

**IMPLEMENTASI KOMBINASI ALGORITMA CNN 1 DAN 2  
DIMENSI PADA *SPEECH EMOTION RECOGNITION***

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**MUKHLISHIN M ARSYAD**

**20.11.3325**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2025**

**IMPLEMENTASI KOMBINASI ALGORITMA CNN 1 DAN 2  
DIMENSI PADA *SPEECH EMOTION RECOGNITION***

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi *Informatika*



disusun oleh  
**MUKHLISHIN M ARSYAD**  
**20.11.3325**

Kepada

**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS AMIKOM YOGYAKARTA**  
**YOGYAKARTA**  
**2025**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**IMPLEMENTASI KOMBINASI ALGORITMA CNN 1 DAN 2 DIMENSI  
PADA *SPEECH EMOTION RECOGNITION***

yang disusun dan diajukan oleh

**MUKHLISHIN M ARSYAD**

**20.11.3325**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 24 Februari 2025

Dosen Pembimbing,



**Arif Akbarul Huda, S.Si.M.Eng.**  
**NIK. 190302287**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**IMPLEMENTASI KOMBINASI ALGORITMA CNN 1 DAN 2 DIMENSI**  
**PADA *SPEECH EMOTION RECOGNITION***

yang disusun dan diajukan oleh

**MUKHLISHIN M ARSYAD**

20.11.3325

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 24 Februari 2025

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

Anna Baita, S.Kom., M.Kom.  
NIK. 190302290

Novi Prisma Yunita, M.Kom  
NIK. 190302526

Arif Akbarul Huda, S.Si., M.Eng.  
NIK. 190302287



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 24 Februari 2025

**DEKAN FAKULTAS ILMU KOMPUTER**



Prof. Dr. Kusriani, M.Kom.  
NIK. 190302106

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : MUKHLISHIN M ARSYAD  
NIM : 20.11.3325

Menyatakan bahwa Skripsi dengan judul berikut:

### IMPLEMENTASI KOMBINASI ALGORITMA CNN 1 DAN 2 DIMENSI PADA SPEECH EMOTION RECOGNITION

Dosen Pembimbing : Arif Akbarul Huda, S.Si.M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 7 Februari 2025

Yang Menyatakan,

  
Mukhlishin M Arsyad

## KATA PENGANTAR

Puji syukur kepada Allah SWT Tuhan Yang Maha Esa, yang telah memberikan kekuatan, petunjuk, serta kelancaran dalam proses pengerjaan skripsi sehingga dapat terselesaikan dengan tepat waktu.

Penulisan skripsi yang berjudul “IMPLEMENTASI KOMBINASI ALGORITMA CNN 1 DAN 2 DIMENSI PADA SPEECH EMOTION RECOGNITION” merupakan salah satu tonggak sejarah dalam perjalanan akademik bagi penulis serta menjadi salah satu persyaratan dalam menempuh gelar strata 1 (S-1) Program Studi Informatika.

Dalam melakukan penelitian dan penulisan skripsi ini, penulis dengan sadar bahwasanya tanpa dukungan dan bantuan dari berbagai pihak penelitian ini tidak dapat berjalan secara maksimal. Maka dari itu, penulis ingin menyampaikan rasa terima kasih kepada:

1. Allah SWT atas rahmat dan karunia-Nya sehingga penelitian ini dapat berjalan dengan lancar
2. Ibu dan Bapak serta saudara-saudara penulis yang selalu mendukung dan mendukung baik secara moril maupun materi kepada penulis serta yang paling penting adalah memberikan doa untuk penulis. Terima kasih banyak atas semuanya sehingga penulis dapat sampai pada titik ini.
3. Bapak Arif Akbarul Huda, S.Si.M.Eng. yang telah bersedia membimbing dan berdiskusi serta memberikan pengertian kepada penulis. Terima kasih banyak karena sudah bersedia juga untuk meluangkan waktu kepada penulis dalam penyusunan skripsi ini.
4. Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta
5. Hanif Al Fatta, S.Kom., M.Kom., Ph.D. selaku Dekan Fakultas Ilmu Komputer sebelumnya dan Prof. Dr. Kusrini, M.Kom. selaku Dekan Fakultas Ilmu Komputer saat ini di Universitas Amikom Yogyakarta
6. Windha mega Pradnya Dhuhita, M.Kom. selaku Kaprodi Informatika Universitas Amikom Yogyakarta

