

DISERTASI

MODEL DETEKSI STRES MELALUI ANALISIS SUARA DALAM LINGKUNGAN NON-STASIONER MULTI SPEAKER DENGAN PENDEKATAN DEEP LEARNING

*Stress Detection Model Through Speech Analysis in Multi Speaker
Non-Stationary Environment with Deep Learning Approach*

PHIE CHYAN

D053211003



PROGRAM STUDI DOKTOR ILMU TEKNIK ELEKTRO
DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024



PENGAJUAN DISERTASI

MODEL DETEKSI STRES MELALUI ANALISIS SUARA DALAM LINGKUNGAN NON-STASIONER MULTI SPEAKER DENGAN PENDEKATAN DEEP LEARNING

Disertasi
Sebagai Salah Satu Syarat untuk Mencapai Gelar Doktor
Program Studi Ilmu Teknik Elektro

Disusun dan diajukan oleh

**PHIE CHYAN
D053211003**

Kepada

**FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2023**



DISERTASI

MODEL DETEKSI STRES MELALUI ANALISIS SUARA DALAM LINGKUNGAN NON-STASIONER MULTI SPEAKER DENGAN PENDEKATAN DEEP LEARNING

PHIE CHYAN
NIM : D053211003

Telah dipertahankan di hadapan Panitia Ujian Disertasi yang dibentuk dalam rangka penyelesaian studi pada Program Doktor Teknik Elektro Fakultas Teknik Universitas Hasanuddin
Pada tanggal 15 Januari 2024
Dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui
Promotor



Prof. Dr. Ir. H. Andani Achmad, MT
NIP. 196012311987031022

Co-Promotor



Dr. Ir. Ingrid Nurtanio, MT
NIP. 196108131988112001

Co-Promotor



Prof. Dr. Eng. Intan Sari Areni, ST, MT
NIP. 1975020032000122002

Dekan Fakultas Teknik
Universitas Hasanuddin



Eng. Ir. M. Isran Ramli, ST, MT.
197309262000121002

Ketua Program Studi
S3 Teknik Elektro



Prof. Dr. Ir. H. Andani Achmad, MT
NIP. 196012311987031022



Optimized using
trial version
www.balesio.com



Balai
Sertifikasi
Elektronik

- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE
- UU ITE No 11 Tahun 2008 Pasal 5 Ayat 1

"Informasi Elektronik dan/atau Dokumen Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"

PERNYATAAN KEASLIAN DISERTASI DAN PELIMPAHAN HAK CIPTA

Yang bertanda tangan di bawah ini

Nama : Phie Chyan

Nomor Mahasiswa : D053211003

Program Studi : S3 Teknik Elektro

Dengan ini menyatakan bahwa, disertasi berjudul “Model Deteksi Stres Melalui Analisis Suara Dalam Lingkungan Non-Stasioner Multi Speaker Dengan Pendekatan Deep Learning” adalah benar karya saya dengan arahan dari komisi pembimbing (Prof. Dr. Ir. H. Andani Achmad, M.T., sebagai promotor, Dr. Ir. Ingrid Nurtanio, MT sebagai co-promotor-1 dan Prof. Dr.Eng. Intan Sari Areni, S.T., M.T.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka disertasi ini. Sebagian dari isi disertasi ini telah dipublikasikan di prosiding (2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering, DOI:10.1109/ICITISEE57756.2022.10057845), Sebagai artikel dengan judul “A Deep Learning Approach for Stress Detection Through Speech with Audio Feature Analysis” kemudian di Jurnal (International Journal of Innovative Computing, Information and Control, Vol.20, No.1, 2023) sebagai artikel dengan judul “Stress Detection of Children Through Speech Signals in Multi-Speaker Environment Using Deep Learning”, di jurnal (JOIV : International Journal on Informatics Visualization, Vol.7, No.4, 2023) sebagai artikel dengan judul “Hybrid Deep Learning Approach For Stress Detection Model Through Speech Signal” dan terakhir di prosiding (2023 7th International Conference on Information Technology, Information Systems and Electrical Engineering) sebagai artikel dengan judul “Multi-Stage Approach for Stress Detection Using Speech Lexical Analysis”.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa disertasi ini kepada Universitas Hasanuddin

Gowa, November 2023

Yang menyatakan



Phie Chyan



KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Tuhan Yang Maha Esa, yang atas rahmat dan karunianya sehingga penulisan penelitian disertasi dengan judul **“MODEL DETEKSI STRES MELALUI ANALISIS SUARA DALAM LINGKUNGAN NON-STASIONER MULTI SPEAKER DENGAN PENDEKATAN DEEP LEARNING”** dapat penulis susun dan selesaikan dengan baik.

Dalam proses penyusunan disertasi ini yang penuh tantangan dalam mewujudkan gagasan-gagasan ke dalam sebuah hasil penyusunan disertasi yang lengkap, penulis banyak mendapatkan dukungan dari berbagai pihak yang telah membantu penulis dalam menyusun disertasi yang sesuai dengan kaidah-kaidah keilmiahan yang dipersyaratkan. Untuk itu secara khusus penulis menyampaikan apresiasi dan terima kasih secara tertulis kepada:

1. Bapak Prof. Dr.Ir. H. Andani Achmad, M.T., selaku promotor, Ibu Dr. Ir. Ingrid Nurtanio, M.T, Selaku co-promotor 1 dan ibu Prof. Dr.Eng. Intan Sari Areni, S.T., M.T., selaku co-promotor 2 yang telah meluangkan waktu, tenaga dan pikiran dalam membimbing dan memberi saran yang konstruktif sehingga disertasi ini dapat tersusun dengan baik dan terstruktur.
2. Bapak Prof. Dr. Eng. Ir. Muh. Isran Ramli, ST., MT., IPM., ASEAN Eng., Bapak Prof. Ir. Lukito Edi Nugroho, M.Sc., PhD, Bapak Dr. Ir. Amil Ahmad Ilham, ST., M.IT, Bapak Dr. Muhammad Niswar, ST., M.IT, Bapak Dr.Eng. Ady Wahyudi Paundu, S.T, M.T, dan Bapak Dr. Zulkifli Tahir, S.T.,M.Sc selaku komisi penguji yang telah berkontribusi dalam memberikan saran dan masukan untuk penyusunan disertasi ini.
3. Rektor Universitas Hasanuddin, Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc., Dekan Fakultas Teknik, Bapak Prof. Dr.Eng. Ir. Muhammad Isran Ramli, S.T., M.T., IPM., ASEAN Eng., Ketua Departemen Teknik Elektro, Ibu Dr. Eng. Ir. Dewiani, M.T., IPM, Ketua Program Studi S3 Teknik Elektro, Bapak Prof. Dr. Ir. H. Andani Achmad, M.T., dan



seluruh dosen serta staf S3 program studi Teknik Elektro yang telah memberikan dukungan dan layanan terbaik selama proses studi penulis.

4. Ketua Dewan Pembina Yayasan Perguruan Tinggi Atma Jaya Makassar, Bapak John Chanda Syarif, Ketua Dewan Pengawas Yayasan Perguruan Tinggi Atma Jaya Makassar, Bapak Dr. Drs, Yakobus K. Ditti, M.M., Ketua Pengurus Yayasan Perguruan Tinggi Atma Jaya Makassar, Ibu Lita Limpo, S.E., M.Si., Ph.D.
5. Rektor Universitas Atma Jaya Makassar, Bapak Dr. Wihalminus Sombo'layuk, S.E., M.Si, beserta wakil rektor, Dekan Fakultas Teknologi Informasi, Bapak Sean Coonery Sumarta, ST, M.Eng., Wakil Dekan Fakultas Teknologi Informasi, Ibu Elisabeth, S.Kom., M.MSI., Ketua Program Studi Informatika, Bapak Alfredo Gormantara, S.Kom, M.Kom dan seluruh kolega dosen serta tenaga kependidikan di Universitas Atma Jaya Makassar yang telah mendukung penulis selama dalam melaksanakan studi.
6. Para sahabat, Teman seangkatan S3 Elektro UNHAS dan Seluruh Mahasiswa S3 Teknik Elektro UNHAS. Terima kasih atas bantuan dan dukungannya selama ini kepada penulis.

Ucapan terima kasih yang sebesar-besarnya kepada orang tua, istri, anak dan beserta seluruh keluarga yang telah memberikan semangat dan dukungan moril kepada penulis selama menempuh studi.

Meskipun dalam penyusunan dan penulisan disertasi ini penulis telah berusaha maksimal namun tidak menutup kemungkinan masih terdapat berbagai kekurangan baik dari aspek penulisan maupun substansinya. Oleh ksarena itu saran, masukan dan kritik yang membangun sangat penulis harapkan demi perbaikan ke arah yang lebih baik lagi.

Akhir kata penulis mengucapkan terima kasih dan mohon maaf atas segala kekurangan yang ada. Semoga rahmat Tuhan Yang Maha Esa selalu menyertai kita semua.

Makassar, November 2023

Penulis



ABSTRAK

PHIE CHYAN. Model Deteksi Stres Melalui Analisis Suara Dalam Lingkungan Non-Stasioner Multi Speaker dengan Pendekatan Deep Learning (dibimbing oleh **Andani Achmad, Ingrid Nurtanio, Intan Sari Areni**)

Stres merupakan suatu kondisi psikologis yang memerlukan penanganan yang tepat karena berpotensi menimbulkan dampak jangka panjang terhadap kesehatan dan kemampuan kognitif. Hal ini relevan terutama ketika mempertimbangkan anak-anak usia pra dan awal sekolah, dimana stres dapat menimbulkan berbagai dampak buruk bagi fisik dan psikis mereka. Penggunaan suara ucapan merupakan salah satu pendekatan yang dapat digunakan untuk mendeteksi stres. Pendekatan ini memiliki kelebihan yang memungkinkan deteksi dilakukan tanpa menimbulkan ketidaknyamanan pada subjek dan tidak mensyaratkan tingkat kemampuan kognitif tertentu pada subjek yang dideteksi. Saat ini meskipun penelitian terkait deteksi stres melalui suara sudah banyak dilakukan akan tetapi umumnya akuisisi suara dilakukan dalam lingkungan stasioner yang membatasi potensi deteksi stres dilakukan dalam situasi yang riil dimana kebanyakan aktivitas sosial manusia dilakukan dalam lingkungan non stasioner dimana terdapat banyak derau dan suara pembicara yang saling tumpang tindih.

