DAFTAR PUSTAKA

Alghamdi, N.S. *et al.* (2020) 'Predicting Depression Symptoms in an Arabic Psychological Forum', *IEEE Access*, 8, pp. 57317–57334. Available at: https://doi.org/10.1109/ACCESS.2020.2981834.

Blanco, V. *et al.* (2021) 'Symptoms of Depression, Anxiety, and Stress and Prevalence of Major Depression and Its Predictors in Female University Students', *International Journal of Environmental Research and Public Health*, 18(11). Available at: https://doi.org/10.3390/ijerph18115845.

Bucci, M. *et al.* (2016) 'Toxic Stress in Children and Adolescents', *Advances in Pediatrics*, 63(1), pp. 403–428. Available at: https://doi.org/10.1016/j.yapd.2016.04.002.

Can, Y.S. *et al.* (2020) 'Personal Stress-Level Clustering and Decision-Level Smoothing to Enhance the Performance of Ambulatory Stress Detection with Smartwatches', *IEEE Access*, 8, pp. 38146–38163. Available at: https://doi.org/10.1109/ACCESS.2020.2975351.

Chauhan, M., Vora, S. V. and Dabhi, D. (2017) 'Effective stress detection using physiological parameters', *Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017*, 2018-Janua, pp. 1–6. Available at: https://doi.org/10.1109/ICIIECS.2017.8275853.

Chen, Z., Luo, Y. and Mesgarani, N. (2017) 'Deep attractor network for single-microphone speaker separation', *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2017*, 2(1), pp. 246–250.

Choi, Y. *et al.* (2017) 'A biological signal-based stress monitoring framework for children using wearable devices', *Sensors (Switzerland)*, 17(9), pp. 1–16. Available at: https://doi.org/10.3390/s17091936.

Chyan, P. (2021) 'Automatic monitoring system for the elderly based on internet of things', *IOP Conference Series: Materials Science and Engineering* [Preprint]. Available at: https://doi.org/10.1088/1757-899x/1088/1/012041.

Chyan, P. *et al.* (2022) 'A Deep Learning Approach for Stress Detection Through Speech with Audio Feature Analysis', in *The 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE-2022)*. IEEE, pp. 269–273.

Défossez, A., Synnaeve, G. and Adi, Y. (2020) 'Real time speech enhancement in waveform domain', *Proceedings of the Annual Conference of the ernational Speech Communication Association, INTERSPEECH* eprint].

. and Ni Teoh, A. (2021) 'Real-time Stress Detection Model and Voice

Analysis: An Integrated VR-based Game for Training Public Speaking Skills', *IEEE Conference on Games*, pp. 1–4.

Douzas, G. *et al.* (2019) 'Imbalanced learning in land cover classification: Improving minority classes' prediction accuracy using the geometric SMOTE algorithm', *Remote Sensing*, 11(24). Available at: https://doi.org/10.3390/rs11243040.

Epel, E.S. *et al.* (2018) 'More than a feeling: A unified view of stress measurement for population science', *Frontiers in Neuroendocrinology*, 49(December 2017), pp. 146–169. Available at: https://doi.org/10.1016/j.yfrne.2018.03.001.

Garmezy, N., Masten, A.S. and Tellegen, A. (1984) 'The study of stress and competence in children: a building block for developmental psychopathology.', *Child development*, 55(1), pp. 97–111. Available at: https://doi.org/10.1111/j.1467-8624.1984.tb00276.x.

Gedam, S. and Paul, S. (2021) 'A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques', *IEEE Access*, 9, pp. 84045–84066. Available at: https://doi.org/10.1109/ACCESS.2021.3085502.

Gratch, J. *et al.* (2014) 'The distress analysis interview corpus of human and computer interviews', *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, pp. 3123–3128.

Han, H., Byun, K. and Kang, H.G. (2018) 'A deep learning-based stress detection algorithm with speech signal', *AVSU 2018 - Proceedings of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia, Co-located with MM 2018*, pp. 11–15. Available at: https://doi.org/10.1145/3264869.3264875.

Hannibal, K.E. and Bishop, M.D. (2014) 'Chronic Stress, Cortisol Dysfunction, and Pain: A Psychoneuroendocrine Rationale for Stress Management in Pain Rehabilitation', *Physical Therapy*, 94(12), pp. 1816–1825.

Haq, S. and Jackson, P.J.B. (2009) 'Speaker-Dependent Audio-Visual Emotion Recognition', in *Int'l Conf. on Auditory-Visual Speech Processing*, pp. 53–58.

Healey, J.A. and Picard, R.W. (2005) 'Detecting stress during real-world driving tasks using physiological sensors', *IEEE Transactions on Intelligent Transportation Systems* [Preprint]. Available at: https://doi.org/10.1109/TITS.2005.848368.

John R *et al.* (2016) 'DEEP CLUSTERING : DISCRIMINATIVE EMBEDDINGS FOR SEGMENTATION AND SEPARATION Mitsubishi Electric Research Laboratories (MERL), Cambridge , MA 02139 , USA',

ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pp. 31–35.

Hershey, John R. *et al.* (2016) 'Deep clustering: Discriminative embeddings for segmentation and separation', *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2016-May, pp. 31–35. Available at: https://doi.org/10.1109/ICASSP.2016.7471631.

Huang, Z. *et al.* (2022) 'Investigating Self-Supervised Learning for Speech Enhancement and Separation', *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6837–6841. Available at: https://doi.org/10.1109/icassp43922.2022.9746303.

Jason, C.A. and Kumar, S. (2020) 'An Appraisal on Speech and Emotion Recognition Technologies based on Machine Learning', *International Journal of Recent Technology and Engineering*, 8(5), pp. 2266–2276. Available at: https://doi.org/10.35940/ijrte.e5715.018520.

Kaczmarek, M. and Trambacz-Oleszak, S. (2021) 'School-related stressors and the intensity of perceived stress experienced by adolescents in Poland', *International Journal of Environmental Research and Public Health*, 18(22). Available at: https://doi.org/10.3390/ijerph182211791.

Karrouri, R. *et al.* (2021) 'Major depressive disorder: Validated treatments and future challenges', *World Journal of Clinical Cases*, 9(31), pp. 9350–9367. Available at: https://doi.org/10.12998/wjcc.v9.i31.9350.

Kejriwal, J., Beňuš, Š. and Trnka, M. (2022) 'Stress detection using non-semantic speech representation', in *2022 32nd International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 1–5. Available at: https://doi.org/10.1109/RADIOELEKTRONIKA54537.2022.9764916.

Kim, T.Y., Měsíček, L. and Kim, S.H. (2021) 'Modeling of Child Stress-State Identification Based on Biometric Information in Mobile Environment', *Mobile Information Systems*, 2021. Available at: https://doi.org/10.1155/2021/5531770.

König, A. *et al.* (2021) 'Measuring stress in health professionals over the phone using automatic speech analysis during the COVID-19 pandemic: Observational Pilot study', *Journal of Medical Internet Research*, 23(4), pp. 1–14. Available at: https://doi.org/10.2196/24191.

Kopin, I.J., Eisenhofer, G. and Goldstein, D. (1988) 'Sympathoadrenal medullary system and stress.', *Advances in experimental medicine and biology*, 245, 11–23. Available at: https://doi.org/10.1007/978-1-4899-2064-5_2.

, K. *et al.* (2019) 'Detecting moments of stress from measurements of arable physiological sensors', *Sensors (Switzerland)*, 19(17). Available

at: https://doi.org/10.3390/s19173805.

LeMoult, J. (2020) 'From Stress to Depression: Bringing Together Cognitive and Biological Science', *Current Directions in Psychological Science*, 29(6), pp. 592–598. Available at: https://doi.org/10.1177/0963721420964039.

Li, Q. *et al.* (2021) 'Research Progress in the Field of Image Completion', *Proceedings - 2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture, AIAM 2021*, pp. 398–402. Available at: https://doi.org/10.1109/AIAM54119.2021.00086.

Li, R. and Liu, Z. (2020) 'Stress detection using deep neural networks', *BMC Medical Informatics and Decision Making*, 20(11), pp. 1–11. Available at: https://doi.org/10.1186/s12911-020-01299-4.

Lu, H. *et al.* (2012) 'StressSense: Detecting stress in unconstrained acoustic environments using smartphones', *UbiComp'12 - Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 351–360. Available at: https://doi.org/10.1145/2370216.2370270.

Luo, Y., Chen, Z. and Yoshioka, T. (2020) 'Dual-Path RNN: Efficient Long Sequence Modeling for Time-Domain Single-Channel Speech Separation', *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2020-May, pp. 46–50. Available at: https://doi.org/10.1109/ICASSP40776.2020.9054266.

Luo, Y. and Mesgarani, N. (2018) 'TASNET : TIME-DOMAIN AUDIO SEPARATION NETWORK FOR REAL-TIME , SINGLE-CHANNEL SPEECH SEPARATION Yi Luo Nima Mesgarani Department of Electrical Engineering , Columbia University , New York , NY', *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 696–700.

Luo, Y. and Mesgarani, N. (2019) 'Conv-TasNet: Surpassing Ideal Time-Frequency Magnitude Masking for Speech Separation', *IEEE/ACM Transactions on Audio Speech and Language Processing*, 27(8), pp. 1256–1266. Available at: https://doi.org/10.1109/TASLP.2019.2915167.

Madhavi, I. *et al.* (2020) 'A Deep Learning Approach for Work Related Stress Detection from Audio Streams in Cyber Physical Environments', *IEEE Symposium on Emerging Technologies and Factory Automation, ETFA*, 2020-Septe, pp. 929–936. Available at: https://doi.org/10.1109/ETFA46521.2020.9212098.

Mather, M. and Lighthall, N.R. (2012) 'Both Risk and Reward are Processed Differently in Decisions Made Under Stress.', *Current directions in psychological science*, 21(2), pp. 36–41. Available at: https://doi.org/10.1177/0963721411429452.

Matsuo, N., Hayakawa, S. and Harada, S. (2015) 'Technology to detect levels of stress based on voice information', *Fujitsu Scientific and Technical Journal*, 51(4), pp. 48–54.

Mohler-kuo, M. *et al.* (2021) 'Stress and mental health among children/adolescents, their parents, and young adults during the first COVID-19 lockdown in Switzerland', *International Journal of Environmental Research and Public Health*, 18(9). Available at: https://doi.org/10.3390/ijerph18094668.

Morgado, P. and Cerqueira, J. (2018) 'The Impact of Stress on Cognition and Motivation', *Front. Behav. Neurosci.* [Preprint]. Available at: https://doi.org/10.1038/mp.2015.196.

Nachmani, E., Adi, Y. and Wolf, L. (2020) 'Voice separation with an unknown number of multiple speakers', *37th International Conference on Machine Learning, ICML 2020*, PartF16814, pp. 7121–7132.

Narvaez Linares, N.F. *et al.* (2020) 'A systematic review of the Trier Social Stress Test methodology: Issues in promoting study comparison and replicable research', *Neurobiology of Stress*, 13, p. 100235. Available at: https://doi.org/10.1016/j.ynstr.2020.100235.

Noe, S.M. *et al.* (2022) 'Automatic Detection and Tracking of Mounting Behavior in Cattle Using a Deep Learning-Based Instance Segmentation Model', *International Journal of Innovative Computing, Information and Control*, 18(1), pp. 211–220. Available at: https://doi.org/10.24507/ijicic.18.01.211.

Paulmann, S. *et al.* (2016) 'How psychological stress affects emotional prosody', *PLoS ONE*, 11(11), pp. 1–21. Available at: https://doi.org/10.1371/journal.pone.0165022.

Pisanski, K. and Sorokowski, P. (2021) 'Human Stress Detection: Cortisol Levels in Stressed Speakers Predict Voice-Based Judgments of Stress', *Perception*, 50(1), pp. 80–87. Available at: https://doi.org/10.1177/0301006620978378.

Punjabi, S.K. *et al.* (2019) 'Smart Intelligent System for Women and Child Security', in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018*. Available at: https://doi.org/10.1109/IEMCON.2018.8614929.

Van Puyvelde, M. *et al.* (2018) 'Voice stress analysis: A new framework for voice and effort in human performance', *Frontiers in Psychology*, 9(NOV), pp. 1–25. Available at: https://doi.org/10.3389/fpsyg.2018.01994.

min *et al.* (2018) 'Past review, current progress, and challenges ahead on cocktail party problem', *Frontiers of Information Technology and ctronic Engineering*, 19(1), pp. 40–63. Available at: ɔs://doi.org/10.1631/FITEE.1700814.

Radford, A. *et al.* (2023) 'Robust Speech Recognition via Large-Scale Weak Supervision', in A. Krause et al. (eds) *Proceedings of the 40th International Conference on Machine Learning*. PMLR (Proceedings of Machine Learning Research), pp. 28492–28518. Available at: https://proceedings.mlr.press/v202/radford23a.html.

Rafique, N. *et al.* (2019) 'Comparing levels of psychological stress and its inducing factors among medical students', *Journal of Taibah University Medical Sciences*, 14(6), pp. 488–494. Available at: https://doi.org/10.1016/j.jtumed.2019.11.002.

Rejaibi, E. *et al.* (2022) 'MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech', *Biomedical Signal Processing and Control*, 71, pp. 1–14. Available at: https://doi.org/10.1016/j.bspc.2021.103107.

Ren, Q., Li, Y. and Chen, D.G. (2021) 'Measurement invariance of the Kessler Psychological Distress Scale (K10) among children of Chinese rural-to-urban migrant workers', *Brain and Behavior*, 11(12), pp. 1–10. Available at: https://doi.org/10.1002/brb3.2417.

Rohmadi, M. *et al.* (2020) 'Case Study: Exploring Golden Age Students' Ability and Identifying Learning Activities in Kindergarten', *Proceedings of the First Brawijaya International Conference on Social and Political Sciences, BSPACE, 26-28 November, 2019, Malang, East Java, Indonesia* [Preprint]. Available at: https://doi.org/10.4108/eai.26-11-2019.2295218.

Saputri, M.S., Mahendra, R. and Adriani, M. (2019) 'Emotion Classification on Indonesian Twitter Dataset', *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*, pp. 90–95. Available at: https://doi.org/10.1109/IALP.2018.8629262.

Saranya, A., Venkatesh, C. and Kumar, S.S. (2016) *DESIGN AND IMPLEMENTATION OF AUTOMATIC CHILD MONITORING (ACM) SYSTEM USING WIRELESS NETWORK*, *International Journal of Computer Science and Mobile Computing*.

Shi, H.P., Cao, J.H. and Liu, X. (2011) 'Blind source separation for non-stationary signal based on time-frequency analysis', *Proceedings - 2011 4th International Conference on Intelligent Networks and Intelligent Systems, ICINIS 2011*, pp. 45–48. Available at: https://doi.org/10.1109/ICINIS.2011.12.

Slavich, G.M., Taylor, S. and Picard, R.W. (2019) 'Stress measurement using [illegible]ech: Recent advancements, validation issues, and ethical and privacy [illegible]siderations', *Stress*, 22(4), pp. 408–413. Available at: [illegible]s://doi.org/10.1080/10253890.2019.1584180.

[illegible]. *et al.* (2022) 'Age at onset of mental disorders worldwide: large-scale

meta-analysis of 192 epidemiological studies', *Molecular Psychiatry*, 27(1), pp. 281–295. Available at: https://doi.org/10.1038/s41380-021-01161-7.

Tomba, K. *et al.* (2018) 'Stress detection through speech analysis', in *ICETE 2018 - Proceedings of the 15th International Joint Conference on e-Business and Telecommunications*. Available at: https://doi.org/10.5220/0006855803940398.

Turcan, E. and McKeown, K. (2019) 'Dreaddit: A reddit dataset for stress analysis in social media', *LOUHI@EMNLP 2019 - 10th International Workshop on Health Text Mining and Information Analysis, Proceedings*, pp. 97–107. Available at: https://doi.org/10.18653/v1/d19-6213.

Vallejo, M.A. *et al.* (2018) 'Determining factors for stress perception assessed with the Perceived Stress Scale (PSS-4) in Spanish and other European samples', *Frontiers in Psychology*, 9(JAN). Available at: https://doi.org/10.3389/fpsyg.2018.00037.

Vandana, Marriwala, N. and Chaudhary, D. (2023) 'A hybrid model for depression detection using deep learning', *Measurement: Sensors*, 25(November 2022), p. 100587. Available at: https://doi.org/10.1016/j.measen.2022.100587.

Wang, D. and Chen, J. (2022) 'Supervised Speech Separation Based on Deep', *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–27. Available at: https://doi.org/10.1109/ICASSP43922.2022.9746303.

Wemm, S. and Wulfert, E. (2017) 'Effects of Acute Stress on Decision Making', *Physiology & behavior*, 176(3), pp. 139–148. Available at: https://doi.org/10.1007/s10484-016-9347-8.Effects.

Wijayakusuma, A. *et al.* (2021) 'Implementation of Real-Time Speech Separation Model Using Time-Domain Audio Separation Network (TasNet) and Dual-Path Recurrent Neural Network (DPRNN)', *Procedia Computer Science*, 179(2020), pp. 762–772. Available at: https://doi.org/10.1016/j.procs.2021.01.065.

World Health Organization (WHO) (2021) *Mental Health*.

Yaribeygi, H. *et al.* (2017) 'The impact of stress on body function: A review.', *EXCLI Journal*, 16, pp. 1057–1072.

Lampiran 1. Publikasi karya ilmiah dan korespondensinya

[6th ICITISEE 2022] Your paper #1570867307 ('A Deep Learning Approach for Stress Detection Through Speech with Audio Feature Analysis')

EDAS Conference Manager <help@edas.info>

on behalf of

6th ICITISEE 2022 (icitisee@amikom.ac.id) <icitisee=amikom.ac.id@edas.info>

Sat 12/3/2022 4:53 PM

To:Phie Chyan <phie_chyan@lecturer.uajm.ac.id>;Andani Achmad <andani@unhas.ac.id>;Ingrid Nurtanio <ingrid@unhas.ac.id>;Intan Sari Areni <intan@unhas.ac.id>

Dear Mr. Phie Chyan:

Congratulations - your paper #1570867307 ('A Deep Learning Approach for Stress Detection Through Speech with Audio Feature Analysis') for 6th ICITISEE 2022 has been **accepted** and will be presented in the session titled __.

The reviews are below or can be found at [../../showPaper.php?m=1570867307]1570867307.

Review 1

Relevance and timeliness: Rate the importance and timeliness of the topic addressed in the paper within its area of research.

Average (3)

Technical content and scientific rigour: Rate the technical content of the paper. (e.g. completeness of the analysis or simulation study, thoroughness of the treatise, accuracy of the models, etc), its soundness and scientific rigour.

Good (4)

Novelty and originality: Rate the novelty and originality of the ideas or results presented in the paper.

Average (3)

Quality of presentation: Rate the paper organization, the clearness of text and figures, the completeness and accuracy of references

Average (3)

ıdation: How do you rate your recommendation?

:cept. (2)

Detailed comments: Please justify your recommendation and suggest improvements in technical content or presentation.

-This research is very interesting and quite enough of novelty. -Figure 1 is too small, redraw or enlarge the figure to make accessible and readable.

Review 2

Relevance and timeliness: Rate the importance and timeliness of the topic addressed in the paper within its area of research.

Excellent (5)

Technical content and scientific rigour: Rate the technical content of the paper. (e.g. completeness of the analysis or simulation study, thoroughness of the treatise, accuracy of the models, etc), its soundness and scientific rigour.

Good (4)

Novelty and originality: Rate the novelty and originality of the ideas or results presented in the paper.

Average (3)

Quality of presentation: Rate the paper organization, the clearness of text and figures, the completeness and accuracy of references

Good (4)

Recommendation: How do you rate your recommendation?

Definite Accept. (4)

Detailed comments: Please justify your recommendation and suggest improvements in technical content or presentation.

- a good and clear introduction
- good literature review
- clear methodology and result research
- make clear the knowledge contribution
- how do you judge 97.1% the best accuracy in this topic area?

Review 3

Relevance and timeliness: Rate the importance and timeliness of the topic addressed in the paper within its area of research.

Technical content and scientific rigour: Rate the technical content of the paper. (e.g. completeness of the analysis or simulation study, thoroughness of the treatise, accuracy of the models, etc), its soundness and scientific rigour.

Good (4)

Novelty and originality: Rate the novelty and originality of the ideas or results presented in the paper.

Average (3)

Quality of presentation: Rate the paper organization, the clearness of text and figures, the completeness and accuracy of references

Good (4)

Recommendation: How do you rate your recommendation?

Accept. (3)

Detailed comments: Please justify your recommendation and suggest improvements in technical content or presentation.

it is better to add relevant references to be able to deepen the theory and results

Regards,

UNIVERSITAS AMIKOM YOGYAKARTA

ICITISEE 2022

CERTIFICATE

OF PARTICIPATION

THIS CERTIFICATE IS AWARDED TO

Phie Chyan

for the contribution as **Presenter**

In the 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE 2022)
"Applying Data Sciences and Artificial Intelligence Technologies for Enviromental Sustainability"

D. I. Yogyakarta (Indonesia)
December 13, 2022

ICITISEE

Prof. Dr. Kusrini, M.Kom

The General Chair of 6th ICITISEE 2022

7th ICITISEE 2023] Your paper #1570957029 ('Multi-Stage Approach for Stress Detection Using Speech Lexical Analysis')

icitisee=ami...@edas.info

Tue, Nov 7, 10:18 AM

to Phie, me, Ingrid, Intan

Dear Mr. Phie Chyan:
Congratulations - your paper #1570957029 ('Multi-Stage Approach for Stress Detection Using Speech Lexical Analysis') for 7th ICITISEE 2023 has been **accepted** and will be presented in the session titled __.
The reviews are below or can be found at 1570957029.

Review 1

Relevance and timeliness: Rate the importance and timeliness of the topic addressed in the paper within its area of research.
Excellent (5)

Technical content and scientific rigour: Rate the technical content of the paper. (e.g. completeness of the analysis or simulation study, thoroughness of the treatise, accuracy of the models, etc), its soundness and scientific rigour.
Good (4)

Novelty and originality: Rate the novelty and originality of the ideas or results presented in the paper.
Good (4)

Quality of presentation: Rate the paper organization, the clearness of text and figures, the completeness and accuracy of references
Excellent (5)

Recommendation: How do you rate your recommendation?
Definite Accept. (4)

Detailed comments: Please justify your recommendation and suggest improvements in technical content or presentation.
add to the literature review regarding the use of voice and NLP to measure stress levels. This was done to underline the novelty of the paper.

