Understanding Obsessive-Compulsive Disorder Through Human Skin Textures

Yazhen Zhu[1,2,3], Jian Chen[2,3,4], Yuwei Sun[5], and Wei Wang[2,3,6(✉)]

[1] Royal College of Art, London SW7 2EU, UK
[2] Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China
chenj589@mail2.sysu.edu.cn, ehomewang@ieee.org
[3] Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Shenzhen, MSU-BITUniversity, Shenzhen 518172, Guangdong, China
[4] School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen 518000, China
[5] Columbia University, New York, NY 10027, USA
ys3371@tc.columbia.edu
[6] School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China

Abstract. Obsessive-Compulsive Disorder (OCD) is a complex and heterogeneous mental health condition that challenges our understanding of its underlying mechanisms. This paper explores the potential connection between OCD and human skin textures, particularly Excoriation Disorder (chronic skin-picking), through a comprehensive analysis of existing literature. Investigating cognitive aspects, memory impairments, and potential neurobiological factors contributing to this association, the study also examines the role of human-computer interaction (HCI) in data analysis and treatment approaches, with a focus on skin texture-related aspects. Additionally, the thesis delves into two entry points for understanding OCD through human skin texture. OCD's clinical manifestations involve compulsive repetitive movements, where memory disorders lead individuals to hyperfocus on event details, causing behaviors of constantly enlarging objects, leaving traces of skin texture on them. Drawing inspiration from Exposure and Response Prevention (ERP) therapy, the paper proposes magnifying skin texture details to simulate ERP, exposing patients to imperfections and reducing perfectionistic tendencies. Secondly, related OCD symptoms, like compulsive skin peeling, leave specific skin marks, providing potential clues for identifying OCD characteristics and patterns. This innovative approach offers valuable insights into the complexities of OCD, highlighting the significance of human skin texture in understanding and treating the disorder. By integrating cognitive and neurobiological aspects, this study provides a comprehensive perspective on the intriguing relationship between OCD and human skin textures, contributing to advancements in OCD research and intervention.

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 85–95, 2024.
https://doi.org/10.1007/978-3-031-65126-7_9

11. Zakaria, C., Yilmaz, G., Mammen, P.M., Chee, M., Shenoy, P., Balan, R.: Sleepmore: inferring sleep duration at scale via multi-device wifi sensing. Proc. ACM Interact. Mob. Wearable Ubiquit. Technol. **6**(4), 1–32 (2023)
12. Yu, B., et al.: Wifi-sleep: sleep stage monitoring using commodity wi-fi devices. IEEE Internet Things J. **8**(18), 13900–13913 (2021)
13. Yang, X., Yu, X., Xie, L., Xue, H., Zhou, M., Jiang, Q.: Sleep apnea monitoring system based on commodity wifi devices. Comput. Mater. Cont **2**(69), 2793–2806 (2021)
14. Liu, W., Chang, S., Liu, Y., Zhang, H.: Wi-PSG: detecting rhythmic movement disorder using cots wifi. IEEE Internet Things J. **8**(6), 4681–4696 (2020)
15. Ridolfi, M., Kaya, A., Berkvens, R., Weyn, M., Joseph, W., Poorter, E.D.: Self-calibration and collaborative localization for UWB positioning systems: a survey and future research directions. ACM Comput. Surv. (CSUR) **54**(4), 1–27 (2021)
16. Li, S., Wang, Z., Zhang, F., Jin, B.: Fine-grained respiration monitoring during overnight sleep using IR-UWB radar. In: Hara, T., Yamaguchi, H. (eds.) International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services, pp. 84–101. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-94822-1_5
17. Han, W., Dai, S., Yuce, M.R.: Real-time contactless respiration monitoring from a radar sensor using image processing method. IEEE Sens. J. **22**(19), 19020–19029 (2022)
18. Kwon, H.B., et al.: Attention-based LSTM for non-contact sleep stage classification using IR-UWB radar. IEEE J. Biomed. Health Inform. **25**(10), 3844–3853 (2021)
19. Atlas, D., Srivastava, R., Sekhon, R.S.: Doppler radar characteristics of precipitation at vertical incidence. Rev. Geophys. **11**(1), 1–35 (1973)
20. Baboli, M., Singh, A., Soll, B., Boric-Lubecke, O., Lubecke, V.M.: Wireless sleep apnea detection using continuous wave quadrature doppler radar. IEEE Sens. J. **20**(1), 538–545 (2019)
21. Islam, S.M.M., Lubecke, V.M.: Sleep posture recognition with a dual-frequency microwave doppler radar and machine learning classifiers. IEEE Sensors Lett. **6**(3), 1–4 (2022)
22. Rahman, T., et al.: DoppleSleep: a contactless unobtrusive sleep sensing system using short-range doppler radar. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 39–50 (2015)
23. Chen, L., Xiong, J., Chen, X., Lee, S.I., Zhang, D., Yan, T., Fang, D.: Lungtrack: towards contactless and zero dead-zone respiration monitoring with commodity RFIDS. Proc. ACM Interact. Mob. Wearable Ubiquit. Technol. **3**(3), 1–22 (2019)
24. Liu, C., Xiong, J., Cai, L., Feng, L., Chen, X., Fang, D.: Beyond respiration: contactless sleep sound-activity recognition using RF signals. Proc. ACM Interact. Mob. Wearable Ubiquit. Technol. **3**(3), 1–22 (2019)
25. Zhao, M., Yue, S., Katabi, D., Jaakkola, T.S., Bianchi, M.T.: Learning sleep stages from radio signals: a conditional adversarial architecture. In: International Conference on Machine Learning, pp. 4100–4109. PMLR (2017)

In summary, each technology has its own merits and considerations. Contactless sensing also leaves much to be desired, such as greater noise immunity to the varying light conditions of different indoor environments. At the same time, because contactless sensing can capture more information, it faces more serious privacy issues. The choice depends on specific requirements, budget constraints, and the desired level of monitoring accuracy. Besides, more research can focus on how to combine these two methods for better performance and less cost.

5 Conclusion

In this work, we review the existing sleep monitoring methods based on Wifi sensors and wireless sensors. Then we make a comparative analysis between these two methods for a better illustration of wireless sensors used in the field of sleep monitoring. Through the summary of the existing methods, we can better find the direction for the follow-up research. However, in addition to wifi sensors and radar, acoustic and optical sensors are also beginning to be used in this field. Therefore, it is our future work to further summarize and analyze the advantages and disadvantages of these methods.

References

1. Perez-Pozuelo, I., et al.: The future of sleep health: a data-driven revolution in sleep science and medicine. NPJ Digital Med. **3**(1), 42 (2020)
2. Bertisch, S.M., et al.: Insomnia with objective short sleep duration and risk of incident cardiovascular disease and all-cause mortality: sleep heart health study. Sleep 41(6), zsy047 (2018)
3. Zhou, Q., Zhang, M., Hu, D.: Dose-response association between sleep duration and obesity risk: a systematic review and meta-analysis of prospective cohort studies. Sleep Breathing **23**, 1035–1045 (2019)
4. Palagini, L., Hertenstein, E., Riemann, D., Nissen, C.: Sleep, insomnia and mental health. J. Sleep Res. **31**(4), e13628 (2022)
5. Rundo, J.V., Downey, R., III.: Polysomnography. Handb. Clin. Neurol. **160**, 381–392 (2019)
6. Engstrøm, M., Rugland, E., Heier, M.S.: Polysomnography (PSG) for studying sleep disorders. Tidsskrift for den Norske laegeforening: tidsskrift for praktisk medicin, ny raekke **133**(1), 58–62 (2013)
7. Rottenberg, F., Nguyen, T.-H., Dricot, J.-M., Horlin, F., Louveaux, J.: CSI-based versus RSS-based secret-key generation under correlated eavesdropping. IEEE Trans. Commun. **69**(3), 1868–1881 (2020)
8. Chen, Z., Zhang, L., Jiang, C., Cao, Z., Cui, W.: Wifi CSI based passive human activity recognition using attention based BLSTM. IEEE Trans. Mob. Comput. **18**(11), 2714–2724 (2018)
9. Gui, L., Ma, C., Sheng, B., Guo, Z., Cai, J., Xiao, F.: In-home monitoring sleep turnover activities and breath rate via wifi signals. IEEE Syst. J. **17**, 2355–2365 (2022)
10. Liu, J., Chen, Y., Wang, Y., Chen, X., Cheng, J., Yang, J.: Monitoring vital signs and postures during sleep using wifi signals. IEEE Internet Things J. **5**(3), 2071–2084 (2018)

Doppler radar is widely used in the field of sleep detection due to its excellent ability to measure target displacement remotely. Doppler radar can capture the information of chest displacement due to respiration or heartbeat through the transmitted microwave signals and analyze it through the Doppler effect [19]. A contactless system named PRMS using quadrature microwave doppler radar to monitor sleep apnea events in real time. The system contains a real-time actigraphy and sleep apnea detection algorithm [20]. A novel sleep posture recognition technique is proposed, which employs classifiers that are amenable to optimization through Bayesian hyperparameter tuning. These classifiers operate on data from a dual-frequency monostatic continuous-wave radar system [21]. DopplerSleep, a contact sleep sensing system, uses a single Doppler sensor to track sleep quality. DopplerSleep can monitor both body movements and tiny chest and heart movements, and the system has been experimentally validated to perform well on sleep stage classification tasks [22].

RF signals are widely used for contactless motion and vital signs monitoring in the field of sleep monitoring. Radio Frequency Identification (RFID) is a contactless communication technology that enables two-way data exchange for identification and data transfer using RF signals with flexibility and low cost. A respiration monitoring system with RFID sensors called LungTrack is proposed to achieve dual objective monitoring with an accuracy of above 93% for two targets at a distance of 10 cm at least [23]. TagSleep is a sleep posture recognition system using the concept of two-layer sensing with RFID sensors [24]. A model combining a convolutional network and recurrent neural network is trained on the RF-measured sleep dataset with an adversarial training regime [25].

4 Comparative Analysis

WiFi sensors and other wireless sensors, as non-interference devices, offer both advantages and disadvantages in sleep monitoring. Figure 1 shows a comparison between these two methods. WiFi sensors typically utilize wireless signals and receivers to track variables such as breathing, body movement, and sleeping positions. These sensors analyze movement patterns and breathing rates by observing changes in WiFi signals. They are cost-effective and easy to deploy, but privacy concerns may arise.

On the other hand, radar technology emits high-frequency pulse signals and measures the time it takes for the signals to bounce back. This enables accurate positioning and tracking of objects, including monitoring human movements and breathing patterns during sleep. Radar provides precise distance and position measurements, boasting high accuracy and reliability. However, radar requires specialized hardware and incurs higher costs. Both UWB and doppler radars described previously are capable of real-time sleep monitoring with a high degree of accuracy, but there is the problem of higher equipment costs and more demanding deployment conditions during equipment placement.

While RFID technology offers advantages like low power consumption and affordability, it may have limitations when it comes to more detailed sleep analysis and breathing monitoring.

wifi sensors are used for obstructive sleep apnea (OSA) detection and rhythmic movement disorder (RMD) detection. An intelligent apnea monitoring system can utilizing linear fitting and wavelet transform eliminate the phase error of CSI. The system uses commodity wifi, which is better able to eliminate interference from changes in sleeping posture [13]. A sleep monitoring system named Wi-PSG is proposed to utilize CSI from Wifi infrastructures for RMD-related movement detection, which can achieve an accuracy of above 92% for different RMD movement classifications [14].

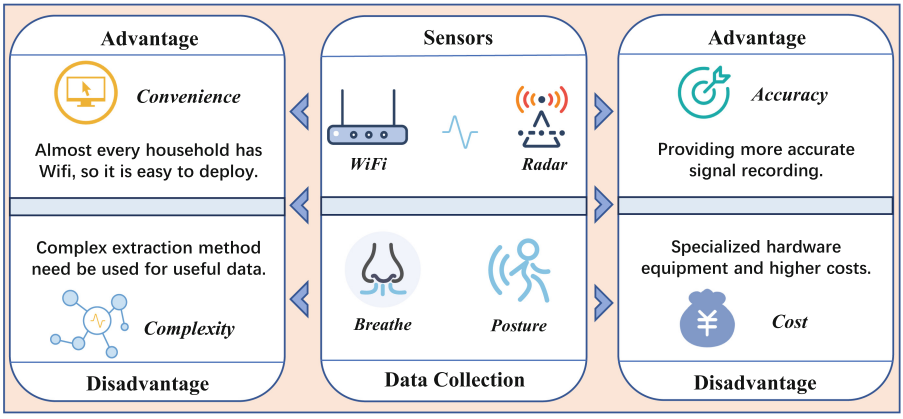

Fig. 1. Comparison between WiFi sensors and radars.

3 Sleep Detection Based on Wireless Sensor

Wireless radars are the most widely used sensors in sleep detection based on wireless sensors. Systems with wireless sensors are usually used for vital signs detection during sleep and sleep quality detection. The main sensors used in these systems are Ultrawideband (UWB) radar, Doppler radar, and Radio Frequency (RF) sensors.

UWB radar is commonly utilized for precise localization, employing low energy levels for short-range and high-bandwidth communications across the radio spectrum [15]. The required sleep information can be extracted by the UWB radar sensor penetrating the clothes and quilt. A fine-grained prototype for overnight respiration monitoring is proposed by exploiting the complementarity between the amplitude and phase of the radar signal [16]. Four respiration patterns are recognized during overnight sleep in this method. Another image processing method converts the raw signals collected by the UWB radar into a 2-D heatmap image and then an image-processing algorithm is used to capture respiratory information for respiratory motion measure [17]. An attention-based LSTM model is proposed to use the vital signs detected remotely by an impulse-radio UWB radar for sleep stage classification [18].

[6], However, the recording of PSG always needs expensive equipment and keep lots of contact with the subjects' body which bring discomfort. These drawbacks make it unsuitable for daily life sleep monitoring.

With the development of information techniques, more and more wireless sensors are used for sleep monitoring. There are already a lot of wearable devices used for sleep monitoring, but they also face resistance because of the discomfort brought to subjects and instability during sleep. Contactless sensors can effectively address the problem that invasive sensors bring natural sleep difficulties. There are various contactless sensors used in sleep monitoring now. The main of them are wireless sensors. Wifi sensor is also a kind of wireless sensor but since it has received more attention than other wireless sensors, it is put in a separate category.

Since wireless sensors are now widely used in sleep monitoring and have shown great potential, it is meaningful to review sleep monitoring research based on wifi sensors and wireless sensors. This can help develop contactless devices to achieve stable, safe, and non-contact sleep detection. In this work, we will first review the main sleep detection methods based on wifi sensors and wireless sensors respectively, and then a comparative analysis is made to summarize the difference between wifi sensors and wireless used in sleep monitoring. Finally, we provide a conclusion of our work.

2 Sleep Detection Based on Wifi Sensor

Wifi-based sleep monitoring activities are generally carried out through high precision indoor positioning, and the commonly used methods include Received Signal Strength (RSS) and Received Signal Strength (CSI) [7]. With the development of the technology, the CSI technique has demonstrated greater stability and accuracy and has become the more mainstream method nowadays. While using wifi sensors for sleep monitoring, CSI can be used to capture the effect of sleep activity contained by the Wifi signals [8].

Existing methods that use Wifi sensors to monitor sleep quality include heart rate monitoring and respiration monitoring [9]. A method is proposed to track the breathing rate and heart rate during sleep with Wifi [10]. They exploit to utilize the fine-grained channel information of existing Wifi networks to extract the minute movements that come with breathing and heartbeats. Wifi network activity is also used in a sleep-tracking approach called SleepMore which utilizes machine learning methods [11]. SleepMore constructs a semi-personalized random forest model to make a classification of the network activity behavior and the results are divided into sleep and awake states in minute dimensions. The experimental results show that SleepMore achieves an indistinguishable result with the Oura ring baseline within a 5% uncertainty rate.

Wifi sensors are also used for sleep stage classification and sleep-related disorders detection. An advanced signal processing and fusion method is proposed to extract accurate respiration and body movement for four-stage sleep classification, which achieves an accuracy of 81.1% [12]. In disorders monitoring,

Review of Sleep Monitoring Research Based on Wireless Sensor

Yuzhu Hu[1,2,3], Jian Chen[1,2,3], Shen Zhao[1(✉)], Kexin Tan[2], Kuai Yu[2], and Wei Wang[2,3,4]

[1] School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen 518000, China
{huyzh27,chenj589}@mail2.sysu.edu.cn, z-s-06@163.com
[2] Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China
{1120200259,1120200296}@smbu.edu.cn, ehomewang@ieee.org
[3] Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Shenzhen MSU-BITUniversity, Shenzhen 518172, Guangdong, China
[4] School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China

Abstract. Since sleep quality is crucial to human health, sleep monitoring has become a hot spot in the field of smart healthcare. Previous methods depend on polysomnography and wearable devices need immediate contact with the subject, which brings discomfort. Contactless sensors can address this issue. The most common contactless sensors used in sleep monitoring are wireless sensors (including radar and WiFi). To clarify the research in this area, we summarized the existing sleep monitoring methods based on WiFi sensors and wireless radar and made a comparison. The conclusion shows that the two kinds of methods have advantages and disadvantages, so the development of complementary methods is very promising for sleep monitoring.

Keywords: Sleep monitoring · contactless sensors · wireless sensing

1 Introduction

Sleep is one of the most important basic life activities of human beings, and it is also an important basis for maintaining physical and mental health [1]. Chronic poor sleep has also been linked to cardiovascular disease, obesity, and even some mental health problems [2–4]. Therefore, sleep monitoring is important for health status monitoring and is now become a hot topic for research.

Polysomnography (PSG) is the most widely used tool to monitor sleep, and it is regarded as the gold standard to detect sleep-related breathing disorders [5]. PSG can provide comprehensive information on sleep stages on the basis of Electroencephalography (EEG) activity, eye movements, and muscular tension

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 79–84, 2024.
https://doi.org/10.1007/978-3-031-65126-7_8

15. Vos, T., et al.: Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. The Lancet **396**(10258), 1204–1222 (2020)
16. Walker, E.R., McGee, R.E., Druss, B.G.: Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry **72**(4), 334–341 (2015)

developing countries (Fig. 2); anxiety contributes to the development of depression and must be taken into account as well.

Considering all above we can suggest the authorities to take more measures to ease the burden and stress of the depraved people. As other studies showed [4,6,8], low-income group are at the higher risk of getting depression and having worse health condition in general [3,5], so, some government financial help is better be provided (subsidiaries, money allowance, etc.).

Acknowledgment. This work is supported by the Shenzhen Science and Technology Innovation Commission (Stabilisation Support Programme).

References

1. Chesney, E., Goodwin, G.M., Fazel, S.: Risks of all-cause and suicide mortality in mental disorders: a meta-review. World Psychiatry **13**(2), 153–160 (2014)
2. Cuijpers, P., Vogelzangs, N., Twisk, J., Kleiboer, A., Li, J., Penninx, B.W.: Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. Am. J. Psychiatry **171**(4), 453–462 (2014)
3. Diener, E., Biswas-Diener, R.: Will money increase subjective well-being? Soc. Indic. Res. **57**, 119–169 (2002)
4. Dwyer, R.J., Dunn, E.W.: Wealth redistribution promotes happiness. Proc. Natl. Acad. Sci. **119**(46), e2211123119 (2022)
5. Headey, B., Muffels, R., Wooden, M.: Money does not buy happiness: or does it? a reassessment based on the combined effects of wealth, income and consumption. Soc. Indic. Res. **87**, 65–82 (2008)
6. Kahneman, D., Deaton, A.: High income improves evaluation of life but not emotional well-being. Proc. Natl. Acad. Sci. **107**(38), 16489–16493 (2010)
7. Kartaev, P.S.: How to teach econometrics to economists: bachelor level..... 72 macroeconomic policy. Sci. Res. Fac. Econ. Electron. J. **11**(2), 72–90 (2019)
8. Killingsworth, M.A.: Experienced well-being rises with income, even above $75,000 per year. Proc. Natl. Acad. Sci. **118**(4), e2016976118 (2021)
9. Patel, V., et al.: Addressing the burden of mental, neurological, and substance use disorders: key messages from disease control priorities. The Lancet **387**(10028), 1672–1685 (2016)
10. Pearce, M., et al.: Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA Psychiatry **79**, 550–559 (2022)
11. Reger, M.A., Stanley, I.H., Joiner, T.E.: Suicide mortality and coronavirus disease 2019-a perfect storm? JAMA Psychiatry **77**(11), 1093–1094 (2020)
12. Son, J., Shin, J.: Bimodal effects of sunlight on major depressive disorder. Compr. Psychiatry **108**, 152232 (2021)
13. Stock, J.H., Watson, M.W.: Introduction to Econometrics, vol. 104. Addison Wesley Boston (2003)
14. Viswanathan, M., et al.: Screening for depression and suicide risk in children and adolescents: updated evidence report and systematic review for the us preventive services task force. JAMA **328**(15), 1543–1556 (2022)

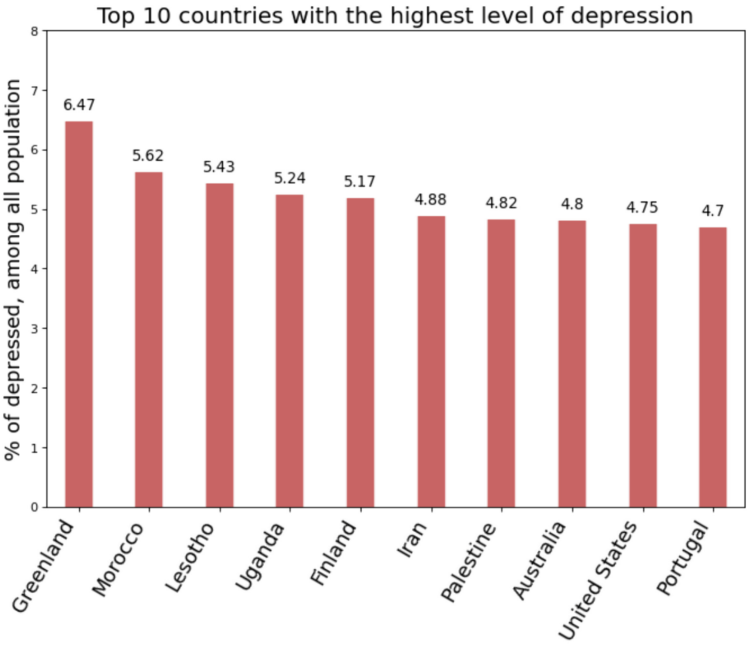

Fig. 2. Top 10 countries with the highest level of depression

As can be noticed, top ten includes mainly developing countries where GDP per capita quite small. The exceptions are Finland, Australia, United States, Portugal and Greenland as a special region of Denmark, where the GDP per capita is medium or higher. In case with Finland and Greenland, such higher level of depression could be explained by two factors: 1) isolated and low populated communities, as can be seen from the depression model, the population size is significant factor; 2) the lack of sunny days, what negatively effects on mood and emotional conditions [12]. As for other countries, the further deep analysis is required.

