

BAB I

PENDAHULUAN

1.1 Latar Belakang

Depresi bukan lagi menjadi sebuah hal yang tabu, kebanyakan masyarakat saat ini mengetahui apa itu depresi tetapi mereka tidak memahami bagaimana cara menyikapinya. Jenjang mahasiswa memiliki risiko tinggi dalam mengalami masalah kesehatan mental karena mereka harus beradaptasi dengan lingkungan sosial baru, menghadapi tekanan untuk meraih karier, dan menghadapi tantangan finansial. Dari proses adaptasi yang dilalui ini dapat mempengaruhi bagaimana cara mereka berpikir dan bertindak. Depresi didefinisikan sebagai gangguan suasana hati yang mempengaruhi pikiran, perasaan, dan perilaku seseorang. Hal ini menjadi perhatian utama dalam kesehatan mental global dikarenakan dapat mempengaruhi dalam bagaimana seseorang mengatasi tekanan, berkomunikasi dengan orang lain, membuat keputusan, dan bahkan menciptakan dorongan untuk melukai diri sendiri[1]. Menurut data dari World Health Organization (WHO), depresi merupakan gangguan kesehatan mental yang umum dan mempengaruhi sekitar 5% orang dewasa secara global [1]. Depresi cenderung lebih umum terjadi pada perempuan dibandingkan laki-laki. Secara global, lebih dari 10% perempuan yang sedang hamil atau baru melahirkan mengalami depresi. Angka kejadian depresi diperkirakan mencapai 280 juta orang. Setiap tahun, lebih dari 700.000 orang meninggal akibat bunuh diri, menjadikannya sebagai penyebab kematian keempat tertinggi di kalangan usia 15-29 tahun[1]. Mendeteksi secara dini gejala depresi sangat penting dilakukan karena dapat membantu dalam intervensi lebih awal serta pengelolaan yang lebih efektif.

Depresi dapat disebabkan oleh berbagai faktor, yaitu faktor biologis, psikologis, dan sosial. Faktor-faktor lainnya seperti biologi otak, genetika, lingkungan, dan kondisi kesehatan fisik dan mental lainnya berkontribusi pada pengembangan depresi[2]. Menurut DBSA (Depression and Bipolar Support Alliance)[3], penyebab kecacatan utama di seluruh dunia maupun pengaruh utama pada beban penyakit di global yaitu depresi. Angka kematian bunuh diri yang

disebabkan oleh depresi yang dilaporkan oleh Mental Health U.S White House Conference di tahun 1999 mencapai 30,000 kematian akibat depresi. Diantara angka kematian tersebut rate bunuh diri akibat depresi untuk dewasa mencapai lebih dari 50%, hingga dua per tiga kasus bunuh diri pada orang lanjut usia disebabkan depresi yang tidak ditangani atau juga yang *misdiagnosed*[3].

WHO mencatat depresi dan kecemasan menyebabkan kerugian 12 miliar hari kerja dan biaya US\$ 1 triliun per tahun. Risiko bunuh diri meningkat dengan depresi yang juga berkorelasi dengan penyakit serius. Di tahun 2019, 970 juta orang mengalami gangguan mental, terutama depresi dan kecemasan, yang meningkat signifikan selama pandemi COVID-19[4]. Bunuh diri sering terjadi karena faktor sosioekonomi, dengan lebih dari 700.000 kasus tiap tahun, dan banyak di negara berpenghasilan rendah dan menengah[5].