Review 2

Relevance and timeliness: Rate the importance and timeliness of the topic addressed in the paper within its area of research.
Good (4)

ontent and scientific rigour: Rate the technical content of the paper. (e.g. completeness ysis or simulation study, thoroughness of the treatise, accuracy of the models, etc), its and scientific rigour.
)

Novelty and originality: Rate the novelty and originality of the ideas or results presented in the paper.
Average (3)

Quality of presentation: Rate the paper organization, the clearness of text and figures, the completeness and accuracy of references
Average (3)

Recommendation: How do you rate your recommendation?
Accept. (3)

Detailed comments: Please justify your recommendation and suggest improvements in technical content or presentation.
Please describe the experimental results at the two steps in question. A. First Stage: B. Second stage

Review 3

Relevance and timeliness: Rate the importance and timeliness of the topic addressed in the paper within its area of research.
Average (3)

Technical content and scientific rigour: Rate the technical content of the paper. (e.g. completeness of the analysis or simulation study, thoroughness of the treatise, accuracy of the models, etc), its soundness and scientific rigour.
Average (3)

Novelty and originality: Rate the novelty and originality of the ideas or results presented in the paper.
Average (3)

Quality of presentation: Rate the paper organization, the clearness of text and figures, the completeness and accuracy of references
Average (3)

Recommendation: How do you rate your recommendation?
Accept. (3)

Detailed comments: Please justify your recommendation and suggest improvements in technical content or presentation.
Please provide an explanation for each image, graph and table as to why it is displayed in the paper
The resolution of some images needs to be improved to make them easier to understand
At the end of the discussion there is only a table, there is no explanation of the final results of the research.
Regards,
Rizki Wahyudi
TPC Chair of ICITISEE 2023

[JOIV] Editor Decision

Alde Alanda aldealanda@gmail.com via joiv.org

Sun, Sep 10, 9:04 PM

to me

Andani Achmad:

We have reached a decision regarding your submission to JOIV : International Journal on Informatics Visualization, "Hybrid Deep Learning Approach For Stress Detection Model Through Speech Signal".

Our decision is: Revisions Required

Alde Alanda
(Scopus ID: 57203718850); Politeknik Negeri Padang, Sumatera Barat
Phone 81267775707
Fax 81267775707
aldealanda@gmail.com

Alde Alanda

Reviewer A:

- Enhance the utilization of the English language in composing this manuscript. It is advisable to utilize a third-party entity in order to achieve more favorable outcomes.
- In the last part of the introduction, it is important to provide a comprehensive explanation of the merits inherent in conducting stress detection through your research, as well as a thorough analysis of the distinctions compared to prior studies.
- It is advisable to offer a more precise outline of stress in children, given that the manner in which stress manifests can vary throughout different developmental stages. It would be advantageous to incorporate a comparative analysis of the data with the insights of child psychology experts.
- The positioning of Table 1 and Figure 5 in your post should be adjusted as they currently overlap, impeding the reader's ability to effectively comprehend the content in a synchronized manner.
- The level of detail provided in the explanation of the model architecture should be enhanced. For instance, the precise procedures involved in each component of the model, together with the rationales underlying the use of CNN and GSOM as analytical approaches.
- The article discusses the precision and F1-score metrics of the model. In hance reader comprehension, it would be advantageous to furnish and elucidations of these measurements.
he conclusion

Andani Achmad <andani@unhas.ac.id>

Sat, Oct 7, 8:42 AM

to Alde

Dear Editorial Team of JOIV : International Journal on Informatics Visualization

We would like to thank the Editorial Team of JOIV for allowing us to submit a revised version of our manuscript titled "Hybrid Deep Learning Approach For Stress Detection Model Through Speech Signal" and for the time and effort spent reviewing our work. The reviewer's helpful and insightful comments allowed us to improve our paper substantially. Therefore, we carefully revised the manuscript to reflect our objectives better and address all reviewer's comments.

Here is a point-by-point response to the reviewers' comments and concerns.

1. Enhance the utilization of the English language in composing this manuscript. It is advisable to utilize a third-party entity in order to achieve more favorable outcomes.

   - We use third-party professional proofreading services to proofread our manuscript according to reviewers' advice.

2. In the last part of the introduction, it is important to provide a comprehensive explanation of the merits inherent in conducting stress detection through your research, as well as a thorough analysis of the distinctions compared to prior studies.

   - We have comprehensively explained the merits inherent in conducting stress detection through our research and analyzed the distinction compared to prior studies in the last paragraph of the introduction.

3. It is advisable to offer a more precise outline of stress in children, given that the manner in which stress manifests can vary throughout different developmental stages. It would be advantageous to incorporate a comparative analysis of the data with the insights of child psychology experts.

   - We have explained in greater detail how stress manifests in kids, particularly those in pre- and early-school age, based on the literature and the empirical experience of child psychologists, which we describe in the second paragraph of the introduction part and the first part of the materials and methods section.

4. The positioning of Table 1 and Figure 5 in your post should be adjusted as they currently overlap, impeding the reader's ability to effectively comprehend the content in a synchronized manner.

   - We have adjusted the positions of Table 1 and Figure 5 to make it easier for readers to comprehend the content in a synchronized manner.

The level of detail provided in the explanation of the model architecture should be enhanced. For instance, the precise procedures involved in each component of the model, together with the rationales underlying the use of CNN and GSOM as analytical approaches.

- We have added a detailed explanation of the model architecture, together with rationales underlying the use of CNN and GSOM as the analytical approach in section B. Architecture of proposed model in Material and Method Section.

6. The article discusses the precision and F1-score metrics of the model. In order to enhance reader comprehension, it would be advantageous to furnish definitions and elucidations of these measurements.

- As suggested by reviewers to enhance reader understanding, we have added a detailed explanation of the metrics used in model performance evaluation in the final paragraph of section II materials and methods.

7. Improve the conclusion.

- We have improved the conclusion section of our manuscript.

We have submitted our revised article at: https://joiv.org/index.php/joiv/author/submissionReview/2026 also we attached a copy of our revised manuscript with this email.

We hope that the improvements we have made are appropriate and we look forward to hearing from the editorial team regarding our submission.

Thanks to the editorial team of JOIV: International Journal on Informatics Visualization.

[JOIV] Editor Decision

Alde Alanda via joiv.org

Mon, Oct 30, 10:55 AM

to me

Andani Achmad:

We have reached a decision regarding your submission to JOIV : International Journal on Informatics Visualization, "Hybrid Deep Learning Approach For Stress Detection Model Through Speech Signal".

Our decision is to: Accept Submission

Starting July 2023. publication fees shall be implemented to all accepted papers. For more details, please email to joiv [at] pnp.ac.id. This journal charges the
following author fees (Article Publication Fee):

- Indonesian authors: 4.500.000 IDR per article (Regular)
- Indonesian authors: 6.500.000 IDR per article (Fast Track)

- International authors: 380 USD per article (Regular)
- International authors: 480 USD per article (Fast track)

This fee includes:

- DOI registration for each paper
- Checking the article similarity by turnitin
- English proofreading

Joiv

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

31 October 2023

Dear Andani Achmad

Hasanuddin University

Email: andani@unhas.ac.id

RE: JOURNAL ACCEPTANCE LETTER

We are happy to inform you that the International Journal on Informatics Visualization (JOIV) has been indexed in Scopus. The Scientific committee of JOIV agrees that the following manuscript is **accepted** for publication in JOIV

Title	Hybrid Deep Learning Approach For Stress Detection Model Through Speech Signal
Author	Phie Chyan, Andani Achmad, Ingrid Nurtanio, Intan Sari Areni

Thank you for your contribution the International Journal on Informatics Visualization (JOIV) and we look forward to receiving further submissions from you.

Sincerely

Regards,

Rahmat Hidayat

Editor in Chief
International Journal on Informatics Visualization
http:/joiv.org

Review Result: IJICIC-2304-002 (1)

office@ijicic.net

Tue, Jul 4, 1:30 PM

to me, chyanp21d, ingrid, intan

Dear Prof. Andani Achmad,

Reference No.: IJICIC-2304-002

Title: Stress Detection Through Speech Signal on Multi-Speaker Environment Using Deep Learning

Author(s): Andani Achmad, Phie Chyan, Ingrid Nurtanio, Intan Sari Areni

The paper above you submitted for possible publication in International Journal of Innovative Computing, Information and Control (IJICIC), has been reviewed by an Associate Editor and reviewers. Based on the Associate Editor's recommendation with which I concur (see the bottom of this email), I am sorry to inform you that your paper is not publishable in its current form. However, it may be publishable after extensive revision and rewriting. If you decide to do this, I would suggest that you carefully consider the comments of the Associate Editor/reviewers, and submit the revised version and response letter to IJICIC Submission System within three months from the date of this letter.

Thank you for your submission to IJICIC, and we are looking forward to receiving the revision, soon.

Best Regards,

Dr. Yan SHI

Executive Editor, IJICIC

Fellow, The Engineering Academy of Japan

Professor, School of Industrial and Welfare Engineering, Tokai University

9-1-1, Toroku, Kumamoto 862-8652, Japan

Tel.: 81-96-386-2666

E-mail: office@ijicic.net

Comments:

(1) This work holds significant research implications as it highlights the potential of deep learning techniques in advancing solutions for children's psychological stress.

vision of the Title is advised. The inclusion of the keyword "children" is
ıded to emphasize the paper's focus.

troduction should incorporate additional numerical evidence on children's
ical stress to elucidate the motivation of this study.

ntroduction, it is advisable to clearly articulate the limitations of the baseline

methods in detecting psychological stress among children, thereby emphasizing research gaps.
(5) In the last second paragraph of the Introduction, the contribution points by this work are suggested to be highlighted.
(6) It is recommended to interpret all symbols of Equations 1 and 2 in the modified version.
(7) Revisions to Figures 7 and 8, such as adjusting the axis range, are recommended in order to highlight discrepancies in modeling effects among various methods.
(8) It is suggested to add some discussions on the shortages of this work and future directions.
(9) More up-to-date studies are suggested to be cited, such as "Su Myat Noe, Thi Thi Zin, Pyke Tin and Ikuo Kobayashi, Automatic Detection and Tracking of Mounting Behavior in Cattle Using a Deep Learning-Based Instance Segmentation Model, International Journal of Innovative Computing, Information and Control, vol.18, no.1, pp.211-220, 2022".
(10) Some minor errors exist, such as the first occurrence of “RNN” in the article should be referred to by its full name.

Dear Associate Editor and Reviewer of IJICIC,
Thank you for giving us the opportunity to submit a revised version of our manuscript titled “Stress Detection Through Speech Signal on Multi-Speaker Environment Using Deep Learning”. Thank you for the time and effort spent reviewing our work. Your helpful and insightful comments allowed us to improve our paper substantially.
Thanks to your thorough reviews and comment, we were able better focus and ground our study and better frame and highlight our intended goals and takeaways.
Therefore, we carefully revised the manuscript to reflect our objectives better and address all reviewer’s comments. Starting from the title, the introduction, result, and conclusion sections have been made more explicit and clear.
Here is a point-by-point response to the reviewers’ comments and concerns.
1. “This work holds significant research implications as it highlights the potential of deep learning techniques in advancing solutions for children’s psychological stress”.
➢Thank you for your interest in our work. We are pleased to know that our work is considered to have significant research implications.

2. “The revision of the Title is advised. The inclusion of the keyword “children” is recommended to emphasize the paper’s focus.”
➢The focus of our paper is explicitly to detect potential psychological stress in children. Therefore we fully agree with your advice by revising the title of our paper to "Stress Detection of Children Through Speech Signals in Multi-Speaker Environment Using Deep Learning" to emphasize the focus of our paper.

3. “The Introduction should incorporate additional numerical evidence on children’s psychological stress to elucidate the motivation of this study.”
e included additional numerical evidence on the global prevalence of
vith stress-related health problems sourced from the World Health
ion in the second paragraph of the introduction to elucidate this study's
n.

4. “In the Introduction, it is advisable to clearly articulate the limitations of the baseline methods in detecting psychological stress among children, thereby emphasizing research gaps.”

➢We agree with this comment. Therefore in the fourth paragraph of the introduction section, we rewrite the section that discusses the limitations of the baseline methods in detecting psychological stress in children by clearly articulating what has been achieved by previous studies and what is still research problems that need to be resolved related to the detection of psychological stress using speech signal.

5. “In the last second paragraph of the Introduction, the contribution points by this work are suggested to be highlighted.”

➢We have revised the last second paragraph of the introduction section to highlight our contribution. This work proposes a model capable of detecting stress through speech signals in a multi-speaker environment. The results of our work can help parents and

caregivers of children better understand the stressful conditions that occur in children and the potential event or stressor that cause them so that the mitigation process can be carried out more effectively.

6. “It is recommended to interpret all symbols of Equations 1 and 2 in the modified version.”

➢We have interpreted all the symbols in Equations 1 and 2 in the revised manuscript.

7. “Revisions to Figures 7 and 8, such as adjusting the axis range, are recommended in order to highlight discrepancies in modeling effects among various methods.”

➢To emphasize differences in modeling effects among different methods, we changed the axis range of Figures 7 and 8 as suggested.

8. “It is suggested to add some discussions on the shortages of this work and future directions.”

➢In the last paragraph of the conclusion section, we have discussed the shortages and future directions of our work, as suggested by the reviewer.

9. “More up-to-date studies are suggested to be cited, such as "Su Myat Noe, Thi Thi Zin, Pyke Tin and Ikuo Kobayashi, Automatic Detection and Tracking of Mounting Behavior in Cattle Using a Deep Learning-Based Instance Segmentation Model, International Journal of Innovative Computing, Information and Control, vol.18, no.1, pp.211-220, 2022"

➢We have added some more up-to-date citations from related studies, including that suggested by the reviewers in reference number [20].

10. “Some minor errors exist, such as the first occurrence of “RNN” in the article should be referred to by its full name”

➢Thank you for pointing this out. We apologize for the minor mistakes we made. In the revised manuscript, we have fixed several minor errors, including those pointed out by reviewers, such as the use of the abbreviation RNN and CNN. We have replaced them according to their full names in their first occurrence. In addition to that, we also fixed some typos that we still found in the original manuscript.

We look forward to hearing from you regarding our submission and to responding to any further questions and comments you may have.
Best Regards,
Andani Achmad

Acceptance Letter: IJICIC-2304-002

office@ijicic.net

Sun, Jul 30, 8:15 PM

to me, chyanp21d, ingrid, intan

Dear Prof. Andani Achmad,

I am pleased to let you know that your paper submitted to IJICIC below

Reference No.: IJICIC-2304-002

Title: Stress Detection Through Speech Signal on Multi-Speaker Environment Using Deep Learning

Author(s): Andani Achmad, Phie Chyan, Ingrid Nurtanio, Intan Sari Areni

has been accepted for publication in International Journal of Innovative Computing, Information and Control (IJICIC). Please use the IJICIC style files (http://www.ijicic.net) to prepare the final version of your paper. Please note that author biography is also required.

Please send the final version (Word or TeX with PDF), the Copyright Form (http://www.ijicic.net) and the completed Invoice Letter on IJICIC online submission system (http://www.ijicic.net) within three weeks' time. All authors' handwritten signatures are required in Copyright Form.

Please feel free to contact me if you have any questions.

Sincerely yours,

Dr. Yan SHI

Executive Editor, IJICIC

Fellow, The Engineering Academy of Japan

Professor, School of Industrial and Welfare Engineering, Tokai University

9-1-1, Toroku, Kumamoto 862-8652, Japan

Tel.: 81-96-386-2666

E-mail: office@ijicic.net

A Deep Learning Approach for Stress Detection Through Speech with Audio Feature Analysis

Phie Chyan
Department of Electrical Engineering
Hasanuddin University
Gowa, South Sulawesi, Indonesia
chyanp21d@student.unhas.ac.id

Andani Achmad
Department of Electrical Engineering
Hasanuddin University
Gowa, South Sulawesi, Indonesia
Andani@unhas.ac.id

Ingrid Nurtanio
Department of Informatics
Hasanuddin University
Gowa, South Sulawesi, Indonesia
Ingrid@unhas.ac.id

Intan Sari Areni
Department of Electrical Engineering
Hasanuddin University
Gowa, South Sulawesi, Indonesia
intan@unhas.ac.id

2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) | 979-8-3503-9961-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICITISEE57756.2022.10057845

***Abstract*—Stress, a change in psychological reactions from a calm state to an emotional state, is a psychological problem that can negatively impact a person's physical and mental condition. Daily life that is full of pressures can be a stressor that triggers the stress. Various artificial intelligence-based technological approaches are currently used to detect stress through various indicators, one of which is using speech. In this study, a deep learning model based on CNN architecture was developed to detect stress through voice recording using various sound features extracted in the signal domain. The performance evaluation of the model was demonstrated using an open-source dataset (Crema-D and TESS), and the best accuracy value obtained was 97.1% in performing binary classification on stressed and unstressed labelled speech. The highest accuracy was obtained from experiments using various combinations of sound features in the signal domain using a combination of Mel Spectrogram and MFCC features. This evaluation result shows that the deep learning model with the appropriate sound feature extraction can accurately detect stress through voice recording.**

Keywords—Stress detection, deep learning, Audio feature, CNN architecture, Speech Signal

I. INTRODUCTION

According to the general definition, stress is characterized by a change in psychological reactions from calm to emotional. According to a psychology review, stress can be divided into eustress and distress. Eustress refers to a good emotional state, such as joy and excitement, while distress leads to negative emotions, but from these two terms, the term stress is more often used to describe a state of distress which symbolizes negative emotional states such as anger, anxiety, sadness, fear, pain, and nervousness [1]. Various existing research results describe a close correlation between stressful conditions with a decrease in the effectiveness of decision-making abilities [2], a decrease in cognitive performance and motivation to a decrease in awareness of the surrounding environment [3], [4].

The implementation of technology that is commonly used in detecting stress is through direct measurement of the body's electrical signal indicators (bio-signals) using various sensors e y. This method based on the bio-s l because it generally provides a b n other methods [5]. Although it h ccuracy, stress detection based on b es sensors or instruments that r e human's body. Recording data directly from the subject's body requires sensors or instruments to be worm or attached to the subject's body [6]. Although this method is generally harmless and not invasive, the use of various devices and sensors attached to the body can cause discomfort to the subject and be a source of stress in itself.

Voice (speech) from many studies on the psychology of stress concluded that it could be used as a marker of stress conditions experienced by humans because the sound output is a psychophysiological response that is part of the human integrative psychophysiological stress system. Stress reactivity is a complex integration of sympathetic and parasympathetic control in the brain. Psychological stress induces many effects on the body, including increased muscle tension, respiratory rate, and salivary levels, which can affect vocal production [7], [8]. Under psychological stress, voice pitch (acoustic correlation of fundamental frequencies) generally increases with increasing subglottal pressure and vocal intensity [9]. The increase in voice pitch and several other voice parameters, such as intonation and speech prosody, occurs under conditions of psychological stress, so voice features can be a promising biomarker in detecting stress levels [10]. The use of speech signals to detect stress has advantages and disadvantages. Speech signals can be easily acquired through a microphone without needing to be directly attached to the human body, so this method is more convenient for the subject. With this convenience, building an extensive database to train stress detection system models is possible. However, the extraction process and the selection of the sound features must be appropriately chosen to provide good accuracy results [11].

In this paper, we use a deep learning approach using the Convolutional Neural Network architecture to detect stress through a speech from audio recording. In addition, various low-level sound features will be extracted to be used in the classification process to see discriminant sound features in detecting stress through speech. The audio stream used as a dataset in this research comes from the open source datasets Crema-D and TESS, which will then be expanded with augmentation techniques to produce new synthetic data samples by adding small perturbations on the initial training set to produce a model with better generalization capabilities.

The rest of the paper is organized as follows. Section II presents related work related to technology and methods for detecting physiological stress. Section III describes the

methodology used in this study, including the stress detection model used. Section IV documents the experiments and the results obtained, and finally, section V hand out the conclusions of this paper.

II. Related Works

Research to develop stress detection models using artificial intelligence discussed in these related works is generally divided into two main focuses: stress detection using biosignal data and stress detection using speech.

A. Biosignal Based Stress Detection

Medically, stress conditions will affect the autonomic nervous system, a comprehensive regulatory system that controls various body functions such as heart rate, respiration, saliva and sweat. Stress conditions can generally be characterized by increased heart rate and blood pressure, more active sweat glands, and changes in metabolic activity, which can cause frequent urination, diarrhoea, nausea and other metabolic problems in some people [12]. Based on these conditions, stress can be detected by directly measuring various body response signals (biosignals).

Several related studies include detecting stress through heart rate (Heart Rate) and skin resistance (Galvanic Skin Response) [13], [14], detection and monitoring of stress through variable heart rate (Heart rate variable) and skin conductivity (Electrodermal Activity) through smartwatch devices [4], [15]. Others produced a stress detection method using heart rhythm data through an electrocardiogram (ECG) device combined with blood pressure measurements. In addition, respiration rate is also commonly used as an indicator in detecting stress [16].

These studies that use various instruments and sensors to detect the response signal of the human body when experiencing stress have advantages in terms of accuracy obtained compared to methods using non-biosignal data with an increasing tendency for accuracy when using a combination of several biosignal data. The combination of heart rate data, skin resistance and skin temperature are discriminant parameters that produce the highest accuracy in detecting stress based on several studies [5]. Generally, biosignal data is obtained through one or more wearable devices worn on the body. After that, the raw data is pre-processed using filters to remove noise and artefacts. Then the steps continued to the feature extraction and selection process. Some of the commonly used classifiers in this study are K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), and Random Forest (RF) with K-fold cross-validation and leave-one-subject-out implemented for model validation.

B. Stress Detection Using Speech Signal

Several related works related to stress detection through audio streams include stress detection models, which are integrated with virtual reality-based simulations to practice public speaking skills [17]. In that study, stress detection models based on voice analysis combined with VR s he experience as if he was giving a eople. Then another study related t stress of medical personnel on t utomated conversation analysis d emic [18]. This study has the a ss through voice on a telephone communication network. Furthermore, another study tried to analyze the effect of stress on cortisol hormone levels [9]. This study tried to see how far the impact of stress was measured through voice analysis regarded on the increase in the cortisol hormone.

The spectral and temporal features of the audio stream represent the short-time spectrum and temporal evolution of the audio signal, respectively. Spectral features are generated by transforming time-domain signals into frequency-domain signals to identify rhythm, pitch and melody information. Mel Frequency Cepstral Coefficient (MFCC) is a representation of the short-term power spectrum of a sound based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. The pitch of the human voice that reflects the fundamental frequency of the vocal cords is said to be the most studied acoustic feature for stress detection [19]. Temporal feature extraction is relatively straightforward with energy-based and zero crossing of audio signals, which can indicate a person's stress level.

Generally, there are two approaches in feature extraction methods: manual audio feature extraction using low-level features based on statistical methods and feeding raw audio input directly to the deep neural network for automatic feature learning [20].