4 Discussion

This large-scale study based on worldwide panel data about depression showed that people who live in countries with low GDP per capita are more vulnerable to depression. We find that the relationship between depression and GDP per capita is strongly negative, and because of analyses of huge massive of date, the results are universal. At the same time, the connection between depression and anxiety disorders is strongly positive, thus, the following conclusions could be made: in countries with lower GDP per capita, more people tend to suffer from depression. Actually, this fact can be proved even statistically: the majority of the countries in the list of top 10 countries with high level of depression are

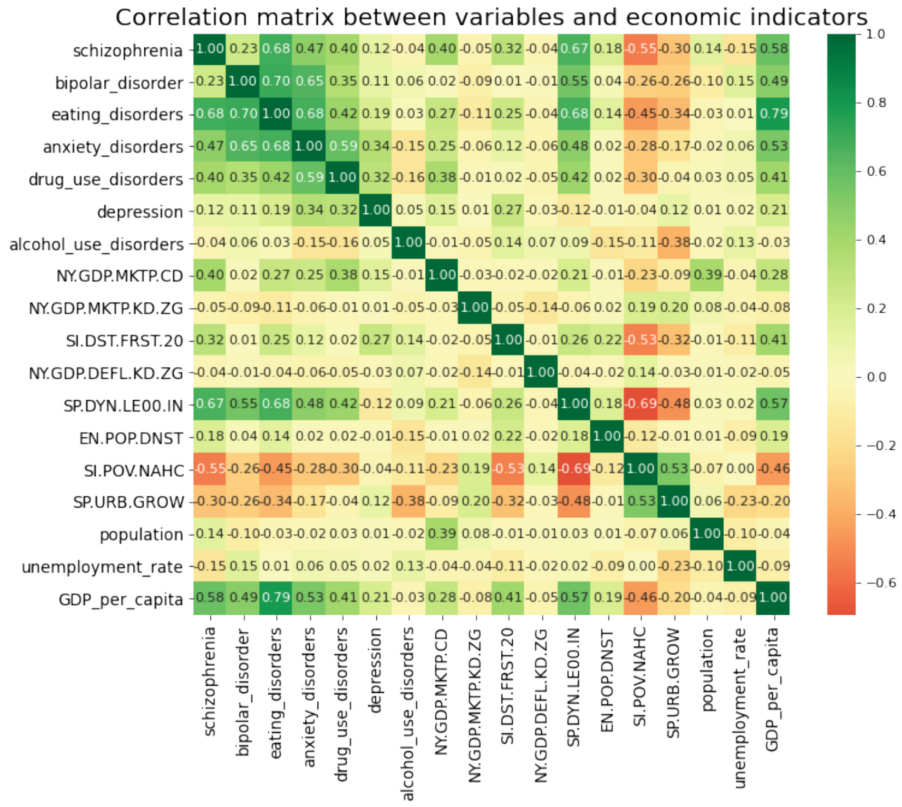

Fig. 1. Correlation matrix between variables and economic indicators

- GDP per capita - all things being equal, with an increase in GDP by one dollar, the number of people suffering from depression decreases by 2,74*e-6%
- Income share held by lowest 20% - all things being equal, with an increase in Income share held by lowest 20% by one dollar, the number of people suffering from depression decreases by 0,024%
- Life expectancy at birth, total - all things being equal, with an increase in life expectancy at birth, by one year, the number of people suffering from depression decreases by 0,013%
- Population - all things being equal, with an increase in population by one people, the number of people suffering from depression decreases by 1,05*e-9%
- Anxiety disorders - all things being equal, with an increase in anxiety disorders by one percent, the number of people suffering from depression increases by 0,35%

Now countries that have the highest rates of depression are shown in Fig. 2.

– SI.POV.NAHC - Poverty headcount ratio at national poverty lines (% of population)
– SP.URB.GROW - Urban population growth (annual %)
– Unemployment - Unemployment rate, (% of work force)
– GDP_PER_CAPITA - GDP per capita, (current US$)

2.3 Panel Study

In order to avoid omitted variable bias, we took regressors from different spheres [7,13]: pure economic, social and urban. We also have added the variable of control - Anxiety, as, by all means, anxiety disorders influence on the development of depression and other mental disorders. We conducted all measures using special econometric program Gretl.

3 Results

We calculated the correlation between all mental disorders and economic indicators as Fig. 1.

At the same time, we got the following depression model as Tabel 1.

Table 1. Depression Model

	Coefficient	St. error	t-statistics	p-value	
const	325,388	0,707587	4,599	¡0,0001	***
anxiety disorders	0,354180	0,145535	2,434	0,0168	**
NYGDPMKTPCD	0,000000	0,000000	4,352	¡0,0001	***
NYGDPMKTPKDZG	0,00112238	0,00100377	1,118	0,2663	
SIDSTFRST20	−0,0239324	0,00750013	−3,191	0,0019	***
NYGDPDEFLKDZG	−0,000230614	0,000412541	−0,5590	0,5775	
SPDYNLE00IN	−0,0134092	0,00500963	−2,677	0,0088	***
ENPOPDNST	0,000163380	0,000295864	0,5522	0,5821	
SIPOVNAHC	−0,00107081	0,00145502	−0,7359	0,4636	
SPURBGROW	−0,00724004	0,00900659	−0,8039	0,4235	
population	−1,04580e−09	1.73E−05	−6,051	< 0,0001	***
unemployment_rate	−0,00358409	0,00199508	−1,796	0,0756	*
GDP_PER_CAPITA	−2,74364e−06	9.21E−02	−2,978	0,0037	***

The LSDV R-square for this model is 0,9956, '*' means that variable is significant on 10%, '**' - 5%, and '***' - 1%. Therefore, we could interpret four variables of interest (on 5%):

World Health Association, around 280 million of people worldwide are suffering from depression, moreover, the World Health Organization assumes that 5% of men and 9% of female experience depressive disorders in their lifetime [10, 15]. Depression can lead to the development of other illnesses what effect on premature mortality [1,2,16] and even increase the suicide rates [9,11,14], that is why it is crucial for authorities to be aware of development of such illnesses. The innovation of this work is that it includes factors and figures from different spheres and examine their impact on the development of depression and other mental disorders. This allows us to broaden our thinking and to make more clear judgments [7,13]. Particularly, in addition to social-economic indicators, we also added urban population growth in our list of economic indicators, what allows to see the big picture. This article is aimed to determine how the main economic indicators are connected with mental disorders. After establishing the relationships, it will be possible to judge whether the country at the risk of mass depression. We believe that with the help of our research local authorities will be able to identify the upcoming health threats more effectively, and, what is the key point, much earlier, thus, many human lives would be improved or even saved.

2 Methods

This is a panel study which includes data from 196 countries throughout 27 years. In our research we mainly used econometrics and ordinary least square (OLS) analysis to make proper models. All implemented models have passed the Ramsey Test, the check for heteroscedasticity and multicollinearity, thus, all described models are trustful. Besides, in case with the depression analysis, the Fixed Effects model was used due to take into account each country peculiarity [7,13].

2.1 Dependent Variables

In addition to Depression, we also considered the following types of mental diseases: Schizophrenia, Bipolar disorder, Eating disorders, Anxiety disorders, Drug use disorders, Alcohol use disorders. All variables are examined as % of all population.

2.2 Economic Indicators

For each variable we make an econometric model with the following regressors:

- NY.GDP.MKTP.CD - GDP (current US$)
- NY.GDP.MKTP.KD.ZG - GDP growth (annual %)
- SI.DST.FRST.20 - Income share held by lowest 20%
- NY.GDP.DEFL.KD.ZG - Inflation, GDP deflator (annual %)
- SP.DYN.LE00.IN - Life expectancy at birth, total (years)
- EN.POP.DNST - Population density (people per sq. km of land area)

Identification of Economic Factors for Mass Depression Based on Panel Study and Machine Learning

Iaroslava Pravolamskaya[1,2,3,4], Jian Chen[3,4,5], and Wei Wang[3,4,6(✉)]

[1] Faculty of Economics, Shenzhen MSU-BIT University, Shenzhen 518172, China
[2] Faculty of Economics, Lomonosov Moscow State University, Moscow 119991, Russia
[3] Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China
[4] Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Shenzhen, MSU-BITUniversity, Shenzhen 518172, Guangdong, China
[5] School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen 518000, China
chenj589@mail2.sysu.edu.cn
[6] School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China
ehomewang@ieee.org

Abstract. Panel study and machine learning are important tools for analyzing various aspects of the economy. They allow researchers to study the dynamics of changes in different economic indicators, such as GDP, inflation, unemployment, etc. In addition, these tools can be used to determine causal relationships between social, economic and psychological factors what can allow us to predict the development of the economy and changes in people's life in the future. However, previous works in this sphere studied the connections between income and happiness, not taking into account the relationships between economic indicators and mental disorders. This article is aimed to analyze the relationship between economic factors and the level of mass depression based on a panel study and machine learning methods. Experimental results based on panel study and machine learning demonstrate effectiveness of our proposed econometric model.

Keywords: Panel Study Machine Learning Depression Identification Economic Factors Econometrics Models

1 Introduction

The increasing number of people suffering from depression and other mental diseases is one of the most challenging issues in the 21 century. According to

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 72–78, 2024.
https://doi.org/10.1007/978-3-031-65126-7_7

14. Koyama, K., Hoshikawa, H., Kojima, G.: Eddy current nondestructive testing for carbon fiber-reinforced composites. J. Press. Vessel. Technol. **135**(4), 041501 (2013)
15. Kostopoulos, V., Vavouliotis, A., Karapappas, P., Tsotra, P., Paipetis, A.: Damage monitoring of carbon fiber reinforced laminates using resistance measurements. Improving sensitivity using carbon nanotube doped epoxy matrix system. J. Intell. Mater. Syst. Struct. **20**(9), 1025–1034 (2009). https://doi.org/10.1177/1045389X08099993

conducted to get the correct data and then compared to draw conclusions. Damage to the carbon/glass blend and fracture of the carbon fibers was observed by using Three Point Bending method. The pictures show that where pressure is applied the upper layers are damaged by shear stresses leading to kinking and the lower layers are damaged mainly in the form of delamination leading to failure.

In conclusion, this study has been designed, experimented and concluded that it is feasible to monitor the electrical conductivity of this hybrid carbon/glass fiber blend and that this composite fiber can also be seen as a self-sensor.

Acknowledgments. This project is supported by the funding of Guangdong Province Key Laboratory of Intelligent Detection in Complex Environment of Aerospace, Land and Sea. (2022KSYS016).

References

1. Maleque, M.A., Salit, M.S.: Materials Selection and Design. SM, Springer, Singapore (2013). https://doi.org/10.1007/978-981-4560-38-2
2. Jalalvand, M., Czél, G., Wisnom, M.R.: Damage analysis of pseudo-ductile thin-ply UD hybrid composites – A new analytical method. Compos. A Appl. Sci. Manuf. **69**, 83–93 (2015). https://doi.org/10.1016/j.compositesa.2014.11.006
3. Sauer, M.: Composites Market Report 2019—The Global CF-und CC-Market 2019: Market Developments, Trends, Outlook and Challenges. Composites United eV, Berlin, Deutschland (2019)
4. Czél, G., Wisnom, M.R.: Demonstration of pseudo-ductility in high performance glass/epoxy composites by hybridization with thin-ply carbon prepreg. Compos. A Appl. Sci. Manuf. **52**, 23–30 (2013)
5. David-West, O., et al.: A review of structural health monitoring techniques as applied to composite structures. Struct. Durability Health Monit. **11**(2), 91–147 (2017)
6. Rev, T., et al.: A simple and robust approach for visual overload indication-UD thin-ply hybrid composite sensors. Compos. A Appl. Sci. Manuf. **121**, 376–385 (2019)
7. Chapuis, B.: Introduction to structural health monitoring. In: Chapuis, B., Sjerve, E. (eds.) Sensors, Algorithms and Applications for Structural Health Monitoring. IC, pp. 1–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-69233-3_1
8. Vavouliotis, A., Paipetis, A., Kostopoulos, V.: On the fatigue life prediction of CFRP laminates using the electrical resistance change method. Compos. Sci. Technol. **1**(5), 630–642 (2011)
9. Smith, R.: Composite defects and their detection. Mater. Sci. Eng. **3**(1), 103–143 (2009)
10. Gregor Trtnik, M.G.: Recent advances of ultrasonic testing of cement based materials at early ages. Ultrasonics **54**, 66–75 (2013)
11. Song, S., Jing, J., Cheng, W.: Online monitoring system for macro-fatigue characteristics of glass fiber composite materials based on machine vision. IEEE Trans. Instrum. Meas. **71**, 1–12 (2022)
12. Manjunatha, P.A.: Vision-based and data-driven analytical and experimental studies into condition assessment and change detection of evolving civil, mechanical and aerospace infrastructures. Doctoral dissertation, University of Southern California (2022)
13. Bayraktar, E., Antolovich, S.D., Bathias, C.: New developments in non-destructive controls of the composite materials and applications in manufacturing engineering. J. Mater. Process. Technol. **206**(1–3), 30–44 (2008). https://doi.org/10.1016/j.jmatprotec.2007.12.001

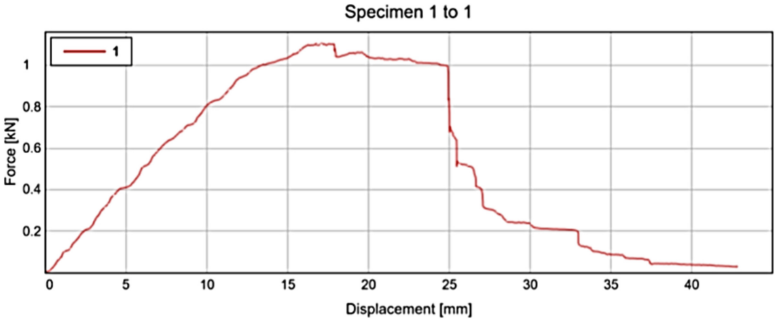

Fig. 11. 10 mm experimental group displacement and Force curve

to the phenomenon of pseudo-stretchability. Analysis of the sample damage showed that shear damage dominated at the upper end of the sample, while delamination dominated from the middle to the lower plies, shown as Fig. 12. However, the main change in resistance in this test was due to the fracture of the thin carbon fibers, which was mainly due to tensile stresses, while the delamination of the lower plies was mainly due to shear stresses, so the design of this test was reasonable.

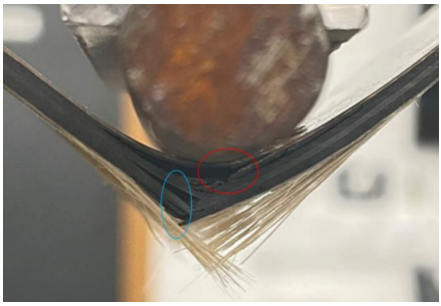

Fig. 12. Injury patterns under three -point bending

6 Conclusion

In this project, a hybrid thin-layer carbon/glass fiber self-sensing method is proposed. It is innovative in that it changes the traditional case of applying the carbon fibers directly to the object to be sensed. Also, by using an S -shape instead of the traditional direct strip, it allows for greater coverage and a larger area to be monitored than just partial detection, while its more holistic nature makes it more effective for monitoring a whole plane rather than monitoring a broken location, and also has a greater improvement in monitoring the effects of certain unseen damage. Regarding the experiment, this experiment uses the controlled variable method to create differences for different variables. By designing groups of different widths as well as different styles, several experiments were

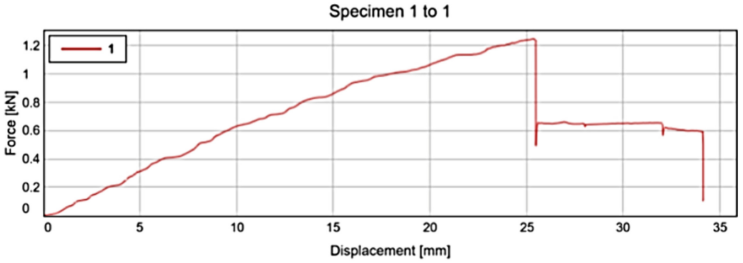

Fig. 9. 5 mm experimental group displacement and Force curve

in the 25–30 mm zone, the resistance begins to rise, indicating that the carbon fiber body is further destroyed in this zone, when in the 30–35 zone, the carbon fiber is completely destroyed and the resistance rises to 4000 Ω (Fig. 8). When the carbon fibers are completely destroyed, the loading force is removed and the fibers spring back, at this time some of the fibers reduce in resistance because the stress is reunited (Fig. 9).

10 mm Bending Test

The 10 mm three-point bending test is also primarily a comparison with 5 mm, observing the change in resistance of two different widths of carbon fiber to determine which is more appropriate.

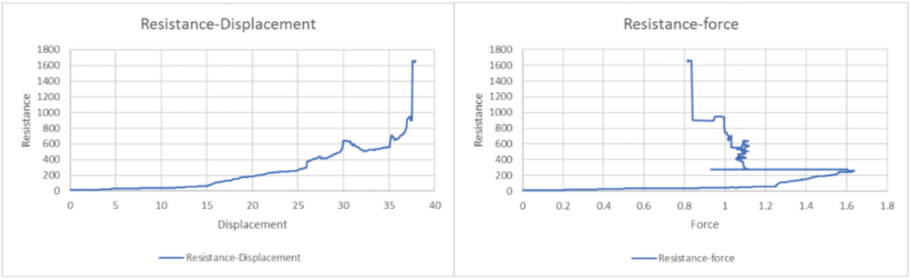

Fig. 10. 10 mm experimental group resistance displacement curve(left) and 10 mm experimental group resistance Force curve(right)

By comparing the two sets of plots, it can be found that the 5 mm images of resistance and Force are relatively similar to the 10 mm images on the three -point bending method, but on the resistance displacement curve, it is obvious that the rising trend of the 10 mm curve is smoother, so after the comparison, it is more recommended to use a 10 mm wide thin layerof carbon fiber as a self-sensor (Figs. 10 and 11).

Damage Mode Analysis

In this section, the damage pattern of the experimental product and the image in the above figure will be analyzed in detail, as the damage to the sample occurred gradually over the course of the test and this fiber hybridization slowed down the catastrophic rate and so led

Table 1. Mechanical properties of curing

Properties	Numerical value	Unit
Tg Onset(DMA)	**140**	°C
Tensile Strength	**645**	MPa
Compressive Strength	**515**	MPa
Flexural Strength	**882**	MPa
Flexural Modulus	**60.1**	GPa
Interlaminar Shear Strength	**69.8**	MPa
Tg Peak(DMA)	**148**	°C

of plain carbon fiber strips alone. As the fiber orientation was also considered in this carbon fiber experiment to affect the magnitude of the current, a unidirectional thin layer of carbon fiber was used in this case so that the consistency of the current could be maintained throughout.

In the three-point bending test, since the three -point bending test causes large shear stresses, data were collected from the start to sample failure and finally the changes in resistance and the reasons for these changes were analyzed in conjunction with the changes in the curves.

5 mm Bending Test

In this section the experimental data on the 5 mm three-point bending method is described. Unlike the above, as this design is a hybrid design, the standard T700 carbon fiber bending performance criteria above can only be used as a reference value, so according to the experimental process, the bending performance is significantly lower compared to T700, only around 800 Mpa, so it is speculated that it is possible that the mixture of glass fiber and thin-layered carbon fiber has affected the bending performance.

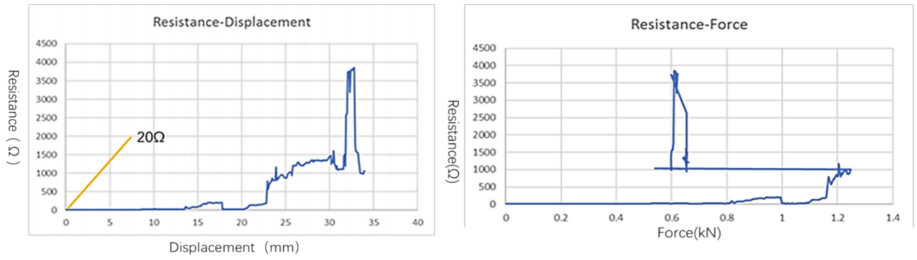

Fig. 8. 5 mm experimental group resistance displacement curve(top) and 5 mm experimental group resistance Force curve(bottom)

According to the data we can see that there is a relatively obvious increase in resistance after the indentation test, as can be seen from the graph, at 25 mm of the experiment is the maximum stress, when the carbon fiber begins to destroy, it can be concluded that

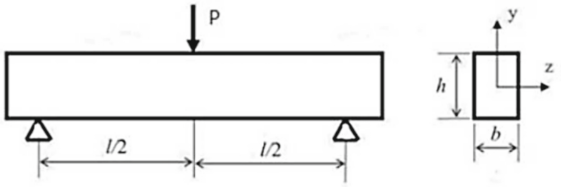

Fig. 7. Three-point bending method model

In the three-point bending test, when viewed from the front, it can simply be seen as a simply supported beam subjected to a concentrated pressure. Three-point bending should theoretically result in a linear distribution of positive stresses along the beam in the cross-sectional area when loaded.

$$\sigma = \frac{M}{I_z} y \tag{1}$$

where σ is the stress, M is the moment, Iz is the moment of inertia of the cross-section to the z-axis and y is the distance in the cross-section to the y-axis. The maximum positive stress at the danger point of the beam is:

$$\sigma_{Max} = \frac{M_{Max}}{I_z} y_{Max} \tag{2}$$

For rectangular section specimens:

$$M = \frac{P \times l}{4} \qquad I_z = \frac{bh^3}{12} \tag{3}$$

Substituting Eqs. (3) into (2) yields the new equation

$$\sigma_{bb} = \frac{3P \times l}{2bh^2} \tag{4}$$

where P is the load and L is the span, b is for width, h is for thickness.

In the case of a sample based on this equation, the maximum shear stress is calculated a

$$s\sigma_{bb} = \frac{3P \times l}{2bh^2} = \frac{3 \times 1.5\,KN \times 150\,mm}{2 \times 50\,mm \times 9\,mm^2} = 800\,Mpa \tag{5}$$

According to the Table 1, its standard value is 880 Mpa, However, as this design contains other fibers of different thicknesses or patterns, this data can only be used as a reference value for the main body of the sample, so in principle the maximum acceptable shear stress for this design should be lower than this value.

5 Results

This experiment focused on the fabrication process of the self-sensor, which was designed using an innovative mixture of carbon fiber and thin layers of E glass fiber, and investigated the advantages and differences between this combination and the use

widely used in destructive testing due to its simple construction and the fact that it does not require much manipulation. For this test, the sample is placed on a jig and a multimeter is connected to the two sections of copper to read the resistance data. The movement speed of 7 mm/min is entered into the control of the hydraulic press and the test is started (Fig. 4).

Fig. 4. Three -point bending test.

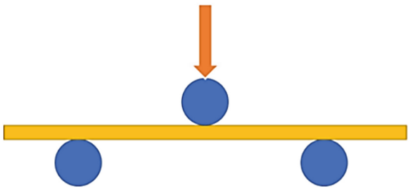

Fig. 5. Injury patterns under three-point bending

The three-point bending process is achieved mainly by applying pressure to the tip, during which the sample undergoes a process of gradual destruction. The following diagram shows the principle of three -point bending (Fig. 5).

During the three-point bending experiment, the sample started to break gradually when it was loaded to a high enough stress. As this sample was a mixed sample, the surface glass fiber started to break when it was loaded to 0.7 KN, the glass fiber broke completely when it was loaded to 1.2 KN, then the load was reduced to 0.6 KN and then the carbon fiber started to break gradually.

The images show that the entire damage process is produced gradually, and based on the experimental images it can be seen that the samples start with damage and end up with damage (Fig. 6).

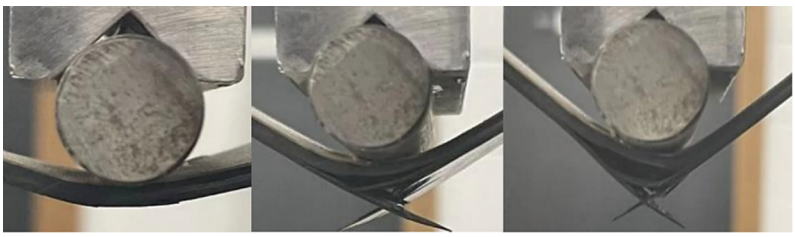

Fig. 6. The process of three-point bending test

Theory of Three-Point Bending Test

When bending deformation occurs in the three-point bending method, the fibers near the bottom elongate and those near the top shorten. According to the planar hypothesis, the fiber state changes gradually from stretching to compression along the height of the cross section from the bottom to the top, then there must be a layer in between where the length of the fiber remains constant, this layer is called the neutral layer (Fig. 7).

known as multi -directional (MD) fibers (Fig. 2). This multi -axial material has better tensile and compressive resistance than uniaxial material, but because it is manufactured at an angle, it is less malleable.