Misdiagnosis antara gangguan bipolar dan depresi mayor adalah masalah yang serius dan umum terjadi[6]. Salah diagnosis depresi adalah masalah yang signifikan dalam kesehatan mental. Dalam sebuah penelitian di Ethiopia, ditemukan bahwa 57.60% pasien yang didiagnosis dengan gangguan depresif mayor (MDD) sebenarnya salah didiagnosis. Pasien dengan MDD sering kali salah didiagnosis sebagai skizofrenia atau gangguan bipolar. Salah diagnosis ini terjadi karena gejala depresi sering tumpang tindih dengan kondisi lain seperti skizofrenia dan gangguan bipolar. Selain itu, kurangnya informasi yang dilaporkan oleh pasien dan proses diagnostik yang tidak komprehensif juga berkontribusi pada kesalahan ini. Tekanan eksternal dan perbedaan budaya dalam mengekspresikan gejala juga mempengaruhi bagaimana depresi dikenali dan didiagnosis oleh profesional kesehatan. Secara keseluruhan, sekitar 39.16% dari pasien dengan gangguan psikiatrik berat mengalami salah diagnosis. Pengenalan algoritma prediktif ini diharapkan dapat membantu dokter dalam mengidentifikasi pasien yang berisiko tinggi dan memberikan perawatan yang lebih tepat waktu dan efektif, sehingga mengurangi durasi penyakit yang tidak diobati dan dampak negatif dari pengobatan yang tidak tepat.

Penelitian oleh Cai et al [7]. menyajikan dataset multi-modal (MODMA) terbuka untuk analisis gangguan mental, termasuk data EEG dan audio dari pasien depresi yang didiagnosis secara klinis dan kontrol normal yang cocok. Pasien dipilih dan didiagnosis oleh psikiater profesional. Data EEG mencakup sinyal EEG tradisional dan kolektor EEG 3-elektroda yang inovatif. Data audio direkam selama wawancara, membaca, dan deskripsi gambar. Memanfaatkan dataset MODMA, yang merupakan kumpulan data multi-modal untuk analisis gangguan mental. Meskipun dataset ini mencakup data EEG yang direkam menggunakan kap elektroda 128-elektroda dan kolektor EEG 3-elektroda, fokus utama penelitian ini terletak pada penggunaan data audio. Dataset audio ini terdiri dari rekaman suara dari 52 subjek yang telah didiagnosis secara klinis dengan depresi, beserta dengan kontrol normal yang sepadan, yang dikumpulkan selama wawancara, membaca, dan deskripsi gambar.

He et al [8]. mengembangkan metode baru untuk analisis depresi dari suara dengan menggabungkan fitur yang diolah secara manual dan fitur yang dipelajari secara mendalam menggunakan Convolutional Neural Networks (CNN). Metode ini dirancang untuk mengukur tingkat depresi secara efektif, menawarkan pendekatan yang mengatasi keterbatasan fitur yang dibuat secara manual dan memanfaatkan kekuatan CNN dalam mengenali pola dengan menggunakan hasil dari ekstraksi audio menjadi spektrum gambar. Hasil uji coba pada dataset AVEC2013 menunjukkan bahwa model deep learning menghasilkan MAE sebesar 8,4832 dan RMSE sebesar 10,4561, sementara pada dataset AVEC2014, model ini menghasilkan MAE sebesar 8,6014 dan RMSE sebesar 10,4413. Penelitian ini menegaskan bahwa kombinasi fitur yang dipelajari secara mendalam dan fitur yang diolah secara manual dapat secara efektif mengukur tingkat keparahan depresi dari suara, menawarkan alat yang lebih akurat dan efisien untuk diagnosis otomatis depresi.

Dalam penelitian oleh Liu et. al [9], di Gansu *Provincial Key Laboratory of Wearable Computing*, mengembangkan metode pengenalan depresi berbasis suara menggunakan decision tree yang inovatif. Studi ini menggunakan sampel 52 subjek yang terdiri dari 23 pasien depresi dan 29 kontrol sehat. Metode ini

berhasil mencapai tingkat akurasi 75,8% untuk model berbasis gender pria dan 68,5% untuk model berbasis gender wanita. Hasil ini menunjukkan bahwa model decision tree yang diusulkan dapat meningkatkan performa klasifikasi depresi secara signifikan, memberikan alternatif yang lebih objektif dan efektif dibandingkan metode tradisional yang bergantung pada laporan pasien dan wawancara psikiater yang mungkin bias subyektif. Meskipun nama MODMA dataset tidak disebutkan secara eksplisit dalam paper Liu [9], karakteristik dataset yang mereka gunakan sangat mirip dengan deskripsi yang diberikan dalam paper Cai et al. (2020). Kedua paper ini memiliki penulis yang sama dan menggunakan metode pengumpulan data yang identik, yang menunjukkan bahwa dataset yang digunakan adalah MODMA dataset. Referensi Website MODMA: "Paper Liu et al. (2020) terdaftar di situs web MODMA sebagai contoh penggunaan dataset ini. Ini menunjukkan bahwa peneliti tersebut memang menggunakan MODMA dataset