III. Methodology

This study used audio streams from open-source datasets, namely the Crowd-Sourced Emotional Model Actor Dataset (Crema-D) and the Toronto Emotional Speech Set (TESS), which consisted of 480 and 2800 audio speech files, respectively. Each sound file reflects one type of emotion: anger, sadness, disgust, happiness, fear, pleasant surprise and neutral. After the data preparation, various low-level features are extracted from each audio file. The results are then trained into the proposed model, as illustrated in Fig. 1. The output of the classifier model based on the Convolutional Neural Network (CNN) is the prediction of stress status, whether the speech is predicted to be in a state of stressed or unstressed conditions. The open-source library Librosa was used for audio analysis in this study.

A. Data Preparation and Augmentation

The open-source datasets used in this study, Crema-D and TESS, contain sound files that reflect various types of emotions. According to the review of psychology discussed previously, stress is characterized by changes in psychological reactions. Stress or distress leads to negative emotional changes, such as anger, sadness, fear, and nervousness, while eustress leads to positive emotional changes, such as joy and excitement. Based on this theory, the relabeling process is based on the voice data in the dataset used. Audio files with angry, sad, disgusted, and fearful labels are relabeled as audio with a stressed label, while those with the labels happy, pleasant, surprised, and neutral are relabeled as audio with unstressed label. After the relabeling process, 1840 stressed and 1440 unstressed audio files were obtained. For audio analysis in this study, we included feature extraction and statistical modelling. However, the line between the two has blurred in recent years with the application of deep learning methods. Many open-source libraries are available for audio feature extraction, such as Librosa, Essentia, aubio, Madmom, and Marsyas. Different libraries can produce different numerical representations for the same feature (e.g., Mel

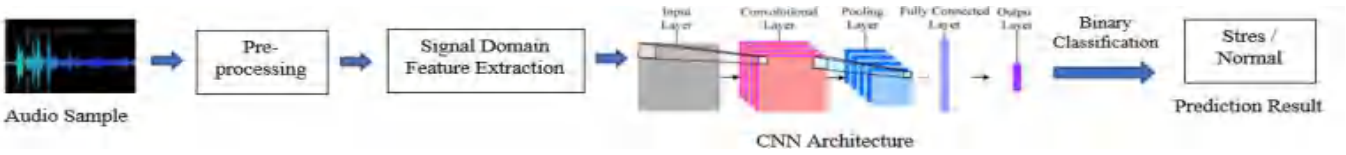


Fig. 1. Deep learning based stress detection model with CNN architecture

spectrum). For this study, Librosa was selected as the library used for audio feature extraction.

Data augmentation is carried out to create new synthetic data samples to enrich the dataset by adding small perturbations on the initial training set so that the resulting model has better generalization capabilities to various variations of audio files. In this study, each audio file from the dataset was made with three variations: the original version, the version with additional noise injection and the version with changes in pitch and shifting time so that the total file becomes 9840 files. Fig.2. shows the comparison of the audio file waveform against the various modification variants carried out in the augmentation process.

B. Feature Extraction

Feature extraction is essential in converting the audio signal into a format the model can understand. Various property descriptions can be obtained through the extracted audio features, which can then be fed into the model. The useful feature in the signal domain category will be used in stress detection needs because it consists of various essential features related to audio in general. Signal domain features consist of time domain, frequency domain and cepstral representation. The time domain is a feature extracted directly from the audio file waveform. The frequency domain is a feature that focuses on the frequency component of the audio signal. Generally, the signal is converted from the time domain to the frequency domain using a Fourier transform, and cepstral representation is a feature that combines the time and frequency components of the audio signal obtained by applying a short-time Fourier transform to the time domain waveform.

Based on the criteria discussed, the stress detection model uses each audio feature representing each representation of the signal domain feature as presented in Table I. That method aims to obtain optimal prediction accuracy in stress detection and to see which sound features are discriminant for predicting stress conditions.

TABLE I. AUDIO FEATURES USED IN SIGNAL DOMAIN

Signal Domain Feature	Audio Features Used
Time Domain	Amplitude Envelope (AE)
	Zero Crossing Rate (ZCR)
	Root Mean Square (RMS)
	Spectral Centroid (SC)
	Spectral Rolloff (SR)
	Spectral Bandwidth (SB)
	quency Cepstral Coefficient (MFCC)
	Mel Spectogram

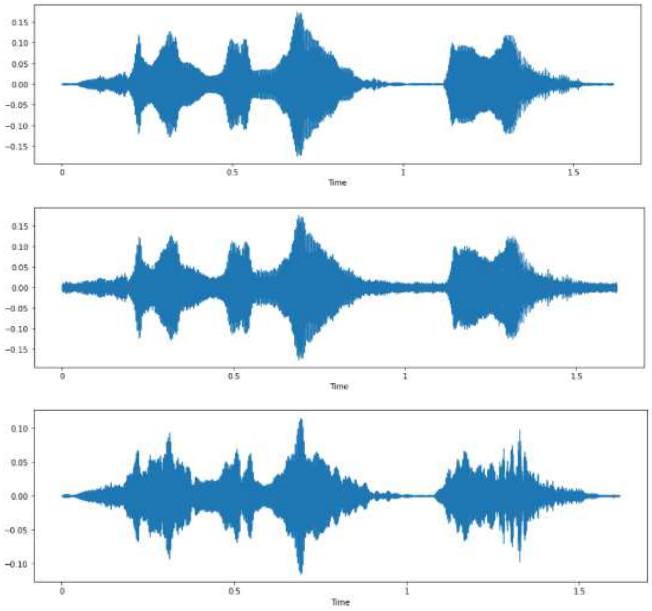


Fig.2. Waveform from original audio file (top), with noise injection (middle), and pitch modification (bottom)

IV. EXPERIMENTS AND RESULTS

Using an open source dataset that has been expanded with the augmentation process, the total number of audio files is 9840, consisting of 5520 files representing stressed status and 4320 files representing unstressed status. For the training processes, the dataset is divided by a ratio of 75:25 split on training: test randomly. The CNN-based architectural model used is a sequential model with the convolution 1D (Conv1D) type corresponding to the time series data. For the training and validation process, a softmax activation model is configured with a dense layer with two neurons. CNN is trained for 50 epochs with a batch size of 64 using the adam optimizer and will dynamically adjust the learning rate if the learning process stagnates.

Furthermore, to evaluate the model's performance, an accuracy measure commonly used in classification is carried out by determining the ratio value of how many outputs are classified correctly compared to the overall classification output produced. The states that can occur to the prediction results are True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) [21] , [22]. TP states that the prediction results and actual label show that the speech is stressed. FP states a state where the prediction results state that the speech is in a state of stress, but the actual speech label is in an unstressed condition. FN states a state where the prediction results state that the speech is in an unstressed state, but in fact, it is in a state of stressed condition, and finally, TN states a state where the predicted results and the actual label of the speech are unstressed. Recall, precision, and F1-Score values are also calculated to get a more detailed evaluation of the model. Recall describes how well the model calculates the

ratio of correctly labelled data to all audio data that is actually under stress. Precision calculates the ratio value of the number of correctly labelled data versus all audio data predicted to be in the stressed condition. F1-Score is a weighted comparison of the average precision and recall.

The results of the tests on CNN-based deep learning models on various types of audio features extracted for stress detection were carried out by measuring the performance of each audio feature individually and in combination for each feature category in the signal domain. Using deep learning and CNN models for the classifier, the best accuracy result obtained was 97.1% obtained with the cepstral domain feature using a combination of Mel spectrogram and MFCC. In contrast, the worst accuracy is obtained using the frequency domain feature. The combination of SC, SR and SB features obtained an accuracy of 57.6%. The detailed results of the test are presented in Fig.3.

Based on the evaluation test of the three feature categories in the signal domain, the cepstral domain shows significantly better accuracy than the time domain and frequency domain features in the model used for stress detection. That is because cepstral features such as Mel spectrogram and MFCC consist of composite feature sets that describe various information from the extracted sound characteristics. The MFCC used in this study consists of 20 parameter features that make up the formant (sound quality characteristic component) and timbre, while the Mel Spectrogram is a spectrogram whose frequency is converted to Mel Scale, which is a logarithmic transformation of the signal frequency. The Mel scale is very useful for machine learning because it directly represents the human perception of sound. Furthermore, to find out the most discriminant features to be used in the stress detection model, a single feature was used for each of the Mel spectrogram and MFCC, with the results presented in Fig.4.

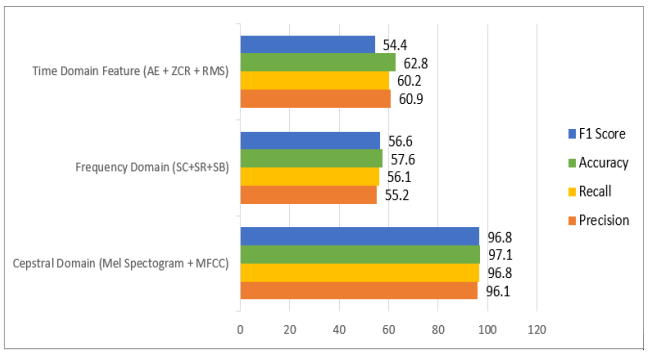
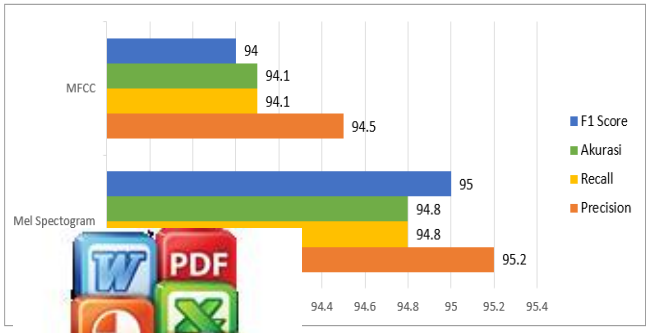


Fig.3. Stress detection model evaluation results



performance using MFCC and Mel
gram features

Based on the tests carried out on the stress detection model, the Mel Spectrogram feature has slightly better accuracy than MFCC. Hence, in Stress detection using a CNN-based deep learning model, the Mel spectrogram feature is the most discriminant audio feature in detecting stress conditions in audio speech.

V. Conclusion

Stress is a psychological condition characterized by changes in psychological reactions towards the emergence of various negative emotions. Voice is one of the excellent biomarkers to be used in detecting stress in the human subject. In this study, audio streams from open-source datasets were used. Various important audio features are extracted from each audio file for stress detection models using a deep learning approach with CNN architecture.

Based on the evaluation of the model's performance in performing binary classification (Stress vs unstressed), the best accuracy obtained was 97.1% using the cepstral feature in the signal domain, which is a combination of Mel spectrogram and MFCC. For future work, research can be continued by developing a more advanced model that can classify the stress level experienced by a person through his voice instead just a binary condition (stress or unstressed).

References

[1] Y. Choi, Y. M. Jeon, L. Wang, and K. Kim, "A biological signal-based stress monitoring framework for children using wearable devices," Sensors (Switzerland), vol. 17, no. 9, pp. 1–16, 2017.

[2] S. Wemm and E. Wulfert, "Effects of Acute Stress on Decision Making," Physiol. Behav., vol. 176, no. 3, pp. 139–148, 2017.

[3] P. Morgado and J. Cerqueira, "The Impact of Stress on Cognition and Motivation," Front. Behav. Neurosci., 2018.

[4] T. Y. Kim, L. Měsíček, and S. H. Kim, "Modeling of Child Stress-State Identification Based on Biometric Information in Mobile Environment," Mob. Inf. Syst., vol. 2021, 2021.

[5] S. Gedam and S. Paul, "A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques," IEEE Access, vol. 9, pp. 84045–84066, 2021.

[6] [H. Han, K. Byun, and H. G. Kang, "A deep learning-based stress detection algorithm with speech signal," AVSU 2018 - Proc. 2018 Work. Audio-v. Scene Underst. Immersive Multimedia, Co-located with MM 2018, pp. 11–15, 2018.

[7] [M. Van Puyvelde, X. Neyt, F. McGlone, and N. Pattyn, "Voice stress analysis: A new framework for voice and effort in human performance," Front. Psychol., vol. 9, no. NOV, pp. 1–25, 2018.

[8] G. M. Slavich, S. Taylor, and R. W. Picard, "Stress measurement using speech: Recent advancements, validation issues, and ethical and privacy considerations," Stress, vol. 22, no. 4, pp. 408–413, 2019.

[9] K. Pisanski and P. Sorokowski, "Human Stress Detection: Cortisol Levels in Stressed Speakers Predict Voice-Based Judgments of Stress," Perception, vol. 50, no. 1, pp. 80–87, 2021.

[10] N. Matsuo, S. Hayakawa, and S. Harada, "Technology to detect levels of stress based on voice information," Fujitsu Sci. Tech. J., vol. 51, no. 4, pp. 48–54, 2015.

[11] R. Li and Z. Liu, "Stress detection using deep neural networks," BMC Med. Inform. Decis. Mak., vol. 20, no. 11, pp. 1–11, 2020.

[12] K. Kyriakou et al., "Detecting moments of stress from measurements of wearable physiological sensors," Sensors (Switzerland), vol. 19, no. 17, 2019.

[13] A. De Santos Sierra, C. Sánchez Ávila, J. Guerra Casanova, and G. Bailador Del Pozo, "A stress-detection system based on physiological signals and fuzzy logic," IEEE Trans. Ind. Electron., vol. 58, no. 10, pp. 4857–4865, 2011.

[14] M. Chauhan, S. V. Vora, and D. Dabhi, “Effective stress detection using physiological parameters,” Proc. 2017 Int. Conf. Innov. Information, Embed. Commun. Syst. ICIIECS 2017

[15] Y. S. Can, N. Chalabianloo, D. Ekiz, J. Fernandez-Alvarez, G. Riva, and C. Ersoy, “Personal Stress-Level Clustering and Decision-Level Smoothing to Enhance the Performance of Ambulatory Stress Detection with Smartwatches,” IEEE Access, vol. 8, pp. 38146–38163, 2020.

[16] N. Attaran, A. Puranik, J. Brooks, and T. Mohsenin, “Embedded Low-Power Processor for Personalized Stress Detection,” IEEE Trans. Circuits Syst. II Express Briefs, vol. 65, no. 12, pp. 2032–2036, 2018.

[17] R. Dillon and A. Ni Teoh, “Real-time Stress Detection Model and Voice Analysis: An Integrated VR-based Game for Training Public Speaking Skills,” IEEE Conf. Games, pp. 1–4, 2021.

[18] A. König et al., “Measuring stress in health professionals over the phone using automatic speech analysis during the COVID-19 pandemic: Observational Pilot study,” J. Med. Internet Res., vol. 23, no. 4, pp. 1–14, 2021.

[19] H. Lu et al., “StressSense: Detecting stress in unconstrained acoustic environments using smartphones,” UbiComp’12 - Proc. 2012 ACM Conf. Ubiquitous Comput., pp. 351–360, 2012.

[20] C. A. Jason and S. Kumar, “An Appraisal on Speech and Emotion Recognition Technologies based on Machine Learning,” Int. J. Recent Technol. Eng., vol. 8, no. 5, pp. 2266–2276, 2020.

[21] I. Madhavi, S. Chamishka, R. Nawaratne, V. Nanayakkara, D. Alahakoon, and D. De Silva, “A Deep Learning Approach for Work Related Stress Detection from Audio Streams in Cyber Physical Environments,” IEEE Symp. Emerg. Technol. Fact. Autom. ETFA, vol. 2020-Septe, pp. 929–936, 2020.

[22] P. Chyan, “Design of intelligent camera-based security system with image enhancement support,” *J. Phys. Conf. Ser.*, vol. 1341, p. 42009, Oct. 2019.

2023 IEEE 7th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) | 979-8-3503-8226-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICITISEE58992.2023.10404529

Multi-Stage Approach for Stress Detection Using Speech Lexical Analysis

Phie Chyan
Department of Electrical Engineering
Hasanuddin University
Gowa, South Sulawesi, Indonesia
chyanp21d@student.unhas.ac.id

Andani Achmad
Department of Electrical Engineering
Hasanuddin University
Gowa, South Sulawesi, Indonesia
andani@unhas.ac.id

Ingrid Nurtanio
Department of Informatics
Hasanuddin University
Gowa, South Sulawesi, Indonesia
ingrid@unhas.ac.id

Intan Sari Areni
Department of Electrical Engineering
Hasanuddin University
Gowa, South Sulawesi, Indonesia
intan@unhas.ac.id

***Abstract*— Stress is one of the psychological problems that people most often experience. Stress that is not managed well and occurs for a prolonged time can lead to depression, negatively impacting a person's physical and mental health. Detecting and identifying stress is challenging, but it is necessary to provide appropriate early treatment for someone who experiences it. Many researchers have explored various technology-based approaches to detecting stress based on these needs. One approach currently being actively researched is stress detection technology using voice signals as an indicator to detect stress in a person. However, of these various studies, studies have yet to focus on exploring the use of voice properties optimally in terms of its properties as a signal wave and its output as a linguistic product. In this paper, we propose a multi-stage stress detection model through speech using a combination of speech signal characteristics and lexical aspects of speech. We employ the Mel-Spectrogram for the sound feature extraction required for stress detection from speech signals. Then, using Word2Vec and TF-IDF as input representations, we developed an NLP model for stress detection via lexical aspects of speech. The evaluation results of our proposed multi-stage stress detection model were able to detect stress well by showing an average accuracy and F1-Score of 91.7 and 91.2, respectively.**

Keywords*— *stress detection, speech signal, lexical analysis, mel-spectrogram, word2vec

I. Introduction

One of the psychological problems that people most often experience is stress, and stress that is not managed well and lasts for a prolonged time can cause depression [1]. This stressful event that leads to depression is a mental problem that cannot be taken lightly. Some symptoms related to depression are feelings of hopelessness, demotivation, withdrawal from social interactions, and mood swings, which occur repeatedly and cause the sufferer to experience psychological and physical damage depending on the severity of the problem [2], [3]. At a high level of severity, depression can affect brain activity and affect a person's behaviour, way of thinking and decision-making ability.

a tress, which leads to depression, p icted by WHO, has a fairly high p timated that around 280 million s m depression. The incidence of h l with depression, is also quite e ,000 cases reported worldwide uses of the high number of cases of depression is the lack of availability of early treatment services, which means sufferers do not receive adequate treatment [5]. Some of the challenges faced include the subject's lack of awareness of the psychological problems they are experiencing and also the difficulty for the subject's close relatives to recognize these psychological problems [6]. A person's stressful state is generally reflected in their emotional expression and can be detected from their facial expressions and voice. Negative emotions such as anger, sadness, nervousness and fear that mark stressful situations can be indicators of psychological problems experienced by a person [7].

The general instrument used to diagnose stress is interview-based psychological assessments [8], [9]. This conventional method is unsuitable because it requires much patient background data. Besides that, this method requires honesty from the patient, where patients who experience psychological problems are generally vulnerable and try to hide or deny their psychological problems. Apart from interview-based methods, other methods rely on technology-based approaches using biometric body measurements such as heart rate, blood pressure, skin resistance, and other parameters [7], [10], [11]. Regarding the precision attained, this method is excellent in identifying stress. However, this method is not practical, as it needs various tools and sensors to identify different biometric traits. Often, this equipment can only be used in clinics or hospitals. In addition to these issues, using this procedure, which necessitates the attachment of numerous devices to the body, may be uncomfortable for certain patients, particularly those with allergies or skin sensitivity issues [12], [13].

Voice is a biomarker that can be used to detect stress because the sound output is a response from the brain's sympathetic and parasympathetic nerve circuits. Stressful episodes experienced by a person can affect the control of these nerves, which also control various muscles in the body and will trigger effects such as muscle tension, respiratory control and saliva production, which will affect sound reproduction [14], [15]. Changes in the body's working mechanisms when a person experiences physical and mental stress will eventually be translated into changes in various parameters in the voice, such as volume, pitch and speech prosody [16]. Apart from the vocal output, someone who experiences stress tends to communicate using language that contains negative expressions following the negative emotions they are experiencing. For example, they use

sentences expressing sadness, anger, despair and various other negative expressions [17]. Based on the numerous parameters outlined, voice and lexical output can serve as effective biomarkers for stress identification. The ease of acquiring voice makes it possible to build a large dataset to support the stress detection model being built, which is one advantage of this method over baseline methods based on psychological assessments and biosignal parameters. Another advantage is that in terms of subject comfort, voice acquisition does not necessarily interfere with the subject's activities, preventing psychological effects like panic or nervousness before an examination.

In some recent research, the potential of voice as a promising biomarker in detecting stress has been explored [15], [18]. It is just that the use of sound features in these studies is more inclined towards extracting sound signal properties using artificial intelligence support to study signal patterns from the voice of someone experiencing stress, but the use of sound signal properties is generally not followed by analysis of lexical aspects as linguistic output. Whereas, as previously discussed, a person's state of stress can also be observed from what a person says because the words or sentences spoken reflect the feelings they are experiencing. Referring to the potential of voice, which is integrated with lexical analysis in stress detection modeling, in this research, we propose a multi-stage stress detection model through voice using signal characteristics and lexical aspects of speech.

The main contribution of this research is to propose a multi-stage stress detection model using signal features and lexical aspects of speech. The subject's voice will be captured by a microphone and sent into the model as input. The voice is recorded as a waveform, from which the signal domain feature will extract the sound features later. The results are then fed into a classification model based on a convolutional neural network that will temporarily predict the subject's stress status. Next, the model will analyze speech based on lexical features to identify potential stress in the subject's words or sentences. The final subject's stress status will be determined more precisely based on the findings of this proposed stress detection model.

In order to structure the presentation in this paper, the remainder of this paper is organized as follows: Section II discusses work related to this research topic. Section III explains the proposed methodology and model. Section IV documents the experiments and discussion based on the results obtained, and finally, Section V states the conclusions of this paper along with the continuation of future research directions.

II. Related Work

This section provides a broad explanation of the works relevant to this research's focus on numerous state-of-the-art studies on voice-based stress detection.

A. Stress Detection Methodology

n hological issue with long-term
s hysical and mental health. It has
d work on different methods of
b el. Nowadays, using interview-
d ient tools is the gold standard for
[el of psychological stress [15], ncluding the Patient Health Questionnaire and the Perceived Stress Scale (Morgan) (PHQ-8), are currently used. This interview-based examination is less efficient because it requires much time and money, and responders must meet specific cognitive aptitude criteria to take the test. Furthermore, this approach needs more validity and specificity [15].

Biological processes or biosignal-based measurement methods are another often utilized technique besides interview assessment-based techniques. This stress detection method offers advantages over interview-based assessment methods in preventing self-bias. In theory, this technique measures a variety of bodily reactions that stress-related stimuli cause. This biosignal-based detection method frequently uses heart rate, skin conductivity, blood pressure, and hormones like cortisol and adrenocorticotropics [20]–[22]. Depending on the type of parameter needed, various sensors or equipment are often employed and attached to the subject's body to get the necessary biosignal values. Currently, various more user-friendly devices are being developed, such as smartwatches that can monitor blood pressure and heart rate. However, in general, this biosignal measurement-based method is still relatively invasive, requiring blood or saliva samples from the subject [16].