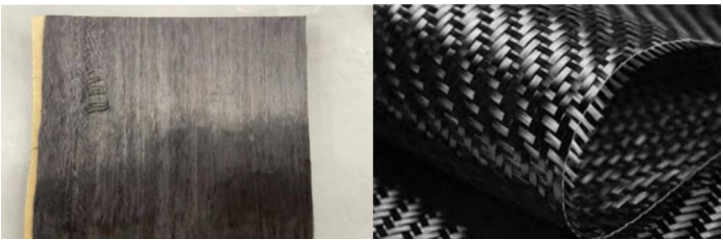

Fig. 2. UD carbon fiber(A), MD carbon fiber(B)

Experiment Test Group

The carbon fibers in the experimental group will be linked to each other, showing S-shaped connections, which then means that the data from the experimental group will affect each other. This control group can be used to see if the resistance will be affected by the occurrence of fiber breaks. Having established that the resistance will change due to fiber breakage, then this experimental group has the advantage that only two electrodes are needed to complete the experiment due to the large area it covers. The main reason for using two different widths of samples was to see the rate of change in resistance by comparing the two sizes of 5 mm and 10 mm. In the graph below, sample 1 is the control group of 10 mm, sample 2 is the control group of 5 mm, sample 3 is the experimental group of 5 mm and sample 4 is the experimental group of 10 mm, shown as Fig. 3.

Fig. 3. Kinds of simple.

Three-Point Bending Test

After the indentation test, the material is tested using the three-point bending method, which is one of the simplest and most effective methods of testing laminates and is

4 Research Methodology

It describes the methods and techniques used in the selection of materials, the preparation process, and the design of the testing process for this project. In particular, first, it will be described in detail what materials are used to prepare the samples, as well as the preparation process and methods. Then, it will be described the test process of the experiment and other equipment used in the experimental process, last, it is going to be described the detail of the whole experiment and the theoretical data will be given, including the theoretical currents and the theoretical stresses generated by the experiment.

Material Selection

In addition, for sample preparation we used T700/XC130 unidirectional prepreg carbon fiber as the sample body and S -Glass/913 and M46JB unidirectional prepreg thin carbon fiber as the sensor. The base material was made from T700 carbon fiber manufactured by Toray of Japan. When selecting the substrate material, it was considered that the main carbon fibers are mainly T300 and T700, both of which contain a large amount of carbon, but the overall performance of T700 is significantly better than that of T300. In the selection of the sensing layer, we chose to use a thin carbon fiber sandwiched between the two glass fibers. As the thin carbon fiber chosen, M46JB, has a similar tensile strength to T700, but obviously the compressive strength of M46JB is weaker than that of T700.T In this experiment, glass fibers were chosen to wrap the thin carbon fiber because, as seen in Meisam's model, there are three different damage modes for composites made from high and low strain materials, so in order for the sensing layer of carbon fibers to break before the glass fibers in the isolation layer, a thin layer of carbon fibers with a lower degree of strain than the glass fibers must be used as the induction material (Fig. 1). (Fotouhi, Jalalvand et al., 2017)

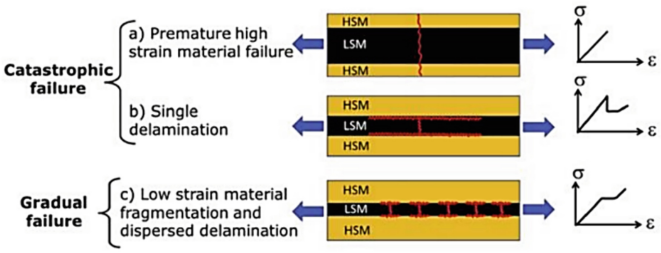

Fig. 1. Possible failure modes in a three layers UD hybrid made from HSM and LSM (red lines show fracture) (a) single crack through the whole specimen, (b) single crack in the LSM followed by instantaneous delamination, and (c) multiple fracture and localised stable pull-out of the LSM

Typically, the basic constituent material of a carbon fiber reinforced material is usually a combination of multiple or unidirectional fiber orientations, rearranged to provide different mechanical properties. A single unidirectional (UD) fiber arrangement, where all the fibers in the resin are aligned in one direction with no voids or breaks. Another type of arrangement is where the fibers are aligned at 0 and 90 degrees, this is

In fields such as aviation and construction, it is important to ensure safety margins, as sudden damage can potentially lead to injury or death as well as huge financial losses. In these important areas, sudden failures as well as small residual load capacities are not allowed. This is why higher safety margins and more conservative structural designs are the dominant design approach in current designs, while another problem with engineered materials is that they break down without prior detectable damage and warning [2].

2 Literature Review

The composite material is made by two or more components. Composite material can be made by using fiber that are cured within a resin. Using a combination of different materials, taking advantage of their strengths and reducing the impact of their weaknesses is an important idea in designing composite materials. The most common types are combining carbon fibers and glass fibers with a thermosetting resin to create either a CFRPs or GFRPs. The use of multiple fiber-reinforced polymer laminations to form a new composite material with enhanced mechanical properties, such as compressive and tensile resistance, and the consideration of how to efficiently monitor the new composite material, based on previous research by scientists, has become an important key, and finally some of the advantages and disadvantages of previous monitoring methods in relation to the composite material in this experiment are presented.

3 Aims and Objective

How to monitor the health of composite materials is an important current research issue. It can effectively contribute to the development of several areas where composites are used, such as aerospace engineering and the automotive industry. However, in the traditional composite fiber industry the damage itself is unpredictable and sudden as it can be caused by different kind of stresses and pseudo-plastic deformations. There are many different methods of detection, but most of them do not reflect the changes in the material in real time and are also too expensive. This project therefore proposes a method to monitor changes in material resistance based on the fact that the resistance of thin carbon fibers changes gradually during damage. Research done by Meisam has shown that damage to fibers varies linearly (Rev, Jalalvand et al. 2019) and therefore the health of the material can be monitored by detecting changes in resistance based on this theory. The aim of this project is to design another sensor based on resistance variation, to experiment in order to check if the sensor is usable, and to design and test the sensor in order to achieve the best possible results, observing through experimentation and results whether it is sensitive and efficient.

Research Objective 1. Research test samples and target sensors based on existing literature and conjecture.

Research Objective 2. Analysis of the change in resistance of the sample and its force and displacement profiles.

Research Objective 3. Using a hybrid S-shape to detect damage on a flat plane and resistance changes monitored using a carbon glass hybrid.

Structural Health Monitoring of Carbon Fiber Composite Lamination Using Electrical Resistance

Guiping Lu[1(✉)], Xiaofeng Zhang[1], Shan Lu[2], Binghua Su[1], Kejun Wang[1], and Jiaran Liang[1]

[1] Beijing Institute of Technology, Zhuhai, Zhuhai, China
344088386@qq.com
[2] BMW Brilliance Automotive Ltd, Shenyang, China

Abstract. It focuses on a composite material made of glass and carbon fibers in this paper. The composite can be actively monitored and controlled by the self-sensing of the carbon fibers. However, due to the high stiffness and brittleness of the composite material, damage often occurs instantaneously. It is difficult to monitor damage patterns and control damage through factors such as fiber type variables and displacement relationships. This is why monitoring the health of composite fibers is an important direction, which has major implications for the aerospace, industrial and automotive sectors. In this project, the main focus is to monitor the electrical conductivity of carbon fibers online by breaking thin layers and observing the changes in their conductivity, and to understand changes in condition through changes in current. In addition, the composite design of this project can be applied to the monitoring of large planar materials, as well as to applications in important areas such as aerospace. In making further comparisons, it can be seen that the 5 mm thin layer of carbon fiber is more sensitive in the process of self-sensing, while the change in resistance is more noticeable when damage is received in the period.

Keywords: Structural Health Monitoring · Carbon Fiber Composite Lamination · Electrical Resistance · Three-point bending test

1 Introduction

There are more and more high-tech products made of composite materials, such as aerospace or automotive, and even the latest batteries. They are becoming increasingly popular due to their outstanding properties, such as their high strength, low weight and fatigue resistance, Composites are combinations of two or more materials with different physical behavior and chemical states. In particular in this test, the materials used are fiber reinforced polymers. As the properties of composites are usually more variable, engineers consider their design structure and components to minimize failure during the design of composites [1].

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 61–71, 2024.
https://doi.org/10.1007/978-3-031-65126-7_6

5. Carrard, V., et al.: Medical student mental health. https://www.kaggle.com/datasets/thedevastator/medical-student-mental-health
6. Davis, C., Martin, G., Kosky, R., O'Hanlon, A.: Early intervention in the mental health of young people: a literature review. In: ERIC (2000)
7. Ediz, B., Ozcakir, A., Bilgel, N.: Depression and anxiety among medical students: examining scores of the beck depression and anxiety inventory and the depression anxiety and stress scale with student characteristics. Cogent Psychol. **4**(1), 1283829 (2017)
8. Eva, E.O., et al.: Prevalence of stress among medical students: a comparative study between public and private medical schools in bangladesh. BMC. Res. Notes **8**(1), 1–7 (2015)
9. Ge, F., Zhang, D., Wu, L., Mu, H.: Predicting psychological state among chinese undergraduate students in the covid-19 epidemic: a longitudinal study using a machine learning. Neuropsychiatric Disease Treat. **16**, 2111–2118 (2020)
10. Ghrouz, A.K., Noohu, M.M., Dilshad Manzar, M., Warren Spence, D., BaHammam, A.S., Pandi-Perumal, S.R.: Physical activity and sleep quality in relation to mental health among college students. Sleep Breathing **23**, 627–634 (2019)
11. Henry, S.K., Grant, M.M., Cropsey, K.L.: Determining the optimal clinical cutoff on the CES-d for depression in a community corrections sample. J. Affect. Disord. **234**, 270–275 (2018)
12. Jungbluth, C., MacFarlane, I.M., Veach, P.M., LeRoy, B.S.: Why is everyone so anxious?: an exploration of stress and anxiety in genetic counseling graduate students. J. Genet. Couns. **20**(3), 270–286 (2011)
13. Mao, Y., Zhang, N., Liu, J., Zhu, B., He, R., Wang, X.: A systematic review of depression and anxiety in medical students in china. BMC Med. Educ. **19**(1), 1–13 (2019)
14. McGorry, P.D., Killackey, E.J.: Early intervention in psychosis: a new evidence based paradigm. Epidemiology Psychiatric Sci. **11**(4), 237–247 (2002)
15. Moutinho, I.L.D., et al.: Depression, stress and anxiety in medical students: a cross-sectional comparison between students from different semesters. Revista da Associação Médica Brasileira **63**, 21–28 (2017)
16. Womble, M., Jennings, S., Schatz, P., Elbin, R.: A-173 clinical cutoffs on the state-trait anxiety inventory for concussion. Arch. Clin. Neuropsychol. **36**(6), 1228–1228 (2021)

Regarding depression: Concerning study duration, age, academic efficacy scores from the MBI questionnaire, and QCAE affective empathy scores, the corresponding p-values are 0.851, 0.626, 0.578, and 0.405, all significantly greater than 0.05. From a statistical standpoint, this indicates that these features do not manifest significant differences at the given level. In other words, we lack sufficient evidence to support significant relationships or disparities between these features and anxiety.

However, in the context of gender, history of psychological counseling, native language, health status, and MBI Cynicism scores, the corresponding p-values are all below 0.05. This implies that these features may possess some degree of correlation, association, or influence with anxiety. In terms of statistical analysis, these divergences suggest that these features might hold a certain impact or role in relation to anxiety emotions.

5 Conclusion

In this study, we investigated the statistical relationships between various factors and the occurrence of psychological disorders, revealing patterns of variation in the proportions of individuals affected by psychological disorders and the severity of these disorders across different populations. We identified several factors closely associated with psychological disorders, with gender, native language, and health status potentially exhibiting more significant correlations with anxiety and depression.

Nevertheless, our study does have certain limitations. The size of the dataset is relatively small, and the number of features is limited, which could potentially impact the accuracy of our conclusions. To arrive at more universally applicable conclusions, we require a more comprehensive dataset of medical student information and a larger sample size.

Acknowledgment. This work is supported by the Shenzhen Science and Technology Innovation Commission (Stabilisation Support Programme).

References

1. Ahad, A., Chahar, P., Haque, E., Bey, A., Jain, M., Raja, W.: Factors affecting the prevalence of stress, anxiety, and depression in undergraduate indian dental students. J. Educ. Health Promot. **10**, 266 (2021)
2. Al-Dabal, B.K., Koura, M.R., Rasheed, P., Al-Sowielem, L., Makki, S.M.: A comparative study of perceived stress among female medical and non-medical university students in dammam, saudi arabia. Sultan Qaboos Univ. Med. J. **10**(2), 231 (2010)
3. Behere, S.P., Yadav, R., Behere, P.B.: A comparative study of stress among students of medicine, engineering, and nursing. Indian J. Psychol. Med. **33**(2), 145–148 (2011)
4. Carrard, Valerie, C., et al.: The relationship between medical students' empathy, mental health, and burnout: a cross-sectional study. Med. Teacher **44**(12), 1392–1399 (2022)

The presence of a job or a partner appears to have limited influence on anxiety or depression.

4.2 Correlation Analysis

Through t-tests and chi-squared tests, we will determine features that exhibit robust correlations with anxiety and depression, as well as those with weaker correlations.

Table 2. Demographic characteristics of college students - Anxiety and Depression

Variable	Anxiety p-Value	Depression p-Value
Gender	0.000***	0.000***
Job	0.439	0.018**
Part	0.242	0.012**
psyt	0.000***	0.000***
year	0.083*	0.084*
age	0.76	**0.626**
glang	0.000***	0.000***
stud_h	**0.987**	**0.851**
health	0.001***	0.000***
qcae_cog	0.216	0.15
qcae_aff	0.074*	**0.405**
mbi_ex	**0.587**	0.053*
mbi_cy	0.005***	0.000***
mbi_ea	**0.529**	**0.578**

Note: ***, **, * represent significance levels of 1%, 5%, and 10%, respectively.

Based on the results from Table 2, the following insights can be derived:

For anxiety disorder: In the examination of study duration, the significance p-value is 0.987; concerning the emotional exhaustion and academic efficacy scores from the MBI questionnaire, the respective p-values are 0.587 and 0.529; regarding the presence of a job, the p-value is 0.439. These outcomes indicate that statistically, the aforementioned features do not exhibit significant differences at the given level. In other words, we lack sufficient evidence to support a significant association or disparity between these features and anxiety.

However, when considering features such as gender, history of psychological counseling, native language, and health status, all respective p-values are below 0.05. This suggests the potential existence of some degree of correlation, association, or influence between these features and anxiety. This statistical divergence implies that these features might have a certain impact or role in relation to anxiety emotions.

Statistical Analysis of Other Variables' Relationship with Psychological Disorders. The statistical graphs illustrating the proportions of anxiety and depression, as well as the average scores of the affected population, varying with different features, are presented in Fig. 2.

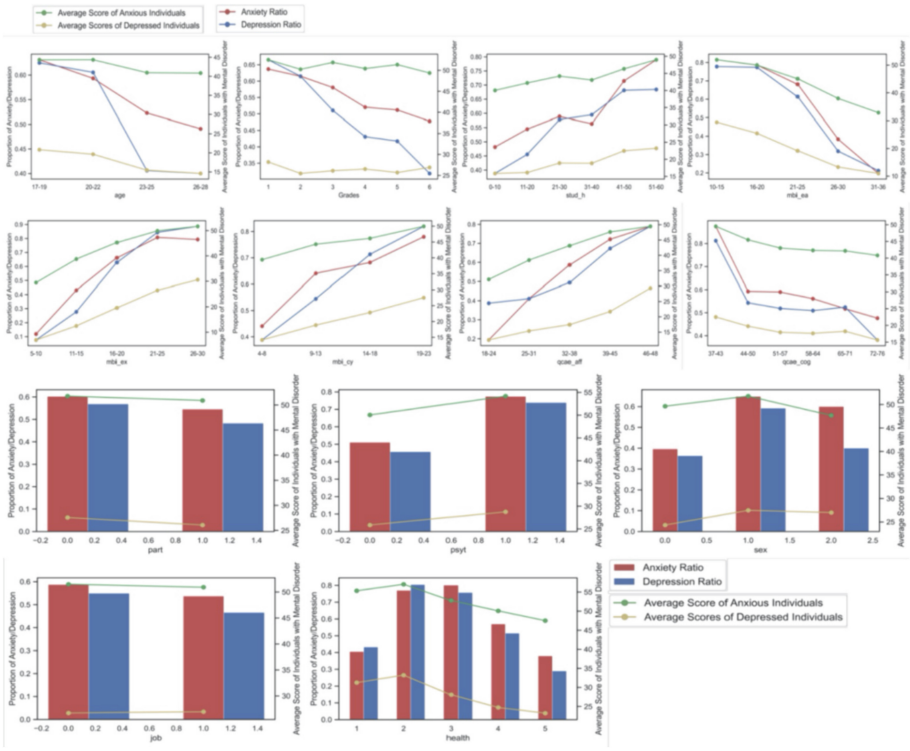

Fig. 2. Anxiety or depression statistical graphs

Features that exhibit a negative correlation with the proportion and severity of individuals with anxiety and depression include: age, academic year, academic efficacy, cognitive empathy, and health status. We observed that as age increases or academic year advances, the proportion of individuals with anxiety or depression decreases. Notably, the proportion of individuals with depression significantly drops after the age of 23. This may be attributed to medical students gradually adapting to the pace of learning, acquiring effective study methods, and consequently reducing the occurrence of anxiety and depression.

Features that show a positive correlation with the proportion and average scores of individuals with anxiety and depression include study duration, emotional exhaustion, cynicism, and affective empathy. We found that individuals with longer study durations exhibit a higher prevalence of psychological disorders, coupled with increased severity.

4 Result and Discussion

4.1 Statistical Analysis

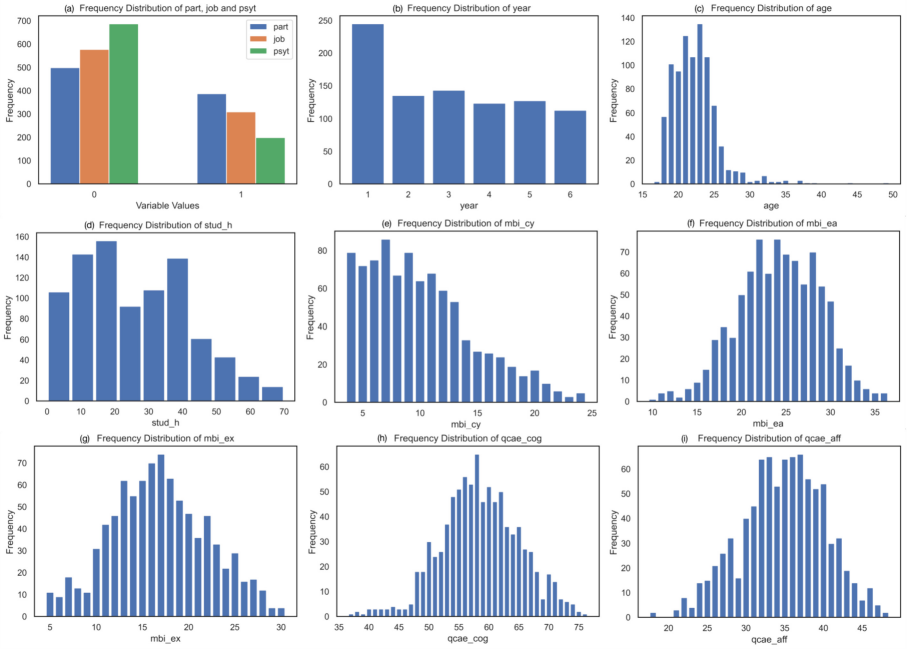

Fig. 1. Frequency Distribution of Each Feature. In Figure (a), the horizontal axis scale of 0 and 1 represents no partner (no job, no psychological treatment) and have a partner (job, psychological treatment) categories, respectively.

Univariate Statistical Analysis. The statistical graphs for each feature are depicted in the Fig. 1.

From Fig. 1(a), it is evident that students without partners outnumber those with partners, and similarly, students without jobs exceed those with jobs. Most students have not undergone psychological therapy over the past year.

Figure 1(b) illustrates that the first-year student count significantly surpasses other academic years, while second to sixth-year students exhibit a more even distribution.

Figure 1(c) indicates that the age distribution of medical students in the sample is concentrated between 18 to 25 years.

Figure 1(d), the highest number of individuals falls within the 10 to 20 h per week study time range. Most individuals do not exceed 40 h of study time per week.

The distributions of other features approximate a normal distribution.

Statistical Description. Creating statistical graphs for individual features provides a more intuitive display of the distribution of each feature's quantity, aiding in gaining a deeper understanding of the overall feature distribution within the sample population.

We have chosen two indicators, the proportion of individuals with psychological disorders and the average scores of the affected population, to depict the quantity and severity of psychological disorders. By visualizing the trends of these two indicators in relation to other features, we can gain a clearer insight into the influence of these features on psychological disorders.

Statistical Inference. This study primarily employs two hypothesis testing methods: the t-test and the chi-squared test, to conduct an analysis of dissimilarities among various features.

The independent samples t-test is utilized to compare differences between categorical and quantitative samples (samples A and B). The main steps are as follows:

1. Hypothesis formulation: The null hypothesis assumes no significant difference between samples A and B, while the alternative hypothesis assumes the presence of a difference.
2. Assumption of sampling distribution: Independent samples A and B are assumed to be approximately normally distributed, satisfying the conditions for t-distribution.
3. Calculation of t-value:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}(\frac{1}{n_1}+\frac{1}{n_2})}}$$

4. Calculation of confidence interval for means: Using the computed t-value, along with sample sizes and confidence level, the confidence interval for means is calculated, allowing for statistical inference regarding mean differences.

The Pearson chi-squared test is employed for analyzing differences between two categorical sample variables. The statistical measure used is

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i},$$

which assesses the disparity between theoretical frequencies and observed values.

commonly utilized, each specifically designed for screening anxiety and depression symptoms, respectively.

3 Methods

3.1 Dataset Introduction

The dataset [4,5] for this study was released in 2020 and encompasses information from 886 medical students. The features comprise individual demographic details ('age', 'year', 'sex', 'glang', 'part', and 'job'), educational aspects ('stud_h', 'mbi_cy', and 'mbi_ea'), psychological conditions ('qcae_cog', 'qcae_aff', and 'mbi_ex'), and physical well-being ('health'). Two labels describing the psychological disorder status are 'stai_t' and 'cesd', which are derived from the STAI-T and CESD questionnaires respectively. These questionnaires are widely employed for screening anxiety and depression patients. The introduction of each feature is shown in the Table 1.

Table 1. Study variables

variable name	description
age	age at questionnaire
year	curriculum year
sex	gender
glang	mother tongue
part	having a partner
job	have a paid job
stud_h	how many hours per week spend on study
health	How satisfied are you with your health
psyt	consulted a psychotherapist or a psychiatrist for health
qcae_cog	QCAE Cognitive empathy score
qcae_aff	QCAE Affective empathy score
mbi_ex	MBI Emotional Exhaustion
mbi_cy	MBI Cynicism
mbi_ea	MBI Academic Efficacy

3.2 Statistical Analysis

This study primarily engages in statistical description and hypothesis testing of the dataset, aiming to identify the relationships between psychological disorders (anxiety or depression) and various features.