Penelitian ini bertujuan untuk mengembangkan model deteksi stres melalui analisis karakteristik suara ucapan di lingkungan non stasioner multi-speaker. Model ini menerima masukan suara dari lingkungan aktivitas, yang terdapat kebisingan dan banyak suara pembicara yang tumpang tindih. Audio yang diperoleh kemudian dipisahkan menggunakan algoritma *speech separation* berbasis *deep learning* dengan arsitektur RNN, menghasilkan keluaran sebagai suara ucapan yang tersegregasi. Suara ucapan tersebut kemudian diekstraksi untuk fitur-fiturnya dan dimasukkan ke model deteksi stres berdasarkan arsitektur CNN, yang memprediksi status stres pembicara. Hasil eksperimen menunjukkan bahwa model yang diusulkan mampu melakukan pemisahan ucapan hingga lima pembicara dan memprediksi status stres subjek dengan akurasi dan F1 score rata-rata masing-masing 95,6% dan 94,8%.

Kata Kunci : Deteksi stres, suara ucapan, *speech separation*, *deep learning*



ABSTRACT

PHIE CHYAN. Stress Detection Model Through Speech Analysis in Multi Speaker Non-Stationary Environment with Deep Learning Approach (Supervised by **Andani Achmad, Ingrid Nurtanio, Intan Sari Areni**)

Stress is a psychological condition that requires appropriate treatment because it has the potential to have long-term impacts on health and cognitive abilities. This is especially relevant when considering pre- and early school-aged children, where stress can have various negative impacts on their physical and psychological well-being. Speech sounds is one approach that can be used to detect stress. This approach has the advantage that it allows detection to be carried out without causing discomfort to the subject and does not require a certain level of cognitive ability in the subject being detected. While research on sound-based stress detection has been conducted, most sound acquisition studies are conducted in stationary environments. This limits the potential of sound-based stress detection in real-world scenarios, where most human social interactions take place in non-stationary environments with lots of noise and overlapping speaker voices.

This research aims to develop a stress detection model through speech sounds analysis in a non stationary multi-speaker environment. This model receives sound input from the activity environment, which contains noise and many overlapping speaker voices. The audio obtained is then separated using a deep learning-based speech separation algorithm with RNN architecture, producing output as segregated speech sounds. The speech sounds are then extracted for their features and fed to a stress detection model based on CNN architecture, which predicts the speaker's stress status. Experimental results show that the proposed model can perform speech separation for up to five speakers and predict the stress status of subjects with an average accuracy and F1-score of 95.6% and 94.8% respectively.

Keywords : Stress detection, Speech, speech separation, deep learning



DAFTAR ISI

Halaman

HALAMAN JUDUL	
PENGAJUAN DISERTASI	
PERSETUJUAN DISERTASI	
PERNYATAAN KEASLIAN DISERTASI	
KATA PENGANTAR.....	i
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI.....	v
DAFTAR TABEL	vii
DAFTAR GAMBAR.....	viii
DAFTAR LAMPIRAN	x
DAFTAR SINGKATAN DAN ARTI SIMBOL	xi
BAB I. PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
1.6 Ruang Lingkup Penelitian	6
BAB II. KERANGKA KONSEPTUAL PENELITIAN	8
2.1 Kerangka Konseptual	8
2.2 Hipotesis Penelitian	23
BAB III. DETEKSI STRES DENGAN PENDEKATAN BERTINGKAT MELALUI SUARA DAN ANALISIS LEKSIKAL	26
3.1 Abstrak	26
Pendahuluan	26
Metodologi	29
Hasil dan Pembahasan.....	39
Kesimpulan.....	45



BAB IV. PENDEKATAN HIBRID BERBASIS DEEP LEARNING UNTUK MODEL DETEKSI STRES MELALUI SINYAL SUARA	46
4.1 Abstrak	46
4.2 Pendahuluan	47
4.3 Metodologi	53
4.4 Hasil dan Pembahasan	62
4.5 Kesimpulan.....	66
BAB V. DETEKSI STRES ANAK MELALUI SINYAL SUARA PADA LINGKUNGAN MULTI-SPEAKER DENGAN PENDEKATAN DEEP LEARNING	68
5.1 Abstrak	68
5.2 Pendahuluan	68
5.3 Metodologi	72
5.4 Hasil dan Pembahasan	81
5.5 Kesimpulan.....	84
BAB VI. PEMBAHASAN UMUM	86
6.1 Deteksi Stres Dengan Pendekatan Bertingkat Melalui Suara dan Analisis Leksikal	88
6.2 Pendekatan Hibrid Berbasis Deep Learning untuk Model Deteksi Stres Melalui Sinyal Suara	89
6.3 Deteksi Stres Anak Melalui Sinyal Suara Pada Lingkungan Multi-Speaker dengan Pendekatan Deep Learning	91
BAB VII. KESIMPULAN DAN SARAN	94
7.1 Kesimpulan.....	94
7.2 Saran.....	94
DAFTAR PUSTAKA	96
LAMPIRAN	



DAFTAR TABEL

Nomor	Halaman
Tabel 1 State of the art penelitian terkait	14
Tabel 2 Rumusan hipotesis dan penelitian terkait.....	24
Tabel 3 Statistik dari dataset Dreddit	39
Tabel 4. Metrik Performa klasifikasi dari model deteksi stres berbasis leksikal .	43
Tabel 5. Perbandingan antara ground truth dengan prediksi model.....	44
Tabel 6. Perbandingan metrik performansi model dari tiga pendekatan	45
Tabel 7. Benchmark kinerja dari berbagai model <i>state-of-the-art</i> menggunakan dataset DAIC-WOZ.....	64
Tabel 8. Fitur audio yang digunakan dalam domain sinyal	80
Tabel 9. Skor Si-SNRi dari berbagai model yang digunakan terhadap jumlah pembicara di dalam sampel audio	82



DAFTAR GAMBAR

Nomor	Halaman
Gambar 1. Kerangka konsep penelitian	13
Gambar 2. Diagram posisi penelitian terhadap <i>state-of-the-art</i>	18
Gambar 3. Diagram fishbone penelitian.....	19
Gambar 4. Milestone dan luaran penelitian.....	21
Gambar 5. Konstruksi penelitian dan hipotesis	24
Gambar 6. Model deteksi stres multi-tahap.....	31
Gambar 7. Import <i>Library</i> yang dibutuhkan	33
Gambar 8. Prosedur relabeling dataset.....	33
Gambar 9. Posisi dataset sampel setelah proses relabeling	34
Gambar 10. Waveplot dan spectrogram dari sampel suara stress	34
Gambar 11. Waveplot dan spectrogram dari sampel suara non-stres.....	35
Gambar 12. Waveplot dari file audio asli, dengan injeksi derau dan modifikasi pitch	36
Gambar 13. Script prosedur transkripsi bahasa indonesia ke bahasa inggris dengan Openai/whisper-medium.....	37
Gambar 14. NLP pipeline untuk model deteksi stres	37
Gambar 15. Import <i>Library</i> model Word2Vec	39
Gambar 16. Load dataset Dreddit.....	39
Gambar 17. Graph yang merepresentasikan nilai loss dan akurasi dari tahap pelatihan dan pengujian.....	40
Gambar 18. Ringkasan dari parameter arsitektur CNN model	40
Gambar 19. Prosedur tokenisasi	41
Gambar 20. Prosedur <i>stemming</i> dan Lemmatization.....	41
Gambar 21. Ekstraksi fitur dengan Word2Vec	42
Gambar 22. Kemiripan kata berdasarkan <i>vector space</i> pada dataset	42
23. Bobot 20 kata teratas pada pesan.....	43
24. Diagram metode ekstraksi fitur berbasis leksikal	43
25. Beragam sensor dan penempatannya pada tubuh	50



Gambar 26. Instrumen PHQ-8 untuk asesmen kesehatan mental	54
Gambar 27. Arsitektur dari model deteksi stres hibrid	56
Gambar 28. Bentuk <i>waveplot</i> dari sampel suara ucapan, sampel asli, dengan injeksi derau buatan, dan modifikasi pitch	58
Gambar 29. Visualisasi dari contoh sampel suara dalam Mel-Spectrogram	59
Gambar 30. Script yang menampilkan model dengan arsitektur CNN	61
Gambar 31. Impor model GSOM dan parameter yang digunakan.....	61
Gambar 32. Grafik yang menunjukkan loss dari proses latih dan uji terhadap jumlah epoch	62
Gambar 33. Grafik yang menunjukkan akurasi dari proses latih dan uji terhadap jumlah epoch	63
Gambar 34. Kinerja model dengan dataset yang dibangun vs dataset DAIC-WOZ.....	64
Gambar 35. Distribusi simpul suara dalam dataset suara anak	65
Gambar 36. Distribusi simpul suara pada dataset DAIC-WOZ	66
Gambar 37. Diagram model sistem.....	72
Gambar 38. Tahap pre-processing <i>silence removal</i>	74
Gambar 39. Model <i>speech separation</i>	76
Gambar 40. Blok Multiply dan Concat (MULCAT).....	76
Gambar 41. Diagram alir dari algoritma seleksi untuk menyesuaikan model yang digunakan dengan jumlah pembicara di dalam sampel suara	77
Gambar 42. Script potongan coding untuk mendeteksi channel kosong pada <i>speech mixture</i>	78
Gambar 43. Representasi audio sebagai sinyal 3 dimensional.....	79
Gambar 44. Model deteksi stres	80
Gambar 45. Hasil evaluasi kinerja model menggunakan berbagai fitur suara pada domain sinyal	83
Gambar 46. Perbandingan kinerja model menggunakan dataset open-source dan dataset yang dibangun	84



DAFTAR LAMPIRAN

Nomor	Halaman
Lampiran 1.	103
Lampiran 2.	134



BAB I PENDAHULUAN

1.1 Latar Belakang

Menurut definisi umum, stres ditandai dengan perubahan reaksi psikologis dari keadaan tenang ke keadaan emosional. Stres menurut tinjauan ilmu psikologi dapat dibedakan menjadi *eustress* dan *distress*. *Eustress* mengarah ke keadaan emosi yang baik seperti gembira dan semangat sementara *distress* mengarah ke emosi negatif, tetapi dari kedua terminologi ini istilah stres lebih sering digunakan untuk menggambarkan kondisi *distress* yang melambangkan keadaan emosi yang negatif seperti marah, gelisah, sedih, takut, sakit dan gugup (Healey and Picard, 2005).