The limitations of the previously discussed techniques increase the possibility of using alternative techniques for stress detection, such as stress detection through speech. Because they are inexpensive and non-intrusive, stress detection techniques based on speech are very promising. When speaking, a person prepares a string of words to be uttered in order to communicate with others and convey their message. Word choice, grammar, and word emphasis are just a few examples of how stressful situations can affect a person's use of language. As a result, language's lexical properties can be used as a stress indicator [23]. In addition to impacting language, stress also alters how the body functions, particularly the parts of the body that produce voice in humans, such as the vocal cords, muscles, and respiratory system [24]. Signal properties that detect various changes in a voice signal can be found by looking at the waveform. Typically, research that suggests a speechs-based stress detection model employs this technique by extracting features that can capture various alterations in the sound signal properties related to stressful situations.

B. Voice-Based Stress Detection Using a Machine Learning Approach

Deep learning is now widely used in many fields due to its ability to identify patterns in extensive data representations. This deep learning support has been applied in many research areas of image processing, computer vision, and signal processing [25], [26]. Research on voice-based stress detection also utilizes deep learning and machine learning. For example, research from [4] and [6] suggests a deep learning-based model that uses the CNN algorithm to detect subjects' stress status through voice. In extracting the features, the research utilized a mel-spectrogram and MFCC (Mel-frequency Cepstral Coefficients) extracted from the acquired sound waveform. These extracted features are then trained into a model to perform stress classification based on the characteristics of the subject's voice. Audio features in domain signals consist of time, frequency, and cepstral domains. Based on the audio feature analysis performed from those studies, it has been determined that the audio features of conversational sounds

are more representative when extracted in the cepstral domain using audio features like Mel-spectrogram and MFCC. As a result, this feature is typically used in studies relating to analyzing and recognizing speech voice.

The voice acquisition process for stress detection is typically carried out in a controlled environment where noise or ambient sounds can be ignored or minimized [27]. The voice of each subject was recorded independently. Because the environment in which people interact is typically one that is filled with a lot of sound and noise, this is done to lessen the pre-processing load that must be applied to the acquired sound. The drawback of a procedure like this is that it makes the subject aware of the procedure being used, which leaves room for information bias. In addition to this issue, it is possible that by the time a person's voice is acquired, their moment of stress may have passed. Many studies have attempted to address this by conducting sound acquisition procedures directly in noisy environments, such as research from [27] and [28], but this method requires a more complex pre-processing method to separate noise, background sounds, and speech sounds so that the sound that goes into the feature extraction stage is entirely pure speech sound. The accuracy results of this procedure-based model are generally lower regardless of how sophisticated the pre-processing method is compared to models where the sound acquisition procedure is carried out in a controlled environment. This is because the noise that enters the feature extraction stage can affect model accuracy.

III. Proposed Methodology

The proposed model is a multi-stage stress detection model that performs hierarchical stress detection. The first stage of the model performs direct voice signal analysis through the voice waveform using low-level features related to the properties of the voice signal, and the second stage of the model performs speech recognition, converting it into text form and then analyzing the linguistic aspects using lexical features to predict the emotional sentiment of the sentence. The model then uses the multi-stage analysis results to predict the subject's stress status.

Fig. 1 demonstrates how the proposed stress detection model works. A microphone records the subject's voice and represents it as a waveform. The waveform is processed incrementally following the preprocessing stage. First, mel-spectrogram is used to extract sound features from the signal domain. After feature extraction, the feature vector obtained from the results is fed to a CNN classifier that is trained to identify stress in voice record. The CNN classifier's output will indicate whether the subject is experiencing stress and use this information for the next stage. According to the classification from the first stage, if the subject's stress status indicates a non-stress state, the model will declare that status as its final prediction status. However, suppose the CNN classifier detects a stress state. In that case, the model will continue stress detection in a subsequent stage by performing speech recognition and translation using a pre-trained Automatic Speech Recognition (ASR) model which will carry out automatic transcription (speech-to-text). The obtained text will be extracted using the Word embedding feature extraction model and statistical weighting based on Word2Vec + TF-IDF against the open-source dataset Dreaddit, a multi-domain text collection database for stress identification. The output of this lexical analysis model will provide a prediction of the subject's stress status based on sentiment analysis of the sentences spoken. If at this stage the model predicts the subject is in a stressful condition then the final prediction of the model is "stress". However, if at this stage the subject is predicted to be in a non-stressed state based on the context of their speeches, then the final prediction of the subject's stress status is "possible stress". A detailed explanation of each stage will be discussed further in this section.

A. First Stage: Stress Detection Through Speech Signal

In order to detect the stress using the properties of speech signals, the model in this study was trained using the open-source datasets Surrey Audio-Visual Expressed Emotion (SAVEE) and the Crowd-Sourced Emotional Model Actor Dataset (Crema-D). This dataset comprises 2800 and 480 voice sample files from various subjects, each representing a different emotion: happiness, pleasant surprises, sadness, anger, disgust, fear, and neutral. Based on psychological theory, negative emotional characteristics such as anger, sadness, disgust, and fear are characteristic emotions often experienced by someone who experiences stress. In contrast, someone who does not experience stress tends to show neutral emotions or joy and happiness [7]. Because in this study, the model only states a person's emotional status in a stressed or non-stressed state, we adjusted by relabeling the dataset. Voice samples that represent negative emotions like anger, sadness, disgust and fear are relabeled with the label "Stress," Conversely, voice samples that were originally labeled as neutral, happy, and pleasant surprises were relabeled as "Non-Stress." After relabelling, we obtained 1840 voice sample files labeled "Stress" and 1440 voice sample files labeled "Non-Stress."

After the relabeling process, we then carry out the data augmentation. Data augmentation is applied to increase the dataset's diversity so that the model has better generalization capabilities over various small perturbations in sound samples. The data augmentation process creates variations of new synthetic data samples based on voice samples from the dataset by injecting artificial noise, modifying pitch, and

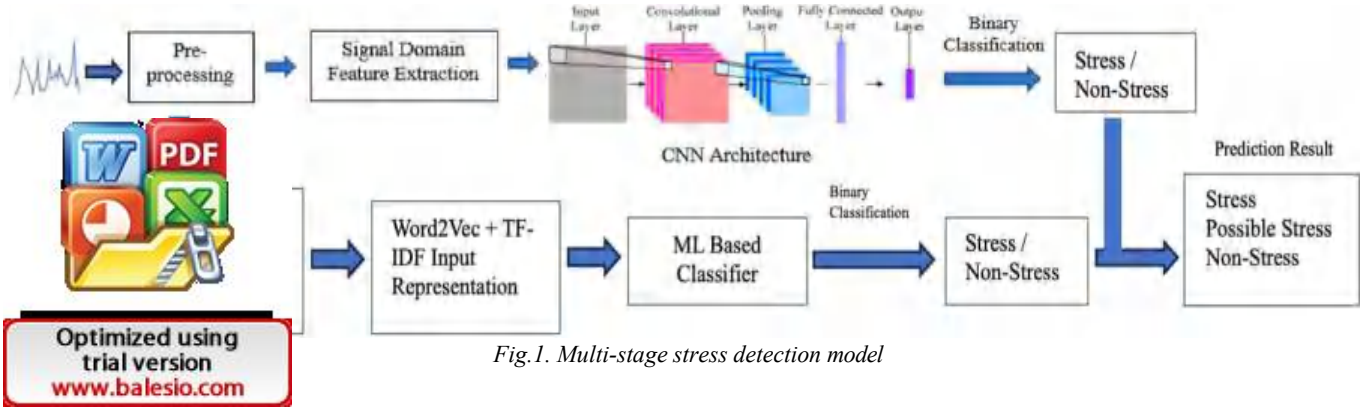


Fig.1. Multi-stage stress detection model

shifting to add a reverb effect to the original sound sample. This process simulates scenarios encountered in most real-world conditions where voice is acquired. Following the augmentation process, an audio feature extraction process is required so that the dataset can be used for model training. This process aims to translate different audio signal properties into a numerical representation that deep learning algorithms can understand. Mel-Spectrogram audio features in the cepstral domain were used in this study to extract audio features. Due to its ability to perceive sound in a manner similar to the human auditory system, this feature is frequently used in speech recognition and audio analysis applications.

B. Second stage: Stress Detection using Lexical Analysis

The voice recognition process is used in the second stage of the stress detection model to determine the final prediction of the subject's stress status by analyzing the speech's lexical features. Whisper-Medium, an OpenAI-developed pre-trained automatic speech recognition (ASR) model, is used for recognition. Whisper is intended to convert spoken language into text (Speech to Text) and supports multilingual capabilities thanks to the extensive data collection used to train the model. The data pipeline process, which includes data preprocessing, transformation, training, and deployment models, is carried out after the acquired voice has been converted into text. This process is crucial in working with Natural Language Processing (NLP) based models [29]. The completed data pipeline process is briefly shown in Fig. 2.

We used the publicly available open-source dataset Dreaddit, a multi-domain social media text data collection for stress identification, to identify stress conditions through text [30]. As shown in Table I, this dataset includes 3553 posts with binary labels (Stress and Non-Stress) from five general problem domain categories that frequently cause stress, including social problems, abuse, anxiety, PTSD, and financial problems. The average length of posts in the dataset is 420 tokens, with 3,553 labeled segments. We use TF-IDF and Word2Vec for input representation. Word2Vec with TF-IDF is a Word Embedding-based model with statistical weighting and dense vector representations of words, or subwords, which help capture semantic relationships between words.

TABLE I. STATISTIC OF DREADDIT DATASET

Domain	Total Post	*Avg. Token / Post*	*Labeled Segment*
Abuse	2,901	402	703
Anxiety	59,208	191	728
Financial	12,517	198	717
PSTD	4,910	265	711
Social	107,908	578	694
Total	**187,444**	**420**	**3,553**

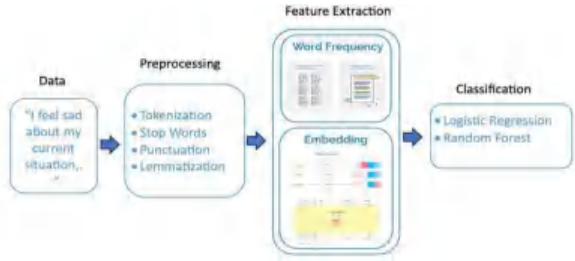


Fig.2.. NLP pipeline for stress detection model

IV. EXPERIMENT AND RESULT

As an experiment from the first stage we used Crema-D and SAVEE datasets to train the speech signal-based stress detection model. After going through a data augmentation process, the dataset increased to 9840 files, consisting of 5520 and 4320, each representing stress and non-stress. To train a CNN-based supervised model with the sequential Conv1D model used, we split the training data and train with a random 75:25 configuration. The CNN architecture consists of 3 convolution layers with a max pooling layer, a flattened layer and two dense layers. For feature extraction, we used the mel-spectrogram and got 131 signal features, which were then used as an input layer with shape (131,1) and produced output (131,256) through 256 channels using a 5x5 filter. The second and third convolutional layers consist of 256 and 128 channels with 5x5 filters, followed by a max-pooling layer. A dense layer with two neuron units with the SoftMax activation function is used in the training process with 50 epochs with a batch size of 4 using the ADAM optimizer. As shown in Fig. 3, the model's accuracy with training data was 0.96 with a loss of 0.05, while for test data it was 0.93 with a loss of 0.16.

In the second stage of the experiment, we trained lexical-based stress detection model with the dreaddit dataset, which contains 3553 data in total, and divided into train and test data with 2883 and 715 data, respectively,. The next step is performing NLP pipeline sequence including tokenization, removing punctuation, removing frequently used words (stopwords), and stemming or lemmatization. Then, Word2Vec is used to extract features with training parameters of vector size=300, window=10, min count=2, and number of workers=10 for 100 iterations. Word embedding, which is a weight vector from hidden layer neurons for each word in the corpus, is obtained after the model has been trained. This weight will establish the semantic relationship between words. Based on the similarity weights assigned to each word, Fig. 4 depicts their semantic relationship. Following word embedding, we weight the words using TF-IDF, and after that, we train the model using random forest and logistic regression classifiers to utilize the lexical features used. The accuracy results from the model training were 70.6% and 69.7% for each classifier. The metric results of each classifier are displayed in Table II.

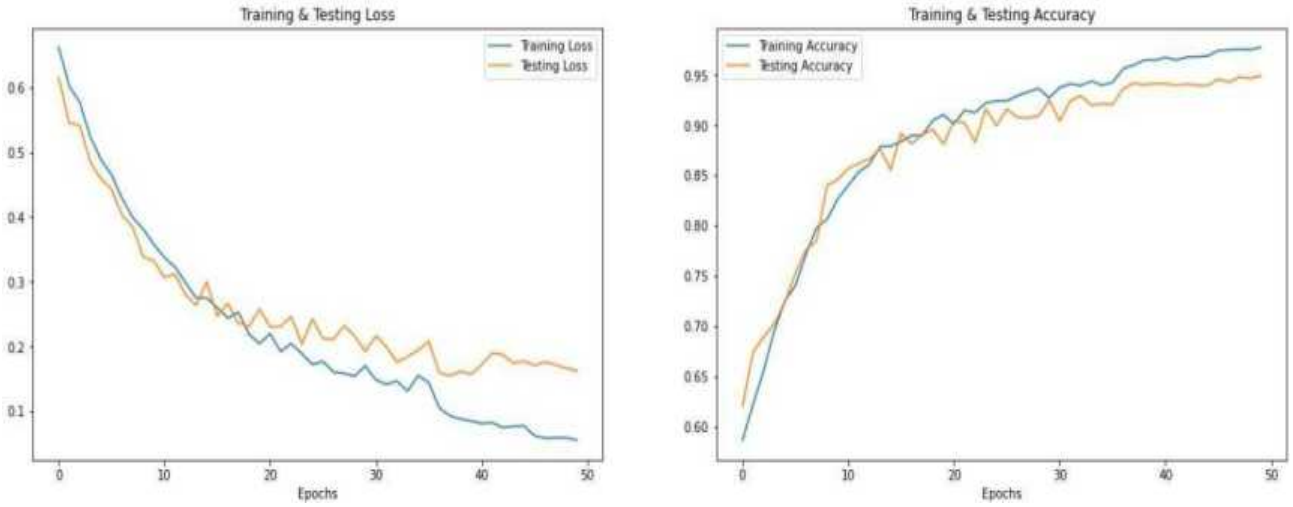


Fig.3. Graph represent model's loss value (left) and accuracy (right)

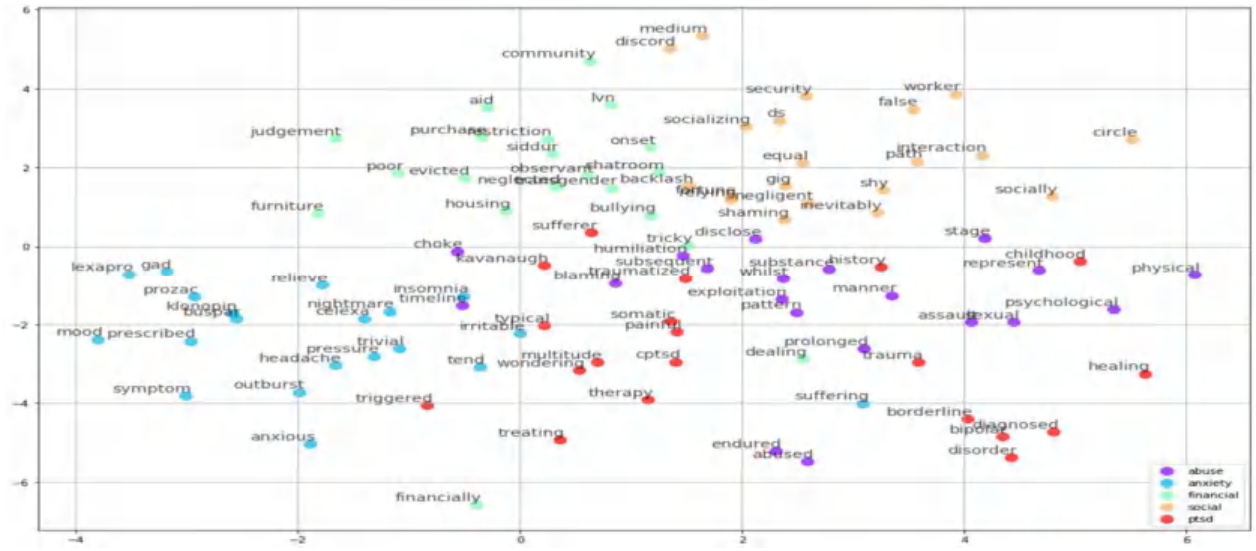


Fig.4. Similar word based on vector space on the dataset

Finally, to conduct a comprehensive test of the stress detection model, we use voice samples from the SAVEE Dataset; we extract speech signal properties to detect stress through speech and also carry out lexical stress detection by capturing the contextual meaning of speech in the voice samples. We used 100 voice samples, each with 50 voice labeled as stress and non-stress, to test our model. Table III shows several samples from the dataset and the prediction results from the model. The prediction results from the model are divided into three categories: Stress, Possible Stress and Non-Stress. In the model performance evaluation, the Stress and Possible Stress categories are counted as Stress. As a result, an average accuracy of 91.7% was obtained, with a precision value of 87.6%, Recall 94% and an average F1-Score of 91.2%. Currently, with an Intel Core i7-10750H based computer test bed with 16GB RAM, the model requires a total processing time of approximately 35 s[illegible] of 5 seconds. So, it is not yet p[illegible] cess in real-time.

Classifier	Log Reg (%)	Random Forest (%)
Precision	72.0	72.0
Recall	78.3	75.6
F1-Score	75.0	73.2

TABLE III. COMPARISON BETWEEN GROUND TRUTH AND MODEL PREDICTION

Speech Sample From Dataset	Ground Truth	Model prediction
"But the ships are very slow now and we don't get so many sailors anymore."	Stress	Stress
"Project development was proceeding too slowly."	Stress	Stress
"He would not carry a briefcase."	Non-Stress	Possible Stress
"She had your dark suit and greasy washwater all year."	Non-Stress	Non-Stress

[illegible] PERFORMANCE METRIC OF LEXICAL-[illegible] DETECTION MODEL

[illegible] Reg (%)	Random Forest (%)
70.6	69.7

V. CONCLUSION

In this paper, we propose a multi-stage stress detection model using a machine learning approach to detect stress through signal characteristics of speech output and combined

with lexical analysis of spoken sentences. Better generalization capabilities in detecting stress through speech as a signal output and as a linguistic product are provided by our stress detection model, which combines these two methods in a multi-stage manner. Based on the model's evaluation, it produced good results, with an average F1-Score accuracy of 91.7% and 91.2%. In future work, we will concentrate on enhancing the model's performance to reach nearly real-time processing levels, allowing it to detect and monitor the subject's stress condition and making it possible to create a model that is more capable of identifying the subject's stressful moments as well as potential stressors or stress state triggers.

ACKNOWLEDGMENT

We want to express our gratitude to the Ministry of Education, Culture, Research, and Technology for supporting the funding of this research through the Penelitian Disertasi Doktor (PDD) 2023 grant scheme.

REFERENCES

[1] J. LeMoult, "From Stress to Depression: Bringing Together Cognitive and Biological Science," *Curr. Dir. Psychol. Sci.*, vol. 29, no. 6, pp. 592–598, 2020, doi: 10.1177/0963721420964039.

[2] N. S. Alghamdi, H. A. Hosni Mahmoud, A. Abraham, S. A. Alanazi, and L. García-Hernández, "Predicting Depression Symptoms in an Arabic Psychological Forum," *IEEE Access*, vol. 8, pp. 57317–57334, 2020, doi: 10.1109/ACCESS.2020.2981834.

[3] V. Blanco, M. Salmerón, P. Otero, and F. L. Vázquez, "Symptoms of Depression, Anxiety, and Stress and Prevalence of Major Depression and Its Predictors in Female University Students," *Int. J. Environ. Res. Public Health*, vol. 18, no. 11, 2021, doi: 10.3390/ijerph18115845.

[4] Vandana, N. Marriwala, and D. Chaudhary, "A hybrid model for depression detection using deep learning," *Meas. Sensors*, vol. 25, no. November 2022, p. 100587, 2023, doi: 10.1016/j.measen.2022.100587.

[5] R. Karrouri, Z. Hammani, Y. Otheman, and R. Benjelloun, "Major depressive disorder: Validated treatments and future challenges," *World J. Clin. Cases*, vol. 9, no. 31, pp. 9350–9367, 2021, doi: 10.12998/wjcc.v9.i31.9350.

[6] P. Chyan, A. Andani, I. Nurtanio, and I. Areni, "A Deep Learning Approach for Stress Detection Through Speech with Audio Feature Analysis," in *The 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE-2022)*, IEEE, 2022, pp. 269–273.

[7] Y. Choi, Y. M. Jeon, L. Wang, and K. Kim, "A biological signal-based stress monitoring framework for children using wearable devices," *Sensors (Switzerland)*, vol. 17, no. 9, pp. 1–16, 2017, doi: 10.3390/s17091936.

[8] M. A. Vallejo, L. Vallejo-Slocker, E. G. Fernández-Abascal, and G. Mañanes, "Determining factors for stress perception assessed with the Perceived Stress Scale (PSS-4) in Spanish and other European samples," *Front. Psychol.*, vol. 9, no. JAN, 2018, doi: 10.3389/fpsyg.2018.00037.

[9] M. Kaczmarek and S. Trambacz-Oleszak, "School-related stressors and the intensity of perceived stress experienced by adolescents in Poland," *Int. J. Environ. Res. Public Health*, vol. 18, no. 22, 2021, doi: 10.3390/ijerph182211791.

[10] S. Gedam and S. Paul, "A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques," *IEEE Access*, vol. 9, pp. 84045–84066, 2021, doi: 10.1109/ACCESS.2021.3085502.

[11] T. Y. Kim, L. Mžíček, and S. H. Kim, "Modeling of Child Stress-
on Biometric Information in Mobile
Syst., vol. 2021, 2021, doi:

[
etecting stress in unconstrained acoustic
ones," *UbiComp'12 - Proc. 2012 ACM
ut.*, pp. 351–360, 2012, doi:

[13] Y. S. Can, N. Chalabianloo, D. Ekiz, J. Fernandez-Alvarez, G. Riva, and C. Ersoy, "Personal Stress-Level Clustering and Decision-Level Smoothing to Enhance the Performance of Ambulatory Stress Detection with Smartwatches," *IEEE Access*, vol. 8, pp. 38146–38163, 2020, doi: 10.1109/ACCESS.2020.2975351.

