SKRIPSI

PENGENALAN EMOSI MANUSIA MELALUI WAJAH DAN SUARA DENGAN MENGGUNAKAN CNN

Disusun untuk Memenuhi Syarat Mendapatkan Gelar Sarjana Teknik pada
Jurusan Teknik Elektro Fakultas Teknik
Universitas Sriwijaya

Oleh:

FAZRUN ARROFIQ

03041281722048

JURUSAN TEKNIK ELEKTRO

FAKULTAS TEKNIK

UNIVERSITAS SRIWIJAYA

2022

LEMBAR PENGESAHAN
PENGENALAN EMOSI MANUSIA MELALUI WAJAH DAN SUARA DENGAN MENGGUNAKAN CNN

SKRIPSI

Disusun untuk Memenuhi Syarat Mendapatkan Gelar Sarjana Teknik pada

Jurusan Teknik Elektro Fakultas Teknik

Universitas Sriwijaya

Oleh :

FAZRUN ARROFIQ

03041281722048

Palembang, 25 Juli 2022

Mengetahui ,

Ketua Jurusan Teknik Elektro

Menyetujui

Pembimbing Utama

Muhammad Abu Bakar Sidik, S. T., M. Eng., Ph.D.

NIP : 197108141999031005

Dr. Eng. Suci Dwijayanti, S.T., M.S.

NIP. 198407302008122001

HALAMAN PERNYATAAN INTEGRITAS

Yang bertanda tangan di bawah ini :

Nama : Fazrun Arrofiq
NIM : 03041281722048
Fakultas : Teknik
Jurusan/Prodi : Teknik Elektro
Universitas : Universitas Sriwijaya

Hasil Pengecekan *software iThenticate/Turnitin* : 6%

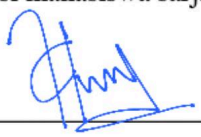
Menyatakan bahwa tugas akhir saya yang berjudul "Pengenalan Emosi Manusia Melalui Wajah dan Suara Dengan Menggunakan CNN" merupakan hasil karya sendiri dan benar keasliannya. Apabila ternyata dikemudian hari ditemukan unsur penjiplakan/plagiat dalam karya ilmiah ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian penyataan ini saya buat dengan sebenarnya dan tanpa paksaan.

Palembang, 25 Juli 2022

Fazrun Arrofiq
NIM. 03041281722048

Saya sebagai pembimbing dengan ini menyatakan bahwa saya telah membaca dan menyetujui skripsi ini dan dalam pandangan saya ruang lingkup dan kualitas skripsi ini mencukupi sebagai skripsi mahasiswa sarjana strata satu (S1).

Tanda Tangan : 

Pembimbing Utama : Dr. Eng. Suci Dwijayanti, S.T., M.S.

Tanggal : 25 / Juli / 2022

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Sriwijaya, saya yang bertanda tangan di bawah ini :

Nama : Fazrun Arrofiq
NIM 03041281722048
Fakultas : Teknik
Jurusan/Prodi : Teknik Elektro
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Sriwijaya **Hak Bebeas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul :

Pengenalan Emosi Manusia Melalui Wajah dan Suara Dengan Menggunakan CNN

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Sriwijaya berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tulisan saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian penyataan ini saya buat dengan sebenarnya.

Dibuat di Palembang
Pada tanggal : 25 Juli 2022
Yang Menyatakan,

Fazrun Arrofiq
NIM. 03041281722048

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah memberikan rahmat dan karunia-Nya kepada penulis, keluarga, sahabat, maupun tenaga pendidik yang sudah memberikan ilmu dan pengalaman kepada penulis, sehingga penulis dapat menyelesaikan skripsi ini dengan judul "Pengenalan Emosi Manusia Melalui Wajah dan Suara Dengan Menggunakan CNN" sebagai salah satu syarat untuk mendapatkan gelar Sarjana Teknik pada Jurusan Teknik Elektro Fakultas Teknik Universitas Sriwijaya.

