the ability of characterizing implicit driver status varies from feature to feature. In this case, information compensation also occurs among different features (Hasan & Kim, 2019). Nevertheless, how it happens among various features in driver state detection remains to be further studied.

As for driver state detection, traditional ML-based and DL-based algorithms make decision on each time step based on the current input state, which is inconsistent with the actual driving process. Driving is a long-lasting and interactive activity. During this process, drivers make decisions based on not only the current received information but also the previous information (Liang et al., 2007). Hence, it is vital to detect driver state according to the long-time context dependency. LSTM has been demonstrated to keep both short-term and long-term memory of sequential data (Hochreiter & Schmidhuber, 1997; Masood et al., 2024) and can be used to detect driver state. Furthermore, BiLSTM has also been utilized to memorize both forward and backward long and short-term valuable context information and is successfully applied for text classification and sleep apnea detection (G. Liu & Guo, 2019).

1.2 Aims of the research

Motivated by the above challenges in previous research, this dissertation aims to explore the implicit information of physiological signals changing with driver state stimulated by different tasks while driving. Besides, the dissertation also aims to detect driver state effectively based on the found fluctuation rules and LSTM.

Article I focuses on mining for the implicit information of EEG signals while driver distraction from the viewpoint of complexity and investigating how different features compensate for each other when detecting driver distraction with fused features. In this study, multiscale entropy on relative time scales (MSE) is adopted to manifest the dynamic fluctuations of EEG so as to eliminate the influence of residual noise on the results. The distraction detection result of MSE is then compared with other four conventional entropy features. Finally, to analyze the principle of feature compensation, RF is employed to detect driver distraction with fused features.

Article II is the improvement and enhancement of **Article I**, which aims to explore and learn the bidirectional distraction information of two modalities signals with MSE in sliding windows and BiLSTM to improve the detection accuracy. This study also aims to validate the behavioral changes while driver distraction by analyzing the vehicle behavioral data. In this study, MSE feature is firstly extracted from EEG to determine the most distraction position. Then, the statistical features of vehicle behavioral signals are calculated to validate the changes of driving performance after distraction. Afterwards, a BiLSTM model is established to detect driver distraction. We finally demonstrate the efficiency and superiority of the proposed framework by comparing with traditional features and classifiers.

Article III proposes a novel framework to probe into the dynamic distraction information contained in multiple modalities physiological signals with multiscale entropy on absolute time scales (MSaE) and identify driver distraction with LSTM. A resampling approach on the basis of entropy is used to determine the proper downsampling frequency of multi-modality physiological signals (i.e., EEG, ECG, and EMG). MSaE is then calculated in a sliding window to extract the distraction information. Thereafter, ReliefF selected from five traditional feature selectors is utilized for determining the optimal features. Finally, the selected features are fed into a LSTM classifier to recognize driver distraction. The proposed framework is validated and shows its potential in effectively detect driver distraction with multiple signals.

Article IV centers around manifesting the complexity information of EEG signals in different emotions with fused entropy features and achieving the goal of cross-subject emotion classification in high reliability with BiLSTM. Firstly, five entropy features are extracted from EEG signals to excavate the complex emotional information. A BiLSTM classifier is then trained with the fused features to recognize different emotional states. The results show the feasibility of the presented approach.

1.3 Research design and structure of the dissertation

The research design proposed in this dissertation is described to address the mentioned challenges in driver's abnormal state detection. In order to analyze driver status timely and reduce the effect of residual noise, MSE is employed to explore the fluctuations of physiological signals and detect driver's abnormal states robustly. After that, resampling problem arises in multiscale calculation, and MSaE of physiological signals is recruited to eliminate the influence of resampling. Once the resampling problem is resolved, the next difficulty is to find out the appropriate downsampling rate to keep high temporal resolution as well as save computational cost. An entropy-based approach is then proposed based on MSaE to determine the proper downsampling frequency of each physiological signals. As for how different features compensate for each other while detecting driver state, it is studied with RF. Last but not least, the long-term temporal dependency is also important to automatically detect abnormal driver state. Therefore, a novel framework based on BiLSTM is introduced to memorize the contextual information in signals for recognizing driver's abnormal state. Figure 1 shows the technical flowchart of the dissertation and Figure 2 shows the relationship between the included articles, mainly used signals, features, classifiers and the goals of them in this dissertation.

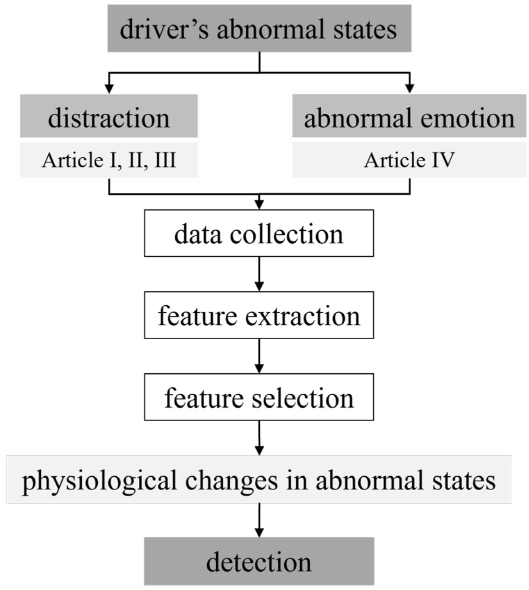


FIGURE 1 The technical flowchart of the dissertation.

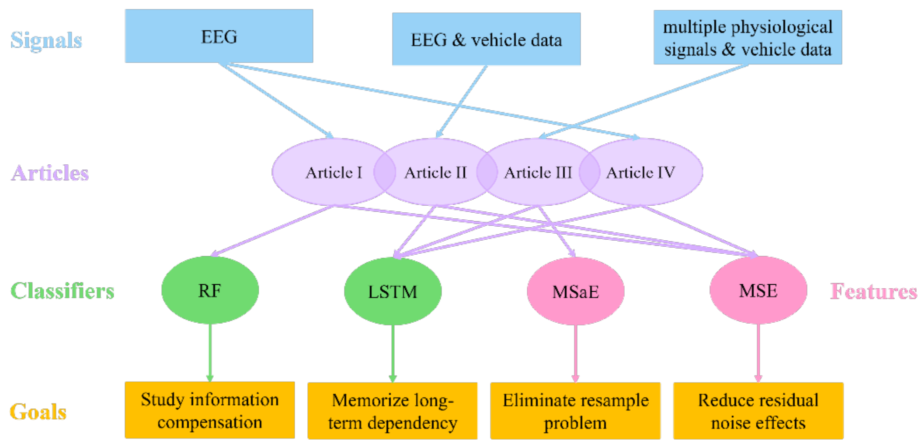


FIGURE 2 The used signals, features, classifiers, and goals in the dissertation.

The dissertation is organized as follows. The motivations and aims of this study are illustrated in Chapter 1. Chapter 2 introduces the background and related findings of driver state analysis and detection from the perspectives of concepts, data, and feature types as well as detection algorithms. Chapter 3 firstly describes the two driver distraction experiments and one emotion dataset involved in the dissertation, then the methodologies used in the research are detailed introduced. Chapter 4 summarizes the four included articles comprehensively and lists the contributions of authors in each article. Chapter 5 discusses and concludes the findings and limitations of this dissertation, as well as the future directions.

2 BACKGROUND

This chapter addresses the relevant research on driver distraction, emotion analysis and the classifiers used in the detection tasks.

2.1 Driver distraction

There are a variety of definitions for driver distraction in the literature. For instance, Horberry et al. (2006) believe that driver distraction occurs when the attention is shifted away from the task of driving to a triggering event. Stutts et al. (2001) hold the view that distraction appears when some reasons inside or outside the vehicle delay the information processing of driving safely and prompt drivers to shift their attention away from driving. All these definitions correlate driver distraction with distributing attention by activities unrelated to driving.

2.1.1 Categories of driver distraction

The current research about driver distraction mechanism can be categorized into four types: cognitive, manual, visual and audio distraction (Kashevnik et al., 2021). Cognitive distraction usually means that drivers are deep in meditation and that their minds are shifted from driving. Manual distraction happens when driver's hands are taken off from the steering wheel because of eating, drinking, operating devices, etc. With regard to visual distraction, looking at anything instead of the road and relevant traffic conditions while driving will induce driver distraction. When a driver is distracted by the acoustic events like car radio and music, auditory distraction occurs. Distracting secondary tasks are commonly used to induce one kind of distraction in order to analyze its mechanism. But in fact, secondary tasks can lead to the driver distracted by a combination of more than one certain type of distraction, which will generate a higher probability of collisions. For example, when drivers are required to talking in situational dialogues, they firstly listen to the words and then think about the answers, which involves cognitive distraction and auditory distraction. Since driver distraction comes into being by the interaction among the driver, vehicle, and environment, different types of distraction usually appear in a concomitant manner in driving scenarios (R. Wang et al., 2021). Hence, it is necessary to analyze and detect driver distraction without dividing it into different types.

2.1.2 Sensing techniques in driver distraction analysis

Various sensing techniques have been utilized to collect signals to explore driver states and detect distraction, and they can be roughly divided into two groups, i.e., non-intrusive and intrusive sensors. Non-intrusive sensors (including internal and external video cameras, Global Position System (GPS), and inertial measurement unit) have no effect on drivers and is suitable for large commercial systems built into vehicles (Marchegiani & Posner, 2018). The visual signals obtained by internal video cameras are most commonly used in the literature. Dua et al. (2020) proposed an AutoRate system to monitor driver's attention levels by utilizing various driver's facial and eye behaviors (such as yawns, facial landmarks, eye closure and so on) and compared the consistency of the results with human annotators. Al Shalfan et al. established a vision-based architecture to classify driver distraction on the basis of the movements of arms and hands and indicated the relevance of body behaviors to driver distraction (Alshalfan & Zakariah, 2021). Visual signals of drivers can be collected conveniently without impacts on drivers and the equipment is cheap, but these measures are sensitive to individual behavioral habits, illumination, and facial occlusion. The behavioral data of vehicles (like lateral position, steering wheel angel and grip force, etc.) is also widely adopted to study driver distraction. Yadawadkar et al. (2018) focused on building a classification system based on the statistical features of various naturalistic driving signals like acceleration and longitudinal distance and succeed to detect distraction, attention and drowsiness with the proposed system. The sensors gathering these signals are easily put in different components of the vehicle without making drivers feel uncomfortable. However, there is a delay in response to driver state with vehicle behavioral signals. Besides, the weather and road conditions may have effects on the driving behaviors.

As for the intrusive sensors, physiological sensors are usually used to directly collect physiological signals (like electroencephalography (EEG), electromyography (EMG), and electrocardiogram (ECG)) from human body. Fan et al. (2022) presented a method to detect distraction based on the energy features of EEG signals and validated the stability and feasibility of the proposed method. Sahayadhas et al. (2014) probed into the changes of drivers' hypovigilance states like distraction and fatigue with statistical features of EMG signals, and then constructed a universal system to detect driver inattention and alert the drivers. Physiological signals are the records of the real-time electrical potentials generated by the brain, muscles and heart etc., which reflect driver's real internal states in a faster and more accurate way than behaviors. But these sensors are usually attached to the skin surface of the driver and may be uncomfortable to wear.

Recently, the sensing technology and portable equipment have been greatly developed, which has largely reduced the impact of the physiological sensors on drivers. There is research that study distraction with portable equipment in the literature. For instance, Yang et al. (2021) collected the ECG signals with comfortable flex sensors to classify driver workload and investigated the influence of temporal variation and individual differences on it. Moreover, it has been demonstrated more sufficient information can be explored with multiple signals than that of single signals. Under this circumstance, more and more researchers start to study driver distraction adopting multiple kinds of signals. Das et al. (2022) introduced a multimodal dataset of driver distraction analysis including signals collected from visual, linguistic, near infrared and physiological modalities. They extracted features from different modality of signals to detect driver distraction and compared their roles in identifying driver distraction. Lechner et al. (2019) established a lightweight framework for exploring driver distraction with multi-sensor and multi-device and applied it in a preliminary driving experiment involving GPS, head movement and heart rate data. These studies have demonstrated the possibility of combining different modality signals together for the sake of analyzing driver distraction.

2.1.3 Analysis of driver distraction

As for how to analyze and detect driver distraction with different signals, it is necessary to extract discriminative indicators from signals especially for physiological signals, which are always collected with eye blinks, body movements, power line interface and so on. A variety of features have been recruited to explore the distraction information in different kinds of signals in previous research. The features in time domain like standard deviation and mean value are most widely used in excavating the changes of vehicle behavioral signals while distraction. The maximum left- and right-side lane departure and the instantaneous absolute steering angle were utilized to analyze the relationship between driving performance and driver distraction (Pavlidis et al., 2016). Their results illustrate that the mean absolute steering as well as the lane departure show increasing tendencies after distraction. The speed deviation and lane keeping offset features were calculated to find out the changes in vehicle behavioral signals when a driver is distracted and then used to detect driver distraction (Aksjonov et al., 2019). There are studies analyzing driver distraction with the frequency domain features of vehicle behavioral signals as well. For example, the cepstral analysis of the gas and brake pedal pressure signals and the vehicle velocity signals were applied to capture the distraction information (Öztürk & Erzin, 2012). Much research has employed facial landmarks to study driver distraction with reference to visual signals. The percentage of eye closed over time, eye or mouth occlusion and face orientation were proved to be associated with driver attention and can be used to analyze distraction (Smith et al., 2003). Among all facial features, eye movements related features are the most popular metrics in distraction analysis since they can intuitively reflect what a driver is focusing on (Song et al., 2013). The eye glance pattern while using cell phone was compared with that of normal driving process (Flannagan et al., 2012). According to the results, the duration and direction of drivers' glances can be used to recognize distraction.

When it comes to physiological signals, most of the research has extracted the time domain and frequency domain features to explore the distraction information in the literature. The energy values of four EEG rhythms were calculated in different brain regions to investigate the differences of brain activities among brain areas while performing distracting tasks, which provided a reference for detecting distraction with EEG (G. Li et al., 2023). Conventional statistical features (like median, standard deviation, and skewness etc.) and spectral features (like energy and first-order spectral moment, etc.) of ECG and EMG signals were employed to classify cognitive and visual distraction (Sahayadhas et al., 2015). The changing patterns of physiological signals in time domain and frequency domain can be surely manifested by these indicators during distraction. However, the physiological states of drivers change all the way in the driving process, which are accompanied by alteration of the stability and orderliness of physiological signals (Z. Zhang et al., 2018). So, there is not only temporal fluctuations and frequency information but also a large amount of complexity information in brain, heart, and other physiological activities (Gao et al., 2018).

Physiological signals record the real electrical activities generated by nervous system and reflect the actual internal states of human body. Once the complexity features are overlooked, a great deal of valuable information about the driver's physiological status may be missed.

Nowadays, entropy has been regarded as a measure of complexity and adopted to mine the complexity information contained in physiological signals in many fields (Dehzangi et al., 2018; Lv et al., 2020; Sharma et al., 2022). Ashok et al. (2017) proposed to use the Shannon entropy and approximate entropy (AE) of EEG to manifest the complexity difference between normal and epileptic states. They found that the value of AE decreases obviously in epileptic process. In a study of obstructive sleep apnea (OSA) (Zarei & Asl, 2018), four entropy features and other time domain and frequency domain features were calculated from ECG signal to detect OSA. The results proved that entropy-based features are better than traditional features in exploring the hidden information of ECG. Najmeh et al. recruited Shannon entropy to quantify the complexity information of EEG and GSR signals while rest and listening to music pieces (Pakniyat & Namazi, 2022). The feature value changes with the music significantly and a strong correlation is found between the variations of EEG and GSR in different conditions.

Although entropy features have been used to analyze human physiological status, challenges still exist in the analysis of physiological signals with these complexity-base features. It is known that the noise contamination consisted in signals cannot be completely eliminated no matter which approach is selected for preprocessing and that they will be partly retained instead (J. Urigüen & Zapirain, 2015). As a result, the residual noise is involved in the calculation of entropy features like AE, SE, and FE, which increases the computation time and weaken the robustness of the results. What's more, high temporal resolution of signals is necessary to analyze the subtle fluctuations in driver status and detect them timely. But it will result in the increase in the sampling rate and data size and then decrease the computational efficiency. For the purposes of keeping high temporal resolution and improving the calculation efficiency at the same time, most of the researchers adopt the way of resampling ahead of analysis. Yet, the work of investigating the effect of sampling frequency on SE had demonstrated the correlation between entropy value and sampling rate and indicated that a decrease appears in SE value with the increase of sampling frequency (Fallahtafti et al., 2021; Raffalt et al., 2019). Hence, the accompanied resampling problem occurs when calculating entropy features after downsampling.

In general, most of the current research on driver distraction is firstly dividing it into four categories. Then, select one or several specific types to study separately based on the visual-based features or vehicle behavioral features. There is also research analyzing and detecting driver distraction with the time domain and frequency domain features of physiological signals. Nevertheless, the vehicle behavioral signals and visual signals are delayed responses to driver states, and the complexity of physiological signals is usually neglected by the features in time domain and frequency domain. Under the circumstances, it is necessary to further study driver distraction based on complexity features of physiological signals.

2.2 Driver emotion

Emotion is a complex psychological concept that can be described from different perspectives. It is regarded as the tool of evaluating experience and preparing to act on specific situations from the viewpoint of behaviors (Niedenthal & Ric, 2006), while emotion is the response to the changes in physical conditions thinking of it physiologically (Phelps, 2004). There is also definition in terms of cognitive theory that considers emotion as the perceptual response to the environment and the result of the interaction between human and external environment (Lazarus, 1993). Since the derivation of emotion relates to the environment, physical conditions and individual cognition, emotion can be concluded that it is the psychological and physiological response induced by perceiving the internal and external stimuli. Current research has pointed that driving behaviors and performance can be influenced by driver's emotion directly, since emotion plays an important role in the perception, decision-making, learning and memory processes (Picard, 2003). In this way, it is bound up with driving safety and driving experience. Current research about emotion is usually carried out from three dimensions according to its definition that are subjective reports, external performance, and physiological responses.

2.2.1 Subjective reports

Subjective reports mainly include questionnaires and interviews, which collect individual's subjective data about their feelings by professional scales and oral inquiries (Koslouski et al., 2022). Mesken et al. (2005) utilized the sensation seeking scale to study the effect of anger emotion on cognitive bias in traffic and found that driver's judgement can be interfered by emotional state. The research based on subjective evaluation is easy to be conducted with low cost, but it is difficult to be reproduced and analyze the data precisely. Moreover, the results are largely influenced by the external environment and subjects themselves. Hence, it is often used to validate the results of other emotion analysis approaches.

2.2.2 External performance

Methods based on external performance usually adopt facial expressions, voice tones, body behaviors, and vehicle behavioral signals to study emotions because these signals can reflect human emotions intuitively. Jeong et al. (2020) proposed a driver facial expression recognition model to monitor driver emotional status with the facial geometric features, which can be embedded into low-power systems. Kessous et al. (2010) designed a speech interaction experiment to arouse different emotions to record the body gesture, facial expression, and speech data. A multimodal emotion recognition approach was then promoted with features of external behaviors. Although these signals are easy to be collected without effect on drivers, the applicability and validity need be further considered. Since body movements are apparently confined to very limited space, drivers are not allowed to act at their willingness. Besides, the audio signals are intermittent during driving process, which means that the emotion status cannot be detected without a conversation. Additionally, external expressions can be intentionally controlled by human to conceal their true feelings.

2.2.3 Physiological responses

Different from external performance, physiological signals reflect the real internal states of human and cannot be hidden deliberately. An emotion recognition model was developed based on the physiological changes while listening to emotional music (Kim & André, 2008). The authors extracted various features of EMG, ECG, respiration and skin conductivity from different aspects such as geometric analysis, time and frequency domains and so on. In another study, the heart rate under fear, happiness and anger emotions was investigated by meta-analysis, which demonstrated that different emotions are associated with the acceleration of heart rate (Cacioppo et al., 2000). The EEG signals of drivers were induced by audio and video stimuli to analyze their emotional states (Gamage et al., 2022). In this work, the changing patterns of EEG was explored and fed into a support vector machine (SVM) to monitor four emotions. These studies prove the possibility to study driver emotion with physiological signals. Unlike EMG and ECG etc., EEG is generated by the neutral nervous system and measures the dynamic neuro-electrical activities with higher resolution in real time. Thus, it has aroused much attention to recognize driver emotion with brain activities.

A variety of features have been used in contemporary research to manifest the emotion information from EEG signals including time domain features, frequency domain features, non-linear features, and time-frequency features. Fan et al. (2018) extracted the statistical features, fractal dimension, power features as well as the higher order crossings features from EEG to build an affective states and mental workload recognition model in simulated driving environment. Gamage et al. (2022) calculated the statistical features, power spectral density (PSD), and wavelet coefficients of EEG for identifying the angry, calm, sad and fear emotions while driving. EEG signals, as introduced in Chapter 2.1, contain a large amount of valuable information of different emotions. Consequently, although the traditionally used features can characterize the emotion patterns in some aspect, the complexity of EEG in various emotions also need to be explored.

Features from the perspective of entropy have been validated that they can mine for the complexity of physiological signals and have been used in disease diagnosis successfully (Ashok et al., 2017; Cao et al., 2015; Zarei & Asl, 2018). Researchers in emotion recognition area also attempt to employ entropy features to find the dynamic fluctuation patterns of EEG under different emotional states. For instance, the SE of EEG and blood volume pulse (BVP) signals were utilized to characterizing angry driving in an on-road experiment (Wan et al., 2015). The results illustrated that the SE value of EEG is obvious smaller during angry driving than normal driving while a contrary phenomenon is observed in SE of BVP. The AE feature of EEG was calculated in order to be recruited in probing into the dynamic emotional information in EEG and identifying four emotional states (T. Chen et al., 2018). Yet, just as the analysis of driver distraction, there is challenge in entropy-based emotion recognition caused by the residual noise in EEG signals. Besides, which entropy feature is better for detecting driver emotion remains unclear. Moreover, information compensation has been observed among different features during pattern recognition in previous study (Hasan & Kim, 2019). So, it needs to be verified whether integrating various entropy features can enhance the emotion recognition model.

Overall, there are a large amount of research on emotion recognition but quite a bit of the present driver emotion research focuses on the anger and fear emotions. As for the detection of driver emotion, the self-report measures are individual dependent and easily influenced by internal and external factors. The features based on driver's external performance are delayed to the internal status and can be concealed artificially. Physiological signals have also been used in emotion recognition with the advantages of instantaneity and veracity, but the traditional features are sensitive to residual noise. In this way, challenge remains in driver emotion detection with physiological signals.

2.3 Driver's abnormal state detection

There are a great number of algorithms in the literature utilized to detect driver's abnormal state, and they can be roughly converged into two groups, i.e., machine learning (ML) and deep learning (DL) methods.

2.3.1 Machine learning-based driver state detection

Conventional approaches such as K-nearest neighbor (KNN), artificial neural networks (ANN) and SVM have been commonly used in driver state identification. Murugan et al. proposed to detect different driver states with KNN, SVM and ensemble classifiers by collecting the ECG signals in a simulated driving experiment (Murugan et al., 2020). They also compared the results of the three classifiers utilizing the heart rate and heart rate variability (HRV) features and proved that ensemble method performs better in multi-class tasks. Naurois et al. introduced two models for driver drowsiness detection and prediction based on ANN and multiple sources data including physiological, facial, and behavioral signals (Jacobé de Naurois et al., 2019). The results showed the possibility to predict the dynamics of driver state. Gwak et al. proposed to detect the alertness of drivers with physiological signals and behavioral signals based on ML (Gwak et al., 2018). They compared the performance of KNN, SVM, random forest (RF), and logistic regression and found that RF method is superior to the others in classifying alertness and drowsiness.

2.3.2 Deep learning-based driver state detection

As a branch of ML, DL has the advantage of learning large datasets with deep neural networks and has been declared to outperform many traditional ML algorithms in the classification of time series (Ismail Fawaz et al., 2019). Thus, it is now widely used to detect human mental states. Li et al. (2016) proposed a mental status recognition method based on nonlinear features of EEG and deep belief network (DBN) and showed that the proposed method can distinguish three kinds of mental states. Kose et al. (2019) introduced a driver state monitoring system utilizing CNN and body movement images. They extracted the temporal and spatial information from images to classify driver's movement decision and mental states. Although these classifiers are commonly used in mental status detection, they can just make decision on the basis of the current input sample and are not able to learn the contextual information in sequential data (Kouchak & Gaffar, 2019).

Driving is a continuous process with features of lasting long and contextual dependency. The current decision is made not only based on the current state but also several previous states. Therefore, the long-term time dependency is important for detecting driver state. On this occasion, it is necessary to improve the performance of algorithms by adding memory to neural networks (Kouchak & Gaffar, 2021). Recurrent neural network (RNN) is a method that can memorize the previous input states and then decide the output. Chamishka et al. (2022) adopted RNN to capture the context of the emotional conversational audio signals and evaluated its capability in real time emotion recognition. However, RNN learns the short-term information of time series rather than keeping long-term memory as the gradient of the loss function either decays or explodes with time (Wollmer et al., 2011). For enhancing the ability of RNN, long short-term memory (LSTM) network is proposed as a variant of RNN, which learns both long and short-term contextual dependency in sequential data (Hochreiter & Schmidhuber, 1997). The vanishing gradient problem in RNN is overcame by adding four gates in each neural cell. The optimal information related to the detection task can then be learnt making it popular in many research fields. Masood et al. (2024) proposed to monitor the neurological stress with LSTM based on physiological data and demonstrated that it is better than traditional ML algorithms in classification tasks. Abbasi et al. (2019) developed a DL framework using LSTM to detect epilepsy with EEG signals and compared its results with SVM. They found that it generates a higher accuracy than that of SVM. Moreover, bidirectional long short-term memory network (BiLSTM) has also aroused the interest of researchers recently since it can learn both the forward and the backward context information. For example, a BiLSTM model was presented to enhance semantic learning and the detection performance in text classification problem (G. Liu & Guo, 2019). The performance of the proposed approach was measured in seven public datasets and was better than those of traditional classifiers.

3 MATERIALS AND METHODS

In this chapter, the data used for driver distraction detection is firstly introduced, which is collected from a realistic driving experiment and a simulated driving experiment, respectively. The dataset for emotion recognition is then described in detail. Thereafter, the methods adopted in the analysis and detection procedures are illustrated.

3.1 Driver distraction datasets

To obtain the needed data (i.e., physiological signals and vehicle behavioral signals) for driver distraction detection, we designed two driving experiments including one in real driving environment and the other in driving simulator. **Article I** and **Article II** share the same dataset that comes from the real road driving environment, while the data for **Article III** is collected in the simulated driving experiment. The detailed descriptions of the two experiments are listed in this section.

3.1.1 Experiment design in real environment

This study was reviewed and approved by the Ethics Committee, Dalian University of Technology.

Participants

Six experienced drivers with driving license were recruited in the experiment, who are physically and mentally in good health. All of them are right-handed and have normal or corrected to normal vision and hearing. Besides, it is also a basis that the subjects should be skilled in using smartphones and WeChat (a popular online chat APP). Additionally, they were told to have a good sleep and not to smoke, take medicine, or drink alcohol, tea, and coffee the day before the experiment. What's more, we inspected the qualification of each subject and got the informed consent from them before the experiment. The subjects were also given oral and written instructions about the upcoming experiment. After the experiment, each participant was paid the reward.

Apparatus

A straight road of Dalian University of Technology was selected to conduct the real-world experiment. A portable and wearable headband, and a Mangold-10 multipurpose polygraph with features of portable and wireless were utilized to gather the EEG signals from the brain scalp with a sample rate of 256 Hz. Since it has been proved that the occipital brain area is correlated with mental status in the literature (Kumar et al., 2020), and the adopted headband has little influence on drivers, the electrodes inside the headband are placed on O1 and O2 on the basis of the International 10-20 System.

Furthermore, previous research shows that car signals also provide valuable information to study driver state and that driving performance changes with driver state (Pavlidis et al., 2016). To validate these findings, sensors were set up in the experiment car so as to obtain the needed vehicle behavioral data. In this experiment, the speed and deceleration data were gathered, and the sampling rate was 50 Hz. The experiment scene is shown in Figure 3.

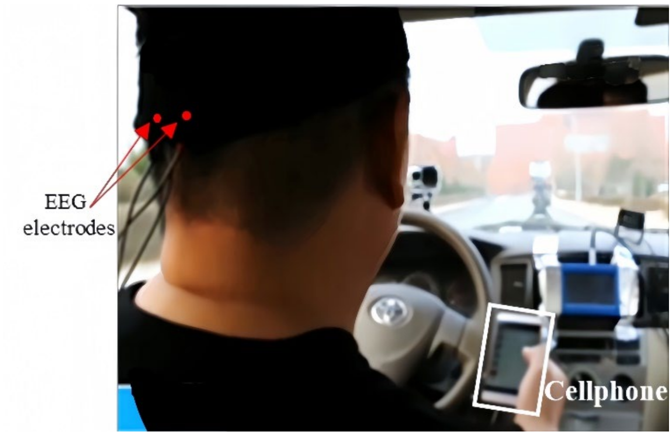


FIGURE 3 The distraction experiment in realistic driving environment.

Procedure

In this experiment, a task of using cellphone while driving was employed to induce driver distraction. The task can be illustrated broadly as: messages were sent to the subject's cellphone while driving on the road, and then the subject was required to use cellphone for no less than 3 seconds.

There were six driving trials in the experiment including one normal driving trials (i.e., focusing on the road all the way) and five distracted trials (i.e., using cellphone while driving). In normal driving, the trial duration was set as at least 6 seconds. As for the distracted driving trials, the duration was about 20 seconds and the cellphone usage task started on around 12 seconds in each trial. A short interval between trials was designed to relax the subjects.

An experimenter sat inside the car along with the experiment going on. His work was to give hints about cellphone usage task start and end to the subjects by sending messages to them in distracted driving trials. The experimenter could not do anything except for the above required. Subjects should pay full attention to safe driving in normal driving trial, while the distracting task was performed in distracted trials. They were asked to drive with full attention at the beginning of these trials, messages would then be sent to their cellphone seconds later. After that, they needed to check the messages for at least three seconds. What's more, a foam obstacle was finally thrown to the road by another experimenter in the end in order to observe the subjects' reactions. During the whole experiment, EEG signals and vehicle behavioral data were acquired.

3.1.2 Experiment design in simulated environment

This study was reviewed and approved by the Ethics Committee, Liaoning Normal University.

Participants

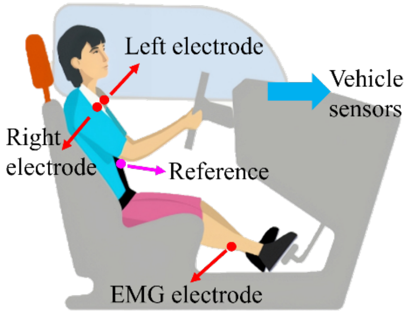
There were sixty experienced and right-handed drivers with driving license participating in the study. They are in good health both mentally and physically and have normal hearing and visual acuity. Besides, they are experienced in using smartphone and online chat APP. In order to avoid the effects of sleep rhythm, the experiment was conducted in the morning, and subjects were told to get enough sleep the night before the experiment. Moreover, they are not allowed to smoke, take medicine, or drink coffee, alcohol and tea the day before the experiment. Prior to the experiment, the qualification of each subject was verified and written informed consent was obtained from subjects. They were introduced the upcoming experiment as well before the start of the experiment. After the experiment, each participant was paid the reward.

Apparatus

The experiment was performed in the laboratory of Liaoning Normal University, which is a simulated driving environment. The driving scenarios were displayed using a 120° viewing screen with Xuan Love QJ-3A1 driving simulator that is composed of control system, simulated cockpit, interactive visual system, exterior accessories and so on (see Figure 4(a)).

The ANT Data Recording System and the 64-channel electrode cap in accordance with the International 10-20 System were adopted to acquire the EEG signals in the experiment. Besides, the EMG and ECG signals were also collected along with EEG. The EMG electrode was laid on the soleus muscle of the right lower limb because it is closely related to the activities of foot and lower leg. The brake needs to be controlled by right foot and leg while driving, so the tension of soleus muscle can reflect driving state. A chest lead III showed in Figure 4(b) was used to obtain the ECG signals. The three electrodes were separately located below the left clavicle (positive electrode), below the right clavicle (negative electrode), and below the right rib (ground reference). A sample rate of 2000 Hz was set to acquire the physiological signals. Additionally, the vehicle behavioral data i.e., the lane position variability (LPV) and velocity (V) was collected with a sampling frequency of 1 Hz to evaluate the changing patterns of driving performance after distraction.

(a)



(b)

FIGURE 4 The experiment apparatus in simulated driving environment. (a) The Xuan Love QJ-3A1 driving simulator and electrode cap. (b) The diagram of ECG, EMG and vehicle sensors.

Procedure

In this simulated experiment, the answering cellphone task was recruited to stimulate driver distraction that subjects should keep talking on the phone while distracted driving process.

Before the formal experiment, all subjects took part in the practice procedure for acquainting themselves with the simulator. The formal experiment lasted for about 60 minutes consisting of 6 blocks. In each block, the drivers pay full attention to driving at the first seven minutes (i.e., normal driving process), and then they would be called by an experimenter in the last three minutes (i.e., distracted driving process). In this condition, they were supposed to talking with the experimenter until the block ends. The protocol of the formal experiment is shown in Figure 5. The EEG, ECG, EMG as well as the vehicle behavioral signals were obtained after the experiment.

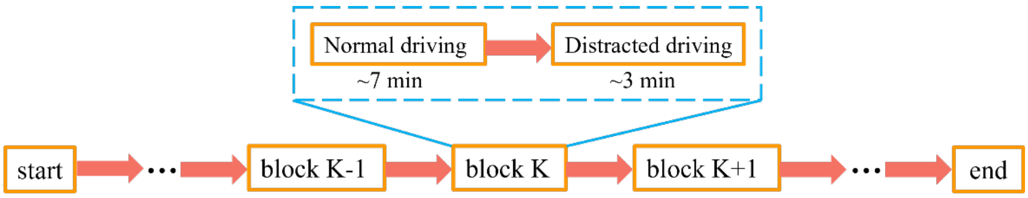


FIGURE 5 The protocol of the formal experiment.

3.2 Emotion dataset

To obtain the EEG signals for emotion recognition, we employed the SJTU Emotion EEG Dataset (SEED), which is a collection of EEG data for multiple purposes emotion study (W. Liu et al., 2022; W.-L. Zheng & Lu, 2015), for analysis in **Article IV**. The detailed descriptions of the dataset are listed in this section.

Participants and materials

The experiment was reviewed and approved by the Ethics Committee, Shanghai Jiao Tong University. Twenty-three mentally and physically healthy subjects consisting of fifteen Chinese and eight French were recruited to gather the EEG data while watching different emotional film clips in their native language. All the subjects were detailed introduced the experiment to make sure they understood it totally prior to the experiment. The written content was obtained from each one of them before the experiment.

The emotional film clips were chosen based on the work of Scharfer et al. so as to induce three kinds of emotions (i.e., negative, neutral and positive) in the experiment (Schaefer et al., 2010). There were five and seven film excerpts of every emotion for Chinese and French subjects separately, and the film duration was limited to about two minutes. A 62-channel NeuroScan System was employed to collect the EEG data in consistent with International 10-20 System. The sampling frequency was kept at 1000 Hz.

Procedure

The experiment was conduct in the laboratory environment. The subjects watched the successive film excerpts sitting comfortably. Each one participated in the experiment once, and there was a total of 15 trials for Chinese subject and 21 trials for French subjects. Before each film clip, there was a hint picture lasting for five seconds to remind the start of the film. After each film segment, the subjects were asked to report their emotions stimulated by the film clip in the 45 seconds self-assessment interval by filling in a questionnaire, which would be regarded as the label to validate the results of emotion recognition. The protocol of the experiment is shown in Figure 6.

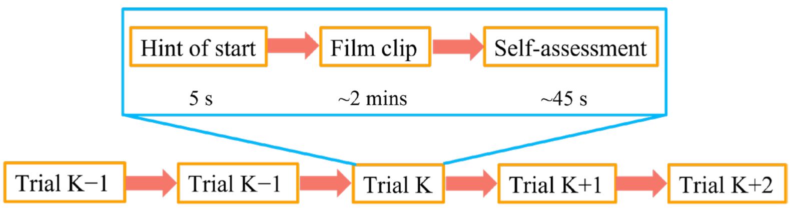


FIGURE 6 The protocol of the emotion experiment.

3.3 Methodology

This section presents the methods used to analyze and detect driver state with the collected physiological signals and vehicle behavioral signals. These methods can be categorized into four groups including preprocessing approaches, features utilized in the dissertation, feature selectors and classifiers used to detect driver state.

3.3.1 Preprocessing

The segments of EEG, ECG, EMG and vehicle behavioral signals corresponding to the duration of each formal trial were extracted from the continuous dataset, which is comprised of signals not only while performing the formal trial but also the preparation and rest intervals. The power line is then eliminated with a bandstop filter of 50 Hz. Thereafter, we extracted five rhythms from the EEG signals using wavelet decomposition, as previous research has pointed out the activities of different frequency bands of EEG are related to the mental status (Nie et al., 2011). Finally, a method based on wavelet transform (WT) was adopted to reject the effects of artifacts in each rhythm.

WT is a method that analyzing and characterizing non-stationary signals in both time domain and frequency domain. Since it can adapt to the requirements of time-frequency analysis and focus on the details of signals, it has been widely applied to signal processing (Rhif et al., 2019). In order to extract the five rhythms from EEG, we should firstly decompose EEG with a mother wavelet $\psi(t)$. The mother wavelet can then be constructed as the family of function $\psi_{j,k}(t)$ by dyadic shifts and dilations:

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi\left(2^{-j}t - k\right) \qquad k,j \in Z \tag{1}$$

The signal $S(t)$ is then described as

$$S(t) = \sum_{k} s_j(k)\phi_{j,k}(t) + \sum_{k} d_j(k)\,\psi_{j,k}(t) \tag{2}$$

where $\phi(t)$ is the scaling function, $s_j(k)$ and $d_j(k)$ are the approximate coefficients and detailed coefficients at jth level, respectively.

Since the obtained EEG rhythms were mixed with artifacts like body movements and blinks, a WT-based approach was employed to remove them. The approximate coefficients and detailed coefficients i.e., $s_j(k)$ and $d_j(k)$ stand for the correlation of the selected mother wavelet and the signal. According to the algorithm, a larger coefficient occurs when the artifact appears (C. Zhang et al., 2018). In this condition, a threshold can then be selected to reduce the abnormal large coefficients and reject the influence of artifacts. The threshold is defined as follows:

$$T_j = \text{mean}\left(C_j\right) + 2 \times \text{std}\left(C_j\right) \tag{3}$$

where C_j is the coefficient at jth level. The value of coefficient is reduced to half of the original value when it is larger than T_j, after which a new set of coefficients are used to obtain the signal without artifacts.

3.3.2 Feature extraction

Various features were calculated in this dissertation including entropy features, time-domain features as well as frequency-domain features:

Multiscale entropy

Multiscale entropy in relative time scales (MSE), firstly proposed by Costa et al. (2002), was an extension of SE. It is calculated in different time scales to characterize the complexity of signals, which can decrease the effects of residual noise retained in physiological signals on the results (Costa et al., 2005). So, it can be used to explore the dynamic patterns in physiological signals along with the changes of driver state. There are two steps to obtain the MSE feature of the physiological signals including the coarse-graining step and SE calculation step. The two steps are introduced in detail in the following content:

Step 1. Considering a signal $\{x_1, \ldots, x_i, \ldots, x_N\}$, the consecutive time series $\{y^{(\tau)}\}$ after coarse graining at relative time scales τ is supposed to be constructed firstly. This can be achieved by dividing the signal into sequential windows without overlapping every τ data points and calculating the average value of these data points within each window (see Figure 7). Each element of the constructed time series $\{y^{(\tau)}\}$ is calculated according to

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i \,, \quad 1 \le j \le N/\tau \,, \tag{4}$$

where N is the length of the signal, N/τ is the length of each coarse-grained time series.

Step 2. Calculate the SE value of the obtained time series $\{y^{(\tau)}\}$. For a given time series $\{y_1, \ldots, y_j, \ldots, y_n\}$, the m-dimensional vector $Y^m(i)$ can be obtained from

$$Y^m(i) = [y(i), y(i+1), \ \ldots, y(i+m-1)], \quad 1 \le i \le n-m. \tag{5}$$

Then, d measuring the distance between $Y^m(i)$ and $Y^m(j)$ can be calculated with

$$d = max[\,|y(i+k) - y(j+k)|\,] \,, \\ 0 \le k \le m-1, i \ne j, 1 \le j \le n-m \tag{6}$$

After that, $B_i^m(r)$ is calculated with

$$B_i^m(r) = \frac{\{\text{the number of } d < r, i \ne j\}}{(n-m-1)} \,, \tag{7}$$

which represents the number of $d < r$ for each i. Following this, the average value of $B_i^m(r)$ set can be obtained with

$$B^m(r) = \frac{1}{n-m} \sum_{i=1}^{n-m} B_i^m(r) \,, \tag{8}$$

Finally, we should calculate the m+1-dimensional vector and repeat this process to get $B^{m+1}(r)$. After all these procedures, the SE can be obtained with

$$\mathrm{SE}(m,r) = \lim_{n\to\infty}\left[-\ln\frac{B^{m+1}(r)}{B^{m}(r)}\right], \tag{9}$$

where r stands for the predefined tolerable distance. When n is finite, SE can be obtained according to

$$\mathrm{SE}(m,r,n) = -\ln\left[\frac{B^{m+1}(r)}{B^{m}(r)}\right]. \tag{10}$$

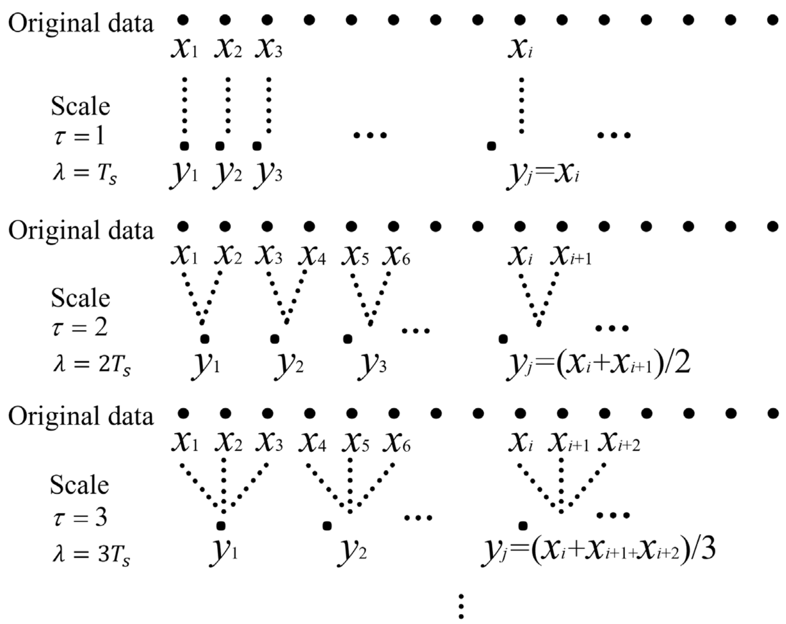


FIGURE 7 Schematic illustration of the coarse-graining process.

It has been validated that the entropy value in single time scale can be influenced by sampling frequency in the literature (Raffalt et al., 2019). Besides, the stretching and compressing effects will come into being while calculating MSE with the resampling of signals (Fallahtafti et al., 2021; J. Zheng et al., 2023). Hence, to overcome the challenges, we explored a variation of MSE, which calculates multiscale entropy in absolute time scale λ (MSaE) rather than in relative time scale. The unit of λ is second, and it can be expressed as

$$\lambda = \tau \cdot T_s = \tau / f_s \tag{11}$$

where T_s is the original sampling period and f_s is the original sampling frequency.

Therefore, the coarse-graining step can be illustrated that the time series $\{y^{(\tau)}\}$ is constructed by means of calculating the mean value in each non-overlapping time window with a length of λ seconds (see Figure 7). It should be noted that the signal's time duration t_D keeps unchanged after resampling, which means $t_{D\_d}=t_D$. After the coarse-graining step, the same step as MSE (i.e., calculating the SE value for $\{y^{(\tau)}\}$) can be utilized to calculate MSaE.

After introducing MSE and MSaE, the entropy-based method to find the appropriate downsample frequency for a signal can be described as the following five steps:

Step 1. Plot the fitting curve of entropy-scale (see Figure 8) after calculating the SE value in various scale τ to find τ_{peak} where the entropy value arrives its maximum value indicating the high relevance of entropy value and time scale (Borowiec et al., 2014).

Step 2. The peak time λ_{peak} when the entropy value is the highest can then be obtained with (11).

Step 3. As λ_{peak} cannot be changed after downsampling, the peak time after downsampling can be expressed as $\lambda_{peak_d} = \lambda_{peak}$.

Step 4. The frequency after downsampling f_{s_d} can then be described as

$$f_{s_d} = \frac{\tau_{peak_d}}{\lambda_{peak_d}} = \frac{\tau_{peak_d}}{\lambda_{peak}} \tag{12}$$

Step 5. For a given peak time scale after downsampling τ_{peak_d}, the corresponding downsampling frequency can be obtained.