1 Introduction

The psychological well-being of college students has garnered extensive attention. The university phase, which signifies the transition between academic and social realms [10], marks the initial steps of students venturing into the societal arena. However, due to factors such as uncertainty about the future, substantial academic pressure, challenges in interpersonal relationships, and insufficient self-confidence, college students are susceptible to experiencing psychological health issues such as anxiety and depression [12].

Among various academic disciplines, medical students particularly warrant significant concern as they encounter heightened psychological health challenges [2,3,8]. Their prolonged academic duration, substantial academic pressures, and the weight of future employment prospects create a formidable environment. Moreover, the daily exposure to patients' ailments and suffering brings about negative emotions, thereby increasing the likelihood of psychological health problems.

Despite the plethora of research focusing on factors contributing to psychological disorders, there remains a relative scarcity of investigations concentrating on medical students. Consequently, this study primarily revolves around medical students as a specific sample group, delving into their prevalence and severity of psychological disorders. Concurrently, we aim to discern potential factors contributing to the onset of psychological ailments and analyze the varying degrees of correlation between these factors and psychological disorders.

2 Related Work

Many researchers have initiated investigations into the psychological well-being of medical students. Medical students exhibit higher levels of depression, anxiety, and stress symptoms [7,15]. Such psychological disorders as anxiety and depression can potentially have adverse effects on medical students' personal and professional lives, leading to issues like insomnia and even triggering thoughts of suicide [9].

Mao *et al.* [13] found that the occurrence of depression and anxiety among medical students is influenced by a variety of factors, including individual characteristics, socioeconomic status, and environmental factors such as gender, academic year, family structure, family income, parental educational background, and social support. Additionally, Ahad *et al.* [1] revealed that age, gender, employment status, and accommodation situation are significant factors affecting stress levels among medical students. Notably, female students tend to experience higher stress levels, and those engaged in clinical internships face greater stress compared to pre-internship periods. It's noteworthy that the findings by Moutinho *et al.* [15] emphasize significant variations in the psychological well-being of medical students across different semesters.

Early detection and treatment of mental disorders are crucial for achieving favorable recovery outcomes and reducing the risk of relapse [6,14]. Typically, questionnaire surveys are employed for the early screening of anxiety and depression patients. Among these, the STAI-T [16] and CESD [11] questionnaires are

Analysis of Factors Related to Anxiety and Depression in Medical Students

Zheng Jinfang[1,2,3], Pan Jiachen[1,2,3], Zhang Peiyi[1,2,3], Xiao Yi[2,3], and Wang Wei[2,3,4(✉)]

[1] Faculty of Engineering, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China
1120200266@smbu.edu.cn

[2] Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China
xiaoyi@smbu.edu.cn, ehomewang@ieee.org

[3] Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China

[4] School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China

Abstract. The psychological well-being of university students, particularly those pursuing medical education, has garnered widespread attention. These students hold the potential to shape the future of societal progress, with medical students shouldering a crucial responsibility for the development of overall community health. However, many medical students are susceptible to psychological disorders such as anxiety and depression due to high levels of stress. While numerous studies have investigated factors contributing to the prevalence of psychological ailments in the general population, there has been a limited focus on analyzing this phenomenon specifically among medical students. This study utilizes a sample of 886 medical students, gathering information regarding their personal backgrounds, academic pursuits, psychological states, and physical health conditions. The aim is to discern which subgroups have a higher prevalence of anxiety or depression. Employing statistical analysis, the relationships between various factors and the occurrence of psychological disorders are examined. Through differential analysis, factors with a stronger correlation to psychological disorders are identified. Notably, factors like study duration and emotional fatigue exhibit a positive association with anxiety and depression, while factors such as academic year and academic efficacy demonstrate a negative correlation. Furthermore, gender and health status exhibit robust correlations with the manifestation of anxiety and depression.

Keywords: Anxiety · depression · medical students · correlative factors

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 52–60, 2024.
https://doi.org/10.1007/978-3-031-65126-7_5

34. Williams, J.B., First, M.: Diagnostic and statistical manual of mental disorders. In: Encyclopedia of Social Work (2013)
35. Wongkoblap, A., Vadillo, M.A., Curcin, V., et al.: Deep learning with anaphora resolution for the detection of tweeters with depression: Algorithm development and validation study. JMIR Mental Health **8**(8), e19824 (2021)
36. Wu, J., Zhou, Z., Wang, Y., Li, Y., Xu, X., Uchida, Y.: Multi-feature and multi-instance learning with anti-overfitting strategy for engagement intensity prediction. In: 2019 International Conference on Multimodal Interaction, pp. 582–588 (2019)
37. Yoon, J., Kang, C., Kim, S., Han, J.: D-vlog: Multimodal vlog dataset for depression detection. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 12226–12234 (2022)
38. Zhang, H., et al.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18802–18812 (2022)
39. Zheng, W., Yan, L., Wang, F.Y.: Two birds with one stone: knowledge-embedded temporal convolutional transformer for depression detection and emotion recognition. IEEE Trans. Affect. Comput. **14**(4), 2595–2613 (2023)
40. Zhou, L., Liu, Z., Yuan, X., Shangguan, Z., Li, Y., Hu, B.: Caiinet: neural network based on contextual attention and information interaction mechanism for depression detection. Digit. Sig. Process. **137**, 103986 (2023)
41. Zhu, Y., Shang, Y., Shao, Z., Guo, G.: Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. IEEE Trans. Affect. Comput. **9**(4), 578–584 (2017)
42. Zogan, H., Razzak, I., Wang, X., Jameel, S., Xu, G.: Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. World Wide Web **25**(1), 281–304 (2022)

14. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning, pp. 2127–2136. PMLR (2018)
15. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
16. Lin, C., et al.: Sensemood: depression detection on social media. In: Proceedings of the 2020 International Conference on Multimedia Retrieval, pp. 407–411 (2020)
17. Mann, P., Matsushima, E.H., Paes, A.: Detecting depression from social media data as a multiple-instance learning task. In: 2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 1–8. IEEE (2022)
18. de Melo, W.C., Granger, E., Hadid, A.: A deep multiscale spatiotemporal network for assessing depression from facial dynamics. IEEE Trans. Affect. Comput. **13**(3), 1581–1592 (2020)
19. de Melo, W.C., Granger, E., Lopez, M.B.: MDN: a deep maximization-differentiation network for spatio-temporal depression detection. IEEE Trans. Affect. Comput. 14(1), 578–590 (2021)
20. Meng, Y., Bridge, J., Addison, C., Wang, M., Merritt, C., Franks, S., Mackey, M., Messenger, S., Sun, R., Fitzmaurice, T., et al.: Bilateral adaptive graph convolutional network on CT based covid-19 diagnosis with uncertainty-aware consensus-assisted multiple instance learning. Med. Image Anal. **84**, 102722 (2023)
21. Mitra, V., et al.: The SRI avec-2014 evaluation system. In: Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, pp. 93–101 (2014)
22. Organization, W.H., et al.: Depression and other common mental disorders: global health estimates. World Health Organization, Technical Report (2017)
23. Safa, R., Bayat, P., Moghtader, L.: Automatic detection of depression symptoms in twitter using multimodal analysis. J. Supercomput. **78**(4), 4709–4744 (2022)
24. Saldanha, O.L., et al.: Self-supervised attention-based deep learning for pan-cancer mutation prediction from histopathology. NPJ Precision Oncol. **7**(1), 35 (2023)
25. Salekin, A., Eberle, J.W., Glenn, J.J., Teachman, B.A., Stankovic, J.A.: A weakly supervised learning framework for detecting social anxiety and depression. Proc. ACM Interact. Mobile Wearable Ubiquitous Technol. **2**(2), 1–26 (2018)
26. Shangguan, Z., Liu, Z., Li, G., Chen, Q., Ding, Z., Hu, B.: Dual-stream multiple instance learning for depression detection with facial expression videos. IEEE Trans. Neural Syst. Rehabil. Eng. **31**, 554–563 (2022)
27. Song, H., Kim, M., Park, D., Shin, Y., Lee, J.G.: Learning from noisy labels with deep neural networks: a survey. IEEE Trans. Neural Netw. Learn. Syst. **34**(11), 8135–8153 (2022)
28. Sotelo, J.L., Nemeroff, C.B.: Depression as a systemic disease. Personalized Med. Psychiatry **1**, 11–25 (2017)
29. Vahia, V.N.: Diagnostic and statistical manual of mental disorders 5: a quick glance. Indian J. Psychiatry **55**(3), 220 (2013)
30. Valstar, M., et al.: Avec 2014: 3d dimensional affect and depression recognition challenge. In: Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, pp. 3–10 (2014)
31. Valstar, M., et al.: Avec 2013: the continuous audio/visual emotion and depression recognition challenge. In: Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, pp. 3–10 (2013)
32. Verhoeven, J.E., Révész, D., Epel, E.S., Lin, J., Wolkowitz, O.M., Penninx, B.W.: Major depressive disorder and accelerated cellular aging: results from a large psychiatric cohort study. Mol. Psychiatry **19**(8), 895–901 (2014)
33. Wang, T., Li, C., Wu, C., Zhao, C., Sun, J., Peng, H., Hu, X., Hu, B.: A gait assessment framework for depression detection using kinect sensors. IEEE Sens. J. **21**(3), 3260–3270 (2020)

detection task, and the best performance results illustrate the effectiveness and superiority of our proposed method. We hope that our work can add more effective contributions to the field of weakly supervised depression detection. In future work, we hope to add more modal social media such as text for depression detection.

Acknowledgment. This work was supported by the National Nature Science Foundationof China (62102266, 62231020, 62272317), Tencent "Rhinoceros Birds"-Scientific Research Foundation for Young Teachers of Shenzhen University, Public Technology Platform of Shenzhen City (GGFW2018021118145859), Shenzhen Science and Technology Innovation Commission (R2020A045), Natural Science Foundation of Guangdong Province-Outstanding Youth-Program (2019B151502018), Pearl River Talent Recruitment Program ofGuangdong Province (2019ZT08X603, 2019JC01X235), National Key R&DProgram of China (2020YFA0908700), and the Natural Science and Engineering Research Council of Canada (corresponding author: Xiping Hu,huxp@bit.edu.cn).

References

1. Al Jazaery, M., Guo, G.: Video-based depression level analysis by encoding deep spatiotemporal features. IEEE Trans. Affect. Comput. **12**(1), 262–268 (2018)
2. Alghowinem, S., Goecke, R., Wagner, M., Parker, G., Breakspear, M.: Eye movement analysis for depression detection. In: 2013 IEEE International Conference on Image Processing, pp. 4220–4224. IEEE (2013)
3. Bourke, C., Douglas, K., Porter, R.: Processing of facial emotion expression in major depression: a review. Aust. NZ. J. Psychiatry **44**(8), 681–696 (2010)
4. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T.F.: A review of depression and suicide risk assessment using speech analysis. Speech Commun. **71**, 10–49 (2015)
5. Cummins, N., Sethu, V., Epps, J., Schnieder, S., Krajewski, J.: Analysis of acoustic space variability in speech affected by depression. Speech Commun. **75**, 27–49 (2015)
6. Feng, J., Zhou, Z.H.: Deep miml network. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31 (2017)
7. Ge, Y., Zhou, Q., Wang, X., Shen, C., Wang, Z., Li, H.: Point-teaching: weakly semi-supervised object detection with point annotations. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 667–675 (2023)
8. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. **18**(5–6), 602–610 (2005)
9. Gui, T., et al.: Cooperative multimodal approach to depression detection in twitter. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 110–117 (2019)
10. Hashimoto, N., et al.: Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3852–3861 (2020)
11. Hendra, C., Pratanwanich, P.N., Wan, Y.K., Goh, W.S., Thiery, A., Göke, J.: Detection of m6a from direct RNA sequencing using a multiple instance learning framework. Nat. Methods **19**(12), 1590–1598 (2022)
12. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
13. Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H.: Patch-based convolutional neural network for whole slide tissue image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2424–2433 (2016)

As show in Fig. 1, the max pooling operation performs the worst, but is comparable to machine learning-based methods. The mean pooling operation outperforms the max pooling operation and achieves comparable results to multiple deep network based methods. In contrast, our proposed attention-based pooling operation achieves the best result, which shows that the attention mechanism effectively improves the performance of the MIL framework.

4.5 Effect of Instance Temporal Size

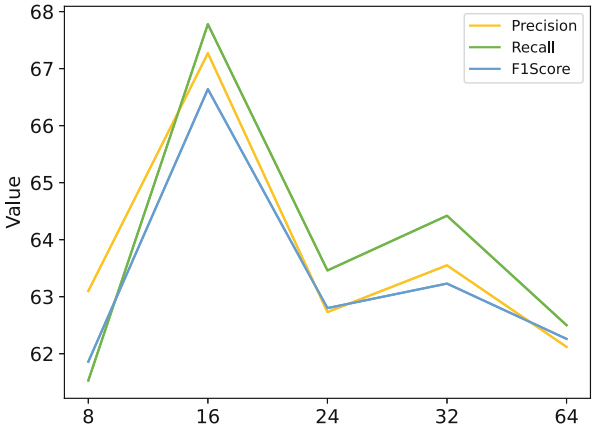

Fig. 2. Evaluation of varying instance temporal size on D-Vlog dataset

To assess the effect of instance time segments size on the method, we construct time segments k of different sizes for the study. As shown in Fig. 2, the best results are achieved in all metrics when k = 16. Moreover, it is worth noting that the results of the model do not exhibit linear change when the value of k increases or decreases, demonstrating that the smaller or larger time segments are not appropriate in depression detection. Technically, the size of the time segment determines the length of the time information contained in the instance. When the size of time segment decreases, continuous time series entering the time window is too short to perform sufficient feature aggregation through the multiple instance pooling layer. When the size of time segment increases, the redundancy of too much temporal information may affect the training of the model.

5 Conclusion

In this study, we perform an attention-based multiple instance learning method to detect depression using social media. We conduct sufficient experiments on the D-Vlog dataset and report the state-of-the-art model performance compared to us. The experimental results show the advantages and potential of multiple instance learning in depression

Table 1. Evaluation of the proposed methods on D-Vlog dataset

Method	Precision(%)	Recall(%)	F1Score(%)
LR	54.86	54.72	54.78
SVM	53.10	55.19	52.97
RF	57.69	58.49	57.84
KNN-Fusion	57.86	59.43	54.25
BiLSTM	60.81	61.79	59.70
TFN	61.39	62.26	61.00
Transformer_Concat	62.51	63.21	61.10
Transformer_Add	59.11	60.38	58.11
Transformer_Multiply	63.48	64.15	63.09
Depression_Detector	65.40	65.57	63.50
Temporal Convolutional	65.40	64.70	65.00
CAIINET	66.57	66.98	66.56
Ours	**67.27**	**67.77**	**66.64**

4.4 Comparison with the MIL Methods

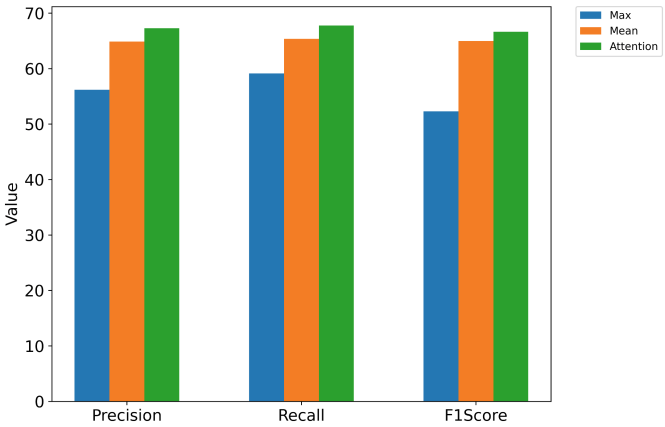

Fig. 1. Evaluation of the MIL methods on D-Vlog dataset

In order to explore the potential of MIL in depression detection and compare with the attention-based MIL method used in this paper, we present the methods based on max pooling [6] and mean pooling [6] as comparison experiments. Notably, the max pooling and mean pooling operation select the feature with the highest and average feature among the instance to obtain the bag-level feature.

4 Experiments

4.1 Experimental Dataset

In this work, we use the D-Vlog dataset [37] collected from YouTube, which contains 961 vlogs videos from 816 subjects composed of 322 males and 639 females. The dataset has a total of 505 depressed subjects and 406 healthy controls, and the average length of the vlog is 596 s. According to the ratio of 7:1:2, the dataset is divided into training set, validation set, and test set, respectively. The preliminary label assignment of the dataset comes from the title keywords of the vlog. Usually, vlogs containing keywords such as "depression daily vlog", "depression journey vlog" and "depression vlog" are labeled as depressed vlog. In addition, vlogs containing keywords such as "daily vlog" and "haul vlog" are labeled as non-depressed vlogs. Then, two tasks are used to ensure the plausibility of labels. First, videos that do not conform to the "vlog" format (e.g., videos without appearance) are removed. Second, specific annotators are assigned to judge whether the subjects have depression by watching the vlog videos with automatic text generation. For privacy protection considerations, D-Vlog only provides the features of the extracted voice and facial expression in the video, which are the 15-dimensional extended Geneva Minimalistic Acoustic Parameter Set and the 68-dimensional facial landmarks, respectively.

4.2 Experimental Setup

In this paper, the size of the time segment that constitutes the instance is 16, and the total length of the bag is limited to 596. All models are trained for 30 epochs, using Adam [15] as the optimizer with learning rate, weight decay and eps are set to 1e-4, 5e-4, and 1e-8, respectively and the batch size is set to 16. We report weighted average precision, recall, and F1 score to evaluate model performance. In order to prevent overfitting, the model uses an early stopping mechanism during training. All experiments are implemented in pytorch, running on a server with NVIDIA 1660 s and 16 GB RAM.

4.3 Comparison with the Previous State-of-the-Art Models

We compare with current state-of-the-art methods to evaluate the effectiveness of our method, and the results are shown in Table 1. Specifically, the recent methods for comparison include: 10 methods using in the D-Vlog dataset [37] as the baseline, the Knowledge-Embedded Temporal Convolutional Transformer method proposed by Zheng et al. [39] and the CAIINET method proposed by Zhou et al. [40]. The traditional machine learning methods including LR, SVM and RF don't perform well on the D-vlog dataset, which is due to the lack of nonlinear fitting ability of machine learning. Moreover, corresponding deep learning methods including BISTM, TFN and Depression Detector achieve better performance compared to machine learning methods.

Compared with baseline, our proposed method improves at least 1.87%, 2.2% and 3.14% on the weighted average precision, recall and F1 score metrics. In addition, compared with the recently proposed Knowledge-Embedded Temporal Convolutional Transformer method and the CAIINET method, our proposed method has achieved the highest results in all metrics, indicating the effectiveness of the proposed method.

$$F_{max} = W_2 h_{max} \tag{6}$$

where W_1 and W_2 represent trainable weights respectively. Moreover, in order to integrate the information of the obtained vectors F_{mean} and F_{max}, we concat them and use a one-dimensional convolution operation to obtain the contextual kernel α:

$$\alpha = f_c\left([F_{mean} : F_{max}]\right) \tag{7}$$

where f_c represents the convolution operation with convolution kernel size is 1. Formally, by combining the context kernel α with the final output O of BILSTM, we obtain the instance features with temporal context. This step can be formulated into:

$$z = \sum_{t=1}^{T} a_t O_{w,t} \tag{8}$$

where,

$$a_t = \frac{\exp\left(\alpha O_{w,t}^{\top}\right)}{\sum_{\tau=1}^{T} \exp\left(\alpha O_{w,\tau}^{\top}\right)} \tag{9}$$

Technically, a_t is the attention weight to indicate the effectiveness of the BiLSTM output feature. Also, instance feature with temporal information is obtained by combining the attention weight with the output feature, which helps to articulate the dynamic information of depressed patients.

3.3 AD-MIL

Recently, many studies have attempted to use attention mechanisms to integrate them into the MIL framework [11,14,38]. Notably, Ilse et al. [14] demonstrate that MIL based on attention pooling can achieve better performance compared to conventional multiple instance pooling such as max pooling and mean pooling. Inspired by these, we use attention pooling to aggregate the instance features obtained in the previous section.

Formally, we denote $Z = \{z_1, \ldots, z_M\}$ as a bag of M instance features, and attention-based MIL pooling can be defined as:

$$e = \sum_{m=1}^{M} b_m z_m \tag{10}$$

with,

$$b_m = \frac{\exp\left\{q^{\top} \tanh\left(V z_m^{\top}\right)\right\}}{\sum_{k=1}^{M} \exp\left\{q^{\top} \tanh\left(V z_k^{\top}\right)\right\}} \tag{11}$$

where q and V are trainable parameters and hyperbolic tangent $\tanh(\cdot)$ is the element-wise non-linearity. In addition, b_m represents as an attention weight indicating the contribution of a given instance to the prediction of the whole bag. Therefore, different attention weights can be used as an implicit feature selection to make the final bag features more informative.

m-th instance of i-th bag. Furthermore, each instance s_{im} is assumed to have implicit label $y_m \in \{0,1\}$ to represent negative or positive, which is not given in practice due to labeling difficulties.

Traditional MIL meets the following constraints: A bag is positive if there is at least one positive instance, while a negative bag is only if all the instances making up the bag are negative. Formally, it follows that

$$Y = \begin{cases} 0, & \text{iff } \sum_m y_m = 0 \\ 1, & \text{otherwise.} \end{cases} \tag{1}$$

However, in the case of our work, there will be cases where both positive and negative instances are included in one bag, so assumption here is not strict. Hence, we propose an attention-based algorithm for depression detection by introducing a looser version of the attention mechanism to assign implicit weights to instances.

3.2 AD-LSTM

The proposed AD-LSTM module first uses Bi-directional LSTM (BiLSTM) [8] to extract the temporal information of the two directions of LSTM [12] as output, and then uses the attention mechanism to integrate the semantic features with temporal information to obtain the feature of the instance.

We develope BiLSTM combining information in both directions of LSTM at the same time to obtain richer semantic information in the instance. Notably, each layer of BiLSTM consists of LSTM in two directions, and the outputs of the layer are as follows:

$$h_{i,t} = l_f\left(O_{i-1,t}\right) \tag{2}$$

$$H_{i,T-t} = l_b\left(O_{i-1,T-t}\right) \tag{3}$$

$$O_{i,t} = \left[h_{i,t}, H_{i,T-p}\right] \tag{4}$$

where T represents the total length of the segment. l_f and l_b represent the forward and backward LSTM models, respectively. $h_{i,t}$ and $H_{i,T-t}$ represent the output of the i-th layer at the time t of the forward LSTM and the output of the i-th layer at the $T-t$ time of the backward LSTM. Then, we add the forward and backward features of the last layer w of BILSTM to get $O = \{O_{w,1}, \ldots, O_{w,T}\}$, and we connect the forward and backward output features of BILSTM at time T to form contextual feature $h = \{h_{1,T}, H_{1,T}, \ldots, h_{w,T}, H_{w,T}\}$. Similar to the feature map in CNN, each $h_{i,T}$ in h represents the feature of BILSTM at the last time. Therefore, in order to obtain rich contextual information, we use the mean pooling operation and max pooling operation, which is commonly used in spatial information processing, to obtain context features h_{mean} and h_{max}. Further, we use the two-layer network model to introduce the non-linear operations of the two pooling features:

$$F_{mean} = W_1 h_{mean} \tag{5}$$

2.2 Weak Supervision and Multiple Instance Learning

Multiple instance learning (MIL) is a form of weakly supervised learning, which is used to deal with model training under insufficient labels. Typically, in multiple instance learning, the model only receives coarse-grained bag-level labels, and the labels of the instances that make up the bag are unknown. According to the different MIL settings, the current MIL algorithm can be divided into instance-level [13] and bag-level [10]. Due to the difficulties and high costs in the actual labeling process, specific annotators can only assign bag-level labels in the context. Hence, MIL has been widely used in many fields including object detection [7], pneumonia detection [20] and tumor detection [24].