Skripsi ini terselesaikan dengan bantuan dari berbagai pihak selama masa penyusunan skripsi. Dengan begitu penulis ingin mengucapkan terima kasih setulus-tulusnya kepada:

1. Bapak Muhammad Abu Bakar Sidik, S.T., M.Eng., Ph.D. selaku Ketua Jurusan Teknik Elektro Universitas Sriwijaya dan Ibu Dr. Eng. Suci Dwijayanti S.T., M.S. selaku Sekretaris Jurusan Teknik Elektro Universitas Sriwijaya.
2. Ibu Dr. Eng. Suci Dwijayanti S.T., M.S. selaku dosen pembimbing utama tugas akhir yang telah bersabar dalam memberikan bimbingan kepada penulis sehingga proses penulisan berjalan dengan baik.
3. Bapak Dr. Bhakti Yudho Suprapto, S.T., M.T. dan Ibu Dr. Eng. Suci Dwijayanti S.T., M.S. yang telah memberikan pencerahan dalam pemilihan tema pada bimbingan tugas akhir ini serta memberikan dukungan dan bimbingan sehingga tugas akhir dapat diselesaikan dengan lancar.
4. Dosen pembimbing akademik, ibu Hermawati, ST, MT yang telah memberikan arahan serta bimbingan kepada penulis selama masa perkuliahan.
5. Seluruh staff pengajar Jurusan Teknik Elektro Fakultas Teknik Universitas Sriwijaya yang telah memberikan ilmu pengetahuan yang tak ternilai selama penulis menempuh masa perkuliahan.

6. Orang tua penulis, Nasrun Simatupang dan If Ezmi, yang selalu memberikan doa, semangat, kasih sayang, nasehat kepada penulis yang sangat berharga dalam peroses perkuliahan dan pembuatan skripsi ini.
7. Kakak penulis Vina oktaviany dan Afrianty Ramadhani yang telah memberikan dukungan selama perkuliahan dan penyusunan skripsi ini.
8. Keluarga Sukan Agung Perdana yang telah memperbolehkan penulis menetapi rumahnya selama masa penulisan skripsi ini.
9. Irvine Valiant Fanthony, Markus Hermawan, M. Fauzan Nugraha, Sukan Agung Perdana, M Zulfikar Firdaus, M. Zaid Haritsyah, Annisa DR dan teman-teman satu angkatan konsentrasi Teknik kendali dan Komputer yang telah banyak membantu pada masa perkuliahan dan dalam menyelesaikan skripsi.
10. Teman – teman ersiz kost Vendra, Dicky, Jodi, Sukan, Togar, Niko yang selalu membantu, menyemangati dan menjadi teman yang tak terlupakan selama masa perkuliahan terutama saat di kosan.
11. Seluruh responden Agit, Fauzi, Yuda, Novan dan Uni Vina yang terlah membantu dalam penyelesaian skripsi ini.
12. Semua pihak yang tidak dapat penulis sebutkan satu-persatu yang telah memberikan doa dan motivasi sehingga dapat terselesaikannya skripsi ini.

Di dalam penyusunan skripsi ini, masih terdapat kekurangan karena keterbatasan penyusun, oleh karena itu penyusun mengharapkan kritik dan saran yang membangun agar dapat menjadi evaluasi dan berguna untuk penyusun dimasa yang akan datang.

Palembang, 25 Juli 2022

Fazrun Arrofiq

NIM. 03041281722048

ABSTRAK

PENGENALAN EMOSI MANUSIA MELALUI WAJAH DAN SUARA DENGAN MENGGUNAKAN CNN

(Fazrun Arrofiq, 03041281722048, 2022, 73 Halaman)

Zaman ini, pekerjaan sulit dapat dilakukan oleh benda yang dibuat manusia yang dinamakan robot. Salah satunya merupakan robot yang menyerupai manusia yang dinamakan *humanoid robot*. Untuk menyerupai manusia *humanoid robot* sebisanya mampu untuk mengenali emosi manusia, karena emosi merupakan hal umum yang dimiliki manusia berupa emosi senang, sedih, marah, tenang,dan terkejut. Maka penelitian ini akan mengembangkan model CNN yang dapat mengenali emosi melalui gabungan wajah dan suara karena penelitian sebelumnya hanya menggunakan salah satu karakteristik yang dimiliki manusia saja. Untuk mengenali emosi, penelitian ini menggunakan arsitektur CNN berupa VGG dan AlexNet. Dengan menggunakan bahasa pemrograman *python* untuk melatih model CNN, didapatkan hasil bahwa model mampu mengenali emosi manusia melalui wajah atapun melalui suara dengan nilai akurasi terbaik sebesar 76,4% pada model wajah VGG saja dan 75,9% pada model gabungan dalam mengenali wajah dengan emosi netral. Dan akurasi terbaik pada pengenalan suara hanya dapat membaca emosi marah 100% pada model suara VGG saja dan 100% emosi senang pada model gabungannya. Namun hal ini dapat diatasi dengan menggunakan input suara sepertihalnya model *tensorflow* dengan akurasi yang didapat mencapai 90% untuk keseluruhan model. Dengan ini dapat disimpulkan bahwa model dapat mengenali emosi manusia dengan input berupa wajah maupun suara dengan performansi arsitektur VGG lebih baik dibandingkan arsitektur AlexNet.