7. Bapak/Ibu Dosen pada Program Studi Informatika Universitas Amikom Yogyakarta
8. Ibu Ike Verawati yang secara tidak langsung membuat saya kembali menyukai matematika dengan segala macam perhitungannya dan juga Ibu Zaenab yang juga mengarahkan kemana sebaiknya penelitian ini.
9. Teman-teman senior di asrama Babasal yang sudah penulis anggap sebagai kakak-kakak penulis sendiri. Terima kasih banyak atas segala kebersamaan dan cerita-ceritanya.
10. Sahabat-sahabat terdekat penulis yaitu Asrul sances, pak sudarmono, yayi S.Ikom, jiran tato, yayang koyo, nopal arab dan beberapa yang tidak dapat disebutkan telah membantu peneliti baik secara langsung dan tidak langsung. Jasa kalian abadi.
11. Teman-teman penulis di kelas 20-S1IF01 yang selama ini telah saling membantu sama lain. Terima kasih dan semoga kita dapat bertemu di masa depan yang cerah dan gemilang.
12. Semua pihak yang telah membantu penulis yang telah membantu baik secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan satu per satu. Penulis mengucapkan banyak-banyak terima kasih semoga kehidupan kalian semua dilancarkan.
13. Mukhlisin, orang yang sudah berjuang dengan hebat walaupun kadang sangat malas namun berhasil sampai titik ini. Mari bersama kita lebih giat belajar, konsisten, pergi ke tempat-tempat baru, wujudkan masa depan yang gemilang, dan temukan semua O yang menggajal pikiran kita selama ini serta mari kita hadapi segala rintangan yang akan datang.