Recently, several works of MIL has been applied for depression detection. Concretely, in the use of weakly supervised learning framework, Salekin et al. [25] proposed a MIL method to identify depression from voice speech containing labels of depressed patients without providing specific segments of symptoms. Shangguan et al. [26] proposed a dual-stream MIL deep network to identify depression by using raw facial expressions. In addition, extensive works have used MIL for detecting depression on social media due to its superiority. Wongkoblap et al. [35] proposed two multiple instance learning models to predict depression using textual information from Twitter. Moreover, a MIL method for detecting depression using students' posts from university was presented by Mann et al. [17]. They performed theoretical and experimental analysis by using Transformer and LSTM model on the dataset of university students.

Previous work using MIL to identify depression in social media has mainly focused on text information, and few works have used the information of subjects' facial expressions and voices to identify depression. Since the facial expressions and voice can express the mental state of the subjects and the effectiveness of MIL in the detection of depression, it is very necessary to establish a model for detecting depression using MIL based on these two modalities. Inspired by the work of these pioneers, we aim to expand the scope of the current literature on depression detection through the application of MIL and attention mechanisms.

3 Methods

We propose a weakly supervised learning model for the depression detection task in a single end-to-end deep network. Concretely, our model receives the vocal features and visual features extracted by OpenSmile and Dlib respectively, and then the AD-LSTM module extracts the temporal information within the instances. Finally, the AD-MIL module integrates the information of the instance for identifying depression. In this section we present the formulation of MIL and the details of the proposed MIL model.

3.1 Preliminaries

The MIL algorithm receives N labeled sample pairs $\mathrm{D} = \{(S_1, Y_1), \ldots, (S_N, Y_N)\}$, where S_i (i from 1 to N) is the whole bag and Y_i is $\{0, 1\}$ for binary classification of depression and health. Also, $S_i = \{s_{i1}, s_{i2}, \ldots, s_{iM}\}$ consists of M instances where s_{im} represents the

approaches using weakly supervised learning. In Sect. 3, we introduce the relevant preliminaries and the details of our proposed model. We provide the datasets, experimental settings and results used in the experiments in Sect. 4. Finally, we conclude our work in Sect. 5.

2 Related Work

This section briefly reviews the related methods of depression detection and weakly supervised learning.

2.1 Deep Learning for Depression Detection

Since the emerging applications in affective computing, the deep learning-based methods can use behavior signals for depression detection. The datasets for depression detection tasks can be divided into task-specific collection and non-specific task collection. In a specific task, the process of data collection comes from recording subjects completing a certain task according to the requirements of the examiner, such as answering some specific questions or discussing the certain topics. In a non-specific task, the process of data collection comes from external information, such as, voice, video, and text of individuals on the Internet.

AVEC2013 [31] and AVEC2014 [30] are typical task-specific datasets which focus on video modalities. For example, Zhu et al. [41] proposed a two-stream deep network to detect depression by considering the appearance and movements of subjects. By leveraging the optical flow of dynamic information of facial expressions, they improved the performance of the model. Similarly, Jazaery et al. [1] used a convolutional 3D network (C3D) to capture spatio-temporal information and to learn the features of continuous segments through Recurrent Neural Network (RNN). To reduce the model size for depression detection, Melo et al. [19] proposed the 2D deep network (a.k.a., MDN) to capture the spatio-temporal information in facial videos. By embedding the maximization block and difference block in the 2D deep network, the model captured the subtle changes and sudden transitions between face expression, and achieved comparable performance to 3D deep network.

Since a considerable number of users share recent life emotions and states on the Internet, social media can provide data information under non-specific tasks for depression detection. There are several approaches that use multi-modal data of social media for depression detection. For example, Safa et al. [23] used the biological features, features generated by analyzing user profile pictures, and banner images to detect depression. By using image and text information posted by users on social media, Gui et al. [9] introduced a new collaborative multi-agent reinforcement learning method to predict depression. Zagan et al. [42] presented a novel interpretable depression detection framework, the Hierarchical Attention Network, which used textual, behavioral, temporal, and semantic aspects of social media features for deep learning. Moreover, a deep visual-textual multimodal learning system dubbed SenseMood was proposed to predict the mental state of the users on social networks. Lin *et al.* [16] used CNN and Bert to extract deep representations of pictures and text on social media, which were combined for further depression classification.

of depression mainly relies on the subjective and complex reports of the subjects and the professional judgment of the psychiatrists. For example, the clinician rating scales (e.g., Hamilton rating scale) require rigorous training of raters. The self-rating depression scales (e.g., Self-rating depression scale) rely on accurate description, assessment, and expression of subjects and may change the purpose of their report [29]. Due to the lack of medical resources, the great harm of depression, and the large number of patients, the subjective assessment and diagnosis cannot meet the current demands for depression diagnosis. In this vein, the automatic detection of depression has attracted ever-increasing research attention due to objectivity, fast deployment, and long endurance.

With the advancement of affective computing, previous studies use behavioral signals as objective indicators to conduct research on depression detection, which provides an objective and effective way for auxiliary diagnosis of depression. Many current research outcomes have shown that common behavioral signals can be used as objective indicators for depression detection, such as, eye movement [2], voice [4], gait [33], and facial expression [3].

Different from eye movements and gaits that need to be collected during professional experiments, the leveraged facial expressions and voices in our research obtained through more relaxed methodologies (e.g., social media). In this paper, we use social media data collected from vlog of people documenting their daily lives on the Internet. Compared with data collected from the experimental environments, vlog data has three advantages: 1) easier to obtain; 2) larger quantity; and 3) consistent increase in volume. The three advantages of the vlog data allow to build a more generalized model and explore the ability to articulate the datasets in the wild. In general, a vlog dataset has both facial expressions and voice modalities, and there are dedicated annotators to judge whether the subjects are depressed. However, a complete piece of vlog data has only one binary classification sparse label, because it is impractical for annotators to accurately label the symptoms reflecting depression at a fine-grained level. The traditional depression detection methods [1,5,18,21] assign the same coarse-grained labels to the training instances (video clips or single frames) and may lead to overfitting and corresponding performance loss [27,36].

To address this challenge, we propose a weakly supervised method to identify the binary classification of the subjects (depression or health) using vlog data. Our model takes temporal segments of a certain size as instance input, and uses AD-LSTM to extract temporal contextual information to obtain instance-wise representations. Then, AD-MIL views the vlog video of each subject as a bag containing multiple instances that may be positive or negative. More specifically, the AD-MIL model first uses an attention mechanism to identify the contribution weights of instances to the final classification, and then obtains individual subject representations by combining the weights with instance representations. Technically, using the attention mechanism can effectively alleviate the impact of instances that are not related to the classification label and integrate the information of the whole bag. We conduct a series of experiments on the D-vlog dataset [37], and the experimental results show that our proposed method exceeds the state-of-the-art works, indicating the effectiveness of the proposed method.

The remaining parts of this work are organized as follows. In Sect. 2, we present recent approaches to automatic depression detection using deep learning as well as

Automatic Depression Detection Using Attention-Based Deep Multiple Instance Learning

Zixuan Shangguan[1], Xiaxi Li[2], Yanjie Dong[2], and Xiaoyan Yuan[1(✉)]

[1] Beijing Institute of Technology, Beijing, China
yxy_newly@163.com
[2] Shenzhen MSU-BIT University, Shenzhen, China
{1120200239,ydong}@smbu.edu.cn

Abstract. Depression is a serious mental illness and one of the leading causes of suicide worldwide. However, the social prejudice and the lack of psychiatrists for depression lead to a significant number of depressed patients without accurate diagnosis and subsequent serious consequences. With the rise of social media, previous studies have found that the information of depressed patients on social media can be analyzed to automatically detect depression for auxiliary diagnosis. In the context of weakly supervised learning framework, a multiple instance learning (MIL) method is proposed to identify depression from social media with visual and vocal information. By leveraging the state-of-the-art attention-based deep LSTM (AD-LSTM), the proposed MIL method can handle the problem with sparse labels (i.e., one label for a long-term sequence of visual information). More specifically, the AD-LSTM module is used to process a fixed-length visual and vocal segments to extract temporal representations of instances, and the AD-MIL module is used to aggregate the obtained temporal representations for individual subject predictions. Compared with current benchmarks, our experiments demonstrate that our proposed MIL method can achieve the best weighted average precision, recall and F1 score with the corresponding values as 66.56%, 66.98% and 66.55%, respectively. The numerical results illustrate that the potential and effectiveness of our proposed MIL method in the field of depression detection.

Keywords: Depression Detection · Multiple Instance Learning · Social media

1 Introduction

Major depressive disorder (a.k.a.. depression) is a critical mental illness with serious consequences for individual physical and mental health. More than 300 million people worldwide, which is equivalent to 4.4% of the global population, are currently suffering from varying degrees of depression [22]. Depressed people often exhibit low mood, loss of interest in practice, sleep disturbance, loss of appetite, lack of self-confidence, loss of energy, and inability to concentrate [34]. In addition, depression increases the risk of diabetes, heart disease, Alzheimer's and, in more severe cases, suicide [28,32].

Accurate diagnosis of depression can be effectively controlled and treated through psychological consulting and psychotropic medication. However, the current diagnosis

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 40–51, 2024.
https://doi.org/10.1007/978-3-031-65126-7_4

29. Shen, G., et al.: Depression detection via harvesting social media: a multimodal dictionary learning solution. In: IJCAI, pp. 3838–3844 (2017)
30. Shen, G., e al.: Depression detection via harvesting social media: a multimodal dictionary learning solution. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, IJCAI 2017, pp. 3838-3844. AAAI Press (2017)
31. Tadesse, M.M., Lin, H., Xu, B., Yang, L.: Detection of suicide ideation in social media forums using deep learning. Algorithms **13**(1), 7 (2019)
32. Trotzek, M., Koitka, S., Friedrich, C.M.: Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences (2018)
33. Trotzek, M., Koitka, S., Friedrich, C.M.: Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. IEEE Trans. Knowl. Data Eng. **32**(3), 588–601 (2018)
34. Wang, N., etal.: Learning models for suicide prediction from social media posts. arXiv preprint arXiv:2105.03315 (2021)
35. Wang, Y., Wang, Z., Li, C., Zhang, Y., Wang, H.: A multitask deep learning approach for user depression detection on sina weibo. arXiv preprint arXiv:2008.11708 (2020)
36. Yang, T., et al.: Fine-grained depression analysis based on chinese micro-blog reviews. Inf. Process. Manage. **58**(6), 102681 (2021)
37. Yao, X., Yu, G., Tang, J., Zhang, J.: Extracting depressive symptoms and their associations from an online depression community. Comput. Hum. Behav. **120**, 106734 (2021)
38. Zhou, S., Zhao, Y., Bian, J., Haynos, A.F., Zhang, R., et al.: Exploring eating disorder topics on twitter: machine learning approach. JMIR Med. Inform. **8**(10), e18273 (2020)
39. Zogan, H., Razzak, I., Jameel, S., Xu, G.: Depressionnet: a novel summarization boosted deep framework for depression detection on social media. ArXiv **abs/2105.10878** (2021)

13. Gui, T., et al.: Cooperative multimodal approach to depression detection in twitter. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'19/IAAI'19/EAAI'19, AAAI Press (2019). https://doi.org/10.1609/aaai.v33i01.3301110
14. Holt-Lunstad, J., Smith, T.B., Baker, M., Harris, T., Stephenson, D.: Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect. Psychol. Sci. J. Assoc. Psychol. Sci. **10**(2), 227 (2015)
15. Kessler, R.C., et al.: Lifetime prevalence and age-of-onset distributions of mental disorders in the world health organization's world mental health survey initiative. World Psychiatry **6**(3), 168 (2007)
16. Kohler, C.G., Hoffman, L.J., Eastman, L.B., Healey, K., Moberg, P.J.: Facial emotion perception in depression and bipolar disorder: a quantitative review. Psychiatry Res. **188**(3), 303–309 (2011)
17. Lewis, M., et al.: Bart: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7871–7880. Association for Computational Linguistics, July 2020. https://doi.org/10.18653/v1/2020.acl-main.703, null ; Conference date: 05-07-2020 Through 10-07-2020
18. Li, X., Sun, X., Meng, Y., Liang, J., Wu, F., Li, J.: Dice loss for data-imbalanced NLP tasks. arXiv preprint arXiv:1911.02855 (2019)
19. Lin, H., Jia, J., Nie, L., Shen, G., Chua, T.S.: What does social media say about your stress? In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, pp. 3775-3781. AAAI Press (2016)
20. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
21. Malhi, G.S., Mann, J.J.: Depression. The Lancet **392** (2019)
22. Malhotra, A., Jindal, R.: Deep learning techniques for suicide and depression detection from online social media: a scoping review. Appl. Soft Comput. **130**, 109713 (2022)
23. Meng, Y., Li, M., Li, X., Wu, W., Li, J.: Dsreg: using distant supervision as a regularizer. arXiv preprint arXiv:1905.11658 (2019)
24. Park, M., Cha, C., Cha, M.: Depressive moods of users portrayed in twitter. In: Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining, SIGKDD 2012, pp. 1–8 (2012)
25. Pyszczynski, T., Holt, K., Greenberg, J.: Depression, self-focused attention, and expectancies for positive and negative future life events for self and others. J. Pers. Soc. Psychol. **52**(5), 994 (1987)
26. Ríssola, E.A., Losada, D.E., Crestani, F.: A survey of computational methods for online mental state assessment on social media. ACM Trans. Comput. Healthcare **2**(2), 1–31 (2021)
27. Salas-Zárate, R., Alor-Hernández, G., Salas-Zárate, M.D.P., Paredes-Valverde, M.A., Bustos-López, M., Sánchez-Cervantes, J.L.: Detecting depression signs on social media: a systematic literature review. In: Healthcare, vol. 10, p. 291. MDPI (2022)
28. Sekulić, I., Strube, M.: Adapting deep learning methods for mental health prediction on social media. arXiv preprint arXiv:2003.07634 (2020)

5 Conclusions

In this work, we have devised a versatile framework for processing Twitter data to facilitate the diagnosis of early-stage depression. Through an extensive study, we have ascertained that Twitter textual content, user profile information, and historical posting data all hold profound significance in diagnosing depression. Consequently, we have proposed a comprehensive model that amalgamates these inputs and conducted empirical validations to evince the efficacy of our approach. Moreover, we addressed the issue of imbalanced data pertaining to depression patients by exploring several diverse methodologies, culminating in commendable achievements.

Acknowledgement. This work is supported by the Natural Science Foundation of Guangdong Province of China (No. 2021A1515011905)

References

1. Ahmed, U., Mukhiya, S.K., Srivastava, G., Lamo, Y., Lin, J.C.W.: Attention-based deep entropy active learning using lexical algorithm for mental health treatment. Front. Psychol. **12**, 642347 (2021)
2. Beck, A.T.: Cognitive Therapy of Depression. Guilford Press, New York (1979)
3. Belmaker, R.H., Agam, G.: Major depressive disorder. New England J. Med. Mech. Disease **385**, 47–60 (2008)
4. Birmaher, B., Ryan, N.D., Williamson, D.E., Brent, D.A., Kaufman, J.: Childhood and adolescent depression: a review of the past 10 years. part ii. J. Am. Acad. Child Adolescent Psychiatry **35**(11), 1427–1439 (1996)
5. Brent, A.D.: Course and outcome of child and adolescent major depressive disorder. Child Adolescent Psych. Clin. North Am. **11**(3), 619–637 (2002)
6. Carlson, G.A.: The challenge of diagnosing depression in childhood and adolescence. J. Affect. Disord. **61**(supp-S1), S3–S8 (2000)
7. Castillo-Sánchez, G., Marques, G., Dorronzoro, E., Rivera-Romero, O., Franco-Martín, M., De la Torre-Díez, I.: Suicide risk assessment using machine learning and social networks: a scoping review. J. Med. Syst. **44**(12), 205 (2020)
8. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
9. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
10. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
11. Fine, A., Crutchley, P., Blase, J., Carroll, J., Coppersmith, G.: Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data. In: Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science, pp. 50–54 (2020)
12. Ghosh, S., Anwar, T.: Depression intensity estimation via social media: a deep learning approach. IEEE Trans. Comput. Soc. Syst. **8**(6), 1465–1474 (2021)

Table 3. The impact of different information

User information	Prec	Rec	F1	Acc
Historical tweet	80.2%	**89.0**%	**84.3**%	83.5%
Historical tweet+BiGRU	**85.8**%	82.8%	84.2%	**84.5**%

the model's robustness. In practical scenarios, this inclusion can mitigate the risk of misidentifying healthy individuals as depressed patients. Furthermore, as user information tends to display greater individuality and lacks a discernible pattern, our user information extraction module can still capture its distinctive features, thereby contributing to incremental improvements in the overall model.

4.4 The Experiment on Data Imbalance

To investigate the efficacy of traditional methods for handling imbalanced data in depression detection, we randomly extracted two imbalanced datasets from the original dataset. The first dataset had a ratio of depressed patients to non-depressed patients of 1:4, while the second dataset had a ratio of 1:2. Subsequently, we conducted a comparative analysis of three methods on these two datasets: the direct usage of cross-entropy loss without any imbalance treatment, the employment of Focal Loss for handling imbalance, and the utilization of Dice Loss for the same purpose. The experimental outcomes are presented in Table 4, with the evaluation metric being the F1 score.

Table 4. The effects of different methods for handling data imbalance.

Imbalanced ratio	CE	Focal Loss	Dice loss
1:4	33.3%	**79.5**%	75.7%
1:2	66.3%	79.8%	**81.8**%

From Table 4, it can be observed that when the imbalance ratio reaches 1:4, the model fails to learn useful classification knowledge from the limited positive samples when using cross-entropy loss without imbalance handling. Even under a 1:2 imbalance ratio, the results are significantly unsatisfactory. Focal Loss, in comparison to cross-entropy, demonstrates a 46.2% and 13.5% increase in F1 scores at 1:4 and 1:2 imbalance ratios, respectively. Similarly, Dice Loss shows a 42.4% and 15.5% improvement in F1 scores at 1:4 and 1:2 imbalance ratios. This verifies the effectiveness of traditional data imbalance handling methods for depression classification in our model. Additionally, it is evident that under higher imbalance conditions (1:4), Focal Loss outperforms, achieving an F1 score of 79.5%, whereas when the imbalance ratio reduces to 1:2, the performance of Dice Loss is superior, achieving an F1 score of 81.8%.

Historical Tweet Model. We employed the large-scale language pretraining model BERT and attention-based bidirectional LSTM to construct a historical tweet feature extraction model. Subsequently, we conducted experimental comparisons on each module, and the results are presented in Table 2. Note: due to better performance with straightforward extraction during data processing, all experiments were performed on data obtained through straightforward extraction.

Table 2. Historical tweet model

model	Prec	Rec	F1	Acc
BERT	77.7%	82.8%	80.1%	79.5%
BERT+BiLstm	**81.6**%	83.3%	82.4%	82.3%
BERT+StackedBiLstm	80.2%	**89.0**%	**84.3**%	**83.5**%

According to Table 2, it can be observed that the utilization of BERT in conjunction with the StackedBiLSTM model yields the most favorable results when processing textual features. Following this, the employment of BERT in combination with BiLSTM ranks second in performance. We believe this is due to a certain temporal correlation in the user's tweet data. Due to BiLSTM's capacity to maintain "memory," the model with an added BiLSTM layer outperforms the classification model solely relying on BERT. Furthermore, it is evident that the StackedBiLSTM model outperforms the BiLSTM model in terms of recall, F1 score, and accuracy, surpassing it by 5.7%, 1.9%, and 1.2%, respectively. However, it should be noted that the accuracy is reduced by 1.4% . We posit that this could be attributed to an excessive focus on contextual information, leading to the inadvertent capture of some depression-irrelevant data and thereby increasing the likelihood of misidentifying depression patients.

The Impact of User Information. We investigated the impact of user information on the classification of users with depression in our experiment. We extracted personal information from users, including the number of individuals they follow, the count of their followers, the quantity of friends, and the number of tweets sent in the past month. For feature extraction, we utilized a Bidirectional Gated Recurrent Unit (BiGRU) to complement the historical tweet features. We explored the experimental outcomes achieved by solely using historical tweets and by amalgamating user information with historical tweets. The model employed in this study is depicted in the aforementioned model diagram in Fig. 1, and the results are presented in Table 3.

From Table 3, it can be observed that the incorporation of user information and historical text enhances the model's precision and accuracy, surpassing the historical tweet model by 4.6% and 1%, respectively. However, the recall rate decreased by 6.2%. We contend that the inclusion of user information enhances

4.2 Hyperparameter Configuration

We attempted to employ the BERT pre-trained model as our Encoder model. For the tweet extraction model (a BiLSTM classification model with attention mechanism), we utilized a 2-layer BiLSTM with hidden layer neural units set to 128. As for the user behavior model, we opted for a single-layer BiGRU with hidden layer neural units set to 128. All experiments were conducted on an RTX3090GPU using the Pytorch framework. We employed the SGD optimizer for training with specific parameters: learning rate (lr) = 0.001, momentum = 0.9, and weight decay = 0.0004. Additionally, we employed a warm-up strategy to reach the initial learning rate. We performed a total of 30 epochs, and during the 10th and 20th epochs, we applied a learning rate decay with a rate of 0.1.

We introduced two approaches to process tweets: simple tweet extraction and K-means-based tweet filtering. To assess the performance of our model, we employed metrics such as accuracy, precision, recall, and F1 score. Additionally, to examine the impact of various components in the model, we conducted an ablation analysis. In experiments involving imbalanced data, we set $\gamma = 2$ in the Focal Loss and $\alpha = 0.9$ in the Dice Loss, with $\epsilon = 1e^{-4}$.

4.3 Ablation Study

Method of Data Processing. We have presented three distinct approaches for data processing: a straightforward tweet extraction method and a tweet filtering based on K-means clustering. These methods were subjected to experimental comparison. All models employed the BERT pre-trained model in conjunction with a single layer of Bidirectional LSTM (BiLSTM). The results are presented in Table 1.

Table 1. Module for Data Processing

Data processing	Prec	Rec	F1	Acc
simplistic tweet extraction	**81.6**%	83.3%	**82.4**%	**82.3**%
K-means	76.6%	**83.5**%	79.9%	79.0%

As indicated in Table 1, the employment of a simplistic tweet extraction approach exhibits superior performance compared to the utilization of the K-means extraction approach, yielding improvements of 5%, 2.5%, and 3.3% in accuracy, F1 score, and precision, respectively, over the BiLSTM model. In terms of recall, the difference between the two methods is negligible. We hypothesize that this discrepancy could be attributed to K-means clustering, which identifies text closely related to the classification but, at the same time, disrupts contextual coherence to some degree, leading to the deterioration of results.

the model to pay less attention to them. From a derivative perspective, once the model correctly classifies the current sample (just passed the 0.5 threshold), Dice Loss leads the model to pay less attention to it, unlike cross-entropy, which encourages the model to approach the two endpoints, 0 or 1. This effectively prevents the training of the model from being dominated by numerous straightforward samples.

3.4 User Information Model

After considering user behavioral data, we extracted relevant features pertaining to their social interactions, such as the number of followers and friends. Furthermore, we took into account user-generated actions, including the quantity of tweets posted and tweets favorited. These extracted features were utilized as inputs for the Bidirectional Gated Recurrent Unit (BiGRU) [9].

Both GRU and LSTM employ gating mechanisms to capture interdependencies among inputs, with GRU being a simplified variant of LSTM. Given the relatively straightforward nature of user behavioral data, we posit that the Bidirectional GRU is better suited for capturing relationships among these features.