Teknologi pendukung keamanan dan pengawasan anak saat ini sudah banyak dikembangkan yang pada umumnya bertindak atas respon terhadap sesuatu kejadian yang terjadi pada anak, misalnya meninggalkan area pengawasan (Saranya, Venkatesh and Kumar, 2016), terjatuh (Chyan, 2021), dan menghadapi bahaya (Punjabi *et al.*, 2019). Kelemahan dari metode seperti ini adalah ketidakmampuan untuk mendeteksi potensi gangguan yang mungkin dihadapi oleh anak sejak dini yang salah satunya disebabkan oleh masalah psikologis seperti stres. Masalah stres merupakan hal yang dapat mengganggu tumbuh kembang seorang anak. Berbagai hasil penelitian yang ada menggambarkan korelasi erat antara keadaan stres dengan penurunan efektifitas kemampuan pengambilan keputusan (Wemm and Wulfert, 2017), penurunan kinerja kognitif dan motivasi (Morgado and Cerqueira, 2018) hingga penurunan *awareness* terhadap situasi lingkungan sekitar (Healey and Picard, 2005).

Pentingnya pengamatan dan deteksi dini terhadap kondisi stres ini menarik minat peneliti dari berbagai disiplin ilmu termasuk dalam bidang teknologi informasi dan komunikasi. Saat ini implementasi teknologi dalam mendeteksi dan memonitor stress adalah dengan metode yang melibatkan pengukuran langsung indikator sinyal respon tubuh (*biosignal*) dengan menggunakan beragam sensor yang disematkan pada tubuh manusia. Metode tersebut memiliki kelebihan akurasi yang diperoleh. Berbagai penelitian yang memanfaatkan metode ini memperoleh akurasi yang cukup tinggi dalam mendeteksi stress



dengan kecendrungan akurasi yang semakin meningkat bila menggunakan kombinasi dari beberapa data *biosignal*. Kombinasi data denyut jantung, resistansi kulit dan temperatur kulit merupakan parameter diskriminan yang menghasilkan akurasi paling tinggi dalam mendeteksi stres berdasarkan beberapa penelitian yang dikaji (Gedam and Paul, 2021).

Meski memiliki kelebihan dari sisi akurasi yang cukup tinggi penelitian-penelitian tersebut menggunakan data *biosignal* yang membutuhkan sensor atau instrumen yang merekam data langsung dari tubuh subjek. Perekaman data langsung dari tubuh subjek membutuhkan sensor atau instrumen tersebut harus dilekatkan atau ditempelkan pada tubuh subjek (Han, Byun and Kang, 2018). Meskipun metode seperti ini secara umum tidak berbahaya dan bukan merupakan metode invasive, tetapi penggunaan berbagai perangkat yang dilekatkan pada tubuh dapat menimbulkan ketidaknyamanan dari responden dan dapat menjadi sumber stres itu sendiri. Dari sisi efisiensi solusi pemanfaatan teknologi ini juga cukup sulit untuk diimplementasikan secara riil karena banyaknya peralatan dan sensor yang dibutuhkan serta harus disesuaikan dengan banyaknya subjek yang perlu dimonitor (Li and Liu, 2020).

Suara ucapan (*speech*) dari berbagai penelitian terhadap psikologi stres disimpulkan dapat dijadikan sebagai penanda terhadap kondisi stres yang dialami manusia karena output suara sendiri adalah respon *psikofisiologis* yang merupakan bagian dari sistem stres *psikofisiologis* integratif manusia dan reaktivitas stres adalah integrasi kompleks dari kontrol simpatis dan parasimpatis pada otak manusia. Stres psikologis menginduksi banyak efek pada tubuh, termasuk peningkatan ketegangan otot, peningkatan laju pernapasan, dan perubahan tingkat air liur, yang pada gilirannya dapat mempengaruhi produksi vokal (Van Puyvelde *et al.*, 2018; Slavich, Taylor and Picard, 2019). Di bawah tekanan psikologis, *voice pitch* (korelasi akustik dari frekuensi dasar) umumnya meningkat bersamaan dengan peningkatan tekanan *subglotal* dan intensitas vocal (Pisanski and Sorokowski, 2021). Peningkatan *voice pitch* dan beberapa parameter suara lainnya

intonasi dan *speech prosody* terjadi dalam kondisi stres psikologis fitur suara menjadi *bio marker* yang baik dalam mendeteksi level stres. tahun terakhir ini, penelitian terkait dengan pemanfaatan suara sebagai



biomarker dalam mendeteksi stres sudah mulai banyak dilakukan dan umumnya penelitian tersebut dilakukan dalam lingkungan terkontrol dengan derau latar belakang yang dapat diabaikan atau berada dalam lingkungan stasioner dimana jenis frekuensi derau bersifat konstan dalam periode waktu tertentu (Han, Byun and Kang, 2018). Saat ini berbagai penelitian yang memanfaatkan properti sinyal suara untuk deteksi stres secara rata-rata belum memiliki akurasi setinggi dari model deteksi stres yang berbasis parameter biosignal. Salah satu penyebabnya adalah penelitian deteksi stres dengan memanfaatkan sinyal suara relatif lebih baru dibandingkan dengan pendekatan berbasis biosignal yang lebih matang karena telah menjadi topik penelitian dalam jangka waktu yang lama. Meskipun demikian disisi lain deteksi stress melalui suara sebagai biomarkernya menjadi peluang penelitian yang potensial bagi peneliti, antara lain dalam hal untuk mencapai akurasi deteksi sebaik model deteksi stres yang umum digunakan.

Model deteksi stres pada anak merupakan salah satu solusi dalam teknologi pendukung keamanan dan pengawasan anak yang menjadi urgensi dari penelitian ini. Berdasarkan dampak yang telah dibahas sebelumnya terhadap kondisi fisik dan mental anak, stres merupakan suatu kondisi yang harus dapat dideteksi dan ditangani sejak dini apabila terjadi pada seorang anak. Kesulitan dalam diagnosa stres pada anak adalah terkait dengan kemampuan verbal mereka yang masih terbatas untuk dapat berkomunikasi secara efektif terhadap permasalahan yang mengganggu mereka kepada orang tua atau pengasuh mereka (Choi *et al.*, 2017; Kim, Měsíček and Kim, 2021), terlebih lagi di usia mereka seorang anak umumnya belum *aware* terhadap kondisi psikologis yang mereka alami dan dampak negatif dari permasalahan psikologis ini bila tidak ditangani akan menyebabkan masalah stres yang kronis seperti depresi yang akan sangat mempengaruhi tumbuh kembang seorang anak. Dari sudut pandang lain ditinjau dari sisi pengasuh dalam lingkungan dimana terdapat banyak anak yang harus diawasi misalnya dalam skenario sekolah PAUD, taman bermain, jasa *daycare* dimana rasio jumlah pengawas tidak sebanding dengan jumlah anak, kegiatan



san terhadap setiap individu anak menjadi hal yang sulit untuk dilakukan. Hal ini harus menerus terlebih lagi bila menyangkut kondisi psikologis yang tidak dapat dikenali secara visual.

Stasioneritas adalah cara untuk menggambarkan karakteristik proses pembangkitan sinyal, yang selanjutnya dibedakan kedalam stationer dan non-stationer. Perbedaan utama antara sinyal stationer dan sinyal non-stationer adalah bahwa sifat properti sinyal yang diproses stationer tidak berubah terhadap waktu sedangkan sinyal non-stationer sifat propertinya tidak konsisten terhadap waktu. Lingkungan belajar dan bermain anak pada umumnya merupakan lingkungan non-stasioner yang terdapat banyak sumber suara dengan beragam frekuensi yang dinamis. Suara dari banyak orang yang berbicara kemudian suara *ambient* / latar belakang yang dikategorikan sebagai derau seperti suara instrumen, mesin dan sebagainya. Kebutuhan monitoring membutuhkan model deteksi stres harus dapat mengakuisisi suara secara langsung dari lingkungan non stasioner tersebut. Tantangan dalam melakukan pemodelan dan deteksi stres melalui analisis suara pada lingkungan non stasioner semacam ini membutuhkan pendekatan terkini dalam bidang *audio / speech separation*, sampai saat ini penelitian di bidang ini masih sedang berkembang, permasalahan yang juga dikenal dengan istilah *cocktail party problem* merupakan kemampuan untuk melacak (*tracing*) dan merekognisi suara dari subjek spesifik ketika banyak subjek yang berbicara bersamaan khususnya dalam lingkungan dimana terdapat suara latar belakang (derau) (Qian *et al.*, 2018). Permasalahan ini walaupun mudah bagi manusia karena indera manusia mampu untuk memisahkan sinyal yang berasal dari banyak sumber dan fokus untuk merekognisi dan menjejaki satu sumber tertentu tapi bagi komputer merupakan tugas yang menantang karena kesulitan dalam implementasinya (Luo and Mesgarani, 2018; Wang and Chen, 2022), diperlukan pendekatan berbasis kecerdasan buatan untuk mampu memisahkan input berupa derau, distorsi, dan suara yang saling *overlap* untuk dapat melakukan analisis suara secara optimal. Dengan demikian, *speech separation* merupakan tugas yang mendasar sebelum masuk ke tahap analisis sinyal suara dan hal ini juga ditargetkan untuk menjadi salah satu kontribusi ilmiah dari penelitian ini. Berdasarkan permasalahan yang telah dikemukakan, peneliti mengangkat

ini dengan judul “*Model Deteksi Stres Melalui Analisis Suara dalam an Non-Stasioner Multi Speaker dengan Pendekatan Deep Learning*”



1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah dari penelitian ini adalah:

1. Bagaimana memodelkan karakteristik suara untuk deteksi stres pada lingkungan non stasioner multi-speaker?
2. Bagaimana melakukan separasi dan segregasi suara dari audio kanal tunggal untuk mendukung model deteksi stress dalam lingkungan non stasioner *multi-speaker* ?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah:

1. Menghasilkan model deteksi stres melalui suara yang mampu mendukung deteksi stres pada lingkungan non-stasioner *multi-speaker* dengan akurasi yang ditingkatkan melalui berbagai aspek properti suara.
2. Melakukan modifikasi metode separasi dan segregasi suara dari *audio* kanal tunggal dalam lingkungan non stasioner untuk mendukung model deteksi stres *multi-speaker*.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah:

1. Membantu pengasuh atau pengawas anak dalam memperoleh gambaran terhadap kondisi psikologis anak.
2. Memberi kontribusi dalam solusi separasi dan rekognisi suara dari audio kanal tunggal dalam lingkungan non stasioner.
3. Memberi kontribusi pada bidang penelitian terkait deteksi dan monitoring stres melalui analisis suara.