Yogyakarta, 7 Februari 2025

Penulis

## DAFTAR ISI

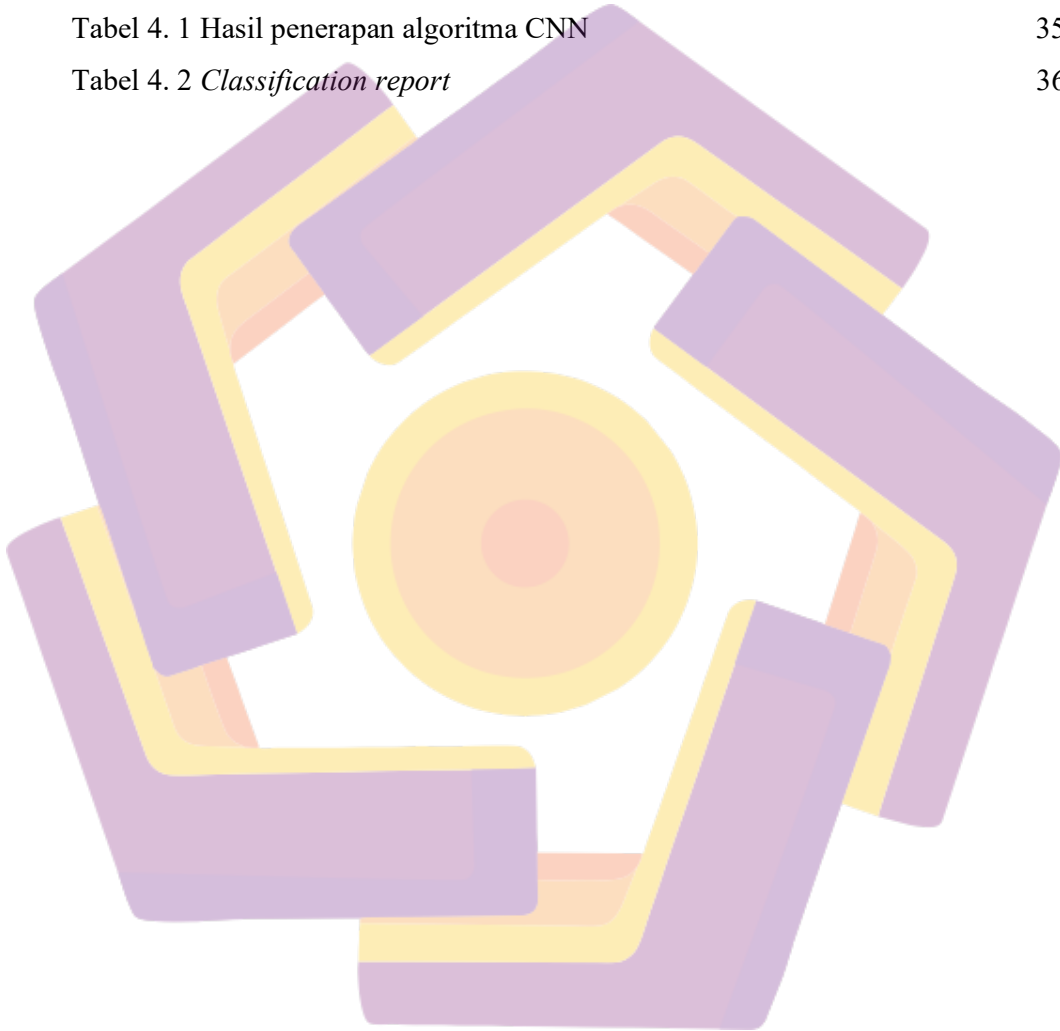
|                                                 |      |
|-------------------------------------------------|------|
| HALAMAN JUDUL                                   | i    |
| HALAMAN PERSETUJUAN                             | ii   |
| HALAMAN PENGESAHAN                              | iii  |
| HALAMAN PERNYATAAN KEASLIAN SKRIPSI             | iv   |
| KATA PENGANTAR                                  | v    |
| DAFTAR ISI                                      | vii  |
| DAFTAR TABEL                                    | x    |
| DAFTAR GAMBAR                                   | xi   |
| DAFTAR LAMPIRAN                                 | xii  |
| DAFTAR LAMBANG DAN SINGKATAN                    | xiii |
| DAFTAR ISTILAH                                  | xiv  |
| INTISARI                                        | xv   |
| <i>ABSTRACT</i>                                 | xvi  |
| BAB I PENDAHULUAN                               | 1    |
| 1.1 Latar Belakang                              | 1    |
| 1.2 Rumusan Masalah                             | 3    |
| 1.3 Batasan Masalah                             | 3    |
| 1.4 Tujuan Penelitian                           | 3    |
| 1.5 Manfaat Penelitian                          | 4    |
| 1.6 Sistematika Penulisan                       | 4    |
| BAB II TINJAUAN PUSTAKA                         | 6    |
| 2.1 Studi Literatur                             | 6    |
| 2.2 Dasar Teori                                 | 13   |
| 2.2.1 <i>Convolutional Neural Network (CNN)</i> | 13   |