Kata Kunci: Pengenalan Emosi Wajah, Pengenalan Emosi Suara, CNN, VGG, AlexNet, *Tensorflow*

Palembang, 25 Juli 2022

Mengetahui,	**Menyetujui**
Ketua Jurusan Teknik Elektro	**Pembimbing Utama**
Muhammad Abu Bakar Sidik, S. T., M. Eng., Ph.D.	**Dr. Eng. Suci Dwijayanti, S.T., M.S.**
NIP. 197108141999031005	**NIP. 198407302008122001**

ABSTRACT

RECOGNITION OF HUMAN EMOTIONS THROUGH FACE AND VOICE USING CNN

(Fazrun Arrofiq, 03041281722048, 2022, 73 Pages)

These days, difficult jobs can be done by man-made objects called robots. One of them is a robot that resembles a human called a humanoid robot. To resemble a human, a humanoid robot should be able to recognize human emotions as much as possible, because emotions are common in humans in the form of happy, sad, angry, calm, and surprised emotions. So this study will develop a CNN model that can recognize emotions through a combination of faces and voices because previous studies only used one of the characteristics possessed by humans. To recognize emotions, this research uses CNN architecture in the form of VGG and AlexNet. By using the python programming language to train the CNN model, the results show that the model is able to recognize human emotions through the face or through the voice with the best accuracy value of 76.4% on the VGG facial model only and 75.9% on the combined model in recognizing faces with emotions. neutral. And the best accuracy on voice recognition can only read 100% angry emotions on the VGG voice model and 100% happy emotions on the combined model. However, this can be overcome by using voice input as well as the tensorflow model with an accuracy of up to 90% for the entire model. With this it can be concluded that the model can recognize human emotions with input in the form of faces and voices with the VGG architecture performing better than the AlexNet architecture.

Keyword: Face Emotion Recognition, Speech emotion Recognition, CNN, VGG, AlexNet, *Tensorflow*

Palembang, 25 Juli 2022

Mengetahui,
Ketua Jurusan Teknik Elektro

Menyetujui
Pembimbing Utama

Muhammad Abu Bakar Sidik, S. T., M. Eng., Ph.D.
NIP : 197108141999031005

Dr. Eng. Suci Dwijayanti, S.T., M.S.
NIP. 198407302008122001

DAFTAR ISI

HALAMAN SAMPUL i
LEMBAR PENGESAHAN ii
HALAMAN PERNYATAAN INTEGRITAS iii
LEMBAR PERNYATAAN DOSEN iv
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS v
KATA PENGANTAR vi
ABSTRAK viii
ABSTRACT ix
DAFTAR ISI x
DAFTAR GAMBAR xiii
DAFTAR TABEL xv
BAB I 1
1.1 Latar Belakang Penelitian 1
1.2 Permasalahan 3
1.3 Tujuan Penelitian 3
1.4 Pembatasan Masalah 3
1.5 Keaslian Penelitian 4
BAB II 6
2.1 State of The Art 6
2.2 Teori pendukung 11
2.2.1 *Face Emotion Recognition* 11
2.2.2 *Speech Emotion Recognition* 12
2.2.2.1. *Prepocessing* 13
2.2.2.2. Ekstrasi Ciri *Spectrogram* 14
2.2.3 CNN 15
BAB III 19