1.5 Batasan Masalah

Sesuai dengan fokus penelitian yang dilakukan maka batasan masalah

ini adalah:

Dataset suara yang digunakan dalam penelitian ini berasal dari subjek anak-anak usia pra dan awal sekolah (5-7 tahun) yang disusun untuk



mendukung model yang dibangun. Selain menggunakan dataset suara anak, penelitian juga memanfaatkan beberapa *open-source* dataset dengan tujuan membandingkan kinerja model dengan berbagai model *state-of-the-art*.

2. Penelitian ini berfokus pada model deteksi stres. Pencegahan dan mitigasi yang diperlukan terkait kondisi stres yang terjadi berada di luar ruang lingkup dari penelitian ini.

1.6 Ruang Lingkup Penelitian

Berdasarkan hasil studi literatur yang dilakukan terkait dengan berbagai penelitian yang berfokus pada deteksi stres menggunakan pemanfaatan teknologi dan kecerdasan buatan diperoleh kesimpulan bahwa saat ini berbagai pendekatan telah dimanfaatkan dalam upaya mendukung deteksi stres secara akurat terhadap manusia. Dari berbagai pendekatan tersebut, metode pengukuran *biosignal* paling banyak dilakukan untuk deteksi stres. Selain metode arus utama tersebut, berbagai penelitian terbaru juga berusaha mengeksplorasi berbagai pendekatan lain yang dapat digunakan dalam deteksi stress seperti melalui wajah, ekspresi, suara dan sebagainya. Penelitian terkait deteksi stres melalui suara saat ini masih dalam tahap awal sehingga masih memiliki ruang yang luas untuk dieksplorasi lebih jauh terutama dengan ketersediaan dukungan dari metode kecerdasan buatan seperti *machine learning* dan *deep learning*.

Ruang lingkup dari penelitian ini adalah pada pemanfaatan berbagai fitur suara dalam mendeteksi stres pada manusia. Penggunaan suara sebagai *biomarker* dalam deteksi stres memiliki banyak kelebihan dibanding pendekatan lain, diantaranya adalah suara dapat dengan mudah diakuisisi dan tidak bersifat invasif dalam proses akuisisinya sehingga potensi intervensi terhadap subjek juga akan minimal dan tidak berpengaruh terhadap kondisi stres yang mungkin dialaminya. Metode ini sangat cocok untuk diaplikasikan dalam kelompok demografi tertentu seperti misalnya anak-anak. Seperti yang telah dibahas bahwa stres pada anak-



rupakan suatu permasalahan yang kompleks dan membutuhkan an khusus yang berbeda dibandingkan pada orang dewasa. Kemampuan asi dan kognitif yang masih terbatas pada anak-anak secara umum

membuat mereka sulit untuk mengenali masalah psikologis yang terjadi pada diri mereka, termasuk juga oleh orang tua atau pengasuh mereka. Berdasarkan berbagai masalah dan juga potensi yang dihadapi maka penelitian ini berfokus dalam mengeksplorasi metode deteksi stres dengan pemanfaatan *deep learning* dengan fokus untuk mendeteksi stres pada subjek anak-anak.

Novelty atau kebaruan yang diharapkan dalam penelitian ini sehingga memberikan kontribusi bagi ilmu pengetahuan dan teknologi khususnya terkait dengan pemanfaatan teknologi dan kecerdasan buatan dalam mendukung kesehatan mental sebagai berikut:

1. Model deteksi stres melalui pemanfaatan fitur sinyal dan aspek leksikal dari suara:

Kontribusi utama yang ditunjukkan dalam penelitian ini adalah model deteksi stress melalui suara dengan pendekatan yang komprehensif dalam memanfaatkan suara dari sisi properti sinyal dan dari sisi aspek linguistiknya untuk mendeteksi stres dari subjek melalui karakteristik vokal dan makna kontekstual dari ucapannya.

2. Metode separasi dan segregasi suara dalam mendukung model deteksi stres :

Salah satu kebaruan yang diklaim dari penelitian ini adalah modifikasi metode separasi dan segregasi suara yang dihasilkan untuk mendukung model deteksi stres dalam lingkungan multi-speaker. Metode ini memungkinkan suara diakuisisi langsung dari lingkungan non-stasioner dimana terdapat suara yang saling *overlap* satu sama lain.



BAB II

KERANGKA KONSEPTUAL DAN HIPOTESIS PENELITIAN

2.1 Kerangka Konseptual

Permasalahan yang diangkat dalam penelitian adalah keadaan stress yang merupakan salah satu masalah mental yang dihadapi oleh banyak orang tanpa memandang usia. Menurut (Kopin, Eisenhofer and Goldstein, 1988; Healey and Picard, 2005) stres dapat dibagi menjadi dua konsep stres: "eustres" yang merupakan stres positif dan "distress" yang merupakan stres negatif. Selain itu menurut jenis dan tingkat keparahannya stres dapat dibedakan menjadi tiga kategori: stres akut, stres episodik dan stres kronis. Diskriminasi ini didasarkan pada waktu paparan stresor. Stres akut adalah respons bawaan individu ketika terpapar stresor dalam waktu singkat. Biasanya, stres ini tidak membawa konsekuensi negatif. Stres episodik biasanya ditemukan ketika seorang individu mengalami kehidupan yang sangat menegangkan. Stres ini muncul ketika situasi stres sering terjadi tetapi berhenti dari waktu ke waktu. Stres kronis adalah stres yang muncul terus-menerus. Stres ini ditemukan ketika individu menghadapi stresor yang berasal dari keluarga atau lingkungan kerja. Biasanya, bagi seorang individu, stres kronis dianggap cukup berbahaya. Paparan stres yang berkepanjangan dapat menghasilkan peningkatan tingkat stres kronis yang menyebabkan penurunan pada fungsi imunitas dan berbagai sistem fisiologis. Selain itu stres juga mempengaruhi fungsi kognitif dan dapat membawa berbagai efek negatif yang menurunkan fungsi kognitif (Mather and Lighthall, 2012; Morgado and Cerqueira, 2018).

Secara medis, kondisi stress akan mempengaruhi sistem saraf otonom yaitu suatu sistem regulasi komprehensif yang mengontrol beragam fungsi tubuh seperti detak jantung, pernafasan, air liur dan keringat sehingga kondisi stres umumnya dapat ditandai dengan meningkatnya denyut jantung dan tekanan darah,

keringat yang lebih aktif, perubahan aktifitas metabolik yang pada orang dapat menyebabkan sering berkemih, diare, mual dan masalah lainnya (Bucci *et al.*, 2016). Peningkatan kadar cortisol, dikenal sebagai



hormon stres, yang mendadak dalam waktu singkat pada tubuh juga dapat digunakan sebagai penanda kondisi stres yang terjadi pada seseorang (Bucci *et al.*, 2016; Wemm and Wulfert, 2017; Pisanski and Sorokowski, 2021).

Pencetus stress (*stressor*) bertindak sebagai stimulus untuk menimbulkan respon fisiologis, psikologis, dan perilaku yang kompleks dari seorang individu. Respons ini dapat menyebabkan berbagai perubahan emosi, kognisi dan perilaku. Saat ini pengelompokan *stressor* masih menjadi pertanyaan peneliti dalam hal apakah otak manusia itu sendiri memang mengelompokkan jenis stres berdasar *stressornya* dan memberikan respon stres yang spesifik terhadap masing-masing *stressor* yang berbeda. Saat ini belum ada konsensus yang dapat diterima secara umum mengenai hal ini. Biasanya, peneliti yang mendukung teori adanya pengelompokan stres berargumentasi bahwa pengelompokan stres bergantung pada diskriminasi otak, ada dua jenis stres. Jenis pertama adalah yang menyebabkan gangguan aktual dari status fisiologis, seperti perdarahan, tantangan kekebalan atau infeksi. Stresor ini biasanya disebut stres fisik atau sistemik. Jenis kedua adalah yang mempengaruhi keadaan mental individu dalam menghadapi situasi saat ini atau yang akan datang, seperti konflik sosial, keributan dan stimulus lingkungan yang tidak memuaskan. Stresor ini biasanya disebut stres psikologis atau emosional (Wemm and Wulfert, 2017).

Dalam hal metode deteksi stres, secara konvensional, untuk menilai keadaan stres individu dilakukan melalui assesment penilaian psikologis oleh seorang psikolog. Subjek diminta untuk mengisi kuesioner standar (misalnya, *Perceived Stress Questionnaire*). Dengan menganalisis kuesioner, tingkat stres dinilai pada beberapa skala, sehingga tingkat stres individu dapat diukur (Vallejo *et al.*, 2018). Namun, strategi penilaian subjektif ini memerlukan intervensi manusia, misalnya secara manual menafsirkan pola perilaku dan keadaan afektif yang terkait. Selain itu, untuk dapat memahami dan menjawab pertanyaan asesmen dengan tepat dibutuhkan tingkat kecakapan kognitif tertentu serta sikap terbuka dan jujur yang biasanya pada orang yang mengalami stress terdapat kecendrungan penyangkalan

situasi yang dialami. Untungnya, para peneliti telah menemukan bahwa mengekspresikan keadaan afektif melalui banyak cara. Oleh karena itu, tubuh telah dipelajari untuk menemukan sumber terpercaya untuk



memahami keadaan afektif. Ekspresi tubuh seperti ekspresi wajah, suara, dan respons fisiologis adalah ekspresi tubuh yang diselidiki secara luas untuk mengukur keadaan afektif, misalnya stres (Slavich, Taylor and Picard, 2019; Gedam and Paul, 2021).