|                             |                                                         |    |
|-----------------------------|---------------------------------------------------------|----|
| 2.2.2                       | Speech Emotion Recognition                              | 14 |
| 2.2.3                       | Data Segmentation                                       | 14 |
| 2.2.4                       | Data Augmentation                                       | 15 |
| 2.2.5                       | Ekstraksi Fitur                                         | 15 |
| 2.2.6                       | Dimensi                                                 | 19 |
| 2.2.7                       | Label Encoding                                          | 19 |
| BAB III METODE PENELITIAN   |                                                         | 20 |
| 3.1                         | Objek Penelitian                                        | 20 |
| 3.2                         | Alur Penelitian                                         | 20 |
| 3.2.1                       | Studi Literatur                                         | 21 |
| 3.2.2                       | Pengumpulan Data                                        | 21 |
| 3.2.3                       | Proses pengolahan data                                  | 22 |
| 3.2.4                       | Perangkaian Model CNN kombinasi 1 dan 2 dimensi         | 22 |
| 3.2.5                       | <i>Train, Validation and Testing Model</i>              | 24 |
| 3.2.6                       | Implementasi ke Application Programming Interface (API) | 24 |
| 3.3                         | Alat dan Bahan                                          | 25 |
| 3.3.1                       | Hardware                                                | 25 |
| 3.3.2                       | Software                                                | 26 |
| BAB IV HASIL DAN PEMBAHASAN |                                                         | 28 |
| 4.1                         | Implementasi                                            | 28 |
| 4.1.1                       | Import Library                                          | 28 |
| 4.1.2                       | <i>Variable Declare</i>                                 | 29 |
| 4.1.3                       | <i>Pre-processing &amp; Exploratory Data Analysis</i>   | 29 |
| 4.1.4                       | <i>Processing</i>                                       | 30 |
| 4.1.5                       | <i>Feature Extraction</i>                               | 31 |

|               |                               |    |
|---------------|-------------------------------|----|
| 4.1.6         | Pembuatan kombinasi model     | 32 |
| 4.1.7         | Penerapan Model ke dalam API  | 32 |
| 4.2           | Hasil                         | 33 |
| 4.2.1         | Rangkaian Model Kombinasi CNN | 33 |
| 4.2.2         | CNN 1 dimensi                 | 33 |
| 4.2.3         | CNN 2 dimensi                 | 33 |
| 4.2.4         | CNN Kombinasi 1 dan 2 dimensi | 34 |
| 4.3           | Evaluasi                      | 35 |
| BAB V PENUTUP |                               | 37 |
| 5.1           | Kesimpulan                    | 37 |
| 5.2           | Saran                         | 38 |
| REFERENSI     |                               | 39 |
| LAMPIRAN      |                               | 41 |

## DAFTAR TABEL

|                                                |    |
|------------------------------------------------|----|
| Tabel 2. 1 Keaslian Penelitian                 | 10 |
| Tabel 3. 1 Data yang akan diproses             | 21 |
| Tabel 3. 2 Spesifikasi Hardware yang digunakan | 26 |
| Tabel 4. 1 Hasil penerapan algoritma CNN       | 35 |
| Tabel 4. 2 <i>Classification report</i>        | 36 |