Dalam analisis sinyal audio, mel-spectrogram menjadi alat yang sangat penting karena kemampuannya untuk merepresentasikan spektrum frekuensi sinyal audio dalam skala mel, yang lebih sesuai dengan persepsi pendengaran manusia. Mel-spectrogram digunakan dalam berbagai aplikasi, termasuk pengenalan suara, analisis musik, dan deteksi kondisi kesehatan melalui suara, seperti deteksi dini depresi. Mel-spectrogram menghitung dan mencari skala mel melalui beberapa langkah penting. Proses dimulai dengan memuat sinyal audio yang akan dianalisis[16]. Sinyal audio ini biasanya disimpan dalam format file seperti .wav atau .mp3. Setelah sinyal dimuat, parameter utama seperti panjang jendela (window length), panjang tumpang tindih (overlap length), dan jumlah titik FFT (FFT length) diatur untuk memastikan resolusi yang diinginkan. Transformasi Fourier Cepat (FFT) kemudian diterapkan pada setiap jendela waktu untuk mendapatkan spektrum frekuensi dari sinyal. FFT mengubah sinyal dari domain waktu ke domain frekuensi, memungkinkan analisis lebih lanjut. Setelah itu, spektrum frekuensi dipetakan ke skala mel menggunakan serangkaian filter segitiga yang disusun pada skala mel. Filter ini dirancang untuk mencerminkan cara telinga manusia merespons frekuensi. Energi dari setiap filter segitiga dihitung, dan nilai energi ini kemudian dikonversi dari skala linier ke skala

logaritma. Konversi ini dilakukan untuk mencocokkan persepsi intensitas suara manusia yang bersifat logaritmik. Hasil akhirnya adalah mel-spectrogram yang memvisualisasikan distribusi energi frekuensi dalam skala mel. Mel-spectrogram ini memberikan representasi yang lebih baik dari spektrum frekuensi yang sesuai dengan persepsi pendengaran manusia.

Penggunaan sinyal dalam bentuk mel spectrogram telah menjadi metode yang sangat efektif dibandingkan dengan sinyal audio 1D. Mel spectrogram adalah representasi visual dari sebuah sinyal yang menunjukkan variasi energi pada berbagai pita frekuensi seiring waktu, juga dikenal sebagai representasi waktu-frekuensi dari sinyal[17]. Mel spectrogram dihasilkan dengan membagi sinyal menjadi segmen-segmen kecil dari jendela waktu yang sangat pendek dan kemudian menghitung spektrum daya pada setiap segmen waktu menggunakan metode seperti Short-Term Fourier Transform (STFT) atau filter bandpass. Salah satu keunggulan utama dari mel spectrogram adalah kemampuannya untuk menangkap informasi temporal dan spektral dari sinyal audio. Ini memungkinkan ekstraksi fitur yang lebih kompleks, yang sangat berguna dalam tugas-tugas pengenalan pola dan klasifikasi. Dengan representasi 2D ini, kita dapat menerapkan teknik-teknik pengolahan citra canggih seperti Convolutional Neural Networks (CNNs), yang telah terbukti sangat efektif dalam pengenalan pola pada data gambar[17]. CNN dapat mendeteksi pola yang rumit dalam data audio yang mungkin terlewatkan jika hanya menggunakan sinyal 1D. Selain itu, skala mel pada mel spectrogram dirancang untuk meniru cara telinga manusia merespons frekuensi suara yang berbeda. Ini berarti fitur yang diekstrak dari mel spectrogram lebih sesuai dengan persepsi manusia terhadap suara, yang dapat meningkatkan akurasi model dalam tugas-tugas yang melibatkan pengenalan atau klasifikasi suara manusia. Oleh karena itu, penggunaan mel spectrogram dalam analisis sinyal audio menawarkan berbagai keuntungan yang membuatnya lebih unggul dibandingkan dengan metode pemrosesan sinyal 1D tradisional.