[14] M. Van Puyvelde, X. Neyt, F. McGlone, and N. Pattyn, "Voice stress analysis: A new framework for voice and effort in human performance," *Front. Psychol.*, vol. 9, no. NOV, pp. 1–25, 2018, doi: 10.3389/fpsyg.2018.01994.

[15] G. M. Slavich, S. Taylor, and R. W. Picard, "Stress measurement using speech: Recent advancements, validation issues, and ethical and privacy considerations," *Stress*, vol. 22, no. 4, pp. 408–413, 2019, doi: 10.1080/10253890.2019.1584180.

[16] K. Pisanski and P. Sorokowski, "Human Stress Detection: Cortisol Levels in Stressed Speakers Predict Voice-Based Judgments of Stress," *Perception*, vol. 50, no. 1, pp. 80–87, 2021, doi: 10.1177/0301006620978378.

[17] G. Rao, Y. Zhang, L. Zhang, Q. Cong, and Z. Feng, "MGL-CNN: A Hierarchical Posts Representations Model for Identifying Depressed Individuals in Online Forums," *IEEE Access*, vol. 8, pp. 32395–32403, 2020, doi: 10.1109/ACCESS.2020.2973737.

[18] C. A. Jason and S. Kumar, "An Appraisal on Speech and Emotion Recognition Technologies based on Machine Learning," *Int. J. Recent Technol. Eng.*, vol. 8, no. 5, pp. 2266–2276, 2020, doi: 10.35940/ijrte.e5715.018520.

[19] E. S. Epel *et al.*, "More than a feeling: A unified view of stress measurement for population science," *Front. Neuroendocrinol.*, vol. 49, no. December 2017, pp. 146–169, 2018, doi: 10.1016/j.yfrne.2018.03.001.

[20] A. De Santos Sierra, C. Sánchez Ávila, J. Guerra Casanova, and G. Bailador Del Pozo, "A stress-detection system based on physiological signals and fuzzy logic," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4857–4865, 2011, doi: 10.1109/TIE.2010.2103538.

[21] M. Chauhan, S. V. Vora, and D. Dabhi, "Effective stress detection using physiological parameters," *Proc. 2017 Int. Conf. Innov. Information, Embed. Commun. Syst. ICIIECS 2017*, vol. 2018-Janua, pp. 1–6, 2017, doi: 10.1109/ICIIECS.2017.8275853.

[22] N. Attaran, A. Puranik, J. Brooks, and T. Mohsenin, "Embedded Low-Power Processor for Personalized Stress Detection," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 65, no. 12, pp. 2032–2036, 2018, doi: 10.1109/TCSII.2018.2799821.

[23] M. S. Saputri, R. Mahendra, and M. Adriani, "Emotion Classification on Indonesian Twitter Dataset," *Proc. 2018 Int. Conf. Asian Lang. Process. IALP 2018*, pp. 90–95, 2019, doi: 10.1109/IALP.2018.8629262.

[24] S. Paulmann, D. Furnes, A. M. Bøkenes, and P. J. Cozzolino, "How psychological stress affects emotional prosody," *PLoS One*, vol. 11, no. 11, pp. 1–21, 2016, doi: 10.1371/journal.pone.0165022.

[25] P. Chyan, "Design of intelligent camera-based security system with image enhancement support," *J. Phys. Conf. Ser.*, vol. 1341, p. 42009, Oct. 2019, doi: 10.1088/1742-6596/1341/4/042009.

[26] P. Chyan, "Image Enhancement Based On Bee Colony Algorithm," *J. Eng. Appl. Sci.*, vol. 14, no. 1, pp. 43–49, 2019.

[27] I. Madhavi, S. Chamishka, R. Nawaratne, V. Nanayakkara, D. Alahakoon, and D. De Silva, "A Deep Learning Approach for Work Related Stress Detection from Audio Streams in Cyber Physical Environments," *IEEE Symp. Emerg. Technol. Fact. Autom. ETFA*, vol. 2020-Septe, pp. 929–936, 2020, doi: 10.1109/ETFA46521.2020.9212098.

[28] A. König *et al.*, "Measuring stress in health professionals over the phone using automatic speech analysis during the COVID-19 pandemic: Observational Pilot study," *J. Med. Internet Res.*, vol. 23, no. 4, pp. 1–14, 2021, doi: 10.2196/24191.

[29] C. Yang, X. Lai, Z. Hu, Y. Liu, and P. Shen, "Depression Tendency Screening Use Text Based Emotional Analysis Technique," *J. Phys. Conf. Ser.*, vol. 1237, no. 3, 2019, doi: 10.1088/1742-6596/1237/3/032035.

[30] E. Turcan and K. McKeown, "Dreaddit: A reddit dataset for stress analysis in social media," *LOUHI@EMNLP 2019 - 10th Int. Work. Heal. Text Min. Inf. Anal. Proc.*, pp. 97–107, 2019, doi: 10.18653/v1/d19-6213.

JOIV : Int. J. Inform. Visualization, 7(4) - December 2023 2474-2480

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Hybrid Deep Learning Approach for Stress Detection Model Through Speech Signal

Phie Chyan [a,c], Andani Achmad [a,*], Ingrid Nurtanio [b], Intan Sari Areni [a]

[a] Department of Electrical Engineering, Hasanuddin University, Bontomarannu, Gowa, 92171, Indonesia
[b] Department of Informatics, Hasanuddin University, Bontomarannu, Gowa, 92171, Indonesia
[c] Department of Informatics, Atma Jaya Makassar University, Tamalate, Makassar, 90224, Indonesia
Corresponding author: *andani@unhas.ac.id

***Abstract*— Stress is a psychological condition that requires proper treatment due to its potential long-term effects on health and cognitive faculties. This is particularly pertinent when considering pre- and early-school-age children, where stress can yield a range of adverse effects. Furthermore, detection in children requires a particular approach different from adults because of their physical and cognitive limitations. Traditional approaches, such as psychological assessments or the measurement of biosignal parameters prove ineffective in this context. Speech is also one of the approaches used to detect stress without causing discomfort to the subject and does not require prerequisites for a certain level of cognitive ability. Therefore, this study introduced a hybrid deep learning approach using supervised and unsupervised learning in a stress detection model. The model predicted the stress state of the subject and provided positional data point analysis in the form of a cluster map to obtain information on the degree using CNN and GSOM algorithms. The results showed an average accuracy and F1 score of 94.7% and 95%, using the children's voice dataset. To compare with the state-of-the-art, model were tested with the open-source DAIC Woz dataset and obtained average accuracy and F1 scores of 89% and 88%. The cluster map generated by GSOM further underscored the discerning capability in identifying stress and quantifying the degree experienced by the subjects, based on their speech patterns.**

***Keywords*— Stress detection; speech processing; deep learning; CNN; GSOM.**

Manuscript received 2 Aug. 2023; revised 10 Sep. 2023; accepted 30 Oct. 2023. Date of publication 31 Dec. 2023.

I. Introduction

Stress is a psychological problem that arises in response to the challenges experienced in daily life, affecting individuals of all ages [1], [2]. Stressors or stress triggers can stem from internal factors, such as feelings of inferiority and helplessness, or external factors, including bullying and academic pressure at school [3]. Different cross-disciplinary studies provide evidence of the complex relationship between mental development, social environment, and long-term health conditions [4]. The early years of a child's life often called the Golden Age period, represent a critical phase of development where their biological system rapidly as and negative experiences [5]. In ure to stress contributes closely to ns, including heart problems, di h [6]. Different expressions of ne ger, sadness, nervousness, and fe

Furthermore, this stressful condition also negatively influences the human nervous system. Several studies have shown that high-intensity chronic stress can lead to decreased brain mass, cognitive degradation, and memory problems. A growing child's physical and mental development may be adversely affected due to the occurrence of this circumstance [6], [7].

According to WHO data, the prevalence of mental illnesses in children is estimated to be 13% of the population aged 10 to 19 worldwide, which is roughly equal to disorders in adults at 20 % [8]. Therefore, mental health issues, such as stress, can persist into adulthood when the underlying issues are unaddressed [9]. Stress, considered non-threatening in its mild stages, is a prevalent mental condition encountered in everyday life. However, when not managed effectively, the condition can lead to severe mental health issues [7]. Detecting stress in children, particularly those in the pre-early school age group, presents a considerable challenge. This is primarily due to their limited communication skills and a lack

of awareness regarding the various potential mental health issues encountered [10], [11]. The impact of stress on pre-early school-age children can manifest in various ways, influenced by factors such as their personality, environmental context, and specific stressors [12]. Due to the restricted communication and cognitive abilities, children typically respond to stressful situations by exhibiting alterations in behavior and emotions. Common behavioral responses in children experiencing stress include heightened irritability, aggression, or withdrawal. Concurrently, typical emotional responses are nervousness, sadness, anger, and mood fluctuations.

Detection of Stress in children is a multifaceted challenge, necessitating the exploration of technology-driven solutions capable of autonomously identifying stressors in this vulnerable demographic. Based on current technological developments, various approaches can be used to detect stress, including direct measurements of the human body's biosignal parameters using various sensors. Several related studies include detecting stress through heart rate and skin resistance (Galvanic skin response), monitoring stress through variable heart rate (HRV), and skin conductivity (Electrodermal Activity) using smart watch devices [13], [14]. Other studies have produced stress detection approaches using heart rhythm data through an electrocardiogram (ECG) combined with blood pressure measurements [15] and studies using stress detection through a combination of heart rhythm data and respiration rate (Respiration Rate) [16]. Despite the efficacy of these studies in using biosignal measurements to achieve a high level of accuracy in stress detection, their applicability is limited due to the potential discomfort and the inadvertent introduction of additional stressors associated with affixing sensors to a child's body [17].

An alternative child-friendly approach to stress detection includes the analysis of an individual's speech. Based on medical literature reviews, there is a correlation between stress experienced and human vocal reproduction where the conditions affect various body functions and tension of various muscles for supporting vocal reproduction. Therefore, the output of human vocal sounds can be used as a good marker in detecting stressful conditions [18]–[20]. Studies related to stress detection through speech have been carried out in recent years [17], [21]–[28]. These studies have detected stress through speech with fairly good accuracy, between 75-85%. Furthermore, these studies primarily concentrate on binary classifications of stress status, distinguishing between individuals with and without stress [18]. The model that can provide high accuracy in identifying stressful states experienced by the subject with the level of severity is needed to achieve effective management. This is because the treatment of a person's stressful condition depends on the level of stress experienced. At mild levels of stress, individuals do not necessitate medical intervention but are often sufficient to address and manage the underlying stressors that provoke their distress. Conversely, in cases of se s, medical treatment becomes im udes the expertise of specialized m cluding child psychologists, pr eutic assistance [29].

del using a deep learning-based hy integrates supervised and unsupervised approaches. The primary objective is to enhance stress detection accuracy by analyzing speech signals and applying clustering approaches to identify associative relationships between voice characteristics and stress levels. The proposed model constitutes the central contribution of this study. Additionally, a dataset is established for stress detection in children, supporting the contributions in this domain.

According to the proposed detection model, the identification of stress in children can be conducted earlier, enabling the implementation of necessary preventive or therapeutic measures, contingent on the severity level. This proposed model stands apart from previous studies in two key aspects. Firstly, it is distinguished through explicitly constructing a supporting dataset for the stress detection model in children. Secondly, the model uses a novel hybrid approach, combining supervised and unsupervised learning elements to facilitate stress detection through speech analysis. This study is structured into four distinct sections, contributing to a comprehensive understanding of the result, and the introductory section provides the overview. The second section delves into the intricacies of the proposed approach, stating the details for a more comprehensive comprehension. Subsequently, the third section presents the results and engages in a thorough discussion. The fourth and concluding section encapsulates the entirety of the study content.

II. Material and Method

This section describes the dataset, the proposed model, and the performance evaluation approach of the stress detection system.

A. Dataset Preparation

In building the dataset, direct voice samples were obtained from 10 pre-early-school children aged between 5-7 years directly from the school environment, and the parents expressed their consent. The study was accompanied by a child psychologist who designed and supervised the activities based on the Trier Social Stress Test approach. Activities based on the Trier Social Stress Test include working on complex arithmetic questions and public speaking to induce stress [30]. After the activity session, the children were called into an interview session guided by a psychologist. During the session, their voices were recorded, and the psychologist monitored the children directly to observe the symptoms of stress through behavior or gestures. Based on the observations, stress status labeling was carried out on the voice sample recordings in binary, namely Stress or Non-Stressed.

In the case of children experiencing stress, the psychologist documented their levels based on the Kessler standard instrument. This instrument classifies the condition into three categories. The first is mild stress, where the subject exhibits subtle gestures or mild stressful behaviors. The second is moderate stress, characterized by symptoms and noticeable stressful behaviors. The third is severe stress, which indicates intense and pronounced symptoms and behaviors [29], [31]. The voice recordings from each subject were cut into sound samples with a duration of 1 and 2 seconds. After the

validation and labeling process, 106 and 142 voice samples were labeled as Stressed and Non-Stressed.

An open-source dataset known as The Distress Assessment Interview Corpus, abbreviated as DAIC-WOZ, was also used to evaluate the model developed against various state-of-the-art counterparts [32]. The dataset consisted of a voice sample and questionnaire answers from participating subjects labeled with a degree of stress level according to the standard Patient Health Questionnaire (PHQ-8). This questionnaire consisted of eight question items that measured various aspects of depression with a scale of 0 to 3, with response options of "not at all," "several days," "more than half days," and "nearly every day." From the eight questions, a score of 0-24 was obtained, which indicated the degree of stress. This dataset consisted of 59 and 130 samples for Stressed and Non-stressed subjects.

B. The Architecture of the Proposed Model

The proposed model consists of a combination of supervised and unsupervised learning approaches. The supervised approach uses the Convolution neural network (CNN) architecture. In the proposed model, CNN is based on its effectiveness in various tasks related to audio classification, such as speech recognition. This is because of its ability to understand various attribute representations in audio spectrograms, handle various input sizes, and use pre-trained model and transfer learning to provide good classification performance. CNN architecture consists of a Convolution layer, a max-pooling layer, and a fully connected layer [28].

Furthermore, the convolution layer provides the main framework for the entire neural network and a set of kernels to be learned in the training process. The max pooling layer functions to reduce the feature map's dimensions, reducing the data's size and the number of parameters to be studied. The linear layer connects each neuron from the previous to the next layer. Meanwhile, the ReLu activation functions to overcome the vanishing gradient problem, the flattened layer between the convolution and the fully connected layer, the dense layer using SoftMax as activation function, and the drop-out layer, reducing the overfitting problem for the unsupervised approach using GSOM (Growing Self-Organizing Map). GSOM is a type of unsupervised neural network used for dimensionality reduction and visualization of high-dimensional data to support models conducting audio data clustering. The model can categorize voice samples into clusters characterized by shared attributes using the support provided by GSOM, a variant of Self-Organizing Maps (SOMs) to elucidate the topological properties in data.

In addition, providing a visual representation of cluster data is needed for stress analysis. The approach consists of four nodes as the initial configuration and learning using rules based on Euclidean distance. New nodes are formed when quantitation errors accumulate ahead of the growth threshold value. Combining two approaches based on supervised and unsupervised using CNN and GSOM architectures is a hybrid used in developing a stress detection model. Figure 1 summarizes the stress detection model using this hybrid approach.

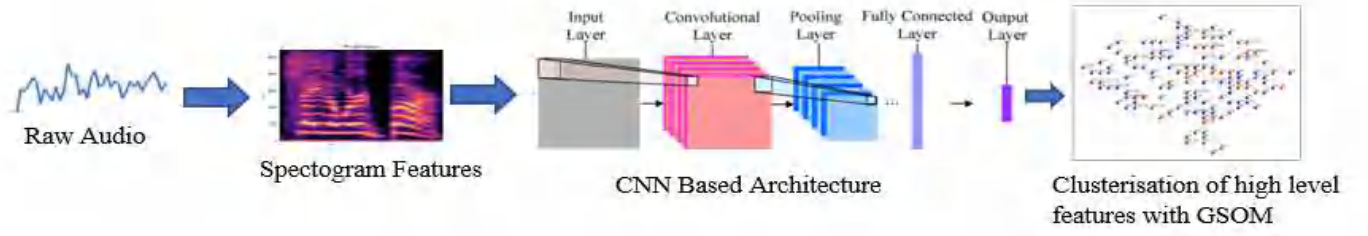


Fig. 1 Architecture of proposed model

The model receives input in the form of voice recordings from the dataset, and the sound samples are converted into a spectrogram, a visual representation of changes in the frequency of signals over time. Before conducting feature extraction, segmentation and data cleaning processes are carried out. This process is one of the critical stages since the sounds acquired are from subjects in a classroom environment susceptible to various noises. Segmentation is conducted to separate the sound signal from noise, including the detected silences on the recording. This study uses the Librosa library to perform segmentation, sound analysis, and feature extraction.

The data balancing procedure is undertaken to rectify the imbalance within the dataset. In the case of DAIC-WOZ, the number of voice samples categorized as non-stressed exceeds str an a twofold margin. Similarly, th bits an imbalance, where non-str greater than stressed voice. To ad a imbalance is mitigated by in sed labeled samples through the us proaches. Subsequently, feature extraction is performed on the audio spectrogram to convert the audio signal into a format comprehensible by model. After training, model can classify speech into binary labels, namely `stressed and non-stressed. The output from CNN in the form of high-dimensional features is fed to GSOM to conduct stress clustering by generating and visualizing feature maps to determine the characteristics of model in identifying sounds with stress categories.

C. Data Augmentation and Feature Extraction

The issue of imbalanced data warrants attention due to its potential to introduce bias into the training process, impacting the accuracy of the proposed model's output. In the context of this study, the children's voice dataset and the Daic-Woz dataset encountered challenges associated with data imbalance. The children's voice dataset contains 106 and 142 sound files labeled stressed and non-stressed. Similarly, in the open-source dataset DAIC-WOZ, there are 59 and 130 sound files labeled stressed and non-stressed, respectively. A way to overcome this imbalance problem is to use data augmentation approaches [33]. In addition to overcoming the problem, this

augmentation approach is also useful for increasing the diversity of datasets. In this study, data augmentation is carried out by creating a new synthetic audio file, which is a variation of the original audio file. To do this, we add two new samples of synthetic data for each original audio file. Using the Librosa audio library, we injected artificial noise as the first audio variation and performed pitch shifting as the second audio variation. So that the augmentation process is as bias-free as possible, we do several things to ensure this, including maintaining data balance.

In this case, we ensure that the proportion of each audio file representing both categories, namely stress and non-stress, is balanced so that learning is not biased towards one category. In addition, we ensure that each original audio has two synthetic variations produced with the same method so that the proportion of data after the augmentation process remains consistently maintained. Figure 2 shows the audio signal changes through the graphical waveform of the augmentation process. After the augmentation process on the children's voice dataset, 300 sound samples were obtained, consisting of 145 and 142 labeled stressed and non-stressed. For the open-source DAIC-WOZ dataset, after the augmentation process, a total of 300 samples were obtained, consisting of 142 and 158 sound samples labeled stressed and non-stressed.

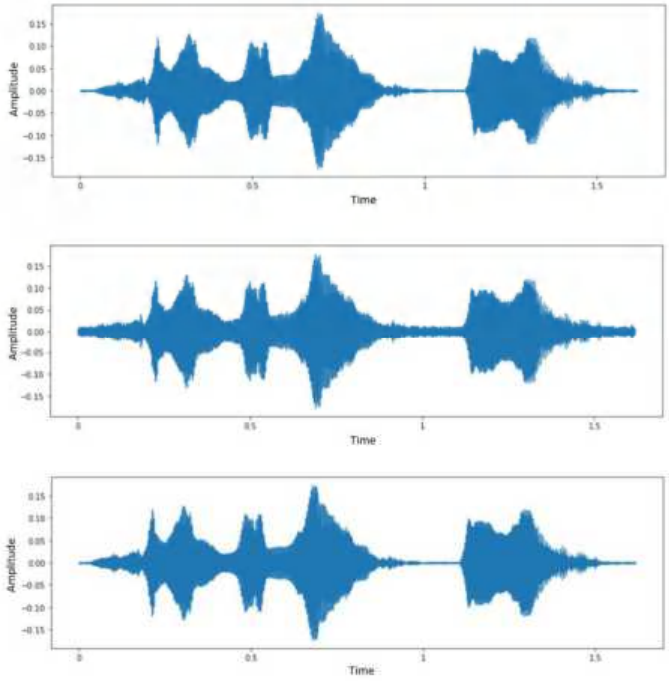


Fig. 2 Waveform of augmented speech sample, original sample (top), with artificial noise (mid), and pitch shifting (bottom)

Audio feature extraction converts the audio signal into a vector form the model understands. Various audio applications, such as audio classification, speech recognition, speech separation, and audio fingerprinting, require proper audio feature extraction to produce good performance. Based on [illegible] udio features can be divided into Lo[illegible] High-level. High-level audio fea[illegible] rhythms, are abstract features hu[illegible] omprehend or understand. The se[illegible] mid-level audio features, such as [illegible] MFCC. Low-level audio features ar[illegible] d directly from audio, and these features can only be understood by machines. The features include amplitude envelope, energy, spectral centroid, and zero crossing rate. In addition to the level of abstraction, audio features also depend on the signal domain, which states how the perspective of the signal is represented.

The signal domain is divided into time, frequency, and cepstral [34]. Audio features in the time domain are extracted directly from the waveform represented in time. The amplitude of the sound signal is measured as a function of time, and examples of audio features in the time domain are Zero crossing rate, amplitude envelope, and Root mean square energy. Audio features in the frequency domain are signal characteristics that describe the analysis of the mathematical function of signal to frequency. Signals are converted from the time domain using the Fourier transform, and some examples of audio features in the frequency domain are spectral centroids, band energy ratio, and spectral flux. The audio features in the cepstral domain are obtained by performing an inverse Fourier transform of the logarithm spectrum Fourier. Mel Frequency cepstral coefficients (MFCC) and Mel-spectrogram belong to this cepstral domain.

Furthermore, the Mel-spectrogram represents the cepstral domain for feature extraction in the stress detection model. The Mel-spectrogram has the advantage of imitating human auditory perception suitable for sound analysis. Figure 3 shows an example visualization of sound samples in the dataset represented in the mel-spectrogram.