## DAFTAR GAMBAR

|                                                                                                     |    |
|-----------------------------------------------------------------------------------------------------|----|
| Gambar 2. 1 Contoh CNN 1 dimensi                                                                    | 13 |
| Gambar 2. 2 Contoh CNN 2 Dimensi                                                                    | 13 |
| Gambar 2. 3 Penerapan <i>Speech Emotion Recognition</i>                                             | 14 |
| Gambar 2. 4 Ilustrasi gelombang yang dibiaskan pada prisma                                          | 15 |
| Gambar 2. 5 Mengubah audio menjadi spectrogram dengan metode stft                                   | 17 |
| Gambar 2. 6 Visualisasi koefisien dari MFCC                                                         | 18 |
| Gambar 2. 7 Visualisasi frekuensi pada setiap nada dari waktu ke waktu                              | 18 |
| Gambar 2. 8 Dimensi data                                                                            | 19 |
| Gambar 2. 9 Contoh pelabelan encoding                                                               | 19 |
| Gambar 3. 1 Alur penelitian                                                                         | 20 |
| Gambar 3. 2 Alur proses pengolahan data                                                             | 22 |
| Gambar 3. 3 Rangkaian model kombinasi CNN                                                           | 23 |
| Gambar 3. 4 Alur simulasi penerapan API                                                             | 25 |
| Gambar 4. 1 Alur program.....                                                                       | 28 |
| Gambar 4. 2 Distribusi data berdasarkan label .....                                                 | 29 |
| Gambar 4. 3 Distribusi Audio berdasarkan durasi.....                                                | 30 |
| Gambar 4. 4 Contoh visualisasi Audio dari setiap label.....                                         | 30 |
| Gambar 4. 5 Grafik akurasi <i>training</i> dan <i>validation</i> CNN 2 dimensi.....                 | 33 |
| Gambar 4. 6 Grafik akurasi <i>training</i> dan <i>validation</i> CNN 2 dimensi.....                 | 34 |
| Gambar 4. 7 Grafik akurasi <i>training</i> dan <i>validation</i> kombinasi CNN 1 dan 2 dimensi..... | 34 |
| Gambar 4. 8 <i>Confusion matrix</i> .....                                                           | 35 |
| Gambar 4. 9 Potongan klip film dan respon dari API dalam bentuk JSON .....                          | 36 |

## DAFTAR LAMPIRAN

|                                              |    |
|----------------------------------------------|----|
| Lampiran 1. <i>Library import code</i>       | 41 |
| Lampiran 2. <i>Declare variable code</i>     | 41 |
| Lampiran 3. <i>Processing &amp; EDA code</i> | 41 |
| Lampiran 4. <i>Label encoding code</i>       | 42 |
| Lampiran 5. <i>Segmentation code</i>         | 42 |
| Lampiran 6. <i>Augmentation code</i>         | 42 |
| Lampiran 7. <i>Features extraction code</i>  | 43 |
| Lampiran 8. <i>Combined model code</i>       | 43 |
| Lampiran 9. <i>API implement code</i>        | 44 |
| Lampiran 10. <i>Rangkaian model</i>          | 45 |

## DAFTAR LAMBANG DAN SINGKATAN



|          |                                                                                           |
|----------|-------------------------------------------------------------------------------------------|
| CNN      | <i>Convolutional Neural Network</i>                                                       |
| 1D       | <i>1 Dimension</i>                                                                        |
| 2D       | <i>2 Dimension</i>                                                                        |
| EDA      | <i>Exploratory Data Analysis</i>                                                          |
| SVM      | <i>Support Vector Machines</i>                                                            |
| SER      | <i>Speech <b>E</b>motion Recognition</i>                                                  |
| RAVDESS  | <i>Ryerson Audio-Visual Database of Emotional Speech and Song</i>                         |
| ASVP-ESD | <i>Audio, Speech, and Vision Processing Lab Emotional Sound database</i>                  |
| TESS     | <i>Toronto emotional speech set</i>                                                       |
| AA-CBGRU | <i>Attention <b>A</b>ugmented Convolutional <b>B</b>idirectional Gated Recurrent Unit</i> |
| GLAM     | <i>Global-Aware <b>M</b>ulti-scale</i>                                                    |
| AI       | <i>Artificial Intelligence</i>                                                            |
| IEMOCAP  | <i>The Interactive Emotional Dyadic Motion Capture</i>                                    |
| MFCC     | <i>Mel-Frequency Cepstral Coefficients</i>                                                |
| STFT     | <i>Short-Time Fourier Transformation</i>                                                  |
| MLP      | <i>Multi-Layer <b>P</b>erptron</i>                                                        |
| API      | <i>Application Programming <b>I</b>nterface</i>                                           |
| JSON     | <i><b>J</b>avascript <b>O</b>bject <b>N</b>otation</i>                                    |
| LSTM     | <i>Long short-term memory</i>                                                             |
| API      | <i>Application Programming Interface</i>                                                  |