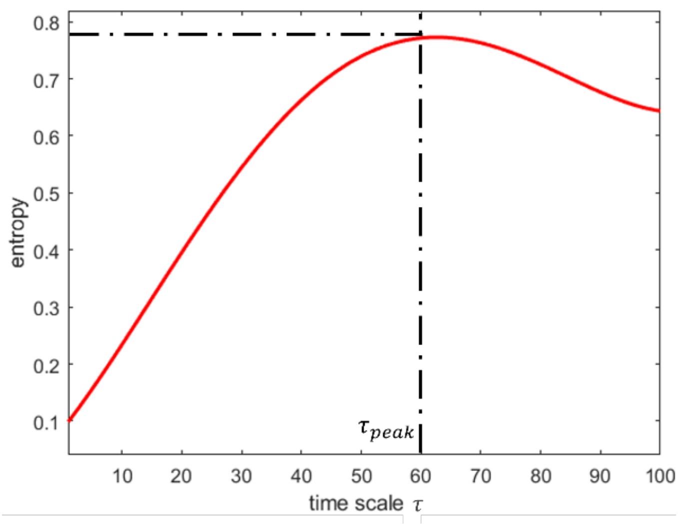


FIGURE 8 The entropy-scale curve of the original alpha rhythm with fs=2000 Hz.

Complexity features

Four traditional entropy features in single time scale were also calculated in this dissertation for the purpose of comparison with multiscale entropy. These features are briefly introduced in Table 1, which can also refer to (Feutrill & Roughan, 2021; Lv et al., 2020; Zarei & Asl, 2018).

TABLE 1 The entropy features used for driver state analysis.

No.	Feature	Formula	Comments
1	Approximate entropy (AE)	$\mathrm{AE}(m,r,N) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} log\, C_i^m(r) - \frac{1}{N-m}\sum_{i=1}^{N-m} log\, C_i^{m+1}(r)$ where $C_i^m(r) = \frac{B_i^m}{N-m+1}$	AE is a measure of the complexity and the statistical quantization characteristic of the signal. B_i^m is the number of matches of dimension m. The definitions of m, r, and N are the same with SE.
2	Fuzzy entropy (FE)	$\mathrm{FE}(m,n,r,N) = ln\frac{O^m(n,r)}{O^{m+1}(n,r)}$ where $O^m(n,r) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{j=1,j\neq i}^{N-m} D_{ij}^m\right)$	In the calculation of FE, the exponential function is used to measure the similarity D_{ij}^m of two vectors i.e., $X^m(i)$ and $X^m(j)$. n is the fuzzy exponent.
3	Rényi Entropy (RE)	$\mathrm{RE} = \frac{1}{1-q} log\left(\sum_{i=1}^{N} p(i)^q\right)$ $q \geq 0$ & $q \neq 1$	For signal $X=\{x_1, \ldots, x_i, \ldots, x_N\}$, $p(i)$ is the probability of choosing x_i in X and $\sum_{i=1}^{N} p(i) = 1$. q is the entropic index.
4	Differential Entropy (DE)	$\mathrm{DE} = -\int_a^b f(x)\, log(f(x))\, dx$	DE is the extension of Shannon entropy. $f(x)$ is the probability density function of the signal. $[a,b]$ is the value interval.

Time-domain features

The waveforms of signals fluctuate along with experiment going on, which could reflect the changes of signals over time. Time-domain features could manifest the changing patterns of signals by analyzing the waveforms of signals (G. Zhang et al., 2020). In this case, the time-domain features were also extracted from the obtain physiological and vehicle behavioral signal. Table 2 shows the used time-domain features in the dissertation.

TABLE 2 The time-domain features for analyzing driver state

No.	Feature	Formula
1	Mean value	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$
2	Standard deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$
3	Skewess	$\text{Skewess} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3 \Big/ (\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2)^{3/2}$
4	Kurtosis	$\text{Kurtosis} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^4 \Big/ (\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2)^2$
5	Zero crossing	$\text{ZeroCross} = \sum_{i=1}^{N-1} \mathbb{1}\{x_i x_{i-1} < 0\}$
6	Root mean square	$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} {x_i}^2}$

Frequency-domain features

Since the spectrum information of signals varies with the experiment conditions, the information related to driver state can be revealed in frequency domain (Heathers, 2014; Phinyomark et al., 2012; G. Zhang et al., 2020). Therefore, the frequency-domain features were calculated to explore the spectral information and compare with multiscale entropy features as listed in Table 3.

TABLE 3 The frequency-domain features for analyzing driver state

No.	Feature	Formula
1	Amplitude spectrum	$\text{AMP} = \int_0^T x(t)\, e^{-i\omega t} dt$, $x(t)$ is the signal, T is the signal period.
2	Power spectrum density	$\text{PSD} = \frac{1}{T}\lvert\text{AMP}\rvert^2$
3	Mean frequency	$\text{FMEAN} = \int_0^{+\infty} f \cdot \text{PSD}(f)\, df \Big/ \int_0^{+\infty} \text{PSD}(f)\, df$
4	Median frequency	$\text{FMED} = \frac{1}{2}\int_0^{+\infty} \text{PSD}(f)\, df$
5	Power of frequency band $[f_1, f_2]$	$\text{P} = \int_{f_1}^{f_2} \text{PSD}(f)\, df$

3.3.3 Feature selection

The extracted features were in various numerical ranges that could not be compared directly. Thus, min-max normalization was used to normalize each feature to the range of [0,1] before inputting the features into classifiers to detect driver state.

$$y = \frac{(y_{max} - y_{min}) * (x - x_{min})}{x_{max} - x_{min}} + y_{min} \tag{13}$$

where y is the value after normalization, x is the value before normalization, x_{max} and x_{min} stand for the maximum and minimum values of x separately, y_{max} and y_{min} represent the maximum and minimum values after normalization, respectively.

In this dissertation, various features from three perspectives were extracted from the gathered signals, which may increase the training time and generate overfitting issue at the classification step to some extent. In order to reduce the computation cost and find the optimal features for detecting driver state, five conventional feature selectors were utilized here.

ReliefF algorithm

ReliefF, a robust algorithm, was firstly proposed by Kononenko in 1994 to deal with noisy and missing data (Kononenko, 1994). It can distinguish the most important features by assigning different weights to features according to how important each feature is relevant to each category (Y. Zhang et al., 2019). In order to study the correlation between feature and category, the nearest neighbors of each feature will be selected from all classes by calculating the Manhattan Distance. High weights appear when the feature is highly related to the category. In this way, the most discriminative feature can be selected among all features.

Non-negative matrix factorization (NMF)

Non-negative matrix factorization (NMF), a widely used method for data analysis, deals with high-dimensional matrix by representing it with two low dimensional matrices (Lee & Seung, 2000). One of the low-dimensional matrices is the non-negative basis, and the other is the weights matrix. After factorization, the valuable information in the original matrix can be kept intact. Then, discriminative features can then be selected based on the obtained weights matrix (Gupta & Xiao, 2011).

Mutual information

Mutual information (MI) can be used to quantify the correlation of the feature in regard to the corresponding categories by means of calculating the information contained in one variable about another variable (Vergara & Estévez, 2014). The value of MI is larger when the feature is more related to the corresponding category. Thereby, the feature selection step based on MI can be described as to find the most relevant feature with largest MI value to the class label. Since it can detect the nonlinear relationship between features and classes and analyze multiple-dimensional feature sets, it is commonly used in feature selection (Doquire & Verleysen, 2013).

Neighborhood component analysis

Neighborhood component analysis (NCA), a supervised learning algorithm, selects features according to the distance metric. The vector indicating feature weights is obtained by optimizing the leave-one-out classification accuracy to the maximum with a regularization term, and the information contained in the feature set retains in this process (Raghu & Sriraam, 2018). After that, the significance of each feature can then be observed directly with feature ranking. The discriminative feature can then be selected. What's more, the increase of irrelevant features has little influence on the results.

Sequential forward selection

The process of sequential forward selection (SFS) can be illustrated as a bottom-up search process (Marcano-Cedeño et al., 2010). It starts with an initial feature matrix with empty value, and then add in features one by one with evaluation functions based on the principle of minimizing the mean square error. During each iteration, one feature is chosen from the rest of the features to add to the predefined selected feature set. The selection iteration continues until the classification accuracy does not change with the number increase in the feature set. Then, the corresponding feature set is determined as the optimal feature set. Because of its simplicity and quick calculation, it has been widely used for reducing matrix dimension.

3.3.4 Classification

To detect driver state and find out how different features compensate for each other while detecting driver state with integrate features, RF is firstly adopted in this dissertation. It was also used to compare the feature selection results of the five feature selectors so that the feature set selected by the most effective feature selection methods can be used in the latter detection procedure. In addition, LSTM-based classifier was employed to learn the contextual information in the signals obtained in driving process and improve the state detection accuracy. The details of the two classifiers are described in this section.

Random forest

RF is a widely used ensemble learning algorithm to classify various tasks developed by Breiman (Breiman, 2001). It outputs the classification results by constructing plenty of decision trees and integrating the decisions by most of the trees. According to the bagging method utilized in RF, it can deal with large number of variables quickly and balance the errors for unbalanced dataset. Moreover, the importance of each input variable can be evaluated in classification tasks. It nowadays has been recruited in different fields for classification and feature selection (R. Zhang et al., 2018). Considering this, it is used not only to detect driver state and estimate the importance of each input feature but also to evaluate the most efficient feature selectors among the five feature selection algorithms in this dissertation.

As an ensemble learning approach, RF ensures that the trained model performs well by exploring the interdependency between features and avoiding overfitting. The principle of random data and variable selection need to be followed aiming at the two tasks (W. Chen et al., 2014). The procedure of RF algorithm can be illustrated as: firstly, select the bootstrap sample Bs in the training set T. Then built the decision trees Tr_b on Bs. In this step, one-third samples are used to estimate the classification error and evaluate the importance of each variable, which are defined as the out-of-bag (OOB) data. Thereafter, the variable candidate sets with a number of M are chosen at random among the holistic variable set in each split. The best way of splitting can then be selected from the candidate sets and be carried out at the node. In order to minimize the bias, the trees grow to the maximum extent and cannot be pruned. Repeat these steps until the minimum classification error is obtained. Thus, the decision tree set $\{Tr_b\}_1^N$ can be obtained. Finally, evaluate the trained RF model in the test set and output most of the trees' results. Algorithm 1 shows its pseudo-code.

Algorithm 1

Input: training set T
N how many decision trees will be built
M how many variables will be chosen for splitting at each node
Training: for each i=1:N **do**
1. Select the bootstrap sample Bs from T.
2. Build decision tree Tr_b on Bs.
3. Select M variable candidate sets randomly at each node of Tr_b.
4. Find the best splitting way among M sets.
5. Build tree Tr_b without pruning.

end for
Output: $\{Tr_b\}_1^N$ the ensemble of trees
X testing set
Classification: Assume $C_b(X)$ is the classification result of each tree. The result of RF $C(X)$ can then be expressed as: $C(X)$=majority vote $\{C_b(X)\}_1^N$

LSTM

LSTM is an extension of RNN, which is proposed to address the vanishing gradient problem and learn both long and short-term contextual information in signals (Karim et al., 2018). It achieves these aims by four neural networks interacting in specific ways in the memory cell as shown in Figure 9. BiLSTM is a variation of LSTM that can learn the long-term context both in forward direction and backward direction. It works by integrating two LSTM layers that consist of memory cells. One layer propagates information from front to back while the other is in the opposite direction.

There are mainly three steps in the information propagation process and each step is implemented with a so-called gate i.e., forget gate, input gate, and output gate (P. Liu et al., 2022). After inputting the data, it will be transferred to the three gates with a sigmoid activation function. In the first step, the information contained in the cell state C_{t-1} is checked to decide what kind of information should be overwritten, which is fulfilled with the forget gate using

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{14}$$

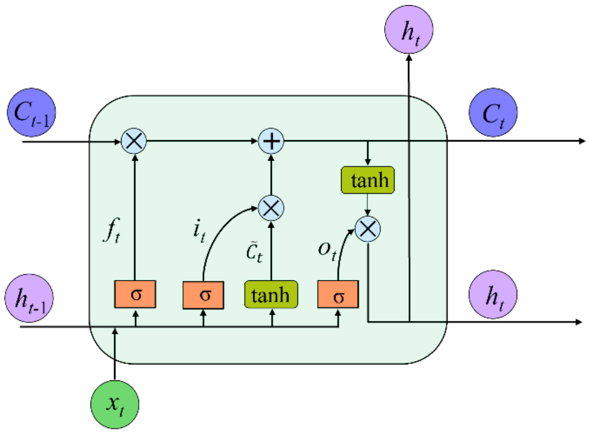


FIGURE 9 The details of LSTM memory cell.

Afterwards, the new information learned from the current inputs should be explored with input gate. During this process, a sigmoid activation function is used to find the updated information with

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{15}$$

Then, the new candidate cell state $\widetilde{C}_t$ is created with a tanh layer based on

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{16}$$

In this way, the old cell state C_{t-1} can be replaced by the new cell state C_t with

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{17}$$

In the last step, the output gate and a tanh activation function are used to decide the cell output h_t as follows:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{18}$$

$$h_t = o_t * \tanh(C_t) \tag{19}$$

In equations (14) to (19), σ and tanh stand for the sigmoid and tanh activation functions separately, x_t is the feature in time t. Besides, W, b and h denote the weight, bias and hidden state of each gate, respectively.

4 OVERVIEW OF INCLUDED ARTICLES

This chapter is the overview of the included articles including the objective, methods, results, and conclusion. The contributions of authors in each article are elucidated as well.

4.1 Article I: Driver distraction detection based on EEG feature fusion using Random Forest

Xin Zuo, Chi Zhang, Jian Zhao, Timo Hämäläinen, and Fengyu Cong. (2024). Driver distraction detection based on EEG feature fusion using Random Forest. In *2023 International Conference on Biomedical Imaging, Signal Processing (ICBSP)*, pp. 104-109. ACM.

Objective

Driver distraction is reported as one of the primary inducements for road crashes (Pandurov, 2023). Under this circumstance, it is important to detect driver distraction in time so as to alert them to pay attention to the driving activity and keep safe. EEG has been regarded as a reliable indicator to detect driver mental state (G. Li et al., 2023). The time domain and frequency domain features of EEG have been widely used to explore the distraction information in EEG. However, there are still challenges in mining the valuable distraction information in EEG because the abundant complexity information is to some extent overlooked in realistic driving environment. The residual noise retained in preprocessed EEG can also decrease the robustness of the results. Besides, it needs to be further studied how different features provide compensation information to each other while detecting driver distraction with fused features. Aiming at the challenges, a driver distraction detection method is proposed on the basis of RF and the entropy feature fusion of EEG in realistic driving scenarios.

Methods

A driving experiment was designed and executed in a real straight road in Dalian University of Technology. The distraction task was to use cellphone while driving. The EEG data of 6 subjects was gathered from O1 and O2 electrodes with sampling frequency of 256 Hz. Firstly, the interested EEG segments corresponding to each driving process were extracted from raw data and the power line of 50 Hz was filtered. Then, alpha rhythm (8-13 Hz) was roughly extracted from EEG segments utilizing wavelet decomposition. EEG was decomposed into four levels followed by reconstructing alpha rhythm with the detailed coefficients at fourth level. In this process, db6 wavelet was selected to decompose the EEG signals because of the similarity between the two waveforms. After that, the WT-based method was adopted to remove artifacts in EEG.

After preprocessing, five entropy features were calculated with a non-overlapping sliding window to explore the distraction information in alpha rhythm including AE, FE, SE, DE, and MSE with time scale of five. Finally, RF was adopted not only to detect driver distraction with each single feature but also to weigh the importance of the five features when detecting driver distraction with fused features.

Results

When classifying driver distraction with signal feature, DE with an accuracy of 72.9% performs better than the other four features. MSE followed by FE and SE ranks the second. The accuracies of them are 68.22%,65.42%, and 63.55%, respectively. AE shows the lowest detection accuracy that reaches only 58.88%. As for detecting driver distraction with fused features, the performance of RF is enhanced and achieves about 80% accuracy. Figure 10 shows the importance of each feature for detecting driver distraction with the fusion of five feature. It is clear that MSE contributes the most among the five entropy features. DE is the second important feature for driver distraction detection. AE still ranks the last one among all features.

Conclusion and discussion

The results illustrate that DE is a better choice to mine the complexity of EEG than other features when detecting driver distraction with single type of feature. In addition, the classification accuracy of RF can be improved with fused multiple features, which validates that different features compensate for each other by evaluating the weights of them. What's more, the weight of MSE is larger than that of DE when using fused features to detect driver distraction. This finding is inconsistent with the results of detecting driver distraction using single type feature. The reason may be that RF classifier cannot learn and memorize the contextual information in different time scales while recognizing driver distraction with single MSE features.

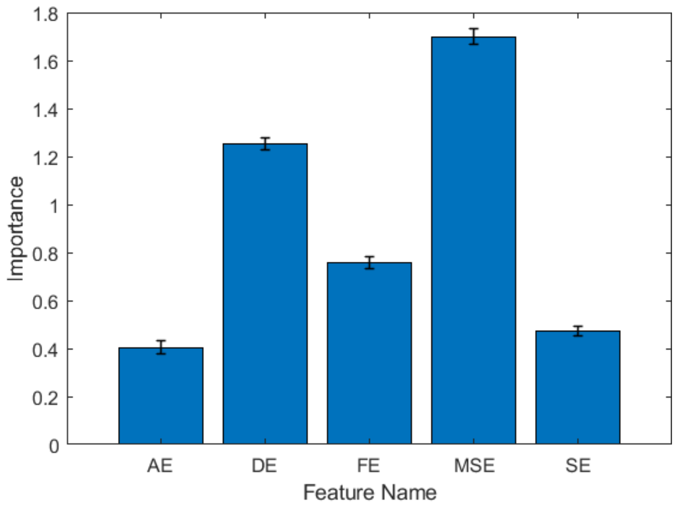


FIGURE 10 The importance of features in distraction detection using fused features. Error bar shows the standard deviation.

Contribution

Xin Zuo conceived the paper, developed the algorithm, analyzed the data, and wrote and revised the manuscript. Chi Zhang contributed to the conceptualization of the paper, data collection, methodology, writing-editing. Jian Zhao, Timo Hämäläinen, and Fengyu Cong supervised the work and edited the manuscript.

4.2 Article II: Driver distraction detection using bidirectional long short-term network based on multiscale entropy of EEG

Xin Zuo, Chi Zhang, Fengyu Cong, Jian Zhao and Timo Hämäläinen. (2022). Driver distraction detection using bidirectional long short-term network based on multiscale entropy of EEG. *IEEE Transactions on Intelligent Transportation Systems*, 23(10), 19309-19322.

Objective

Entropy features can be used to reflect the complexity of EEG signals and detect driver distraction as proved in **Article I**. However, challenges still exist in detecting driver distraction with entropy features. Driving is a continuous long duration activity, the current state of driver is not only affected by the current environment but also the previous status (Kouchak & Gaffar, 2021). So, the long-term contextual information in signals should also be considered. Besides, how entropy feature like MSE changes with driver state is remains unknown. Additionally, vehicle behavioral data can provide information about driver distraction as well. Hence, it is also important to study the driving performance while distraction. In this article, a novel framework based on the MSE in a sliding window and BiLSTM is presented to mine the distraction information of EEG and to detect driver distraction based on hybrid signals.

Methods

The same dataset and preprocessing methods as **Article I** were used in this paper. MSE feature was extracted firstly from EEG to explore the fluctuation patterns after distraction and determine the most distraction position (DP). Then, statistical analysis was performed on vehicle behavioral data (i.e., speed and deceleration) to find out whether changes appear in driving performance before and after distraction. Thereafter, BiLSTM and four other classifiers were utilized to learn and memorize the long and short-term context in MSE and other traditional features to detect driver distraction. The overall architecture of the proposed framework is shown in Figure 11.

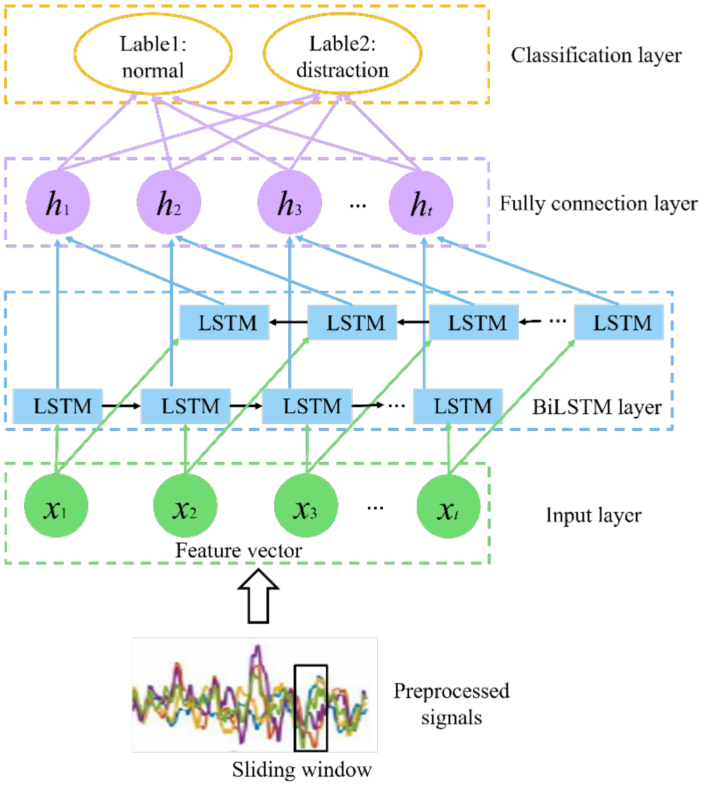


FIGURE 11 Schematic illustration of the BiLSTM framework.

Results

The activity of alpha rhythm obviously increases after the distraction task start. As for the 5-scale MSE, Figure 12 shows that its waveform fluctuates significantly in distracted trials and decreases to the trough soon after the task. The time when MSE reaches its minimum value is then defined as DP. As for the statistical analysis of vehicle behavioral data, significant difference exists in each trial for the speed and deceleration data before and after DP (p<0.05). The mean value and standard deviation were also analyzed, and the results is shown in Figure 13. It is clear that the mean speed tends to decrease after distraction while the deceleration increases after distraction. Moreover, the standard deviations of speed and deceleration data become greater after distraction. Driver distraction was also recognized with different features and classifiers as listed in Table 4. The results show that BiLSTM is superior to other classifiers followed by LSTM. When single type of feature was used to detect driver distraction, the highest accuracy is obtained with MSE reaching at 91.83%, which is apparently better than other traditional entropy features of EEG and vehicle statistical features. If integrating the features of EEG and vehicle data to detect distraction, the performance of BiLSTM can be further enhanced, peaking at 92.48%.

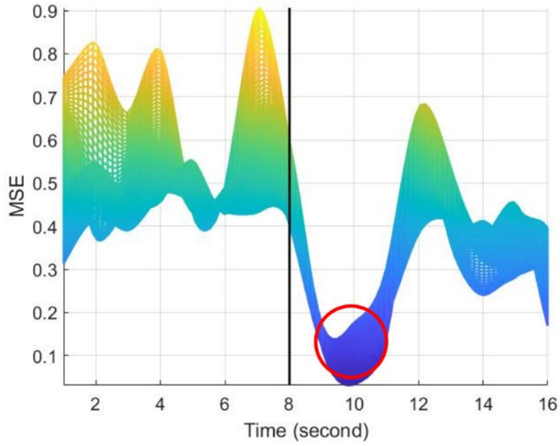


FIGURE 12 The MSE results of alpha band in distracted driving trial. The black solid line shows the onset of using cellphone and the red circle is the distraction position of this trial.

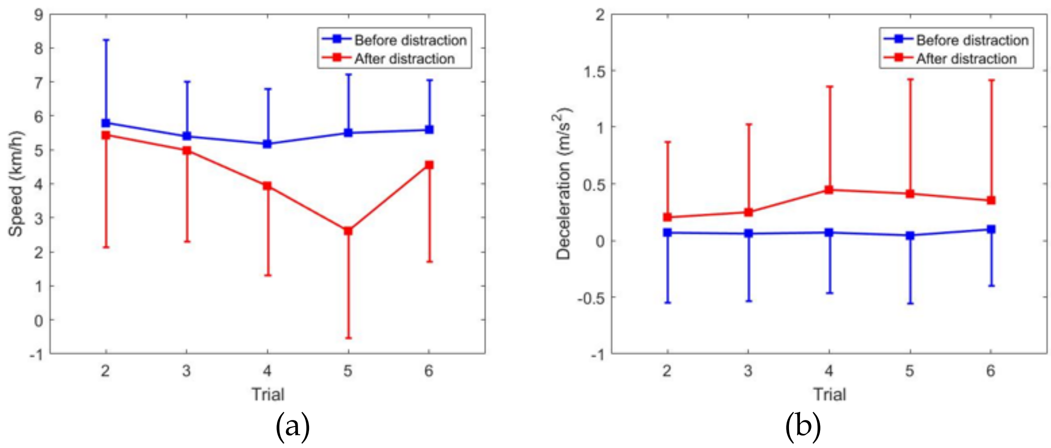


FIGURE 13 The statistical results of the vehicle data. (a) The mean value of speed. (b) The absolute value of the deceleration mean value. Error bar shows the standard deviation.

TABLE 4 The accuracies of different classifiers for different features (%)

Feature	AE	DE	FE	SE	MSE	VS	VS+MSE
BiLSTM	83.29	82.67	76.35	67.01	**91.83**	89.85	**92.48**
LSTM	82.24	81.31	71.03	63.55	89.72	88.79	91.59
CNN	62.62	73.63	62.01	60.32	73.9	67.29	78.5
SVM	52.94	54.6	54.27	56.75	66.72	74.85	77.76
KNN	69.45	71.93	67.12	59.05	65.34	76.84	77.07

Conclusion and discussion

The proposed framework based on MSE of EEG and BiLSTM classifier has been applied in driver distraction in this study. It proves that MSE outperforms other traditional features in mining the complexity of EEG and that BiLSTM is better in learning the contextual information while driving than other classifiers. Besides, the model performance can also be enhanced with features of hybrid signals. Additionally, the value of MSE show a decreasing tendency after distraction. The phenomenon can be illustrated as the complexity of alpha rhythm decrease after distraction. Drivers are more likely to brake and drive in a lower speed after distraction to keep safe. Furthermore, distraction also has negative effect on their ability to control the vehicle, which can be indicated by the increased standard deviation of speed and deceleration data.

Contribution

Xin Zuo conceived the paper, developed the algorithm, analyzed the data, and wrote and revised the manuscript. Chi Zhang contributed to the conceptualization of the paper, data collection, methodology, writing-editing. Fengyu Cong, Jian Zhao, and Timo Hämäläinen supervised the work and edited the manuscript.

4.3 Article III: Driver distraction detection based on MSaE of multi-modality physiological signals

Xin Zuo, Chi Zhang, Fengyu Cong, Jian Zhao and Timo Hämäläinen. (2023). Driver distraction detection based on MSaE of multi-modality physiological signals. *IEEE Transactions on Intelligent Transportation Systems*. Under review.

Objective

According to the work in **Article I** and **Article II**, the complex driver distraction information in EEG can be explored with MSE and the detection accuracy can be improved when utilizing the EEG and vehicle behavioral signals at the same time. Nevertheless, previous research has pointed out that traditional features of physiological signals are sensitive to the retained residual noise and that stretching/compressing effect occurs when extract features at multiple time scales (Costa et al., 2005; J. Zheng et al., 2023). Moreover, various physiological signals can be gathered to study driver distraction, but which one is better in driver distraction detection and whether different physiological features can provide complementary information for each other to improve the detection accuracy still need to be studied. In this article, a driver distraction detection framework is proposed based on MSaE of multiple modalities physiological signals.

Methods

A simulated driving experiment was conducted to collect the needed data including EEG, ECG, EMG, and vehicle behavioral data. For the EEG signals, alpha rhythm was extracted with the commonly used db6 wavelet. Then, the entropy-based method was utilized to determine the proper downsampling rate of EEG for saving computational cost. Afterwards, the EEG data was downsampled to 200 Hz and the WT-based method was adopted to remove the artifacts in EEG. As for the preprocessing of ECG, the downsampled frequency of 256 Hz was obtained with the same downsampling method as EEG. Afterwards, a bandpass filter of 0.7 Hz to 40 Hz was used to reduce the influence of the breath and movements induced baseline drift in ECG. When it comes to EMG data, a bandpass filter with cutoff frequencies at 20 Hz and 500 Hz was employed to obtain the interested EMG signals. Then, a sliding window with length of 125 milliseconds was designed to average the data.

After preprocessing, MSaE and 12 other commonly used features were calculated for EEG. 10 features were extracted from ECG including MSaE, MSE, and 8 features in time-domain and frequency-domain. As for EMG, a total of 8 features were used to analyze driver distraction. The vehicle behavioral signals were also analyzed by calculating the mean and standard deviation of V and LPV.

Thereafter, five conventional feature selectors were used to reduce the redundancy of the large feature set and save time. Following this step, RF was adopted to compare the performance of the five feature selectors to distinguish the most discriminative features. Finally, the selected feature sets were fed into a LSTM classifier to detect driver distraction.

Results

As shown in Figure 14, a minimum value of MSaE occurs soon after the distraction task start, which is obviously smaller than the mean value of MSaE. In contrast, the MSaE value of ECG shows a rising trend after starting to use cellphone. The waveform also fluctuates more apparently after distraction as can be seen in Figure 15. As for EMG, the MSaE value slightly descends after distraction until reaching the minimum value in Figure 16. The statistical results of vehicle behavioral data are similar to that in **Article II**. The mean V declines while the mean LPV increases after distraction. Augments appear in the standard deviation of V as well as LPV. The weight of MSaE is obviously larger than other features and MSE ranks the second for each physiological signal obtained with ReliefF algothrim. Then the selected MSaE feature was used to detect driver distraction utilizing LSTM. Its result is also compared with other features (see Table 5 and Table 6). It is clear that MSaE outperform MSE in recognizing driver distraction with single type signal. Additionally, the MSaE feature of EEG is better than EMG and ECG. Moreover, the performance of the trained model can be enhanced with features extracted from multi-modality signals.

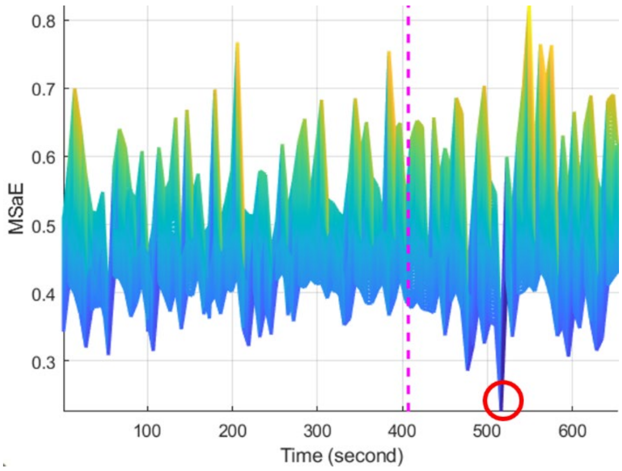


FIGURE 14 The MSaE results of alpha rhythm. The magenta dash line shows the onset of using mobile phone. The red circle is the minimum MSaE value.

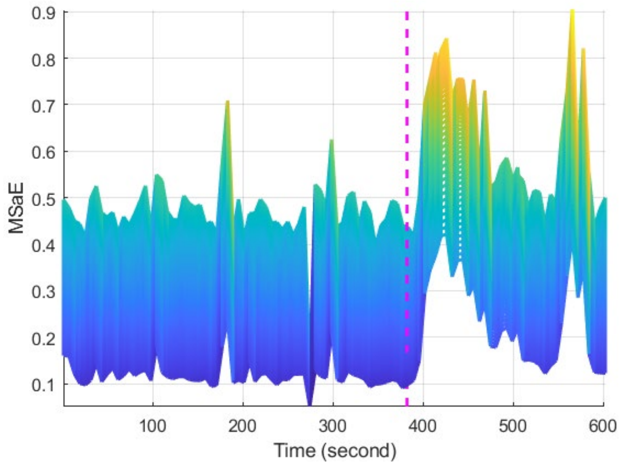


FIGURE 15 The MSaE results of ECG. The magenta dash line shows the onset of using mobile phone.

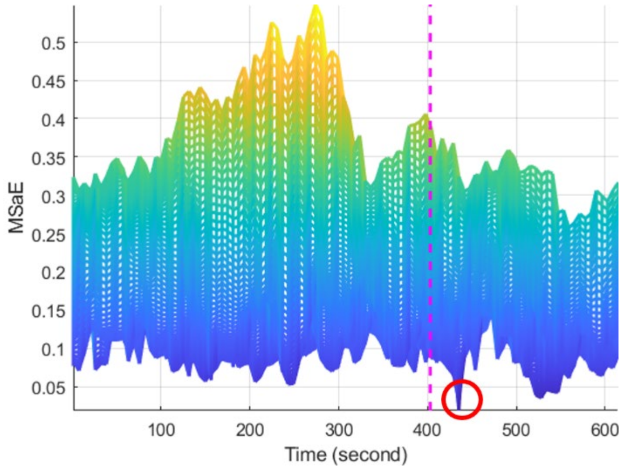


FIGURE 16 The MSaE results of EMG. The magenta dash line shows the onset of using mobile phone. The red circle is the minimum MSaE value.

TABLE 5 The accuracies of different physiological features using LSTM (%)

	EEG	**ECG**	**EMG**	**ALL**
MSaE	74.64	67.09	64.81	78.32
MSE	72.47	66.25	63.74	74.16
All features	63.69	66.11	64.85	63.86

TABLE 6 The accuracies of vehicle and multi-modality features using LSTM (%)

	Vehicle	**Multi-modality**
Accuracy	72.61	81.27

Conclusion and discussion

The results validate the effectiveness and accuracy of the proposed driver distraction detection framework. The MSaE feature of EEG, ECG, and EMG changes obviously in distracted driving and is better than any other calculated physiological features. In addition, the classification results can also be improved with multi-modality signals. In the distracted driving process, the complexity of EEG declines while the heart rate becomes less stable. Furthermore, driver's ability to control the vehicle and muscle recedes after distraction that is indicated with the decreased velocity, increased variability of LPV and V, as well as the reductive EMG complexity.

Contribution

Xin Zuo conceived the paper, developed the algorithm, analyzed the data, and wrote and revised the manuscript. Chi Zhang contributed to the conceptualization of the paper, data collection, methodology, writing-editing. Fengyu Cong, Jian Zhao, and Timo Hämäläinen supervised the work and edited the manuscript.

4.4 Article IV: Cross-subject emotion recognition using fused entropy features of EEG

Xin Zuo, Chi Zhang, Timo Hämäläinen, Hanbing Gao, Yu Fu and Fengyu Cong. (2022). Cross-subject emotion recognition using fused entropy features of EEG. *Entropy*. 24(9), 1281.

Objective

Previous studies have demonstrated that the emotion state of drivers has effect on their performances and behaviors and that it is bound up with driving safety (Picard, 2003). EEG is generated by the spontaneous electrical activity of neutral nervous system, so high interest has been aroused in classifying emotions with it. Various entropy features of EEG have been used in detecting emotion state. However, the ability of manifesting the implicit information of EEG varies with features. In order to reliably detect different emotions in time and alert drivers to keep safe, a framework for cross-subject emotion recognition using the fused entropy features of EEG and BiLSTM is proposed.

Methods

The SEED dataset was used in this article. 12 electrodes in the lateral temporal brain region were selected from the EEG dataset, as previous studies have proved that not all electrodes are relevant to emotion and that more active brain activities can be found in the lateral temporal region experiencing emotional fluctuations (Almahasneh et al., 2014; W.-L. Zheng & Lu, 2015). The selected electrodes were FT7, T7, TP7, P7, C5, CP5, FT8, T8, TP8, P8, C6, and CP6. The data was also downsampled to 256 Hz to reduce the computation and save time. Then the five rhythms including delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-50 Hz) were roughly extracted by decomposing EEG with db6. In the end, the WT-based method was used to eliminate the artifacts in five frequency bands.

Then, MSE and four traditional entropy features (i.e., FE, AE, DE, RE) were utilized to mine for the dynamic emotion information in EEG. After that, BiLSTM classifiers were trained so as to learn the long-term dependency and mutual effect of different features. The classifiers were trained not only with each single feature but also the fused features to compare the performance of BiLSTM.

Results

The activity of gamma rhythm after preprocessing is apparently different under different emotions as shown in Figure 17(a). It is much more active than other two states while in positive state, followed by negative emotion. The waveform of DE shares a similar pattern with FE that largest values appear in positive state and that lowest values occur in neutral state. The fluctuations of AE and RD show analogous tendency, which are in contrast with those of DE and FE. A slightly growing trend can be observed in the MSE feature when the emotion state converts from neutral to positive. As for the results of emotion recognition, MSE achieves the best performance with BiLSTM peaking at 67.9% when recognizing emotion status with single type feature. If fused features are used to detect emotion at the same, the accuracy is further improved to 70.05%.

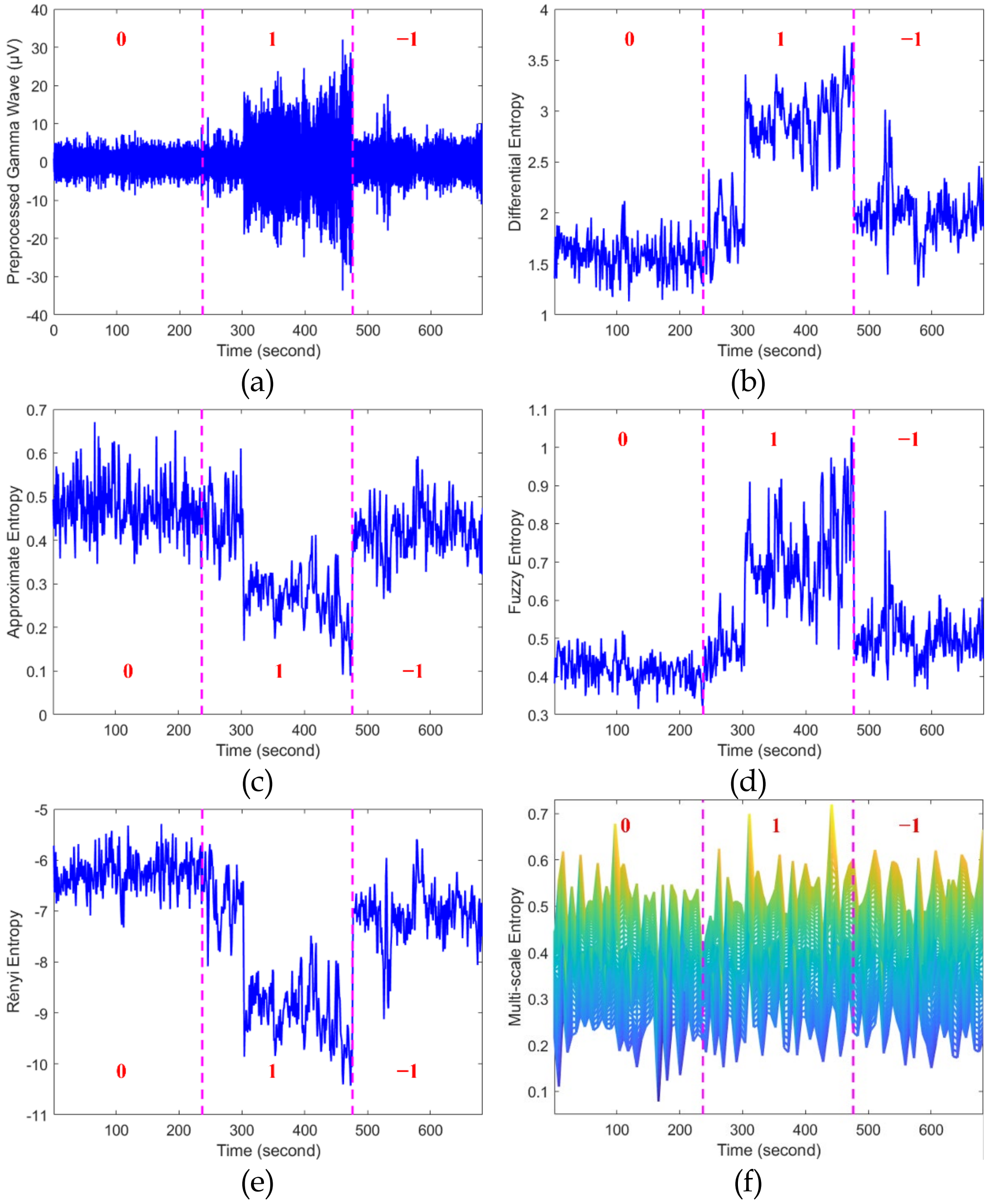


FIGURE 17 The results of EEG analysis. "1", "0", and "-1" are the positive, neutral, and negative emotions, respectively. The magenta dashed lines show the boundaries of different emotions. (a) Preprocessed gamma rhythm. (b) DE. (c) AE. (d) FE. (e) RE. (f) MSE.

Conclusion and discussion

This study validates the feasibility to study and recognize various emotions with the proposed framework. It shows that the MSE feature is much more effective in detecting the emotion status than other conventional entropy features if single type of EEG feature is used in emotion recognition. What's more, different features indeed provide complementary information for each other that is conducive to the enhancement of the classifier's performance.

Contribution

Xin Zuo conceived the paper, developed the algorithm, analyzed the data, and wrote and revised the manuscript. Chi Zhang contributed to the conceptualization of the paper, methodology, and writing-editing. Timo Hämäläinen and Fengyu Cong supervised the work and edited the manuscript. Hanbing Gao and Yu Fu downloaded the data.

5 CONCLUSION AND DISCUSSION

This dissertation aims to explore the complex information contained in physiological signals induced by driver state fluctuation using entropy features and effectively detect driver state for driving safety. The findings of the dissertation are firstly summarized in this chapter. Then, the limitations of the current work are discussed. Finally, some prospective future directions based on the current research are considered.

5.1 Summary of the findings

This dissertation is mainly focused on the detection of abnormal driver state (i.e., driver distraction and emotion) with physiological signals because of their negative effects on road safety. **Articles I**, **II**, and **III** pay close attention to driver distraction detection with ML and DL algorithms. **Article IV** studies emotion recognition based on entropy features and DL algorithm.

Article I proposes a ML method for driver distraction detection in realistic driving environment from the view of complexity of EEG signals. One challenge of the proposed method lies in how to reduce the effect of residual noise on results. In this article, the MSE feature is adopted to manifest the complexity of EEG and reduce the residual noise's influence by calculating entropy in multiple time scales. Then, how different features compensate for each other while detecting distraction with fused features is investigated with the RF algorithm. The results demonstrate that RF can distribute corresponding weights to different features according to their importance in detecting distraction and that the contribution of MSE is obviously the most among all features. Thus, it is useful to reduce the effect of residual noise with MSE feature. But the classification accuracy still needs to be improved.

Article II designs a framework for driver distraction detection based on the MSE feature of EEG and BiLSTM. The fluctuation pattern of alpha rhythm is studied with MSE. The value of MSE declines obviously after distraction and sooner a minimum value appears. Besides, how driver's behaviors and performance change with distraction is also studied with the statistical analysis of behavioral data. The result proves that drivers tend to brake and decrease the speed after distraction to keep safe and that their vehicle controlling ability is significantly influenced in this condition. What's more, the BiLSTM classifer can greatly increase the detection accuracy by means of the long- and short-term contextual information in signals compared to other traditional ML algorithms.

Article III presents a driver distraction detection framework based on the MSaE feature of multi-modality physiological signals gathered in a simulated driving experiment. The biggest challenge in this work is that the entropy values change with sample frequency, thus resampling problem occurs when the signal is downsampled for improving calculation efficiency and saving computational cost. To overcome this problem, MSaE is recruited to explore the complexity of EEG, ECG, and EMG. Next, an entropy-based resampling method is applied to find the appropriate downsampling rate for each signal. The MSaE waveform of EEG shows a similar pattern with the MSE of EEG in **Article II**. MSaE of ECG increases after distraction, which indicates the heart rate is less stable under this circumstance. In addition, driver's ability to control the muscle and vehicle is also greatly affect by distraction. Moreover, it is easy to observe the changes of driver state with MSaE of physiological signals, which is beneficial to explore the complexity and overcome the resampling problem compared to other features. Lastly, more valuable information about driver distraction can be learn with LSTM when incorporating multiple modalities signals, thus contributing to the increasement of detection accuracy.

Article IV introduces a cross-subject emotion recognition framework utilizing fused entropy features of EEG and BiLSTM. The results show that the waveforms of different features fluctuate in various patterns and that MSE is superior to other conventional entropy features for investigating the complex emotion information in EEG. As for the emotion recognition results, the highest accuracy is achieved with MSE while single type of features is used for classification. Additionally, the classifier's performance can be promoted by means of integrating various features. This result also demonstrates that information compensation takes place between features.

In conclusion, this dissertation systematically studies the fluctuation patterns of physiological signals with multiscale entropy of optimized sampling rate in different driver status to eliminate the influence of residual noise and resampling. Then, based on the found patterns, it proposes to automatically detect abnormal driver state with LSTM. The presented framework shows the potential to investigate and detect abnormal mental state with multimodal signals.