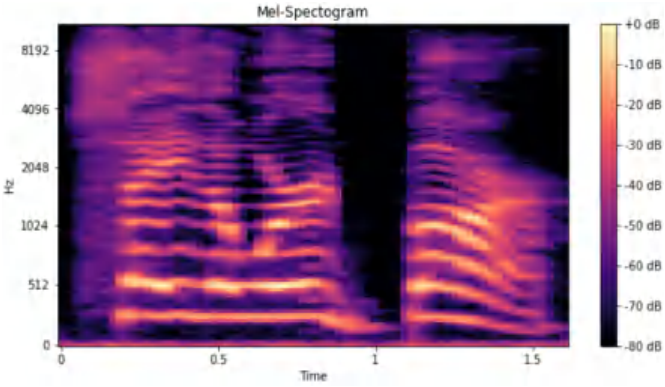


Fig. 3 Visualization of one of the voice samples in mel-spectrogram

D. Stress Detection Hybrid Model

Based on the architecture of the proposed model, the stress detection model uses a hybrid approach through supervised and unsupervised learning. CNN architecture in the proposed model consists of 3 convolution layers mediated by a max pooling layer to maintain the dominant features of the feature map, with a flattened layer followed by two dense layers. In the Mel-spectrogram, a total of 131 features are extracted, and these features are used as inputs into the initial layer, with dimensions (131,1). The output of this layer undergoes a transformation resulting in a (131, 256) shape, achieved through the use of 256 channels with 5x5 filters. Therefore, a dimension-reduction process is implemented using a max-pooling layer with a 5x5 filter and a stride of 2.

The second and third convolutional layers consist of 256 and 128 channels using 5x5 filters. These layers are followed by max-pooling layers, which maintain the same configuration as the initial max-pooling layer. Furthermore, a flattened layer converts the feature map to a linear form.

There is a dense layer with 32 neuron units, and the dropout layer is used with a value of 0.3. The SoftMax layer is configured with two units of neurons in the training phase for 50 epochs and a batch size of 50 using the ADAM optimizer. The training and test distribution is 75:25, with training data of 225 voice samples with 110 and 115 labeled stress and non-stressed. There are 75 voice samples for the test data, with 35 and 40 voice samples labeled Stress and Non-Stressed. In the training process, the ReduceLRonPlateau configuration is used to reduce the learning rate when the metric values begin to slope.

The output from CNN becomes input data to GSOM and is categorized into stress and non-stressed clusters. GSOM is initialized with four parallel nodes forming a quadrilateral with 131 dimensions of input data. The model is dynamically organized until the 131-dimensional mapping to 2-dimensional space is completed. Parameters for GSOM are set to 50 learning iterations, threshold 75, and spread factor from 0.1 to 0.9.

To evaluate model's performance, several metrics are used, namely accuracy, which measures the accuracy of predictions by calculating the ratio between correct predictions compared to those made by model. Furthermore, recall, precision, and F1-Score are also used to understand the effects of model on stress detection. The metric recall offers insight into model's ability to predict stress by examining the ratio of samples correctly predicted as stress to the total number of samples labeled as stress. Precision provides the ratio of correct stress predictions to the entire sample predicted as stress, and F1-Score indicates model performance by combining precision and recall values in a single value, balancing the trade-off between the recall and precision.

III. Result and Discussion

The system platform used is a Windows 10 64-bit computer with Intel Core I7 2.8 GHz CPU specifications, 16GB of RAM, and NVIDIA GTX 1060 4GB GPU, and model training process takes 4 min 25 s. The accuracy of the training model and validation with 50 epochs is 0.97 and 0.94 with a loss of 0.05 and 0.16, respectively. The graph of training accuracy, training loss, validation accuracy, and validation loss is shown in Figure 4. Based on the graph, models have a good learning rate with increasing model accuracy as the number of epochs increases before sloping at 40 epochs. The parameter loss values from the training and validation phases are also very small at 0.05 and 0.16, indicating an effective learning model.

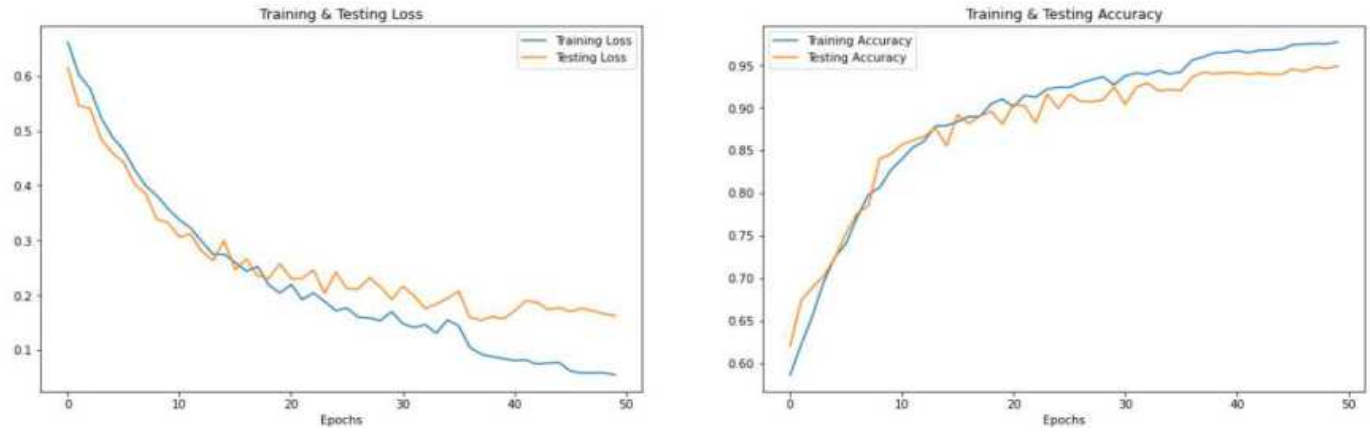
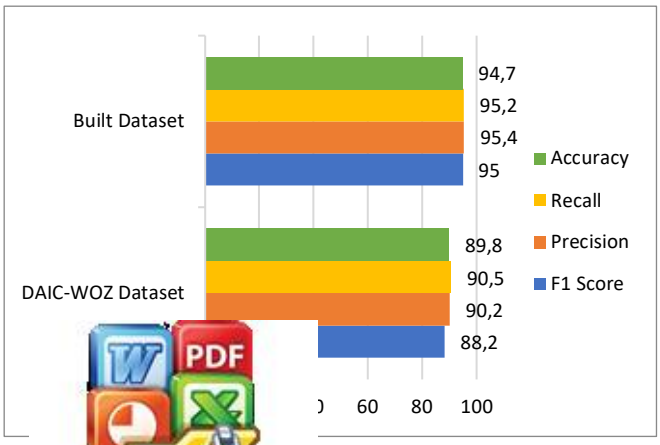


Fig. 4 Graph represents training and testing accuracy and loss vs number of epochs

Model performance results were measured using evaluation metrics for the constructed dataset and from the DAIC-WOZ dataset as presented in Figure 5.



built dataset vs DAIC WOZ dataset

The benchmarking process was carried out to assess the proposed model's performance compared to the state-of-the-art model. Several deep learning model using the DAIC-WOZ dataset were examined, including model discussed in references [26] and [27]. These models use CNN with feature extraction from image spectrograms. Additionally, model [28] was evaluated, which uses a combination of CNN and LSTM while adopting MFCC as the extracted feature. Based on the performance benchmark results in Table 1, the proposed model succeeded in exceeding the performance of the deep learning-based model.

TABLE I
PERFORMANCE BENCHMARK OF THE STATE-OF-THE-ART MODEL USING THE DAIC-WOZ DATASET

Model	Accuracy	F1-Score (N)	F1-Score (S)
RNN [26]	76 %	85%	45%
CNN [27]	-	70%	52%
CNN+LSTM [28]	76%	82%	64%
Our Model	89%	99%	78%

For the DAIC-WOZ dataset of 46 data, 21, 11, and 5 samples have a PHQ-8 score between 0 to 4, 5 to 9, and 10-14 showing non-stressed psychological, mild stress, and moderate symptoms, while the remaining 9 have a PHQ score above 15, indicating severe stress. From 21 sample subjects who showed non-stressed psychological conditions based on a PHQ-8 score <= 4, model correctly classified 19 samples as non-stressed conditions. Meanwhile, 14 samples with a PHQ-8 score >= 10 indicated a real state of stress, and model correctly classified 11 of the samples. These results show model has high sensitivity and specificity in detecting stress.

In the augmentation process for training the classification model using CNN algorithm, synthesis data was added to increase the voice dataset labeled with stress to overcome the problem of imbalanced data. In cluster analysis using GSOM, feature vectors from the original dataset were used without augmentation to avoid bias. The groups of these feature vectors were divided into stressed and non-stressed clusters. Each node in the cluster was labeled according to the data label in the dataset. Nodes from data marked as stressed and non-stressed were given red and blue labels before plotting onto the cluster map. High-density areas with nodes labeled stressed and non-stressed were then identified as stressed and non-stressed clusters.

Fig. 6 shows the analysis of the distribution of nodes on the cluster map. The area marked with a rectangular marker indicates the collection of voice nodes of the subject under stress conditions, and the surrounding area shows the density of subject nodes under stress conditions. Meanwhile, the subject nodes under non-stress conditions are concentrated from clusters with stress nodes. From the comparisons made according to the child's stress level, dense areas with a distribution of stress nodes originated from child subjects with moderate and severe levels. Meanwhile, stress subject nodes scattered in the minority into dense areas with non-stressed are mostly subjects with a mild level.

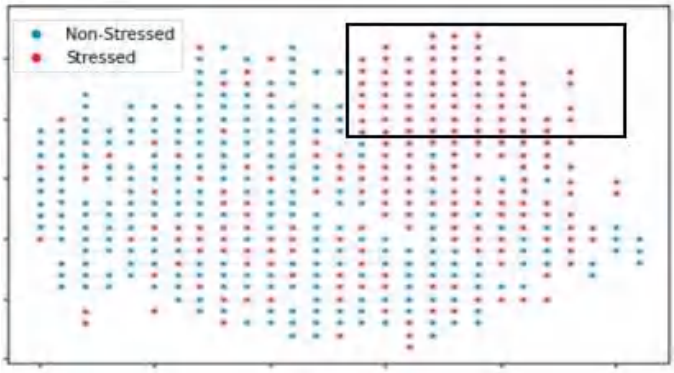


Fig. 6 Distribution node of built dataset

The cluster analysis of the DAIC-WOZ dataset also shows similar results (Fig.7). From the comparisons made according to the degree of stress, dense areas with a distribution of nodes came from subjects with a high PHQ-8 score level above 10, indicating a medium to a high degree. Furthermore, areas de ssed nodes were almost entirely fr)-8 score levels below 10. The pr ood differentiation ability in id ree of stress level of the subject th

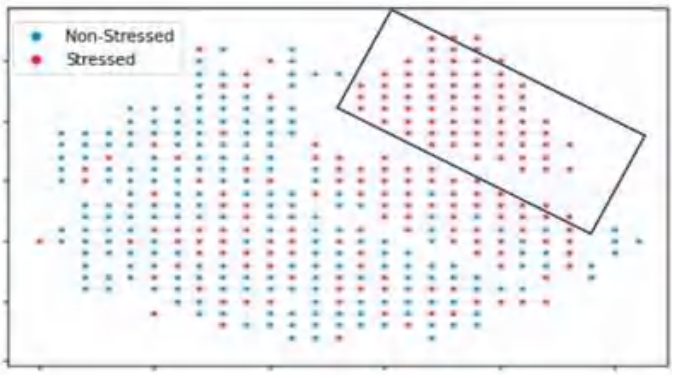


Fig. 7 Distribution node of DAIC-WOZ dataset cluster map

IV. Conclusion

In conclusion, the average accuracy and F1-Score results were 94.7% and 95%, respectively, using the child voice dataset built to support this study. For benchmark testing compared to the state-of-the-art model using the DAIC-WOZ dataset, the model also obtained results that exceeded these results with accuracy and an average F1-score of 89.8 and 88.2. Therefore, the model had generalization abilities over various voice samples. According to the cluster analysis, the model had good differentiation capabilities in identifying the subject's stress level with the unsupervised learning approach using GSOM. This was conducted by grouping nodes representing voice samples into appropriate clusters based on similarity.

Future studies could analyze the viability of detecting and monitoring stress in real-time from the child's activity environment. This will provide more information on the source of stress and stressors in children. However, a challenge to be overcome was optimizing model processing time in more complex audio pre-processing and audio separation approaches. Future model development can also integrate natural language processing methods into the model to detect stress through lexical speech analysis. With this integration, the model can detect stress more accurately by combining the signal and contextual properties of the subject's speech.

Acknowledgment

The authors are grateful to the Directorate of Research, Technology, and Community Service, Director General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for funding this study through the doctoral dissertation research grant (PDD) scheme.

References

[1] E. S. Epel et al., “More than a feeling: A unified view of stress measurement for population science,” Frontiers in Neuroendocrinology, vol. 49, pp. 146–169, Apr. 2018, doi:10.1016/j.yfrne.2018.03.001.

[2] M. Bucci, S. S. Marques, D. Oh, and N. B. Harris, “Toxic Stress in Children and Adolescents,” Advances in Pediatrics, vol. 63, no. 1, pp. 403–428, Aug. 2016, doi: 10.1016/j.yapd.2016.04.002.

[3] M. Kaczmarek and S. Trambacz-Oleszak, “School-Related Stressors and the Intensity of Perceived Stress Experienced by Adolescents in Poland,” International Journal of Environmental Research and Public Health, vol. 18, no. 22, p. 11791, Nov. 2021, doi:10.3390/ijerph182211791.

[4] N. Garmezy, A. S. Masten, and A. Tellegen, “The Study of Stress and Competence in Children: A Building Block for Developmental Psychopathology,” Child Development, vol. 55, no. 1, p. 97, Feb. 1984, doi: 10.2307/1129837.
[5] M. Rohmadi, M. Sudaryanto, C. Ulya, H. Akbariski, and U. Putri, “Case Study: Exploring Golden Age Students’ Ability and Identifying Learning Activities in Kindergarten,” Proceedings of the Proceedings of the First Brawijaya International Conference on Social and Political Sciences, BSPACE, 26-28 November, 2019, Malang, East Java, Indonesia, 2020, doi: 10.4108/eai.26-11-2019.2295218.
[6] H. Yaribeygi, Y. Panahi, H. Sahraei, T. P. Johnston, and A. Sahebkar, "The impact of stress on body function: A review.," *EXCLI J.*, vol. 16, pp. 1057–1072, 2017.
[7] P. Morgado and J. J. Cerqueira, Eds., The Impact of Stress on Cognition and Motivation. Frontiers Media SA, 2019. doi:10.3389/978-2-88945-774-8.
[8] M. Solmi et al., “Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies,” Molecular Psychiatry, vol. 27, no. 1, pp. 281–295, Jun. 2021, doi:10.1038/s41380-021-01161-7.
[9] M. Mohler-Kuo, S. Dzemaili, S. Foster, L. Werlen, and S. Walitza, “Stress and Mental Health among Children/Adolescents, Their Parents, and Young Adults during the First COVID-19 Lockdown in Switzerland,” International Journal of Environmental Research and Public Health, vol. 18, no. 9, p. 4668, Apr. 2021, doi:10.3390/ijerph18094668.
[10] Y. Choi, Y.-M. Jeon, L. Wang, and K. Kim, “A Biological Signal-Based Stress Monitoring Framework for Children Using Wearable Devices,” Sensors, vol. 17, no. 9, p. 1936, Aug. 2017, doi:10.3390/s17091936.
[11] T.-Y. Kim, L. Měsíček, and S.-H. Kim, “Modeling of Child Stress-State Identification Based on Biometric Information in Mobile Environment,” Mobile Information Systems, vol. 2021, pp. 1–13, Apr. 2021, doi: 10.1155/2021/5531770.
[12] K. E. Smith and S. D. Pollak, “Early life stress and development: potential mechanisms for adverse outcomes,” Journal of Neurodevelopmental Disorders, vol. 12, no. 1, Dec. 2020, doi:10.1186/s11689-020-09337-y.
[13] Y. S. Can, N. Chalabianloo, D. Ekiz, J. Fernandez-Alvarez, G. Riva, and C. Ersoy, “Personal Stress-Level Clustering and Decision-Level Smoothing to Enhance the Performance of Ambulatory Stress Detection With Smartwatches,” IEEE Access, vol. 8, pp. 38146–38163, 2020, doi: 10.1109/access.2020.2975351.
[14] K. Kyriakou et al., “Detecting Moments of Stress from Measurements of Wearable Physiological Sensors,” Sensors, vol. 19, no. 17, p. 3805, Sep. 2019, doi: 10.3390/s19173805.
[15] S. Gedam and S. Paul, “A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques,” IEEE Access, vol. 9, pp. 84045–84066, 2021, doi: 10.1109/access.2021.3085502.
[16] M. Chauhan, S. V. Vora, and D. Dabhi, “Effective stress detection using physiological parameters,” 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Mar. 2017, doi: 10.1109/iciiecs.2017.8275853.
[17] P. Chyan, A. Andani, I. Nurtanio, and I. Areni, "A Deep Learning Approach for Stress Detection Through Speech with Audio Feature Analysis," in *The 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE-2022)*, IEEE, 2022, pp. 269–273.
[18] G. M. Slavich, S. Taylor, and R. W. Picard, “Stress measurement using speech: Recent advancements, validation issues, and ethical and privacy considerations,” Stress, vol. 22, no. 4, pp. 408–413, Apr. 2019, doi: 10.1080/10253890.2019.1584180.
[19] S. Paulmann, D. Furnes, A. M. Bøkenes, and P. J. Cozzolino, “How Psychological Stress Affects Emotional Prosody,” PLOS ONE, vol. 11, no. 11, p. e0165022, Nov. 2016, doi:10.1371/journal.pone.0165022.
[20] K. Pisanski and P. Sorokowski, “Human Stress Detection: Cortisol Levels in Stressed Speakers Predict Voice-Based Judgments of Stress,” Perception, vol. 50, no. 1, pp. 80–87, Dec. 2020, doi:10.1177/0301006620978378.
[21] K. Tomba, J. Dumoulin, E. Mugellini, O. Abou Khaled, and S. Hawila, “Stress Detection Through Speech Analysis,” Proceedings of the 15th International Joint Conference on e-Business and Telecommunications, 2018, doi: 10.5220/0006855803940398.
[22] H. K. Shin, H. Han, K. Byun, and H. G. Kang, "Speaker-invariant Psychological Stress Detection Using Attention-based Network," *2020 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA ASC 2020 - Proc.*, no. December, pp. 308–313, 2020.
[23] R. Dillon and A. Ni Teoh, "Real-time Stress Detection Model and Voice Analysis: An Integrated VR-based Game for Training Public Speaking Skills," *IEEE Conf. Games*, pp. 1–4, 2021.
[24] I. Madhavi, S. Chamishka, R. Nawaratne, V. Nanayakkara, D. Alahakoon, and D. De Silva, “A Deep Learning Approach for Work Related Stress Detection from Audio Streams in Cyber Physical Environments,” 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Sep. 2020, doi: 10.1109/etfa46521.2020.9212098.
[25] A. König et al., “Measuring Stress in Health Professionals Over the Phone Using Automatic Speech Analysis During the COVID-19 Pandemic: Observational Pilot Study,” Journal of Medical Internet Research, vol. 23, no. 4, p. e24191, Apr. 2021, doi: 10.2196/24191.
[26] E. Rejaibi, A. Komaty, F. Meriaudeau, S. Agrebi, and A. Othmani, “MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech,” Biomedical Signal Processing and Control, vol. 71, p. 103107, Jan. 2022, doi:10.1016/j.bspc.2021.103107.
[27] G. Douzas, F. Bacao, J. Fonseca, and M. Khudinyan, “Imbalanced Learning in Land Cover Classification: Improving Minority Classes’ Prediction Accuracy Using the Geometric SMOTE Algorithm,” Remote Sensing, vol. 11, no. 24, p. 3040, Dec. 2019, doi:10.3390/rs11243040.
[28] Vandana, N. Marriwala, and D. Chaudhary, “A hybrid model for depression detection using deep learning,” Measurement: Sensors, vol. 25, p. 100587, Feb. 2023, doi: 10.1016/j.measen.2022.100587.
[29] N. Rafique, L. I. Al-Asoom, R. Latif, A. Al Sunni, and S. Wasi, “Comparing levels of psychological stress and its inducing factors among medical students,” Journal of Taibah University Medical Sciences, vol. 14, no. 6, pp. 488–494, Dec. 2019, doi: 10.1016/j.jtumed.2019.11.002.
[30] N. F. Narvaez Linares, V. Charron, A. J. Ouimet, P. R. Labelle, and H. Plamondon, “A systematic review of the Trier Social Stress Test methodology: Issues in promoting study comparison and replicable research,” Neurobiology of Stress, vol. 13, p. 100235, Nov. 2020, doi:10.1016/j.ynstr.2020.100235.
[31] Q. Ren, Y. Li, and D. Chen, “Measurement invariance of the Kessler Psychological Distress Scale (K10) among children of Chinese rural-to-urban migrant workers,” Brain and Behavior, vol. 11, no. 12, Nov. 2021, doi: 10.1002/brb3.2417.
[32] J. Gratch *et al.*, "The distress analysis interview corpus of human and computer interviews," *Proc. 9th Int. Conf. Lang. Resour. Eval. Lr. 2014*, pp. 3123–3128, 2014.
[33] A. Défossez, G. Synnaeve, and Y. Adi, "Real-time speech enhancement in the waveform domain," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, 2020.
[34] S. He-Ping, C. Ji-Hua, and L. Xiao, “Blind Source Separation for Non-stationary Signal Based on Time-Frequency Analysis,” 2011 4th International Conference on Intelligent Networks and Intelligent Systems, Nov. 2011, doi: 10.1109/icinis.2011.12.

International Journal of Innovative
Computing, Information and Control
Volume **19**, Number **6**, December **2023**

ICIC International ©2023 ISSN 1349-4198
pp. **1983–1994**

STRESS DETECTION OF CHILDREN THROUGH SPEECH SIGNALS IN MULTI-SPEAKER ENVIRONMENT USING DEEP LEARNING

Phie Chyan[1], Andani Achmad[1,*], Ingrid Nurtanio[2] and Intan Sari Areni[1]

[1]Department of Electrical Engineering
[2]Department of Informatics
Hasanuddin University
Jalan Poros Malino Km. 6, Gowa 92171, Indonesia
chyanp21d@student.unhas.ac.id; {ingrid; intan}@unhas.ac.id
*Corresponding author: andani@unhas.ac.id

Received April 2023; revised July 2023

Abstract. *Stress is a psychological problem that can affect anyone, including children. Detecting stress in children is a complex problem because they are generally unaware of the psychological problems they are experiencing and have verbal limitations that affect their communication skills with their parents. One of the biomarkers that can be used to detect stress is the voice (speech signal). The use of speech in stress detection has advantages in terms of convenience for the subject and ease of acquisition. This study proposes a stress detection model through speech signals in a multi-speaker environment. This model accepts audio input from the classroom environment, where there is noise and many speakers' voices overlap. The audio acquired is then separated using a speech separation algorithm based on an RNN architecture, producing output as segregated speech. The speech is then extracted for features and fed to the stress detection model based on CNN architecture, which predicts the speaker's stress status. The experimental results show that the proposed model is capable of speech separation with up to five speakers and predicts the stress status of the subject with an average accuracy of 95.6%.*

Keywords: Stress detection, Child mental stress, Speech separation, Speech signal, Deep learning

1. **Introduction.** Psychological stress is a change in a person's emotional state as a response to the pressures faced in everyday life. According to the psychological review, stress or distress is a condition that triggers a person to express negative emotions such as anger, sadness, panic, fear, and anxiousness [1]. Several studies related to psychological stress show that prolonged stress is closely correlated with decreased cognitive ability, motivation, decision-making skills, and spatial awareness of a person [1-4].