## DAFTAR ISTILAH

*Speech Emotion Recognition*

Sistem yang dapat mengenali emosi secara otomatis dengan menggunakan *Machine Learning*

*Feature Extraction*

Tahap mengeluarkan / mengambil informasi penting yang berada pada data

*Augmentation*

Proses memperkaya data dengan mengubah data dan dibuat salinannya tanpa mengubah data asli

Konvolusi

Salah satu operasi matematika yang digunakan untuk menggabungkan dua fungsi yang akan membentuk sebuah fungsi baru

*Neural Network*

Sistem komputasi yang meniru bagaimana cara kerja dari otak manusia

Segmentasi

Memecah data yang ada pada hal ini audio agar memiliki panjang durasi yang sama

*Overfitting*

Kondisi model AI terlalu ideal pada data *training* namun tidak pada data baru

*Spectral*

Representasi audio yang terkandung pada suara

*Fourier Transform*

Sebuah teknik atau persamaan pada matematika yang dapat mengubah sinyal dari domain waktu ke domain frekuensi

## INTISARI

Pengenalan emosi dari ucapan atau biasa disebut *Speech Emotion Recognition (SER)* adalah salah satu bidang dari kecerdasan buatan yang penelitiannya saat ini sedang berkembang. Penerapan dari teknologi ini cukup luas. Beberapa penelitian telah dilakukan dan memberikan hasil yang signifikan. Namun, teknologi ini memiliki beberapa tantangan yang cukup serius dikarenakan karakteristik suara manusia yang variatif, juga faktor kebisingan atau *noise*. Selain itu, beberapa hal teknis menjadi tantangan dalam pengembangan teknologi ini seperti pemilihan fitur ekstraksi, *augmentation*, dan penanganan perbedaan durasi dari *dataset* yang ada. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan Algoritma *Convolutional Neural Network (CNN)* namun menggunakan pendekatan yang berbeda yaitu dengan menggabungkan hasil ekstraksi 1 dan 2 dimensi dari *dataset* yang ada. Pada penelitian ini terdapat 3 *dataset* yang umumnya sudah dipakai oleh beberapa penelitian yang sudah pernah ada sebelumnya. Beberapa *dataset* itu adalah *Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)*, *The Audio, Speech, and Vision Processing Lab Emotional Sound database (ASVP-ESD)*, *Toronto emotional speech set (TESS)* yang berisi rekaman suara dengan berbagai macam emosi yang sudah dilabeli dengan emosi itu sendiri. Hasil pengujian yang didapatkan setelah dilakukan implementasi dan pengujian dengan CNN 1 dan 2 dimensi didapatkan akurasi mencapai lebih dari 90%.

**Kata kunci:** Pengenalan emosi dari ucapan, *Deep Learning*, *Convolutional Neural Network*, Ekstraksi Fitur, Kombinasi dimensi

## ***ABSTRACT***

*Speech Emotion Recognition (SER) is one of the fields of artificial intelligence that the research is currently developing. The application of this technology is quite broad. several studies have been conducted and provided significant results. However, this technology has some serious challenges due to the varied characteristics of the human voice, as well as the noise factor. Additionally, some technical issues become a challenge in the development of this technology such as the selection of feature extraction, augmentation, and handling duration differences from existing datasets. Therefore, this research aims to implement the Convolutional Neural Network (CNN) algorithm but using a different approach by combining the results of 1 and 2-dimensional extraction from existing datasets. In this research, there are 3 datasets that have generally been used by several previous studies. Some of the datasets are Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), The Audio, Speech, and Vision Processing Lab Emotional Sound database (ASVP-ESD), Toronto emotional speech set (TESS) which contains voice recordings with various emotions that have been labeled with the emotion itself. The results obtained after implementation and testing with CNN 1 and 2 dimensions obtained accuracy reached more than 90%.*

**Keyword:** *Speech Emotion Recognition, Deep Learning, Convolutional Neural Network, Features Extraction, Combined Dimension*