Subsequently, the features derived from the Bidirectional GRU were fed into a fully connected layer. The resulting output from this layer serves as a guiding factor for classifying users within the historical posting model.

4 Experimental Setup

4.1 Dataset

We have reprocessed the extensive publicly available depression dataset proposed by [29]. These tweets were collected and labeled by the authors on Twitter, while also retrieving the user's historical tweets within a month. The dataset consists of three parts: (1) **Depression Patient Dataset D1**, comprising 2506 labeled samples of depressed users and their tweets; (2) **Non-depression Patient Dataset D2**, comprising 4166 labeled samples of non-depressed users and their tweets; and (3) **Depression Patient Candidate Dataset D3**. The author constructed a large-scale unlabeled depression candidate dataset containing 58,810 samples. In our experiments, we only utilized the labeled datasets: D1 and D2. We preprocessed the datasets by removing users with fewer than ten posting histories, users without an anchor tweet, or users posting tweets in languages other than English. Additionally, we removed emojis from the data to eliminate any impact on the experimental results, thus ensuring that we have sufficient statistical information related to each user. Finally, for balanced data experiments, we considered only 4000 user samples, with 2000 samples each for depressed and non-depressed users. For unbalanced data experiments, we explored the ratios of depressed users to non-depressed users at 1:2 and 1:4, with 2000 samples for non-depressed users in both cases. For testing purposes, we randomly divided the dataset into a training set (80%) and a test set (20%).

The mechanism of attention allows assigning distinct weights to each input feature and reflects the correlation between features and outcomes.

Data Imbalance. The phenomenon of data imbalance is quite common within social media datasets. This imbalance gives rise to the following two issues:

(1) Disparity between training and testing procedures: Under the influence of imbalanced data, models tend to converge towards points that strongly favor classes with the majority labels. This, in effect, creates a disparity between the training and testing processes. During training, each training instance contributes equally to the objective function, whereas during testing, F1 equally weighs the contributions of positive and negative samples.
(2) Excessive impact of simple negative samples on the model: As pointed out by [23], an abundance of negative samples implies a large quantity of straightforward negatives. Consequently, an overwhelming proportion of the loss stems from these numerous simple negative samples, thereby dominating the gradients and hindering the model from adequately learning how to differentiate between positive samples and challenging negative samples. Both cross-entropy (CE) and maximum likelihood estimation (MLE), which are extensively utilized loss functions in machine learning, fail to address these two issues.

Focal Loss and Dice Loss are two deliberately designed loss functions aimed at mitigating the imbalance between positive and negative samples during the one-stage classification process.

The primary objective of Focal Loss is to diminish the weight of easy samples, thereby focusing the training on the negation of difficult samples. To be more precise, Focal Loss introduces a modulation factor $(1-p_t)^\gamma$ into the cross-entropy loss, where $\gamma \geq 0$ represents an adjustable focal parameter. The general form of Focal Loss can be expressed as follows:

$$FL(p_t) = -(1-p_t)^\gamma log(p_t). \tag{4}$$

Dice Loss, on the other hand, contemplates the classification task from a distinct perspective. In this framework, categorizing a sample as negative is contingent solely upon its probability being less than 0.5; there is absolutely no need to drive it towards 0. Furthermore, considering that the primary objective is to mitigate the data imbalance issue within the dataset and, consequently, enhance the effectiveness of the F1 evaluation metric, Dice Loss is designed to exert a direct impact on F1.

Consequently, the general formulation of Dice Loss has been derived as follows:

$$Dice(p_t) = \frac{2(1-p_t)p_t \cdot y_t + \epsilon}{(1-p_t)p_t + y_t + \epsilon}, \tag{5}$$

where p_t represents the estimated probability, incorporated ϵ acts as a smoothing term, and y_t denotes the true label. The term $(1-p_t)$ serves as a scaling factor. For uncomplicated samples (when p_t approaches 1 or 0), $(1-p_t)p_t$ prompts

of tweets articulating their emotions over several days. Considering the token length constraints inherent in BERT [10] and the acknowledged temporal impact on a user's narrative, we judiciously adopt a straightforward strategy: selecting a user's most recent 20 tweets as the primary method for tweet processing. In cases where a user has fewer than 20 tweets, we employ padding techniques to ensure completeness of the dataset.

Tweet Filtering Based on K-Means Clustering. We acknowledge that not all of a user's posts are necessarily relevant to the point we focus on. For instance, a depressed individual might also publish tweets expressing positive emotions, such as 'Today's weather is lovely!' In order to mitigate the influence of such tweets on our determination of a user's depressive status, we endeavor to filter out tweets that more accurately portray the user's identity. Since we cannot introduce user labels during the processing phase, we adopt an unsupervised approach to analyze users' historical posts. Consequently, we employ the K-means clustering method as our second approach to tweet processing. We select a user's most recent 50 posts, tokenize them using BERT, apply the K-means algorithm to cluster the tweets into two categories, and then extract the 20 tweets closest to the cluster centroids. Should there be an insufficient number of tweets remaining, we will once again utilize padding to complete the dataset.

3.3 Historical Posting Information Model

We employed a pre-trained BERT model and a bidirectional LSTM (BiLSTM) based on an attention mechanism to process the input, capturing sequential information such as sentence context. Moreover, in light of the minority representation of depression patients, we tackled the prevalent issue of imbalanced data in the depression dataset by adopting the Focal Loss and Dice Loss techniques, as introduced in the works by [18,20], respectively.

Classification Module Based on Pre-trained Bidirectional LSTM. From the embedding layer of BERT, the extracted features are passed to the Bidirectional Long Short-Term Memory (BiLSTM), which is an RNN designed to capture sequential information and the long-term dependencies within sentences. Comprising the Bidirectional LSTM are the forward and backward LSTMs, each one independently updating the input x_i at time t:

$$forward(h_t) = LSTM(x_t, forward(h_{t-1})). \tag{1}$$

$$forward(h_t) = LSTM(x_t, backward(h_{t-1})). \tag{2}$$

After BiLSTM processing, the hidden state h_t at time t is a concatenation of the states $\overrightarrow{h}$ and $\overleftarrow{h}$ obtained from the forward LSTM and backward LSTM, respectively. The representation of the i-th word is as follows:

$$h_t = forward(h_{t-1}). \oplus backward(h_{t-1}). \tag{3}$$

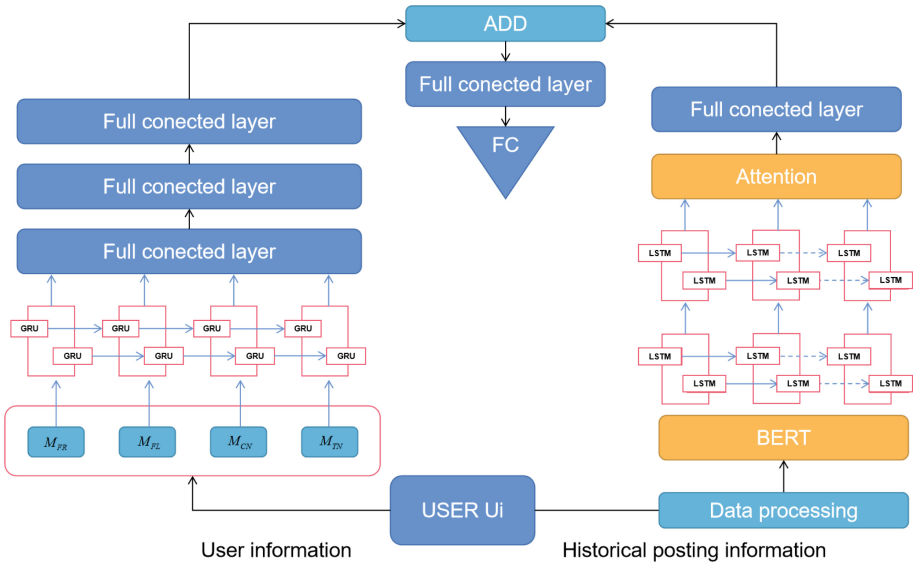

Fig. 1. We have integrated the user information and the user's past tweets of each user to forecast their labels. The user's historical tweets are processed through a tweet processing procedure and then handled by a classification model based on pre-training and bidirectional LSTM. The user's information, on the other hand, is processed using a model based on bidirectional GRU.

extraction, tweet filtering based on K-means clustering. Furthermore, we also incorporate user behavioral metrics, including the number of friends, followers, favorited contents, and the frequency of posts within a month, denoted as $M_{FR}, M_{FL}, M_{CN}, M_{TN}$ respectively.

3.2 Data Processing

The analysis of users' historical tweets constitutes a pivotal aspect of the depression assessment process. Consequently, our efforts are directed towards scrutinizing the past tweets of individuals experiencing depression, in comparison to those of mentally healthy users. This endeavor seeks to afford us a more profound understanding of the behavioral patterns manifested by individuals grappling with depression. Following this analysis, our objective is to leverage these insights to enhance the efficacy of depression detection methodologies.

Conducting Direct Tweet Extraction. Primarily, our initial consideration revolves around a streamlined approach to tweet handling, specifically involving the extraction of a defined quantity of tweets to encapsulate the entirety of a user's Twitter activity. We posit that the temporal dynamics of a user's posts exert a substantial influence on the thematic content of their tweets. To illustrate, an individual experiencing depression may consecutively share a series

to effectively capture both local contextual features and long-range dependencies [31,38].

Moreover, attention mechanisms [1,34,37] are employed to enable models to focus on the most salient aspects of the input. Additionally, multi-task learning is harnessed to jointly train models alongside other auxiliary tasks, such as statistical feature classification [35] and depression causation prediction [36], which yield supplementary insights for depression detection.

Recently, researchers have gathered data through online surveys and online social discourse, leveraging the substantial number of users and tweets on social platforms to obtain an ample and longitudinally sampled dataset. This approach effectively addresses the previously mentioned issue of collecting depression-related data. One study collected data from over 4,000 individuals, encompassing both depressed and non-depressed users on Twitter. [30] The dataset comprised over 2,000 samples from individuals diagnosed with depression, encompassing personal information and an anchor tweet to determine their depressive status. Additionally, all tweets posted within the 30 days preceding the anchor tweet were collected, based on the clinical characteristic of prolonged mood despondency among depression patients. For non-depressed users, relevant personal information and tweets within the same 30-day period were collected. However, this dataset did not account for the issue of data imbalance. Building upon this dataset, [39] further integrated personal information features and textual features, effectively removing redundant features through the utilization of k-means algorithm and the Bart summarization model [17], thereby improving the accuracy of depression identification. Our work builds upon the aforementioned research, focusing primarily on addressing data redundancy and data imbalance issues.

3 Method

In this section, we present our model for depression. Our model utilizes both the users' personal information and their historical posts as the foundation for detecting depression. Figure 1 illustrates our model diagram.

3.1 Task Definition

For social media datasets, users' posts often exhibit redundancy, irrelevance to their status, and may even contain unusable information, posing significant challenges for researchers to effectively extract user information. Herein, we have established the relevant symbols as defined in the article. We assume a user, denoted as U_i, has a total of n tweets in their history: $[T_1, T_2, \ldots, T_t, \ldots, T_n]$, where the t-th tweet represents the user's recent post. Our objective is to determine whether user U_i is a depression sufferer, for which we define the label as $y_i \in depression, undepression$. To achieve our goal, we amalgamate each user's profile information with their historical posts. We explore three approaches to process a user's historical tweets to address this issue: conducting direct tweet

extract activity information typically mandate real-time tracking for a duration exceeding two weeks. This extended timeframe, coupled with the requisite high level of patient cooperation, introduces significant costs and complexity into the research process.

Efforts to enhance the precision and applicability of AI-based depression identification must contend with these multifaceted challenges.

Our contributions are as follows:

1. We present a comprehensive depression assessment model that concurrently leverages users' historical tweets and their personal information. To achieve this, we meticulously devise distinct models for both historical tweets and user information. Our approach involves the integration of two discrete methodologies, one dedicated to handling multiple tweets and the other addressing the inherent challenge of imbalanced depression data.
2. We reprocess the depression tweet dataset to enhance its practical utility.
3. Through rigorous experimentation, we showcase the efficacy of our model in discriminating depression, yielding compelling empirical results and partially alleviating the associated challenges.

2 Related Work

In the field of psychology, early scholars have observed the theoretical correlations between mental health conditions and specific linguistic attributes, such as the presence of "depressive language" [2] advanced cognitive therapy and emphasized the significance of the frequency of negatively-valenced words, while other researchers [25] focused on the utilization of first-person pronouns and the patients' negative anticipations. Subsequent empirical investigations have validated these hypotheses and revealed associations between specific linguistic characteristics and the mental states of patients. Consequently, numerous studies utilize social media as a rich source of textual data, employing online user-generated posts for the manual analysis of mental health conditions. [7,26].

However, due to the burgeoning volume of online texts and the sensitivity of mental health conditions, manual text analysis and large-scale psychiatric interventions are no longer tenable. Consequently, Natural Language Processing (NLP) and text mining technologies have been harnessed to automate the analysis of mental health from social media data. While these approaches are not intended for definitive diagnoses, they do offer assistance in early detection [11,22,27]

Advancements in the realm of deep learning also bolster tasks related to mental health. The most recent methodologies employ deep learning models to automatically capture latent semantic information without the need for explicit feature engineering. Some studies utilize Convolutional Neural Networks (CNN) [33] or Recurrent Neural Networks (RNN) [8], including Long Short-Term Memory (LSTM) [12] and Gated Recurrent Unit (GRU) [28], to discern depression. Researchers also explore hybrid architectures combining CNN and RNN

1 Introduction

With the advancement and progress of society, people's material living standards are constantly improving, and psychological issues are receiving increasing attention. Psychological disorders are prevalent among young individuals, with approximately 75% of cases emerging during adolescence [15]. According to estimates by the World Health Organization, depression is one of the most prevalent psychological disorders, and by the year 2030, depression is projected to become the leading burden of disease globally [16,21].

Depression is characterized by significant and enduring mood melancholy, with symptoms encompassing sleep disturbances, appetite changes, and mental turmoil [3–5]. Despite its high prevalence, there is evidence indicating that 60% of severely depressed adolescents do not receive treatment.

Depression possesses a covert nature, and its occurrence is influenced by intricate factors such as heredity, gender, living environment, and physical ailments, rendering its diagnosis exceedingly challenging [4–6]. Presently, an accurate diagnosis of depression necessitates psychiatric practitioners to employ systematic inquiries, psychiatric examinations, and supplementary assessments, such as the Hamilton Rating Scale for Depression (HAMD) and the Patient Health Questionnaire-9 (PHQ-9) self-rating scale. Thus, the diagnostic evaluation heavily relies on patients' self-reported severity of depressive symptoms or clinical judgment regarding symptom severity. However, the advent of artificial intelligence-based approaches has presented the potential for objective diagnosis.

We focus on depression detection based on social networks. Recently, [14] discovered that a lack of social interaction increases the risk of depression. [24] analyzed the behavior and language usage of depressed users on Twitter. People's tweets on social networks such as Facebook, Twitter, and Weibo can be used to assess the risk of various mental health issues, such as depression and anxiety. [32] employed lemmatization tools to vectorize more recent tweets, reducing redundant features. [13] utilized a multimodal model and applied reinforcement learning to merge textual and image features of tweets, thereby improving the accuracy of depression identification [19].

The efficacy of depression identification through artificial intelligence remains suboptimal, encountering several noteworthy challenges. These challenges may be succinctly outlined as follows: **Limited Sample Size:** The recruitment of patients poses a substantial hurdle due to ethical concerns within the medical field. Consequently, a pervasive issue in depression studies is the constraint imposed by small sample sizes. This limitation complicates the attainment of definitive conclusions regarding individual-level depression diagnoses. **Data Complexity:** The datasets employed in these studies often exhibit a profusion of irrelevant features and noise. This characteristic not only augments the computational intricacy of algorithms but also compromises the predictive performance of models. Additionally, an inherent imbalance exists within the dataset, stemming from a lower representation of depression cases. **Temporal Constraints:** Extracting nuanced daily life characteristics of patients necessitates a protracted timeline. For instance, methodologies reliant on mobile devices to

Sentiment Analysis Based on Social Media - Early Stress and Depression Detection

Zixuan Li[1,3], Yuxuan Hu[1,3], Chenwei Zhang[1,3], Chengming Li[1,2(✉)], and Xiping Hu[1,2(✉)]

[1] Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China
{lizx76,huyx55,zhangshw7}@mail2.sysu.edu.cn, {licm,huxp}@smbu.edu.cn
[2] Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China
[3] Sun Yat-sen University, Shenzhen, Guangdong, China

Abstract. Depression has recently gained significant attention as a condition marked by persistent and profound mood disturbances. Extensive research suggests that depression can influence individuals' online speech behavior, manifested through the use of depressive language and a reduction in posting frequency. Our system seamlessly integrates various sources of information, including historical tweets and user profile data. Concerning historical tweets, we propose two methods to navigate the extensive and intricate user tweet history. Our findings indicate that these methods yield more pertinent user information. Subsequently, we input this information into our meticulously constructed deep learning classification model. This model is built upon a pre-trained BERT (Bidirectional Encoder Representations from Transformers) and a bidirectional LSTM (Long Short-Term Memory) model that incorporates attention mechanisms. In the context of user information, we extract relevant details and directly incorporate them into a deep learning model based on bidirectional GRU (Gated Recurrent Unit) and MLP (Multi-Layer Perceptron). Concurrently, to address the challenge of imbalanced depression datasets, we introduce Focal Loss and Dice Loss. Our experimental results underscore the effectiveness of these loss functions in our model. To validate the efficacy of our system, we reprocess the depression tweet dataset and conduct a series of experiments. Through these experiments, we conclusively demonstrate the robustness of our model, effectively mitigating the challenge of sample imbalance to a considerable extent.

Keywords: Deep learning · Social network · Depression recognition · Data imbalance

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 26–39, 2024.
https://doi.org/10.1007/978-3-031-65126-7_3

26. Perkins, T.K., Kern, L.R.: Widths of hydraulic fractures. J. Petrol. Technol. **13**(09), 937–949 (1961)
27. Nordgren, R.P.: Propagation of a vertical hydraulic fracture. Soc. Petrol. Eng. J. **12**(04), 306–314 (1972)
28. Hudson, J.A.: A critical examination of indirect tensile strength tests for brittle rocks [Ph. D. Thesis]. University of Minnesota, Minneapolis (1984)
29. Wang, C., Wang, R., Wang, C.: Development of multiple-diameter core hydraulic fracturing machine to test tensile strength of rocks. Chin. J. Rock Mech. Eng. **36**(S1), 3321–3331 (2017)
30. Cuisiat, F.D., Haimson, B.C.: Scale effects in rock mass stressmeasurements. Int. J. Rock Mech. Min. Sci. Geomech. Abst. **29**(2), 99–117 (1992)
31. Hou, B., Chen, M., Wan, C., Sun, T.: Laboratory studies of fracture geometry in multistage hydraulic fracturing under triaxial stresses. Chem. Technol. Fuels Oils **53**(2), 219–226 (2017)
32. Park, J.Y., Tuell, G.: Conceptual design of the CZMIL data processing system (DPS): algorithms and software for fusing lidar, hyperspectral data, and digital images. Proc Spie **7695**(5), 731–739 (2010)
33. Qin, H.: Constructions of uniform designs with mixed levels. Acta Math. Appl. Sin. **28**(4), 704–712 (2005)
34. Montgomery, D.C., Peck, E.A.: Introduction to linear regression analysis (1982)
35. Schmitt, D.R., Zoback, M.D.: Diminished pore pressure in low-porosity crystalline rock under tensional failure; apparent strengthening by dilatancy. J. Geophys. Res. **97**(B1), 273–288 (1992)
36. Ito, T., Satoh, T., Kato, H.: Deep Rock Stress Measurement by Hydraulic Fracturing Method Taking Account of System Compliance Effect. Xie Furen. CRC Press, Boca Raton (2010)
37. Zhu, X., Zhang, J., Feng, J.: Multiobjective particle swarm optimization based on PAM and uniform design. Math. Probl. Eng. **2015**, 1–17 (2015)
38. Yang, L., Pan, F., Weifeng, J., Shengwei, S., Yong, Z., Tao, Z.: Predictive method of nonlinear system based on artificial neural network and svm. Oxidat. Commun. **39**(1Appa), 1226–1235 (2016)

4. Haimson, B.C.: Hydraulic Fracturing in Porous and Nonporous Rock and its Potential for Determining In Situ Stresses at Great Depth. University of Minnesota, Minneapolis (1968)
5. Haimson, B.C., Fairhurst, C.: Initiation and extension of hydraulic fractures in rocks. Soc. Petrol. Eng **9**, 310–318 (1967)
6. Von Schonfeldt, H., Fairhurst, C.: Field experiments on hydraulic fracturing. Soc. Petrol. Eng. AIME **12**(1), 69–77 (1972)
7. Wang, C.: Brief review and outlook of main estimate and measurement methods for in-situ stresses in rock mass. Geol. Bull. China **60**(5), 971–996 (2014)
8. Wang, J., Li, H., Wang, Y., Li, Y., Jiang, B., Luo, W.: A new model to predict productivity of multiple-fractured horizontal well in naturally fractured reservoirs. Math. Probl. Eng. **2015**, 1–9 (2015)
9. Xie, F., Chen, Q.: Study on the Crustal Stress Environment in China. Geological Press, Beijing (2003)
10. Jaeger, J.C., Cook, N.G.W., Zimmerman, R.W.: Fundamentals of Rock Mechanics. Blackwell Publishing, London (2007)
11. Wang, C., Song, C., Xing, B.: Compliance of drilling-rod system for hydro-fracturing in situ stress measurement and its effect on measurements at great depth. Geoscience **26**(4), 808–816 (2012)
12. Zoback, M.D., Pollard, D.D.: Hydraulic fracture propagation and the interpretation of pressure-time records for in-situ stress determinations. In: 19th US Symposium on Rock Mechanics (USRMS), pp. 14–22. American Rock Mechanics Association (1978)
13. Ito, T., Hayashi, K.: Physical background to the breakdown pressure in hydraulic fracturing tectonic stress measurements. Int. J. Rock Mech. Mining Sci. Geomech. Abst. **28**(4), 285–293 (1991)
14. Chang, C., Jo, Y., Oh, Y., Lee, T.J., Kim, K.: Hydraulic fracturing in situ stress estimations in a potential geothermal site, Seokmo Island, South Korea. Rock Mech. Rock Eng. **47**(5), 1793–1808 (2014)
15. Zhou, L., Ding, L., Guo, Q.: Experimental study of absolute rock stress measurements under different fracture media. Rock Soil Mech. **10**, 2869–2876 (2013)
16. Zhang, J.: Analysis of the Hydromechanics Factors Impact on Hydraulic Fracturing In-situ Stress Measurement. China University of Geosciences, Beijing (2018)
17. Matsunaga, I., Kobayashi, H., Sasaki, S.: Studying hydraulic fracturing mechanism by laboratory experiments with acoustic emission monitoring. Int. J. Rock Mech. Min. Sci. Geomech. Abst. **7**, 909–912 (1993)
18. Ishida, T., Chen, Q., Mizuta, Y.: Effect of injected water on hydraulic fracturing deduced from acoustic emission monitoring. In: Seismicity Associated with Mines, Reservoirs and Fluid Injections. Birkhäuser, Basel (1997)
19. Fang, K.: Uniform Design and Uniform Design Table. Science Press, Beijing (1994)
20. Wang, Z., Fang, K.: Measures of uniformity for uniform designs with qualitative factors. Math. Stat. Manag. **19**(3), 28–32 (2000)
21. Myers, R.H.: Classical and Modern Regression with Applications, 2nd edn. Duxbury Press, Belomont (1994)
22. Liu, Y., Wang, C., Wang, J., Ji, W.: Optimization research on thermal error compensation of FOG in deep mining using uniform mixed-data design method. Math. Probl. Eng. **2019**, 1–6 (2019)
23. Zhang, L., Cai, X.: Uniformity masks design method based on the shadow matrix for coating materials with different condensation characteristics. Sci. World J. **2013**, 1–4 (2013)
24. Khristianovich, S.A., Zheltov, Y.P.: Formation of vertical fractures by means of a highly viscous fluid. In: Proceedings 4th World Petroleum Congress, pp 579–586 (1955)
25. Geertsma, J., De Klerk, F.: A rapid method of predicting width and extent of hydraulically induced fractures. J. Petrol. Technol. **21**(12), 1571–1581 (1969)

experimental results. Meanwhile, neural networks and deep learning algorithms are also considered to analyze and predict rock fracturing values, verifying the accuracy of the rock fracturing model [38].