Suara manusia membawa banyak informasi, yang dapat dicirikan oleh dua bagian utama. Bagian pertama adalah informasi linguistik dimana ujaran-ujaran yang diucapkan terstruktur menurut kaidah-kaidah pengucapan bahasa yang digunakan. Bagian kedua adalah informasi non-linguistik yang juga disebut informasi paralinguistik. Informasi nonlinguistik meliputi intonasi, kualitas suara, *prosodi*, ritme, dan jeda bicara. Mirip dengan modalitas non-verbal lainnya seperti ekspresi wajah, tatapan mata, kedipan mata dan pelebaran pupil, isi pidato non-verbal ini selalu dapat mengungkapkan pesan subjek, misalnya, keadaan afektifnya. Analisis stres melalui suara berawal dari konsep bahwa ketika seseorang berada di bawah stres, terutama jika seseorang berada situasi yang menegangkan, detak jantungnya meningkat seiring detak jantung yang semakin cepat untuk mengirim lebih banyak darah ke otot. Darah dialihkan ke otot dan mempersiapkan mereka dalam situasi untuk melawan atau melarikan diri (*Fight or flee*). Hal ini meningkatkan getaran otot, yang disebut tremor otot mikro (MMT). Otot-otot yang membentuk saluran vokal dapat mengirimkan getaran melalui ucapan yang mempengaruhi karakteristik vokal dari seseorang yang mengalami situasi tersebut. Berbagai karakteristik dari vokal tersebut dapat dianalisis dalam berbagai parameter seperti volume, pitch, frekuensi, jitter, energi dan sebagainya.

Speech separation atau separasi suara adalah termasuk area penelitian yang masih berkembang yang mencoba memecahkan masalah yang lebih dikenal dengan nama “*Cocktail Party Problem*” dari sinyal audio, yaitu, untuk memisahkan sinyal individu yang direkam menggunakan mikrofon tunggal atau ganda. Pemisahan ucapan dari kebisingan latar belakang dapat disebut sebagai *speech enhancement* atau *speech to background separation*, sedangkan pemisahan



dari sinyal suara percakapan lain dapat disebut sebagai *speech on*, *speech to speech separation* atau cukup disebut sebagai *speech n* (Huang *et al.*, 2022). Seiring dengan kemajuan perkembangan metode

deep learning, *Speech Separation* juga mengalami perkembangan yang pesat. Kebanyakan pendekatan terkini dari *framework speech separation* adalah berbasis *supervised learning* dan dapat dikategorikan ke dalam metode *frequency domain* dan *time domain*. Metode *frequency domain* mengestimasi *time-frequency* (T-F) *mask* untuk setiap sumber berdasarkan fitur *Short Time Fourier Transform* (STFT) dan merekonstruksi sumber individual menggunakan *Inverse Short Time Fourier Transform* (iSTFT). Beberapa diantaranya adalah *deep clustering* (John R. Hershey *et al.*, 2016), *uPIT*, *Deep Attractor Network* (Chen, Luo and Mesgarani, 2017) dan sebagainya. Metode *time domain* mengambil gabungan *waveform* sebagai input dan secara langsung memprediksi *waveform* dari sumber lain menggunakan arsitektur *encoder-decoder*. Metode yang lebih baru untuk pemisahan *speech-background* adalah menggunakan teknik pembelajaran untuk mempelajari karakteristik ucapan dan kebisingan dari satu set data pelatihan dan menggunakan informasi ini pada waktu pengujian. Salah satu teknik paling awal yang memanfaatkan data pelatihan adalah *Nonnegative Matrix Factorization* (NMF). Dalam NMF, dapat dilakukan pelatihan kamus secara terpisah dari data suara dan derau dan menggabungkannya untuk membentuk kamus gabungan yang digunakan pada waktu pengujian. Setelah melakukan faktorisasi matriks pada data uji, dapat diperoleh estimasi dari setiap sumber.

Deep Learning adalah algoritma pembelajaran mesin yang menggunakan banyak lapisan untuk secara progresif mengekstrak fitur tingkat yang lebih tinggi dari raw input. Misalnya, dalam pemrosesan gambar, lapisan bawah dapat mengidentifikasi tepi, sedangkan lapisan yang lebih tinggi dapat mengidentifikasi konsep yang relevan dengan manusia seperti angka atau huruf atau wajah. Sebagian besar model *deep learning* modern didasarkan pada jaringan saraf tiruan, khususnya *Convolutional Neural Network* (CNN), meskipun model tersebut juga dapat menyertakan rumus proposisional atau variabel laten yang diatur secara berlapis dalam model generatif yang dalam seperti node dalam *deep believe network* dan mesin *Boltzmann*. Kata "dalam" pada "*deep learning*"

pada jumlah lapisan yang melaluinya data ditransformasikan. Lebih sistem *deep learning* memiliki kedalaman *Credit Assignment Path* yang substansial. CAP adalah rantai transformasi dari input ke output.



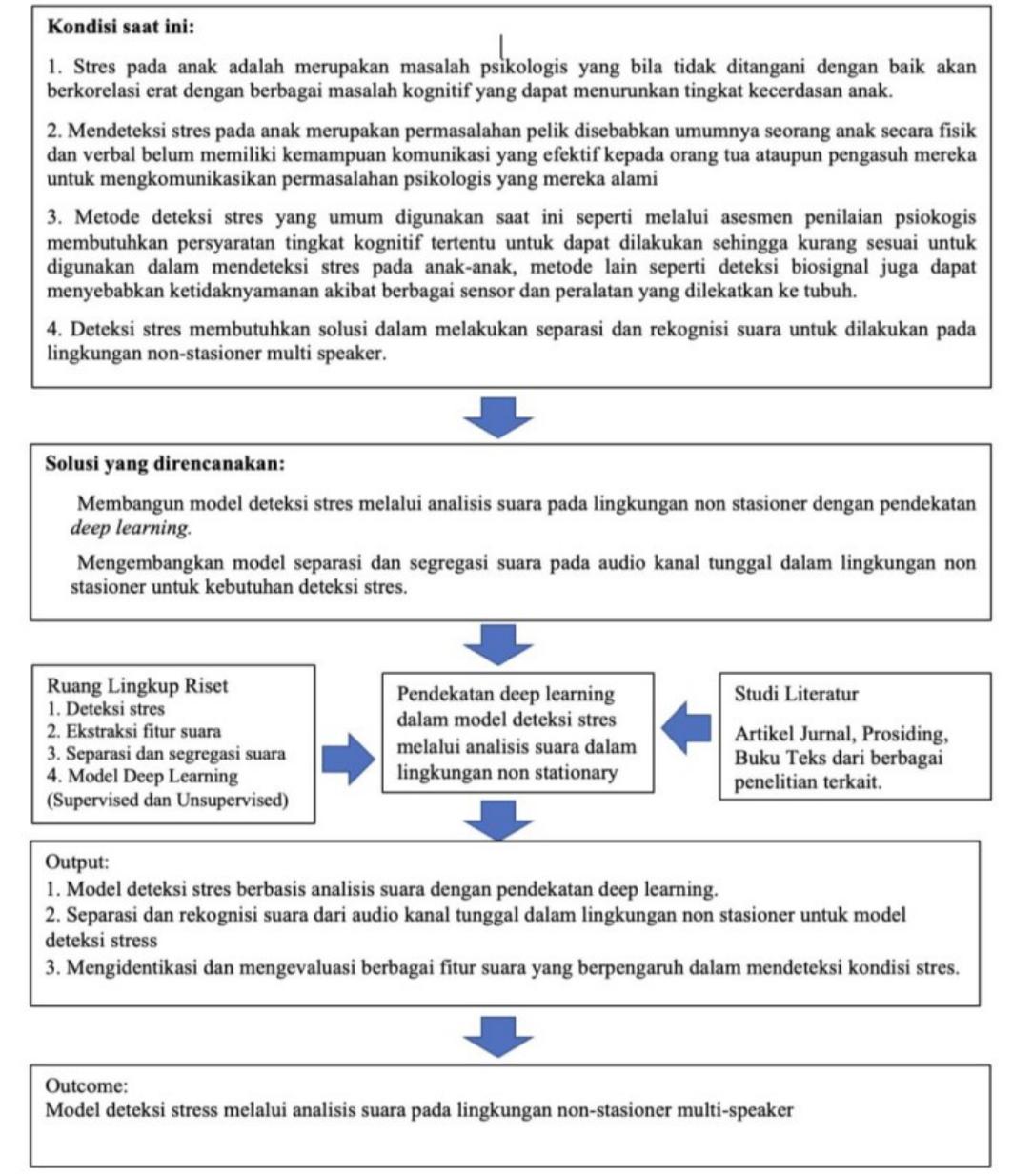
CAP menggambarkan hubungan kausal yang berpotensi antara input dan output. Untuk jaringan saraf umpan maju, kedalaman CAP adalah kedalaman jaringan dan merupakan jumlah lapisan tersembunyi ditambah satu (karena lapisan keluaran juga diparameterisasi). Untuk jaringan saraf berulang, di mana sinyal dapat merambat melalui lapisan lebih dari sekali, kedalaman CAP berpotensi tidak terbatas. Tidak ada ambang batas kedalaman yang disepakati secara universal yang memisahkan pembelajaran dangkal dari *deep learning*, tetapi sebagian besar peneliti setuju bahwa *deep learning* melibatkan kedalaman CAP yang lebih tinggi dari 2. CAP kedalaman 2 telah terbukti menjadi pendekatan universal dalam arti bahwa ia dapat meniru fungsi apa pun. Di luar itu, lebih banyak lapisan tidak menambah kemampuan fungsi aproksimator jaringan. Model dalam ($CAP > 2$) mampu mengekstrak fitur yang lebih baik daripada model dangkal dan karenanya, lapisan tambahan membantu mempelajari fitur secara efektif.

Sinyal merupakan kuantitas terukur dalam sistem fisik yang dihasilkan didalam sistem. Sinyal dapat direpresentasikan dalam bentuk grafik yang diperoleh secara digital. Pengelompokan sinyal dapat dilakukan dalam berbagai perspektif, misalnya dalam hal kontinuitas sinyal dapat dibedakan kedalam bentuk kontinyu dan diskrit, Perioditas dimana sinyal dapat dibedakan kedalam bentuk periodik atau aperiodik, probabilitas dimana sinyal dapat diklasikasi dalam bentuk deterministik atau random, dan dalam hal sifat stationeritas dimana sinyal dapat diklasifikasi kedalam bentuk stationer atau non-stationer (Wang and Chen, 2022). Sinyal juga dapat diperlakukan sebagai pengamatan atau catatan dari suatu kejadian. Stasioneritas adalah cara untuk menggambarkan karakteristik proses pembangkitan sinyal, yang selanjutnya dibedakan kedalam stationer dan non-stationer. Perbedaan utama antara sinyal stationer dan sinyal non-stationer adalah bahwa sifat properti sinyal yang diproses stationer tidak berubah terhadap waktu sedangkan sinyal non-stationer sifat propertinya tidak konsisten terhadap waktu (Défossez, Synnaeve and Adi, 2020). Suara percakapan dapat dianggap sebagai bentuk sinyal non-stasioner, selain itu gelombang sinus *multitone* dalam frekuensi

juga masuk dalam golongan sinyal non-stasioner.