5.2 Research limitations

Although the dissertation achieves to study how driver state affects driver's responses and behaviors and detect driver's abnormal state successfully, limitations still exist in this study. Firstly, to detect driver distraction, two driving experiments are designed and conducted to induce distraction related to cellphone usage and gather the needed data. In the first distraction experiment, only six subjects were recruited to take part in the experiment, then EEG and vehicle behavioral data were collected. Even though high detection accuracy has been achieved, the dataset is to some degree kind of small that need to be further validated with more subjects and signals. In the second simulated distraction experiment, the dataset involves sixty subjects and four kinds of signals to detect distraction. But it has been pointed out that realistic driving environment with more uncertain factors is more complex than simulated environment. Additionally, driver distraction is only induced by cellphone usage task, how the proposed framework works in detecting distraction induced by other tasks remains uncertain. So, applying the proposed framework to real driving dataset with more distraction tasks should be further studied.

Secondly, the emotion data selected from SEED contains stimuli and subjects of two native languages i.e., Chinese and French. The stimuli number of the two groups subjects also differs. In **Article IV**, the emotion recognition framework is proposed with the assumption that the differences of stimuli number and native language would not influence the results, since the emotion of a subject is induced by its native language film clips and the categories of stimuli are the same in the two groups subjects. However, there may be effects caused by these differences without drawing attention. Besides, only EEG is utilized to recognize emotion status in this study. It is meaningful to detect driver emotion with more modalities signals so as to integrate their merits.

Thirdly, the artifacts in EEG were removed with a WT-based approach, which requires a threshold to minish the wavelet coefficients. But to what extent the coefficients should be reduced is still very subjective. What's more, the EEG signal was analyzed from the perspectives of time-domain, frequency-domain, as well as complexity. Yet, the spatial information and the interaction among different electrodes is to some extent overlooked. Besides, this dissertation focuses on detecting driver's abnormal state with hand-crafted features, it is necessary to compare the results with the up-to-date end-to-end methods.

Lastly, this dissertation proposes to detect driver's abnormal state with multimodal signals and studies how each kind of signal is influenced by driver's state fluctuation. But how different signals are correlated and whether the interrelationship between the central and peripheral nervous systems are related to driver's state still need to be further explored.

5.3 Future directions

According to the above-described limitations in this dissertation, there are several promising directions in the future work.

Firstly, a driving experiment should be designed to induce driver distraction in real road with more kinds of tasks (like eating and operating devices), subjects and modalities of signals such as blood pressure, respiration, skin conductivity and so on. Then, the proposed driver distraction framework should be tested not only on the obtained dataset but also on various dataset publicly available in order to validate its accuracy and efficiency.

Next, another driving experiment bringing in emotional fluctuations of drivers should be designed in real driving environment. If any words are in the stimuli within various categories, The stimuli should be in the same language and emotions of subjects are induced by their native language. Additionally, it is also a promising direction to analyze driver emotion with multi-modality signals.

Following this, different modality of signals should not only be used to analyze the physiological changes while abnormal driving states independently, but also be used to explore the interrelationships among them. For instance, Are there correlations between different signals in different driving states? Whether and how are the correlations related to driving states? So, it will be an interesting topic to study the interrelationships among multiple modalities signals.

In addition, different objective and interpretable preprocessing methods should be used and compared to eliminate the artifacts in physiological signals like singular spectrum analysis, which is an approach to remove irrelevant artifacts by decomposing signals into interpretable individual components. Then, the spatial information of EEG and how to evaluate the interaction between electrodes should be further studied with state-of-the-art algorithms. To make the proposed framework more reliable, comparison between the current hand-crafted feature engineering and end-to-end algorithms could be made.

Finally, there are also other DL algorithms to learn the long-term context dependency in time series like Transformers, which have showed superiority in natural language processing and computer vision with its multi-head self-attention and parallel inputting mechanisms. Thus, it is potential to apply Transformers and such kinds of DL algorithms to driver's abnormal state detection for exploring the most suitable classifier in this field. Moreover, it is believed that DL is a black box when using it to detect driver state. Hence, it is interesting and prospective to improve the transparency and interpretability of the algorithm in the future as well.

YHTEENVETO (SUMMARY IN FINNISH)

Tämä väitöskirja keskittyy kuljettajan poikkeavan tilan (eli kuljettajan häiriön ja tunnetilan) havaitsemiseen fysiologisten signaalien avulla niiden negatiivisten liikenneturvallisuusvaikutusten vuoksi. Artikkelit I, II ja III kiinnittävät erityistä huomiota kuljettajan häiriön havaitsemiseen koneoppimisen (Machine Learning, ML) ja syvällisen oppimisen (Deep Learning, DL) algoritmeilla. Artikkeli IV tutkii tunnetilan tunnistusta entropiapohjaisten piirteiden ja DL-algoritmin avulla.

Artikkeli I esittää ML-menetelmän kuljettajan häiriön havaitsemiseksi realistisessa ajoympäristössä EEG-signaalien monimutkaisuuden näkökulmasta. Ehdotetun menetelmän haasteena on, miten vähentää jäännösäänien vaikutusta tuloksiin. Tässä artikkelissa moniskaalaentropiapiirre otetaan käyttöön EEG:n monimutkaisuuden ilmentämiseksi ja jäännösäänien vaikutuksen vähentämiseksi. Entropia lasketaan useilla aikaskaaloilla. Sitten tutkitaan käsittelemällä yhdistettyjä piirteitä Random Forest -algoritmin avulla, miten erilaiset piirteet kompensoivat toisiaan häiriön havaitsemisessa. Tulokset osoittavat, että tämä algoritmi voi jakaa vastaavat painot eri piirteille niiden tärkeyden perusteella häiriön havaitsemisessa ja että moniskaalaentropian osuus on selvästi suurin kaikista piirteistä. Siksi on hyödyllistä vähentää jäännösäänien vaikutusta moniskaalaentropiapiirteen avulla, mutta luokittelutarkkuutta on vielä parannettava.

Artikkeli II suunnittelee mallikehyksen kuljettajan häiriön havaitsemiseksi EEG:n moniskaalaentropiapiirteen ja kaksisuuntaisen pitkäaikaisen lyhyen aikavälin muistin perusteella. Alfa-rytmin vaihtelukuvioita tutkitaan moniskaalaentropian avulla. Moniskaalaentropian arvo laskee selvästi häiriön jälkeen, ja minimiarvo ilmenee jo aiemmin. Lisäksi kuljettajan käyttäytymistä ja suoritusta tutkitaan häiriön yhteydessä käyttäytymisdatan tilastollisen analyysin avulla. Tulos osoittaa, että kuljettajat taipuvat jarruttamaan ja vähentämään nopeutta häiriön jälkeen pysyäkseen turvassa ja että heidän ajoneuvon ohjauskykynsä on merkittävästi parempi tässä tilassa. Lisäksi kaksisuuntaisen pitkäaikaisen lyhyen aikavälin muistin luokittelija voi merkittävästi lisätä signaalien havaitsemistarkkuutta pitkän ja lyhyen aikavälin kontekstin tiedoilla verrattuna muihin perinteisiin ML-algoritmeihin.

Artikkeli III esittää kuljettajan häiriön havaitsemisen mallikehyksen moniskaalaentropian absoluuttisten aikaskaalojen perusteella monimuotoisista fysiologisista signaaleista, jotka on kerätty simuloidussa ajokokeessa. Suurin haaste tässä työssä on se, että entropia-arvot muuttuvat näköradan kanssa. Uudelleennäytteistämisongelma ilmenee, kun signaalia alennetaan näytteenottoa varten laskennan tehokkuuden parantamiseksi ja laskennallisten kustannusten säästämiseksi. Tämän ongelman ratkaisemiseksi otetaan käyttöön moniskaalaentropia absoluuttisilla aikaskaaloilla tutkittaessa EEG:n, EKG:n ja elektromyografian (EMG) monimutkaisuutta. Seuraavaksi entropiaan perustuvaa uudelleennäytteistämismenetelmää sovelletaan kullekin signaalille sopivan näköradan löytämiseksi. EEG:n moniskaalaentropia absoluuttisilla aikaskaaloilla (aaltomuoto) näyttää samanlaiselta kuin EEG:n moniskaalaentropia artikkelissa II. Se kasvaa häiriön jälkeen, mikä osoittaa sykkeen olevan vähemmän vakaa tällaisessa tilassa.

Lisäksi kuljettajan kyky hallita lihaksia ja ajoneuvoa vaikuttaa suuresti häiriöön. Kuljettajan tilan muutosten havaitseminen moniskaalaentropian absoluuttisten aikaskaalojen avulla fysiologisissa signaaleissa on helppoa, mikä on hyödyllistä monimutkaisuuden tutkimiseksi ja uudelleennäytteistämisongelman voittamiseksi muihin piirteisiin verrattuna. Pitkäaikaisen lyhyen aikavälin muistin avulla voidaan saada arvokasta lisätietoa kuljettajan häiriöstä, kun otetaan käyttöön useita modaliteettisignaaleja, mikä edistää havaitsemistarkkuuden parantumista.

Artikkeli IV esittelee monitieteellisen tunteentunnistuskehyksen, joka hyödyntää EEG:n sulautettuja entropiapiirteitä ja kaksisuuntaista pitkäaikaista lyhyen aikavälin muistia. Tulokset osoittavat, että erilaisten piirteiden aaltomuodot vaihtelevat erilaisissa malleissa ja että moniskaalaentropia on muita perinteisiä entropiapiirteitä parempi monimutkaisen tunneinformaation tutkimisessa EEG:ssä. Mitä tunnistustuloksiin tulee, korkein tarkkuus saavutetaan moniskaalaentropialla, kun yhtä piirrettä käytetään luokittelussa. Lisäksi luokittelijan suorituskykyä voidaan parantaa eri piirteitä integroimalla. Tämä tulos osoittaa myös, että tietojen kompensaatiota tapahtuu piirteiden välillä.

Yhteenvetona voidaan todeta, että tämä väitöskirja tutkii järjestelmällisesti fysiologisten signaalien vaihtelumalleja moniskaalaentropian avulla optimoidulla näköradalla eri kuljettajatiloissa poistaakseen jäännösäänien ja uudelleennäytteistämisen vaikutuksen. Tämän perusteella voidaan ehdottaa kuljettajan poikkeavan tilan automaattista havaitsemista pitkäaikaisen lyhyen aikavälin muistin avulla. Esitetty mallikehys osoittaa mahdollisuuden multimodaalisia signaaleja käyttäen havaita ja tutkia poikkeavaa mielentilaa.

REFERENCES

Abbasi, M. U., Rashad, A., Basalamah, A., & Tariq, M. (2019). Detection of Epilepsy Seizures in Neo-Natal EEG Using LSTM Architecture. *IEEE Access*, *7*, 179074–179085. https://doi.org/10.1109/ACCESS.2019.2959234

Aksjonov, A., Nedoma, P., Vodovozov, V., Petlenkov, E., & Herrmann, M. (2019). Detection and Evaluation of Driver Distraction Using Machine Learning and Fuzzy Logic. *IEEE Transactions on Intelligent Transportation Systems*, *20*(6), 2048–2059. https://doi.org/10.1109/TITS.2018.2857222

Almahasneh, H., Chooi, W.-T., Kamel, N., & Malik, A. (2014). Deep in thought while driving: An EEG study on drivers' cognitive distraction. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 218–226. https://doi.org/10.1016/j.trf.2014.08.001

Alshalfan, K., & Zakariah, M. (2021). Detecting Driver Distraction Using Deep-Learning Approach. *Computers, Materials & Continua*, *68*(1), 689–704. https://doi.org/10.32604/cmc.2021.015989

Ashok, S., Raj, A., Shashank, P., & Mahalakshmi, P. (2017). Epileptic seizure detection using DWT-based approximate entropy, Shannon entropy and support vector machine: a case study. *Journal of Medical Engineering & Technology*, *42*(1), 1–8. https://doi.org/10.1080/03091902.2017.1394389

Borowiec, M., Rysak, A., Betts, D. N., Bowen, C. R., Kim, H. A., & Litak, G. (2014). Complex response of a bistable laminated plate: Multiscale entropy analysis. *The European Physical Journal Plus*, *129*(10), 211. https://doi.org/10.1140/epjp/i2014-14211-3

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cacioppo, J. T., Klein, D. J., Berntson, G. G., & Hatfield, E. (2000). The psychophysiology of emotion. In R. Lewis, J.M. Haviland-Jones (Eds.), *The Handbook of Emotion (2nd ed.)*, (pp. 173-191). Guilford Press.

Cao, Y., Cai, L., Wang, J., Wang, R., Yu, H., Cao, Y., & Liu, J. (2015). Characterization of complexity in the electroencephalograph activity of Alzheimer's disease based on fuzzy entropy. *Chaos (Woodbury, N.Y.)*, *25*(8), 083116. https://doi.org/10.1063/1.4929148

Chamishka, S., Madhavi, I., Nawaratne, R., Alahakoon, D., Silva, D., Chilamkurti, N., & Nanayakkara, V. (2022). A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling. *Multimedia Tools and Applications*, *81*, 1–22. https://doi.org/10.1007/s11042-022-13363-4

Chen, T., Ju, S., Yuan, X., Elhoseny, M., Ren, F., Fan, M., & Chen, Z. (2018). Emotion recognition using empirical mode decomposition and approximation entropy. *Computers & Electrical Engineering*, *72*, 383–392. https://doi.org/10.1016/j.compeleceng.2018.09.022

Chen, W., Wang, Y., Cao, G., Chen, G., & Gu, Q. (2014). A random forest model based classification scheme for neonatal amplitude-integrated EEG.

Biomedical Engineering Online, *13*(Suppl 2), S4. https://doi.org/10.1186/1475-925X-13-S2-S4

Costa, M., Goldberger, A. L., & Peng, C.-K. (2002). Multiscale Entropy Analysis of Complex Physiologic Time Series. *Physical Review Letters*, *89*(6), 68102. https://doi.org/10.1103/PhysRevLett.89.068102

Costa, M., Goldberger, A. L., & Peng, C.-K. (2005). Multiscale entropy analysis of biological signals. *Physical Review E*, *71*(2), 21906. https://doi.org/10.1103/PhysRevE.71.021906

Das, K., Papakostas, M., Riani, K., Gasiorowski, A., Abouelenien, M., Burzo, M., & Mihalcea, R. (2022). Detection and Recognition of Driver Distraction Using Multimodal Signals. *ACM Transactions on Interactive Intelligent Systems*, *12*(4), 1-28. https://doi.org/10.1145/3519267

(Groot-)Mesken, J., Hagenzieker, M., & Rothengatter, J. A. (2005). Effects of Emotions on Optimism Bias and Illusion of Control in Traffic. In *Traffic and Transport Psychology Theory and Application,* (pp. 203–213). Elsevier. https://doi.org/10.1016/B978-008044379-9/50172-2

Dehzangi, O., Sahu, V., Taherisadr, M., & Galster, S. (2018). Multi-modal system to detect on-the-road driver distraction. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2191–2196. IEEE. https://doi.org/10.1109/ITSC.2018.8569893

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver Inattention Monitoring System for Intelligent Vehicles: A Review. *IEEE Transactions on Intelligent Transportation Systems*, *12*(2), 596–614. https://doi.org/10.1109/TITS.2010.2092770

Doquire, G., & Verleysen, M. (2013). Mutual information-based feature selection for multilabel classification. *Neurocomputing*, *122*, 148–155. https://doi.org/10.1016/j.neucom.2013.06.035

Dua, I., Nambi, A. U., Jawahar, C. V, & Padmanabhan, V. N. (2020). Evaluation and Visualization of Driver Inattention Rating From Facial Features. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *2*(2), 98–108. https://doi.org/10.1109/TBIOM.2019.2962132

Fallahtafti, F., Wurdeman, S. R., & Yentes, J. M. (2021). Sampling rate influences the regularity analysis of temporal domain measures of walking more than spatial domain measures. *Gait & Posture*, *88*, 216–220. https://doi.org/10.1016/j.gaitpost.2021.05.031

Fan, C., Peng, Y., Peng, S., Zhang, H., Wu, Y., & Kwong, S. (2022). Detection of Train Driver Fatigue and Distraction Based on Forehead EEG: A Time-Series Ensemble Learning Method. *IEEE Transactions on Intelligent Transportation Systems*, *23*(8), 13559–13569. https://doi.org/10.1109/TITS.2021.3125737

Fan, J., Wade, J., Key, A., Warren, Z., & Sarkar, N. (2018). EEG-Based Affect and Workload Recognition in a Virtual Driving Environment for ASD Intervention. *IEEE Transactions on Biomedical Engineering*, *65*(1), 43–51. https://doi.org/10.1109/TBME.2017.2693157

Feutrill, A., & Roughan, M. (2021). A Review of Shannon and Differential Entropy Rate Estimation. *Entropy*, *23*(8). https://doi.org/10.3390/e23081046
Flannagan, C. A., Bao, S., & Klinich, K.D. (2012). Driver distraction from cell phone use and potential for self-limiting behavior. *Technical report UMTRI-2012-36,* University of Michigan, Ann Arbor, Transportation Research Institute. https://deepblue.lib.umich.edu/handle/2027.42/108381
Gamage, T. A., Kalansooriya, L. P., & Sandamali, E. R. C. (2022). An Emotion Classification Model for Driver Emotion Recognition Using Electroencephalography (EEG). In *2022 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, pp. 76–82. IEEE. https://doi.org/10.1109/SCSE56529.2022.9905108
Gao, Z., Zhang, K., Dang, W., Yang, Y., Wang, Z., Duan, H., & Chen, G. (2018). An adaptive optimal-Kernel time-frequency representation-based complex network method for characterizing fatigued behavior using the SSVEP-based BCI system. *Knowledge-Based Systems*, *152*, 163–171. https://doi.org/10.1016/j.knosys.2018.04.013
Guettas, A., Ayad, S., & Kazar, O. (2020). Driver State Monitoring System: A Review. In *BDIoT' 19: Proceedings of the 4th International Conference on Big Data and Internet of Things,* pp. 1-7 (Article No. 28). ACM. https://doi.org/10.1145/3372938.3372966
Gupta, M. Das, & Xiao, J. (2011). Non-negative matrix factorization as a feature selection tool for maximum margin classifiers. In *CVPR' 11: Conference on Computer Vision and Pattern Recognition*, pp. 2841–2848. IEEE. https://doi.org/10.1109/CVPR.2011.5995492
Gwak, J., Shino, M., & Hirao, A. (2018). Early Detection of Driver Drowsiness Utilizing Machine Learning based on Physiological Signals, Behavioral Measures, and Driving Performance. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1794–1800. IEEE. https://doi.org/10.1109/ITSC.2018.8569493
Hasan, M. J., & Kim, J. (2019). A Hybrid Feature Pool-Based Emotional Stress State Detection Algorithm Using EEG Signals. *Brain Sciences*, *9*(12), 376. https://doi.org/10.3390/brainsci9120376
He, D., Donmez, B., Liu, C. C., & Plataniotis, K. N. (2019). High Cognitive Load Assessment in Drivers Through Wireless Electroencephalography and the Validation of a Modified N-Back Task. *IEEE Transactions on Human-Machine Systems*, *49*(4), 362–371. https://doi.org/10.1109/THMS.2019.2917194
Healey, J., & Picard, R. (2000). SmartCar: detecting driver stress. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Vol. 4, pp. 218–221. IEEE. https://doi.org/10.1109/ICPR.2000.902898
Heathers, J. A. J. (2014). Everything Hertz: methodological issues in short-term frequency-domain HRV. *Frontiers in Physiology*, *5*. https://doi.org/10.3389/fphys.2014.00177

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention*, *38*(1), 185–191. https://doi.org/10.1016/j.aap.2005.09.007

Hossain, M., Rahman, M., Islam, M., Akhter, A., Uddin, M. A., & Paul, B. K. (2022). Automatic Driver Distraction Detection using Deep Convolutional Neural Networks. *Intelligent Systems with Applications*, *14*, 200075. https://doi.org/10.1016/j.iswa.2022.200075

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, *33*(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1

Jacobé de Naurois, C., Bourdin, C., Stratulat, A., Diaz, E., & Vercher, J.-L. (2019). Detection and prediction of driver drowsiness using artificial neural network models. *Accident Analysis & Prevention*, *126*, 95–104. https://doi.org/10.1016/j.aap.2017.11.038

Jeong, M., Nam, J., & Ko, B. C. (2020). Lightweight Multilayer Random Forests for Monitoring Driver Emotional Status. *IEEE Access*, *8*, 60344–60354. https://doi.org/10.1109/ACCESS.2020.2983202

Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2018). LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access*, *6*, 1662–1669. https://doi.org/10.1109/ACCESS.2017.2779939

Kashevnik, A., Shchedrin, R., Kaiser, C., & Stocker, A. (2021). Driver Distraction Detection Methods: A Literature Review and Framework. *IEEE Access*, *9*, 60063–60076. https://doi.org/10.1109/ACCESS.2021.3073599

Kessous, L., Castellano, G., & Caridakis, G. (2010). Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, *3*(1), 33–48. https://doi.org/10.1007/s12193-009-0025-5

Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(12), 2067–2083. https://doi.org/10.1109/TPAMI.2008.26

Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. In F. Bergadano & L. De Raedt (Eds.), *Machine Learning: ECML-94* (pp. 171–182). Springer Berlin Heidelberg.

Kose, N., Köpüklü, O., Unnervik, A., & Rigoll, G. (2019). Real-Time Driver State Monitoring Using a CNN Based Spatio-Temporal Approach. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC),* pp. 3236-3242. IEEE. https://doi.org/10.1109/ITSC.2019.8917460

Koslouski, J., Wilson-Mendenhall, C., Parsafar, P., Goldberg, S., Martin, M., & Chafouleas, S. (2022). Measuring emotional well-being through subjective report: a scoping review of reviews. *BMJ Open*, *12*(12), e062120. https://doi.org/10.1136/bmjopen-2022-062120

Kouchak, S. M., & Gaffar, A. (2019). Using Bidirectional Long Short Term Memory with Attention Layer to Estimate Driver Behavior. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pp. 315–320. IEEE. https://doi.org/10.1109/ICMLA.2019.00059

Kouchak, S. M., & Gaffar, A. (2021). Detecting Driver Behavior Using Stacked Long Short Term Memory Network With Attention Layer. *IEEE Transactions on Intelligent Transportation Systems*, *22*(6), 3420–3429. https://doi.org/10.1109/TITS.2020.2986697

Kumar, S. P., Selvaraj, J., Krishnakumar, R., & Sahayadhas, A. (2020). Detecting Distraction in Drivers using Electroencephalogram (EEG) Signals. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 635–639. IEEE. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-000118

Laker, R. C., Xu, P., Ryall, K. A., Sujkowski, A., Kenwood, B. M., Chain, K. H., Zhang, M., Royal, M. A., Hoehn, K. L., Driscoll, M., Adler, P. N., Wessells, R. J., Saucerman, J. J., & Yan, Z. (2014). A Novel MitoTimer Reporter Gene for Mitochondrial Content, Structure, Stress, and Damage in Vivo*. *Journal of Biological Chemistry*, *289*(17), 12005–12015. https://doi.org/10.1074/jbc.M113.530527

Lazarus, R. S. (1993). From Psychological Stress to the Emotions: A History of Changing Outlooks. *Annual Review of Psychology*, *44*(1), 1–22. https://doi.org/10.1146/annurev.ps.44.020193.000245

Lechner, G., Fellmann, M., Festl, A., Kaiser, C., Kalayci, T. E., Spitzer, M., & Stocker, A. (2019). A Lightweight Framework for Multi-device Integration and Multi-sensor Fusion to Explore Driver Distraction. In P. Giorgini & B. Weber (Eds.), *Advanced Information Systems Engineering* (pp. 80–95). Springer International Publishing.

Lee, D., & Seung, H. (2000). Algorithms for Non-negative Matrix Factorization. In *NIPS'00: Proceedings of the 13th International Conference on Neural Information Processing Systems,* pp. 535–541. MIT Press.

Li, G., Wu, X., Eichberger, A., Green, P., Olaverri-Monreal, C., Yan, W., Qin, Y., & Li, Y. (2023). Drivers' EEG Responses to Different Distraction Tasks. *Automotive Innovation*, *6*(1), 20–31. https://doi.org/10.1007/s42154-022-00206-z

Li, P., Jiang, W., & Su, F. (2016). 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)*, pp. 367–370. IEEE. https://doi.org/10.1109/EMBC.2016.7590716

Liang, Y., Lee, J. D., & Reyes, M. L. (2007). Nonintrusive Detection of Driver Cognitive Distraction in Real Time Using Bayesian Networks. *Transportation Research Record*, *2018*(1), 1–8. https://doi.org/10.3141/2018-01

Liu, G., & Guo, J. (2019). Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing*, *337*, 325–338. https://doi.org/10.1016/j.neucom.2019.01.078
Liu, P., Sun, X., Han, Y., He, Z., Zhang, W., & Wu, C. (2022). Arrhythmia classification of LSTM autoencoder based on time series anomaly detection. *Biomedical Signal Processing and Control*, *71*, 103228. https://doi.org/10.1016/j.bspc.2021.103228
Liu, W., Zheng, W.-L., Li, Z., Wu, S.-Y., Gan, L., & Lu, B.-L. (2022). Identifying similarities and differences in emotion recognition with EEG and eye movements among Chinese, German, and French People. *Journal of Neural Engineering*, *19*(2), 026012. https://doi.org/10.1088/1741-2552/ac5c8d
Lyv, W., Gonçalves, R., Guo, F., Torrao, G., Radhakrishnan, V., Puente Guillen, P., Louw, T., & Merat, N. (2020). Applying Entropy to Understand Drivers' Uncertainty during Car-following. In *Proceedings of the joint meeting of the 12th International Conference on Measuring Behavior and the 6th Seminar on Behavioral Methods,* Vol. 1, pp. 43-51.
Marcano-Cedeño, A., Quintanilla-Domínguez, J., Cortina-Januchs, M. G., & Andina, D. (2010). Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network. In *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, pp. 2845–2850. IEEE. https://doi.org/10.1109/IECON.2010.5675075
Marchegiani, L., & Posner, I. (2018). Long-Term Driving Behaviour Modelling for Driver Identification. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 913–919. IEEE. https://doi.org/10.1109/ITSC.2018.8569610
Masood, S., Khan, R., Abd El-Latif, A. A., & Ahmad, M. (2024). An FCN-LSTM model for neurological status detection from non-invasive multivariate sensor data. *Neural Computing and Applications*, *36*, 77-93. https://doi.org/10.1007/s00521-022-07117-4
Murugan, S., Selvaraj, J., & Sahayadhas, A. (2020). Detection and analysis: driver state with electrocardiogram (ECG). *Physical and Engineering Sciences in Medicine*, *43*(2), 525–537. https://doi.org/10.1007/s13246-020-00853-8
NHTSA. (2021). FARS (fatality analysis reporting system) encyclopedia. National Highway Traffic Safety Administration, Washington, DC. https://www-fars.nhtsa.dot.gov/Main/index.aspx
NHTSA. (2022). Overview of the 2021 Crash Investigation Sampling System. *Traffic Safety Facts Research Note,* National Highway Traffic Safety Administration, Washington, DC. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813397
Nie, D., Wang, X.-W., Shi, L.-C., & Lu, B.-L. (2011). EEG-based emotion recognition during watching movies. In *2011 5th International IEEE/EMBS Conference on Neural Engineering*, pp. 667–670. IEEE. https://doi.org/10.1109/NER.2011.5910636
Niedenthal, P. M., & Ric F. (2006). Psychology of emotion. New York: Psychology Press.

Öztürk, E., & Erzin, E. (2012). Driver Status Identification from Driving Behavior Signals. In J. H. L. Hansen, P. Boyraz, K. Takeda, & H. Abut (Eds.), *Digital Signal Processing for In-Vehicle Systems and Safety* (pp. 31–55). Springer New York. https://doi.org/10.1007/978-1-4419-9607-7_3

Pakniyat, N., & Namazi, H. (2022). Decoding the coupling between the brain and skin reactions in auditory stimulation by information-based analysis of EEG and GSR signals. *Technology and Health Care*, *30*(3), 623–632. https://doi.org/10.3233/THC-213052

Pandurov, M. (2023). 13 Crucial Texting and Driving Statistics for Canada. https://reviewlution.ca/resources/texting-and-driving-statistics-canada/

Pavlidis, I., Dcosta, M., Taamneh, S., Manser, M., Ferris, T., Wunderlich, R., Akleman, E., & Tsiamyrtzis, P. (2016). Dissecting Driver Behaviors Under Cognitive, Emotional, Sensorimotor, and Mixed Stressors. *Scientific Reports*, *6*(1), 25651. https://doi.org/10.1038/srep25651

Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, *14*(2), 198–202. https://doi.org/10.1016/j.conb.2004.03.015

Phinyomark, A., Thongpanja, S., Hu, H., Phukpattaranont, P., & Limsakul, C. (2012). The Usefulness of Mean and Median Frequencies in Electromyography Analysis. In *Computational Intelligence in Electromyography Analysis - A Perspective on Current Applications and Future Challenges* (pp. 195–220). InTech. https://doi.org/10.5772/50639

Picard, R. W. (2003). Affective computing: challenges. *International Journal of Human-Computer Studies*, *59*(1-2), 55–64. https://doi.org/10.1016/S1071-5819(03)00052-1

Precht, L., Keinath, A., & Krems, J. F. (2017). Identifying the main factors contributing to driving errors and traffic violations – Results from naturalistic driving data. *Transportation Research Part F: Traffic Psychology and Behaviour*, *49*, 49–92. https://doi.org/10.1016/j.trf.2017.06.002

Raffalt, P., Mccamley, J., Denton, W., & Yentes, J. (2019). Sampling frequency influences sample entropy of kinematics during walking. *Medical & Biological Engineering & Computing*, *57,* 759-764. https://doi.org/10.1007/s11517-018-1920-2

Raghu, S., & Sriraam, N. (2018). Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms. *Expert Systems with Applications*, *113*, 18–32. https://doi.org/10.1016/j.eswa.2018.06.031

Ramzan, M., Khan, H. U., Awan, S. M., Ismail, A., Ilyas, M., & Mahmood, A. (2019). A Survey on State-of-the-Art Drowsiness Detection Techniques. *IEEE Access*, *7*, 61904–61919. https://doi.org/10.1109/ACCESS.2019.2914373

Rhif, M., Abbes, A., Farah, I., Martinez, B., & Sang, Y. (2019). Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Applied Sciences*, *9*(7), 1345. https://doi.org/10.3390/app9071345

Sahayadhas, A., Sundaraj, K., Murugappan, M., & Palaniappan, R. (2015). A physiological measures-based method for detecting inattention in drivers

using machine learning approach. *Biocybernetics and Biomedical Engineering*, *35*(3), 198–205. https://doi.org/10.1016/j.bbe.2014.12.002
Sahayadhas, A., Sundaraj, K., & Murugappan, P. (2014). Electromyogram signal based hypovigilance detection. *Biomedical Research (India)*, *25*(3), 281–288.
Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, *24*(7), 1153–1172. https://doi.org/10.1080/02699930903274322
Shahverdy, M., Fathy, M., Berangi, R., & Sabokrou, M. (2020). Driver behavior detection and classification using deep convolutional neural networks. *Expert Systems with Applications*, *149*, 113240. https://doi.org/10.1016/j.eswa.2020.113240
Sharma, L. D., Bohat, V. K., Habib, M., Al-Zoubi, A. M., Faris, H., & Aljarah, I. (2022). Evolutionary inspired approach for mental stress detection using EEG signal. *Expert Systems with Applications*, *197*, 116634. https://doi.org/10.1016/j.eswa.2022.116634
Smith, P., Shah, M., & Lobo, N. da V. (2003). Determining driver visual attention with one camera. *IEEE Transactions on Intelligent Transportation Systems*, *4*(4), 205–218. https://doi.org/10.1109/TITS.2003.821342
Song, F., Tan, X., Chen, S., & Zhou, Z.-H. (2013). A literature survey on robust and efficient eye localization in real-life scenarios. *Pattern Recognition*, *46*(12), 3157–3173. https://doi.org/10.1016/j.patcog.2013.05.009
Stutts, J., Reinfurt, D., & Rodgman, E. (2001). The role of driver distraction in crashes: An analysis of 1995-1999 Crashworthiness Data System Data. *Annual Proceedings / Association for the Advancement of Automotive Medicine*, *45*, 287–301.
Urigüen, J., & Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *Journal of Neural Engineering*, *12*(3), 031001. https://doi.org/10.1088/1741-2560/12/3/031001
Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing and Applications*, *24*(1), 175–186. https://doi.org/10.1007/s00521-013-1368-0
Wan, P., Wen, J., & Wu, C. (2015). A discriminating method of driving anger based on sample entropy of EEG and BVP. In *2015 International Conference on Transportation Information and Safety (ICTIS)*, pp. 156–161. IEEE. https://doi.org/10.1109/ICTIS.2015.7232093
Wang, R., Huang, L., & Wang, C. (2021). Distracted driving detection by sensing the hand gripping of the phone. In *MobiCom '21: Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pp. 828-830. ACM. https://doi.org/10.1145/3447993.3482861
WHO (2023). Road traffic injuries. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
Wollmer, M., Blaschke, C., Schindl, T., Schuller, B., Farber, B., Mayer, S., & Trefflich, B. (2011). Online Driver Distraction Detection Using Long Short-

Term Memory. *IEEE Transactions on Intelligent Transportation Systems*, *12*(2), 574–582. https://doi.org/10.1109/TITS.2011.2119483
Yadawadkar, S., Mayer, B., Lokegaonkar, S., Islam, M. R., Ramakrishnan, N., Song, M., & Mollenhauer, M. (2018). Identifying Distracted and Drowsy Drivers Using Naturalistic Driving Data. In *2018 IEEE International Conference on Big Data (Big Data)*, pp. 2019–2026. IEEE. https://doi.org/10.1109/BigData.2018.8622612
Yang, S., Kuo, J., Lenné, M., Fitzharris, M., Horbery, T., Blay, K., Wood, D., Mulvihill, C., & Truche, C. (2021). The Impacts of Temporal Variation and Individual Differences in Driver Cognitive Workload on ECG-Based Detection. *Human Factors*, *63*(5), 772-787. https://doi.org/10.1177/0018720821990484
Zarei, A., & Asl, B. (2019). Automatic Detection of Obstructive Sleep Apnea Using Wavelet Transform and Entropy-Based Features From Single-Lead ECG Signal. *IEEE Journal of Biomedical and Health Informatics*, *23*(3), 1011-1021. https://doi.org/10.1109/JBHI.2018.2842919
Zhang, B., Chang, W., & Li, X. (2023). Fatigue driving detection based on spatial-temporal electroencephalogram features and parallel neural networks. *Journal of Transportation Systems Engineering and Information Technology*, *23*(2), 315–325. https://link.cnki.net/doi/10.16097/j.cnki.1009-6744.2023.02.033
Zhang, C., Cong, F., Kujala, T., Liu, W., Liu, J., Parviainen, T., & Ristaniemi, T. (2018). Network Entropy for the Sequence Analysis of Functional Connectivity Graphs of the Brain. *Entropy*, *20*(5), 311. https://doi.org/10.3390/e20050311
Zhang, G., Davoodnia, V., Sepas-Moghaddam, A., Zhang, Y., & Etemad, A. (2020). Classification of Hand Movements From EEG Using a Deep Attention-Based LSTM Network. *IEEE Sensors Journal*, *20*(6), 3113–3122. https://doi.org/10.1109/JSEN.2019.2956998
Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*, *59*, 103–126. https://doi.org/10.1016/j.inffus.2020.01.011
Zhang, R., Xiao, X., Liu, Z., Jiang, W., Li, J., Cao, Y., Zhu, Y., & Jiang, D. (2018). A New Motor Imagery EEG Classification Method FB-TRCSP+RF based on CSP and Random Forest. *IEEE Access*, *6,* 44944-44950. https://doi.org/10.1109/ACCESS.2018.2860633
Zhang, W., & Zhang, H. (2021). Research on Distracted Driving Identification of Truck Drivers Based on Simulated Driving Experiment. *IOP Conference Series: Earth and Environmental Science*, *638*(1), 012039. https://doi.org/10.1088/1755-1315/638/1/012039
Zhang, Y., Ren, X., & Zhang, J. (2019). Intrusion detection method based on information gain and ReliefF feature selection. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–5. https://doi.org/10.1109/IJCNN.2019.8851756

Zhang, Z., Wei, S., Zhu, G., Liu, F., Li, Y., Dong, X., Liu, C., & Liu, F. (2018). Efficient sleep classification based on entropy features and a support vector machine classifier. *Physiological Measurement*, *39*(11), 115005. https://doi.org/10.1088/1361-6579/aae943
Zheng, J., Li, Y., Zhai, Y., Zhang, N., Yu, H., Tang, C., Yan, Z., Luo, E., & Xie, K. (2023). Effects of sampling rate on multiscale entropy of electroencephalogram time series. *Biocybernetics and Biomedical Engineering*, *43*(1), 233–245. https://doi.org/10.1016/j.bbe.2022.12.007
Zheng, W.-L., & Lu, B.-L. (2015). Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 162–175. https://doi.org/10.1109/TAMD.2015.2431497

ORIGINAL PAPERS

I

DRIVER DISTRACTION DETECTION BASED ON EEG FEATURE FUSION USING RANDOM FOREST

by

Xin Zuo, Chi Zhang, Jian Zhao, Timo Hämäläinen, and Fengyu Cong 2024

In ICBSP ’23: Proceedings of the 2023 8th International Conference on Biomedical Imaging, Signal Processing, pp. 104-109

https://doi.org/10.1145/3634875.3634890

Driver Distraction Detection Based on EEG Feature Fusion Using Random Forest

Xin Zuo
School of Biomedical Engineering
Dalian University of Technology
Dalian, China
Faculty of Information Technology
University of Jyväskylä
Jyväskylä, Finland
zuoxin93@foxmail.com

Chi Zhang*
School of Biomedical Engineering
Dalian University of Technology
Dalian, China
chizhang@dlut.edu.cn
*Corresponding author

Jian Zhao
School of Automotive Engineering
Dalian University of Technology
Dalian, China
jzhao@dlut.edu.cn

Timo Hämäläinen
Faculty of Information Technology
University of Jyväskylä
Jyväskylä, Finland
timo.t.hamalainen@jyu.fi

Fengyu Cong
School of Biomedical Engineering
Dalian University of Technology
Dalian, China
cong@dlut.edu.cn

***Abstract*—Driver distraction has been one of the primary causes of traffic accidents. Electroencephalography (EEG), a record of the electric potential from the scalp, is considered as a reliable indicator of brain activities. It has been widely used to detect driver distraction. Previous studies have analyzed driver distraction based on time and frequency domain features of EEG. However, challenges still exist in manifesting the distraction information of EEG which contains a large amount of complex information about driver distraction in realistic driving scenarios from the perspective of complexity. In this paper, we propose a driver distraction detection framework using Random Forest (RF) based on the complexity feature fusion of EEG in real driving environment. Five entropy-based features of EEG are firstly extracted with a sliding window. Then, an RF classifier is trained with the extracted features to detect driver distraction. Our results show that differential entropy (DE) with an accuracy of 72.9% achieves the best result while single type feature is applied to detect distraction. The classifier's accuracy is further increased by about 7% using fused multiple features compared with the highest accuracy obtained by single type feature. In terms of feature contribution, we found that the feature with the best distraction detection result by using single type features may not contribute the most when using fused multiple features.**

Keywords—EEG, driver distraction, feature fusion, entropy, random forest

I. Introduction

Driver distraction has been considered to be one of the main causes of car accidents, as it can reduce drivers' ability to manipulate cars and their awareness of potentially dangerous surroundings [1]. The National Highway Traffic Safety Administration (NHTSA) reported that about 2,800 people died and 400,000 were injured in traffic accidents involving driver distraction in 2018 while it rose to 3142 deaths and about 424,000 injuries in 2019 [2], [3]. There is a large number of factors diverting drivers' attention away from driving safely and thus leading to driver distraction, such as mobile phones, passengers, in-vehicle infotainment facilities, and so on. The factor of mobile phone usage ranked even the first among all possible factors [4].

To avoid potentially high-risk situations and prevent the happening of accidents caused by driver distraction, it is important to detect whether a driver is distracted or not. Many methods have been utilized in the literature to study driver distraction. Most of them divide driver distraction into four types (i.e., visual distraction, manual distraction, cognitive distraction, and audio distraction) and mainly focus on one type of distraction [5]-[8]. For example, Le *et al.* [9] designed an n-back digit recall experiment in both simulated and naturalistic driving scenarios to induce cognitive workload. The result showed that high level of distraction would be caused by tasks with high cognitive demand. Although these kinds of experiments can to some extent make contributions to the study of driver distraction, it is usually a combination of two or more distraction types in real driving scenarios. In Le's digit recall experiment, drivers firstly need to hear the voice instructions and bear in mind and then give responses when the same instruction appears. It actually induces both audio distraction and cognitive distraction, which is because driver distraction is caused by the interactions among driver, vehicle, and environment and it usually appears in a form of mixed types in real driving. Hence, challenges still exist in detecting driver distraction efficiently in real traffic.

Many kinds of data have been used to detect driver distraction, such as visual data, physiological data, and vehicle behavioral data in current research [10]-[12]. Physiological signals can provide more reliable information than other data types as they are reflections of the driver's actual internal state. Among all the physiological signals, electroencephalography (EEG) is used more frequently to estimate driver states with the superior performance of representing brain information [13]. For instance, Fan *et al.* [14] collected the EEG data in a simulated driving environment and proposed a time-series ensemble learning method to detect fatigue and distraction based on EEG features. Quantities of research have been done analyzing EEG data from the perspectives of time domain and frequency domain to study driver distraction. Yang *et al.* [15] extracted frequency domain EEG features like power spectral density and log-transformed power of four EEG frequency bands and used them to detect driver distraction. Wang *et al.* [16] utilized the frequency domain features, time domain features as well as time-frequency features to predict the duration of the distraction period and reached a satisfying result. However, EEG signals record electrical activities in the brain regions between pairs of electrodes on the scale. It not only reflects the temporal and spatial information of brain activity but also

This work was supported in part by the National Natural Science Foundation of China under Grant 61703069 and 62001312 and in part by the Fundamental Research Funds for the Central Universities under Grant DUT21GF301.

contains a large amount of complexity information [17]. The traditional most commonly used features may be not enough to manifest the useful complexity information to some degree. Recently, it has been demonstrated that the entropy based methods can explore the complex human state information contained in EEG in many research fields (e.g., sleep staging, disease detection, and mental stress detection). Su *et al.* [18] presented a sleep stage classification system with log energy entropy of EEG and found that the system has high generality which is consistent with the polysomnography records. Wang *et al.* [19] proposed a novel algorithm to predict the preictal state of seizure based on wavelet packet based entropy features of EEG and compared the results with traditional statistical features. The result showed that it reaches a higher classification rate than traditional features. Azami *et al.* [20] extracted the multi-scale entropy (MSE) feature of EEG to observe the dynamic complex brain activity information of Alzheimer's patients and found that MSE could mine for the dynamic EEG changes in an obvious way. Sharma *et al.* [21] extracted sample entropy (SE) and Renyi entropy (RE) at different frequency bands and used them to detect mental stress. Their results show the potential for reliable and timed detection of stress. Zheng *et al.* [22] trained a Deep Belief Network to recognize different emotions with differential entropy (DE) extracted from different brain regions. They found that DE can possess the useful information of EEG and achieve a high emotion classification accuracy.

Although entropy based methods have shown advantages in detecting human states, there is still a challenge in EEG analysis using complexity features. Different features reveal the implicit information of EEG from different aspects [23]-[25]. How the information compensation between different features happens in feature fusion step still needs to be studied. Hence, it is vital to evaluate the importance of different features to improve the classification performance.

A wide range of machine learning methods has been adopted to detect driver distraction in the literature. Random Forest (RF) proposed by Breiman [26] in 2001 is widely applied to classification tasks. It is an algorithm that integrates multiple decision trees according to the idea of ensemble learning. With the superior features of running fast on large databases and estimating variables' importance in classification, it has been used in many fields for classification, feature and channel selection, and so on. Zhang *et al.* [27] presented an advanced RF classifier to select informative features and classify motor imagery EEG with higher accuracy than prevailing approaches. Wang *et al.* [28] proposed an automatic epileptic seizure detection framework using an advanced RF model based on the time-frequency features of EEG and achieved high accuracy in detecting seizures.

In this paper, we propose a driver distraction detection framework based on entropy feature fusion using RF classifier. Non-intrusive wearable EEG sensors are firstly used to gather EEG signals in real driving scenarios. Then, different kinds of entropy based features in a sliding window are calculated to extract the complex distraction information in EEG. After that, the EEG features are fed into RF classifier to detect driver states and to estimate the importance of different features. The results of different kinds of features are finally compared.

The remaining part of the paper is structured as follows. Section II explains the designed experiment. Section III describes the adopted methodologies. The results are shown in Section IV and discussed in Section V. Section VI concludes the paper.

II. Experiment Design

This study was reviewed and approved by the Ethics Committee, Dalian University of Technology. An experiment was conducted on a real straight road at Dalian University of Technology. The Mangold-10 Bluetooth enabled wireless polygraph, a wearable and non-intrusive data acquisition headband, was used to collect EEG data. As the occipital brain region has been demonstrated to be related to driver mental state in previous studies, we put the headband's electrodes on O1 and O2 according to the International 10-20 System. The sample rate was set as 256 Hz.