6 Conclusions

In this paper, we utilized the optimal uniform design method to optimize hydraulic fracturing simulation experiments. The results showed that this design method not only reduced the number of experiments but also improved the uniformity and test effect. It provides a fast and effective way to develop an error compensation formula for various influence elements, aiming to enhance measurement accuracy of the hydrofracture method of geostress measurement.

- The paper proposed an optimal approach for the hydrofracture method of geostress measurement using the uniform design method. Considering these unique properties of the drilling mud, which involves multiple hydraulic elements and values. This paper constructs an experimental plan that can simplify the testing procedure, and decrease implementation fees. As a result, the efficiency of the simulation hydraulic fracturing tests can be significantly enhanced.
- This study examined the impact of various hydrodynamic factors on rock fracturing pressure using test results. Through multivariate regression analysis, an optimal regression model for multiple influencing factors (fracturing fluid viscosity, density, axial load, and injection rate) was obtained. Additionally, the values of instantaneous rock splitting were discussed in depth, and the validity of Perkins-Kern-Nordgren (PKN) classical mechanical model in theoretical analysis was confirmed.

Acknowledgments. This work was supported by Project funded by National Natural Science Youth Foundation of China (41804089), and Geological survey Project of China Geological Survey (DD20230447).

Data Availability. The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest. The authors declare that there are no conflicts of interest regarding the publication of this paper, and the authors confirm that the mentioned received funding in the "Acknowledgment" section did not lead to any conflict of interests regarding the publication of this manuscript.

References

1. Clark, J.B.: A hydraulic process for increasing the productivity of wells. J. Petrol. Technol. **1**(1), 1–8 (1949)
2. Zhao, Z., Guo, J., Ma, S.: The Experimental investigation of hydraulic fracture propagation characteristics in glutenite formation. Adv. Mater. Sci. Eng. **2015**, 1–5 (2015)
3. Hubbert, K.M., Willis, D.G.: Mechanics of hydraulic fracturing. Trans. AIME **210**(1), 153–168 (1957)

To summarize, Eqs. (6) and (7) demonstrate a similar changing pattern as x_1 and x_4, but undergoes an opposite changing pattern in relation to x_2 (fracturing fluid density). The simulation experiments confirmed a strong correlation between rock fracturing pressure and the viscosity, density, and injection rate of the fracturing fluid. These findings support the conclusions of theoretical analysis in Sect. 2. This indicates a good distribution uniformity of experimental points. Additionally, this study validated the efficiency and suitability of the experimental method in establishing a fracture pressure correction formula for various hydrodynamic factors.

5 Discussion

Simulation experiments were conducted to analyze the impact of various factors such as injection rate, density, viscosity, and fracturing fluid medium on hydraulic fracturing. These experiments provide a fast and reliable way to understand the influence of hydrodynamic factors on hydraulic fracturing. A compensation model can be utilized to minimize the interference of hydrodynamic factors and enhance the accuracy of in-situ stress measurement during hydraulic fracturing in practical engineering applications. Consequently, simulation tests have the potential to improve the measurement accuracy of hydraulic fracturing methods. However, this study only considered three fluid mechanics parameters, namely injection rate, fracturing fluid density, and viscosity. Therefore, future research should explore the incorporation of additional hydrodynamic parameters such as hydraulic friction, liquid compressibility, and the effects of different types of fracturing fluid media on the effective fracturing pressure of rocks. These insights are valuable in advancing our understanding of hydraulic fracturing in practical applications.

- In terms of different fracturing fluids, such as hydraulic oil, mud, and aqueous solution, test results from Zhou Longshou (2013) [15] and Zhang Jie, Wang Chenghu et al. (2017) [16] indicate that mud and hydraulic oil lead to higher rock fracturing pressures compared to aqueous solutions. The combination of densities and viscosities of the fracturing fluids greatly affects the rock fracture pressure, while the compressibility of the fracturing fluid also influences the flexibility of the hydraulic fracturing measurement system (Wang Chenghu et al., 2012) [11], thereby affecting the measurement results. This study only considered mud as the fracturing fluid, so future studies should include more representative hydrodynamic parameters and different types of fracturing fluids for a comprehensive analysis of their influence on rock fracture pressure. Additionally, an appropriate correction formula and compensation model should be established for hydraulic fracturing errors under different working conditions.
- The experimental results confirmed that the injection rate of the fracturing fluid has a significant impact on the rock fracturing pressure, with a proportional increase. This finding aligns with the results of hydraulic fracturing tests conducted by several foreign researchers (Ito and Hayashi, 1991; Schmitt et al., 1992; Zo-back et al., 2007) [12–14, 36, 37]. To further enhance the accuracy of future simulation tests and reduce losses associated with hydraulic friction, especially head loss, it is recommended to install a high-precision pressure sensor in the fracturing test section. This enhancement will allow for a better analysis of the influence of injection rate on the

Table 4. Multi-factor VIF value

VIF	Value
x_1	1.0027
x_2	1.0009
x_3	1.0035
x_4	1.0021

polynomial. The Matlab Linear-Model.stepwise function was utilized to perform a multivariate quadratic polynomial regression of the data presented in Table 4. Based on this regression analysis, the stepwise regression method of the LinearModel class object was employed to establish the Eq. (6), to show the relationship between factors (x_1, x_2, x_3, x_4) and fracturing value (y).

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_3^2 + b_6x_1x_4 + b_7x_4^2 \tag{6}$$

Based on the stepwise regression calculation results, the resulting equation for multivariate polynomial regression can be expressed as follows:

$$y = 17.937 + 0.023x_1 - 10.266x_2 + 2.054x_3 + 1.598x_4 - 0.361x_3^2 - 0.067x_1x_4 + 18.535x_4^2 \tag{7}$$

Furthermore, Eq. (8) produces a $p\text{-}value_2 = 0.000405$, and $p\text{-}value_2 << 0.05$ (significance level). Figure 3 illustrates the regression fitting plots of Eqs. (6) and (7), demonstrating a higher degree of fit in the latter. Therefore, Eq. (7) is an optimal fitting formula for (y) and (x_1, x_2, x_3, x_4) in this design.

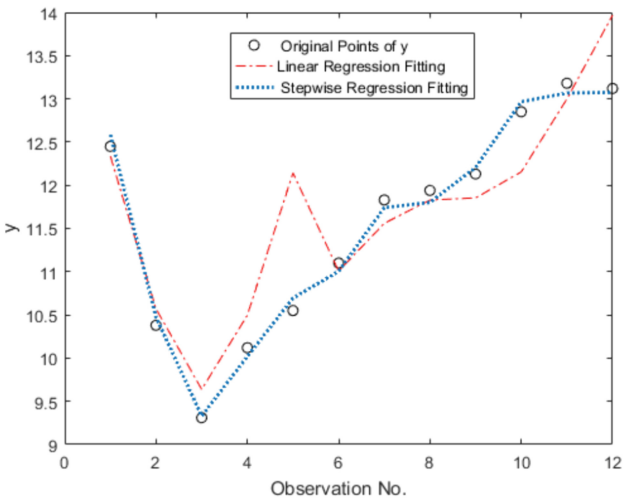

Fig. 3. The optimal fitting formula for linear regression and stepwise regression fitting

4 Results Analysis

4.1 Experimental Results

Table 3 presents results of the experiments using the uniform design method involved in Sect. 3, showcasing the obtained effective rock fracturing value.

Table 3. Table of hydraulic fracturing values of the simulation experiments

No.	Influencing factors				Results
	Viscosity(g/cm^3)	Density(mPa•s)	Axial Compression(MPa)	Injection Rate(MPa/s)	Fracturing pressure(MPa)
1	280	1.4	1.2	0.1	12.46
2	170	1.6	3.6	0.1	10.35
3	70	1.6	1.2	0.55	9.34
4	70	1.2	2.4	0.05	10.12
5	70	1.0	4.8	0.1	10.58
6	150	1.2	4.8	0.4	11.12
7	280	1.6	4.8	0.1	11.85
8	130	1.2	2.4	0.2	11.91
9	150	1.2	3.6	0.2	12.1
10	170	1.2	2.4	0.4	12.88
11	170	1.0	1.2	0.05	13.19
12	280	1.0	3.6	0.55	13.15

4.2 Multivariate Polynomial Regression

Regression analysis is a method used to establish the relationship between the dependent variable y and the independent variables (x_1, x_2,…, x_i) [34–36]. In Eq. (6), y represents actual demonstration value of the rupture pressure, x_1 represents the density, x_2 represents the viscosity, x_3 represents the axial pressure, and x_4 represents the injection speed. Table 4 underwent multiple linear regression and multivariate polynomial regression to determine the respective fitting models. These models were then compared to obtain the optimal fitting formula.

A regression model was subjected to a multicollinearity diagnosis using the variance inflation factor (VIF) method, which resulted in Table 4. Generally, if VIF < 5, there is no collinearity. The independent variables in Table 4 had VIFs below 5, indicating the absence of multicollinearity in the model.

In order to enhance the non-linear terms in the model, a stepwise regression approach was employed to conduct a generalized linear regression analysis using a quadratic

Table 1. Elements and their numerical value

Element	Level	Parameter value
Viscosity	4	70; 150; 170; 280
Density	4	1.0; 1.2; 1.4; 1.6
Axial compression	4	1.2; 2.4; 3.6; 4.8
Injection rate	6	0.05; 0.1; 0.2; 0.25; 0.4; 0.55

In order to account for the numerous elements and their values, using mud as fracturing fluid, an optimized experimental scheme based on mixed-level uniform was developed. Using the Data Processing System (DPS) software [33], a total of 12 experiments were conducted, as shown in Table 2. The constructed optimal mixed-level uniform design table $U_{12}^{*}\left(6 \times 4^{3}\right)$ was subjected to a maximum of 1000 iterations.

Table 2. Table of Influencing elements in $U_{12}^{*}\left(6 \times 4^{3}\right)$

No.	Influencing elements			
	x_1	x_2	x_3	x_4
1	4	3	1	3
2	3	4	3	3
3	1	4	1	5
4	1	3	2	1
5	1	1	4	3
6	2	3	4	6
7	4	4	4	2
8	2	2	2	4
9	2	2	3	1
10	3	2	2	6
11	3	1	1	2
12	4	1	3	5

A smaller D implies better uniformity of the experimental design [19, 20]. By calculating the Eq. (6), we obtained that $D^{*} = 0.1713$ for the $U_{12}^{*}\left(6 \times 4^{3}\right)$, which exhibits a good distribution uniformity.

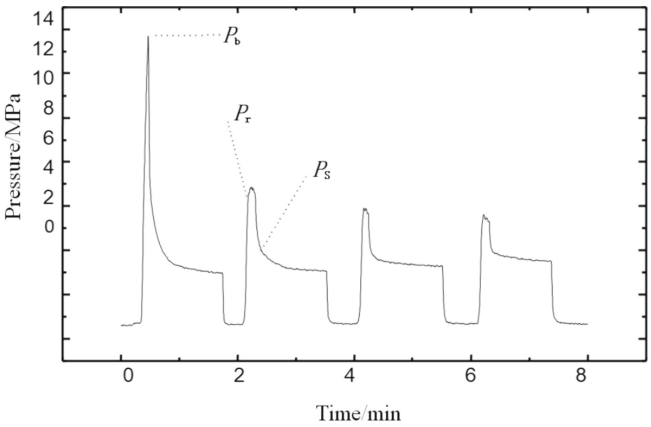

Fig. 2. Typical Pressure-Time Record Curve in Hydraulic Fracturing

and the pressure value at this time is recorded as the instantaneous closure pressure P_s. After releasing the pressure, reloading causes the fracture to reopen, and the pressure value at this time is recorded as the reopening pressure P_r.

According to the elastic theory and the PKN mechanical model, as shown in Fig. 3, the fracturing pressure of the rock in the fracturing section is:

$$P_b = 3\sigma_h - \sigma_H + T \tag{5}$$

Among them, σ_H and σ_h are the maximum and minimum horizontal principal stresses, respectively, and T is the tensile strength of the rock. The fractures induced by hydraulic fracturing are vertical fractures and perpendicular to the direction of the minimum horizontal principal stress. Equation (5) indicates that the fracturing pressure of rocks is independent of the size of the borehole and the elastic modulus of the rock, and is mainly determined by the tensile strength of the rock and the magnitude of the in-situ stress around the borehole.

3 Optimal Design of the Testing

The high pressure fluids are commonly applied in hydraulic fracturing simulation experiments, including clean water, hydraulic fluid, carboxymethyl cellulose (CMC) aqueous solution, and drilling mud [30–32]. The density and viscosity of the mud medium can be adjusted according to the requirements of the simulation experiment. For these fracturing fluid media, only a small number of factors and tests are required, so conventional comprehensive experimental methods can be used for their respective simulation experiments. In contrast, there are more parameters and theirs values in the mud medium, so an optimal design based on uniform design method is suitable for the testing.

The theoretical analysis of the PKN model revealed that when using mud as the fracturing fluid medium in the simulation test, it requires three hydrodynamic factors: density, injection rate, and viscosity, as well as a factor of loading axial compression. Different numerical values of each element are presented in Table 1.

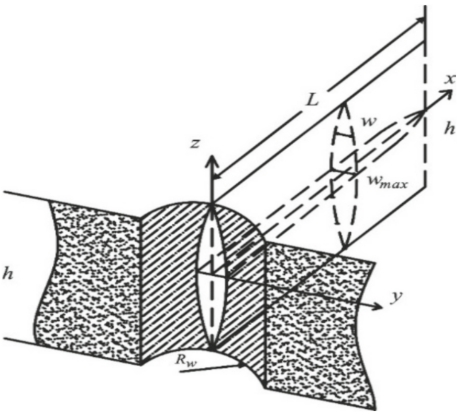

Fig. 1. PKN classical mechanical model

Here, *q(x, t)* represents the volume of fluid flowing through the cross-section of the crack, $q_l(x, t)$ represents the volume of fluid lost per unit length of the crack, and *A(x, t)* represents the cross-sectional area of the crack. When there is no fluid leakage, the length of the crack L, its local width w, and the pore pressure P_w can be calculated [28, 29]:

$$L = 0.68\left[\frac{GQ^3}{(1-\upsilon)\mu h^4}\right]^{\frac{1}{5}} t^{\frac{4}{5}} \tag{2}$$

$$w = 2.5\left[\frac{(1-\upsilon)\mu Q^2}{Gh}\right]^{\frac{1}{5}} t^{\frac{1}{5}} \tag{3}$$

$$p_w = 2.5\left[\frac{G^4\mu Q^2}{(1-\upsilon)^4 h^6}\right]^{\frac{1}{5}} t^{\frac{1}{5}} \tag{4}$$

The following variables are used in this context: G (shear modulus), υ (Poisson ratio), h (length), Q (injection rate), and μ (viscosity).

2.2 Principles of Hydraulic Fracturing Measurement

The basic principle of in-situ stress measurement based on hydraulic fracturing involves placing drill rods and packers into a borehole using a drilling rig to measure their positions. Fluid is injected into the packers through a loading control system, isolating a test section within the borehole, and the fluid is further injected into the test section until fracturing occurs.

As shown in Fig. 2, the first highest pressure value is recorded as the fracturing pressure P_b. Then the pressure drops rapidly to a state of fluid seepage into the fracture and remains constant. At this point, the pump is turned off to stop loading, and the pressure in the fracturing section decreases rapidly, causing the fracture to close quickly. When the fracture is in the near-closed state, the rate of pressure decrease slows down,

of error include: (1) the drill pipe and the packer deformation [9–11]; (2) variations in the determination method of the measurement curve during data analysis [12]; and (3) different category and the associated performance factors of the fracturing fluids [13–17].

Many researchers have dedicated themslves to explore the fracturing fluids impact on rock fracturing. For instance, Ito (1991) and Chang (2014) suggested that increasing the injection rate of fracturing fluid and considering factors like flow rate, viscosity, and density can enhance the tensile strength of the rock. Zhou et al. (2013) and Zhang (2018) conducted tests using different density mud media as fracturing fluids and observed significant variations in rock fracturing behavior. Matsagaga (1993) and Ishida et al. (1997) verified the impact of fracturing fluid viscosity on rock fracturing through oil drilling experiments. Wang (2012) and Zhou (2013) analyzed the error in stress measurement caused by the compressibility of clear water used as a fracturing fluid and its effect on system flexibility. These studies contribute to a better understanding of fluid mechanics factors in accurate rock fracturing measurements.

In summary, the hydrodynamic factors that influence rock fracturing during hydraulic fracturing include flow velocity, viscosity, density, and compressibility. Conducting simulation experiments based on these factors is crucial for understanding their impact on rock fracturing. However, these experiments can be destructive to the testing core, making them complicated and costly to design. To address this, the uniform design method has been proposed as an experimental design approach that evenly spreads test points throughout the range of variables, requiring fewer trials compared to other methods [19, 20]. In particular, the design aims to conduct trials with many experimental factors and a large number of levels, with fewer trials required compared with the orthogonal design or comprehensive design methods [21–23]. In this study, a uniform table for experiment design is used to combine selected hydrodynamic factors of the fracturing fluid with the factor of horizontal pressure. This approach reduces the test times while ensuring their effectiveness and significantly improving efficiency. The results of these experiments are then analyzed to determine the effects of the factors on rock fracturing value.

2 Error Analysis of Hydraulic Fracturing Theory

2.1 PKN Mechanical Model

The borehole used to measure hydraulic fracturing in-situ stress was typically vertical and primarily influenced by the maximum horizontal principal stress, and the minimum stress, and minimum is same as it is [24]. The fracturing crack was vertical because it was perpendicular to the minimum horizontal principal stress plane [25, 26]. The authors used the PKN classical mechanical model [27, 28] to analyze how fluid mechanics affects the fracture crack and its fracture pressure in this paper.

Figure 1 illustrates the PKN fracturing crack model [27, 28]. Nordgren (1972) obtained the fluid's continuity equation in the crack, ignoring the compression properties of the fracturing fluid [28]:

$$\frac{\partial q}{\partial x} + q_t + \frac{\partial q}{\partial t} = 0 \tag{1}$$

Optimal Design of Hydraulic Fracturing Simulation Experiments for In-Situ Stress Measurement

Yang Li[1], Daji Zhang[3], and Yimin Liu[2(✉)]

[1] Institute of Exploration Technology, CGS, Chengdu 611734, China
[2] Chengdu Technological University, Chengdu 611730, China
153973418@qq.com
[3] Chengdu Rail Transit Group Co., Ltd., Chengdu 610036, China

Abstract. In order to account for a large number of hydrodynamic influencing factors with multiple levels in rock fracturing experiments, the uniform design method is frequently utilized instead of conventional methods like comprehensive and orthogonal designs, as they significantly impact the experimental effects. Based on the Perkins-Kern-Nordgren (PKN) model, the influencing factors of injection rate, viscosity, and density of the fracturing fluid, along with their corresponding parameter values or levels, were taken into consideration to construct an optimal table $U_{12}^{*}\left(6 \times 4^{3}\right)$ for experiment design. Subsequently, an optimized experimental scheme was developed. The experimental results based on this design were analyzed using multiple regression analysis to establish an optimal regression equation for the influencing factors (x_1, x_2, x_3, x_4, representing fluid viscosity, density, loading axial compression, and injection rate, respectively) and to determine the corresponding rock fracturing value (y). This indicates a good distribution uniformity of experimental points. Additionally, this study validated the efficiency and suitability of the experimental method in establishing a fracture pressure correction formula for various hydrodynamic factors, and it is also a precise approach for geostress measurements.

Keywords: Uniform design method · Mixed-level · In-situ stress measurement · Rock fracturing

1 Introduction

The stress stored in the interior of a rock mass without disturbance is referred to as in-situ stress, which has multiple sources and is influenced by various factors, resulting in a complex and variable distribution of stress in the Earth's crust [1]. Hydraulic fracturing is a crucial technique for measuring in-situ stress in various geological structures such as hydropower stations, tunnels, chambers et al. [2–4]. This approach offers an efficient test procedure, along with straightforward data analysis procedures. However, several factors can influence the accuracy of rock fracturing measurements [5–7]. The primary sources

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 15–25, 2024.
https://doi.org/10.1007/978-3-031-65126-7_2

33. Xia, R., Liu, Y.: A multi-task learning framework for emotion recognition using 2d continuous space. IEEE Trans. Affect. Comput. **8**(1), 3–14 (2017). https://doi.org/10.1109/TAFFC.2015.2512598
34. Xie, B.: Research on key technology of Mandarin speech emotion recognition. Ph.D. thesis, Zhejiang University (2006)
35. Xu, Y., Su, H., Ma, G., Liu, X.: A novel dual-modal emotion recognition algorithm with fusing hybrid features of audio signal and speech context. Complex Intell. Syst. **9**(1), 951–963 (2023)
36. Yu, W., et al.: Ch-sims: a chinese multimodal sentiment analysis dataset with fine-grained annotation of modality. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 3718–3727 (2020)

16. Li, Y., Zhao, T., Kawahara, T., et al.: Improved end-to-end speech emotion recognition using self attention mechanism and multitask learning. In: Interspeech, pp. 2803–2807 (2019)
17. Liu, A.T., Yang, S.W., Chi, P.H., Hsu, P.C., Lee, H.V.: Mockingjay: unsupervised speech representation learning with deep bidirectional transformer encoders. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6419–6423. IEEE (2020)
18. Liu, Y., et al.: Make acoustic and visual cues matter: Ch-sims v2. 0 dataset and AV-Mixup consistent module. In: Proceedings of the 2022 International Conference on Multimodal Interaction, pp. 247–258 (2022)
19. Livingstone, S.R., Russo, F.A.: The ryerson audio-visual database of emotional speech and song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in north American English. PLoS ONE **13**(5), e0196391 (2018)
20. Logan, B., et al.: Mel frequency cepstral coefficients for music modeling. In: Ismir, vol. 270, p. 11. Plymouth, MA (2000)
21. Luna-Jiménez, C., Griol, D., Callejas, Z., Kleinlein, R., Montero, J.M., Fernández-Martínez, F.: Multimodal emotion recognition on ravdess dataset using transfer learning. Sensors **21**(22), 7665 (2021)
22. McFee, B., et al.: librosa: audio and music signal analysis in python. In: Proceedings of the 14th Python in Science Conference, vol. 8, pp. 18–25 (2015)
23. Nwe, T.L., Foo, S.W., De Silva, L.C.: Classification of stress in speech using linear and nonlinear features. In: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP 2003), vol. 2, pp. II–9. IEEE (2003)
24. Oh, K.J., Lee, D., Ko, B., Choi, H.J.: A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. In: 2017 18th IEEE International Conference on Mobile Data Management (MDM), pp. 371–375. IEEE (2017)
25. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2009)
26. Pravena, D., Govind, D.: Significance of incorporating excitation source parameters for improved emotion recognition from speech and electroglottographic signals. Int. J. Speech Technol. **20**(4), 787–797 (2017)
27. Singh, A., Liu, H., Plumbley, M.D.: E-panns: sound recognition using efficient pre-trained audio neural networks. arXiv preprint arXiv:2305.18665 (2023)
28. Subakan, C., Ravanelli, M., Cornell, S., Bronzi, M., Zhong, J.: Attention is all you need in speech separation. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 21–25. IEEE (2021)
29. Triantafyllopoulos, A., Schuller, B.W.: The role of task and acoustic similarity in audio transfer learning: Insights from the speech emotion recognition case. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7268–7272. IEEE (2021)
30. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. **30** (2017)
31. Wu, T., Peng, J., Zhang, W., Zhang, H., Tan, S., Yi, F., Ma, C., Huang, Y.: Video sentiment analysis with bimodal information-augmented multi-head attention. Knowl.-Based Syst. **235**, 107676 (2022)
32. Xi, Y., Li, P., Song, Y., Jiang, Y., Dai, L.: Speaker to emotion: domain adaptation for speech emotion recognition with residual adapters. In: 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 513–518. IEEE (2019)

5 Conclusions

In this paper, we effectively applied transfer learning to fine-tune the English pre-trained model, achieving a notable improvement in the F1 score to 0.46, significantly surpassing the baseline of 24%. For future research, we will continue exploring the cross-corpus SER domain and further investigating other deep learning techniques to enhance the performance of the transfer learning models in emotion recognition.