3.1 Studi Literatur 19
3.2 Pengambilan Data 19
3.3 Perancangan Sistem 20
3.4 Pengujian 22
BAB IV 24
4.1 Pengambilan Data 24
4.2 *Preprocessing Data* 28
4.3 Arsitektur CNN 30
4.3.1 VGG Net 30
4.3.2 Alex Net 33
4.4 Pelatihan CNN 34
4.4.1 Pelatihan Pengenalan Emosi dari Citra Wajah 34
4.4.1.1 VGG 35
4.4.1.2 Alex Net 36
4.4.2 Pelatihan Pengenalan Emosi dari Suara 37
4.4.2.1 VGG 38
4.4.2.2 Alex Net 39
4.4.3 Pengenalan Emosi dari Kombinasi Wajah dan Suara 40
4.4.3.1 VGG 41
4.4.3.2 Alex Net 42
4.5 Pengujian Model CNN 43
4.5.1 Wajah 44
4.5.2 Suara 50
4.6 Pengujian Model Suara Menggunakan *Pre-processing* Tensorflow 56
BAB V 62
5.1 Kesimpulan 62
5.2 Saran 62
DAFTAR PUSTAKA 64

LAMPIRAN 68

DAFTAR GAMBAR

Gambar 2.1 Flowchart dalam menentukan emosi wajah kelompok 10
Gambar 2.2 Sampel spectrogram yang dihasilkan oleh BAGAN orisinil (a) dan metode yang disarankan (b) 10
Gambar 2.3 Jenis - jenis citra digital (a) Citra biner, (b) Citra *greyscale* dan (c) Citra RGB 11
Gambar 2.4 Contoh emosi yang dapat dilihat melalui ekspresi wajah (dari kiri ke kanan yaitu marah, senang, netral dan sedih) [15] 12
Gambar 2.5 Ilustrasi arsitektur CNN untuk mengklasifikasi sebuah gambar 16
Gambar 2.6 Conv2D 17
Gambar 2.7 Fungsi ReLU 18
Gambar 2.8 *Max pooling layer* 18
Gambar 3.1 *Flowchart* penelitian 21
Gambar 4.1 Gambar wajah yang telah dipilih sesuai frame (a) ekspresi marah, (b) ekspresi netral, (c) ekspresi sedih, (d) ekspresi senang, dan (e) ekspresi terkejut. 25
Gambar 4.2 Gambar suara yang telah dipotong (a) ekspresi marah, (b) ekspresi netral, (c) ekspresi sedih, (d) ekspresi senang, dan (e) ekspresi terkejut 28
Gambar 4.3 Contoh gambar suara yang telah diperoses menjadi *spectrogram* 28
Gambar 4.4 Contoh gambar yang telah dilakukan augmentasi, (a)Rescale, (b)Rotation range, (c)Shear, (d)Zoom, (e)Widht shift, (f)Hight shift, (g)Horizontal flip, (h)Fill mode 29
Gambar 4.5 Grafik akurasi dan *loss* model VGG untuk emosi wajah sebanyak 100 *epoch* 35
Gambar 4.6 Grafik akurasi dan *loss* model VGG untuk emosi wajah sebanyak 200 *epoch* 35
Gambar 4.7 Grafik akurasi dan loss model Alex Net untuk emosi wajah sebanyak 100 *epoch* 36

Gambar 4.8 Grafik akurasi dan loss model Alex Net untuk emosi wajah sebanyak 200 *epoch* 37
Gambar 4.9 Grafik akurasi dan loss 38
Gambar 4.10 Grafik akurasi dan loss model VGG untuk emosi suara sebanyak 200 epoch 38
Gambar 4.11 Grafik akurasi dan loss model Alex Net untuk emosi suara sebanyak 100 epoch 40
Gambar 4.12 Grafik akurasi dan loss model Alex Net untuk emosi suara sebanyak 200 epoch 40
Gambar 4.13 Grafik akurasi dan loss model VGG untuk emosi wajah dan suara sebanyak 100 epoch 41
Gambar 4.14 Grafik akurasi dan loss model VGG untuk emosi wajah dan suara sebanyak 200 epoch 42
Gambar 4.15 Grafik akurasi dan loss model Alex Net untuk emosi wajah dan suara sebanyak 100 epoch 43
Gambar 4.16 Grafik akurasi dan loss model Alex Net untuk emosi wajah dan suara sebanyak 200 epoch 43
Gambar 4.17 Gambar kiri ekspresi netral (terbaca netral) dan gambar kanan emosi terkejut (terbaca netral) 50