We recruited six experienced right-handed drivers to participate in the experiment. All of them are mental health and have normal vision and auditory. Besides, they are also required to be experienced in using smartphones. In addition, all subjects are banned from smoking, and consuming drinks containing caffeine and alcohol the day before the experiment. Prior to participating in the experiment, we verified each subject's qualification and obtained the informed consent from them. What's more, written and oral instructions about the experiment were illustrated to all subjects.

The experiment contains one normal driving trail lasting for about 6 seconds and five distracted driving trials with a duration of about 20 seconds. In the normal trial, the subjects were supposed to pay full attention to driving while there were distracting factors in the distracted driving trials. In these trials, they would firstly focus on driving, then they would receive cellphone messages from the experimenter few seconds later. After that, subjects were asked to check the message for at least three seconds. Finally, they need to react to the obstacles that appeared on the road at the end of the trail. The EEG data was gathered from the car starting to stopping.

III. Methodology

It can be divided into three steps to analyze the EEG data including preprocessing, artifacts removal, and feature extraction.

A. Preprocessing

The EEG segments of each trial were extracted from the raw EEG signals at first. Alpha frequency band was then obtained using wavelet decomposition method, since alpha rhythm has been proved to be correlated highly with distraction [29].

Wavelet transform is widely used to extract sub-bands of EEG with the character of multi-resolution. To decompose the signal, a mother wavelet $\psi(t)$ is firstly utilized. Then, the signal can be expressed according to scaled and shifted versions of $\psi(t)$ and a corresponding scaling function $\phi(t)$ [30]. The discrete $\psi(t)$ can be expressed as

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi\left(2^{-j}t-k\right), \qquad k, j \in \mathrm{Z} \tag{1}$$

The signal $S(t)$ then is defined as

$$S(t) = \sum_k s_j(k)\phi_{j,k}(t) + \sum_k d_j(k)\psi_{j,k}(t) \tag{2}$$

where $s_j(k)$ and $d_j(k)$ are the approximate and detailed coefficients at level j.

B. Artifacts Removal

The obtained alpha frequency band contains artifacts like blinks that need to be removed in this step. A wavelet-based method was applied in this paper. According to large coefficients usually generated at the places where artifacts appear, we can decrease these large coefficients by thresholding technique [31], [32]. The threshold can be defined as

$$T_j = mean(C_j) + 2 \times \mathrm{std}(C_j) \tag{3}$$

where C_j is the wavelet coefficient at jth level of decomposition. If the value of any coefficient is greater than the defined threshold, it will be halved. A new set of wavelet coefficients are then reconstructed to obtain the artifacts removed signal.

C. Feature Extraction

We extracted five entropy based features to explore the distraction information of alpha frequency band including approximate entropy (AE), fuzzy entropy (FE), SE, DE, and MSE. As the first three features are more frequently used in the literature than the other two features, we mainly introduce the algorithm of DE and MSE in this part.

1) Differential entropy

DE is an extension of Shannon entropy so that it can be used to reflect the complexity of continuous variables [33]. It has been validated that DE is more accurate than features like energy spectrum and asymmetrical features in recognizing different emotion types based on EEG [34]. The calculation formula of DE is

$$\mathrm{DE} = -\int_a^b f(x)\log\left(f(x)\right)\mathrm{d}x \tag{4}$$

where $f(x)$ represents the probability density function of the continuous variable and $[a,b]$ shows the taking value interval. If the variable obeys Gaussian distribution $N(\mu,\sigma^2)$ approximately, its DE can then be expressed as

$$\mathrm{DE} = -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}\right)\mathrm{d}x = \frac{1}{2}\log 2\pi e\sigma^2 \tag{5}$$

2) Multi-scale entropy

MSE can mine for the complexity information of signals in different time scales [20]. It involves two steps in MSE feature calculation: the coarse-graining process and SE calculation. The algorithm can be detailed as follows:

In the first step, for EEG signal $\{x_1, \ldots, x_i, \ldots, x_N\}$, a consecutive coarse-grained time series $\{y^{(\tau)}\}$ should be constructed corresponding to the scale factor τ. The coarse-grained time series $\{y^{(\tau)}\}$ is defined as

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i\,, \quad 1 \le j \le N/\tau \tag{6}$$

In the second step, The SE of time series $\{y^{(\tau)}\}$ is then calculated according to the following sub-steps.

- An m dimension vector $Y^m(i)$ can be made up firstly for time series $\{y_1, \ldots, y_j, \ldots, y_n\}$,

$$Y^m(i) = [y(i), y(i+1), \ldots, y(i+m-1)],\ 1 \le i \le n-m \tag{7}$$

- Define d as the absolute value of the maximum difference between the corresponding elements in vectors $Y^m(i)$ and $Y^m(j)$,

$$\begin{aligned} &d = \max[|\, y(i+k) - y(j+k)\,|], \\ &0 \le k \le m-1, i \ne j, 1 \le j \le n-m \end{aligned} \tag{8}$$

- Then count the number of $d<r$ for each i where r is the given threshold and $B_i^m(r)$ can be expressed as

$$B_i^m(r) = \frac{\{\text{the number of } d < r,\, i \ne j\}}{(n-m-1)} \tag{9}$$

- The set of $B_i^m(r)$ are then averaged and the average value $B^m(r)$ is defined by

$$B^m(r) = \frac{1}{n-m}\sum_{i=1}^{n-m} B_i^m(r) \tag{10}$$

- Add the dimension by 1 and repeat the above process, then $B^{m+1}(r)$ is obtained. After all the steps, SE is calculated by

$$\mathrm{SE}(m,r) = \lim_{n\to\infty}\left[-\ln\frac{B^{m+1}(r)}{B^m(r)}\right] \tag{11}$$

D. Random Forest Classifier

Algorithm 1

Input: T the training set
N the number of decision trees to be built
M the number of variables chosen for splitting at each node
Training: for each i=1:N **do**
1. Draw a bootstrap sample Bs from T.
2. Build tree Tr_b on bootstrap sample Bs.
3. Randomly select M candidate sets at each node of tree Tr_b, and find the best split among M sets.
4. Build tree Tr_b without pruning.
end for
Output: the ensemble of trees $\{Tr_b\}_1^N$
X the testing set
Classification: Assume $C_b(X)$ is the classification result of each tree. Then the result of RF
$C(X)$=majority vote $\{C_b(X)\}_1^N$

RF is a typical bagging model integrating multiple decision trees according to the idea of ensemble learning. In order to ensure the generalization ability of the model, the principles of random data and feature selection are followed while building each tree [35]. It works as follows [26]: bootstrap sample Bs is selected from the training set T at first, and decision trees Tr_b can then be built on the bootstrap samples. During this step, there is one-third of the samples are left called out-of-bag (OOB) data to calculate the classification error and to get estimates of variable importance in the classification step. After that, M variable candidate sets are randomly selected from the whole variable set at each split. Then select the best splitting way from M candidate sets and split at the node. To ensure a low bias, each tree is grown to the largest extent without pruning. After this step, the RF tree will repeat the above steps recursively until it is large enough to obtain the minimum classification error and then all decision tress $\{Tr_b\}_1^N$ are obtained. Finally, the trained RF classifier can be used to classify the testing set by voting for all trees' results. The pseudo-code of RF is shown in Algorithm 1.

IV. Results

After obtaining the five entropy based features of alpha frequency band from all subjects, an RF model was trained using Algorithm 1. The data of five subjects was selected as the training set and the remained data was used as the testing set. RF classifier adopted the feature matrixes and corresponding label vectors of the training set to optimize the model parameters and then output the binary classification results of the testing set. In this paper, to compare the performance of RF using single type feature with that of multiple features, we trained classifiers for each type of feature and fused multiple features, respectively. The results of different features are shown in Table I. "ALL" stands for all the five entropy features of EEG.

TABLE I. The mean accuracies of different features (%)

Features	AE	DE	FE	MSE	SE	ALL
RF	58.88	72.9	65.42	68.22	63.55	79.51

As for the results of single type EEG features, it is shown in Table I that the mean accuracy of DE reaches 72.9%, which is obviously higher than the results of the other four entropy

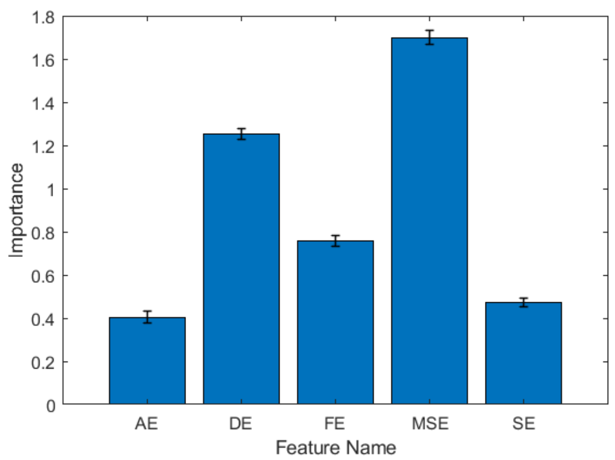


Fig. 1. The importance of features in distraction detection using fused features. Error bar shows the standard deviation.

features. MSE, followed by FE and SE, ranks second with a classification accuracy of 68.22%. The AE feature leads to the lowest accuracy of the RF model, which is only 58.88%. The model performance increases significantly when using multi-features to detect driver distraction peaking at 79.51%. Furthermore, with RF's ability to output the importance of different features during classification, we also estimated each feature's importance in the feature fusion distraction detection process. The results are shown in Fig. 1.

It is clearly shown in Fig. 1 that there are significant differences in the importance of different EEG features for detecting driver distraction. MSE shows greater importance on the feature fusion distraction detection than the other four features. DE following MSE is the second important feature to detect distraction. It is not in accordance with the classification results based on single type feature in which DE display a higher accuracy than MSE. Besides, AE still shows the lowest importance among all features and the importance of FE and SE lies between the importance of AE and DE.

V. Discussion

Driver distraction has drawn a growing concern in recent years with the widespread usage of smartphones and advanced in-vehicle infotainment facilities [7]. An RF model to detect driver distraction is trained using five kinds of complexity based EEG features. The results are compared not only among single type features but also between fused multiple features and single type features.

The results of driver distraction detection in Table I indicate that the mean accuracy of DE achieves the highest than other single type entropy based features. It is consistent with the results in [22], as DE has been proved to be a better feature to recognize human mental states. Moreover, the classification accuracy of multiple features, with an accuracy of 79.51%, is notably greater than that of any single type feature. Since different features can compensate for the inadequacy of each other [23], more useful information of distraction of several different kinds of EEG features can be learn by RF model. Thus, a better performance can be reached in detecting driver distraction. The accuracy in our study is better than that in the research in [36], who adopted the power spectral density and coherence indicators of EEG to detect distraction and reached an accuracy of 73.4% with RF. A similar accuracy of about 80% was obtained for driver state detection in [37] with time domain and frequency domain features. By estimating the importance of different features in distraction detection using multiple features, we can know from Fig. 1 that MSE ranks the first among all features and DE is apparently less important than MSE on the classification results. The finding is not accordant with the classification results utilizing single type features in this paper but it corresponds to the results in our previous work. The BiLSTM model achieved the highest accuracy based on the MSE feature of EEG in [38], which might be because of the advantages of BiLSTM to learn the bidirectional long and short-term dependency of EEG. In this case, the MSE feature of EEG may reveal more contextual information in EEG and thus leading to the highest importance in the feature fusion distraction detection. Furthermore, the results in Table I and Fig. 1 also shows that single type of feature with which the best classification result is obtained may be not necessarily the feature with the most contribution after feature fusion.

VI. Conslusion

In this paper, we propose a driver distraction detection framework applying the RF classifier based on fused complexity features of EEG. It proves that DE feature is the best choice to explore the complex distraction information in EEG than other entropy features used in the literature while detecting driver distraction based on single type feature. Besides, the classifier's performance is greatly enhanced by fusing different EEG features, which demonstrates that different features can provide complementary distraction information of EEG. Additionally, MSE contributes the most among all features to detect driver distraction by fused features. It confirms that a feature achieving the best distraction detection result while using single type features may not contribute the most for driver distraction detection utilizing multiple features. Our work provides a machine learning method to detect driver distraction in real driving situations from the perspective of complexity features of EEG signals. It is useful for mining the complex dynamic brain activity information and driver distraction detection systems in real traffic.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (grant number: 61703069 and 62001312) and in part by the Fundamental Research Funds for the Central Universities (grant number: DUT21GF301).

REFERENCES

[1] B. Darrow, “Distracted driving is now an epidemic in the U.S.,” Fortune, Sep. 14, 2016, [online], Available: http://fortune.com/2016/09/14/distracted-driving-epidemic/

[2] "Distracted driving 2018," National Highway Traffic Safety Administration, Washington, USA. Apr. 2020. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812926

[3] "Distracted driving 2019," National Highway Traffic Safety Administration, Washington, USA. Apr. 2021. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813111

[4] "Research on distracted driving in China," Ford Motor Company, and Tsinghua University, Beijing 2017.

[5] D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 7, pp. 1239-1258, Jul. 2010.

[6] J. D. Lee, "Dynamics of driver distraction: The process of engaging and disengaging," Ann Adv Automot Med, vol. 58, pp. 24-32, Mar. 2014.

[7] A. Gaffar and S. M. Kouchak, "Quantitative driving safety assessment using interaction design benchmarking," in Proc. IEEE SmartWorld Ubiquitous Intell. Comput. Adv. Trusted Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), San Francisco, CA, USA, 2017, pp. 1-8.

[8] N. Lavie, "Attention, distraction, and cognitive control under load," Curr. Dir. Psychol., vol. 19, no. 3, pp. 143-148, 2010.

[9] A. S. Le, T. Suzukic, and A. Hirofumi, "Evaluating driver cognitive distraction by eye tracking: From simulator to driving," Transp. Research Interdiscip. Persp., vol. 9, pp. 100307, 2021.

[10] M., Gjoreski, M., Ž Gams, M., Luštrek, P., Genc, J., U Garbas, and T., Hassan, "Machine learning and end-to-end deep learning for monitoring driver distractions from physiological and visual signals," IEEE Access, vol. 8, pp. 70590-70603, 2020.

[11] M. Taherisadr, P. Asnani, S. Galster, and O. Dehzangi, "ECG-based driver inattention identification during naturalistic driving using Mel-frequency cepstrum 2-D transform and convolutional neural networks," Smart Health, vol. 9-10, pp. 50-61, 2018.

[12] I. Jegham, A. Ben Khalifa, I. Alouani, and M. A. Mahjoub, "A novel public dataset for multimodal multiview and multispectral driver distraction analysis: 3MDAD," Signal Process.: Image Commun., vol. 88, pp. 115960, Aug. 2020.

[13] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 2, pp. 596-614, Jun. 2011.

[14] C. Fan et al., "Detection of train driver fatigue and distraction based on forehead EEG: a time-series ensemble learning method," IEEE Trans. Intell. Transp. Syst., pp. 1-11, Nov. 2021.

[15] L. Yang, W. Guan, R. Ma, and X. Li, "Comparison among driving state prediction models for car-following condition based on EEG and driving features," Accid. Anal. Prev., vol. 133, pp. 105296, 2019.

[16] S. Wang, Y. Zhang, C. Wu, F. Darvas, and W. A. Chaowalitwongse, "Online prediction of driver distraction based on brain activity patterns," IEEE Trans. Intell. Transp., vol. 16, no. 1, pp. 136-150, 2015.

[17] Z. Gao et al., "An adaptive optimal-Kernel time-frequency representation-based complex network method for characterizing fatigued behavior using the SSVEP-based BCI system," Knowl.-Based Syst., vol. 152, pp. 163-171, 2018.

[18] C. Su, L. Ko, J. Juang, and C. Hsu, "Hybrid sleep stage classification for clinical practices across different polysomnography systems using frontal EEG," in Processes. vol. 9, 2021.

[19] Y. Wang, J. Cao, X. Lai, and D. Hu, "Epileptic state classification for seizure prediction with wavelet packet features and random forest," in 2019 Chinese Control And Decision Conference (CCDC), pp. 3983-3987, 2019.

[20] H. Azami, D. Abásolo, S. Simons, and J. Escudero, " Univariate and multivariate generalized multiscale entropy to characterise EEG signals in Alzheimer’s disease," Entropy. vol. 19, 2017.

[21] L. D. Sharma, V. K. Bohat, M. Habib, A. M. Al-Zoubi, H. Faris, and I. Aljarah, "Evolutionary inspired approach for mental stress detection using EEG signal," Expert Syst. Appl., vol. 197, p. 116634, 2022.

[22] W. Zheng, J. Zhu, Y. Peng, and B. Lu, "EEG-based emotion classification using deep belief networks," in IEEE Int. Conf. Multimed. Expo, Chengdu, China, 2014, pp. 1-6.

[23] H. Liu and L. Yu, “Toward integrating feature selection algorithms for classification and clustering,” IEEE Trans. Knowl. Data En., vol. 17, pp. 491-502, April 2005.

[24] R. Desai, P. Porob, P. Rebelo, D. R. Edla, and A. Bablani, "EEG data classification for mental state analysis using wavelet packet transform and Gaussian process classifier," Wireless Pers. Commun., vol. 115, no. 3, pp. 2149-2169, 2020.

[25] M. J. Hasan and J. Kim, "A hybrid feature pool-based emotional stress state detection algorithm using EEG signals," in Brain Sciences. vol. 9, 2019.

[26] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5-32, 2001.

[27] R. Zhang et al., "A new motor imagery EEG classification method FB-TRCSP+RF based on CSP and random forest," IEEE Access, vol. 6, pp. 44944-44950, 2018.

[28] X. Wang, G. Gong, N. Li, and S. Qiu, "Detection analysis of epileptic EEG using a novel random forest model combined with grid search optimization," Front. Hum. Neurosci., vol. 13, 2019.

[29] J. Murphy, C. Devue, P. M. Corballis, and G. M. Grimshaw, " Proactive control of emotional distraction: evidence from EEG alpha suppression," Front. Hum. Neurosci., vol. 14, 2020.

[30] O. Rioul and M. Vetierli, "Wavelets and signal processing," IEEE Signal Proc. Mag., vol. 8, no. 4, pp. 14-38, Oct. 1991.

[31] C. Zhang et al., "Network entropy for the sequence analysis of functional connectivity graphs of the brain," Entropy, vol. 20, no. 5, pp. 311, Apr. 2018.

[32] C. Zhang, L. Sun, F. Cong, and T. Ristaniemi, "Spatio-temporal dynamical analysis of brain activity during mental fatigue process," IEEE Trans. Cogn. Dev. Syst., vol. 13, no. 3, pp. 593-606, 2021.

[33] A. Feutrill and M. Roughan, "A review of shannon and differential entropy rate estimation," Entropy. vol. 23, 2021.

[34] R. Duan, J. Zhu, and B. Lu, "Differential entropy feature for EEG-based emotion classification," in 2013 6th International IEEE/EMBS Confer. on Neural Engineering (NER), 2013, pp. 81-84.

[35] W. Chen, Y. Wang, G. Cao, G. Chen, and Q. Gu, "A random forest model based classification scheme for neonatal amplitude-integrated EEG," Biomed. Eng. Online, vol. 13, no. 2, p. S4, 2014.

[36] N. M. Yusoff, “Visual Cognitive Driving Distraction Detection using EEG,” Ph.D. dissertation, Ecole nationale supérieure d'arts et métiers-ENSAM, Université de technologie de Petronas, Seri Iskandar, Perak, Malaisie, 2019.

[37] L. Yang, and Q. Zhao, “An aggressive driving state recognition model using EEG based on stacking ensemble learning,” J. Trans. Saf. Secur., pp. 1-22, May. 2023.

[38] X. Zuo, C. Zhang, F. Cong, J. Zhao, and T. Hämäläinen, "Driver distraction detection using bidirectional long short-term network based on multiscale entropy of EEG," IEEE Trans. Intell. Transp., pp. 1-14, 2022.

II

DRIVER DISTRACTION DETECTION USING BIDIRECTIONAL LONG SHORT-TERM NETWORK BASED ON MULTISCALE ENTROPY OF EEG

by

Xin Zuo, Chi Zhang, Fengyu Cong, Jian Zhao, and Timo Hämäläinen 2022

IEEE Transactions on Intelligent Transportation Systems, 23(10), 19309-19322

http://doi.org/10.1109/TITS.2022.3159602

Driver Distraction Detection Using Bidirectional Long Short-term Network Based on Multi-scale Entropy of EEG

Xin Zuo, Chi Zhang, Fengyu Cong, Jian Zhao, and Timo Hämäläinen

***Abstract*—Driver distraction diverting drivers' attention to unrelated tasks and decreasing the ability to control vehicles, has aroused widespread concern about driving safety. Previous studies have found that driving performance decreases after distraction and have used vehicle behavioral features to detect distraction. But how brain activity changes while distraction remains unknown. Electroencephalography (EEG), a reliable indicator of brain activities has been widely employed in many fields. However, challenges still exist in mining the distraction information of EEG in realistic driving scenarios with uncertain information. In this paper, we propose a novel framework based on Multi-scale entropy (MSE) in a sliding window and Bidirectional Long Short-term Memory Network (BiLSTM) to explore the distraction information of EEG to detect driver distraction based on multi-modality signals in real traffic. Firstly, MSE with sliding window is implemented to extract the EEG features to determine the distraction position. Statistical analysis of vehicle behavioral data is then performed to validate driving performance indeed changes around distraction position. Finally, we use BiLSTM to detect driver distraction with MSE and other traditional features. Our results show that MSE notably decreases after distraction. Consistent with the result of MSE, driving performance significantly deviates from the normal state after distraction. Besides, BiLSTM performance of MSE outperforms other entropy-based methods and is better than behavioral features. Additionally, the accuracy is improved again after adding MSE feature to behavioral features with a 3% increasement. The proposed framework is useful for mining brain activity information and driver distraction detection applications in realistic driving scenarios.**

***Index Terms*—Driver distraction, EEG, driving performance, MSE, BiLSTM**

I. Introduction

NOWADAYS, the traffic system is highly developed with the increasing number of cars on road. Unfortunately, traffic accidents have become frequent. The World Health Organization reported that over 1.35 million people were killed and about 50 million were injured due to traffic accidents all over the world in 2018 [1]. According to the National Highway Traffic Safety Administration (NHTSA), one of the major contributory factors of traffic accidents is driver distraction [2].

Driver distraction is a diversion of attention away from activities critical for safe driving (i.e., the task of driving) toward a competing activity (e.g., using a cell phone) [3]. In a survey released by Ford Motor Company and Tsinghua University in 2017, almost 39% of respondents caused or nearly caused an accident because of distraction [4]. Due to the use of cell phones and advanced infotainment systems in cars, drivers' attention is often taken away from roads while driving, thus reducing their abilities to control the vehicles and to aware of the surroundings causing more accidents [5]-[7]. What's more, it can also increase the reaction time to the upcoming obstacles [8]. Using cell phones even topped the list for distracted driving reported in 2017 [4].

The existing research about driver distraction mechanism usually could be divided into four different types: manual distraction, audio distraction, visual distraction, and cognitive distraction [9]-[12]. In the previous studies, the subjects were usually asked to perform a specific secondary task while driving for a certain type of distraction to obtain distracted data and then to analyze driver distraction. For instance, "operate devices" tasks are usually used to get the manual distraction signals. Wollmer *et al.* [13] chose eight tasks (e.g., adjust radio sound settings, switch the TV mode and so on) as manual distraction conditions to get the vehicle behavioral signals. They found that tasks with different levels of difficulty would cause different degrees of distraction. As for cognitive distractions, Anh Son *et al.* [14] set an n-back task of digit recall in a simulated situation as well as in a naturalistic situation to impose cognitive workload. The result showed that tasks accompanied by high

This work was supported in part by the National Natural Science Foundation of China under Grant 61703069 and in part by the Fundamental Research Funds for the Central Universities under Grant DUT18RC(4)035. Xin Zuo and Chi Zhang have contributed equally to this work. *(Corresponding author: Chi Zhang; Jian Zhao)*

Xin Zuo and Fengyu Cong are with the School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Linggong Road #2, Dalian 116024, China, and also with the Faculty of Information Technology, University of Jyväskylä, Mattilanniemi 2, Jyväskylä FIN-40014, Finland (e-mail: zuoxin93@foxmail.com; cong@dlut.edu.cn).

Chi Zhang is with the School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Linggong Road #2, Dalian 116024, China (e-mail: chizhang@dlut.edu.cn).

Jian Zhao is with the School of Automotive Engineering, Faculty of Vehicle Engineering and Mechanics, Dalian University of Technology, Linggong Road #2, Dalian 116024, China (e-mail: jzhao@dlut.edu.cn).

Timo Hämäläinen is with the Faculty of Information Technology, University of Jyväskylä, Mattilanniemi 2, Jyväskylä FIN-40014, Finland (e-mail: timo.t.hamalainen@jyu.fi).

cognitive demand had effects on the drivers' eye involuntary movement and would cause a high level of distraction. Although these experiment designs of different kinds of distraction tasks mainly lead to a specific type of distraction to some degree and contribute to the research of distraction mechanism, it usually involves more than one type of distraction in realistic driving situations. For example, when drivers are asked to adjust the radio sound settings for collecting the manual distraction signals in a distraction experiment, they firstly should find where the button is and then turn it to the required place. This process involves not only the manual distraction but also the visual distraction. As driver distraction is a product of the driver-vehicle-environment interaction, its forms are not fixed and usually a combination of different types of distraction in real scenarios. Thus, there is still challenge in the detection of driver distraction in real driving scenarios.

In fact, previous studies have explored many kinds of sensing technologies to detect driver distraction. A commonly used technology is video camera capturing drivers' facial and body behaviors (e.g., gaze movement and head pose) [15]. This kind of method can easily collect the visual data and conveniently detect distraction. However, the results are sensitive to the illumination, facial occlusion, and drivers' behavioral habits. The Controller Area Network-Bus (CAN-Bus) data providing the vehicle behavioral information is also widely utilized in the field. It mainly includes the speed, lateral position, and steering wheel angle, etc. [16]-[19]. The facilities of this kind of method are easy to obtain and quite low cost, but the signals are subject-dependent and influenced by the weather and traffic conditions easily [20]. There is also some work that has been done by using the microphone to collect the acoustic signals for driver distraction detection [21], [22]. The performance of this approach is acceptable, but it just works for audio distraction. Moreover, wearable sensors have also been employed to get the human's physiological signals such as electroencephalography (EEG), electrocardiogram (ECG), and electrooculography (EOG) [23]-[25]. EEG is the predominant and most used signal among all physiological signals. Although physiological signals provide more reliable results for representing drivers' real internal state, the data collecting process is intrusive and may to some degree affect drivers' behaviors.

In recent years, with the development of portable and less intrusive equipment as well as the multi-sensor collection techniques, more and more researchers tend to use hybrid signals to study driver distraction, for it is widely agreed that no single signal alone could provide sufficient information about driver distraction [26],. Li *et al.* [27] collected data from video cameras, microphone arrays and CAN-Bus to model drivers' behavior while executing secondary tasks. Zhang *et al.* [28] utilized vehicle behavioral signal, EMG, acoustic signal as well as visual signal for detecting driver distraction. Lechner *et al.* [29] designed a lightweight framework involving signals of driver movement and GPS position to recognize driver inattentiveness. In addition, Almahasneh *et al.* [30] conducted a simulated driving experiment to study how EEG and driving performance changes because of cognitive secondary tasks. They found that the effects of driver distraction can be clearly seen in the lane keeping ability and accidents occurrence level. As EEG provides reliable information of brain activities, and vehicle behavioral signals reflect the changes of driving performance, it is obvious that the system performance would be improved if EEG and vehicle behavioral signals are employed at the same time to develop driver distraction detection system. In this context, we propose a multi-modality driver distraction detection framework in real driving scenarios based on EEG and vehicle behavioral signals.

The paper is organized as follows. Section II lists the related works about the literature review. Section III introduces the accomplished experiment details and the captured signals used in our research. The adopted methodologies are described in Section IV. Results of the study are presented in Section V and discussed in Section VI. Finally, Section VII concludes the paper.

II. RELATED WORK

There are two major parts in driver distraction detection including the feature extraction part and classification part. Various features are adopted to explore distraction information existing in different types of data. Many researchers analyzing driver distraction based on EEG in the literature, and they usually extract the frequency domain, or the time domain features of EEG to mine the distraction information. Fan *et al*. [31] calculated the energies of different EEG rhythms and their ratios as frequency domain features of EEG and used them for distraction detection. Yang *et al*. [32] extracted the power spectral density and log-transformed power of four EEG waves to evaluate the distraction detection performance. Barus *et al.* [33] used not only frequency domain features but also time domain features of EEG like kurtosis and Hurst Exponent to detect drivers' cognitive load. However, it still only achieved about 70% accuracy. As we all know, EEG signals are recorded directly on the scalp surface and the reflection of the driver's internal electrical activity originated by the brain [34]. But they are also quite complex containing a large amount of information [35]. The conventional features can surely represent the frequency or time domain features of EEG, but how to manifest the complexity of EEG still needs to be further studied. It can reflect the non-linear dynamic changes of the brain activities and manifest the complex distraction information by analyzing the EEG signals of the distracted drivers from the perspective of complexity. The complexity-based algorithm is currently widely utilized in many other areas (e.g., fatigue analysis, emotion classification and sleep staging) and has shown advantages. For example, Gao *et al.* [36] implemented the wavelet entropy to investigate the EEG-based fatigue driving and found that a significant difference exists between the alert and fatigue states. Zheng *et al.* [37] trained an advanced deep learning model with differential entropy. The results showed that differential entropy possesses accurate and stable information of EEG data for emotion classification. Tang [38] applied sample entropy and fuzzy entropy to represent the features of the sleeping EEG data. He demonstrated that the two kinds of features could effectively improve the accuracy of sleep staging. In these studies, the dynamic changes of EEG during fatigue and sleep are reflected through the complexity-based features.

In spite of the existing advantages, a significant challenge still remains in the distracted EEG analyzing procedure based on complexity features. Actually, no matter what kinds of preprocessing methods are adopted, the artifacts cannot be eliminated completely and will still exist to a certain extent [39]. In this case, the residual noise will be included in the complexity of EEG while calculating the complexity-based feature, and the robustness of the obtained result will be relatively poor. Multi-scale entropy (MSE) can reduce the influence of residual noise on the results by calculating complexity features in different time scales [40] and has been successfully used in many fields. Azami *et al.* [41] calculated the MSE feature as well as variate MSE features of the EEG signals to observe the dynamical complex properties in the EEG signals gathered from Alzheimer's disease (AD). They found that MSE could characterize the EEG changes in a detailed way. Luo *et al.* [42] proposed a method based on MSE to detect driver fatigue. The result showed that MSE can obviously present the fatigue features and effectively improve the accuracy of fatigue detection.

As for how to recognize driver distraction, there are many classification techniques utilized in the literature to detect whether a driver is distracted or not. Traditional methods like support vector machines (SVM) and multiple adaptive regression trees (MART) are widely employed in various research areas. Liao *et al.* [43] proposed a method to detect cognitive distraction based on the optimal features extracted by SVM and classify driver state based on SVM. It also compared the SVM performance between two different driving situations. Wu *et al.* [44] used SVM to recognize flight operating patterns based on physiological parameters and reached an average accuracy of 0.84. Torkkola *et al.* [45] described an approach based on MART to find the inattention duration while driving according to the vehicle data. It could detect about 80% of the driver inattention time segments. Besides, deep learning methods has been applied to recognize mental status in the literature. In the work of Wu *et al.* [46], they proposed a stacked contractive sparse antoencoders network to detect the mental status of pilots. What's more, they also designed a gamma deep belief network to study the cognitive status of pilots, which could learn the EEG features with simplest network structure [47]. However, traditional deep learning methods usually learn the information in a single time point and it has been revealed that the time dependencies are critical in predicting human's mental status [48]. Recurrent Neutral Network (RNN) is a typical deep learning method with memory that could keep the information from the contexts and then make decisions. However, the vanishing gradient problem occurs when the input data is too long (i.e., to keep long-term memory) [49]. As a variant of RNN, Long Short-term Memory Network (LSTM) has the property of capturing both short and long-term dependencies, which has been successfully applied to many time-series classification tasks such as driver identification, seizure detection and driver behavior classification [50]-[52]. It is realized by adding memory blocks in the hidden unit to mine for and store critical information for classification over long time periods [53]. Kouchak *et al.* [54] proposed a distraction recognition method based on LSTM and validated that it outperformed multilayer neural network (MLP) for considering dependency between input data. Wollmer *et al.* [13] used LSTM to model the long-term dependency in vehicle behavioral data for detecting driver distraction. They also made a comparison with SVM and found the classification accuracy of LSTM was obviously higher than that of SVM. Recently, the Bidirectional Long Short-term Memory Network (BiLSTM), an improvement of LSTM, has been proved to achieve better performance than traditional one directional LSTM in fields of sleep apnea detection and text classification [55], [56]. As BiLSTM learns long-term dependencies both from former time steps to later time steps and from later time steps to former time steps, it could learn and store more useful information thus improving the performance of the model [57].

In this paper, we propose a framework for driver distraction detection based on MSE with a sliding window and vehicle features. Our approach, using BiLSTM, is to model the bidirectional contextual information in EEG and vehicle behavioral data captured in real scenarios. To collect the distracted EEG and vehicle behavioral signals, a distracted driving experiment is firstly performed in realistic driving situations. The MSE in a sliding window is then implemented to extract the features of the captured EEG signals. Statistical analysis is performed on the vehicle behavioral data to find out whether significant differences appear in driving behaviors before and after distraction. After that, BiLSTM classifier is utilized to learn the time dependent relationships in the extracted MSE and vehicle statistical features and to detect driver states. Finally, the classification accuracy of BiLSTM is compared with four different types of traditional classifiers.

Fig. 1. The distraction experiment scene.

III. EXPERIMENT DESIGN

In order to collect the data reflecting the physiological and vehicle behavioral changes of the distracted drivers, we conducted an experiment in realistic driving scenarios. This section is a description of the participants, the data collection system, and the procedure.

A. Participants

This study was reviewed and approved by Ethics Committee, Dalian University of Technology. There were six right-handed subjects without mental illness or neurological diseases involved in the experiment. All subjects have normal or corrected to normal vision and normal auditory. A driving license and driving experience are required for each subject. All of the subjects owe smartphones and are experienced in using WeChat (an online chatting APP in China). What's more, they are banned from consuming coffee, tea, alcohol as well as smoking the day before the experiment. The qualification of each subject was verified and informed consent from each subject was obtained prior to the experiment.

B. Data collection system

The experiment was conducted on a real straight road at Dalian University of Technology. The Mangold-10 Bluetooth enabled wireless multipurpose polygraph, a portable and non-intrusive data acquisition headband, was used to collect drivers' EEG signals. It transmitted the EEG data via wireless Bluetooth. As the headband is designed to have little effects on drivers' behaviors, and more importantly, previous studies have demonstrated that the occipital brain region is related to driver mental state [58], [59], we put the electrodes on O1 and O2 in accordance with the International 10-20 System. The sampling rate was kept at 256 Hz.

As the car signals could provide useful information about the vehicle's behavior, a car equipped with sensors was used as the experimental car. The vehicle behavioral data including speed and deceleration with a sampling frequency of 50 Hz was analyzed in our present study. Fig. 1 shows the experiment scene.

C. Procedure

All subjects were given written and oral instructions on the driving experiment. To obtain the data of the distracted drivers, a "cellphone use" task was set as distracting factor. The distracting task could be described simply as: The drivers were asked to use WeChat for at least 3 seconds when they drove to half of the distance.

Each subject participated in two sessions amounted to six trials of the experiment. The first driving session was one normal driving trial (i.e., driving without distracting task) which lasted for at least 6 seconds. The second session included five distracted driving trials (i.e., performing the "cellphone use" task while driving), each trial lasted for around 20 seconds and the task began at about 12 seconds. There was a short break after each trial.

During the experiment, one experimenter was in the car together with the subject and gave hints for the start and end of the task. In the normal driving process, subjects were asked to drive down the road with full attention. However, they were supposed to drive normally at first in the distracted driving process, few seconds later the experimenter would send cellphone messages to them. After receiving messages, they had to check the messages for 3 seconds at least. In addition, another experimenter would throw a quadrate foam box to the road while each trial was going to end, and subjects were required to react to the obstacle as soon as possible. The EEG signal and the vehicle behavioral signals were recorded all the way from the car starting to stopping.

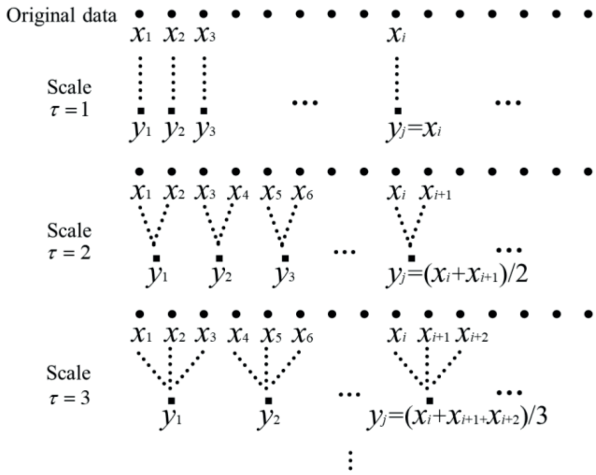


Fig. 2. Schematic illustration of the coarse graining process.

IV. METHODOLOGY

A. Analysis of EEG data

The process of EEG analysis contains three steps: preprocessing, artifacts removal and feature extraction.

1) Preprocessing

We first extracted the EEG segments corresponding to the duration of each trial in our study. Then, the alpha frequency band was obtained applying wavelet decomposition, as previous studies have demonstrated that the alpha frequency band is highly correlated with distraction [30], [60].

Wavelet transform is a time-frequency analysis method, which can reflect the local features of signals both in the time and frequency domain. And with the property of multi-resolution, it is widely used to analyze non-stationary signals [61]. A mother wavelet $\psi(t)$, in order to decompose the signal, is utilized in this method. The signal can be decomposed and expressed in terms of scaled and shifted versions of $\psi(t)$ and a corresponding scaling function $\phi(t)$ in discrete domain [62]. The discrete mother wavelet is represented as

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi\left(2^{-j}t-k\right), \qquad k, j \in Z \tag{1}$$

The signal $S(t)$ then can be expressed as

$$S(t) = \sum_k s_j(k)\phi_{j,k}(t) + \sum_k d_j(k)\psi_{j,k}(t) \tag{2}$$

where $s_j(k)$ and $d_j(k)$ are the approximate and detailed coefficients at level j.

In this paper, the EEG signal has been decomposed into 4 levels in which the detailed component at level 4 roughly represents the alpha band (8-13 Hz). Since db6 (Daubechies family) is similar to the EEG signal in our case as shown in Fig. 3, it is selected as the mother wavelet.

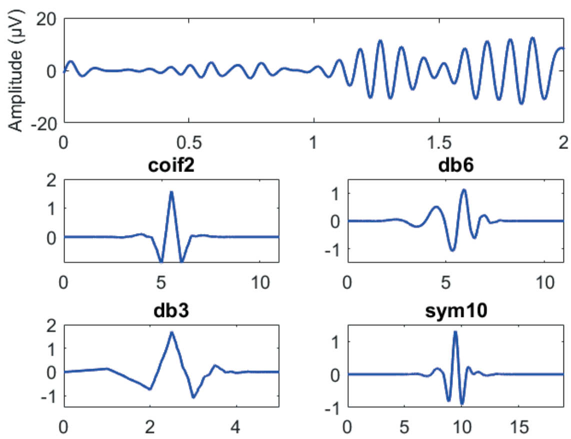


Fig. 3. Alpha wave and typical mother wavelets.

2) Artifacts removal

After preprocessing, the artifacts (e.g., the blinks) in the alpha band were then removed using a wavelet-based technique. The wavelet coefficients mentioned above represent the correlation between the signal and the selected mother wavelet. High amplitude coefficients will be generated at places where artifacts present. We can eliminate these kinds of coefficients utilizing a thresholding technique. It has been proven to be effective in the analysis of driver fatigue [63], [64]. The threshold can be defined as

$$T_j = mean(C_j) + 2 \times \mathrm{std}(C_j) \tag{3}$$

where C_j represents the wavelet coefficient at the jth level of wavelet decomposition. If the value of any coefficient is greater than the computed threshold, it is halved. Then the new set of wavelet coefficients are reconstructed to obtain the wavelet-corrected signal.

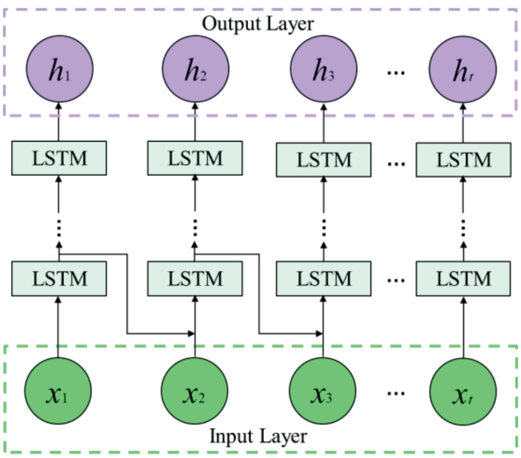


Fig. 4. The basic topological structure of LSTM.

3) Feature extraction

MSE extends the idea of Sample entropy (SE) to several time scales and is an effective method to quantify the complexity of a time series over different time scales. Time series with large fluctuation will produce a larger MSE value, which is considered to have high complexity. Similarly, a highly regular time series will generate lower entropy. This method was first proposed by Costa *et al.* in 2002 [65].

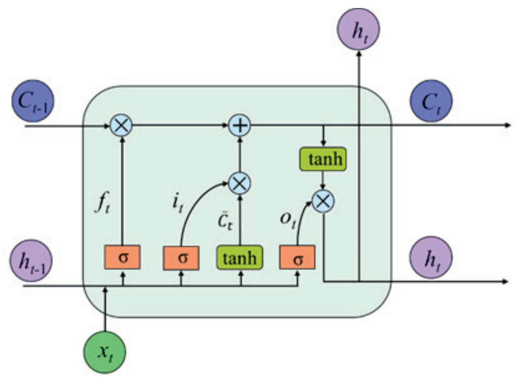


Fig. 5. The details of a LSTM cell.

There are two steps in MSE analysis: coarse graining and SE calculation. Considering the EEG signal $\{x_1, \ldots, x_i, \ldots, x_N\}$, we should construct a consecutive coarse-grained time series $\{y^{(\tau)}\}$, corresponding to the time scale factor τ: Firstly, the original EEG signal is divided into non-overlapping windows of length τ, then the data points inside each window are averaged (see Fig. 2). Each coarse-grained time series can be defined as

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \le j \le N/\tau \tag{4}$$

After the coarse graining procedure, SE is calculated for the obtained time series $\{y^{(\tau)}\}$. For a time series $\{y_1, \ldots, y_j, \ldots, y_n\}$, it can be made up into an m dimension vector $Y^m(i) = [y(i), y(i+1), \ldots, y(i+m-1)],\ 1 \le i \le n-m$. And d, the distance between $Y^m(i)$ and $Y^m(j)$ is defined as

$$\begin{gathered} d = \max[|\, y(i+k) - y(j+k) \,|], \\ 0 \le k \le m-1, i \ne j, 1 \le j \le n-m \end{gathered} \tag{5}$$

Then count the number of $d < r$ for each i, and $B_i^m(r)$ can be expressed as

$$B_i^m(r) = \frac{\{\text{the number of } d < r, i \ne j\}}{(n-m-1)} \tag{6}$$

where r is the given tolerable distance. The set of $B_i^m(r)$ are then averaged and the average value can be calculated by

$$B^m(r) = \frac{1}{n-m} \sum_{i=1}^{n-m} B_i^m(r) \tag{7}$$

Add the dimension by 1 to form an m+1 dimension vector and repeat the above process, then we can get $B^{m+1}(r)$. After all the procedures, the basic definition of SE is given by

$$\mathrm{SE}(m, r) = \lim_{n \to \infty} \left[-\ln \frac{B^{m+1}(r)}{B^m(r)} \right] \tag{8}$$

When n is finite, it can be calculated by the following expression:

$$\mathrm{SE}(m,r,n) = -\ln\left[\frac{B^{m+1}(r)}{B^{m}(r)}\right] \tag{9}$$

In Narayan's study, it has been revealed that MSE changes with time scale and there will be a peak indicating the existence of maximum entropy at that time scale, which indicates high correlation exists in time scale and MSE value [66]. In this paper, the maximum MSE appears when time scale is five, then 5-scale MSE with a sliding window is calculated for the extracted alpha frequency band.

B. *Statistical analysis of vehicle behavioral data*

The speed data and the deceleration data, corresponding to the duration of each trial, were further analyzed after the experiment. Statistical analysis was performed in MATLAB. To find out whether there were significant differences between the vehicle behavioral data before and after distraction, we firstly carried out significance tests on the vehicle behavioral data. Then the mean value and the standard deviation of the data were calculated to investigate the changes before and after distraction.