Bagi dokter, sistem ini menawarkan deteksi untuk mendiagnosis pasien. Dengan demikian, sistem deteksi depresi berbasis suara ini tidak hanya

menawarkan cara yang lebih efisien dan akurat dalam mendeteksi depresi tetapi juga relevan dan bermanfaat bagi dokter dan pasien dalam konteks praktis dan klinis. Perbedaan utama antara deteksi dan validasi terletak pada fokus dan tahapnya. Deteksi depresi ini berfokus pada pengenalan awal suatu kondisi untuk intervensi segera, sedangkan validasi berfokus pada penilaian kinerja model untuk memastikan akurasi dan keandalan prediksi. Deteksi ini terjadi di tahap awal identifikasi kondisi, sementara validasi terjadi setelah model dilatih, untuk menguji dan mengevaluasi kinerjanya. Untuk menganalisis mel spectrogram secara efektif, digunakan Convolutional Neural Networks (CNN), salah satu arsitektur CNN yang terkenal dan efektif untuk tugas klasifikasi adalah InceptionV3. InceptionV3 dikembangkan oleh para peneliti di Google sebagai evolusi dari arsitektur Inception sebelumnya (termasuk GoogLeNet). Arsitektur ini dirancang untuk meningkatkan efisiensi komputasi sambil mempertahankan atau bahkan meningkatkan akurasi model.

Lalu setelah itu akan membandingkan model Convolutional Neural Network (CNN) yang menggunakan algoritma ShuffleNet dengan model CNN yang menggunakan transfer learning. ShuffleNet dipilih karena arsitekturnya yang efisien dan kemampuannya untuk mengurangi beban komputasi tanpa mengorbankan performa. Di sisi lain, juga akan menerapkan model CNN berbasis transfer learning menggunakan arsitektur InceptionV3, termasuk InceptionV3. Perbandingan ini dilakukan untuk mengevaluasi performa antara CNN yang dibangun dari awal (non-transfer learning) dengan CNN yang menggunakan transfer learning. Model transfer learning diharapkan dapat memberikan hasil yang lebih baik karena model ini telah dilatih sebelumnya pada dataset besar dan umum, sehingga memiliki kemampuan ekstraksi fitur yang lebih kuat dan generalisasi yang lebih baik.

Dengan membandingkan model pretrained seperti InceptionV3 yang menggunakan transfer learning dan ShuffleNet yang sering digunakan dalam skenario non-transfer learning, penelitian ini dapat mengevaluasi sejauh mana keuntungan dari menggunakan model yang telah dilatih sebelumnya pada dataset besar umum. Pemilihan ini bukan hanya berdasarkan performa akurasi tetapi juga

mempertimbangkan kebutuhan praktis dari implementasi sistem dalam berbagai kondisi perangkat keras dan sumber daya yang berbeda. Dalam konteks aplikasi nyata, tidak semua pengguna atau sistem akan memiliki akses ke perangkat keras yang kuat. ShuffleNet menawarkan alternatif yang lebih ringan dengan tetap memberikan performa yang baik, sehingga relevan untuk dibandingkan dengan InceptionV3 yang lebih berat.

Akurasi adalah metrik yang sangat penting karena menunjukkan seberapa baik model secara keseluruhan dalam membuat prediksi yang benar. Namun, recall juga sangat penting karena kita ingin memastikan bahwa semua kasus depresi yang ada dapat terdeteksi oleh sistem (sensitivitas tinggi). Validation penting untuk memastikan bahwa model tidak overfitting dan dapat diandalkan pada data baru yang tidak terlihat selama pelatihan.