DAFTAR TABEL

Tabel 2.1 Hasil Pengenalan kelas dalam hal presisi, *recall*, *weighted*, *unweighted* dan F1_score untuk *dataset* IEMOCAP 9
Tabel 4.1 Arsitektur dan Parameter VGG 16 30
Tabel 4.2 Arsitektur dan Parameter AlexNet 33
Tabel 4.3 Akurasi model pengenalan emosi melalui wajah dan model gabungan dengan 100 epoch 44
Tabel 4.4 Akurasi model pengenalan emosi melalui wajah dan model gabungan dengan 200 epoch 45
Tabel 4.5 Hasil Pengujian Model Wajah arsitektur VGG 200 epoch 46
Tabel 4.6 Hasil Pengujian Model Gabungan (Wajah) arsitektur VGG 200 epoch 47
Tabel 4.7 Akurasi model pengenalan emosi melalui suara dan model gabungan dengan 100 epoch 51
Tabel 4.8 Akurasi model pengenalan emosi melalui suara dan model gabungan dengan 200 epoch 51
Tabel 4.9 Hasil Pengujian Model Suara Arsitektur VGG 200 epoch 52
Tabel 4.10 Hasil Pengujian Model Gabungan (Suara) Arsitektur VGG 200 epoch ... 54
Tabel 4.11 Akurasi Model Tensorflow dan Model Suara Sebelumnya 57
Tabel 4.12 Hasil Pengujian Model VGG (Suara) 200 epoch 57
Tabel 4.13 Hasil Pengujian Model Alex Net (Suara) 200 epoch 59

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Sekitar tahun 3000 SM, manusia mencoba membuat peralatan mekanika untuk mempermudah pekerjaan fisik yang membutuhkan tenaga ekstra [1]. Seiring perkembangan zaman, manusia mulai menggunakan robot untuk menyelesaikan pekerjaan kompleks tersebut. Hal tersebut ditandai dengan dibuatnya *automata* yang kompleks berbentuk hewan maupun manusia [2] pada tahun 1700-an oleh Jepang. Teknologi robot terus berkembang hingga sekarang, bahkan ada robot yang mempunyai bentuk menyerupai manusia dan dapat menggantikan manusia dalam mengerjakan sesuatu. Salah satunya adalah penemuan *NEO Robot* pada tahun 2008 [3] yang bertujuan sebagai pendamping sosial, bahan pelajaran di kelas, dan berguna untuk membantu rehabilitas anak – anak yang mengalami *autism*. Tipe robot yang menyerupai manusia tersebut digolongkan dalam kategori *humanoid robot*.

Humanoid robot memiliki banyak jenis, salah satunya adalah *Androids.* Robot ini dapat dikatakan sangat mendekati manusia karena bentuk robot *Androids* menyerupai manusia, seperti memiliki material yang mirip daging dan kulit manusia serta memiliki ekspresi dan emosi seperti manusia itu sendiri[4][3].

Kemampuan untuk membaca emosi manusia merupakan faktor penting yang perlu dimiliki oleh *humanoid robot*. Emosi merupakan hal yang umum dimiliki oleh makhluk hidup, terutama manusia. Emosi merupakan keadaan psikologis yang kompleks dengan melibatkan tiga komponen yang berbeda, yaitu pengalaman subyektif, respon fisiologis, dan respon prilaku atau ekspresif [5]. Manusia memiliki enam emosi dasar yang meliputi marah, senang, sedih, tenang, takut dan terkejut [6][7]. Untuk mendeteksi emosi manusia, ada berbagai *input* yang digunakan oleh *humanoid robot* seperti menggunakan visual, audio ataupun keduanya. Dalam

pengaplikasiannya, pengenalan emosi ke dalam robot dapat dilakukan dengan *real-time emotion recognition*[7].

Salah satu robot yang telah menggunakan *emotion recognition* pada pengaplikasiannya adalah Robot Pepper yang dikembangkan oleh *Softbank Robotics* [7]. Dengan menggunakan *AlMood*, Robot Pepper dapat mengetahui emosi seseorang melalui kamera, *microphone*, dan sensor sentuh. Penelitian lain yang dilakukan Silvia Santano Guillén, Luigi Lo Iacono, dan Christian Meder [7] menganalisis tentang akurasi pengenalan emosi otomatis pada robot *humanoid* dengan menggunakan berbagai *database*. Hasil dari penelitian ini menunjukkan bahwa akurasi yang didapatkan tidak terlalu berbeda antar *database* yang digunakan. Namun, pada penelitian selanjutnya disarankan untuk menggunakan variasi input yang lain seperti, detak jantung ataupun suara dan membuat kombinasi dari beberapa input agar akurasi output yang didapat menjadi lebih baik.