References

1. Amiriparian, S., et al.: Snore sound classification using image-based deep spectrum features (2017)
2. Busso, C., et al.: IEMOCAP: interactive emotional dyadic motion capture database. Lang. Resour. Eval. **42**, 335–359 (2008)
3. Chorowski, J.K., Bahdanau, D., Serdyuk, D., Cho, K., Bengio, Y.: Attention-based models for speech recognition. In: Advances in Neural Information Processing Systems, vol. 28 (2015)
4. Gemmeke, J.F., et al.: Audio set: An ontology and human-labeled dataset for audio events. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 776–780. IEEE (2017)
5. Gong, Y., Chung, Y.A., Glass, J.: Ast: audio spectrogram transformer. arXiv preprint arXiv:2104.01778 (2021)
6. Gong, Y., Lai, C.I., Chung, Y.A., Glass, J.: Ssast: self-supervised audio spectrogram transformer. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 10699–10709 (2022)
7. Haq, S., Jackson, P.J.: Multimodal emotion recognition. In: Machine Audition: Principles, Algorithms and Systems, pp. 398–423. IGI Global (2011)
8. Hossain, M.S., Muhammad, G., Song, B., Hassan, M.M., Alelaiwi, A., Alamri, A.: Audio-visual emotion-aware cloud gaming framework. IEEE Trans. Circuits Syst. Video Technol. **25**(12), 2105–2118 (2015)
9. Huang, Z., Dong, M., Mao, Q., Zhan, Y.: Speech emotion recognition using CNN. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 801–804 (2014)
10. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
11. Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., Plumbley, M.D.: PANNs: large-scale pretrained audio neural networks for audio pattern recognition. IEEE/ACM Trans. Audio Speech Lang. Process. **28**, 2880–2894 (2020)
12. Koolagudi, S.G., Murthy, Y.S., Bhaskar, S.P.: Choice of a classifier, based on properties of a dataset: case study-speech emotion recognition. Int. J. Speech Technol. **21**(1), 167–183 (2018)
13. Kunze, J., Kirsch, L., Kurenkov, I., Krug, A., Johannsmeier, J., Stober, S.: Transfer learning for speech recognition on a budget. arXiv preprint arXiv:1706.00290 (2017)
14. Lee, M.C., Chiang, S.Y., Yeh, S.C., Wen, T.F.: Study on emotion recognition and companion chatbot using deep neural network. Multimedia Tools Appl. **79**, 19629–19657 (2020)
15. Li, P., Song, Y., McLoughlin, I.V., Guo, W., Dai, L.R.: An attention pooling based representation learning method for speech emotion recognition (2018)

Table 1. Quantitative evaluation of the different strategies on speech emotion recognition. In bold, the best model.

Classification	train F1 score	val F1 score	train Acc	Val Acc
original frozen CNN10(baseline)	0.2760	0.3125	0.3658	0.3994
original frozen CNN14(baseline)	0.2528	0.2413	0.3644	0.3778
original CNN10	0.6612	0.4432	0.6831	0.4657
original CNN14	0.8822	0.4542	0.8901	**0.4774**
multihead CNN10	0.6297	0.4536	0.6576	0.4684
multihead CNN14	0.9144	0.4642	0.9208	0.4674
multilayer CNN10	**0.7694**	**0.4636**	**0.7871**	0.447
multilayer CNN14	**0.9516**	**0.4652**	**0.9613**	**0.4722**

also observed a large performance gain for valence and a lesser gain for other aspects. The results suggest that while fine-tuning does incur additional computational costs, the benefits it yields in terms of improved performance make it a worthwhile endeavor. The validation set F1 scores for both CNN10 and CNN14 models, when employing the multilayer multi-head attention module, surpass those of the baseline, with the CNN14 model also demonstrating higher accuracy on the validation set. A comparison between different structures reveals that the multilayer multi-head attention modules generally outperform their single-layer counterparts. Specifically, the 'multilayer CNN14' model delivered the best results, achieving optimal performance with the least amount of epochs.

4.4 Future Work

Compared to the baseline, we believe there is ample room to improve the accuracy of the validation set. There are several areas for future improvements. First, we did not adjust the architecture of the pre-trained model, and the limited number of CNN layers may have hindered its ability to recognize emotions arising from emotional correlations in the data fully. Thus, further adjustments to the model architecture and hyperparameters are necessary for better generalization. In addition, we should further explore the linguistic and cultural differences in the datasets. Our target dataset is in Mandarin Chinese, while the baseline dataset is in English. Cross-language disparities may impede significant performance improvements.

Revealing these potential differences between languages and cultures requires further research in multi-task learning and exploring the fields of language and cultural studies. These areas offer significant potential for future research efforts, helping bridge the cross-linguistic gap and improving the performance of deep learning algorithms in specific tasks.

attention layer is passed into the next layer, it first goes through an additional transformation via a fully connected layer. The potential benefit of this could be to provide an additional means to capture and transform more complex patterns in the data.

4 Experiment

4.1 Dataset Setup

In our experiments, we utilized the CH-SIMS v2.0 [18] dataset, which is partitioned into three sets: the training set (80%), the test set (10%), and validation set (10%). The obtained output is categorized into five distinct labels. For feature extraction, the librosa library [22] is employed to extract log mel spectrograms from raw audio data.

4.2 Experimental Setting

In the training experiments, we leverage a pre-trained model on the AudioSet dataset to facilitate transfer learning on an existing dataset. During the fine-tuning phase, we employed both single and triple multi-head self-attentive layers, with the results being labeled as 'multihead' and 'multilayer' respectively. The training process was utilized the Adam optimizer [10] and cross-entropy loss with a batch size of 16.

Results from the two original models (CNN10 and CNN14), with frozen parameters, were served as the baseline for our benchmark. In the fine-tuning phase, the models with multi-head and multi-layer were trained for 200 epochs with an initial learning rate of 1e-4. Each experiment set were conducted ten times with the average results recorded. The best-performing model is selected and saved, conducting experiments on both test sets and validation sets. The recorded results are presented in Table 1.

4.3 Results and Discussion

Table 1 presents the results of experiments conducted on the CH-SIMS2.0 dataset, with the primary evaluation metrics being the F1 score and accuracy (Acc). Remarkably, the fine-tuned models consistently outperform their counterparts with frozen parameters. When compared to other models, our approach delivers highly competitive results. The findings indicate that fine-tuning of parameters significantly enhances the accuracy of audio classification. Therefore, we firmly advocate for implementing parameter fine-tuning as an effective strategy to elevate output performance.

Table 1 provides a summary of the performance exhibited by the various speech emotion recognition models that were tested. When considering the experiments with the frozen initial parameters as the baseline, improvements are observed across all tested results in comparison to the baseline. Notably, we

every convolutional layer, and then ReLU non-linearity is applied to facilitate better training convergence. For CNN10 and CNN14, the convolutional blocks are used in pairs before an average pooling layer is applied. Specifically, CNN10 is composed of 8 convolutional blocks (4 pairs), while CNN14 consists of 12 convolutional blocks (6 pairs). All networks include a penultimate fully connected layer to enhance representation capability, succeeded by a final fully connected layer with 527 units. A sigmoid activation function is applied at this stage to derive the probabilities of each class.

3.3 Multi-head Attention Block

To capture the semantic relevance embedded within the speech signal, the multi-head self-attention mechanism [30] is employed to focus on emotional information from various subspaces. In the multi-head attention mechanism, there are H parallel attention heads, and each of these attention heads calculates a set of attention weights:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{Q^\top \cdot K}{\sqrt{d_k}}\right) \cdot V^\top \tag{1}$$

where: Q, K, and V are the query, key, and value matrices for calculating the multi-attention mechanism. The Softmax function is commonly used to normalize the attention scores and ensure that they represent a valid probability distribution, where the sum of all attention weights is equal to 1.

We use the optimization algorithm Noam for learning rate tuning to achieve better model solutions. By computing similarities between Q and K, the mechanism assigns weights to each query position, determining the significance of the corresponding values.

$$lr = factor \cdot modelsize^{-0.5} \cdot \min(step^{-0.5}, step \cdot warmup^{-1.5}) \tag{2}$$

where: *factor* refers to the initial learning rate size, *model size* denotes the hidden layer dimension, *step* represents the number of optimization steps, and *warmup* denotes the value of the step when the learning rate reaches its maximum value.

Between each layer, we introduced a gate function incorporating the sigmoid function.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

x is the input variable. It converts the output of the model into a probability value between 0 and 1. This gated mechanism helps the model to dynamically adjust the importance of different layers and enables more flexible and adaptive information processing. Furthermore, the use of the sigmoid function ensures a smooth gating operation, avoiding abrupt changes and maintaining stability during the learning process.

After applying several layers of multi-head attention, we performed an fully connected layer (comprised of a linear layer and a ReLU activation function) following each gating mechanism. This means that before the output from each

3.1 CH-SIMS Dataset

To conduct Speech Emotion Recognition (SER) on Chinese speech, we utilized the CH-SIMS v2 training dataset [18]. This dataset comprises 60 original videos, resulting in 2,121 video segments. It offers a diverse range of character backgrounds, covering different age groups, and featuring high-quality recordings. Only Mandarin Chinese speech is included in this dataset.

Compared to version 1.0 [36], the video data in this updated version includes a broader range of scenarios, and the focus is on acoustic and visual features rather than text, encompassing a wider variety of emotional expressions. This aspect serves as a valuable inspiration for our research.

Each video segment in the CH-SIMS v2 dataset has undergone multimodal annotations, further categorized into five emotion categories:

$$\begin{aligned} \text{negative} &: \{-1.0, -0.8\}, \\ \text{weakly negative} &: \{-0.6, -0.4, -0.2\}, \\ \text{neutral} &: \{0.0\}, \\ \text{weakly positive} &: \{0.2, 0.4, 0.6\}, \\ \text{positive} &: \{0.8, 1.0\}. \end{aligned}$$

3.2 Pre-trained Block

Our approach aims to leverage a pre-trained speech recognition network to extract meaningful features from the samples of CH-SIMS. The CNN architectures utilized in our study are adapted from those presented in reference [11]. The PANNs framework houses a diverse of pre-trained models, encompassing various versions of CNN models. These models are trained on extensive audio datasets, empowering them with ability to capture intricate audio feature representations. This capability allows PANNs to efficiently capture and analyze patterns and recognizable features within audio data. We applied its subsample since, within PANNs [11], the CNN-14 model achieves the best performance, and also uses the pre-trained model corresponding here. Following the preprocessing phase, the vocal data is fed into the framework which then internally constructs a frequency-based representation of the recordings. Interestingly, in a related study [27], it was observed that CNN-10 model performs well with some smaller datasets. Consequently, in our experiments, we employed both CNN-10 and CNN-14 models for the feature extraction and embedding.

The audio data undergoes the following preprocessing steps: first, the audio is resampled to 32kHz. Then, a Short-Time Fourier Transform (STFT) is applied with a window size of 1024 frames and a hop size of 320 frames. This process is to obtain spectrograms from the standard time-domain waveforms. Subsequently, Mel filter banks are utilized to the obtained spectrograms. After this, a logarithm operation is performed to derive log Mel spectrograms.

Each of CNN architectures is composed of convolutional layers with a kernel size of 3×3 for CNN10 and CNN14. Batch normalization is applied after

In deep learning research, various studies have adopted transfer learning methodologies, using techniques like embedding extraction and fine-tuning of pre-existing models [13,21], instead of training models from scratch. Both PANNs [11] and DeepSpectrum [1] are highly influential modern libraries designed for audio-based tasks. Among them, PANNs introduce pre-trained audio neural networks for sound event detection. The ability to fix hyperparameters in PANNs provides flexibility to use it as a transfer learning module with pre-existing knowledge. Singh et al. [27] simplified the original PANNs model using a pruning algorithm to remove redundant parameters and reduce computational effort.

To reduce the computational cost, researchers often use pre-trained models with fixed parameters to extract features, and training output layers on the generated embeddings. However, fine-tuning certain layers has been found necessary for specific tasks to achieve excellent performance [11,17]. Earlier layers in convolutional neural networks (CNNs) generally have stronger generalization capabilities than subsequent layers [29], explaining why fine-tuning all layers is essential for achieving good performance. In our study, we aim to explore whether a similar operation is necessary for the model under consideration. We will conduct two experimental designs, freezing the parameters of pre-trained layers or fine-tuning all output layers, to compare the effects of these approaches empirically.

3 Methodology

In our proposed architecture, we have designed two key modules: the pre-trained block and the multi-head attention block. The pre-trained block is a convolutional neural network (CNN) model that we have encapsulated within the PANNs [11] framework. The system's overall structure and the interconnections between these modules are depicted in Fig. 1. In this section, we provide comprehensive explanations of the datasets utilized and the specific application strategies employed for each module.

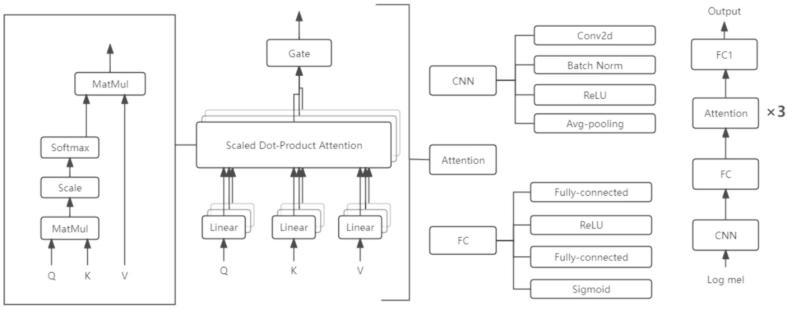

Fig. 1. The structure of proposed multi-head attention block. The number of attention layers will be adjusted to the specific task. 1 and 3 were applied in our experiments.

models, deep learning models do not require handcrafted features but directly learn feature representations from raw data, making them more powerful for processing complex and large-scale datasets and often exhibiting superior performance in specific tasks.

Xu et al. [35] proposed a framework for dual-modal (audio-text) emotion recognition. The framework consists of a parallel convolution module (Pconv) and an attention-based BLSTM [30], with a specific focus on single-modal processing of audio data from the CH-SIMS dataset. By combining Pconv and attention-based BLSTM, the Tensor Fusion Network effectively captures the complementary information from audio and text modalities, enabling more powerful multimodal sentiment analysis. The multiple self-attention mechanism is also a method of sentiment analysis that can enhance modal information [31]. In this paper, we apply transfer learning to a pre-trained CNN model with a multi-head attention mechanism and evaluate the performance of the system in terms of classification accuracy and training time.

2.2 Transformer

The Transformer model possesses several advantages, including its ability to effectively handle long sequences, capture long-range dependencies, and its parallel computing capabilities, making it highly suitable for processing large-scale data. Initially, the transformer model was mainly used in the field of machine translation, but because of its properties, it has gradually been generalized to the field of audio recognition.

In 2015, Chorowski [3] proposed to utilize an attention-based architecture, where the encoder side is a BiRNN structure. This was followed by a study on how transformers can replace RNNs for computation. The combination of CNN and attention mechanism is also a trend in audio emotion recognition, and the self-attention mechanism can express the salient regions of emotion in audio very well [16]. In 2021, Gong et al. Li et al. [15] proposed an Attention pooling method to avoid overfitting of convolutional features input to the fully connected layer. [5]introduced the Audio Spectrogram Transformer (AST), an audio classification model that canceled CNNs. Applying the Transformer encoder output to an audio spectrogram representation. They then proposed a semi-supervised framework [6] that improved the performance of AST by an average of 60.9%.

2.3 Transfer Learning

Transfer learning leveraging knowledge and models learned from one task to improve performance on another related task, reducing the need for extensive training data. It can effectively bypass the time-consuming task of data tagging when discrepancies exist in the feature space or data distribution [25], significantly increasing data mining efficiency. Transfer learning is crucial for multilingual or cross-lingual datasets due to the correlation between languages and speech, enabling the discovery of implicit connections parallelization [28].

collection of data covering Chinese text, images, audio data, and detailed annotations of modality.

In our study, we employed Pretrained Audio Neural Networks (PANNs) that were trained on the comprehensive AudioSet dataset. PANN is a deep learning model architecture crafted for audio data processing, built on the convolutional neural network (CNN) structure. Through fine-tuning on our unique dataset and integrating a multi-head self-attention mechanism, PANNs became more attuned to the specific features of the task and emotional nuances present in speech data, leading to enhanced emotion recognition performance. Our primary contributions include:

- We fine-tuned the pre-trained model on the AudioSet dataset and applied it to CH-SIMS for data preprocessing, yielding results with remarkable generalization capabilities.
- We introduced an architecture that merges CNN with a multi-head attention mechanism, enhancing the model's downstream performance.

2 Related Works

2.1 Speech Emotion Recognition

Over the past nearly three decades, researchers have tried to give machines the ability to understand and express emotions. Currently, the mainstream emotion recognition methods are extracting features that can accurately express emotions and detecting them, either manually or with the help of machines. This field encompasses a wide range of literature and utilizes various English datasets, such as RAVDESS [19], SAVEE [7], and IEMOCAP [2]. AudioSet [4] records a collection of 10-second sound clips including 632 audio event classes and over two million human-tagged clips drawn from YouTube videos. For Chinese language datasets, CH-SIMS [18] is notably prevalent, offering sentiment labels such as Strong Negative, Weak Negative, Neutral, Weak Positive, and Strong Positive. This study contributes to advancing multimodal sentiment analysis and capturing richer representations of sentiment within Chinese language data.

Emotion detection of sound relies on the integration of classical machine learning methods and deep learning techniques. Acoustic features, such as loudness, pitch, and timbre, are extracted and utilized in the algorithm to achieve accurate emotion detection. Spectral features, including Mel Frequency Cepstral Coefficients (MFCC) and their associated features, are also widely used [20]. The demarcation between machine learning and deep learning methodologies primarily resides in their respective approaches to data representations. In machine learning, a set of values is extracted from temporal, frequency, and perceptual domains and then fed into the machine learning algorithm as manually selected or predefined features to establish patterns and relationships for tasks like classification or regression. On the other hand, deep learning employs more complex and elusive algorithms, for example, CNN and attention mechanisms, to automatically learn intricate correlations within data. Unlike the traditional

speeches, such as happiness, sadness, and more. Emotion recognition systems leverage machine learning and deep learning techniques to extract relevant features from speech data, enabling accurate classification of emotions. High-performance SER systems hold significant value across various domains, including human-machine interaction [24], voice assistants [14], and psychological research [8]. They not only help computers better recognize the emotional states of inter-actor, but also pave the way for more personalized and effective human-computer interactions. Advancing SER is one of key objectives in emotion recognition system research. To improve accuracy, researchers employ techniques such as data augmentation and transfer learning, complemented by the use of larger and more diverse speech datasets. These strategies aid in training models proficient at accurately capturing and identifying emotional cues from speech data.

In the task of SER, the objective is to correlate input speech signals with specific emotion categories, thereby determining the underlying expressed emotions. Traditional classification techniques usually rely on probabilistic models, such as the Gaussian mixture model (GMM) [12], hidden Markov model (HMM) [23], and support vector machine (SVM) [26]. However, with the progression of research, various artificial neural network architectures have also been widely utilized, ranging from the simplest multilayer perceptron (MLP) [33], convolutional neural networks (CNNs) [9], to deep architectures like residual neural networks (ResNets) [32] and recurrent neural networks (RNNs) [17] [18]. Particularly, long short-term memory (LSTM)and gated recurrent units (GRU)-based neural networks, which are state-of-the-art solutions in time-sequence modeling, have been ubiquitously applied in speech signal modeling. Additionally, researchers have also proposed various end-to-end architectures aiming to jointly learn both feature extraction and classification [16]. These architectures intensively optimize the identification and association of emotions in speech signals, enhancing the overall performance of SER systems.

SER in Chinese involves identifying and analyzing emotions in Chinese speech data. Chinese-specific speech datasets are used to create diverse databases covering various emotional states. Techniques such as sound signal processing and feature extraction are employed to capture emotion-related features from speech. Machine learning algorithms, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), are used for emotion classification. Recent advancements like transfer learning and data augmentation have shown promising results. [34]

The attention mechanism imitates human attention, selectively focusing on different parts of input data and assigning varying levels of importance. Self-attention, used for sequential data, treats each input element as a query, key, and value. Multi-head self-attention extends this concept by introducing multiple attention heads, enabling the model to capture diverse feature representations and enhancing its expressive power.

Our research focuses on Speech Emotion Recognition (SER) in Chinese. We leveraged the CH-SIMS dataset for our study, which provides a comprehensive

Transfer Learning for Audio-Based Speech Emotion Recognition in Chinese: Leveraging Pretrained Models for Improved Performance

Lanke Zhu[1,2], Xinyue Ma[1,2], Rui Zhang[1,2], and Jianbo Zheng[1,2(✉)]

[1] Artificial Intelligence Research Institute, Shenzhen MSU -BIT University, Shenzhen 518172, Guangdong, China
jianbo.zheng@smbu.edu.cn

[2] Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Shenzhen MSU -BIT University, Shenzhen 518172, Guangdong, China

Abstract. In the field of Speech Emotion Recognition (SER) research, there is a growing emphasis on strengthening model generalization, stepping beyond the traditional classification accuracy metrics. Recent progress in cross-corpus SER has allowed machines to explore relationships among languages from diverse regions. In this paper, we propose an audio emotion recognition model which leverages a pretrained CNN model with a multi-head attention block. To adapt the model for the Chinese dataset CH-SIMS employed in our experiments, we fine-tuned it from a pre-trained English model. The data are categorized into five valence states: negative, weakly negative, neutral, weakly positive and positive. Remarkably, our top-performing model (multi-layer-CNN14) achieves a 24% improvement in accuracy over the baseline. The results highlight the effectiveness of fine-tuning in enhancing speech emotion recognition performance. This study contributes to improving model generalization in transfer learning, nudging us toward a deeper understanding and more accurate recognition of emotions expressed in speech.

Keywords: speech emotion recognition · transfer learning · fine-tuning · attention mechanism · Pretrained audio neural network

1 Introduction

Speech Emotion Recognition(SER) is a vital task in Natural Language Processing (NLP). It aims to detect and recognize the emotions conveyed through

L. Zhu, X. Ma, R. Zhang—These authors contributed equally to this work.

J. Zheng—This work was supported in part by the Shenzhen Sustainable Development Special Project under grant KCXFZ20201221173411032.

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 3–14, 2024.
https://doi.org/10.1007/978-3-031-65126-7_1

E-Health Networks I

Autonomous Vehicles