Jadi, sementara akurasi memberikan gambaran umum tentang kinerja model, recall dan hasil validasi juga sangat penting untuk memastikan bahwa model dapat diandalkan dalam mendeteksi depresi secara konsisten dan tidak hanya menghafal data pelatihan. Dengan demikian, penelitian ini akan memberikan wawasan mendalam mengenai efektivitas penggunaan fitur *Mel-spectrogram* dalam deteksi depresi, serta membandingkan performa berbagai model CNN, baik yang menggunakan algoritma ShuffleNet maupun model transfer learning berbasis InceptionV3.

Pemilihan ShuffleNet dan InceptionV3 dalam penelitian ini didasarkan pada perbedaan signifikan dalam kebutuhan komputasi dan arsitektur yang ditawarkan oleh kedua model tersebut. ShuffleNet dipilih karena efisiensinya yang tinggi dalam komputasi dan penggunaan daya yang rendah, sehingga cocok untuk implementasi pada perangkat dengan sumber daya terbatas. Di sisi lain, InceptionV3 dipilih karena keakuratannya yang tinggi dan kemampuannya dalam menangkap berbagai skala fitur melalui modul Inception, yang telah terbukti efektif dalam berbagai tugas klasifikasi. Penelitian ini tidak hanya mencari akurasi tertinggi, tetapi juga mempertimbangkan performa komputasi dari model-model yang dibandingkan. Dengan melakukan perbandingan ini, penelitian ini berupaya menemukan model yang dapat memberikan hasil yang andal dan akurat, namun

tetap efisien dan praktis untuk digunakan pada perangkat dengan keterbatasan sumber daya.

1.2 Identifikasi Masalah

- 1.2.1 Apakah fitur *mel-spectrogram* dapat digunakan untuk men klasifikasi subjek Major Depressive Disorder dan Healthy Control
- 1.2.2 Apakah Convolutional Neural Networks (CNN) dapat meningkatkan performa dari klasifikasi Major Depressive Disorder dan Healthy Control berdasarkan suara
- 1.2.3 Apakah Transfer Learning dapat meningkatkan performa dari klasifikasi Major Depressive Disorder dan Healthy Control berdasarkan suara

1.3 Batasan Penelitian

- 1.3.1 Penelitian akan terfokus pada analisis audio signal menggunakan machine learning, dengan batasan pada data suara yang dikumpulkan dari subjek yang telah ditentukan.
- 1.3.2 Dataset yang digunakan adalah multi-modal dataset (MODMA) Dataset
- 1.3.3 Studi ini akan membandingkan efektivitas model decision tree dengan model Convolutional Neural Networks (CNN), tanpa melibatkan atau membandingkan dengan model machine learning lain.

1.4 Tujuan Penelitian

- 1.4.1 Menilai efektivitas klasifikasi audio signal dengan menggunakan model Convolutional Neural Networks (CNN).
- 1.4.2 Evaluasi Penggunaan Fitur *mel-spectrogram* dalam Klasifikasi suara Major Depressive Disorder dan Healthy Control.
- 1.4.3 Peningkatan Kinerja Klasifikasi dengan CNN.

1.5 Manfaat Penelitian

- 1.5.1 Memberikan wawasan baru tentang pemanfaatan teknologi machine learning dalam deteksi dini depresi, yang dapat berkontribusi pada intervensi dan pengobatan lebih awal.
- 1.5.2 Meningkatkan pemahaman tentang keefektifan model CNN dalam analisis audio signal untuk deteksi depresi.
- 1.5.3 Memberikan rekomendasi untuk pengembangan alat diagnostik non-invasif berbasis suara yang dapat meningkatkan aksesibilitas layanan kesehatan mental.
- 1.5.4 Berkontribusi pada penelitian di bidang kesehatan mental dengan menyediakan data dan analisis yang dapat mendukung praktik klinis dan kebijakan kesehatan.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian disusun dan dibagi atas lima bab sebagai berikut.

- 1.6.1 BAB 1 PENDAHULUAN
- 1.6.2 BAB 2 TINJAUAN PUSTAKA
- 1.6.3 BAB 3 ANALISIS DAN PERANCANGAN SISTEM
- 1.6.4 BAB 4 ANALISIS DAN PENGUJIAN SISTEM
- 1.6.5 BAB 5 SIMPULAN DAN SARAN