Penelitian yang dilakukan Xusheng Wang, Xing Chen, dan Congjun Cao [8] menggabungkan pengenalan emosi wajah dan pengenalan emosi suara menggunakan metode *convolutional neural network* (CNN) dan *recurrent neural network* (RNN) untuk pengenal ekspresi wajah. Lalu, penelitian tersebut menggunakan kombinasi metode *long short-term memory* (LSTM) dan CNN untuk pengenalan suara. Penelitian ini menggunakan tiga jenis *database*, yaitu RML, AFEW 6.0, dan eNTERFACE'05. Kemudian, hasil dari pengenal ekspresi wajah dan pengenal suara digabungkan dengan menggunakan *bimodal fusion* dan mendapatkan hasil akurasi yang lebih bagus daripada hanya menggunakan ekspresi wajah ataupun pengenal suara saja [8].

Penelitian-penelitian sebelumnya hanya menggunakan salah satu karakteristik yang dimiliki oleh manusia, seperti suara atau wajah saja untuk mengenali emosi. Penelitian yang menggabungkan dua karakteristik memiliki keterbatasan yaitu hanya menggunakan *dataset* sekunder [8]. Sehingga, pada penelitian ini akan dikembangkan sistem pengenalan emosi manusia dengan menggunakan kombinasi suara dan ekspresi

wajah. Metode yang akan digunakan adalah CNN yang telah menunjukkan hasil yang cukup baik pada sejumlah aplikasi [9].

1.2 Permasalahan

Penelitian yang membahas tentang pengenalan emosi umumnya hanya menggunakan salah satu ciri, seperti melalui ekspresi wajah atau intonasi suara saja. Selain itu, penelitian pengenalan emosi yang dilakukan dengan menggunakan ekspresi wajah lebih banyak dibandingkan suara. Sehingga, penelitian ini akan mengembangkan sistem pengenalan emosi menggunakan *input* kombinasi ciri wajah dan suara secara *real time* dengan menggunakan metode CNN.

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah untuk melihat unjuk kerja dari CNN untuk mengenali emosi manusia dengan menggunakan ciri berupa suara dan ekspresi wajah. Pada akhirnya, hasil penelitian ini diharapkan dapat diimplementasikan pada robot *humanoid*.

1.4 Pembatasan Masalah

Agar penelitian ini tidak melebar dari topik yang akan dibahas, maka penelitian ini memiliki batasan masalah yaitu:

1. Bahasa pemrograman yang digunakan dalam penelitian ini adalah *python*.
2. Metode yang digunakan pada penelitian ini adalah CNN tipe *VGG Neural Network*.
3. *Dataset* yang digunakan merupakan data primer.
4. Jarak maksimal untuk melakukan pengujian adalah 60 cm.
5. Emosi yang akan dibaca meliputi senang, sedih, netral, marah, dan terkejut.

1.5 Keaslian Penelitian

Ada banyak penelitian yang membahas mengenai *emotion recognition* dengan menggunakan berbagai metode maupun berbagai macam ciri input. Penelitian yang dilakukan oleh Xusheng Wang, Xing Chen, dan Congjun Cao[8] menggunakan metode pengenalan emosi ucapan berbasis fusi bimodal, dimana terdapat dua ciri berupa pengenalan ekspresi wajah dan sinyal ucapan yang terintegrasi. Pada penelitian tersebut digunakan metode penggabungan antara CNN dan RNN. Penelitian yang menggunakan algoritma fusi ini menggabungkan hasil dari ekspresi wajah dan sinyal suara untuk mengetahui emosi yang ditunjukkan oleh orang yang diperoleh dari input gambar dan suaranya. Tetapi, *dataset* suara yang digunakan memiliki banyak *noise* dari suara di latar belakang sehingga menyebabkan akurasi dalam mengenal emosi melalui suara menjadi berkurang. Hal tersebut mempengaruhi rendahnya akurasi secara keseluruhan dalam metode yang diajukan.