Children are a group that is also vulnerable to stress. Based on data from WHO (World Health Organization), the global prevalence of children with stress-related health problems is estimated to be around 13 percent of the population of children worldwide which, when observed statistically, is more or less balanced with the prevalence of mental problems in adults, which reaches 20 percent of the adult population [5]. This fact may indicate that mental problems experienced in childhood can be carried over into adulthood if no intervention is made to address the underlying problems. The main problem with stress ren is that detecting stress in children is more challenging than detecting Children, especially in pre-and early school age, generally have limited which limit their ability to communicate with their parents or caregivers s problems they face; moreover, a child is generally not aware of stress

icic.19.06.1983

itself as a psychological condition that can have a negative effect on them [6,7]. Stressors or stimuli that trigger stress in children can come from any source, such as family problems, difficulty following lessons, to bullying problems at school. Stress that occurs at this age which is still classified as a golden age, if left untreated, can cause problems that interfere with the growth and development of children [8,9].

Voice (speech) is one of the biomarkers that can be used to detect a state of psychological stress because voice output is a psychophysiological response resulting from the cooperation of approximately one hundred muscles connected by a network of cranial nerves and spinal nerves and is an integrative part of the psychophysiological stress system in humans [10,11]. Psychological stress experienced by a person affects how the body works, such as the emergence of muscular tension, increased respiratory rate, and excessive saliva production, which then affect voice reproduction, such as voice pitch, intonation, speech prosody, and many other sound parameters [12]. Using voice signals to detect stress eliminates the need for sensors or devices to be worn on the body, making it more convenient and safe for children. Voice can also be easily obtained from the microphone, making it possible to build a large database to support the stress detection model. However, the process of selecting and extracting sound features needed to support the model is crucial to obtain good accuracy [13].

Many studies on stress detection through speech have been conducted in recent years. However, this study uses voice acquired in a controlled or stationary environment, where each subject's voice is recorded individually with minimal or no ambient sound [10,14-19]. Although this method can achieve high accuracy, it is unlikely to identify the activity or event that triggers the stressful state (stressor) because the moment of the stressful event might have passed by the time the subject's voice was acquired. In addition, acquiring individual subject voices will increase the number of voice recordings that must be stored and analyzed. To overcome these problems, this study proposes a stress detection model that acquires voice directly from the environment where children learn and play. This environment is generally a non-stationary environment where many sound sources come from the surroundings, for example, the voices of other children or adults and noise from the environment. The solution to the problems presented is the primary focus of this study, which is to detect stress through the speech acquired from a noisy multi-speaker environment where the speaker's voices overlap.

Deep learning is used in a wide range of fields due to its ability to analyze large amounts of data and extract meaningful patterns. With the support of deep learning, areas such as computer vision and signal processing are experiencing very rapid development [20]. In this study, we use a deep learning approach to separate speech signals obtained from a multi-speaker environment. The results of separated speech are in the form of segregated speech from each child in the acquired sound recordings. Then we extract the speech samples to find discriminant features in detecting stress and use these features to train a model that can predict the subject's stress status. This proposed model, which is capable of detecting stress in a multi-speaker non-stationary environment, is the main contribution of this study. The results of this study help caregivers or parents better understand the stress that occurs in children and the potential underlying stressors so that stress mitigation

t more effectively.

of the paper is organized as follows. Section 2 discusses the methodology
y, and the proposed model. Section 3 presents the experimental results,
ıcludes this paper.

Methodology. To detect stress through sound acquired from a multi-
ent, we propose a system model as shown in Figure 1.

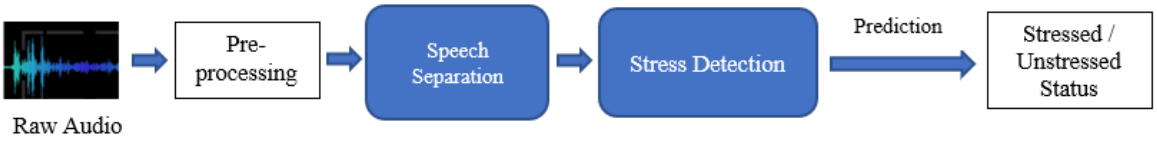


FIGURE 1. Model system diagram

Stress detection using speech in real-world applications requires the acquisition of sound directly from the environment of everyday human interaction. This method aims to get an overview of the causes (stressors) of the stress experienced. Regarding the interaction environment for children, the classroom at school is where children spend time. This environment is generally non-stationary because there are many sound sources with different frequency properties. Besides that, speech is naturally a form of non-stationary signal. In an interactive environment like this, there is more than one subject speaking simultaneously, so a speech separation approach is needed to deal with this problem. The problem referred as the cocktail party problem is the ability to track the voice of each specific subject when many subjects speak simultaneously, especially in an environment with noise [21,22]. Although this problem is easy for humans because the human senses can separate signals originating from multiple sources and focus on recognizing and tracking one particular source, it is a challenging task for computers [23].

In this study, we acquired sound directly from the classroom through a single-channel high-gain microphone placed in the middle of the room. The classroom size is 16 m^2 with five 1st grade elementary school students who are the subjects involved in this study. All the students and their parents agreed to participate in this study. This research was assisted by a child psychologist who designed and implemented various activities using the Trier Social Stress Test method, which aims to emulate stress caused by stressors children often face in everyday life at school [2,24]. Each child will do activities that involve public speaking and answering challenging math problems. Throughout the activity, the psychologist will observe the behaviour and gestures of each child for signs of stress. Before and after the activity, a saliva sample from each child will be taken to measure whether there is an increase in the hormone cortisol (stress hormone). Based on the theory, the cortisol hormone will increase 9 to 15 times from normal baseline levels during periods of stress and will last for several hours afterwards [25]. The results of the psychologist's observations, which were double-validated with the subject's cortisol hormone count, were used to label each child's speech as stressed or unstressed on the audio recording.

We use a deep learning approach to perform speech separation, producing segregated speech from each subject. Each sound recording is then cut into sound samples between 1 and 2 seconds long. After the data validation, we obtained 500 sound samples, each consisting of 223 samples labelled stressed and 277 samples labelled unstressed. Furthermore, for a more detailed discussion, Section 2.1 will explain in detail the speech separation model we use, and Section 2.2 will elaborate on the stress detection model in detail.

2.1. **Speech separation.** Before being fed into the speech separation module, the speech acquired first goes through the preprocessing phase for noise reduction and silence removal. ı in Figure 2 is derived from the latest development of the speech sepa- ed on dual path RNN (Recurrent Neural Network) [26-28]. Input in the signal (waveform) containing speech mixture $x \in R^T$ is entered into the and will produce an output of N dimensional z of size $T' = (2T/L) - 1$ length and L is the encoding compression factor, which then produce

$$z = E(x) \tag{1}$$

where E is a 1D convolutional layer network with kernel size of L and stride of $L/2$, subsequent by non-linear ReLU (Rectified Linear Unit) activation function, x is speech mixture from audio signal, and z is resulting latent representation. The 3D tensor $\nu = [\nu_1, \ldots, \nu_R] \in R^{N \times K \times R}$ is obtained by concatenating each chunk on a single dimension. Then ν will be sent to the separation network Y, which is made up of b RNN blocks [27]. The even blocks B_{2i} will be used along chunk dimensions of size K, while odd blocks B_{2i-1} will be used along the time dependent dimension. The RNN block used contains Multiply and Concat (MULCAT) blocks with two sub networks. It then uses two separate bidirectional LSTM (Long Short-Term Memory) networks to multiply its outputs and combines its inputs to produce a module output using Formula (2)

$$B_i(\nu) = P_i([M_i^1(\nu) \odot M_i^2(\nu), \nu]) \tag{2}$$

where P_i is a learned linear project that converts an input's dimension (ν) into the dimension of the output obtained by concatenating the product of the two LTSMs denoted by M_i^1 and M_i^2, $\odot$ is the element-wise product operation, and B_i is resulting RNN block. An overview of the RNN block pairs can be seen in Figure 3. In this method after each pair of blocks processed requires model to reconstruct the original audio. The 3D tensor via PreLU initialized at 0.25. The decoder is a 1×1 convolution with a C output channel. Each pair of blocks will be decoded using the same PreLU and decoder parameters. We use the add and overlap operators to convert back the tensor to audio. This operator is used to reverse the chunking process and add overlapping frames from the signal.

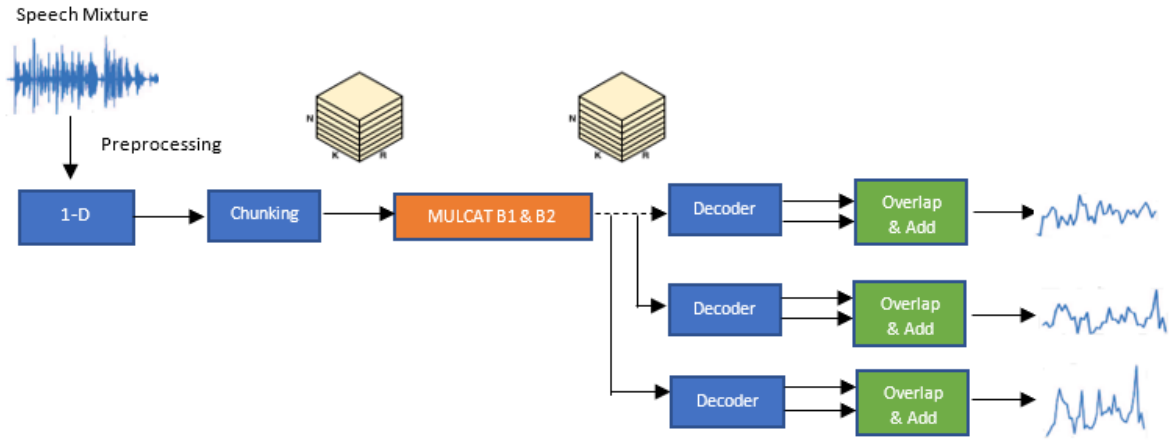


FIGURE 2. Speech separation model

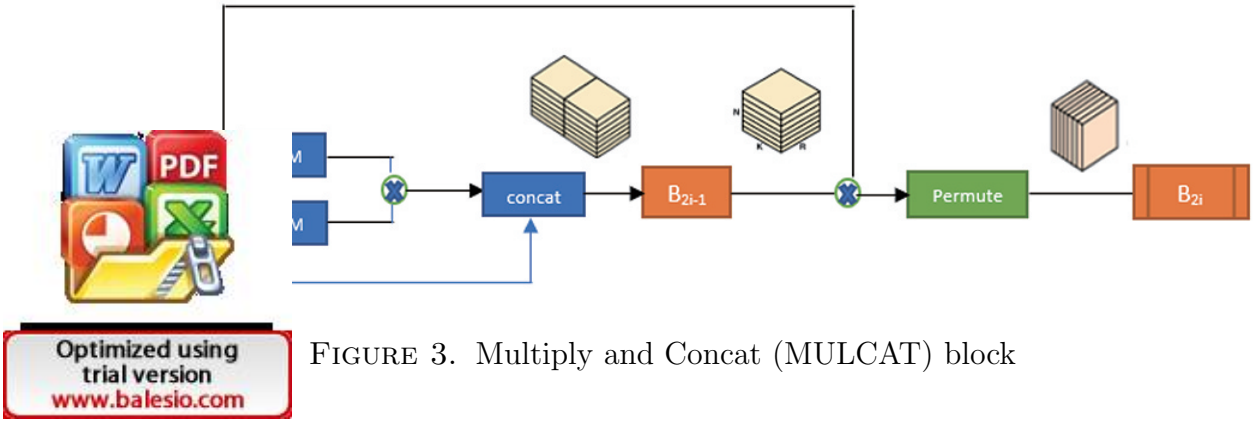


FIGURE 3. Multiply and Concat (MULCAT) block

For the speech separation model in this non-stationary environment, we use the open-source dataset LibriMix, derived from LibriSpeech (clean speech) and WHAM! Noise. We generate two to five speech mixtures, namely LibriMix-2 to LibriMix-5. Each dataset consists of 20 hours of training and 10 hours of validation and test sets. Each speech separation model will be trained with these datasets, which correlate with the number of possible output channels (number of students in the class).

Because the number of speakers from the sound acquired can vary up to 5 speakers, the selection process begins by using a model trained using the largest number of sounds C, which is five. The detection of each output channel is carried out. If an empty channel (silence) is obtained, the model selection is repeated using the model with the $C-1$ output channel. The process will continue to be repeated until all the channels are filled, representing the correct C output model for the speech mixture. Figure 4 shows a flowchart of the model selection algorithm to match the number of speakers in the audio mixture. For implementation, we use batch size 2 with ADAM optimizer. Input kernel size 4. The architecture uses 6 MULCAT blocks where each LSTM layer contains 128 neurons. To evaluate the model, the Si-SNRi metric (Scale-invariant Signal-to-Noise Ratio improvement) is used to measure the quality of the speech separation results [29].

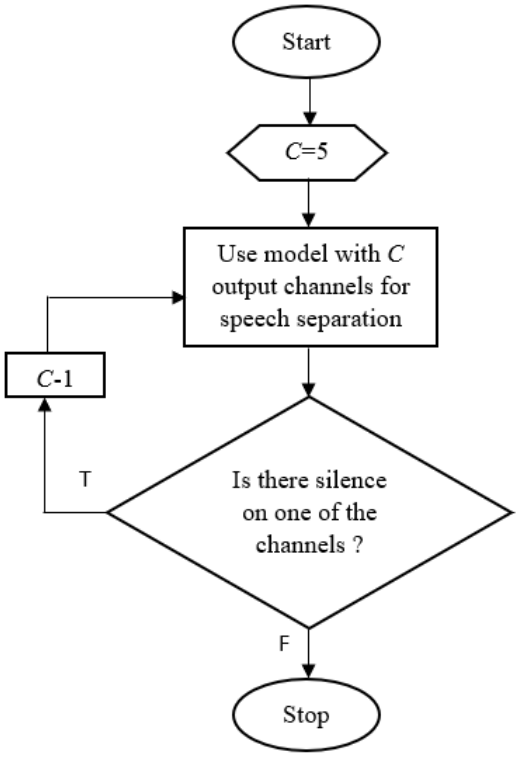


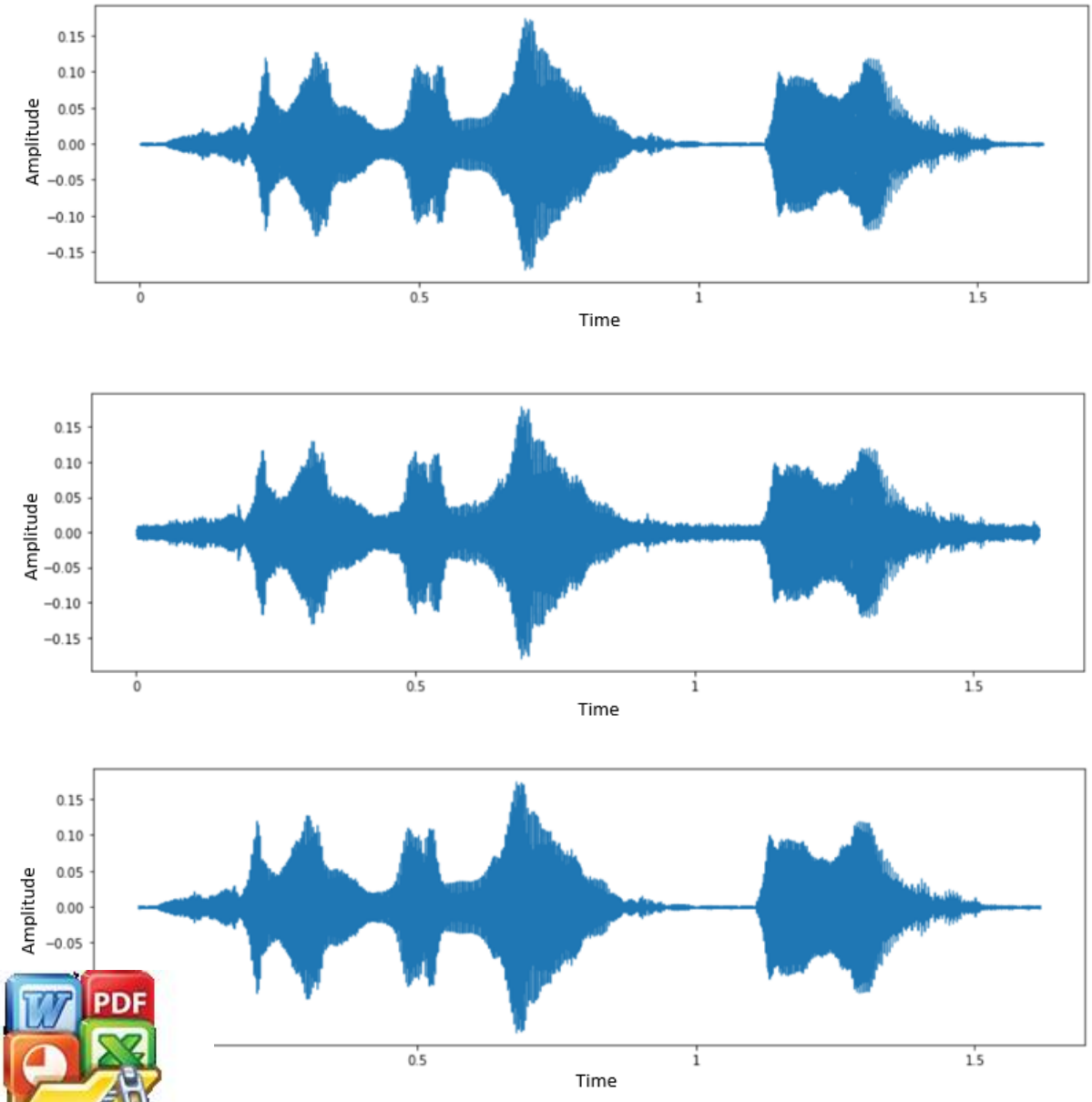
FIGURE 4. Flowchart of model selection to match the number of speakers in audio sample

ection. To build this stress detection model, in addition to using our :, we also use audio data from the open source TESS (Toronto Emotion- nd SAVEE (Surrey Audio-Visual Expressed Emotion) datasets, which d 480 audio files, respectively. Both datasets consist of audio files that ; emotional speeches. Following the review of the psychological theory hange in psychological reactions that are manifested in various negative sed on this theory, we conduct a relabeling process for the dataset used.

The original dataset files labelled sad, angry, disgusted, and fearful were relabeled with the stress label, while files labelled pleasant, pleasantly surprised and neutral were relabeled with the unstressed label. After relabeling and adding our own built dataset, 1,973 data were labelled with stressed labels and 1,807 with unstressed labels.

To enrich the dataset with better generalization capabilities for various noise signal disturbances, we perform a data augmentation process to create new synthetic sample data. Each original audio file is added with two additional variations. The first variation is the original audio file with additional noise injection. The second variation is the original audio with a modified pitch. Thus, 11,340 sound sample files are obtained after this augmentation. Figure 5 shows the waveform of the original audio data and the two additional different versions resulting from the augmentation process.

For the model to perform the classification process, it is necessary to perform feature extraction to convert the audio signal to a format the model can comprehend. Our stress



. The original speech waveform (top), noise added modification
l pitch shifting modification (bottom)

detection model, as shown in Figure 6, is based on the CNN (Convolutional Neural Network) architecture with the Convolution 1D (Conv 1D) sequential model, which consists of 4 convolution layers and two dense layers followed by a flattened layer. Each convolution layer has a max pooling layer to get maximum pattern variations. From each audio sample, various features in the signal domain are extracted related to the speech signal's attributes. This signal domain consists of the time, frequency, and cepstral domain [31,32]. Time domain features refer to characteristics or properties of an audio signal that are derived from the waveform. The frequency domain feature refers to the analytic space in which mathematical functions or signals are conveyed in terms of frequency, rather than time, which result from conversion from the time domain using the Fourier transforms. The cepstral domain is a feature in the cepstral domain obtained by the inverse Fourier transform of the logarithm spectrum of the signal. Cepstrum is often used as a feature vector to represent human voices and musical sounds. So to get the properties of speech, cepstral is a feature widely used, such as for speech recognition needs.

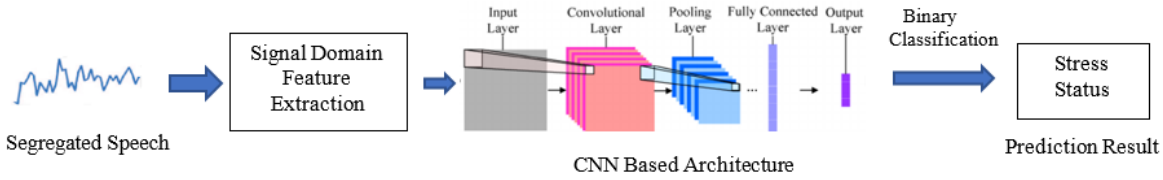


FIGURE 6. Stress detection model

For feature extraction needs to be fed to the model, we will experiment using the cepstral domain features and also a combination of the cepstral domain features with the time domain and frequency domain features to see if there is an increase in accuracy to be obtained by combining the three signal domain features. Table 1 shows the audio features used in each signal domain for feature extraction.

TABLE 1. The audio features used in signal domain

Signal domain feature	Audio features used
Time domain	Amplitude Envelope (AE)
	Zero Crossing Rate (ZCR)
	Root Mean Square (RMS)
Frequency domain	Spectral Centroid (SC)
	Spectral Rolloff (SR)
	Spectral Bandwidth (SB)
Cepstral domain	Mel Spectogram

The model was trained using a 75 : 25 random training – testing split for the training phase. The softmax activation model is used in a dense layer with two neurons. CNN was trained for 40 epochs with a batch size of 32 using the Adam optimizer. In addition to calculating the model's accuracy, three additional metrics are used to evaluate model ecision, Recall, and F1 Score. The Precision (P) metric compares True ıd the number of data predicted to be positive. Recall (R) compares the P) and the number of positive data. The F1 Score (F1) is the harmonic ion and recall.