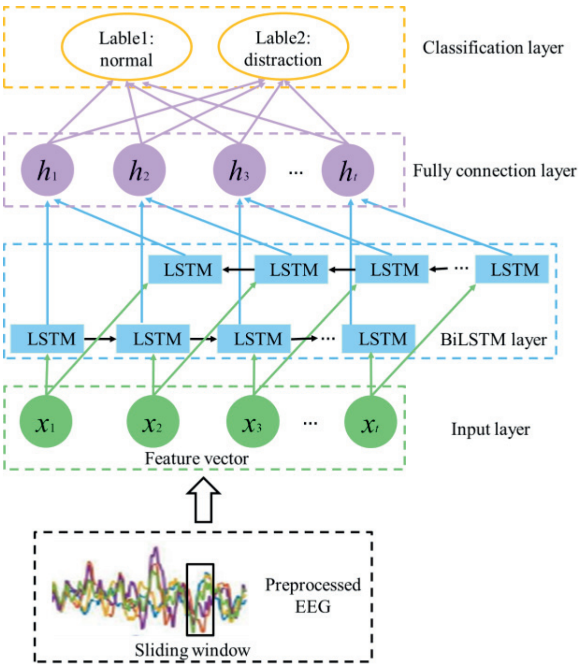


Fig. 6. Schematic illustration of the BiLSTM framework.

C. *BiLSTM*

LSTM is a special kind of RNN, capable of addressing the vanishing gradient problem. It was firstly introduced by Hochreiter *et al.* [53] in 1997. LSTM has two major features compared with RNN [67]. One feature is that it can learn both short and long-term dependencies (i.e., keep both short and long-term memory). The other is that it cannot only add useful information but also remove irrelevant details during the learning process. Fig. 4 is the basic topological structure of LSTM. It consists of a chain of repeating modules of neural networks. The repeating module of LSTM has four neural network layers (see Fig. 5) unlike the standard RNN having one, and they interact in a specific way.

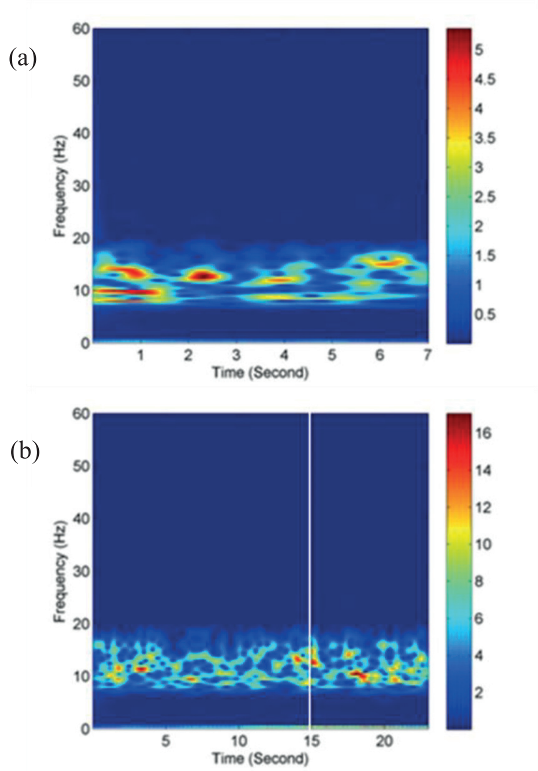


Fig. 7. The time-frequency results of alpha band. (a) Result of normal driving trial. (b) Result of distracted driving trial. The white solid line in (b) shows the onset of using cellphone.

A LSTM cell can add or remove information through structures called "gate". There are totally three types of gates in it: forget gate, input gate and output gate. They work as follows. At first, it is to decide what information should be removed from the cell state by forget gate (10). Then the input gate decides what new information is going to store in the cell state. This step can be divided into three parts. The first part is to use a sigmoid layer to find what is going to be updated using (11). Next is to create a new candidate cell state $\tilde{C}_t$ by a tanh layer (12). After that, the old cell state C_{t-1} can be updated into the new cell state Ct by (13). Finally, the output gate is activated to decide the output h_t of the cell by using (14) and (15).

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{10}$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{11}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{12}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{13}$$

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{14}$$

$$h_t = o_t * \tanh(C_t) \tag{15}$$

In these equations, σ and tanh are the active functions in the cell, W, h, and b represent the weight, hidden state, and bias separately. x_t indicates the EEG feature of time t in Figs. 4-6.

As the traditional the one directional LSTM usually learns the long-term information only from previous time steps to latter time steps, and research has found that the inputs of the latter time steps also contain some information about the inputs of the previous time steps [68]. BiLSTM, an update of LSTM, consists of two layers of LSTM. One layer processes the inputs in a forward direction, and the other learns information from the inputs in a backward direction. Additionally, it can also concatenate the two directions interpretations according to long-term dependency in the inputs. In this study, we use both LSTM and BiLSTM to learn the dependency among the extracted features and compare their performance for driver distraction recognition. The BiLSTM model structure diagram used here is shown in Fig. 6.

V. RESULTS

A. *Analysis of EEG data*

Time-frequency analysis of the extracted alpha frequency band was firstly performed. Fig. 7(a) shows the time-frequency graph of the normal driving trial and Fig. 7(b) is the result of a distracted driving trial. We also calculated the mean absolute amplitudes of alpha band, the results are shown in Fig. 8. During the normal driving process, the activity of alpha band showed a trend of decreasing, while it increased after using cellphone in the distracted driving process as shown in Fig. 7 and Fig. 8. Then the 5-scale MSE feature was calculated to extract the valuable information of the EEG signal. From the MSE result, we can see there is an obvious decrease after using cellphone. Fig. 9 gives the results of the MSE for both normal trial and distracted trial. Fig. 9(a) is the waveform of the normal driving trial, and the result of the distracted driving trial is shown in Fig. 9(b). It can be seen that the waveform in distracted trials fluctuated obviously. The MSE value began to decrease notably after the onset of distraction task and reached the minimum value a few seconds later after the task. Besides, the trough of MSE waveform was obviously lower than the average of MSE. However, there were small and gentle fluctuations in normal trials as shown in Fig. 9(a). The time that MSE reaches its minimum value is defined as the EEG most distraction position (DP) of the subject pointed out in Fig. 9(b). According to the MSE results, the EEG most distraction positions of all subjects could be obtained. The time difference between DP and the onset of using cellphone was then calculated and listed in Table I. Trial 1, which is the normal driving process, is excluded from the table.

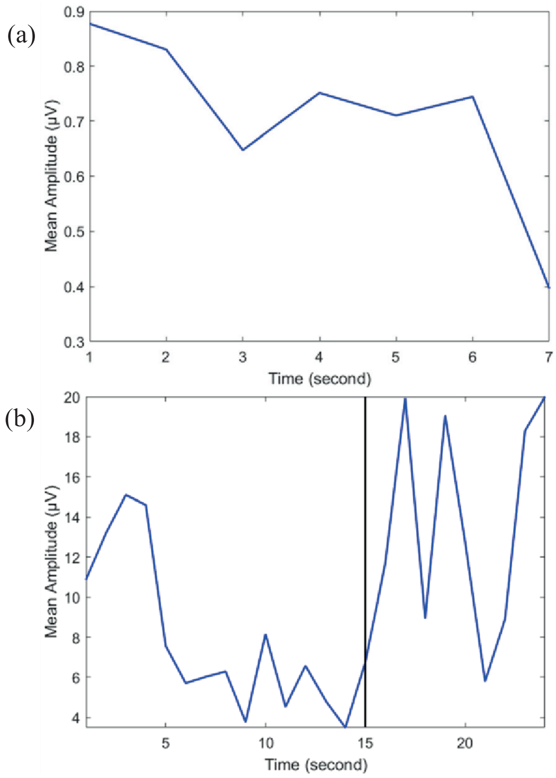


Fig. 8. The mean absolute amplitudes of alpha band. (a) Result of normal driving trial. (b) Result of distracted driving trial. The black solid line in (b) shows the onset of using cellphone.

B. *Statistical analysis of vehicle behavioral data*

The statistical analysis of the obtained vehicle behavioral data (i.e., speed and deceleration data) was performed to validate abnormal changes also appear in driving performance before and after distraction. This section consists of two parts: the first part is to analyze data that before and after the subjects start to use cellphone, the other is before and after DP of the subjects.

1) *Analysis of the vehicle data before and after using cellphone*

In this part, the speed and deceleration data before and after using cellphone was analyzed to investigate the impact of distraction task on the driving performance.

TABLE I
THE TIME DIFFERENCE BETWEEN DP AND THE ONSET OF USING CELLPHONE OF ALL SUBJECTS (S)

Subject \ Trial	2	3	4	5	6
1	1	0	0	1	2
2	4	4	1	5	8
3	1	2	4	8	0
4	1	5	5	7	1
5	5	5	7	5	5
6	2	0	3	7	2

At the beginning of the analysis, we performed unpaired t-test on the vehicle data in each trial to verify whether significant differences exist between the data before and after using cellphone. The significance level is set as 0.05. The t-test results are shown in Table II and Table III. Trial 2 to Trial 6 are distracted driving processes that drivers were asked to use cellphone while driving. h indicates if there are significant differences between the vehicle data before and after distraction. h = 1 means significant differences exist. h = 0 means no significant differences. p represents the probability that the data before and after distraction is distributed identically.

From the results of the speed data, we can see clearly that there are significant differences between the speed before and after using cellphone except for Trial 3. All trials show significant differences between the two conditions in deceleration data shown in Table III.

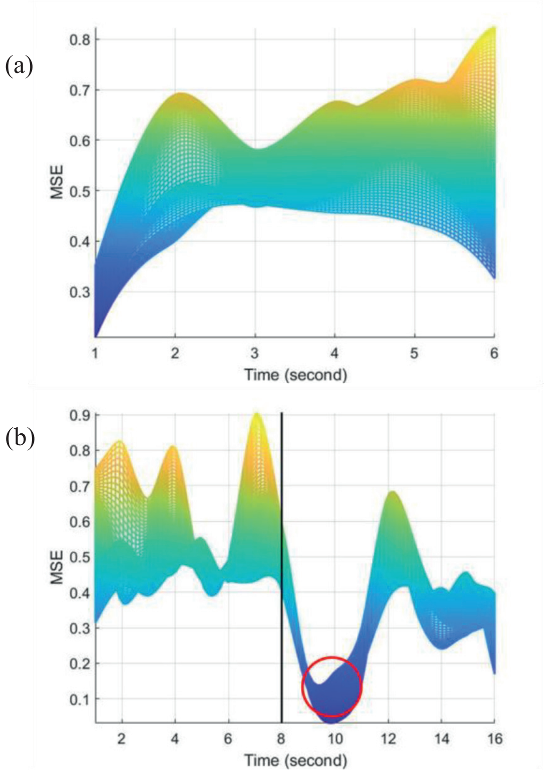


Fig. 9. The MSE results of alpha band. (a) Result of normal driving trial. (b) Result of distracted driving trial. The black solid line shows the onset of using cellphone and the red circle is the distraction position of this trial in (b).

TABLE II
THE *T*-TEST RESULTS OF THE SPEED DATA

Trial	2	3	4	5	6
h	1	0	1	1	1
p	< 0.05	0.935	< 0.05	< 0.05	< 0.05

TABLE III
THE *T*-TEST RESULTS OF THE DECELERATION DATA

Trial	2	3	4	5	6
h	1	1	1	1	1
p	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05

The statistical features (i.e., mean value and standard deviation) of speed and deceleration before and after using cellphone are calculated and listed in Table IV - Table VII separately. "Before" and "After" represent before and after distraction, respectively.

TABLE IV
THE MEAN VALUE OF SPEED (km/h)

Trial	2	3	4	5	6
Before	5.877	5.397	5.145	5.410	5.496
After	5.406	5.331	4.903	4.775	5.094

TABLE V
THE MEAN VALUE OF DECELERATION (m/s^2)

Trial	2	3	4	5	6
Before	-0.131	-0.121	-0.126	-0.134	-0.114
After	0.234	0.246	0.277	0.229	0.232

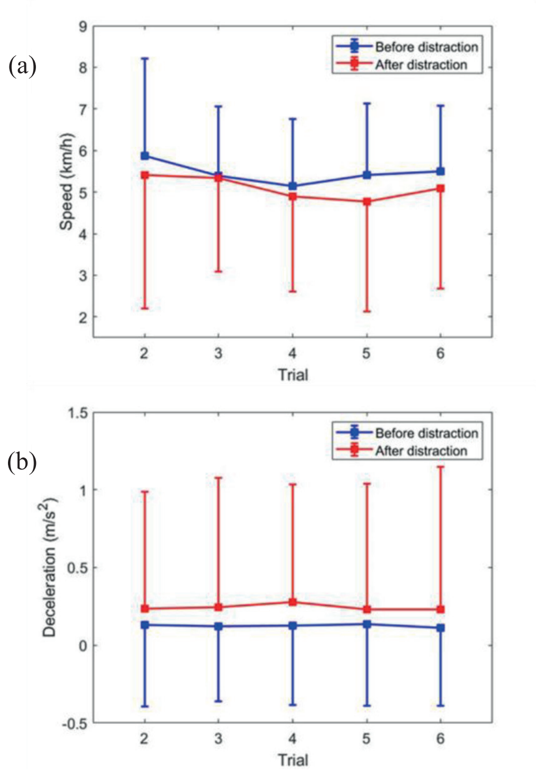


Fig. 10. Statistical results of the vehicle data. (a) The mean value of speed. (b) The absolute value of mean deceleration. Error bar shows the standard deviation.

All of the mean values of speed before using cellphone are greater than those after distraction in Table IV and Trial 3 has the smallest gap between the two conditions. Table V shows that the mean values of deceleration before using cellphone are all negative and that the mean values become positive after distraction. When comparing the absolute values of the deceleration mean values, it is obvious that the absolute values after using cellphone are greater than that of before distraction.

TABLE VI
THE STANDARD DEVIATION OF SPEED

Trial	2	3	4	5	6
Before	2.343	1.657	1.613	1.720	1.583
After	3.199	2.240	2.294	2.641	2.411

The standard deviations of the speed and deceleration data reflect the same trend that all of them becomes greater after using cellphone shown in Table VI and Table VII. The changing patterns of the speed and deceleration data could also be clearly shown in the following error bar figures (see Fig. 10) according to the statistical results.

TABLE VII
THE STANDARD DEVIATION OF DECELERATION

Trial	2	3	4	5	6
Before	0.523	0.484	0.508	0.523	0.502
After	0.752	0.832	0.760	0.810	0.915

2) Analysis of the vehicle data before and after the most distraction position

The vehicle behavioral data before and after the EEG most distraction position (DP) was also analyzed. The analyzing procedure in this part was similar to 1). To explore whether significant differences exist between the vehicle data before and after DP, unpaired *t*-test was firstly performed with the assumption that all subjects are considered as a whole.

The test results are shown in Table VIII and Table IX. It is clear that significant differences do exist in all trials for both speed and deceleration before and after DP.

TABLE VIII
THE *T*-TEST RESULTS OF THE SPEED DATA

Trial	2	3	4	5	6
h	1	1	1	1	1
p	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05

TABLE IX
THE *T*-TEST RESULTS OF THE DECELERATION DATA

Trial	2	3	4	5	6
h	1	1	1	1	1
p	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05

Then the mean value and the standard deviation of the data before and after DP were calculated separately. Table X to Table XIII give the results. We can see from Table X that all of the mean values of speed before DP are obviously greater than that of after DP. Table XI shows that the mean values of deceleration before DP are negative and that after distraction the mean values are positive. Besides, the absolute values of the mean deceleration after DP are greater than before DP. As for the standard deviation, it becomes greater after DP showed in Table XII and Table XIII.

TABLE X
THE MEAN VALUE OF SPEED (km/h)

Trial	2	3	4	5	6
Before	5.795	5.394	5.175	5.495	5.584
After	5.440	4.987	3.938	2.615	4.562

TABLE XI
THE MEAN VALUE OF DECELERATION (m/s^2)

Trial	2	3	4	5	6
Before	-0.068	-0.061	-0.070	-0.045	-0.098
After	0.204	0.249	0.447	0.413	0.353

The error bar figures are drawn as Fig. 11 according to the statistical results. The changing rules mentioned above could be easily seen from the figures. Compared with Fig. 10, the statistical differences of the vehicle behavioral data between before distraction and after distraction increase in Fig. 11.

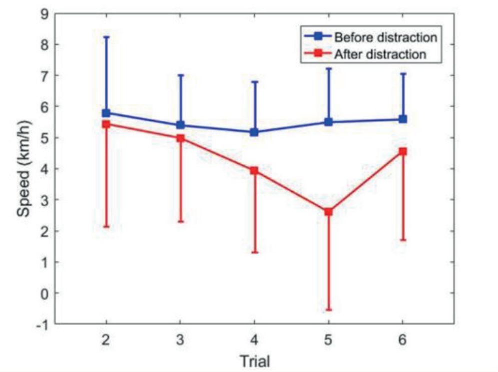
TABLE XII
THE STANDARD DEVIATION OF SPEED

Trial	2	3	4	5	6
Before	2.454	1.631	1.634	1.725	1.475
After	3.286	2.679	2.630	3.134	2.843

TABLE XIII
THE STANDARD DEVIATION OF DECELERATION

Trial	2	3	4	5	6
Before	0.615	0.591	0.534	0.596	0.493
After	0.671	0.780	0.912	1.009	1.067

(a)



(b)

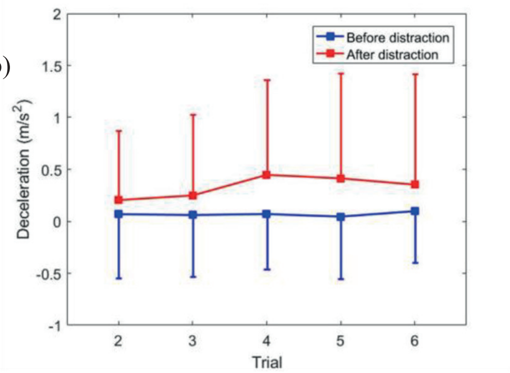


Fig. 11. The statistical results of the vehicle data. (a) The mean value of speed. (b) The absolute value of the deceleration mean value. Error bar shows the standard deviation.

C. *The classification results of BiLSTM*

In this paper, we not only analyzed the dynamic brain activity changes of distracted driving based on MSE and the changes in driving performance but also detected whether a driver is distracted or not using BiLSTM. The detection results were compared with four different types of traditional classifiers, i.e., LSTM, SVM, convolutional neural network (CNN) and *k*-nearest neighbor (kNN). In addition, to be more reliable and convincing, the results were also compared with the results of traditional vehicle behavioral features and another four entropy-based algorithms i.e., Approximate entropy (AE), Differential entropy (DE), Fuzzy entropy (FE) and Sample entropy (SE) of

EEG.

We built and trained a BiLSTM model, which adopted the calculated feature matrixes as inputs and output the category vectors. There are 2 categories in the paper: distraction and non-distraction. The data of five subjects were used for training and that of the remained subject was utilized for testing. In the training process, different numbers of LSTM layers and training iterations were tried to find the best model in the BiLSTM classifier. The classification results of different features are shown in Table XIV. "VS" means statistical features (i.e., mean and standard deviation) of speed and deceleration.

TABLE XIV
THE MEAN ACCURACIES OF DIFFERENT CLASSIFIERS FOR DIFFERENT FEATURES (%)

Feature	AE	DE	FE	SE	MSE	VS	VS+MSE
BiLSTM	83.29	82.67	76.35	67.01	91.83	89.85	92.48
LSTM	82.24	81.31	71.03	63.55	89.72	88.79	91.59
CNN	62.62	73.63	62.01	60.32	73.90	67.29	78.5
SVM	52.94	54.6	54.27	56.75	66.72	74.85	77.76
kNN	69.45	71.93	67.12	59.05	65.34	76.84	77.07

Table XIV shows that the performance of BiLSTM and LSTM are much better than those of the other three common classifiers and that the BiLSTM model is slightly better than the conventional LSTM modal. As for the results of different features using BiLSTM and LSTM, the mean accuracy of the LSTM and BiLSTM using MSE of EEG reaches 89.72% and 91.83%, respectively, which is clearly higher than the results of using vehicle statistical features and other entropy-based methods. The SE feature of EEG leads to the lowest classification accuracy of 63.55% in LSTM and 67.01% in BiLSTM, and the accuracy of the other algorithms lies between the accuracy of SE and MSE. When inputting the features of EEG and vehicle data at the same time, the performance of BiLSTM increases, peaking at 92.48%. The accuracy of LSTM also improved under this condition. In sum, when we just use the EEG features to train the BiLSTM model, the best result is obtained by the MSE feature. Besides, MSE feature of EEG also performs better compared with the result of conventional vehicle behavioral features in BiLSTM, and the mean accuracy is about 3% higher than that of VS when we add MSE feature to the vehicle behavioral features.

VI. DISCUSSION

The concern for driver distraction is growing in recent years with the development of advanced infotainment systems. There are many effects and characteristics of driver distraction [9]-[12]. The distraction information of the dynamic brain activity and the changing rules of vehicle behavioral data before and after distraction are here discussed. A driver distraction classification model using BiLSTM is proposed based on the MSE feature, and the results are compared with four kinds of traditional classifiers as well as other conventional feature extraction methods.

As shown in Fig. 7 and Fig. 8, the activity of the alpha frequency band is related to driver distraction, which increases after being distracted. However, it shows a trend of decreasing with the process of the normal driving trial. The results are consistent with previous studies that the activity of alpha rhythm increases in parietal-occipital brain regions if attentional lapses occur [69], [70]. The important changes in MSE feature of EEG after being distracted can be seen in Fig. 9. The EEG complexity is clearly illustrated by the fluctuation of MSE feature. The MSE value decreases sharply when drivers start to use cellphone (see Fig. 9(b)) compared with normal driving, which indicates that the complexity of the alpha frequency band decreases while distraction. Drivers have to keep high alertness to pay attention to the surroundings like the pedestrians and other cars so that they can drive safely in the normal driving process [71]. In this situation, the brain activity is usually active, and it embodies the relatively high complexity of the alpha frequency band. Contrary to normal driving, drivers' perceptions of driving and the surroundings decrease while using cellphone and then the complexity also decreases, thus leading to the decreased MSE value while distraction. The time difference for each trial in Table I means that it usually takes drivers a few seconds to shift their attention to the task related work. Hence, DP occurs a few seconds later after drivers start to use cellphone.

Statistical analysis of the vehicle behavioral data before and after using cellphone is then performed considering all subjects as a whole. From the *t*-test results [Table II and Table III] of the speed and deceleration data, we confirm that the performance of drivers to control cars is highly affected by the "cellphone use" task, which has been validated in previous studies [72]-[74].

For the statistical analysis results, drivers tend to drive at a lower and much safer speed after beginning to use cellphone shown in Table IV and Fig. 10(a). Many studies have proved that drivers attempt to reduce their workload by decreasing speed while distracted [75], [76], which explained why the mean speed is lower after distraction than that before distraction. Besides, Trial 3 shows the smallest gap between before and after using cellphone. It is because that the changes between the two conditions in this trial are not obvious as listed in Table II. As shown in Table V, the mean deceleration is negative before using cellphone due to the stepwise accelerating stage in this process. However, it becomes positive after using cellphone. A possible reason for the phenomenon is that drivers tend to decrease speed for safety while distracted. What's more, the absolute value of the mean deceleration before using cellphone is apparently lower than after using cellphone in Fig. 10(b), which indicates that distracted drivers often make emergency brakes when obstacles appear. When drivers begin to use cellphone, they are distracted by the task and their abilities to monitor the environment may be reduced, the decision to brake would then be consequently delayed. As a result, drivers will have to brake harder to avoid accidents. This explanation is in accordance with the results of Hancock *et al.* [77]. Their work reported that distracted drivers responded slowly to the traffic lights and had to take stronger braking actions to compensate for the delay in starting braking.

In line with previous driver distraction analysis, the variability of the speed and deceleration data increases after using cellphone in Table VI and Table VII. The change can also be seen from the error bar in Fig. 10. The previous study reported that variability in velocity increased while drivers performing auditory tasks as attention need to be shifted to the task processing streams from focusing on driving leading to the performance decrements [78]. In our study, drivers pay more attention to the task and the brake pedal controlling ability is then weakened. Hence, greater variability occurs in speed and deceleration, which explains why greater standard deviations appear.

In this paper, vehicle behavioral data analysis is discussed not only before and after using cellphone, but also before and after DP. Identical to the analyzing process before and after using cellphone, the speed and deceleration data are analyzed firstly considering all subjects as a whole. Results in Table VIII and Table IX imply that significant changes emerge in driving performance between conditions before and after DP as a consequence of distraction, which is also consistent with previous studies in [72]-[74]. Furthermore, the mean value and standard deviation of speed and deceleration are analyzed to find out how drivers are affected by the "cellphone use" task. As shown in Table X and Fig. 11(a), the mean speed in each trial after DP is visibly greater than that before DP. The results are agreed with the work of Reimer [75] and Mehler [76], pointing out that distracted drivers usually try to decrease speed to reduce workload and keep safe. As for the results of mean deceleration shown in Table XI, the same inference with Table V can be made. the mean deceleration is negative in the accelerating stage before DP, which becomes positive after DP when obstacles abruptly appear in the process. Moreover, compared to Table V, the difference value of the mean deceleration before and after distraction in Table XI is greater, which indicates the performance of controlling the brake pedal after DP is even weaker than after beginning to use cellphone. The absolute value of the mean deceleration is also compared in Fig. 11(b). Note that the ability to monitor the surroundings after DP may be reduced and then the decision on when to brake is delayed. Therefore, drivers have to make harder brake to avoid obstacles [77]. In addition, the variability of speed and deceleration also increases after DP in Table XII and Table XIII, which shows the same rules as the same as in Table VI and Table VII. Previous studies have validated that drivers will shift their attention to the task after distraction [78], thus the ability to handle the brake pedal is weakened. As a result, augmentation variability appears in speed and deceleration.

After mining the valuable information of driver distraction based on MSE feature of EEG and analyzing the changes in driving performance, we finally use BiLSTM to show that driver distraction can be detected with the MSE features. The classification results in Table XIV indicate that the classification accuracy of MSE using BiLSTM is better than traditional vehicle behavioral features and other entropy-based features since it can not only present the complex distraction information of EEG but also reduce the influence of the residual noise on the results [40]. The classification accuracy of MSE is comparable with the research of Li *et al.* [79], which used the temporal and spatial features of the 32-channel EEG signals and reached an accuracy of 92%. Besides, Xie *et al.* [80] also collected six kinds of vehicle signals and smartphone sensor signals to detect driver distraction. The accuracy of VS using BiLSTM in our work is 3% higher than their accuracy obtained from traditional classifiers. The performance of the trained BiLSTM model is further improved with an accuracy of 92.48% when adding MSE features to the statistical features of vehicle behavioral data, which suggests that MSE features could remedy the inadequacy of traditional vehicle behavioral features. It is consistent with the observation in the literature that hybrid signals can provide more sufficient information about driver distraction than one type of signal alone [28]. In addition, the performance of BiLSTM is compared to four conventional classifiers in the study. The results in Table XIV show that BiLSTM, which could learn the bidirectional long-term dependency among the extracted features, is slightly better than traditional one directional LSTM and significantly better than CNN, SVM and kNN. It corresponds to the results in [67] that BiLSTM can decrease the model's train and test error and thus improve the classification accuracy. The reliable results of the study suggest the potential to mine the distraction information in realistic driving environment and to detect driver distraction using MSE and BiLSTM.

The limitation of the study is that only six persons participated in the experiment, so the dataset is a little bit small to some degree. It is difficult to collect the data with driver distraction in realistic driving scenarios. The sample size, while acceptable for distraction detection, had limited statistical power.

VII. CONCLUSION

In this paper, we have applied the BiLSTM model to present a driver distraction detection framework based on the complexity-based MSE feature of EEG. It demonstrates that it is better to use MSE to explore the complex dynamic distraction information of EEG than other features used in the previous studies. Besides, compared to conventional vehicle behavioral features, the model performance is enhanced by adding features of EEG to features of vehicle data. It confirms that the MSE feature can provide complementary information about distracted drivers. For a driver, the MSE value of EEG decreases obviously in the distraction process and the ability to manipulate the vehicle is also greatly influenced, which is manifested in the decreased speed, harder brakes as well as the increased variability of speed and deceleration.

In the future work, an experiment will be designed in driving simulator involving more participants and more types of signals to study driver distraction applying the proposed method. The new dataset containing multi-modality signals provides better opportunities for further investigating the effectiveness of different kinds of signals in detecting driver distraction. What's more, an improvement of the present algorithm will also be explored to detect driver distraction accurately. Another particular interest is to study the influence of the left- and right-handed in the detection performance in the future.

ACKNOWLEDGMENT

We gratefully acknowledge the financial support from the National Natural Science Foundation of China (grant number: 61703069 and 62001312) and the Fundamental Research Funds for the Central Universities (grant number: DUT21GF301).

REFERENCES

[1] "Global status report on road safety 2018: summary," World Health Organization, Geneva, Switzerland, 2018. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/277370/WHO-NMH-NVI-18.20-eng.pdf

[2] "Distracted driving 2013," National Highway Traffic Safety Administration, Washington, USA. Apr. 2015. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812132

[3] M. A. Regan, J. D. Lee, and K. L. Young, *Driver Distraction: Theory, Effects and Mitigation*, CRC Press, 2008. [Online]. Available: https://doi.org/10.1201/9781420007497

[4] "Research on distracted driving in China," Ford Motor Company, and Tsinghua University, Beijing 2017.

[5] S. M. Simmons, A. Hicks, and J. K. Caird, "Safety-critical event risk associated with cell phone tasks as measured in naturalistic driving studies: A systematic review and meta-analysis," *Accid. Anal. Prev.,* vol. 87, pp. 161-169, 2016.

[6] J. Atwood, F. Guo, G. Fitch, and T. A. Dingus, "The driver-level crash risk associated with daily cellphone use and cellphone use while driving," *Accid. Anal. Prev.,* vol. 119, pp. 149-154, 2018.

[7] A. Stelling-Konczak, G. P. van Wee, J. J. F. Commandeur, and M. Hagenzieker, "Mobile phone conversations, listening to music and quiet (electric) cars: Are traffic sounds important for safe cycling?" *Accid. Anal. Prev.,* vol. 106, pp. 10-22, Sep. 2017.

[8] C. Samsa, "Digital billboards 'down under'. Are they distracting to drivers and can industry and regulators work together for a successful road safety outcome?" presented at *the 2015 Australasian Road Safety Conf.*, Gold Coast, Australia, Oct. 14-16, 2015.

[9] D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 32, no. 7, pp. 1239-1258, Jul. 2010.

[10] J. D. Lee, "Dynamics of driver distraction: The process of engaging and disengaging," *Ann Adv Automot Med,* vol. 58, pp. 24-32, Mar. 2014.

[11] A. Gaffar and S. M. Kouchak, "Quantitative driving safety assessment using interaction design benchmarking," in *Proc. IEEE SmartWorld Ubiquitous Intell. Comput. Adv. Trusted Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, San Francisco, CA, USA, 2017, pp. 1-8.

[12] N. Lavie, "Attention, distraction, and cognitive control under load," *Curr. Dir. Psychol.,* vol. 19, no. 3, pp. 143-148, 2010.

[13] M. Wollmer *et al.*, "Online driver distraction detection using long short-term memory," *IEEE Trans. Intell. Transp. Syst.,* vol. 12, no. 2, pp. 574-582, Jun. 2011.

[14] A. S. Lea, T. Suzukic, and A. Hirofumi, "Evaluating driver cognitive distraction by eye tracking: From simulator to driving," *Transp. Research Interdiscip. Persp.*, vol. 9, pp. 100307, 2021.

[15] E. Ohn-Bar, S. Martin, A. Tawari, and M. M. Trivedi, "Head, eye, and hand patterns for driver activity recognition," in *Proc. ICPR, Stockholm*, Sweden, Aug. 2014, pp. 660–665.

[16] I. Jegham, A. Ben Khalifa, I. Alouani, and M. A. Mahjoub, "A novel public dataset for multimodal multiview and multispectral driver distraction analysis: 3MDAD," *Signal Process.: Image Commun.,* vol. 88, pp. 115960, Aug. 2020.

[17] H. M. Eraqi, Y. Abouelnaga, M. H. Saad, and M. N. Moustafa, "Driver distraction identification with an ensemble of convolutional neural networks," *J. Adv. Transp.,* vol. 2019, pp. 1-12, Feb. 2019.

[18] J. H. L. Hansen, C. Busso, Y. Zheng, and A. Sathyanarayana, "Driver modeling for detection and assessment of driver distraction: Examples from the UTDrive test bed," *IEEE Signal Process. Mag.,* vol. 34, no. 4, pp. 130-142, Jul. 2017. doi:10.1109/MSP.2017.2699039.

[19] S. Yadawadkar *et al.*, "Identifying distracted and drowsy drivers using naturalistic driving data," in *IEEE Int. Conf. Big Data*, Seattle, WA, USA, 2018, pp. 2019-2026.

[20] J. Sun, Y. Zhang, and J. Wang, "Detecting driver distraction behavior with naturalistic driving data," *China J. High. Transp.,* vol. 9, no. 33, pp. 11-26, May. 2020.

[21] Y. Zhao, T. Li, and Y. Dong, "A wearable acoustic sensor based driver distraction behaviour recognition," in 2021 *Int. Conf. High Perf. Big Data Intell. Syst.*, Macau, China, 2021, pp. 281-285.

[22] R. Wang, L. Huang, and C. Wang, "Distracted driving detection by sensing the hand gripping of the phone," in *Proc. Annual Int. Conf. Mobile Comput. Net.*, New Orleans, LA, USA, 2021: 828-830.

[23] M. K. Wali, M. Murugappan, and B. Ahmad, "Subtractive fuzzy classifier based driver drowsiness levels classification using EEG," *J. Phys. Ther. Sci.*, vol. 25, no. 9, pp. 1055-1058, 2013.

[24] M. Taherisadr, P. Asnani, S. Galster, and O. Dehzangi, "ECG-based driver inattention identification during naturalistic driving using Mel-frequency cepstrum 2-D transform and convolutional neural networks," *Smart Health*, vol. 9-10, pp. 50-61, 2018.

[25] C. Alessandro, "Automatic detection of saccadic eye movements using EOG for analysing effects of cognitive distraction during driving," M.S. thesis, Dept. Applied. Mechanics., Chalmers Univ. Tech., Goteborg, Sweden, 2017.

[26] D. He, B. Donmez, C. C. Liu, and K. N. Plataniotis, "High cognitive load assessment in drivers through wireless electroencephalography and the validation of a modified N-back task," *IEEE Trans. Hum.-Mach. Syst.,* vol. 49, no. 4, pp. 362-371, Aug. 2019.

[27] N. Li, J. J. Jain, and C. Busso, "Modeling of driver behavior in real world scenarios using multiple noninvasive sensors," *IEEE Trans. Multimedia,* vol. 15, no. 5, pp. 1213-1225, Aug. 2013.

[28] Y. Zhang, Y. Chen, and C. Gao, "Deep unsupervised multi-modal fusion network for detecting driver distraction," *Neurocomputing,* vol. 421, pp. 26-38, 2021.

[29] G. Lechner *et al.*, "A lightweight framework for multi-device integration and multi-sensor fusion to explore driver distraction," in *Adv. Inf. Syst. Eng.*, 2019, pp. 80-95.

[30] H. Almahasneh, W. Chooi, N. Kamel, and A. S. Malik, "Deep in thought while driving: An EEG study on drivers' cognitive distraction," *Transp. Res. Pt. F-Traffic Psychol. Behav.,* vol. 26, pp. 218-226, 2014.

[31] C. Fan *et al.*, "Detection of train driver fatigue and distraction based on forehead EEG: a time-series ensemble learning method," *IEEE Trans. Intell. Transp. Syst.,* pp. 1-11, Nov. 2021.

[32] L. Yang, W. Guan, R. Ma, and X. Li, "Comparison among driving state prediction models for car-following condition based on EEG and driving features," *Accid. Anal. Prev.,* vol. 133, pp. 105296, 2019.

[33] S. Barua, M. U. Ahmed, and S. Begum, "Classifying drivers' cognitive load using EEG signals," *pHealth*, pp. 99-106, May. 2017.

[34] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehav. Rev.,* vol. 44, pp. 58-75, 2014.

[35] Z. Gao *et al.*, "An adaptive optimal-Kernel time-frequency representation-based complex network method for characterizing fatigued behavior using the SSVEP-based BCI system," *Knowledge-Based Syst.,* vol. 152, pp. 163-171, 2018.

[36] Z. Gao *et al.*, "Relative wavelet entropy complex network for improving EEG-based fatigue driving classification," *IEEE Trans. Instrum. Meas.,* vol. 68, no. 7, pp. 2491-2497, Jul. 2019.

[37] W. Zheng, J. Zhu, Y. Peng, and B. Lu, "EEG-based emotion classification using deep belief networks," in *IEEE Int. Conf. Multimed. Expo*, Chengdu, China, 2014, pp. 1-6.

[38] Q. Tang, "Research of automatic sleep staging based on EEG signals," M.S. thesis, Dept. Autom., Guangdong Univ. Technol., Guangzhou, China, 2016.

[39] A. U. Jose, and B. Garcia-Zapirain, "EEG artifact removal−state-of-the-art and guidelines," *J. Neural Eng.,* vol. 12, no. 3, pp. 031001, Apr. 2015.

[40] M. Costa, A. L. Goldberger, and C. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E,* vol. 71, no. 2 Pt 1, pp. 021906, Feb. 2005.

[41] H. Azami, D. Abásolo, S. Simons, and J. Escudero, "Univariate and multivariate generalized multiscale entropy to characterise EEG signals in Alzheimer's disease," *Entropy*. vol. 19, no. 1, pp. 31, 2017.

[42] H. Luo, T. Qiu, C. Liu, and P. Huang, "Research on fatigue driving detection using forehead EEG based on adaptive multi-scale entropy," *Biomed. Signal Process. Control,* vol. 51, pp. 50-58, 2019.

[43] Y. Liao *et al.*, "Detection of driver cognitive distraction: A comparison

study of stop-controlled intersection and speed-limited highway," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1628-1637, Jun. 2016.
[44] E. Q. Wu *et al.*, "Inferring flight performance under different maneuvers with pilot's multi-physiological parameters," *IEEE Trans. Intell. Transp. Syst.*, pp. 1-11, 2021.
[45] K. Torkkola, N. Massey, and C. Wood, "Detecting driver inattention in the absence of driver monitoring sensors," in *2004 Int. Conf. Mach. Learn. Appl.*, Louisville, KY, USA, 2004, pp. 220-226.
[46] E. Q. Wu *et al.*, "Detecting fatigue status of pilots based on deep Learning network using EEG signals," *IEEE Trans. Cognitive Develop. Syst.*, vol. 13, no. 3, pp. 575-585, Sept. 2021.
[47] E. Q. Wu *et al.*, "Nonparametric Bayesian prior inducing deep network for automatic detection of cognitive status," *IEEE Trans. Cyber.*, vol. 51, no. 11, pp. 5483-5496, Nov. 2021.
[48] Y. Liang, J. D. Lee, and M. L. Reyes, "Nonintrusive detection of driver cognitive distraction in real time using Bayesian networks," *Transp. Res. Record,* vol. 2018, no. 1, pp. 1-8, 2007.
[49] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," in *S.C. Kremer, J.F. Kolen (Eds.), A field guide to dynamical recurrent neural networks*, IEEE Press, 2001.
[50] A. Girma, X. Yan, and A. Homaifar, "Driver identification based on vehicle telematics data using LSTM-recurrent neural network," in *ICTAI*, Portland, OR, USA, 2019, pp. 894-902.
[51] U. B. Baloglu, and Ö. Yildirim, " Convolutional long-short term memory networks model for long duration EEG signal classification," *J. Mech. Med. Biol.,* vol. 19, no. 1, pp. 1940005, 2019.
[52] K. Saleh, M. Hossny, and S. Nahavandi, "Driving behavior classification based on sensor data fusion using LSTM recurrent neural networks," in *IEEE 20th ITSC*, Yokohama, Japan, 2017, pp. 1-6.
[53] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural Comput.,* vol. 9, no. 8, pp. 1735-1780, Dec. 1997.
[54] S. M. Kouchak, and A. Gaffar, "Detecting driver behavior using stacked long short term memory network with attention layer," *IEEE Trans. Intell. Transp. Syst.,* vol. 22, no. 6, pp. 3420-3429, Jun. 2021.
[55] T. Mahmud *et al.*, "Sleep apnea setection from variational mode decomposed EEG signal using a hybrid CNN-BiLSTM," *IEEE Access,* vol. 9, pp. 102355-102367, 2021.
[56] G. Liu, and J. Guo. "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing,* vol. 337, pp. 325-338, Apr. 2019.
[57] Q. Sun, C. Wang, Y. Guo, W. Yuan, and R. Fu, "Research on a cognitive distraction recognition model for intelligent driving systems based on real vehicle experiments," *Sensors*. vol. 20, no. 16, pp. 4426, 2020.
[58] S. P. Kumar, J. Selvaraj, R. Krishnakumar, and A. Sahayadhas, "Detecting distraction in drivers using Electroencephalogram (EEG) Signals," in *Proc. Int. Conf. Comput. Meth. Commun. (ICCMC)*, Erode, India, 2020, pp. 635-639.
[59] S. P. Kumar, S. Murugan, J. Selvaraj, and A. Sahayadhas, "Detecting driver mental fatigue based on Electroencephalogram (EEG) signals during simulated driving," *IOP Conf. Ser.: Mater. Sci. Eng.,* vol. 1070, pp. 012096, 2021.
[60] J. Murphy *et al.*, " Proactive control of emotional distraction: evidence from EEG alpha suppression," *Front. Hum. Neurosci.,* vol. 14, pp. 318, 2020.
[61] N. Gurudath, and H. B. Riley, "Drowsy driving detection by EEG analysis using wavelet transform and K-means clustering," *Procedia Comput. Sci.,* vol. 34, pp. 400-409, 2014.
[62] O. Rioul and M. Vetierli, "Wavelets and signal processing," *IEEE Signal Proc. Mag.,* vol. 8, no. 4, pp. 14-38, Oct. 1991.
[63] C. Zhang *et al.*, "Network entropy for the sequence analysis of functional connectivity graphs of the brain," *Entropy,* vol. 20, no. 5, pp. 311, Apr. 2018.
[64] C. Zhang, L. Sun, F. Cong, and T. Ristaniemi, "Spatio-temporal dynamical analysis of brain activity during mental fatigue process," *IEEE Trans. Cogn. Dev. Syst.,* pp. 1-1, Feb. 2020.
[65] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.,* vol. 89, no. 6, pp. 068102, Aug. 2002.
[66] P. S. Narayan, "Understanding multiscale entropy," Sapien Labs, Washington, USA, 2018. [Online]. Available: https://sapienlabs.org/understanding-multiscale-entropy/
[67] S. M. Kouchak and A. Gaffar, "Using bidirectional long short term memory with attention layer to estimate driver behavior," in *18th IEEE ICMLA*, Boca Raton, FL, USA, 2019, pp. 315-320.
[68] A. Fares, S. Zhong, and J. Jiang, "Region level bi-directional deep learning framework for EEG-based image classification," in *IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Madrid, Spain, 2018, pp. 368-373.
[69] P. R. Davidson, R. D. Jones, and M. T. R. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Trans. Biomed. Eng.,* vol. 54, no. 5, pp. 832-839, May. 2007.
[70] M. T. R. Peiris, P. R. Davidson, P. J. Bones, and R. D. Jones, "Detection of lapses in responsiveness from the EEG," *J. Neural Eng.,* vol. 8, no. 1, pp. 016003, Feb. 2011.
[71] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.,* vol. 12, no. 2, pp. 596-614, Jun. 2011.
[72] D. Lamble, T. Kauranen, M. Laakso, and H. Summala, "Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving," *Accid. Anal. Prev.,* vol. 31, no. 6, pp. 617-623, Dec. 1999.
[73] W. J. Horrey, and C. D. Wickens, "Examining the impact of cell phone conversations on driving using meta-analytic techniques," *Hum. Factors,* vol. 48, no. 1, pp. 196-205, Feb. 2006.
[74] J. D. Lee, B. Caven, S. Haake, and T. L. Brown, "Speech-based interaction with in-vehicle computers: the effect of speech-based e-mail on drivers' attention to the roadway," *Hum. Factors,* vol. 43, no. 4, pp. 631-640, Feb. 2001.
[75] B. Reimer, B. Mehler, Y. Wang, and J. F. Coughlin, "A field study on the impact of variations in short-term memory demands on drivers' visual attention and driving performance across three age groups," *Hum. Factors,* vol. 54, no. 3, pp. 454-468, May. 2012.
[76] B. Mehler, B. Reimer, and J. F. Coughlin, "Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task," *Hum. Factors,* vol. 54, no. 3, pp. 396-412, Jun. 2012.
[77] P. A. Hancock, M. Lesch, and L. Simmons, "The distraction effects of phone use during a crucial driving maneuver," *Accid. Anal. Prev.,* vol. 35, no. 4, pp. 501-514, Aug. 2003.
[78] R. Stojan, and C. Voelcker-Rehage, "Neurophysiological correlates of age differences in driving behavior during concurrent subtask performance," *Neuroimage,* vol. 225, pp. 117492, Jan. 2021.
[79] G. Li *et al.*, "A temporal-spatial deep learning approach for driver distraction detection based on EEG signals," *IEEE Trans. Auto. Sci. Eng.*, pp. 1-13, 2021.
[80] J. Xie, A. R. Hilal, and D. Kulic, "Driver distraction recognition based on smartphone sensor data," in *IEEE Int. Conf. Syst., Man, Cyber. (SMC)*, Miyazaki, Japan, 2018, pp. 801-806.

Xin Zuo received the B.S. in Naval Architecture and Ocean Engineering from Harbin Engineering University, China, in 2016, and the M.S. in Design and Manufacture of Ship and Ocean structure from Dalian University of Technology, in 2019. She is currently pursuing the Ph.D. degree in Software and Communications Engineering at University of Jyväskylä, Jyväskylä, Finland.