Penelitian lain yang dilakukan Silvia Santano Guillén, Luigi Lo Iacono, Christian Meder [7] menganalisis akurasi dari pengenalan emosi melalui ekspresi wajah pada *humanoid robot*. Hasil penelitian menunjukkan bahwa pengenalan emosi marah, sedih, dan terkejut lebih sulit untuk diidentifikasi dibandingkan ekspresi senang. Penelitian menyimpulkan bahwa tidak semua emosi dapat dikenali dengan akurasi yang sama meskipun alasan kenapa ekspresi senang mudah untuk diidentifikasi masih belum jelas. Penelitian ini menggunakan implementasi berbasis *deep convolutional neural network* (DCNN) dan solusi lainnya, seperti algoritma *Pepper*, *Google*, dan *Microsoft*. Hasil yang diperoleh DCNN tidak jauh berbeda dari solusi lainnya sehingga dapat disimpulkan bahwa algoritma DCNN memiliki kinerja yang sangat baik.

Mustaqeem, Soonil Kwon [10] meneliti pengenalan emosi melalui suara dengan metode *end-to-end MLT-SER system* berdasarkan 1D *Dilated CNN*. Penelitian ini bertujuan untuk mengenali dan mempelajari fitur dari emosi dari penduduk lokal maupun global melalui sinyal suara secara otomatis. *Dataset* yang digunakan untuk

adalah *Interactive emotional dyadic motion captures* (IEMOCAP) dan EMO-DB SER (*Berlin emotion dataset*) dan akurasi yang didapat untuk masing-masing *dataset* adalah 73% dan 90%. Namun, penelitian ini memiliki keterbatasan hanya diimplementasikan pada data sekunder.

Pada penelitian yang dilakukan oleh Alexandr Rassadin, Alexey Gruzdev, dan Andrey Savchenko [11] digunakan metode *random forest classifiers.* Peneliti menggunakan algoritma tradisonal, yaitu *Viola-Jones cascade classifiers* berdasarkan pada fitur *Haar* dari *library* OpenCV dan *histogram of oriented gradients* (HOG) dari *library* DLIB untuk menemukan wajah dalam sebuah foto grup dan menggunakan CNN untuk mengidentifikasi wajah. Kemudian, keputusan menentukan emosi yang ditampilkan oleh wajah dilakukan dengan menggunakan metode *random forest classifiers*. Hasil yang diperoleh dari penelitian tersebut adalah 75,4% akurasi dalam pengklasifikasian wajah dan 78,53% akurasi dalam pengujiannya. Namun dalam penerapan metode yang diajukan, metode ini bekerja lebih baik pada kelas negatif dan positif daripada kelas netral. Hal ini dikarenakan emosi kelompok yang netral jauh lebih sulit didefinisikan.

Penelitian lain yang dilakukan Aggelina Chatziagapi, Georgios Paraskevopoulos, Dimitris Sgouropoulos, Georgios Pantazopoulos dan yang lainnya [12] menggunakan arsitektur *Generative Adversarial Network* (GAN) untuk mengatasi ketidakseimbangan data pada pengenalan emosi melalui suara, dengan cara menambahkan *class* yang kurang terwakili. Metode yang diajukan ini dapat membuat performa dari pengenalan emosi dengan menggunakan *dataset* IEMOCAP dan FEEL-25*k* meningkat antara 5% sampai 10%.

DAFTAR PUSTAKA

[1] D. L. Gera, *Ancient Greek Ideas on Speech, Language, and Civilization*. 2010.

[2] J. M. Law, *Puppets of nostalgia: The life, death, and rebirth of the Japanese Awaji ningyo tradition*. 2015.

[3] M. A. Miskam, S. Shamsuddin, H. Yussof, I. M. Ariffin, and A. R. Omar, "A questionnaire-based survey: Therapist's response on emotions gestures using humanoid robot for autism," 2016, doi: 10.1109/MHS.2015.7438298.

[4] H. Kumazaki *et al.*, "Android robot-mediated mock job interview sessions for young adults with autism spectrum disorder: A pilot study," *Front. Psychiatry*, vol. 8, no. SEP, 2017, doi: 10.3389/fpsyt.2017.00169.

[5] D. H. Hockenbury and S. E. Hockenbury, *Discovering psychology, 4th ed.* 2007.

[6] L. Chen, W. Su, Y. Feng, M. Wu, J. She, and K. Hirota, "Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction," *Inf. Sci. (Ny).*, vol. 509, pp. 150–163, 2020, doi: 10.1016/j.ins.2019.09.005.