Secara ringkas rangkuman berbagai teori yang mendasari penelitian ini dan telah dibahas pada paragraf-paragraf sebelumnya disajikan dalam sebuah diagram kerangka konsep penelitian yang ditunjukkan pada gambar 1.



Gambar 1 Kerangka konsep penelitian



ate of the art

mendukung penelitian dalam memberikan gambaran tentang angan terkini dalam topik penelitian yang dilakukan, maka

dilakukan studi untuk melihat berbagai penelitian relevan yang telah dilaksanakan oleh berbagai peneliti terdahulu. Berikut pada tabel 1 menyajikan penelitian-penelitian terkait yang relevan dengan penelitian yang dilakukan.

Tabel 1 *State of the art* penelitian terkait

No	Judul	Penulis	Metode/Model
1	Stress Sensor Prototype : Determining the Stress Level in using a Computer through Validated Self-Made Heart Rate (HR) and Galvanic Skin Response (GSR) sensor and Fuzzy Logic Algorithm	Cantara A., & Ceniza A., (2016)	Deteksi Stres dengan biosignal menggunakan parameter HR dan GSR dengan menggunakan algoritma fuzzy logic
2	A biological signal-based stress monitoring framework for children using wearable devices	Choi, Y., Jeon, Y Mi, Wang, L. Kim, K (2017)	Deteksi stress menggunakan kombinasi berbagai parameter biosignal dengan algoritma pembelajaran mesin Decision Tree (DT), Naive Bayes (NB) dan Support Vector Machine (SVM)
3	Effective stress detection using physiological parameters	Chauhan, M., Vora, S.V., Dabhi, D. (2017)	Penelitian ini menganalisis stress mental menggunakan Electrocardiograph (ECG) dalam berbagai posisi dan kondisi mood yang berbeda-beda pada subjek dengan menggunakan klasifikasi Artificial Neural Network (ANN) , SVM, NB dan DT
4	Embedded Low-Power Processor for Personalized Stress Detection	Attaran, N., Puranik, A. Brooks, J. Mohsenin, T. (2018)	Penelitian ini menghasilkan sebuah perangkat wearable untuk membaca berbagai data biosignal dan dengan dukungan klasifikasi pembelajaran mesin SVM dan K-Nearest Neighbors (KNN) untuk mendeteksi stress
	Detection of Stress using Biosensors	Singh, SA, Kumar G., Praveen, R.M	Penelitian ini menggunakan biosensor, spectrophotometer dan light resistance detector menggunakan salivary amylase assay



No	Judul	Penulis	Metode/Model
		Janumala, T (2018)	
6	Detecting moments of stress from measurements of wearable physiological sensors	Kyriakou, K, et al (2019)	Rule based algorithm berdasarkan data GSR dan temperatur kulit
7	Stress detection using deep neural networks	Li, R., Liu, Z (2020)	Dua Deep neural network dengan 1-dimensional CNN dan multilayer perceptron neural network
8	Personal Stress-Level Clustering and Decision-Level Smoothing to Enhance the Performance of Ambulatory Stress Detection with Smartwatches	Can, YS, et al. (2020)	Pendekatan hybrid untuk menentukan level stres personal melalui self report berbasis kuisioner didukung dengan parameter biologis yang diperoleh dari perangkat smartwarch dengan dukungan algoritma pembelajaran mesin
9	Modeling of Child Stress-State Identification Based on Biometric Information in Mobile Environment	Kim,TY, Měsíček, L Kim,SH (2021)	Identifikasi stress pada anak menggunakan pembelajaran mesin dengan metode SVM dan DT
10	Technology to detect levels of stress based on voice information	Matsuo, N, Hayakawa, S Harada, S (2015)	Deteksi pitch dan intensitas suara untuk keyword tertentu
11	Stress Detection Through Speech Analysis	Tomba, K, et al (2018)	Menggunakan pembelajaran mesin untuk mendeteksi stres melalui suara menggunakan mean energy, mean intensity and Mel-Frequency Cepstral Coefficients (MFCCs) sebagai fitur klasifikasi
12	A deep learning-based stress detection algorithm with speech signal	Han, H, Byun, K, Kang, HG (2018)	Algoritma yang diusulkan pertama mengekstrak mel-filterbank coefficients menggunakan data suara preprocessed dan kemudian memprediksi stres output dalam klasifikasi binary (Stres dan non-stres) menggunakan LSTM dan feed-forward network
	Stress detection analysis: A new	Van Puyvelde,	Studi literatur dan analisis akustik



No	Judul	Penulis	Metode/Model
	framework for voice and effort in human performance	M.,et al (2018)	
14	Infant cry language analysis and recognition: An experimental approach	Liu, L., et al (2019)	Mengklasifikasikan suara tangisan bayi berdasarkan arti tangisan menggunakan fitur yang diekstraksi dari audio fitur space antara lain Linear Predictive Coding (LPC), Linear Predictive Cepstral Coefficient (LPCC), Bark Frequency Cepstral Coefficient (BFCC) dan Mel Frequency Cepstral Coefficient (MFCC)
15	An Appraisal on Speech and Emotion Recognition Technologies based on Machine Learning	Jason, C, Kumar, S (2020)	Membandingkan berbagai algoritma pembelajaran mesin dan metode ekstraksi suara LPC dan MFCC
16	Defining and distinguishing infant behavioral states using acoustic cry analysis: is colic painful?	Parga J, Lewin, S, et al (2020)	Melakukan asesment 1000 data tangisan dengan aplikasi mobile. Melatih algoritma penerjemah tangisan dengan mengevaluasi > 6000 fitur akustik
17	A Deep Learning Approach for Work Related Stress Detection from Audio Streams in Cyber Physical Environments	Madhavi I., Chamishka, R., et al (2021)	Metode deep learning berbasis Convolutional Neural Network (CNN) dan Growing Self-Organizing Map (GSOM) dengan teknik sampling dan speaker diarization untuk mereduksi derau
18	Measuring Stress in Health Professionals Over the Phone Using Automatic Speech Analysis During the COVID-19 Pandemic: Observational Pilot Study	Konig A, Riviere K., et al (2021)	Menganalisis sampel suara dari 89 petugas medis profesional pada sambungan telepon terhadap laporan yang bersifat positif, negatif dan netral; Berbagai fitur suara diektraksi dan dikomparasi dengan metode amnesis stres klasik melalui kuisiner standar menggunakan analisis regresi
	-time Stress Detection Model Voice Analysis: Integrated VR-based Game for	Dillon R, Teoh AN, Arushi (2021)	Permainan berbasis Virtual Reality (VR) yang dikolaborasikan dengan algoritma klasifikasi untuk mendeteksi stres secara real time dengan mendeteksi suara pembicara



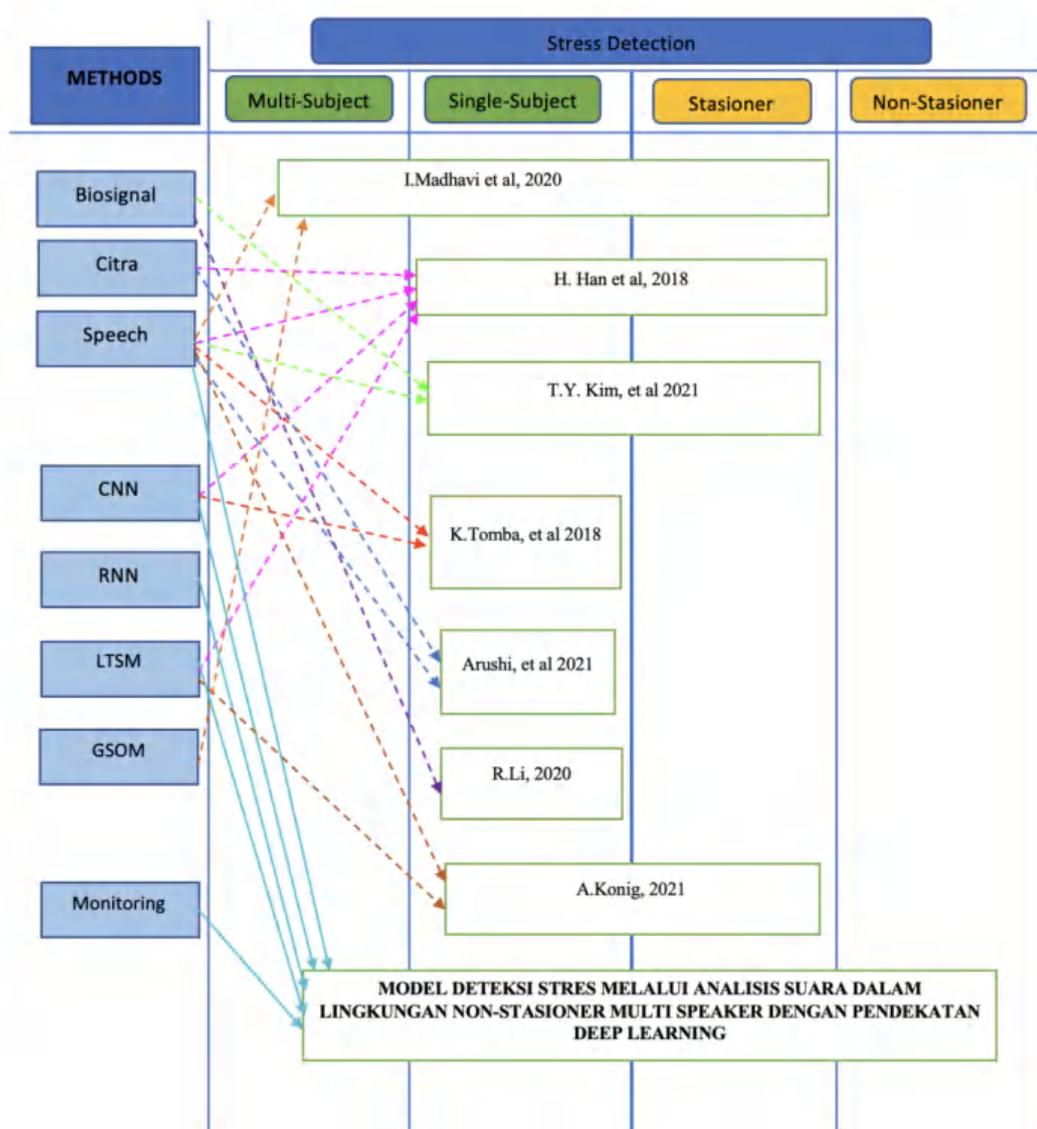
No	Judul	Penulis	Metode/Model
	Training Public Speaking Skills		
20	Human Stress Detection: Cortisol Levels in Stressed Speakers Predict Voice-Based Judgments of Stress	Pisansky K (2021)	Merancang suatu metode untuk mengukur stres berdasarkan suara dengan melakukan analisa akustik dan membandingkan dengan kadar hormon cortisolnya
21	Supervised Speech Separation Based on Deep	Wang DL, Chen J (2017)	Mengkaji berbagai pendekatan separasi suara berbasis supervised learning
22	Deep clustering: Discriminative embeddings for segmentation and separation	Hershey J, Chen Z., et al. (2016)	Deep clustering untuk pemisahan suara dalam permasalahan “cocktail party problem” dimana terdapat jumlah suara dari sumber berbeda yang tidak ditentukan jumlahnya
23	TasNet: Time-domain audio separation network for real-time, single-channel speech separation	Luo Y., Mesgarani N. (2018)	Time-domain Audio Separation Network (TasNet) untuk separasi suara secara real time dan latency rendah
24	Dual-Path RNN: Efficient Long Sequence Modeling for Time-Domain Single-Channel Speech Separation	Luo Y, Chen Z., Yoshioka T (2020)	Dual-path recurrent neural network (DPRNN) yang membagi sekuens yang panjang kedalam potongan-potongan kecil.
25	Implementation of Real-Time Speech Separation Model Using Time-Domain Audio Separation Network (TasNet) and Dual-Path Recurrent Neural Network (DPRNN)	Wijayakusuma A., Gozali D., Widjaja A. (2021)	Implementasi TasNet dan DPRNN untuk pemisahan suara pada multi talker secara real time dalam time domain