Her research interests include driver distraction, biomedical signal processing.

Chi Zhang received the B.S., M.S., and Ph.D. degree of Science in Engineering from Northeastern University, China, in 2010, 2012 and 2016, respectively.

He is an Associate Professor with the School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of

Technology, Dalian, China. His research interests include biomedical signal processing, brain-computer interface, and artificial intelligence.

Fengyu Cong received the B.S. degree in Power and Thermal Dynamic Engineering and the Ph.D. degree in Mechanical Design and Theory from Shanghai Jiao Tong University, China, in 2002 and 2007, and the Ph.D. degree in Mathematical Information Technology from the University of Jyväskylä, Jyväskylä, Finland, in 2010.

He is a Professor in the School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China. His current research interests include brain signal processing, acoustic signal processing, independent component analysis, tensor decomposition, and pattern recognition/machine learning/data mining.

Jian Zhao received the M.S. and Ph.D. degree in Mechanical Engineering from Xidian University, China, in 2006 and 2008, respectively.

He is currently a Professor at the School of Automotive Engineering, Faculty of Vehicle Engineering and Mechanics, Dalian University of Technology, Dalian, China. His current research interests include autonomous driving, MEMS sensors, compliant mechanics, and nonlinear dynamics.

Timo Hämäläinen received the Ph.D. degree in telecommunication from the University of Jyväskylä, Jyväskylä, Finland, in 2002.

In 1997, he joined the University of Jyväskylä, where he is currently a Professor of computer networks. He has more than 25 years research and teaching experience of computer networks. He has led many external funded network management related projects. He has launched and leads master programs with the University of Jyväskylä (SW & Comm. Eng.), and teaches network management related courses. He has more than 200 internationally peer reviewed publications and he has supervised almost 40 Ph.D. dissertation. His research interests include network resource management, IoT, and networking security.

III

DRIVER DISTRACTION DETECTION BASED ON MSAE OF MULTI-MODALITY PHYSIOLOGICAL SIGNALS

by

Xin Zuo, Chi Zhang, Fengyu Cong, Jian Zhao, and Timo Hämäläinen 2023

Submitted to IEEE Transactions on Intelligent Transportation Systems

Request a copy from the author.

Driver Distraction Detection Based on MSaE of Multi-modality Physiological Signals

Xin Zuo, Chi Zhang, Fengyu Cong, Jian Zhao, and Timo Hämäläinen

***Abstract*—Driver distraction, a major cause of traffic crashes, is reported to reduce driving performance and be detected with vehicle behavioral features. It also induces physiological responses. Time and frequency-domain features of physiological signals have been used to study distraction, but they are susceptible to residual noise and overlook the complexity. Moreover, the resampling problem arises while analyzing physiological signals at multiple time scales. This paper proposes a novel framework based on multiscale entropy on absolute time scales (MSaE) and long short-term memory (LSTM) network to mine the distraction information in multi-modality physiological signals and detect distraction automatically. Firstly, an entropy-based resampling method is adopted to find the suitable downsampling rates of electroencephalography (EEG), electrocardiogram (ECG), and electromyography (EMG). Then, calculating entropy with absolute time scales instead of relative time scales in a sliding window is utilized to explore the fluctuations of each signal while distraction. Afterward, ReliefF is selected from conventional feature selectors to identify the optimal feature set for each signal. Finally, LSTM with time dependency is designed to detect driver distraction with the selected feature set. The results illustrate significant distinctions in the MSaE of multiple physiological signals between normal and distracted driving. Additionally, MSaE, superior to traditional features, is selected as the most discriminative feature for each signal in distraction mining. Furthermore, the accuracy is further improved by about 9%, incorporating multi-modality features rather than vehicle behavioral features. This study indicates the potential of employing various signals to understand and detect driver distraction effectively.**

***Index Terms*—Driver distraction, EEG, ECG, EMG, MSaE, multi-modality analysis**

This work was supported in part by the National Natural Science Foundation of China under Grant 61703069, in part by the Fundamental Research Funds for the Central Universities under Grant DUT18RC(4)035 and in part by the Science and Technology Planning Project of Liaoning Province under Grant 2021JH1/10400049. Xin Zuo and Chi Zhang have contributed equally to this work. *(Corresponding author: Chi Zhang; Jian Zhao)*.

Xin Zuo and Fengyu Cong are with the School of Biomedical Engineering, Faculty of Medicine, Dalian University of Technology, Linggong Road #2, Dalian 116024, China, and also with the Faculty of Information Technology, University of Jyväskylä, Mattilanniemi 2, Jyväskylä FIN-40014, Finland (e-mail: zuoxin93@foxmail.com; cong@dlut.edu.cn).

Chi Zhang is with the School of Biomedical Engineering, Faculty of Medicine, Dalian University of Technology, Linggong Road #2, Dalian 116024, China (e-mail: chizhang@dlut.edu.cn).

Jian Zhao is with the School of Automotive Engineering, Faculty of Vehicle Engineering and Mechanics, Dalian University of Technology, Linggong Road #2, Dalian 116024, China (e-mail: jzhao@dlut.edu.cn).

Timo Hämäläinen is with the Faculty of Information Technology, University of Jyväskylä, Mattilanniemi 2, Jyväskylä FIN-40014, Finland (e-mail: timo.t.hamalainen@jyu.fi).

I. INTRODUCTION

TRAFFIC accidents are happening worldwide with increasing frequency in recent years. World Health Organization (WHO) reported in 2022 that the lives of about 1.3 million people are taken away because of road accidents and that between 20 and 50 more people suffer from different levels of injuries every year [1]. According to the statistics of The Daily Star in 2020 [2], there are approximately 90% of traffic accidents occurring as a result of the abnormal behaviors of drivers such as talking with others, operating devices, drinking and eating etc. These kinds of behaviors can increase the drivers' workload and easily divert drivers' attention from the task of driving to unrelated activities. Thus, driver distraction appears. In addition, distraction caused by mobile phones is a growing concern for road safety among various causes of distraction nowadays. In a survey of WHO [1], using mobile phones would significantly increase drivers' exposure to the risk of car crashes by about four times. It can slow down the reaction time of drivers and make it difficult to keep the right forward direction, then leading to accidents. Therefore, it is necessary to detect driver distraction in time and alert drivers to observe the surroundings and focus on driving to reduce such traffic accidents.

A variety of signals have been used for evaluating driver state and detecting driver distraction. The most commonly used signals in the existing research are obtained by non-intrusive sensors, including Global Position System (GPS), external and internal cameras [3], [4]. These sensors can easily collect the vehicle behavioral signals (like steering wheel angle, acceleration, speed, etc.) and driver visual images (like head movement, eye fixation, driving posture, etc.). Driver distraction and situation awareness were evaluated in [5] by analyzing the lane keeping and vehicle velocity data, and the performance of different age, gender, and driving experience groups were compared. Eye movement data was used to study the relationship between response performance and cognitive workload while drivers were talking with passengers in [6]. Although the vehicle behavioral signals are easy to collect and analyze, they are subject dependent and sensitive to traffic conditions and weather [7]. What's more, driver visual images can be influenced by facial occlusion, illumination, and personal driving habits easily even if they can reflect drivers' attention shift intuitively [8].

There is also research studying driver distraction utilizing intrusive sensors to acquire physiological signals like electroencephalography (EEG), electromyography (EMG), and electrocardiogram (ECG). Li *et al.* [9] proposed a deep learning approach to detect driver distraction on the basis of the temporal and spatial information of EEG signals and examined the contributions of temporal and spatial information. Khushaba *et al.* [10] designed a muscle computer interface to reduce distraction appearance by sensing and decoding EMG signals associated with different finger postures while distraction. Taherisadr *et al.* [11] collected ECG signals for driver inattention identification during naturalistic driving. Physiological signals record the real electrical activities of human internal state, which provide a more accurate and faster way to explore the internal state alterations.

Recent studies have demonstrated that none of the signals can provide sufficient information to estimate driver mental state and incorporate signals can compensate for each other [12]–[14]. Moreover, along with the development of sensor technology and less intrusive wearable equipment, multi-modality signals are more and more widely used to study driver distraction. Du *et al.* [15] extracted multi-modality features from speech, facial expression, and car signals to detect driver distraction, and their results show that the predictive accuracy increases by adding more modalities features. Lechner *et al.* [16] proposed a lightweight framework to explore comprehensive detection of driver distraction fusing multi-sensor data from drivers, vehicle, and GPS position, and validated its feasibility. Zhang *et al.* [17] introduced a multi-modal fusion network for detecting driver distraction with acoustic signal, visual signal, EMG, and car motion signals. The results show that multi-modality signals can greatly enhance the performance of the proposed network.

Nonetheless, physiological signals are usually contaminated by breath, blinks, and body movements etc. especially in driving experiments. Residual noise retains after preprocessing and has effect on analyzing physiological signals, since noise contamination cannot be absolutely eliminated by preprocessing [18]. Additionally, driver's physiological status changes dynamically while driving, the rhythmicity and stability of different physiological signals alters along with it [19]. Therefore, it is necessary to mine the distraction information with the complexity of physiological signals. But how their complexity varies between distraction and normal needs to be further studied. To overcome these challenges, we proposed a driver distraction detection framework based on multiscale entropy on absolute time scales (MSaE) of multi-modality physiological signals (i.e., EEG, ECG and EMG signals) in this study.

The remainder of the paper is organized as follows. Section II lists relevant research on the estimation of driver distraction. Section III describes the experimental details and the gathered multi-modality signals. In Section IV, the adopted methodologies are introduced. The findings of our study are presented in Section V and discussed in Section VI. Finally, Section VII summarizes the paper.

II. RELATED WORK

For detecting driver distraction, the popular way is to first extract features from signals and then input the features into classification models. There are various features used to excavate distraction information existing in different signals in the literature. The time-domain features like mean value and standard deviation of vehicle behavioral signals are usually calculated to evaluate changes in drivers' performance due to distraction. Zhang *et al.* [20] selected the mean and standard deviation of speed, acceleration, steering wheel, and lane deviation data as indicators to identify truck driver distraction. Kountouriotis *et al.* [21] calculated the mean speed, lateral position standard deviation and the mean angular acceleration of the steering wheel to analyze the differences between normal and distracted driving. They found that steering wheel indicators are adequate to distinct distraction and non-distraction. Many studies adopt features based on eye movement (e.g., fixation duration, saccade frequency and gaze distribution) of driver visual images to explore the useful information while distraction. Liang *et al.* [22] extracted not only the vehicle behavioral features but also the fixation duration and blink frequency of the eye data and fed them into Bayesian Networks (BNs) model to detect driver distraction. Kircher *et al.* [23] analyzed the glance behavior of drivers and demonstrated that it serves as a good measure of drivers' attention shift, which makes it possible to detect distraction. As for physiological signals, most of the existing research utilize time-domain and frequency-domain features to explore the internal distraction information. Alizadeh *et al.* [24] extracted the power, standard deviation and mean absolute values of four EEG rhythms to compare the performance of various classification methods detecting driver distraction. Deshmukh *et al.* [25] analyzed all possible sub-bands of ECG spectrum and obtained their mean, standard deviation, and power features to identify driver distraction. Sahayadhas *et al.* [26] presented a driver inattention detection system using the statistical and spectral features of EMG and ECG and illustrated that EMG and ECG signals can be adopted to detect inattention. These features can indicate the dynamic changes of the physiological signals in time-domain and frequency-domain while distraction, but the complexity information is neglected to some extent.

Physiological signals are recordings of the electrical activities derived from human body [27]. Extensive amounts of complexity information about human states hide within the fluctuations of signals [28], especially EEG. Since EEG generally records the brain activities of the whole scalp with multi-electrodes rather than in one region. Plenty of valuable information might be overlooked if we do not consider the complexity features. Entropy is a measure of the disorder or randomness of a system, thus it can be selected to manifest the complexity of physiological signals [29]. In this case, entropy features have drawn researchers' attention to use them estimating human states recently. Zhang *et al.* [19] proposed to implement the fuzzy entropy (FE) and sample entropy (SE) of EEG and electrooculogram (EOG) signals to analyze and classify different sleep stages. Their results are slightly better than existing methods. Kumar *et al.* [30] trained an epileptic seizure detection model employing differential entropy (DE) and peak-magnitude of root mean square ratio (PRMS) of EEG. They found that the extracted features could distinguish normal and abnormal seizure EEG with a superior prediction accuracy. Fraiwan *et al.* [31] built a machine learning (ML) model to estimates the visual interest level with multiscale entropy (MSE) of EEG. The results indicate that MSE can quantify EEG complexity to reflect human visual interest. Apart from the advantages, there are still challenges in exploring the complexity of physiological signals applying entropy features. Artifacts cannot be completely eliminated whatever preprocessing approaches are employed [18]. Hence, the calculated entropy features (like FE, SE and DE) involve the residual noise in the complexity of signals, which results in the poor robustness of these features. Besides, researchers have proved that the value of entropy could be affected with sampling rate, e.g., the value of SE will obviously increase while signals are downsampled [32], [33]. Compared to the above single scale entropy features, MSE is better in robustness as it is calculated at multiple time scales to decrease the effect of residual white noise [34]. Nevertheless, resampling problem arises when calculating entropy in different time scales that compressing or stretching effect occurs if the sampling rate of signals is changed [35]. Moreover, the sampling rates of signals have increased along with the advanced sensor technology nowadays. A rise then appears in the data size and computational cost. To reduce the extensive time-consuming calculation, most research apply the downsampling approach prior to analysis. Under the circumstances, how to find the appropriate downsampling frequency and manifest the complexity of physiological signals can be challenging.

As for distraction detection, a variety of classification algorithms have been used in the literature. Deep learning (DL), a branch of ML, has a better understanding of large data with deep neural networks than traditional ML [36]. Recent research has pointed out that DL outperforms many conventional ML approaches for time series classification, and that it appears to be the most promising algorithm to classify temporal signals [37]. Ak *et al.* [38] trained a motor imagery EEG signal classification model for robot control utilizing conventional neural network (CNN) and reached 85% accuracy. Ranganathan *et al.* [39] presented a multimodal emotion recognition method adopting deep belief network (DBN) and validated that better recognition performance could be obtained in subtle emotion expressions with DBN. Driving is a continuous work with the characteristics of long duration and context dependent. Decisions on each action is made upon the current state as well as many previous states, and it has been reported that context dependent is vital for detecting human mental states [22]. Therefore, adding memory to a neural network can learn more information about driver distraction from time series and improve classification performance [40]. Recurrent neural network (RNN) can memorize the previous states of sequential data and decide the output according to the history for a short duration time series [41]. Long short-term memory (LSTM) network, an extension of RNN, achieves to keep both long-term and short-term dependency by adding four gates in each cell [42]. It overcomes the vanishing gradient problem of RNN and has been utilized in many fields. Zhang *et al.* [43] developed a LSTM model to classify hand movement with time and frequency features of EEG and got an accuracy of 83.2%. Hong *et al.* [44] introduced a DL model dealing with LSTM for predicting the development of Alzheimer's Disease (AD). They demonstrated that the performance of LSTM is superior to existing algorithms for the future state prediction of AD. Sun *et al.* [45] investigated sleep staging with ECG and respiratory signals applying LSTM and validated the possibility to adopt LSTM in sleep states estimation.

In this paper, we propose a framework for driver distraction detection based on MSaE of multi-modality physiological signals captured in simulated driving environment. Our method is to determine the appropriate downsampling rates of multi-modality physiological signals with an entropy-based method, to exploit the complexity variations of each signal while distraction with the MSaE feature, and to learn the long-term contextual information in multi-modality features utilizing LSTM. A simulated driving experiment is designed to induce driver distraction and then executed to collect multi-modality signals. After that, MSaE is extracted to investigate drivers' mental states, and a variety of features (including entropy features, time-domain, and frequency-domain features) are calculated for comparison. In addition, different feature selectors are compared to provide a reference for feature reduction and fusion. Finally, the selected optimal features are fed into LSTM model to explore the time dependent relationship in them and detect driver distraction.

III. EXPERIMENT

The experiment is designed to arouse driver distraction by mobile phone usage task in simulated driving environment and collect multi-modality signals to be used for driver distraction detection.

A. Subjects

This study was reviewed and approved by Ethics Committee, Liaoning Normal University. A total of sixty physically and mentally healthy subjects participated in the experiment. All of them are right-handed and have normal or corrected to normal vision and normal hearing. Driving license, driving experience as well as using mobile phone experience are also required. Besides, they are instructed not to smoke and consume any tea, alcohol, coffee, or medicine a day before the experiment. The experiment content was introduced to all participants and written informed consent were signed previous to the experiment.

B. Experiment Design

The experiment was conducted in the laboratory at Liaoning Normal University. The Xuan Love QJ-3A1 car driving simulator was used to display the driving scenario and gather vehicle behavioral signals. It is equipped with interactive visual system, simulated cockpit, electronic control system, customized software and accessory equipment as shown in Fig. 1(a).

As physiological signals can reflect the internal changes of driver state more accurately and timely, a 64-channel electrode cap following the international 10-20 system was utilized to acquire EEG signals. Besides, we also gathered ECG and EMG signals at the same time in the experiment. ECG signal was captured with chest lead III as shown in Fig. 1(b). The EMG electrode was placed along the fiber of the soleus muscle in the right lower limb. The sampling frequency of physiological signals was set as 2000 Hz. To evaluate how distraction affect driving performance, the vehicle behavioral signals were collected in the experiment as well, and the sampling rate was kept at 1 Hz.

In this experiment, driver distraction is designed to be induced by a mobile phone use task that drivers are forced to keep talking on the phone during distraction process.

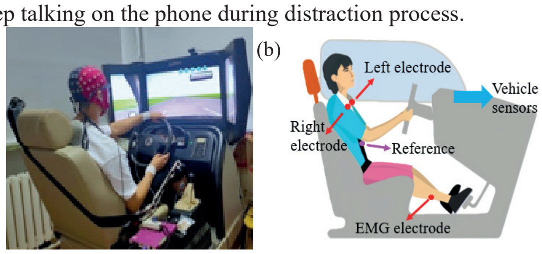
(a) (b)


Fig. 1. Experiment apparatus. (a) The driving simulator and electrode cap. (b) The diagram of ECG, EMG and vehicle sensors.

C. Procedure

There is a practice phase prior to the formal experiment to familiar subjects with simulated driving until they understand the experiment thoroughly. The formal experiment is divided into 6 blocks with a duration of about one hour. Each block lasts for about 10 minutes consisting of approximately 7 minutes normal driving at the beginning and then around 3 minutes distracted driving. Subjects are told to pay full attention to the road in normal driving process, while experimenter call and keep talking with the subjects in distracted driving. Detailed protocol of the experiment is shown in Fig. 2. Vehicle behavioral signals and physiological signals are recorded during the experiment.

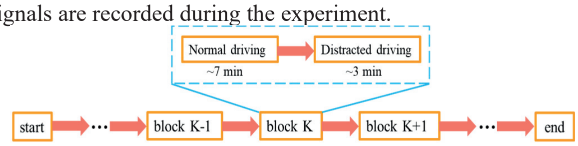


Fig. 2. Protocol of the experiment.

IV. METHODOLOGY

A. Preprocessing and Feature Extraction

In this subsection, we introduce how the gathered multi-modality signals were preprocessed and what kinds of features were extracted in this study.

1) **EEG Analysis**

We first extracted the alpha rhythm of EEG employing wavelet decomposition with 4 levels, as it has been validated that alpha rhythm is highly correlated with distraction [46]. Wavelet transform (WT) has been widely used in analyzing non-stationary time series with the advantages of differentiating time series consisting of different frequencies and retaining both time and frequency information of the signals [47]. It decomposes signals with shifted and scaled versions of a mother wavelet $\psi(t)$ and a scaling function $\phi(t)$. The discrete mother wavelet can be mathematically expressed as

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi(2^{-j}t - k) \qquad k,j \in Z \qquad (1)$$

The original signal $S(t)$ can be defined as

$$S(t) = \sum_k s_j(k)\phi_{j,k}(t) + \sum_k d_j(k)\,\psi_{j,k}(t) \qquad (2)$$

where $s_j(k)$ and $d_j(k)$ represents the approximate and detailed coefficients at j level. In our study, we selected the commonly used db6 wavelet as the mother wavelet.

Then, to reduce the time-consuming feature calculation of EEG, the signals need to be downsampled. We applied an entropy-based approach to find the appropriate downsampling frequency, which is mainly on the basis of MSE and MSaE.

MSE, developed by Costa et *al.* [48], can quantify the complexity of signals over various time scales as an extension of SE. MSE algorithm can be briefly summarized as coarse-graining and SE calculation.

a) For the given EEG signal $\{x_1, \ldots, x_i, \ldots, x_N\}$, consecutive coarse-grained time series $\{y^{(\tau)}\}$ at time scales τ should be constructed at first by averaging the τ data points in successive non-overlapping windows (see Fig. 3). The element of $\{y^{(\tau)}\}$ can be calculated by the following equation

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i \qquad 1 \le j \le N/\tau \qquad (3)$$

where N is the length of the given EEG signal, N/τ is the length of each coarse-grained time series.

b) SE is then calculated for coarse-grained time series. As a measure of the complexity of signals, it has been adopted in EEG analysis and is detailed introduced in [49].

Previous studies have illustrated that the values of single scale entropy features are related to sampling rate and stretching or compressing effect appears in MSE with sampling frequency changes [32], [33], [35]. In the present study, we explored an extension of MSE (i.e., MSaE) that replaces the traditional relative time scale τ with absolute time scale λ with unit second,

$$\lambda = \tau \cdot T_s = \tau / f_s \qquad (4)$$

where T_s and f_s are the sampling period and sampling rate of original EEG, respectively.

In this way, the coarse-graining process can be described as: constructing the coarse-grained time series by averaging the data points every λ seconds in successive non-overlapping windows as shown in Fig. 3. There should be noticed that the time duration t_D of EEG keeps the same before and after downsampling [35], i.e., t_{D_d}=t_D. The rest procedure to calculate MSaE is the same with MSE. Hence, MSaE feature can be obtained by calculating SE of the coarsely grained time series under absolute scale factors.

Now, the entropy-based approach to determine the proper downsampling frequency for the given EEG signal X can be illustrated as:

a) Calculate MSE at different time scale τ and draw the entropy-scale curve to find peak scale τ_{peak}, the scale that MSE reaches its maximum value (see Fig. 4). High correlation between MSE value and time scale τ_{peak} has been demonstrated in [50].
b) Calculate the corresponding peak time λ_{peak} on the basis of (4).
c) Since downsampling has no effect on time duration, λ_{peak} remains unchanged before and after downsampling, i.e., $\lambda_{peak\_d} = \lambda_{peak}$.
d) According to (4), the downsampling rate f_{s_d} can be expressed as

$$f_{s\_d} = \frac{\tau_{peak\_d}}{\lambda_{peak\_d}} = \frac{\tau_{peak\_d}}{\lambda_{peak}} \qquad (5)$$

e) Finally, we can get f_{s_d} for a given peak scale after downsampling τ_{peak_d}.

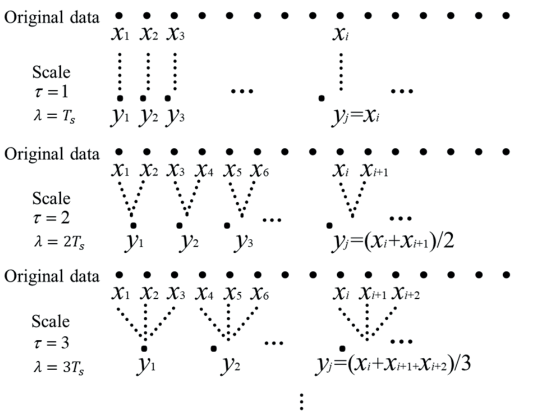


Fig. 3. Schematic illustration of the coarse-graining process.

In the present study, we calculated the entropy-scale curve of alpha rhythm as shown in Fig. 4. τ_{peak} is 60 and λ_{peak} is 0.03 seconds. Let $\tau_{peak\_d} = 6$ referring to the recommended scale range in [51], then the appropriate downsampling rate is calculated as 200 Hz.

After the alpha rhythm was downsampled to 200 Hz, we employed a wavelet-based technique for artifacts removal. The approximate and detailed coefficients mentioned above show the correlation in mother wavelet and the signal, and larger coefficients appear if artifacts exist. These coefficients can be reduced with a threshold and has been used in the analysis of driver fatigue [52]. The threshold is expressed as

$$T_j = mean(C_j) + 2 \times std(C_j) \qquad (6)$$

where C_j is the wavelet coefficient at jth level. If the coefficient's value is larger than the predefined threshold it is halved. Then the signal without artifacts can be obtained with the new coefficients.

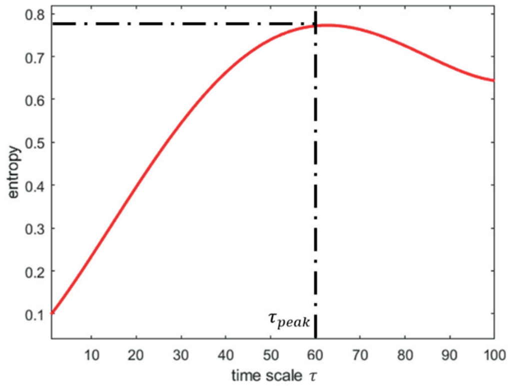


Fig. 4. Entropy-scale curve of the original alpha rhythm with f_s=2000 Hz.

Except for MSaE and MSE, another 11 features of alpha rhythm from time domain, frequency domain and complexity were extracted for comparison. Time-domain features include mean value (Mean), standard deviation (STD), skewness and kurtosis. The amplitude spectrum (AMP) and power spectrum density (PSD) were calculated from the perspective of frequency. As for complexity features, we chose to use DE, FE, approximate entropy (AE), Rényi entropy (RE) and fractal dimension (FD). The detailed algorithms are listed in [29], [43].

2) **ECG Analysis**

ECG signal is the record of the electrical activities of heart, which contains various artifacts such as device, movement, and breath interferences [53]. In order to eliminate these artifacts, it is necessary to preprocess the collected ECG signal. Firstly, the downsampling rate was calculated to decrease the extensive calculation according to the entropy-based resampling approach mentioned above. In this way, the resampling rate of 256 Hz was obtained with a scale of 6. Then, a fourth-order Butterworth band-stop filter of 50 Hz was designed to remove power line noise and a band-pass filter of 0.7 Hz to 40 Hz was applied to eliminate the baseline drift caused by movements and breath. After preprocessing, ECG signal can be used to analyze the activities of heart.

The MSaE and MSE features of ECG in a time window of 5 seconds with 4 seconds overlapping were applied to explore the complexity of ECG. Besides, existing research has pointed out that heart rate variability (HRV) derived from ECG is a reliable indicator for analyzing heart activities [54], [55]. By R-peaks (heart beats) detection, the HRV indicators can be extracted. In this study, we adopted a dynamic threshold approach according to the work of Pan et *al.* [56] to locate R peaks in each block. Thereafter, 8 HRV features from both time domain and frequency domain were extracted in 60 seconds sliding window with 59 seconds overlapping. The time-domain features are the number of heart beats per minute (BPM), the mean distance (IBI) and standard deviation (SDNN) of R-R intervals, the standard deviation (SDSD) and the root mean square (RMSSD) of successive differences between adjacent R-R intervals. The frequency-domain features involve low-frequency (LF, 0.04-0.15 Hz) spectral component, high-frequency (HF, 0.16-0.5 Hz) component and the ratio of LF and HF (LF/HF). A detailed description of these measures can be found in [54], [57].

3) **EMG Analysis**

The data was firstly band-pass filtered with a fourth-order Butterworth band-pass filter and the low and high cutoff frequencies were set as 20 Hz and 500 Hz, respectively in accordance with [58]. A band-stop filter was then utilized to remove the interference of power line, after which a sliding window of 125 milliseconds is employed to average the filter EMG data. Finally, to observe the changes of EMG before and after distraction, we extracted 8 features within a window length of 60 seconds and 59 seconds overlapping. The features are MSaE, MSE, AE, the mean absolute value of amplitude (MAV), the root mean square of EMG amplitude (RMS), zero crossing (ZeroCross), mean power frequency (FMEAN), median frequency (FMED). More details of these measures are described in [59], [60].

4) **Vehicle Behavioral Signals Analysis**

The velocity (V) and lane position variability (LPV) signals were analyzed to validate whether driving performance does alter with distraction. We performed statistical analysis for the signals so as to explore how behavioral alterations happen between normal driving and distracted driving. The mean value and standard deviation of the data were calculated with the same sliding window as ECG and EMG.

B. Feature Selection

After preprocessing and feature extraction of the physiological signals, we get a feature set with 31 kinds of features with the dimension of 811. These indicators are in different numerical ranges, which have effect on the comparison among them. Thereupon, each feature is normalized to [0,1] applying (7) to avoid the numerical influence.

$$y = \frac{(y_{max}-y_{min})*(x-x_{min})}{x_{max}-x_{min}} + y_{min} \quad (7)$$

where y is the normalized value, x is the original value, x_{max} and x_{min} are the maximum and minimum values of the original feature, respectively. y_{max} and y_{min} stand for the corresponding normalized maximum and minimum values separately.

Since the dimension of the extracted feature set for distraction detection is high, it is time-consuming and may induce overfitting problem if put all of them into classifier. On this account, it is necessary to fulfill the feature selection procedure to reduce features' redundancy and improve efficiency. Five commonly used feature selectors were recruited in this study to find out the most discriminative features [61]-[65]. They are briefly illustrated below.

1) **ReliefF Algorithm**

ReliefF algorithm is an efficient feature weighting algorithm that assigns the features with different weights in agreement with the correlation between features and categories [61]. The greater the weight of the feature, the better it is for classification. It explores the correlation by searching for the nearest neighbor of each feature from all categories utilizing Manhattan Distance. As the feature weight vector is stable and ReliefF can do with noise and incomplete problems, it is first selected for feature selection in our study.

2) Non-negative Matrix Factorization (NMF)

NMF is an unsupervised algorithm processing high dimensional data with matrix factorization. It searches for non-negative basis and coefficient matrix to parts-basely represent a non-negative data matrix, which retain the discriminative information of the data. Then, reduce the dimension of the data matrix. Extensive research has proved that NMF is effective in dealing with high dimensional data for feature selection in fields of physiology and neuropsychology [62].

3) Mutual Information (MI)

MI measures the amount of information that one variable has about the other and provides a way to estimate the dependency between features and categories. The feature, with larger MI value, is more discerning to the corresponding category. With advantages of detecting nonlinear relationships between variables and analyzing multidimensional variables, MI has been a popular approach for feature selection [63].

4) Neighborhood Component Analysis (NCA)

NCA is a non-parametric method for selecting features with the goal to maximize the prediction accuracy of classification. It optimizes the leave-one-out classification by means of the first order nearest neighbor, and then obtains the weight of each feature [64]. By analyzing the weight vector with a threshold, the most relevant feature subset can be determined. Besides, the result is barely affected by the increase in the number of irrelevant features.

5) Sequential Forward Selection (SFS)

SFS is a bottom-up search process that begins with an empty feature set and subsequently add features on the basis of minimizing the mean square error [65]. The process continues until the performance achieves the highest, and the corresponding feature subset is regard as the selected features. It is widely used for dimension reduction due to saving computing time and its simplicity.

D. Classification

In order to find out the most discriminative features for distraction detection, we recruited random forest (RF) to compare the performance of different feature selectors. After that, the new set features coming from the best feature selector will be input into a LSTM classifier to detect driver distraction.

1) Random Forest

RF is a kind of bagging model for classification consisting of many randomized decision trees in accordance with ensemble learning technique. Each tree has its local decision, and the final decision is predicted by the majority votes of all trees. The tree is trained with a bootstrap aggregating technique that randomly selected the subsets with replacement from the original training data [66]. Therefore, the stability and robustness of the classifier is increased. In addition, it can also estimate the importance of the inputs. Considering the two merits, it is adopted to evaluate the most efficient feature selector among ReliefF, NMF, MI, NCA and SFS in the present study.

2) LSTM

LSTM, a variant of RNN, addresses the vanishing gradient problem and can keep both long and short-term memory of a longer time series. A memory cell configuring with four neural layers is used to learn long-term dependency and store valuable information in the context (see Fig. 5). The LSTM cell utilizes three so called gates (i.e., forget gate, input gate and output gate) to control the information propagation in the network instead of directly overwriting the cell information in RNN [42].

The input data is transferred to every gate by a sigmoid activation function. The forget gate is firstly used to determine whether there is information need to be discarded from the inner cell state (8). Then, the input gate controls the amount of new information that can be stored in the cell following three steps. Step 1 involves a sigmoid layer exploring any updated information using (9). After that, a tanh layer creates $\widetilde{C}_t$, which is the new candidate cell according to (10). The new cell state C_t substitutes the old cell state C_{t-1} with (11) in the last step. Finally, the output h_t is determined by the output gate by (12) and (13).

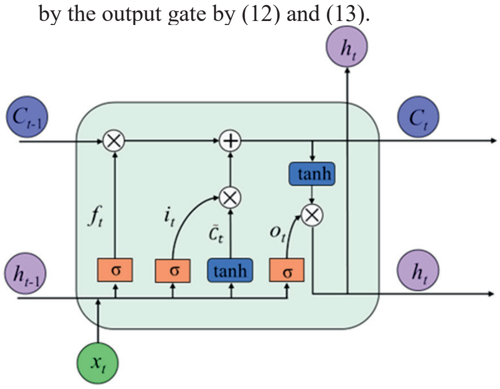


Fig. 5. The basic architecture of LSTM memory cell.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (8)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (9)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (10)$$
$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \quad (11)$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12)$$
$$h_t = o_t * \tanh(C_t) \quad (13)$$

where x_t is the feature of time t, σ and tanh are the sigmoid and tanh activation function in the cell, respectively. W, b and h separately indicate the weight, bias and hidden state of each gate.

V. RESULTS

A. Analysis of EEG Signal

A total of 13 features were calculated from alpha frequency band in the present work, and we mainly focus on the MSaE feature to evaluate the complexity of EEG because of its advantages mentioned above. After determining the appropriate downsampling rate and scale parameter, the MSaE feature was extracted in a sliding time window of 5 seconds with 4 seconds overlapping. As shown in Fig. 6, the waveform of MSaE fluctuates gently in the normal driving process, while it is more obvious in the distracted driving process. A minimum value appears in a while after the onset of the distraction task, and the trough of the waveform is apparently lower than the average MSaE value.

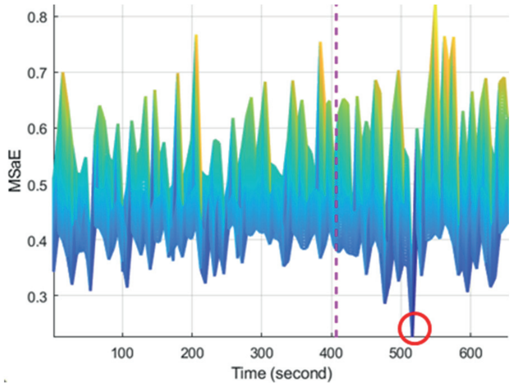


Fig. 6. The MSaE result of alpha rhythm. The magenta dash line shows the onset of using mobile phone, the red circle is the minimum MSaE value.

B. Analysis of ECG Signal

Two entropy features, five time-domain features and three frequency-domain features of ECG were extracted to find out the dynamic fluctuation of heart rate during distraction. Fig. 7 shows the obtained MSaE feature. From an overall perspective, it shows an increasing tendency after the onset of the using mobile phone and rises to its maximum value later. The average MSaE value also augments after the task. In addition, the fluctuation of the waveform is more obvious after distraction than that before distraction.

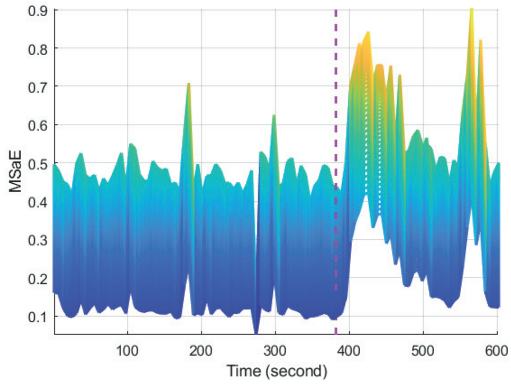


Fig. 7. The MSaE result of ECG. The magenta dash line shows the onset of the distraction task.

C. Analysis of EMG Signal

We currently calculated three entropy, three time-domain and two frequency-domain features for EMG signal so as to study the differences in muscle state between normal and distracted drivers. The MSaE feature is shown in Fig. 8. The values of MSaE show a slightly decreasing trend after drivers start to do the task until the minimum value is reached, which is evidently smaller than the mean value of MSaE in Fig. 8.

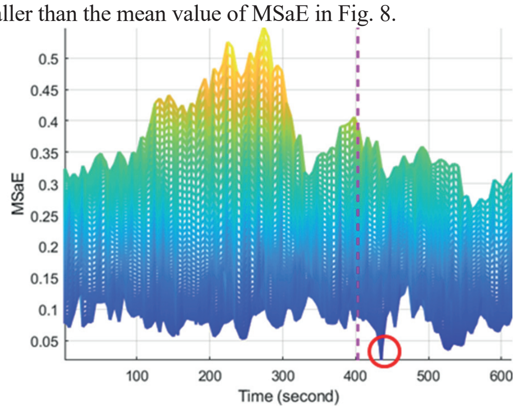


Fig. 8. The MSaE result of EMG. The magenta dash line shows the onset of the distraction task. The red circle is the minimum value.

D. Analysis of Vehicle Behavioral Signals

Statistical analysis was performed for the acquired vehicle behavioral signals to investigate how the driving performance changes after distraction. The mean value and standard deviation were calculated for V and LPV signals and shown in Fig. 9. The mean value of velocity shows a downward trend after using mobile phone in Fig. 9(a), while its standard deviation increases after distraction. In Fig. 9(c), the amplitude of LPV mean value rises due to distraction task. What's more, similar to the V standard deviation, the standard deviation of LPV also augments after distraction in Fig. 9(d).

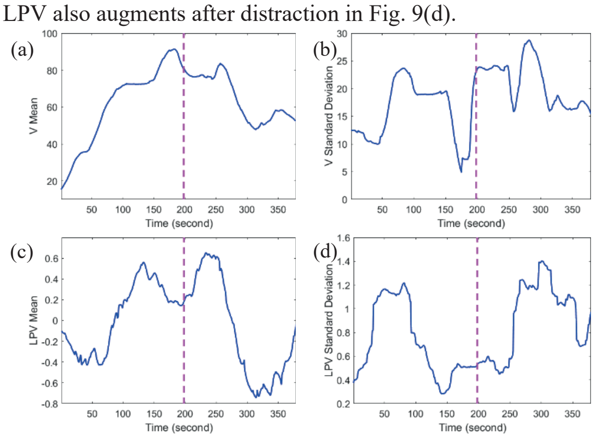


Fig. 9. The features of vehicle behavioral signals. (a) V Mean value. (b) V standard deviation. (c) LPV Mean value. (d) LPV Standard deviation. The magenta dash line shows the onset of the distraction task.

E. Performance Evaluation of Different Feature Selectors

As described in Section IV, there were 31 types of features extracted from EEG, ECG and EMG signals. The dimension of the obtained multi-modality physiological feature set is 811, which may result in overfitting problem. Hence, these features of each modality signal were fed into five feature selectors after normalization to explore the most discriminative features.

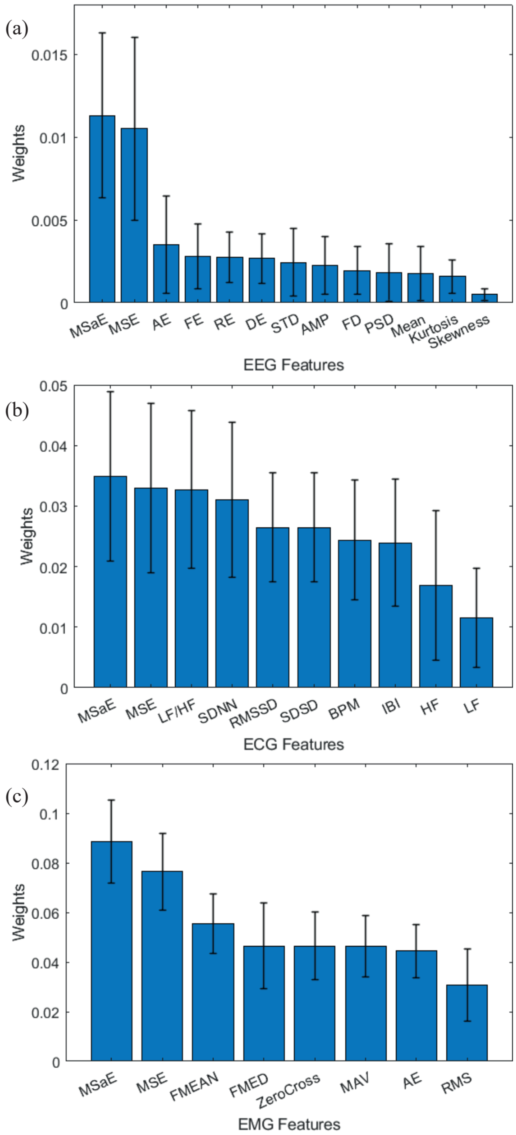


Fig. 10. The mean weights of different features obtained with ReliefF. (a) EEG features. (b) ECG features. (c) EMG features. Error bar shows the standard deviation.

Two steps were adopted in the procedure of EEG feature selection. Firstly, the three most discriminative electrodes were searched among all the 61 electrodes for each EEG feature. Then, the best feature for each electrode could be identified by analyzing the weights of the thirteen features (see Fig. 10(a)). Different from EEG, the electrode selection step was skipped for ECG and EMG features, and one best feature was chosen for each modal signal. Fig. 10 is the feature weights for each signal, it is clear that the weight of MSaE ranks the first among all features followed by MSE for each modal physiological signal utilizing ReliefF. The best feature selected by the feature selectors was then input into a RF classifier to estimate the most effective selector. The corresponding results are shown in Fig. 11. It is clear that the optimal feature selector is ReliefF for the feature set obtained by it displays higher accuracies than those of the other four selectors. On this occasion, the feature set selected by ReliefF was finally utilized to detect whether a driver is distracted or not in the following study.

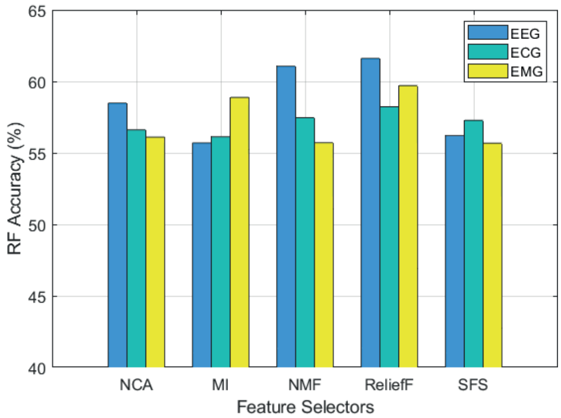


Fig. 11. The Performance of different feature selectors.

F. The Classification Results of LSTM

After acquiring the optimal feature selector and the most discriminative feature for each kind of physiological signal, the new feature set was fed into a LSTM model to detect distraction. There are two categories in our experiment i.e., normal and distraction. The input feature matrix was divided into training set and testing with a ratio of 8:2. In the present work, we conducted a comparative analysis of physiological signals in the context of distraction detection. Specifically, their performance was evaluated not only when employing the most effective feature to detect distraction, but also when using all extracted features as inputs for classification. In addition, we further investigated whether different signals can mutually provide complementary information by adopting the multi-modality feature set for distraction detection. Table I presents the results of the LSTM model with physiological features, and Table II lists the results of vehicle features. In Table I, "MSaE" is the selected best feature for each kind of physiological signal, and "MSE" is for comparative with MSaE. "All features" stands for all the extracted features of ECG and EMG signals, while it means all the features extracted from the selected electrodes of EEG signal. "ALL" is the feature set consisting of EEG, ECG, and EMG features. "Multi-modality" in Table II represents the multi-modality feature set comprising not only physiological features but also vehicle features.

Table I illustrates that MSaE performs better than MSE in detecting distraction for each kind of physiological signal. Its accuracy is also higher than that of utilizing multiple EEG or ECG features. As for EMG, it is 0.04% smaller than that of multiple features. Besides, the MSaE feature of EEG with an accuracy of 74.64% outperforms those of ECG and EMG when only one feature was used for classification in Table I. Similar conclusion can be drawn when inputting MSE to detect distraction. Additionally, the EMG feature exhibits the lowest accuracies in the single feature conditions, with accuracies of 64.81% for MSaE and 63.74% for MSE. The performance of ECG feature lies between EEG and EMG features while employing one feature to detect distraction. When utilizing MSaE of the three modalities physiological signals as inputs at the same time, the performance of LSTM was further enhanced, reaching the peak accuracy of 78.32%. The accuracy of LSTM with multi-modality feature set also increased under the condition of selecting the MSE feature of each signal. Moreover, when adding the multi-modality physiological features to vehicle features, the detection accuracy is promoted again with an augment of about 9% compared to that of vehicle features. In general, the highest accuracy is obtained by the MSaE feature of EEG when employing one feature of single modality physiological signal. In addition, MSaE is superior to the other single or multiple features of physiological signals adopted in our present study. Furthermore, the performance of LSTM can be enhanced if features of multiple modalities signals are simultaneously recruited.