Result. This section will explain the experimental results, performance model's speech separation and stress detection.

3.1. **Speech separation model result.** We evaluate the speech separation model using the SI-SNRi metric to see the quality of the sound separations obtained. The model trained with the sound sample C_m can train audio samples with the number of speakers C as long as $C_m \geq C$. Suppose the number of speakers in the model used exceeds the number of speakers in the audio sample. In that case, the separation results will produce an unused channel containing a silent signal. Utilizing a model that corresponds to the number of speakers in the sound sample will provide more optimal performance; therefore, for sound samples when C is unknown, the model selection algorithm outlined in Section 2 will be utilized, which will select the right model based on channel activity detection. Table 2 shows the obtained SI-SNRi values. Based on these results, the separation performance is better when the model is used on par with the number of speakers in the audio sample.

TABLE 2. SI-SNRi score of various models used for the number of speakers in the audio sample

Used model	The number of speakers in the audio sample			
	2	3	4	5
2-speaker model	17.5	–	–	–
3-speaker model	12.3	13.6	–	–
4-speaker model	9.6	10.8	9.7	–
5-speaker model	6.5	8.7	8.4	7.5

3.2. **Stress detection model result.** We tested the model using each feature in the signal domain. Besides that, we also tried to use a combination of the three signal domain (time, frequency and cepstral) features to see whether combining all signal domain features for the model can improve the accuracy of the stress detection model. The dataset used for the training and test procedures combines the open source dataset and our own built dataset, which aims to provide the model with better generalization ability and applicability in various speech acquisition environments. The results obtained in Figure 7 show that the cepstral domain feature is a discriminant feature for stress detection with an average accuracy of 95.6% and an average F1 Score of 94.8%. The result also confirms that cepstral-based audio features such as the Mel spectrogram correlate with how the human ear perceives sound and are very good at representing various speech signal properties. Based on the results, audio features in the time and frequency domains are less accurate for sound stress detection. Likewise, when the two feature groups are combined with cepstral-based audio features, they do not significantly affect the model's performance.

We also conducted tests to compare the performance of models trained using open-source datasets and trained using our own built dataset. Both models were trained using cepstral domain features as extracted features. The results, as shown in Figure 8, show that using an open-source dataset produces model performance with an average accuracy of 97.4% and an average F1 Score of 96.5% which is significantly better than using our
; with an average accuracy of 88.2% and an average F1 Score of 90%.
in performance obtained is mainly because the open source dataset used
› samples acquired in a controlled room where there is no noise or other
hereas, in our own built dataset, the audio was acquired directly from
ng environment which is a room that is relatively noisy with multiple
simultaneously. In addition, the performance of the speech separation
nces the performance of the stress detection model because the speech

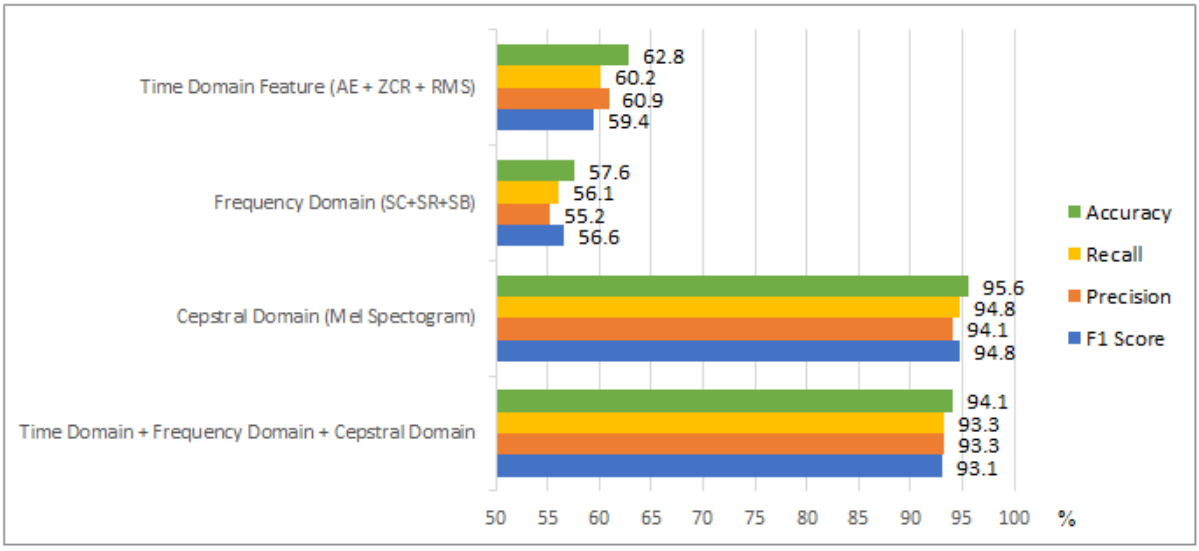


FIGURE 7. Model evaluation results using various audio signal domain features

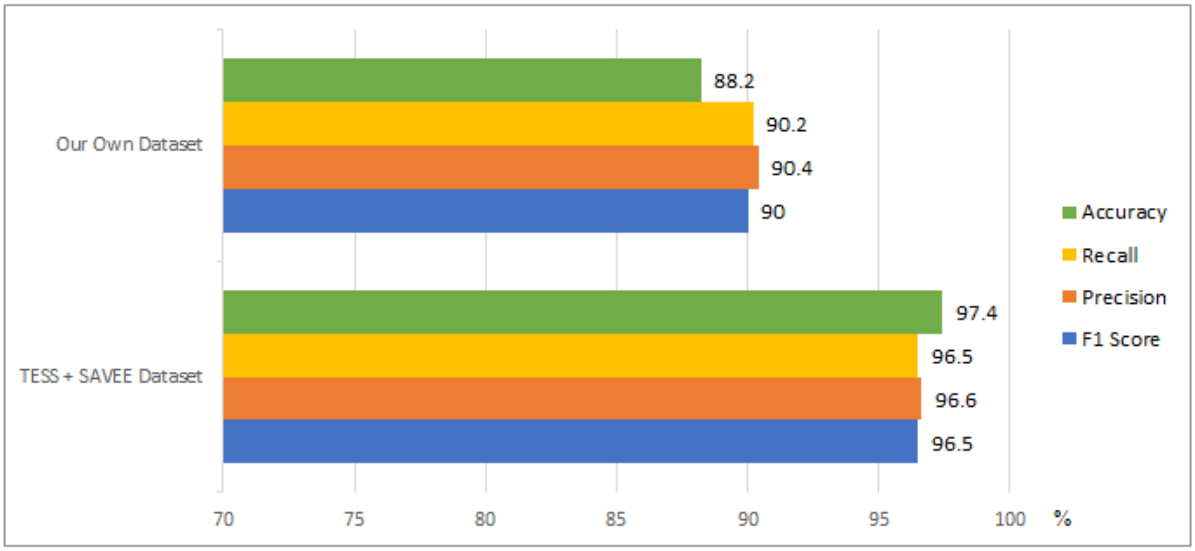


FIGURE 8. The comparison between the model's performance evaluation using open source dataset and our own built dataset

sample in our own built dataset is derived from segregated speech as a product of the speech separation model performed.

4. **Conclusion.** This study proposes a stress detection model through speech in a multi-speaker environment. Audio is acquired directly in the children's learning environment, a non-stationary environment with much noise and many speakers' voices overlapping. The proposed model performs speech separation to produce segregated sounds. The results [illegible] voice samples then are extracted using cepstral audio features that are [illegible] detecting stress. Then, the results are fed to the model to predict the [illegible]he subject. Based on the experimental results, it was found that the [illegible]an detect stress in subjects with high accuracy. Using a combination of [illegible]sets and our own built dataset, we obtain an average accuracy of 95.6% [illegible]1 Score of 94.8%.

We have yet to achieve real-time detection in this proposed model due to high computational costs, especially in speech separation. Also, for the same reason, the number of overlapping voices that can be separated in this model is a maximum of 5 in the acquired voice sample. In future work, we will explore the possibility of optimizing the speech separation algorithm so that the model can achieve real-time detection, which allows the model to detect and monitor the psychological condition of the subject directly through conversation.

Acknowledgment. The grants from Indonesia's Ministry of Education, Culture, Research, and Technology's "Penelitian Disertasi Doktor" Scheme 2023 support this work.

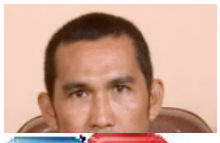
REFERENCES

[1] J. A. Healey and R. W. Picard, Detecting stress during real-world driving tasks using physiological sensors, *IEEE Trans. Intell. Transp. Syst.*, 2005.
[2] S. Wemm and E. Wulfert, Effects of acute stress on decision making, *Physiol. Behav.*, vol.176, no.3, pp.139-148, 2017.
[3] P. Morgado and J. Cerqueira, The impact of stress on cognition and motivation, *Front. Behav. Neurosci.*, 2018.
[4] H. Yaribeygi, Y. Panahi, H. Sahraei, T. P. Johnston and A. Sahebkar, The impact of stress on body function: A review, *EXCLI J.*, vol.16, pp.1057-1072, 2017.
[5] World Health Organization (WHO), *Mental Health*, 2021.
[6] Y. Choi, Y. M. Jeon, L. Wang and K. Kim, A biological signal-based stress monitoring framework for children using wearable devices, *Sensors (Switzerland)*, vol.17, no.9, pp.1-16, 2017.
[7] T. Y. Kim, L. Měsíček and S. H. Kim, Modeling of child stress-state identification based on biometric information in mobile environment, *Mob. Inf. Syst.*, vol.2021, 2021.
[8] M. Bucci, S. S. Marques, D. Oh and N. B. Harris, Toxic stress in children and adolescents, *Adv. Pediatr.*, vol.63, no.1, pp.403-428, 2016.
[9] N. Garmezy, A. S. Masten and A. Tellegen, The study of stress and competence in children: A building block for developmental psychopathology, *Child Dev.*, vol.55, no.1, pp.97-111, 1984.
[10] G. M. Slavich, S. Taylor and R. W. Picard, Stress measurement using speech: Recent advancements, validation issues, and ethical and privacy considerations, *Stress*, vol.22, no.4, pp.408-413, 2019.
[11] S. Paulmann, D. Furnes, A. M. Bøkenes and P. J. Cozzolino, How psychological stress affects emotional prosody, *PLoS One*, vol.11, no.11, pp.1-21, 2016.
[12] K. Pisanski and P. Sorokowski, Human stress detection: Cortisol levels in stressed speakers predict voice-based judgments of stress, *Perception*, vol.50, no.1, pp.80-87, 2021.
[13] P. Chyan, A. Andani, I. Nurtanio and I. Areni, A deep learning approach for stress detection through speech with audio feature analysis, *The 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE2022)*, pp.269-273, 2022.
[14] K. Tomba, J. Dumoulin, E. Mugellini, O. A. Khaled and S. Hawila, Stress detection through speech analysis, *Proc. of the 15th International Joint Conference on e-Business and Telecommunications (ICETE2018)*, 2018.
[15] C. A. Jason and S. Kumar, An appraisal on speech and emotion recognition technologies based on machine learning, *Int. J. Recent Technol. Eng.*, vol.8, no.5, pp.2266-2276, 2020.
[16] A. König et al., Measuring stress in health professionals over the phone using automatic speech analysis during the COVID-19 pandemic: Observational pilot study, *J. Med. Internet Res.*, vol.23, no.4, pp.1-14, 2021.
[17] N. Matsuo, S. Hayakawa and S. Harada, Technology to detect levels of stress based on voice information, *Fujitsu Sci. Tech. J.*, vol.51, no.4, pp.48-54, 2015.
[18] H. Han, K. Byun and H. G. Kang, A deep learning-based stress detection algorithm with speech *the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia* 11-15, 2018.
Beňuš and M. Trnka, Stress detection using non-semantic speech representation, *2022 nal Conference Radioelektronika (RADIOELEKTRONIKA)*, pp.1-5, 2022.
T. Zin, P. Tin and I. Kobayashi, Automatic detection and tracking of mounting tle using a deep learning-based instance segmentation model, *International Journal omputing, Information and Control*, vol.18, no.1, pp.211-220, 2022.

[21] Y. Qian, C. Weng, X. Chang, S. Wang and D. Yu, Past review, current progress, and challenges ahead on the cocktail party problem, *Front. Inf. Technol. Electron. Eng.*, vol.19, no.1, pp.40-63, 2018.
[22] Y. Li, F. Wang, Y. Chen, A. Cichocki and T. Sejnowski, The effects of audiovisual inputs on solving the cocktail party problem in the human brain: An fMRI study, *Cereb. Cortex*, vol.28, no.10, pp.3623-3637, 2018.
[23] J. R. Hershey, Z. Chen, J. Le Roux and S. Watanabe, Deep clustering: Discriminative embeddings for segmentation and separation, *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2016)*, pp.31-35, 2016.
[24] M. A. Vallejo, L. Vallejo-Slocker, E. G. Fernández-Abascal and G. Mañanes, Determining factors for stress perception assessed with the Perceived Stress Scale (PSS-4) in Spanish and other European samples, *Front. Psychol.*, vol.9, no.1, 2018.
[25] K. E. Hannibal and M. D. Bishop, Chronic stress, cortisol dysfunction, and pain: A psychoneuroendocrine rationale for stress management in pain rehabilitation, *Phys. Ther.*, vol.94, no.12, pp.1816-1825, 2014.
[26] Z. Chen, Y. Luo and N. Mesgarani, Deep attractor network for single-microphone speaker separation, *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2017)*, vol.2, no.1, pp.246-250, 2017.
[27] Y. Luo, Z. Chen and T. Yoshioka, Dual-Path RNN: Efficient long sequence modeling for time-domain single-channel speech separation, *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2020)*, pp.46-50, 2020.
[28] E. Nachmani, Y. Adi and L. Wolf, Voice separation with an unknown number of multiple speakers, *The 37th Int. Conf. Mach. Learn. (ICML2020)*, pp.7121-7132, 2020.
[29] A. Wijayakusuma, D. R. Gozali, A. Widjaja and H. Ham, Implementation of real-time speech separation model using time-domain audio separation network (TasNet) and dual-path recurrent neural network (DPRNN), *Procedia Comput. Sci.*, vol.179, pp.762-772, 2021.
[30] E. S. Epel et al., More than a feeling: A unified view of stress measurement for population science, *Front. Neuroendocrinol.*, vol.49, no.12, pp.146-169, 2018.
[31] H. P. Shi, J. H. Cao and X. Liu, Blind source separation for non-stationary signal based on time-frequency analysis, *Proc. of 2011 4th Int. Conf. Intell. Networks Intell. Syst. (ICINIS2011)*, pp.45-48, 2011.
[32] A. Défossez, G. Synnaeve and Y. Adi, Real time speech enhancement in the waveform domain, *arXiv.org*, arXiv: 2006.12847, 2020.

Author Biography

Phie Chyan received the bachelor's degree in Electrical Engineering from Atma Jaya Makassar University, Indonesia, 2004; the master of Computer Science from Gadjah Mada University, Jogjakarta, Indonesia, 2011. His main research interests include computer vision, multimedia processing, and expert system. Now he is pursuing his Ph.D. degree in Hasanuddin University. He is the staff of Department of Informatics, Faculty of Information Technology, Atma Jaya Makassar University.



Andani Achmad received a bachelor's degree in 1986, a master's degree in 2000 and a doctorate in 2010 from Hasanuddin University in Indonesia. His primary topic of study is Electrical Power Engineering. He is a Professor at Hasanuddin University's Department of Electrical Engineering. He is current head of the doctorate program in electrical engineering, Hasanuddin University.

Ingrid Nurtanio received the bachelor's degree in Electrical Engineering from Hasanuddin University, Makassar, Indonesia in 1986. She received her master of technology from Hasanuddin University, Makassar, Indonesia in 2002. She received her doctoral degree from Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia in 2013. Her research interest is digital image processing, computer vision and intelligent system. Currently, she is the staff of Department of Informatics, Faculty of Engineering, Hasanuddin University. She is a member of IAENG and IEEE.

Intan Sari Areni received B.E. and M.E. degrees in Electrical Engineering from Hasanuddin University (UNHAS), Makassar (1999) and Gadjah Mada University (UGM), Jogjakarta (2002), respectively, and received a Doctorate degree from Ehime University, Japan in 2013. Currently, she is a Professor at the Department of Electrical Engineering, Hasanuddin University. Her research interests include multimedia signal processing, telecommunication, biomedical engineering, computer vision, and powerline communication system. She is a member of IEEE and IAENG.

Lampiran 2. Biodata

IDENTITAS DIRI

Nama Lengkap : Phie Chyan
Tempat / Tgl. Lahir : Ujung Pandang, 13 April 1981
Jenis Kelamin : Laki-laki
NIDN : 0913048102
Jab/Pa/Gol.Rg : Lektor Kepala / Pembina / IV/a
Pekerjaan : Dosen Fakultas Teknologi Informasi Univ. Atma Jaya Mks.
Alamat Rumah : Jalan Nuri Lama No. 23c Makassar
No. HP : 0819 4422 1320
Nama Ayah / Ibu : Phie Go Kiong / Henny Lim
Nama Istri : Suryuli

RIWAYAT PENDIDIKAN

Thn Lulus	Jenjang	Perguruan Tinggi	Program Studi
2003	Sarjana	Atma Jaya Makassar	Teknik Elektro
2011	Magister	Universitas Gadjah Mada	Ilmu Komputer

RIWAYAT PEKERJAAN

Periode Tahun	Status Dosen	Perguruan Tinggi
2006- Sekarang	Dosen Tetap	FTI, Univ. Atma Jaya Makassar

KARYA ILMIAH / ARTIKEL JURNAL YANG TELAH DIPUBLIKASIKAN DALAM 5 TAHUN TERAKHIR

No	Tahun	Judul	Jurnal
1	2023	*Stress Detection of Children Through Speech Signals in Multi-Speaker Environment Using Deep Learning*	International Journal of Innovative Computing, Information, and Control Vol.19 No.6, 2023
2	2023	*Hybrid Deep Learning Approach for Stress Detection Model Through Speech Signal*	International Journal on Informatics Visualization Vol. 7, No. 4, 2023
	23	*Image Restoration Using Deep Learning Based Image*	Jurnal Sisfokom (Sistem Informasi Dan Komputer) 12. (3) ISSN 2301-7988

No	Tahun	Judul	Jurnal
		Completion	
4	2023	*Analysis of Supermarket Product Purchase Transactions With the Association Data Mining Method*	Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) 7 (3), 618-627
5	2022	*Segmentasi Kulit Manusia Dengan Ekstraksi Fitur Warna Dan Algoritma GMM-EM*	Jurnal Pendidikan Teknologi Informasi (JUKANTI) 5 (1), 151-156
6	2022	*Pemulihan Citra Berbasis Metode Markov Random Field*	JURIKOM (Jurnal Riset Komputer) 9 (2), 218-223
7	2021	*Sistem Monitoring dan Deteksi Stres Pada Anak Berbasis Wearable Device*	Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) 5 (5), 943-949
8	2021	*Perancangan Learning Management System Sebagai Pendukung Pembelajaran Jarak Jauh*	Rabit: Jurnal Teknologi dan Sistem Informasi Univrab 6 (1), 7-13
9	2021	*Using K-Means Algorithm to Investigate Community Behavior in Treating Waste toward Smart City*	International Journal on Advanced Science, Engineering and Information Technology 11 (4), 1455-1462
10	2020	*Strategi Penerapan Tata Kelola Smart City Dengan Elemen Smart Readiness*	Jurnal Teknologi Industri dan Rekayasa (JTIR) 1 (1), 26-33
11	2020	*Analysis and Design of Waste Management System Using the Spiral Model Towards Smart Cities*	Jurnal Sisforma 6 (2), 41-47
	19	*Decision Support System For Property Investment Selection In Makassar City*	Journal of Engineering and Applied Sciences 14 (23), 8705-8711

No	Tahun	Judul	Jurnal
13	2019	*Image enhancement based on bee colony algorithm*	Journal of Engineering and Applied Sciences 14 (1), 43-49

KARYA ILMIAH / ARTIKEL JURNAL YANG TELAH DIPUBLIKASIKAN DALAM 5 TAHUN TERAKHIR

No	Tahun	Judul	Konferensi Internasional
1	2023	*Multi-Stage Approach for Stress Detection Using Speech Lexical Analysis*	2023 IEEE 7th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)
2	2022	*A Deep Learning Approach for Stress Detection Through Speech with Audio Feature Analysis*	2022 IEEE 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)
3	2021	*Automatic monitoring system for the elderly based on internet of things*	Annual Conference on Computer Science and Engineering Technology (AC2SET) 23rd September 2020, Medan, Indonesia
4	2020	*Geographic Information System for Waste Management for the Development of Smart City Governance*	The 2nd International Conference On Science And Innovated Engineering 9 - 10 November 2019, Malacca, Malaysia
5	2019	*Design of intelligent camera-based security system with image enhancement support*	The 3rd International Conference On Science 26–27 July 2019, Makassar, Indonesia

KARYA BUKU YANG TELAH DIPUBLIKASIKAN DALAM 5 TAHUN TERAKHIR

No	Tahun	Judul	Penerbit
	23	Pengantar Data Mining	PT. Mifandi Mandiri Digital ISBN: 978-623-88688-9-6, 204 halaman

No	Tahun	Judul	Penerbit
2	2023	Pengantar Jaringan Komputer	PT. Mifandi Mandiri Digital ISBN: 978-623-09369-1-3, 184 halaman
3	2023	Sistem Pendukung Keputusan	PT. Mifandi Mandiri Digital ISBN: 978-623-09369-1-3, 120 halaman
4	2022	Pengantar Sistem Informasi	PT. Mifandi Mandiri Digital ISBN: 978-623-09128-9-4 , 169 halaman

PENGALAMAN HIBAH PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT

No	Tahun	Tingkat	Skim Hibah	Pelaksana
1	2023	Nasional	Penelitian Dasar (Penelitian Disertasi Doktor)	Direktorat Riset dan Pengabdian kepada masyarakat (DRPM)
1	2018-2020	Nasional	Penelitian Strategi Nasional Institusi (PSNI)	Direktorat Riset dan Pengabdian kepada masyarakat (DRPM)
2	2017-2019	Nasional	Penelitian Produk Terapan (PPT)	Direktorat Riset dan Pengabdian kepada masyarakat (DRPM)
3	2016	Nasional	IPTEK bagi Masyarakat (IbM)	Direktorat Riset dan Pengabdian kepada masyarakat (DRPM)
4	2015-2016	Nasional	Penelitian Hibah Bersaing (PHB)	Direktorat Riset dan Pengabdian kepada masyarakat (DRPM)
5	2013-2014	Nasional	Penelitian Dosen Pemula	Direktorat Riset dan Pengabdian kepada masyarakat (DRPM)