[7] S. Santano Guillén, L. Lo Iacono, and C. Meder, "Affective Robots: Evaluation of Automatic Emotion Recognition Approaches on a Humanoid Robot towards Emotionally Intelligent Machines Blockchain-driven Collaborative Supply Chains (BC-SC) View project USecureD-Usable Security by Design View project Affect," *World Acad. Sci. Eng. Technol. Int. J. Mech. Mechatronics Eng.*, vol. 12, no. 6, pp. 584–592, 2018, [Online]. Available: https://www.researchgate.net/publication/325688779.

[8] X. Wang, X. Chen, and C. Cao, "Human emotion recognition by optimally fusing facial expression and speech feature," *Signal Process. Image Commun.*,

vol. 84, no. August 2019, p. 115831, 2020, doi: 10.1016/j.image.2020.115831.

[9] M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh, and A. Akan, "Real time emotion recognition from facial expressions using CNN architecture," *TIPTEKNO 2019 - Tip Teknol. Kongresi*, pp. 529–532, 2019, doi: 10.1109/TIPTEKNO.2019.8895215.

[10] Mustaqeem and S. Kwon, "MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach," *Expert Syst. Appl.*, vol. 167, no. October, p. 114177, 2021, doi: 10.1016/j.eswa.2020.114177.

[11] A. Rassadin, A. Gruzdev, and A. Savchenko, "Group-level emotion recognition using transfer learning from face identification," *arXiv*, pp. 3–7, 2017.

[12] A. Chatziagapi *et al.*, "Data augmentation using GANs for speech emotion recognition," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2019-Septe, pp. 171–175, 2019, doi: 10.21437/Interspeech.2019-2561.

[13] Z. A. Khan, T. Hussain, A. Ullah, S. Rho, M. Lee, and S. W. Baik, "Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework," *Sensors (Switzerland)*, vol. 20, no. 5, 2020, doi: 10.3390/s20051399.

[14] L. P. Purnamaningsih, N. K. Suarni, and K. Suranata, "Identifikasi Emosi Melalui Pendeteksian Karakteristik Ekspresi Wajah (Face Expresion) Dalam Rangka Mengentaskan Masalah Siswa Melalui Konseling Individual," p. 2, 2013.

[15] M. Sambare, "FER-2013," 2013. https://www.kaggle.com/msambare/fer2013.

[16] S. M. S. A. Abdullah, S. Y. A. Ameen, M. A. M. Sadeeq, and S. Zeebaree, "Multimodal Emotion Recognition using Deep Learning," *J. Appl. Sci.*

Technol. Trends, vol. 2, no. 02, pp. 52–58, 2021, doi: 10.38094/jastt20291.

[17] M. Lech, M. Stolar, C. Best, and R. Bolia, "Real-Time Speech Emotion Recognition Using a Pre-trained Image Classification Network: Effects of Bandwidth Reduction and Companding," *Front. Comput. Sci.*, vol. 2, no. May, pp. 1–14, 2020, doi: 10.3389/fcomp.2020.00014.

[18] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Commun.*, vol. 116, no. October 2019, pp. 56–76, 2020, doi: 10.1016/j.specom.2019.12.001.

[19] L. O. H. Sagala and A. Harjoko, "Perbandingan Ekstraksi Ciri Full, Blocks, dan Row Mean Spectrogram Image Dalam Mengidentifikasi Pembicara," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 10, no. 1, p. 155, 2014, doi: 10.22146/ijccs.6543.

[20] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way," 2018. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[21] N. Milosevic, "Introduction to Convolutional Neural Networks," *Introd. to Convolutional Neural Networks*, pp. 1–31, 2020, doi: 10.1007/978-1-4842-5648-0.

[22] H. Yonekawa, S. Sato, and H. Nakahara, "A ternary weight binary input convolutional neural network: Realization on the embedded processor," *Proc. Int. Symp. Mult. Log.*, vol. 2018-May, no. August, pp. 174–179, 2018, doi: 10.1109/ISMVL.2018.00038.

[23] M. K. Pichora-Fuller and K. Dupuis, "Toronto emotional speech set (TESS)." Scholars Portal Dataverse, 2020, doi: 10.5683/SP2/E8H2MF.

[24] B. Mcfee *et al.*, “Librosa - audio processing Python library,” *Proc. 14th python Sci. Conf.*, vol. 8, no. Scipy, 2015.