TABLE I
THE MEAN ACCURACIES OF DIFFERENT PHYSIOLOGICAL FEATURES USING LSTM (%)

	EEG	ECG	EMG	ALL
MSaE	74.64	67.09	64.81	78.32
MSE	72.47	66.25	63.74	74.16
All features	63.69	66.11	64.85	63.86

TABLE II
THE MEAN ACCURACIES OF VEHICLE AND MULTI-MODALITY FEATURES USING LSTM (%)

	Vehicle	Multi-modality
Accuracy	72.61	81.27

VI. DISCUSSION

Driver distraction has effect on drivers from multiple aspects and threatens driving safety. It not only causes physiological changes of drivers but also affects the driving performance, which are manifested in the dynamic changes in physiological and vehicle behavioral signals [67]. The dynamic variations in the physiological signals and the distinctions in driving performance between normal and distraction are here discussed. A driver distraction detection model with the ability of memorizing long-term context information is proposed based on the extracted features. The results of single modality features are also compared with those of multi-modality features to validate whether taking the advantage of multi-modality signals is conducive to exploring more characteristics of driver distraction.

As shown in Fig. 6, the MSaE feature of alpha rhythm apparently changes after the start of phone use, which illustrates the changes in EEG complexity. Its waveform fluctuates more obvious in distracted driving than in normal driving. This phenomenon indicates the complexity of alpha frequency band is less stable while distraction. Besides, a trough occurs soon after the onset of the distraction task that represents the complexity of alpha rhythm arrives its minimum at this point. Drivers need to allocate most of the brain resources to pay full attention to the surroundings and keep driving safety in normal driving [68]. In this case, drivers' attention is focused, and brain activity is vibrant with much stable activation degree. The complexity of EEG then changes slightly that embodies the relatively stable MSaE waveform. Different from normal driving process, drivers' alertness reduces because of the distraction task, they also have to handle multiple tasks (i.e., driving and the distraction task) in distraction process, the activation degree of brain activity is thereby not as stable as normal driving. As a result, the fluctuation of MSaE is evident. What's more, it normally takes a while for drivers to shift their attention to distraction task and be distracted to the maximum [67]. The brain activity decreases to the lowest under this condition, therefore, a trough appears in MSaE sooner after the onset of the task.

The MSaE feature of ECG shows an increasing tendency after the onset of the task and arrives its peak later in Fig. 7. The waveform of MSaE also displays a more apparent fluctuation after distraction. In the normal driving, drivers only focus on driving safely and their heart rates are relatively regular and stable. In this occasion, the complexity of ECG slightly changes with driving, thus a stable waveform appears with lower values. While in distracted driving, drivers have to perform multiple tasks to satisfy the needs of answering phone and keeping safe at the same time, which leads to their nerves and decreases the regularity and stability of ECG. As a result, an augmentation occurs in the value of MSaE after distraction, and the fluctuation of the waveform gets more obvious. The dynamic alteration of MSaE is in line with previous study that analyzed ECG-based measures to monitor the physiological changes while distraction [69]. Its results pointed out that entropy features of ECG project higher values while distraction and that the complexity of ECG generates a sign of increasing during distraction. The reason is that multitasking (i.e., driving and secondary task) while distraction increases driver's workload and changes the functional state of heart [70]. Hence, the complexity of ECG signal augments. The results can also be explained from the biological aspect as reported in [71]. The variation of heart rate is brought about the blood regulation executed by heart. During distracted scenario, more oxygenated blood needs to be constantly supplied to the brain in order to satisfy the demands in multitasking. The increment of oxygen demand then increases blood flow, prompting the heart to contract and relax more frequently with lower rhythmicity and ultimately resulting in the changes in ECG feature.

MSaE of EMG signal is also analyzed in the present work. Fig. 8 shows that MSaE decreases after starting to use mobile phone and drops down to the minimum value in a short time after that. The reason why it declines and the minimum comes up may be that the muscle tension gets lower at this stage. Kurt *et al.* [72] argued that EMG signal is in connection with human mental states. High EMG tonus happens if human is awake, and lower EMG tonus arises while drowsiness. In our study, drivers remain vigilant to ensure driving safety with highly focused attention in normal driving. Consequently, they exhibit strong ability of controlling limbs with their muscles in an active and highly tense state, and then increasing the complexity of ECG signal. However, drivers' attention is distracted to answering phones after the onset of distraction task. The limbs control then recedes giving rise to the lower muscle vitality and tension. In this way, decline happens in the complexity of EMG. In addition, drivers normally spend a while to be distracted to the maximum [67], thereby, the muscle tension decreases to the lowest under the circumstances resulting in the minimum of MSaE. Moreover, the phenomenon can be further verified by the changes in vehicle behavioral data in Fig. 9.

The results of statistical analysis for V and LPV data are shown in Fig. 9. Drivers prefer to decrease the velocity so as to keep safe after the beginning of the task shown in Fig. 9(a). The result is in accordance with the existing research that driving speed displays a significant decrease trend when drivers execute secondary tasks [73]. The augmented amplitude of LPV while distraction in Fig. 9(c) reveals that the lateral distance of the vehicle from the road center line increases after distraction. Additionally, the standard deviations of V and LPV also rise as can be seen in Fig. 9(b) and Fig. 9(d). These findings are in line with current driving performance analysis in the literature. Choudhary *et al.* [74] studied the effects of phone use on driving performance and demonstrated that drivers have lower ability of longitudinal control and show higher variations in lane positioning when attention is shifted from driving to task related work. In our experiment, drivers consume more attention resources on the distraction task, the limbs control as well as the muscle tension are thereby weakened (see Fig. 8). Sequentially, diminished ability in lane and velocity keeping occurs. Therefore, the velocity declines and the amplitude of LPV increases after distraction. Moreover, greater standard deviations can be observed in V and LPV.

After analyzing the differences in features of collected signals between normal and distraction, ReliefF was applied to find the most effective feature and minish the feature matrix dimension for each physiological signal. The selected MSaE feature was input into a LSTM model to detect driver distraction and the result was compared with that of MSE and multiple features for each signal as listed in Table I. MSaE, with accuracies of 74.64%, 67.09 % and 64.81% for EEG, ECG and EMG separately, performs better than MSE in Table I. Both of them contribute more than the other features in detecting distraction as can be inferred from Fig. 10. It is due to the mitigated the impact of residual noise on exploring the complexity of physiological signals by multiple scale calculation, and because that MSaE remains unaffected by resampling unlike MSE [34], [35]. Besides, MSaE of EEG contributes a higher accuracy by contrast with that of ECG and EMG since EEG motivated by central nervous system can respond more promptly with higher resolution to changes in mental state than other signals [75]. Furthermore, detecting distraction with the selected MSaE feature outperforms that with multiple features of EEG or ECG, which demonstrates the necessity of feature selection. For the EMG signal, the accuracy of MSaE is comparable with that of multiple features. The results are consistent with previous studies in [76], [77]. Their research has clarified that high-dimensional feature set not always brings with high accuracy in classification task, and that it is necessary to filter out redundant features and select the proper relevant features for saving time and avoiding over-fitting. The performance of trained LSTM model is further enhanced with multi-modality physiological features to 78.32% and 74.16% for MSaE and MSE respectively in Table I. Moreover, the classification accuracy is improved again to 81.27% when adding multi-modality physiological features to vehicle features in Table II, which is about 9% higher than that of vehicle features, thus improving the accuracy. Coinciding with previous studies, different signals manifest human mental states from a variety of aspects, they can compensate for each other and provide more valuable and sufficient information of distraction thus promoting the prediction accuracy [15], [17].

The results in our present work indicate that the discriminating distraction information of physiological signals can be potentially exploited by MSaE and that multiple modalities features can enhance the performance of driver distraction detection. Despite the reliability and superiority of the proposed framework, limitation still exists in the current study. We explored the features of EEG with few electrodes, which to some degree ignores the spatial information of EEG and the interaction between electrodes from various perspectives.

VII. Conclusion

This paper proposed a driver distraction detection method based on MSaE of multi-modality physiological signals. It has validated that the MSaE of multiple physiological signals in distracted driving is apparently different from that in normal driving. Besides, MSaE is better for analyzing the complexity of physiological signals than other conventional features employed in previous research. What's more, the detection accuracy is further improved by adopting features of multi-modality signals compared to single modality feature. The results confirm that incorporate signals supply complementary useful information of driver distraction. When a driver is distracted, physiological and behavioral changes come into being. The MSaE value of EEG signal descends to a trough soon after the task onset. The heart rate tends to be less regularity and stable with the distraction task goes on. Moreover, the ability of muscle and vehicle control also recedes, which can be manifested by the lower complexity of EMG, decreased velocity, and rising variability of V and LPV.

In the future work, we will mine the distraction information of EEG on the basis of brain functional network to improve the ability of exploring the spatial dependencies between electrodes. In addition, singular spectrum analysis, an algorithm of decomposing signals into interpretable individual components, is planned to be adopted to remove artifacts, and then extracting features from the reserved components.

Acknowledgment

We gratefully acknowledge the financial support from the National Natural Science Foundation of China (grant number: 61703069 and 62001312), the Fundamental Research Funds for the Central Universities (grant number: DUT21GF301) and the Science and Technology Planning Project of Liaoning Province (grant number: 2021JH1/10400049).

References

[1] G. Rigas, Y. Goletsis, P. Bougia, and D. I. Fotiadis, "Towards driver's state recognition on real driving conditions," *Int. J. Veh. Technol.*, vol. 2011, Jul. 2011, Art. no. 617210.
[2] J.-S. Wang, C.-W. Lin, and Y.-T. C. Yang, "A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition," *Neurocomputing*, vol. 116, pp. 136–143, Oct. 2013.doi: https://doi.org/10.1016/j.neucom.2011.10.047
[3] E. Ohn-Bar, S. Martin, A. Tawari, and M. M. Trivedi, "Head, eye, and hand patterns for driver activity recognition," in *2014 22nd Int. Conf. Pattern Recognit., Stockholm, Sweden,* Aug. 2014, pp. 660–665.
[4] F. Yan, M. Liu, C. Ding, Y. Wang, and L. Yan, "Driving style recognition based on electroencephalography data from a simulated driving experiment," *Front. Psychol.*, vol. 10, May 2019, [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01254
[5] A. Aksjonov, P. Nedoma, V. Vodovozov, E. Petlenkov, and M. Herrmann, "A method of driver distraction evaluation using fuzzy logic: Phone usage as a driver's secondary activity: Case study," in *2017 XXVI Int. Conf. Inf., Commun. Autom. Technol. (ICAT)*, Sarajevo, Bosnia and Herzegovina, Oct. 2017, pp. 1–6.
[6] G. Tillman, D. Strayer, A. Eidels, and A. Heathcote, "Modeling cognitive load effects of conversation between a passenger and driver," *Atten. Percept. Psychophys.*, vol. 79, no. 6, pp. 1795–1803, May 2017, doi: 10.3758/s13414-017-1337-2.
[7] J. Healey and R. Picard, "SmartCar: detecting driver stress," in *Proc. 15th Int. Conf.Pattern Recognit., ICPR-2000*, Barcelona, Spain, Sep. 2000, pp. 218–221.
[8] B. Zhang, W. Chang, and X. Li, "Fatigue driving detection based on spatial-temporal electroencephalogram features and parallel neural networks," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 23, no. 2, p. 315, 2023.
[9] G. Li, W. Yan, S. Li, X. Qu, W. Chu, and D. Cao, "A temporal–spatial deep learning approach for driver distraction detection based on EEG signals," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 4, pp. 2665–2677, Oct. 2022, doi: 10.1109/TASE.2021.3088897.
[10] R. N. Khushaba, S. Kodagoda, D. Liu, and G. Dissanayake, "Muscle computer interfaces for driver distraction reduction," *Comput. Meth. Programs Biomed.*, vol. 110, no. 2, pp. 137–149, May 2013, doi: https://doi.org/10.1016/j.cmpb.2012.11.002.
[11] M. Taherisadr, P. Asnani, S. Galster, and O. Dehzangi, "ECG-based driver inattention identification during naturalistic driving using mel-frequency cepstrum 2-D transform and convolutional neural networks," *Smart Health*, vol. 9–10, Jul. 2018, doi: 10.1016/j.smhl.2018.07.022.
[12] K. Ryu and R. Myung, "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic," *Int. J. Ind. Ergon.*, vol. 35, no. 11, pp. 991–1009, Nov. 2005, doi: https://doi.org/10.1016/j.ergon.2005.04.005.
[13] B. Mehler, B. Reimer, J. F. Coughlin, and J. A. Dusek, "Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers," *Transp. Res. Rec.*, vol. 2138, no. 1, pp. 6–12, Jan. 2009, doi: https://doi.org/10.3141/2138-0.
[14] D. He, B. Donmez, C. C. Liu, and K. N. Plataniotis, "High cognitive load assessment in drivers through wireless electroencephalography and the validation of a modified N-back task," *IEEE Trans. Hum. Mach. Syst.*, vol. 49, no. 4, pp. 362–371, Jun. 2019, doi: 10.1109/THMS.2019.2917194.
[15] Y. Du, A. Black, L.-P. Morency, and M. Eskenazi, "*Multimodal polynomial fusion for detecting driver distraction*," in Interspeech 2018, Hyderabad,India, Sep. 2018, pp. 611-615.
[16] G. Lechner *et al.*, "A lightweight framework for multi-device integration and multi-sensor fusion to explore driver distraction," in *Adv. Inf. Syst. Eng.*, May 2019, pp. 80–95.

[17]Y. Zhang, Y. Chen, and C. Gao, "Deep unsupervised multi-modal fusion network for detecting driver distraction," *Neurocomputing*, vol. 421, pp. 26–38, Jan. 2021, doi: https://doi.org/10.1016/j.neucom.2020.09.023.
[18]J. Urigüen and B. Zapirain, "EEG artifact removal – State-of-the-art and guidelines," *J. Neural Eng.*, vol. 12, p. 031001, Apr. 2015, doi: 10.1088/1741-2560/12/3/031001.
[19]Z. Zhang *et al.*, "Efficient sleep classification based on entropy features and support vector machine classifier," *Physiol. Meas.*, vol. 39, Oct. 2018, doi: 10.1088/1361-6579/aae943.
[20]W. Zhang and H. Zhang, "Research on distracted driving identification of truck drivers based on simulated driving experiment," *IOP Conf. Ser. Earth Environ. Sci.*, Feb. 2021, pp. 012039.
[21]G. Kountouriotis, P. Spyridakos, O. Carsten, and N. Merat, "Identifying cognitive distraction using steering wheel reversal rates," *Accid. Anal. Prev.*, vol. 96, Jul. 2016, doi: 10.1016/j.aap.2016.07.032.
[22]Y. Liang, J. D. Lee, and M. L. Reyes, "Nonintrusive detection of driver cognitive distraction in real time using bayesian networks," *Transp. Res. Rec.*, vol. 2018, no. 1, pp. 1–8, Jan. 2007, doi: 10.3141/2018-01.
[23]K. Kircher and C. Ahlström, "Evaluation of methods for the assessment of minimum required attention," *Accid. Anal. Prev.*, vol. 114, pp. 40-47, May 2018, doi: 10.1016/j.aap.2017.03.013.
[24]V. Alizadeh and O. Dehzangi, "The impact of secondary tasks on drivers during naturalistic driving: Analysis of EEG dynamics," in *2016 IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Rio de Janeiro, Brazil, Dec. 2016, pp. 2493–2499.
[25]S. V Deshmukh and O. Dehzangi, "ECG-based driver distraction identification using wavelet packet transform and discriminative kernel-based features," in *2017 IEEE Int. Conf.Smart Comput. (SMARTCOMP)*, Hong Kong, China, May. 2017, pp. 1–7.
[26]A. Sahayadhas, K. Sundaraj, M. Murugappan, and R. Palaniappan, "A physiological measures-based method for detecting inattention in drivers using machine learning approach," *Biocybern. Biomed. Eng.*, vol. 35, no. 3, pp. 198–205, 2015, doi: https://doi.org/10.1016/j.bbe.2014.12.002.
[27]E. A. Mattar, H. J. Al-Junaid, and K. N. Al-Mutib, "Electroencephalography features extraction and deep patterns analysis for robotics learning and control through brain-computer interface," in *2019 Int. Conf.Innov. Intell. Inform., Comput., and Technol. (3ICT)*, Sakhier, Bahrain, Nov. 2019, pp. 1–6.
[28]Z. Gao *et al.*, "An adaptive optimal-Kernel time-frequency representation-based complex network method for characterizing fatigued behavior using the SSVEP-based BCI system," *Knowledge-Based Syst.*, vol. 152, pp. 163–171, Jul. 2018, doi: https://doi.org/10.1016/j.knosys.2018.04.013.
[29]W. Lv *et al.*, "Applying Entropy to Understand Drivers' Uncertainty during Car-following.," in *Proc. the joint meeting 12th Int. Conf. Meas. Behav.6th Semin. Behav. Meth.*, Krakow, Poland, Oct. 2021.
[30]M. Ravi Kumar and Y. Srinivasa Rao, "Epileptic seizures classification in EEG signal based on semantic features and variational mode decomposition," *Cluster Comput.*, vol. 22, no. 6, pp. 13521–13531, Feb. 2019, doi: 10.1007/s10586-018-1995-4.
[31]M. Fraiwan, M. Alafeef, and F. Almomani, "Gauging human visual interest using multiscale entropy analysis of EEG signals," *J. Ambient. Intell. Humaniz. Comput.*, vol. 12, no. 2, pp. 2435–2447, 2021, doi: 10.1007/s12652-020-02381-5.
[32]P. Raffalt, J. Mccamley, W. Denton, and J. Yentes, "Sampling frequency influences sample entropy of kinematics during walking," *Med. Biol. Eng. Comput.*, vol. 57, Nov. 2018, doi: 10.1007/s11517-018-1920-2.
[33]F. Fallahtafti, S. R. Wurdeman, and J. M. Yentes, "Sampling rate influences the regularity analysis of temporal domain measures of walking more than spatial domain measures," *Gait Posture*, vol. 88, pp. 216–220, Jul. 2021, doi: https://doi.org/10.1016/j.gaitpost.2021.05.031.
[34]J. Gao, J. Hu, F. Liu, and Y. Cao, "Multiscale entropy analysis of biological signals: a fundamental bi-scaling law," *Front. Comput. Neurosci.*, vol. 9, 2015, [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncom.2015.00064
[35]J. Zheng *et al.*, "Effects of sampling rate on multiscale entropy of electroencephalogram time series," *Biocybern. Biomed. Eng.*, vol. 43, no. 1, pp. 233–245, 2023, doi: https://doi.org/10.1016/j.bbe.2022.12.007.
[36]K. M. Hossain, Md. A. Islam, S. Hossain, A. Nijholt, and M. A. R. Ahad, "Status of deep learning for EEG-based brain–computer interface applications," *Front. Comput. Neurosci.*, vol. 16, Jan. 2023, [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncom.2022.1006763
[37]H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Min. Knowl. Discov.*, vol. 33, no. 4, pp. 917–963, Mar. 2019, doi: 10.1007/s10618-019-00619-1.
[38]A. Ak, V. Topuz, and I. Midi, "Motor imagery EEG signal classification using image processing technique over GoogLeNet deep learning algorithm for controlling the robot manipulator," *Biomed. Signal Process. Control*, vol. 72, p. 103295, Feb. 2022, doi: https://doi.org/10.1016/j.bspc.2021.103295.
[39]H. Ranganathan, S. Chakraborty, and S. Panchanathan, "Multimodal emotion recognition using deep learning architectures," in *2016 IEEE Winter Conf. Appl. Comput. Vis.(WACV)*, Lake Placid, NY, USA, Mar. 2016, pp. 1–9.
[40]S. M. Kouchak and A. Gaffar, "Detecting driver behavior using stacked long short term memory network with attention layer," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3420–3429, Jun. 2021, doi: 10.1109/TITS.2020.2986697.
[41]J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," arXiv preprint arXiv:1412.3555, Dec. 2014.
[42]N. Phutela, D. Relan, G. Gabrani, P. Kumaraguru, and M. Samuel, "Stress classification using brain signals based on LSTM network," *Comput. Intell. Neurosci.*, vol. 2022, p. 7607592, Apr. 2022, doi: 10.1155/2022/7607592.
[43]G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of hand movements from EEG using a deep attention-based LSTM network," *IEEE Sens. J.*, vol. PP, p. 1, Dec. 2019, doi: 10.1109/JSEN.2019.2956998.
[44]X. Hong *et al.*, "Predicting Alzheimer's disease using LSTM," *IEEE Access*, vol. 7, pp. 80893–80901, May. 2019, doi: 10.1109/ACCESS.2019.2919385.
[45]H. Sun *et al.*, "Sleep staging from electrocardiography and respiration with deep learning," *Sleep*, vol. 43, no. 7, p. zsz306, Jul. 2020, doi: 10.1093/sleep/zsz306.
[46]J. Murphy, C. Devue, P. M. Corballis, and G. M. Grimshaw, "Proactive control of emotional distraction: evidence ffrom EEG alpha suppression," *Front. Hum. Neurosci.*, vol. 14, Aug. 2020, [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00318
[47]M. Rhif, A. Ben Abbes, I. R. Farah, B. Martínez, and Y. Sang, "Wavelet transform application for/in non-stationary time-series analysis: A review," *Applied Sciences*, vol. 9, no. 7, pp. 1345, Mar. 2019.
[48]M. Costa, C.-K. Peng, A. L. Goldberger, and J. M. Hausdorff, "Multiscale entropy analysis of human gait dynamics," *Physica A: Stat.Mech. Appl.*, vol. 330, no. 1, pp. 53–60, Dec. 2003, doi: https://doi.org/10.1016/j.physa.2003.08.022.
[49]Y.-H. Wang, I.-Y. Chen, H. Chiueh, and S.-F. Liang, "A low-cost implementation of sample entropy in wearable embedded systems: an example of online analysis for sleep EEG," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, Jan. 2021, doi: 10.1109/TIM.2020.3047488.
[50]M. Borowiec, A. Rysak, D. N. Betts, C. R. Bowen, H. A. Kim, and G. Litak, "Complex response of a bistable laminated plate: Multiscale entropy analysis," *Eur. Phys. J. Plus*, vol. 129, no. 10, p. 211, Oct. 2014, doi: 10.1140/epjp/i2014-14211-3.
[51]M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E.*, vol. 71, no. 2, p. 021906, 2005.
[52]S. Kar, M. Bhagat, and A. Routray, "EEG signal analysis for the assessment and quantification of driver's fatigue," *Transp. Res. Pt. F-Traffic Psychol. Behav.*, vol. 13, no. 5, pp. 297–306, Sep. 2010, doi: https://doi.org/10.1016/j.trf.2010.06.006.
[53]Y. Bai, Y. Guan, and W.-F. Ng, "*Fatigue assessment using ECG and actigraphy sensors*," in *Proc.2020 ACM Int. Symp. Wearable Comput.*, Virtual Event, Mexico, Sep. 2020, pp. 12-16. doi: https://doi.org/10.1145/3410531.3414308.
[54]S. Huang, J. Li, P. Zhang, and W. Zhang, "Detection of mental fatigue state with wearable ECG devices," *Int. J. Med. Inform.*, vol. 119, pp. 39–46, Nov. 2018, doi: https://doi.org/10.1016/j.ijmedinf.2018.08.010.
[55]C. Zhao, M. Zhao, J. Liu, and C. Zheng, "Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator," *Accid. Anal. Prev.*, vol. 45, pp. 83–90, 2012, doi: https://doi.org/10.1016/j.aap.2011.11.019.
[56]J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, 1985, doi: 10.1109/TBME.1985.325532.

[57]A. Noviyanto, S. M. Isa, I. Wasito, and A. Arymurthy, “Selecting features of single lead ECG signal for automatic sleep stages classification using correlation-based feature subset selection,” *Int. J. Comput.Sci. Iss. (IJCSI)*, vol. 8, Sep. 2011.
[58]C. Sapsanis, G. Georgoulas, and A. Tzes, “EMG based classification of basic hand movements based on time-frequency features,” in *21st Mediterr. Conf. Control Autom.*, Platanias, Greece, Jun. 2013, pp. 716–722.
[59]S. Pourmohammadi and A. Maleki, “Stress detection using ECG and EMG signals: A comprehensive study,” *Comput. Meth. Programs Biomed.*, vol. 193, p. 105482, Sep. 2020, doi: https://doi.org/10.1016/j.cmpb.2020.105482.
[60]K. Mahaphonchaikul, D. Sueaseenak, C. Pintavirooj, M. Sangworasil, and S. Tungjitkusolmun, “EMG signal feature extraction based on wavelet transform,” in *ECTI-CON2010: The 2010 ECTI Int. Conf. Electr. Eng./Electronics, Comput., Telecommun. Inf. Technol.*, Chiang Mai, Thailand, May. 2010, pp. 327–331.
[61]Y. Zhang, X. Ren, and J. Zhang, “Intrusion detection method based on information gain and ReliefF feature selection,” in *2019 Int. Joint Conf. Neural Netw. (IJCNN)*, Budapest, Hungary, Jul. 2019, pp. 1–5.
[62]C. Peng, Z. Zhang, Z. Kang, C. Chen, and Q. Cheng, “Nonnegative matrix factorization with local similarity learning,” *Inf. Sci.*, vol. 562, pp. 325–346, Jul. 2021, doi: https://doi.org/10.1016/j.ins.2021.01.087.
[63]G. Doquire and M. Verleysen, “Mutual information-based feature selection for multilabel classification,” *Neurocomputing*, vol. 122, pp. 148–155, Dec. 2013, doi: https://doi.org/10.1016/j.neucom.2013.06.035.
[64]N. S. Malan and S. Sharma, “Motor imagery EEG spectral-spatial feature optimization using dual-tree complex wavelet and neighbourhood component analysis,” *IRBM*, vol. 43, no. 3, pp. 198–209, Jun. 2022, doi: https://doi.org/10.1016/j.irbm.2021.01.002.
[65]A. Marcano-Cedeño, J. Quintanilla-Domínguez, M. G. Cortina-Januchs, and D. Andina, “Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network,” in *IECON 2010 - 36th Annu. Conf. IEEE Ind. Electron. Soc.*, Glendale, AZ, USA, Nov. 2010, pp. 2845–2850.
[66]P. A. Karthick, D. M. Ghosh, and S. Ramakrishnan, “Surface electromyography based muscle fatigue detection using high-resolution time-frequency methods and machine learning algorithms,” *Comput. Meth.s Programs Biomed.*, vol. 154, pp. 45–56, Feb. 2018, doi: https://doi.org/10.1016/j.cmpb.2017.10.024.
[67]H. V. Koay, J. H. Chuah, C.-O. Chow, and Y.-L. Chang, “Detecting and recognizing driver distraction through various data modality using machine learning: A review, recent advances, simplified framework and open challenges (2014–2021),” *Eng. Appl. Artif. Intell.*, vol. 115, p. 105309, 2022, doi: https://doi.org/10.1016/j.engappai.2022.105309.
[68]Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, “Driver inattention monitoring system for intelligent vehicles: a review,” *IEEE Trans. Intel. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011, doi: 10.1109/TITS.2010.2092770.
[69]S. V Deshmukh and O. Dehzangi, “Characterization and identification of driver distraction during naturalistic driving: an analysis of ECG dynamics,” in *Adv. Body Area Netw.* pp. 1–13, 2019, doi: https://doi.org/10.1007/978-3-030-02819-0_1.
[70]L. Yu, X. Sun, and K. Zhang, “Driving distraction analysis by ECG signals: an entropy analysis,” in *Int. Conf. Inter. Des. Glob. Dev.*, Berlin, Heidelberg, Jul. 2011, pp. 258–264.
[71]S. V. Deshmukh, “Study of online driver distraction analysis using ECG-dynamics,” Ph.D. dissertation, Dept. Comput. Inf. Sci., Univ. Michigan-Dearborn, Dearborn, USA, 2018.
[72]M. B. Kurt, N. Sezgin, M. Akin, G. Kirbas, and M. Bayram, “The ANN-based computing of drowsy level,” *Expert Syst. Appl.*, vol. 36, no. 2, Part 1, pp. 2534–2542, Mar. 2009, doi: https://doi.org/10.1016/j.eswa.2008.01.085.
[73]B. Reimer, B. Mehler, Y. Wang, and J. Coughlin, “A field study on the impact of variations in short-term memory demands on drivers’ visual attention and driving performance across three age groups,” *Hum. Factors*, vol. 54, pp. 454–468, May. 2012, doi: 10.1177/0018720812437274.
[74]P. Choudhary and N. R. Velaga, “Effects of phone use on driving performance: A comparative analysis of young and professional drivers,” *Saf. Sci.*, vol. 111, pp. 179–187, Jan. 2019, doi: https://doi.org/10.1016/j.ssci.2018.07.009.
[75]J. Zhang, Z. Yin, P. Chen, and S. Nichele, “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review,” *Inf. Fusion*, vol. 59, pp. 103–126, Jul. 2020, doi: https://doi.org/10.1016/j.inffus.2020.01.011.
[76]A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, “Benchmark for filter methods for feature selection in high-dimensional classification data,” *Comput. Stat. Data Anal.*, vol. 143, p. 106839, Mar. 2020, doi: https://doi.org/10.1016/j.csda.2019.106839.
[77]M. Ghosh, S. Adhikary, K. K. Ghosh, A. Sardar, S. Begum, and R. Sarkar, “Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods,” *Med. Biol. Eng. Comput.*, vol. 57, no. 1, pp. 159–176, 2019, doi: 10.1007/s11517-018-1874-4.

Xin Zuo received the B.S. in Naval Architecture and Ocean Engineering from Harbin Engineering University, China, in 2016, and the M.S. in Design and Manufacture of Ship and Ocean structure from Dalian University of Technology, in 2019. She is currently pursuing the Ph.D. degree in Software and Communications Engineering at University of Jyväskylä, Jyväskylä, Finland.

Her research interests include driver distraction, biomedical signal processing.

Chi Zhang received the B.S., M.S., and Ph.D. degree of Science in Engineering from Northeastern University, China, in 2010, 2012 and 2016, respectively.

He is an Associate Professor with the School of Biomedical Engineering, Faculty of Medicine, Dalian University of Technology, Dalian, China. His research interests include biomedical signal processing, brain-computer interface, and artificial intelligence.

Fengyu Cong received the B.S. degree in Power and Thermal Dynamic Engineering and the Ph.D. degree in Mechanical Design and Theory from Shanghai Jiao Tong University, China, in 2002 and 2007, and the Ph.D. degree in Mathematical Information Technology from the University of Jyväskylä, Jyväskylä, Finland, in 2010.

He is a Professor in the School of Biomedical Engineering, Faculty of Medicine, Dalian University of Technology, Dalian, China. His current research interests include brain signal processing, acoustic signal processing, independent component analysis, tensor decomposition, and pattern recognition/machine learning/data mining.

Jian Zhao received the M.S. and Ph.D. degree in Mechanical Engineering from Xidian University, China, in 2006 and 2008, respectively.

He is currently a Professor at the School of Automotive Engineering, Faculty of Vehicle Engineering and Mechanics, Dalian University of Technology, Dalian, China. His current research interests include autonomous driving, MEMS sensors, compliant mechanics, and nonlinear dynamics.

Timo Hämäläinen received the Ph.D. degree in telecommunication from the University of Jyväskylä, Jyväskylä, Finland, in 2002.

In 1997, he joined the University of Jyväskylä, where he is currently a Professor of computer networks. He has more than 25 years research and teaching experience of computer networks. He has led many external funded network management related projects. He has launched and leads master programs with the University of Jyväskylä (SW & Comm. Eng.), and teaches network management related courses. He has more than 200 internationally peer reviewed publications and he has supervised almost 40 Ph.D. dissertation. His research interests include network resource management, IoT, and networking security.

IV

CROSS-SUBJECT EMOTION RECOGNITION USING FUSED ENTROPY FEATURES OF EEG

by

Xin Zuo, Chi Zhang, Timo Hämäläinen, Hanbing Gao, Yu Fu,
and Fengyu Cong 2022

Entropy, 24(9), 1281

https://doi.org/10.3390/e24091281

entropy

Article

Cross-Subject Emotion Recognition Using Fused Entropy Features of EEG

Xin Zuo [1,2], Chi Zhang [1,3,*], Timo Hämäläinen [2], Hanbing Gao [1], Yu Fu [1] and Fengyu Cong [1,2]

1 School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China
2 Faculty of Information Technology, University of Jyväskylä, 40014 Jyväskylä, Finland
3 Liaoning Key Laboratory of Integrated Circuit and Biomedical Electronic System, Dalian 116024, China
* Correspondence: chizhang@dlut.edu.cn

Abstract: Emotion recognition based on electroencephalography (EEG) has attracted high interest in fields such as health care, user experience evaluation, and human–computer interaction (HCI), as it plays an important role in human daily life. Although various approaches have been proposed to detect emotion states in previous studies, there is still a need to further study the dynamic changes of EEG in different emotions to detect emotion states accurately. Entropy-based features have been proved to be effective in mining the complexity information in EEG in many areas. However, different entropy features vary in revealing the implicit information of EEG. To improve system reliability, in this paper, we propose a framework for EEG-based cross-subject emotion recognition using fused entropy features and a Bidirectional Long Short-term Memory (BiLSTM) network. Features including approximate entropy (AE), fuzzy entropy (FE), Rényi entropy (RE), differential entropy (DE), and multi-scale entropy (MSE) are first calculated to study dynamic emotional information. Then, we train a BiLSTM classifier with the inputs of entropy features to identify different emotions. Our results show that MSE of EEG is more efficient than other single-entropy features in recognizing emotions. The performance of BiLSTM is further improved with an accuracy of 70.05% using fused entropy features compared with that of single-type feature.

Keywords: emotion recognition; EEG; feature fusion; MSE; BiLSTM

check for updates

Citation: Zuo, X.; Zhang, C.; Hämäläinen, T.; Gao, H.; Fu, Y.; Cong, F. Cross-Subject Emotion Recognition Using Fused Entropy Features of EEG. *Entropy* **2022**, *24*, 1281. https://doi.org/10.3390/e24091281

Academic Editors: Raúl Alcaraz, Luca Faes, Leandro Pardo and Boris Ryabko

Received: 20 June 2022
Accepted: 5 September 2022
Published: 11 September 2022

1. Introduction

Emotion is a specific psychological and physiological response generated by perceiving external and inner stimuli. It is a complex state combining thoughts, feelings, and behaviors and is an important part of daily human life [1]. Previous studies have demonstrated that emotion plays a vital role not only in the process of perception, decision making, and communication but also in the learning and memory process [2]. As a result, the measurement and characterization of different emotion states are of great importance to emotion recognition-related studies both theoretically and practically. For example, emotion recognition can be widely used in areas such as health care, distance learning, and user experience evaluation of products, which are closely related to humans [3]. Furthermore, it contributes to the computer ability of emotion recognition and expression in the human–computer interaction (HCI) field [4]. As emotion is often accompanied by high cognitive activities of the brain involving complex psychology and physiology processes [5], further study on how to recognize different emotions accurately is necessary.

There are many kinds of approaches to recognizing emotion states in existing studies. According to the data used in emotion recognition, they can be roughly divided into two categories. One category is based on non-physiological signals, whereas the other is based on physiological signals. Conventional emotion recognition methods based on non-physiological signals usually use facial expressions, behaviors, and voice-based signals, etc.. The features of these signals are more obvious for observation and easier to be extracted. Jain et al. proposed deep convolutional neural networks for observing emotion states based on different facial motions in different image emotions [6]. Meng et al. developed a speech emotion recognition method using spectrum features of speech signals [7]. There is also emotion recognition research combing different types of non-physiological signals. For instance, Kessous et al. studied a multimodal automatic emotion recognition method using the Bayesian classifier based on a mixture of facial expressions, gestures, and acoustic signals, and they found that fusing the multimodal signals would largely increase classification accuracy compared with unimodal systems [8]. Although the data collection process of these methods is easier, their availability and reliability could not be completely guaranteed, as they are mainly affected by two factors [9]. On the one hand, effective non-physiological signals are hard to obtain from participants who have trouble expressing their feelings through body language. On the other hand, participants can deliberately control their expressions, tone, and postures to hide their real feelings. Contrary to the non-physiological measurements, physiological signals are more reliable and effective, as these signals originate from spontaneous activities of the nervous system which cannot be controlled intentionally [1]. The mostly frequently used physiological signals include autonomic nervous system (ANS) signals such as the electrocardiogram (ECG), the electromyogram (EMG), skin resistance, and blood pressure, and neutral nervous system signals such as EEG, functional magnetic resonance imaging (fMRI), and so on. Kim et al. analyzed the multimodal autonomic physiological signals (i.e., ECG, EMG, respiration, and skin conductivity) induced by music and developed a scheme of emotion-specific classification [10]. The brain signals obtained directly from the neutral nervous system can reflect the dynamic neuro-electrical changes in real-time with high resolution, compared to ANS signals which often include a time delay [9]. In addition, the activity of EEG signals varies in different brain regions while emotional processes occur. Particularly, the lateral temporal brain areas are more active than other areas, and the energy of EEG increases for positive emotion, whereas lower energy appears in neutral and negative emotions [11]. Therefore, the emotional changes in different emotion states can be measured by EEG signals in the lateral temporal region. What is more, the equipment of EEG collection is small in size, portable, and much cheaper than that of fMRI. EEG-based emotion recognition has become one of the most prosperous research fields.

To recognize emotion states accurately based on EEG signals, features revealing the dynamic changes of EEG under different emotions should first be extracted. There are four main types of features used in EEG-based emotion recognition [12]: time domain features (e.g., statistical features and auto-regression coefficient), frequency domain features (e.g., power spectral density and energy spectrum), time-frequency features (e.g., wavelet coefficients), and non-linear dynamic features (e.g., fractal dimension and entropy features). The ability of different features varies in reflecting emotion states. Energy-based features in different brain regions have been commonly adopted in emotion recognition. As different brain regions are activated in different emotions, the energy of different frequency bands in these brain regions can be used for emotion recognition [13]. Du et al. selected sound clips of three affective states (i.e., happy (high arousal), afraid (high arousal), and neutral (low arousal)) to explore frontal asymmetry [14]. Their research demonstrated that the right frontal region is more related to high-arousal emotions (i.e., happiness and fear), whereas the left frontal region correlates with low-arousal emotions (i.e., neutral); thus, the energy asymmetry between the left and right brain can be used to classify emotion states. Further, Liu et al. found that there is a correlation between the emotional states and EEG frequency bands and that high-frequency bands contain more emotional information than low-frequency bands [15]. EEG signals, which are a direct reflection of brain activities, are non-stationary signals with a low signal-to-noise ratio, and the activation of EEG and the information it contains varies in different emotions. It is difficult to analyze EEG signals using only traditional time- or frequency-domain features. In recent years, entropy-based features have been proven to manifest more complex dynamic information in EEG than conventional features, leading to a wide use in many fields [16]. Wang et al. extracted the sample entropy (SE) feature of overnight sleeping EEG data utilizing the assisted sliding box algorithm to show the dynamic changes and a reduction of computation time [17]. Chen et al. proposed a method using the approximate entropy (AE) feature and its transformation to identify four human emotions based on EEG with an accuracy of 83.34% [18]. Zheng et al. trained a Deep Belief Network (DBN) to classify three emotions with the input of the differential entropy (DE) feature [19]. Their results showed that the DE feature of certain brain regions could reflect the dynamic changes of EEG in different emotions and can be used for recognizing emotion states. What is more, entropy features deployed to analyze brain states in other areas may also contribute to emotion recognition. For instance, multiscale entropy (MSE), calculating entropy in multiple time scales, has been proven to achieve better robustness of results than conventional features in fields of disease diagnosis and sleep studies [20]. Hadoush et al. adopted MSE to explore patterns in children with mild and severe autism spectrum disorders (ASD) and found that MSE could serve as an effective index for the severity of ASD [21]. Vladimir et al. explored the changes in brain signal complexity across several distinct global states of consciousness using MSE [22]. The results indicated that MSE changes throughout the sleep cycle and is strongly time-scale dependent, which makes it possible to use MSE for sleep staging. However, a challenge still exists in analyzing EEG signals based on entropy features. These different features characterize the EEG implicit complexity information to varying degrees, but it remains unclear which type of entropy feature is more effective for describing emotional states. Further, previous studies demonstrated that any feature could add complementary information to the other features [23]. Hence, there is a necessity to integrate the advantages of different entropy features to enhance the performance of emotion classification.

To take advantage of the EEG features, researchers have trained a variety of classifiers to recognize different types of emotion states. Traditional classifiers such as SVM, K-Nearest Neighbor (KNN), and transfer learning are widely used for emotion classification. Liu et al. established a real-time EEG-based emotion recognition system using SVM that could successfully classify positive and negative emotions with acceptable results [24]. Kolodyazhniy et al. extracted features from physiological signals induced by different emotional film clips and proposed an affective computing approach based on the KNN classifier [25]. Lan et al. [26] utilized the DE feature of EEG and the transfer learning technique to detect three emotions reaching an accuracy of 72.47% in the SEED dataset. Although traditional classifiers have achieved different recognition performance in simple tasks (e.g., 86.43% accuracy for three positive emotions in [24], 83.34% accuracy for four different emotions in [18], 77.5% accuracy via different types of signals in [25]), they are not efficient enough to learn the contextual dependency in a time series and do not perform well in cross-subject emotion recognition [27]. As we all know, human emotions are a continuous time series, and the current emotion state is influenced by both the current stimulus and previous emotions. In this case, it is difficult for traditional classifiers to recognize human emotion only based on the current feature. The Bidirectional Long Short-term Memory (BiLSTM) network has the ability to learn long- and short-term dependency between time steps and to memorize both forward and backward contextual information in a time series, compared with the one-directional Long Short-term Memory (LSTM) network [28] that is widely used in speech synthesis, pathological voice detection, and motion prediction [29–31]. It has been proven to perform effectively in pattern recognition and has been successfully deployed in sequence-to-sequence classification tasks in many fields [32–35]. Mahmud et al. trained an automated BiLSTM model to detect sleep apnea based on EEG and reached a high accuracy on different publicly available datasets [36]. Chang et al. proposed a depression assessment framework based on the spatiotemporal network of EEG and BiLSTM and achieved more than 70% accuracy in the SEED dataset [37].

Two emotional models widely used in the existing research are the circumplex model of affects (CMA) and the discrete emotion model (DEM). CMA defines emotions in a two-dimensional space with arousal and valence. In the work of Posner et al. about CMA, the authors presented that emotion states occur from the cognitive interpretations of core neural sensations and that CMA is a useful tool to study the development of emotion disorders and cognitive underpinnings of affective processing [38]. DEM conversely supposes each discrete emotional state is a different state. In the study of Kılıç et al., the authors proposed that assigning each emotion as a separate discrete state based on DEM is important in recognizing different emotional states and particularly in neuropsychiatric diseases [39].

Motivated by the fact that discrete emotions are vital for emotion recognition and that different entropy features represent implicit EEG complexity in different degrees, we focus on finding out which entropy feature is the best for characterizing three discrete emotional states (i.e., positive, neutral, and negative) and whether integrating different entropy features can enhance the performance of emotion classification by utilizing BiLSTM in the present study. In this paper, we propose a novel framework for cross-subject emotion recognition based on fused entropy features of EEG and BiLSTM. Our approach is to model a BiLSTM classifier based on the fusion of entropy features in EEG induced by different emotional film clips. We first calculate the MSE feature and four other entropy features of EEG to mine the dynamic changes of EEG in different emotion statuses. Then, a BiLSTM classifier is trained to learn the bidirectional time dependency in the extracted EEG features and to realize emotion recognition.

This paper is organized as follows. Section 2 addresses the EEG dataset used in our work. Section 3 details the adopted methodologies. The results of the research are presented in Section 4 and discussed in Section 5. Finally, Section 6 concludes the paper.

2. Data Resource

The EEG data in this study came from the 2020 World Robot Competition—BCI Control Brain Robot Contest. It consisted of two public datasets, namely SEED [11] and SEED-FRA [40] (the SJTU Emotion EEG Dataset, https://bcmi.sjtu.edu.cn/home/seed/, accessed on 30 October 2014), collections for various emotion research purposes using EEG provided by Shanghai Jiao Tong University. Prior to the data collection, the experiment was approved by the Ethics Committee, Shanghai Jiao Tong University. The data were gathered from 23 healthy subjects (15 Chinese and 8 French) while they watched different emotional film clips in their native language. First, 50 cinema managers were asked to fill in a questionnaire in which they were supposed to describe the emotional valence of at least three film excerpts for each emotion state (i.e., positive, neutral, and negative). The cinema managers were selected because they were likely to have significant knowledge about films, which might contribute to creating a large preliminary list of film scenes [41]. Then, the listed emotional film excerpts were discussed and viewed by the cinema managers to rate their valence scale from 1 (sad) to 9 (happy) using the Self-Assessment Manikin (SAM). The mean and standard values of each film excerpt were calculated to analyze the rating results, and the initial pool of film clips was established. After this step, a pilot trial was executed to test whether the selected film clips could elicit the expected emotions. According to the SAM rating results of subjects, the mean and standard values of each film clip in the pilot trial could be obtained. Emotional film excerpts from five Chinese films and seven French films with the largest mean values and similar standard values were finally selected as the positive stimuli. There were also twelve film excerpts with the smallest mean values and approximative standard values chosen to be negative stimuli. As for the neutral stimuli, they consisted of film excerpts whose mean values were close to five (the mean value of the valence scale), and standard values were similar. Before the experiment, subjects were asked to finish the Eysenck Personality Questionnaire (EPQ), and only those with stable moods were selected. There are three types of emotions (i.e., positive, neutral, and negative) included in the experiment, and each type of emotion had five (for Chinese subjects) or seven (for French subjects) corresponding film clips. Each emotional film clip lasted for 2 min.