2.1.2 Posisi penelitian



Dalam penelitian-penelitian terdahulu terkait topik deteksi stres, data *biosignal* lebih banyak dimanfaatkan sebagai indikator dalam stres selain itu beberapa penelitian juga memanfaatkan indikator lain seperti ekspresi wajah dan suara. Penelitian-penelitian

terdahulu yang melakukan deteksi stres melalui suara umumnya melakukan akuisisi suara untuk single subjek/*speaker* dalam lingkungan stasioner. Posisi penelitian yang dilakukan ini memiliki kebaruan dalam hal deteksi stres dalam lingkungan non-stasioner dengan kemampuan deteksi multi subjek/*speaker* melalui model separasi dan segregasi suara. Gambar 2 menunjukkan posisi penelitian terhadap *state of the art*.

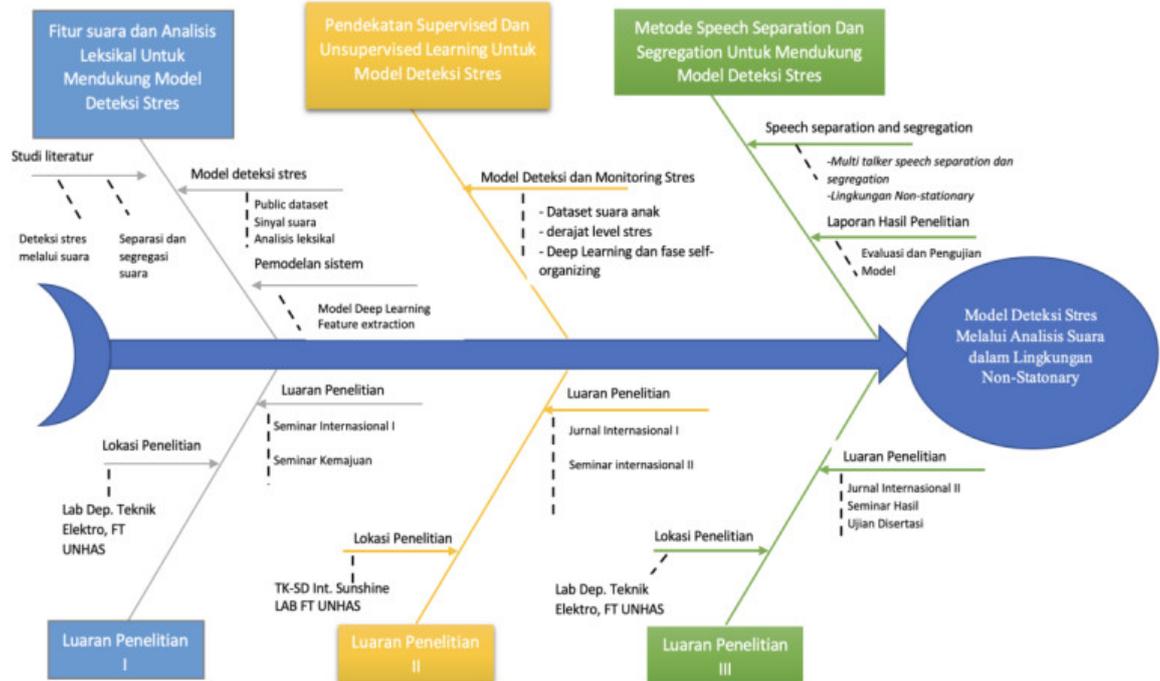


Gambar 2 Diagram posisi penelitian terhadap state of the art



2.1.3 Tahapan penelitian

Dalam menghasilkan luaran penelitian yang diharapkan, maka dilakukan perencanaan dalam bentuk tahapan penelitian sebagaimana yang ditunjukkan pada Gambar 3 beserta target capaian dan luaran penting dalam penelitian yang disajikan dalam milestone penelitian pada Gambar 4



Gambar 3 Diagram fishbone penelitian

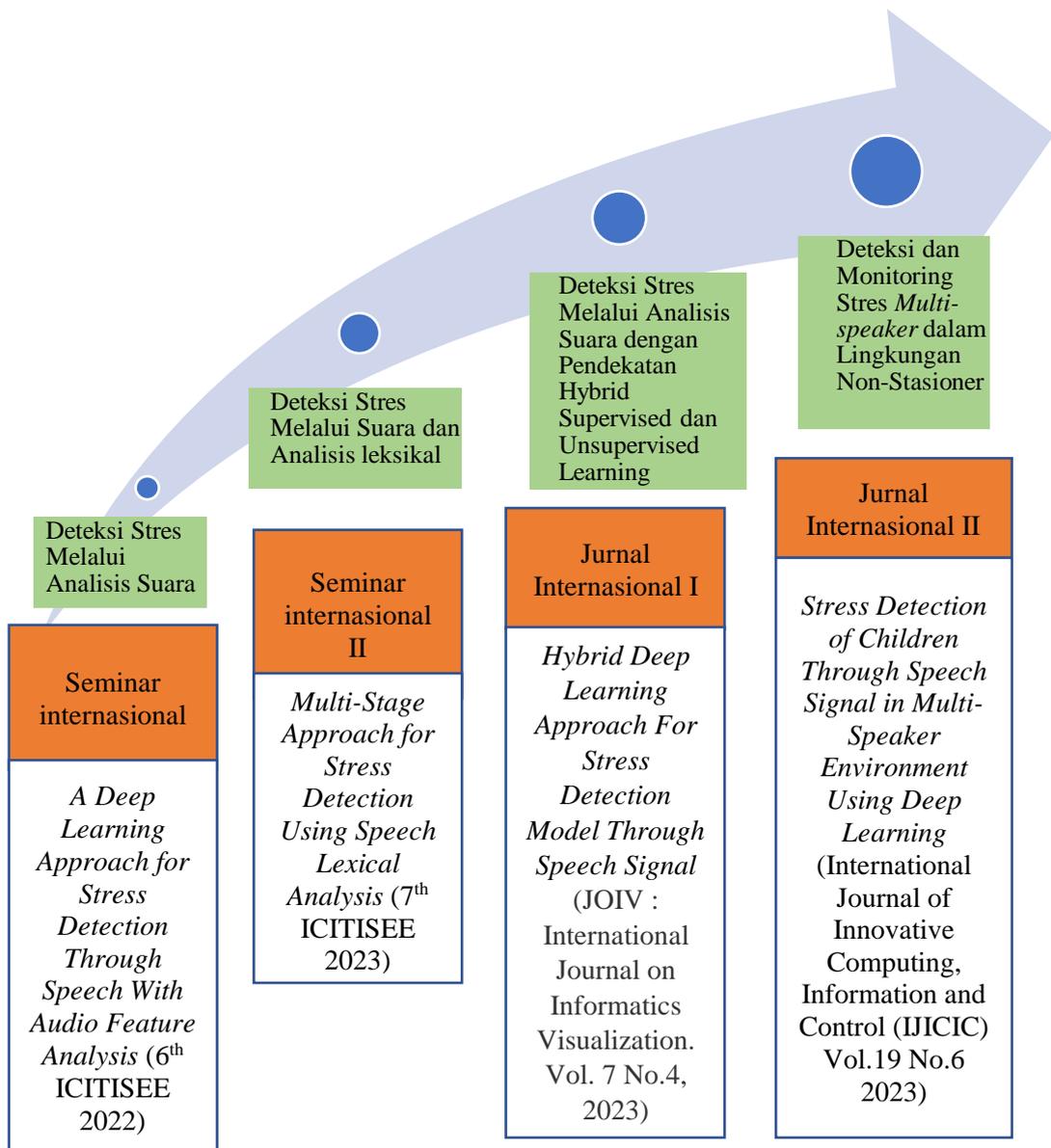
Tahapan pra-penelitian dimulai dari studi literatur untuk mencari berbagai sumber pustaka yang relevan dan terkini mengenai berbagai topik terkait dengan penelitian seperti deteksi dan monitoring stres melalui suara, audio speech processing seperti speech separation, speech enhancement, speech recognition and diarization dan sebagainya. Setelah itu penelitian I masuk ke tahap pengembangan model dengan menganalisis berbagai fitur suara yang relevan dalam deteksi stres untuk digunakan dalam ekstraksi fitur terhadap suara yang dianalisis. Selain fitur suara dilakukan juga analisis leksikal menggunakan pendekatan *Natural Language Processing* (NLP) untuk mendukung model deteksi stress. Topik ini akan menjadi fokus dari Penelitian I. Pada penelitian II, penelitian akan dengan mengeksplorasi model yang telah dikembangkan pada penelitian I, menggunakan pendekatan supervised dan unsupervised learning



untuk pembelajarannya bertujuan untuk meningkatkan akurasi klasifikasi deteksi stres dan mengeksplorasi hubungan asosiatif antara karakteristik suara dan derajat level stress yang dialami seseorang melalui karakteristik suara ucapannya dengan melihat klasterisasi data poin pada vector map. Hubungan asosiatif yang bila eksis dapat mendukung model untuk tidak saja mendeteksi stres tapi juga memprediksi derajat level stres dari seseorang melalui suara.