The experiment was performed in a quiet room. Figure 1 shows the experiment scene. A 62-channel electrode cap arranged according to the international 10–20 system was used to collect the EEG data. The sampling rate was set to 1000 Hz. Before starting, all subjects were given written and oral instructions on the experiment and were asked to stay as still as possible and refrain from moving. In the experiment, the subjects sat comfortably and paid attention to watching the forthcoming film clips. Eight of the subjects watched 21 film clips (i.e., 21 trials), with seven film clips corresponding to each emotion. The other fifteen subjects were shown 15 film clips, and there were five corresponding film clips for each emotion. The detailed protocol of the experiment is shown in Figure 2. A 5 s picture hint was set before each clip, and there was a 45 s interval after each clip, allowing the subjects to report their emotional states concerning the film clips based on their feelings. The self-reported emotional states were then used to validate the emotion classification results of the study. The details about the database can be found in [11,37,40].

Figure 1. The emotion experiment scene.

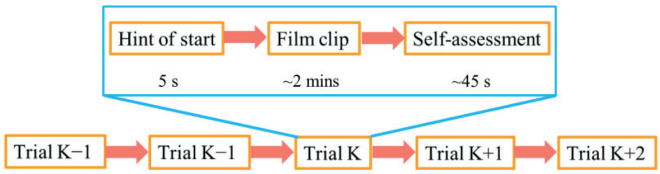


Figure 2. Protocol of the experiment.

3. Methodology

The analysis process of EEG-based emotion recognition includes three steps: data preprocessing, feature extraction, and emotion recognition. This section describes how we dealt with the data in detail. The analyzing process was implemented in MATLAB 2019b.

3.1. Preprocessing

As previous studies have demonstrated that some electrodes are irrelevant to emotion changes [42], and Zheng et al. found that the lateral temporal brain area is activated more than other brain areas in emotion processing [11], we first selected twelve electrodes (i.e., FT7, T7, TP7, P7, C5, CP5, FT8, T8, TP8, P8, C6, CP6) in the lateral temporal brain area for further research in this paper. The EEG data were then down-sampled to 256 Hz to improve the calculation efficiency. After that, the EEG segments corresponding to each film clip's duration were extracted to obtain the entire EEG data from watching all the film clips, as the raw EEG data contained the EEG signals not only while watching the films but also in the preparation and self-assessment stages. To reject interference from the power line, we used a bandstop filter of 50 Hz. Five frequency bands (i.e., delta, theta, alpha, beta, and gamma) of the EEG signals were then roughly extracted, applying wavelet decomposition. Finally, a wavelet-based technique was used to remove the artifacts in each band.

Wavelet transform, which is an effective time-frequency analysis method with the ability for good local representation of signals in the time and frequency domain, is usually used to analyze EEG signals [43]. By decomposing the signal at each level, the detailed and approximate component wavelet coefficients can be obtained corresponding to the level. The wavelet coefficients could reflect the detailed information of the signal as well as the correlation with the mother wavelet. In fact, the coefficients of the artifacts are usually larger than those of a normal EEG signal. Therefore, artifacts were eliminated by setting a threshold value [44]. This wavelet-based method has been validated to be effective in the field of driver fatigue assessment [45,46]. We can calculate the threshold by

$$T_j = mean\left(C_j\right) + 2 \times std\left(C_j\right) \tag{1}$$

where Cj is the wavelet coefficient at the jth level of wavelet decomposition. The value of any coefficient is larger than Tj; it is considered a coefficient of the artifact and halved to eliminate its influence. Then, the wavelet-corrected signal can be reconstructed with the new set of wavelet coefficients. More details about the preprocessing process can be found in Appendix A.

As usually used to resemble EEG signals in the literature [47], db6 was selected as the mother wavelet. The EEG signals from all the twelve electrodes were preprocessed in the same way as mentioned above in this paper. Figure 3 shows the preprocessing results of the gamma frequency band at the FT8 electrode. Figure 3a is a 10 s duration of the original EEG signal with various artifacts. Figure 3b gives the gamma frequency band extracted by wavelet decomposition. The body movements caused large fluctuations in the gamma band were obviously removed, though artifacts induced by blinks still exist. The result of the artifact removal is shown in Figure 3c. It is clear that the wavelet-based thresholding technique can reduce the interference of artifacts in Figure 3b.

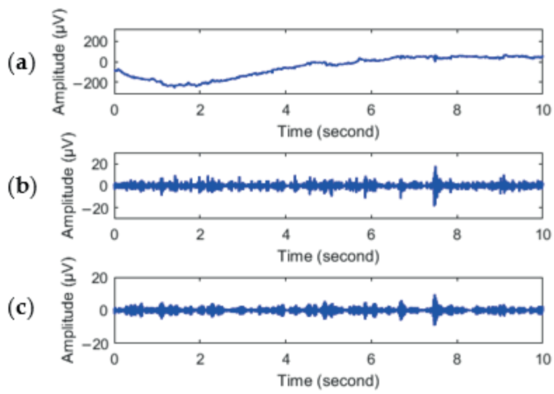


Figure 3. EEG signal preprocessing; the units are μV. (**a**) The original EEG signal; (**b**) the gamma wave obtained by wavelet decomposition; (**c**) the gamma wave after artifact removal.

3.2. Feature Extraction

Five entropy features were calculated to explore the dynamic changes in EEG induced by different emotional film clips including MSE, AE, FE, DE, and Rényi entropy (RE).

3.2.1. Multi-Scale Entropy

MSE, with the ability to reduce the interfere of residual noise on the results by calculating features in different time scales, was chosen to manifest the dynamic changes while subjects were viewing emotional films [48]. It was proposed firstly by Coasta et al. in 2003 [49] that MSE could reflect the complexity of signals in different scale factors by extending the idea of SE to several time scales. For the EEG signal $\{x_1, \ldots, x_i, \ldots, x_N\}$, it is first coarse-grained according to a specified scaling factor τ. In this process, the original signal is divided by sliding windows with a length of τ, and the average value is then calculated in each window to obtain the coarse-grained time series $\{y^{(\tau)}\}$. It is defined as

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \le j \le \frac{N}{\tau}, \tag{2}$$

Then, the SE of the simplified time series $\{y^{(\tau)}\}$ is calculated at each time scale. For more information about the calculation of SE and the parameter setting, please see Appendix B.

3.2.2. Approximate Entropy

AE proposed by Pincus in 1991 is a kind of nonlinear dynamics parameter to measure the complexity and the statistical quantization characteristics of the signal [50]. Due to its effectiveness in reflecting the structure characteristics and complexity information of signals with fewer data points, it is widely used in time series classification studies. Its formula is

$$\mathrm{AE}(m, r, N) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} \log C_i^m(r) - \frac{1}{N-m}\sum_{i=1}^{N-m} \log C_i^{m+1}(r), \tag{3}$$

where $C_i^m(r)$ can be calculated by

$$C_i^m(r) = \frac{B_i^m}{N-m+1}, \tag{4}$$

where B_i^m is the number of matches of dimension m.

The mode dimension m is set as 2, and the tolerance r is equal to the standard deviation of the signal times 0.2. N is the data length, which is set as 256, equals the data points in a 1 s time window without overlap.

3.2.3. Fuzzy Entropy

FE is also a measure of the complexity of signals like AE. Instead of the Heaviside Step Function used in AE, the concept of the fuzzy set is introduced into FE to measure the similarity of two vectors. An exponential function is chosen as the fuzzy function that enables the FE values to change smoothly and continuously with parameters change [51]. In this case, FE is also calculated in our study to make a comparison with other entropy features. It can be calculated by

$$\mathrm{FE}(m, n, r, N) = \ln \frac{O^m(n, r)}{O^{m+1}(n, r)}, \tag{5}$$

where $O^m(n,r)$ is the mean value of the fuzzy membership of the time series with length N in dimension m and tolerance r; n is used to determine the gradient of the similarity tolerance boundary. More details about the definition of $O^m(n,r)$ and the parameters in Equation (5) are described in detail in Appendix B.

3.2.4. Rényi Entropy

RE is a generalization of Shannon entropy, which reflects the time-frequency features and randomness of signals. It is widely used in information theory, such as classification problems. For a given EEG signal $X = \{x_1, \ldots, x_i, \ldots, x_N\}$, its RE can be calculated by

$$\mathrm{RE} = \frac{1}{1-q} \log\left(\sum_{i=1}^{N} p(i)^q\right), \quad q \geq 0 \,\&\, q \neq 1, \tag{6}$$

where q is the entropic index, $p(i)$ is the probability of choosing x_i in X, and $\sum_{i=1}^{N} p(i) = 1$.

According to Kar et al.'s [52] study, we use $q = 2$ to calculate the two-order entropy in a sliding window with a length of one second.

3.2.5. Differential Entropy

As an extension of Shannon entropy, DE can be utilized to reveal the complexity of a time series [53]. Previous study has proven that DE performs better than energy spectrum and asymmetrical features in EEG-based emotion state detection [54]; thus, we calculated DE to represent the changes in EEG signals in different emotional films. It is defined as

$$\mathrm{DE} = -\int_a^b f(x)\log(f(x))dx, \tag{7}$$

where $f(x)$ is the probability density function of the time series and $[a, b]$ is the taking value interval. If the time series is approximately a Gaussian distribution $N(\mu,\sigma^2)$, its DE can then be calculated by

$$\mathrm{DE} = -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}\right) dx = \frac{1}{2}\log 2\pi e\sigma^2, \tag{8}$$

where μ and σ^2 are the expectation and variance of the time series. In our present work, DE is extracted from the signals in a sliding window of a length 1 s without overlap.

3.3. BiLSTM

As an update to LSTM, BiLSTM not only possesses the ability to avoid the receding gradient problem but also memorizes long- and short-term dependency of EEG in a forward direction as well as in a backward direction [55]. As shown in Figure 4, we can see that BiLSTM works in a way that integrates two LSTM together composed of LSTM memory cells.

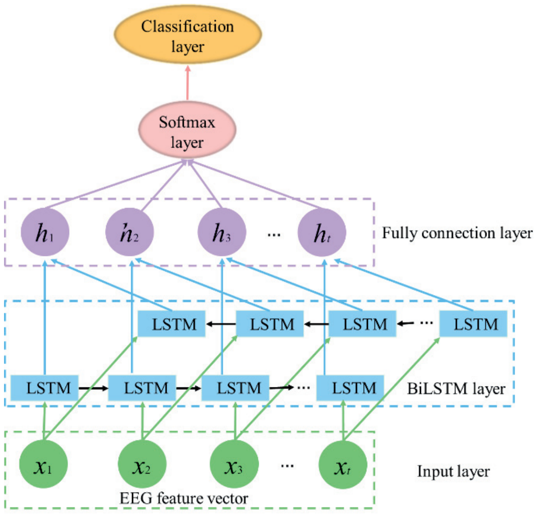


Figure 4. BiLSTM network architecture. It consists of five layers (input layer, BiLSTM layer, fully connection layer, softmax layer, and classification layer). x_t is the EEG feature of time t. h_t is the hidden state of LSTM cell in time t.

LSTM memory cells contain four neural network layers compared to conventional RNN cells with only one layer to model the long-term context. The structures called "gate" consist of neural network layers, and their interactions make it possible for a LSTM memory cell to add or remove information from the cell state [56]. The details about how the memory cell works can be found in Appendix C. Figure 5 shows the structure of a LSTM memory cell.

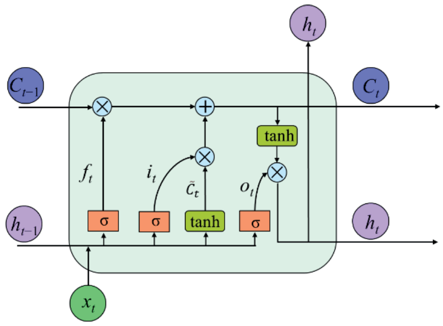


Figure 5. The details of a LSTM memory cell. It contains two kinds of activation functions (σ and tanh). C_t is the LSTM cell state in time t. f_t, $i_{t,}$ and o_t represent the outputs of forget gate, input gate, and output gate in time t separately.

4. Results

4.1. Feature Extraction

In this paper, the five-scale MSE feature was first extracted from the five frequency bands for the selected twelve electrodes after preprocessing. Four other kinds of entropy-based features mentioned in Section 3.2 were also calculated. A non-overlapped sliding window of 1 s was used in the feature extraction procedure. The dimension of the obtained MSE feature matrix for each subject is $300 \times N$, and the other features (i.e., AE, FE, RE, and DE) share the same dimension as $60 \times N$, where N stands for the sampling time. Figure 6 shows parts of the preprocessed gamma frequency band and its entropy features in FT8. The red numbers in Figure 6 are the emotion types of the film clips' duration, which are in accordance with the self-reported emotional states. A positive emotion is marked as "1", and "0" and "-1" represent neutral and negative emotions, respectively. The interval between two adjacent purple dashed lines corresponds to the film clip.

It can be seen from Figure 6 that the waveform of the gamma band after preprocessing varies in different emotional films, and the five entropy-based features change regularly corresponding to the film clips. In Figure 6a, the amplitudes of watching a positive film are obviously larger than those of the other two emotional films. The amplitudes of negative movies rank in second place, followed by those of neutral movies. The fluctuations of DE and FE are similar, as shown in Figure 6b,d. The highest peak values occur during positive emotion film clips, and the lowest valleys appear while subjects watch neutral film clips. The feature values of watching negative film clips are positioned between these two conditions. Additionally, AE and RE share the same waveforms, as can be seen in Figure 6c,e, which are contrary to those of DE and FE. As for the result of MSE in Figure 6f, there is a slightly increasing tendency in the values of MSE when subjects were watching positive and negative films compared with the neutral films, whereas no obvious differences can be seen between positive and negative films in MSE.

4.2. The Classification Results of BiLSTM

BiLSTM is applied to classify the emotion states of the subjects in order to explore the long-term dependency and interplay of the extracted features at different times. We trained BiLSTM models for each kind of feature and the fused entropy features separately. Then the performance of BiLSTM utilizing a single-type feature was compared with that of fused entropy features. Further, the result was also compared with conventional LSTM to make the results more convincing. The five types of feature matrixes obtained in Section 4.1 were first normalized to $(-1, 1)$ to eliminate the effect of individual differences. Then, the normalized feature matrixes could be directly fed into the classifier. As for the output (i.e., the category label vectors), it can be set according to the sequence of the film clips and the results of the self-assessment. The dimension of the label vectors is $1 \times N$, where N is the sampling time. There are in total three categories in this paper: positive, neutral, and negative. The training data come from eighteen subjects selected randomly from all, and the data of the remaining five subjects was set as testing data. The recognition accuracy was defined as the average accuracy of the five subjects in the testing group. The results of different entropy features are shown in Table 1. "ALL" means the five entropy features.

Table 1. The mean accuracies of BiLSTM and LSTM for different features (%).

Feature	AE	FE	RE	DE	MSE	ALL
LSTM	61.1	59.47	54.23	65.09	66.12	67.22
BiLSTM	63.43	61.1	57.15	66.34	67.9	70.05

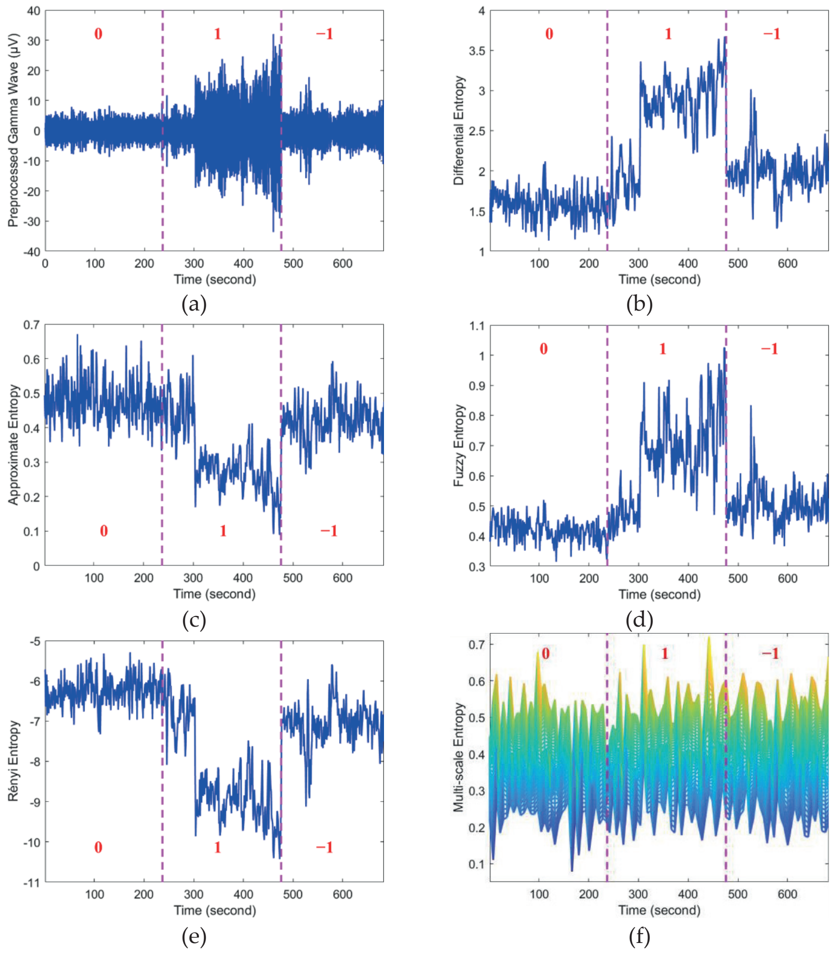


Figure 6. The results of five entropy-based features of EEG. The numbers "1", "0", and "−1" represent positive, neutral, and negative emotions, respectively. The purple dashed lines show the boundaries of different film clips. (**a**) Preprocessed gamma frequency band; (**b**) DE; (**c**) AE; (**d**) FE; (**e**) RE; (**f**) MSE.

From Table 1, it can be seen that the performance of BiLSTM is clearly better than that of LSTM. As for the result comparison the two models based on single-feature inputs, LSTM and BiLSTM with the input of MSE achieves the best result, reaching at 66.12% and 67.9%, respectively, and they are slightly higher than those of DE, which was proven to be a better feature to classify different emotion states in a previous study [11]. The RE feature leads to the lowest accuracy of 54.23% in LSTM and 57.15% in BiLSTM, and the accuracies of AE and FE lie between RE and AE. Furthermore, the two models' performance is apparently enhanced while using fused entropy features to detect emotion states peaking at 67.22% in LSTM and 70.05% in BiLSTM.

5. Discussion

Different types of entropy features of EEG were calculated to explore the dynamic changes of EEG while subjects were watching emotional film clips, and the results of the gamma frequency band in FT8 are shown in Figure 6. It can be seen that the entropy values are significantly different for different emotional films. The higher peaks appear in positive and negative films in the results of DE, FE, and MSE, and the lowest valleys occur in neutral films. Subjects are highly stimulated in emotion when exposed to the positive and negative films; in these cases, the brain activity is usually active and complex (see Figure 6a), and it objectifies a high complexity in the gamma band. However, subjects are not as immersed in the neutral films as during the other two conditions. Thus, there is a decrease in complexity and the lowest entropy values appear. The results are accordant with previous studies showing that greater activities of the gamma band can be found in positive states and the lowest activities are in neutral conditions [15,57]. As for the results of RE, the positive films induce the largest absolute values followed by negative films, and the smallest absolute values are caused by neutral films. The results are consistent among DE, FE, and MSE. RE is a reflection of the amplitude's distribution; a smaller RE is obtained when the EEG amplitude concentrates in a certain subsequence, indicating the signal is more ordered and less complex [58]. Hence, a larger absolute value of RE can be seen in a more active gamma band when subjects are greatly affected by positive and negative films. Subjects are usually more engaged in viewing positive film clips than neutral clips, generating more active brain activities, but the complexity changes in EEG may be not large enough to be detected by AE. Different from the changing rules of the above four features, AE reaches its highest value in neutral emotion and its lowest value in positive emotion, as shown in Figure 6c. As we know, AE was proposed to measure the average logarithmic conditional probability of the new pattern's occurrence in a signal with the dimension change [50]. It introduces self-matches into calculation and will inevitably lead to calculation bias, which can result in an insensitivity to small changes in complexity [59]. Thus, the small complexity changes in brain activity during the positive film's duration may be ignored. It decreases the detected new patterns and leads to the lower AE value in positive emotion.

The extracted entropy features are used to detect emotion states by training a BiLSTM model. The classification results in Table 1 illustrate that the mean accuracy of MSE is slightly higher than that of the other four entropy features using BiLSTM, as it can reveal the complex information of EEG in film watching as well as reduce the influence of the residual noise on the results to some extent [60]. Moreover, the performance of BiLSTM is further improved with multiple entropy features, reaching to 70.05%. Different features can compensate for each other according to the study in [23]. Then, the more useful information of EEG features can be learned by BiLSTM, thus contributing to a higher accuracy in emotion recognition. Additionally, the traditional one-directional LSTM was also trained to make a comparison with BiLSTM. Table 1 indicates that the trained BiLSTM performs better than LSTM since it can learn the long- and short-term dependency among EEG features in a forward direction and in a reverse direction [55]. This finding is consistent with the finding that utilizing the BiLSTM classifier is an efficient way to decrease the train and test error and to increase the classification accuracy [61]. The accuracy is comparable with that in the research in [26], who utilized the DE feature and transfer learning to detect three emotions and reached an accuracy of 72.47% in the SEED dataset. However, our accuracy is lower than that in [11] who used the DE feature and DBN to recognize different emotions in the SEED dataset. This might be because SEED includes only Chinese subjects, and the data we obtained from the competition include not only Chinese but also French subjects (i.e., the data used in our study consist of two datasets: SEED and SEED-FRA). Although the subjects watched their native language films during the experiment to elicit emotional changes more easily, and the stimulus types of the films are the same, there are differences between the Chinese and French subjects, which may lead to the lower accuracy in our study. In a study utilizing the same data as ours, a similar accuracy of about 70% was obtained in [37] for depression recognition.

This study shows the feasibility of recognizing emotion status by deploying multiple entropy features. However, as the dataset we used contains subjects and stimuli of two native languages, and film clips in different emotional categories vary in the degree of induced emotion, there are still limitations in the present work. First, the used dataset consisted of two public datasets, which involved both Chinese and French subjects and stimuli. Each subject watched films in his native language to elicit emotional changes more easily, but the number of film clips for the two languages was different. Since the stimulus types of the films were the same, we assume the different number of film clips for the subjects had no effect on the results in the present work, though influences we do not realize might exist in the results because of the different native speakers and the number of film clips. Further, subjective labeling of participant emotional states was adopted to recognize the emotions in our study, as it is beneficial to the feature analysis for the same type of emotion states and to ensure the reliability of the results by providing more accurate feelings of the subjects. Although most of studies about emotions in the literature choose subjective labeling, there is research that selects objective labeling to label the emotional states, which can be further studied in our future experiment design. Additionally, the results obtained in our present work can be further improved by utilizing some new, more effective algorithms. Electrical Source Imaging (ESI), an emerging algorithm to reconstruct brain or cardiac electrical activity from electrical potentials measured away from the brain, can determine the location of current sources from measurements of voltages [62]. This would be a novel and interesting topic for estimating the cortex brain regions involved in each video viewing to improve the emotion recognition accuracy in our future work.

6. Conclusions

In this paper, we proposed a cross-subject emotion recognition framework based on fused EEG entropy features and a BiLSTM classifier. It demonstrates that MSE is more effective in analyzing the complex emotion information in EEG than the other entropy features when adopting a single EEG feature. What is more, the classification accuracy can be apparently increased by combining all entropy features, which proves that there is information compensation among different types of features.

Future work mainly includes two aspects. One aspect is to extract more features from different perspectives, such as time-frequency domain features and non-linear dynamic features, for feature fusion; the other is enhancing the performance of the classifier with a new feature fusion algorithm or by estimating the cortex brain regions involved in watching emotional film clips by applying Electrical Source Imaging.

Author Contributions: Conceptualization, C.Z.; methodology, X.Z. and C.Z.; software, X.Z.; validation, X.Z.; formal analysis, X.Z.; investigation, X.Z.; data curation, H.G. and Y.F.; writing—original draft preparation, X.Z.; writing—review and editing, C.Z. and X.Z.; supervision, C.Z., T.H., and F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (no. 61703069 and 62001312), the National Foundation in China (no. JCKY2019110B009), the Fundamental Research Funds for the Central Universities (no. DUT21GF301) and the Science and Technology Planning Project of Liaoning Province (no. 2021JH1/10400049).

Institutional Review Board Statement: Not applicable.

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

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: [https://bcmi.sjtu.edu.cn/home/seed/, accessed on 30 October 2014].

Acknowledgments: We would like to thank Shuo Cao from Dalian University of Technology for her support and helpful suggestions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Preprocessing

In this section, we preprocessed the EEG signals to remove the artifacts and obtain five frequency bands. We first downsampled the signals from 1000 Hz to 256 Hz to enhance the calculation efficiency. Then, the EEG segments during film clip watching were extracted. After that, to obtain different frequency band rhythms, a bandstop Butterworth filter of order 4 with cutoff frequencies of 0.19 and 0.2 was used to remove the power line, and wavelet decomposition was applied to roughly obtain the five frequency bands of EEG signals. A wavelet-based technique was chosen to remove the artifacts in the end.

Wavelet transform is a well-known time-frequency analysis algorithm and has been widely adopted to deal with non-stationary signals. The original signals are decomposed and expressed with a scaled and shifted version of the mother wavelet $\psi(t)$ and a scaling function $\phi(t)$ [52]. The discrete mother wavelet can be expressed as

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi\left(2^{-j}t - k\right), \qquad j,k \in Z, \tag{A1}$$

The signal $S(t)$ can then be represented as

$$S(t) = \sum_k s_j(k)\phi_{j,k}(t) + \sum_k d_j(k)\psi_{j,k}(t), \tag{A2}$$

where $s_j(k)$ is the approximate coefficient at the jth level, and $d_j(k)$ represents the detailed coefficient.

In this study, the EEG signal was decomposed into seven levels using db6 (Daubechies family), as the waveform of db6 is similar to the EEG signal and it has been widely used in decomposing EEG in the literature [47]. The delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–50 Hz) frequency bands were roughly obtained by reconstructing the detailed components at levels of seven, five, four, three, and two, respectively.

To eliminate the artifacts' interference, the five frequency band rhythms are decomposed separately into three levels with a 1 s sliding window, and the thresholds of each level are calculated by Equation (1). The coefficient is halved if its value is larger than the calculated threshold. In this way, a new set of signals is generated without artifacts.

Appendix B. Feature Extraction

This section lists some details about the well-known formulas and the parameter setting to extract the entropy features of the five frequency bands, which provides supplementary information for Section 3.2.

After the first coarse graining step in extracting the MSE feature (see Section 3.2.1), the SE is then calculated for the new time series in different scales by the following formula

$$\mathrm{SE}(m,r,N) = -\ln\left[\frac{B^{m+1}(r)}{B^m(r)}\right], \tag{A3}$$

where m is the mode dimension of the data vector, r is the tolerance for similarity matches, and N is the number of data points. $B^m(r)$ and $B^{m+1}(r)$ are the numbers of matches for dimension m and $m + 1$, respectively.

In this paper, the five-scale of the MSE feature is calculated with a 1 s sliding window for the five frequency bands. The mode dimension m is set as 2, and the tolerance r is equal to the standard deviation of the signal times 0.2.

As for FE in Section 3.2.3, the parameter $O^m(n,r)$ in Equation (5) is defined as

$$O^m(n,r) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{j=1,j\neq i}^{N-m} D_{ij}^m\right), \tag{A4}$$

$$D_{ij}^m = f\left(d_{ij}^m, n, r\right), \tag{A5}$$

$$X^m(i) = [x(i), x(i+1), \ldots, x(i+m-1)], \tag{A6}$$

$$X^m(j) = [x(j), x(j+1), \ldots, x(j+m-1)], \tag{A7}$$

where n is the fuzzy exponent, and d_{ij}^m, the distance between vectors $X^m(i)$ and $X^m(j)$, is defined as the absolute value of the maximum difference between the corresponding elements in the two vectors. D_{ij}^m is then calculated by the fuzzy function $f(d_{ij}^m,n,r)$ to show the similarity of vectors $X^m(i)$ and $X^m(j)$.

The same parameters of m, r, and N with AE are used. n is set as 2 according to the research of Chen et al. [51].

Appendix C. BiLSTM Classifier

As mentioned in Section 3.3, BiLSTM is an integration of two opposite-direction LSTM and consists of LSTM memory cells. A LSTM cell contains an input gate, output gate, and a forget gate (see Figure 5), which serve to protect and control the cell state C_t. The first step in a LSTM cell is to select that which should be deleted in C_{t-1}, and it is realized by the forget gate:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right), \tag{A8}$$

Then, the input gate consisting of a sigmoid layer chooses what kind of values need to be updated (Equation (A9)), and a tanh layer creates a new candidate cell state $\widetilde{C}_t$ from Equation (A10). Following these two steps, the new cell state C_t can be obtained from Equation (A11):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{A9}$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{A10}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t, \tag{A11}$$

In the end, the output gate and a tanh layer are activated to decide the final output $h_{t,}$ as expressed in Equations (A12) and (A13):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{A12}$$

$$h_t = o_t * \tanh(C_t) \tag{A13}$$

In the equations, σ and tanh represent the activation functions, x_t denotes the input of the cell (the EEG feature of time t), and W, b and h are the weight, bias, and hidden state of the gates, separately.

In this paper, we trained and tested various BiLSTM and LSTM classifiers with different parameters to find the best parameters for each feature. The Adam optimizer was utilized as the optimizer of the classifier. We selected the data of 18 subjects as the training set, and that of the remaining five subjects was the testing set. The average accuracy of the test set was used to evaluate the classifier's performance. The details of the parameters tested are listed in Table A1.

Table A1. The detailed parameters trained in BiLSTM and LSTM.

Name	Value
Hidden units	[10:150] with step of 10
Epochs	[50:150] with step of 20
Mini batch size	[50:100] with step of 10
Learning rate	0.001

References

1. Shu, L.; Xie, J.; Yang, M.; Li, Z.; Li, Z.; Liao, D.; Xu, X.; Yang, X. A Review of Emotion Recognition Using Physiological Signals. *Sensors* **2018**, *18*, 2074. [CrossRef] [PubMed]
2. Picard, R.W. Affective Computing: Challenges. *Int. J. Hum. Comput. Stud.* **2003**, *59*, 55–64. [CrossRef]
3. Siriwardhana, S.; Kaluarachchi, T.; Billinghurst, M.; Nanayakkara, S. Multimodal Emotion Recognition with Transformer-Based Self Supervised Feature Fusion. *IEEE Access* **2020**, *8*, 176274–176285. [CrossRef]
4. Batbaatar, E.; Li, M.; Ryu, K.H. Semantic-Emotion Neural Network for Emotion Recognition from Text. *IEEE Access* **2019**, *7*, 111866–111878. [CrossRef]
5. Martinez, H.P.; Bengio, Y.; Yannakakis, G.N. Learning Deep Physiological Models of Affect. *IEEE Comput. Intell. Mag.* **2013**, *8*, 20–33. [CrossRef]
6. Jain, D.K.; Shamsolmoali, P.; Sehdev, P. Extended Deep Neural Network for Facial Emotion Recognition. *Pattern Recognit. Lett.* **2019**, *120*, 69–74. [CrossRef]
7. Meng, H.; Yan, T.; Yuan, F.; Wei, H. Speech Emotion Recognition from 3D Log-Mel Spectrograms with Deep Learning Network. *IEEE Access* **2019**, *7*, 125868–125881. [CrossRef]
8. Kessous, L.; Castellano, G.; Caridakis, G. Multimodal Emotion Recognition in Speech-Based Interaction Using Facial Expression, Body Gesture and Acoustic Analysis. *J. Multimodal User Interfaces* **2010**, *3*, 33–48. [CrossRef]
9. Zhang, J.; Yin, Z.; Chen, P.; Nichele, S. Emotion Recognition Using Multi-Modal Data and Machine Learning Techniques: A Tutorial and Review. *Inf. Fusion* **2020**, *59*, 103–126. [CrossRef]
10. Kim, J.; André, E. Emotion Recognition Based on Physiological Changes in Music Listening. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 2067–2083. [CrossRef]
11. Zheng, W.-L.; Lu, B.-L. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [CrossRef]
12. Egger, M.; Ley, M.; Hanke, S. Emotion Recognition from Physiological Signal Analysis: A Review. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 35–55. [CrossRef]
13. Du, R.; Lee, H.J. Power Spectral Performance Analysis of EEG during Emotional Auditory Experiment. In Proceedings of the 2014 International Conference on Audio, Language and Image Processing, Shanghai, China, 7–9 July 2014; pp. 64–68. [CrossRef]
14. Du, R.; Lee, H.J. Frontal Alpha Asymmetry during the Audio Emotional Experiment Revealed by Event-Related Spectral Perturbation. In Proceedings of the 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI), Shenyang, China, 14–16 October 2015; pp. 531–536. [CrossRef]
15. Liu, S.; Meng, J.; Zhang, D.; Yang, J.; Zhao, X.; He, F.; Qi, H.; Ming, D. Emotion Recognition Based on EEG Changes in Movie Viewing. In Proceedings of the 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, France, 22–24 April 2015; pp. 1036–1039. [CrossRef]
16. Mehmood, R.M.; Lee, H.J. A Novel Feature Extraction Method Based on Late Positive Potential for Emotion Recognition in Human Brain Signal Patterns. *Comput. Electr. Eng.* **2016**, *53*, 444–457. [CrossRef]
17. Wang, Y.-H.; Chen, I.-Y.; Chiueh, H.; Liang, S.-F. A Low-Cost Implementation of Sample Entropy in Wearable Embedded Systems: An Example of Online Analysis for Sleep EEG. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
18. Chen, T.; Ju, S.; Yuan, X.; Elhoseny, M.; Ren, F.; Fan, M.; Chen, Z. Emotion Recognition Using Empirical Mode Decomposition and Approximation Entropy. *Comput. Electr. Eng.* **2018**, *72*, 383–392. [CrossRef]
19. Zheng, W.-L.; Guo, H.-T.; Lu, B.-L. Revealing Critical Channels and Frequency Bands for Emotion Recognition from EEG with Deep Belief Network. In Proceedings of the 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, France, 22–24 April 2015; pp. 154–157.
20. Ferrario, M.; Signorini, M.; Magenes, G.; Cerutti, S. Comparison of Entropy-Based Regularity Estimators: Application to the Fetal Heart Rate Signal for the Identification of Fetal Distress. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 119–125. [CrossRef]
21. Hadoush, H.; Alafeef, M.; Abdulhay, E. Brain Complexity in Children with Mild and Severe Autism Spectrum Disorders: Analysis of Multiscale Entropy in EEG. *Brain Topogr.* **2019**, *32*, 914–921. [CrossRef]
22. Miskovic, V.; MacDonald, K.J.; Rhodes, L.J.; Cote, K.A. Changes in EEG Multiscale Entropy and Power-Law Frequency Scaling During the Human Sleep Cycle. *Hum. Brain Mapp.* **2019**, *40*, 538–551. [CrossRef]
23. Hasan, J.; Kim, J.-M. A Hybrid Feature Pool-Based Emotional Stress State Detection Algorithm Using EEG Signals. *Brain Sci.* **2019**, *9*, 376. [CrossRef]
24. Liu, Y.-J.; Yu, M.; Zhao, G.; Song, J.; Ge, Y.; Shi, Y. Real-Time Movie-Induced Discrete Emotion Recognition from EEG Signals. *IEEE Trans. Affect. Comput.* **2017**, *9*, 550–562. [CrossRef]

25. Kolodyazhniy, V.; Kreibig, S.D.; Gross, J.J.; Roth, W.T.; Wilhelm, F.H. An Affective Computing Approach to Physiological Emotion Specificity: Toward Subject-Independent and Stimulus-Independent Classification of Film-Induced Emotions. *Psychophysiology* **2011**, *48*, 908–922. [CrossRef] [PubMed]
26. Lan, Z.; Sourina, O.; Wang, L.; Scherer, R.; Muller-Putz, G.R. Domain Adaptation Techniques for EEG-Based Emotion Recognition: A Comparative Study on Two Public Datasets. *IEEE Trans. Cogn. Dev. Syst.* **2019**, *11*, 85–94. [CrossRef]
27. Wöllmer, M.; Kaiser, M.; Eyben, F.; Schuller, B.; Rigoll, G. LSTM-Modeling of Continuous Emotions in an Audiovisual Affect Recognition Framework. *Image Vis. Comput.* **2013**, *31*, 153–163. [CrossRef]
28. Liu, G.; Guo, J. Bidirectional LSTM with Attention Mechanism and Convolutional Layer for Text Classification. *Neurocomputing* **2019**, *337*, 325–338. [CrossRef]
29. Narendra, N.P.; Alku, P. Glottal Source Information for Pathological Voice Detection. *IEEE Access* **2020**, *8*, 67745–67755. [CrossRef]
30. Bollepalli, B.; Airaksinen, M.; Alku, P. Lombard Speech Synthesis Using Long Short-Term Memory Recurrent Neural Networks. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Orleans, LA, USA, 5–9 March 2017; pp. 5505–5509. [CrossRef]
31. Carrara, F.; Elias, P.; Sedmidubsky, J.; Zezula, P. LSTM-Based Real-Time Action Detection and Prediction in Human Motion Streams. *Multimed. Tools Appl.* **2019**, *78*, 27309–27331. [CrossRef]
32. Sun, Q.; Wang, C.; Guo, Y.; Yuan, W.; Fu, R. Research on a Cognitive Distraction Recognition Model for Intelligent Driving Systems Based on Real Vehicle Experiments. *Sensors* **2020**, *20*, 4426. [CrossRef]
33. Manoharan, T.A.; Radhakrishnan, M. Region-Wise Brain Response Classification of ASD Children Using EEG and BiLSTM RNN. *Clin. EEG Neurosci.* **2021**, 15500594211054990. [CrossRef]
34. Fernando, T.; Denman, S.; Sridharan, S.; Fookes, C. Soft + Hardwired Attention: An LSTM Framework for Human Trajectory Prediction and Abnormal Event Detection. *Neural Netw.* **2018**, *108*, 466–478. [CrossRef]
35. Joshi, V.M.; Ghongade, R.B. EEG Based Emotion Detection Using Fourth Order Spectral Moment and Deep Learning. *Biomed. Signal Process. Control* **2021**, *68*, 102755. [CrossRef]
36. Mahmud, T.; Khan, I.A.; Mahmud, T.I.; Fattah, S.A.; Zhu, W.-P.; Ahmad, M.O. Sleep Apnea Detection from Variational Mode Decomposed EEG Signal Using a Hybrid CNN-BiLSTM. *IEEE Access* **2021**, *9*, 102355–102367. [CrossRef]
37. Chang, H.; Zong, Y.; Zheng, W.; Tang, C.; Zhu, J.; Li, X. Depression Assessment Method: An EEG Emotion Recognition Framework Based on Spatiotemporal Neural Network. *Front. Psychiatry* **2022**, *12*, 837149. [CrossRef] [PubMed]
38. Posner, J.; Russell, J.A.; Peterson, B.S. The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology. *Dev. Psychopathol.* **2005**, *17*, 715–734. [CrossRef] [PubMed]
39. Kılıç, B.; Aydın, S. Classification of Contrasting Discrete Emotional States Indicated by EEG Based Graph Theoretical Network Measures. *Neuroinformatics* **2022**, 1–15. [CrossRef]
40. Liu, W.; Zheng, W.-L.; Li, Z.; Wu, S.-Y.; Gan, L.; Lu, B.-L. Identifying Similarities and Differences in Emotion Recognition with EEG and Eye Movements among Chinese, German, and French People. *J. Neural Eng.* **2022**, *19*, 026012. [CrossRef]
41. Schaefer, A.; Nils, F.; Sanchez, X.; Philippot, P. Assessing the Effectiveness of a Large Database of Emotion-Eliciting Films: A New Tool for Emotion Researchers. *Cogn. Emot.* **2010**, *24*, 1153–1172. [CrossRef]
42. Nie, D.; Wang, X.-W.; Shi, L.-C.; Lu, B.-L. EEG-Based Emotion Recognition during Watching Movies. In Proceedings of the 2011 5th International IEEE/EMBS Conference on Neural Engineering, Cancun, Mexico, 27 April–1 May 2011; pp. 667–670. [CrossRef]
43. Gurudath, N.; Riley, H.B. Drowsy Driving Detection by EEG Analysis Using Wavelet Transform and K-means Clustering. *Procedia Comput. Sci.* **2014**, *34*, 400–409. [CrossRef]
44. Kumar, P.S.; Arumuganathan, R.; Sivakumar, K.; Vimal, C. A Wavelet Based Statistical Method for De-Noising of Ocular Artifacts in EEG Signals. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **2008**, *8*, 87–92.
45. Zhang, C.; Sun, L.; Cong, F.; Ristaniemi, T. Spatiotemporal Dynamical Analysis of Brain Activity During Mental Fatigue Process. *IEEE Trans. Cogn. Dev. Syst.* **2020**, *13*, 593–606. [CrossRef]
46. Zhang, C.; Cong, F.; Kujala, T.; Liu, W.; Liu, J.; Parviainen, T.; Ristaniemi, T. Network Entropy for the Sequence Analysis of Functional Connectivity Graphs of the Brain. *Entropy* **2018**, *20*, 311. [CrossRef]
47. Poorna, S.S.; Raghav, R.; Nandan, A.; Nair, G.J. EEG Based Control—A Study Using Wavelet Features. In Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 19–22 September 2018; pp. 550–553. [CrossRef]
48. Costa, M.; Goldberger, A.L.; Peng, C.-K. Multiscale Entropy Analysis of Biological Signals. *Phys. Rev. E* **2005**, *71*, 21906. [CrossRef] [PubMed]
49. Costa, M.; Peng, C.-K.; Goldberger, A.L.; Hausdorff, J.M. Multiscale Entropy Analysis of Human Gait Dynamics. *Phys. A Stat. Mech. Appl.* **2003**, *330*, 53–60. [CrossRef] [PubMed]
50. Pincus, S.M. Approximate Entropy As a Measure of System Complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [CrossRef] [PubMed]
51. 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]
52. Kar, S.; Bhagat, M.; Routray, A. EEG Signal Analysis for the Assessment and Quantification of Driver's Fatigue. *Transp. Res. Part F Traffic Psychol. Behav.* **2010**, *13*, 297–306. [CrossRef]
53. Feutrill, A.; Roughan, M. A Review of Shannon and Differential Entropy Rate Estimation. *Entropy* **2021**, *23*, 1046. [CrossRef]

54. Duan, R.-N.; Zhu, J.-Y.; Lu, B.-L. Differential Entropy Feature for EEG-Based Emotion Classification. In Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), Diego, CA, USA, 6–8 November 2013; pp. 81–84.
55. Yang, J.; Huang, X.; Wu, H.; Yang, X. EEG-based emotion classification based on Bidirectional Long Short-Term Memory Network. *Procedia Comput. Sci.* **2020**, *174*, 491–504. [CrossRef]
56. Wu, E.Q.; Xiong, P.; Tang, Z.-R.; Li, G.-J.; Song, A.; Zhu, L.-M. Detecting Dynamic Behavior of Brain Fatigue Through 3-D-CNN-LSTM. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**, *52*, 90–100. [CrossRef]
57. Martini, N.; Menicucci, D.; Sebastiani, L.; Bedini, R.; Pingitore, A.; Vanello, N.; Milanesi, M.; Landini, L.; Gemignani, A. The Dynamics of EEG Gamma Responses to Unpleasant Visual Stimuli: From Local Activity to Functional Connectivity. *NeuroImage* **2012**, *60*, 922–932. [CrossRef]
58. Xie, O.; Liu, Z.-T.; Ding, X.-W. Electroencephalogram Emotion Recognition Based on a Stacking Classification Model. In Proceedings of the 2018 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; pp. 5544–5548. [CrossRef]
59. Richman, J.S.; Moorman, J.R. Physiological Time-Series Analysis Using Approximate Entropy and Sample Entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [CrossRef]
60. Bhattacharyya, A.; Tripathy, R.K.; Garg, L.; Pachori, R.B. A Novel Multivariate-Multiscale Approach for Computing EEG Spectral and Temporal Complexity for Human Emotion Recognition. *IEEE Sens. J.* **2021**, *21*, 3579–3591. [CrossRef]
61. Kouchak, S.M.; Gaffar, A. Using Bidirectional Long-Short Term Memory with Attention Layer to Estimate Driver Behavior. In Proceedings of the 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 315–320. [CrossRef]
62. National Research Council (US) and Institute of Medicine (US) Committee on the Mathematics and Physics of Emerging Dynamic Biomedical Imaging. Chapter 8, Electrical Source Imaging. In *Mathematics and Physics of Emerging Biomedical Imaging*; National Academies Press (US): Washington, DC, USA, 1996. Available online: https://www.ncbi.nlm.nih.gov/books/NBK232494/ (accessed on 4 September 2022).