Selanjutnya pada penelitian III, penelitian dilanjutkan dengan fokus untuk memberikan dukungan untuk deteksi stres terhadap multi subjek melalui suara. Tantangan yang dihadapi dalam deteksi stress multi-subjek melalui suara adalah bagaimana melakukan separasi terhadap suara yang diakuisisi sehingga menjadi suara yang tersegregasi untuk masing-masing subjek yang akan dianalisis status stressnya. Untuk itu penelitian III ini berfokus dalam mengembangkan metode speech separation untuk mendukung model deteksi stres. Hasil yang diperoleh akan memungkinkan model untuk mampu menerima masukan suara yang diakuisisi dari lingkungan non-stasioner yang terdapat banyak derau dan suara subjek yang saling *overlap*. Model kemudian dapat melakukan separasi suara untuk menghasilkan suara ucapan yang telah tersegregasi. Suara yang telah tersegrasi tersebut kemudian dianalisis oleh model untuk kemudian dapat diprediksi status stressnya. Dataset pada penelitian ini menggunakan dataset suara anak pra dan awal sekolah dasar berusia antara 5-7 tahun. Selain itu penelitian ini juga memanfaatkan open-source public dataset yang digunakan dalam berbagai penelitian state-of-the-art mengenai deteksi stres melalui suara dengan tujuan untuk membandingkan kinerja model yang dihasilkan terhadap berbagai model tersebut. Luaran dari penelitian ini adalah masing-masing dua artikel pada prosiding seminar internasional terindeks yang membahas mengenai topik penelitian I, kemudian dua artikel pada jurnal internasional terindeks scopus yang isinya masing-masing membahas topik penelitian II dan III. Mileston dari luaran penelitian ini ditampilkan pada gambar 4 berikut





Gambar 4 Milestone dan Luaran Penelitian

2.1.4 Jenis penelitian

Jenis penelitian ini adalah penelitian eksperimental yang merupakan metode sistematis guna membangun hubungan yang mengandung fenomena sebab akibat. Penelitian eksperimen adalah suatu penelitian yang berusaha mencari pengaruh variabel tertentu terhadap variabel lainnya dalam kondisi yang terkontrol secara



nelitian eksperimen merupakan metode inti dari model penelitian yang akan pendekatan kuantitatif. Dalam metode eksperimental, diperlukan iatan utama sebagai prasyarat yaitu kegiatan mengontrol, kegiatan

memanipulasi, dan observasi. Penelitian ini diawali dengan melakukan kajian secara induktif terkait permasalahan yang hendak diatasi kemudian melakukan studi literatur yang relevan terkait dengan bidang penelitian, setelah itu penelitian dilanjutkan dengan pengumpulan data, analisa data dan masuk ke tahap pemodelan dan perancangan. Hasil dari pemodelan dan perancangan kemudian di uji implementasi lalu dilakukan evaluasi melalui analisis dan pengujian.

2.1.5 Sumber data

Sumber data pada penelitian ini terdiri dari data primer dan data sekunder, yaitu sebagai berikut:

1. Untuk menguji performance model dan benchmarking digunakan menggunakan public dataset yang tersedia umum diantaranya adalah RAVDESS (Dillon and Ni Teoh, 2021), SAVEE (Haq and Jackson, 2009), dan DAIC-WOZ (Gratch *et al.*, 2014). Dataset publik yang tersedia ini digunakan untuk menguji model yang akan dikembangkan terhadap model *state-of-the-art*.
2. Data primer akan secara langsung dikumpulkan pada lokasi penelitian secara langsung dengan subjek anak pra dan awal sekolah dasar dengan usia antara 5 sampai 7 tahun. Dalam menyusun dataset peneliti akan melibatkan psikolog anak untuk merancang berbagai kegiatan untuk mengaktifasi atau menginduksi stres pada subjek dengan menggunakan metode *Trier Social Stress Test (TSST)* yang bertujuan untuk mengemulasi stres yang disebabkan oleh *stressor* yang seringkali muncul dalam kehidupan sosial sehari-hari. (misalnya: melakukan public speaking, *mental arithmetic* dan sebagainya). Hasil dari pengambilan data berupa suara percakapan akan divalidasi oleh psikolog anak sebelum diberi label yang sesuai.

2.1.6 Instrumen penelitian



Penelitian ini instrumen penelitian terdiri dari perangkat keras dan perangkat lunak yang digunakan dalam mendukung penelitian ini, yaitu perangkat Keras (*Hardware*)

- a. Personal Komputer (PC) dengan spesifikasi CPU Intel Core I7-8600K
 - b. Memori RAM 16GB
 - c. Graphic Processing Unit (GPU) Nvidia Geforce GTX 1070
 - d. SSD 1 TB
 - e. Monitor VGA 16”.
 - f. Mikrofon high gain
 - g. Sound recorder
2. Perangkat Lunak (Software)
 - a. OS windows 10 Pro
 - b. Python
 - c. Anaconda
 - d. Audition

2.2 Hipotesis Penelitian

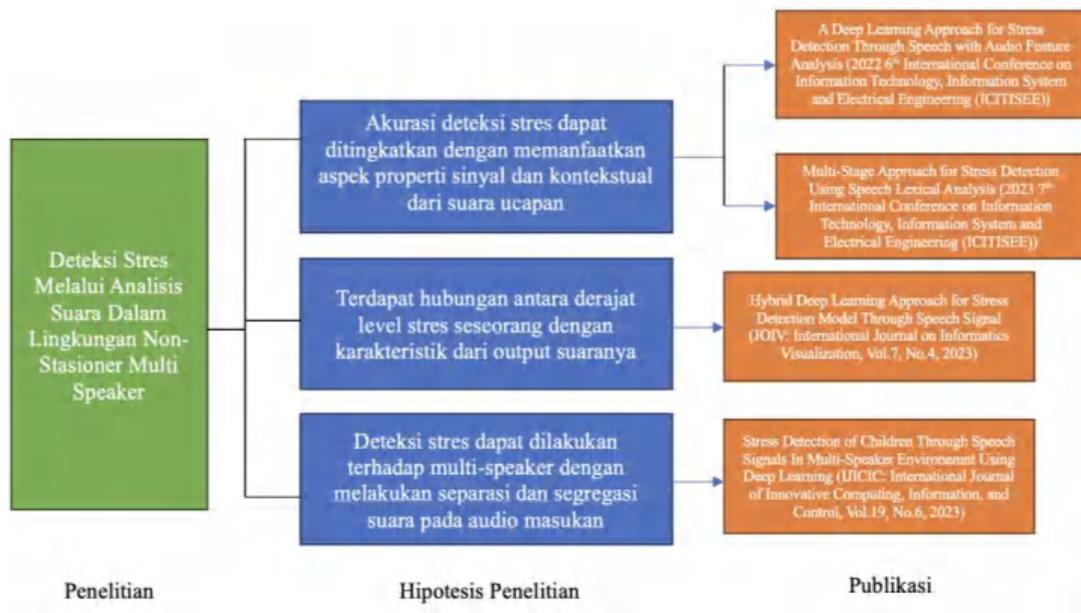
Berdasarkan kajian empirik pada pengembangan konseptual maka hipotesis pada penelitian ini adalah sebagai berikut:

Hipotesis 1: Akurasi deteksi stres dapat ditingkatkan dengan menggabungkan aspek properti sinyal dan kontekstual dari suara ucapan.

Hipotesis 2: Terdapat hubungan antara derajat level stres seseorang dengan karakteristik dari output suaranya.

Hipotesis 3: Deteksi stres dapat dilakukan terhadap multi-speaker dengan melakukan separasi dan segregasi suara pada audio masukan.





Gambar 5 Konstruksi penelitian dan hipotesis

Gambar 5 menunjukkan konstruksi penelitian dengan hipotesis dan publikasi yang mengkaji mengenai pembuktian hipotesis terkait. Berikut tabel 2 menunjukkan rumusan hipotesis dan penelitian pendukung terkait dengan hipotesis yang diangkat.

Tabel 2 Rumusan hipotesis dan penelitian terkait

No	Rumusan	Penelitian Terkait
H1	Akurasi deteksi stres dapat ditingkatkan dengan menggabungkan aspek properti sinyal dan kontekstual dari suara ucapan	(Turcan and McKeown, 2019; Van Puyvelde <i>et al.</i> , 2018; Pisanski and Sorokowski, 2021)
H2	Terdapat hubungan antara derajat level stres seseorang dengan karakteristik dari output	(Matsuo, Hayakawa and Harada, 2015; König <i>et al.</i> , 2021; Han, Byun and Kang, 2018; Tomba <i>et al.</i> , 2018; Madhavi <i>et al.</i> , 2020)



No	Rumusan	Penelitian Terkait
	suaranya	
H3	Deteksi stres dapat dilakukan terhadap multi-speaker dengan melakukan separasi dan segregasi suara pada audio masukan	(John R Hershey <i>et al.</i> , 2016; Luo and Mesgarani, 2018, 2019; Chen, Luo and Mesgarani, 2017; Wijayakusuma <i>et al.</i> , 2021; Huang <i>et al.</i> , 2022)

