

## Klasifikasi dan Pengenalan Emosi dari Ekspresi Wajah Menggunakan CNN-BiLSTM dengan Teknik *Data Augmentation*

Sutarti<sup>1\*</sup>, Fariza Syaqqialloh<sup>2</sup>

<sup>1</sup>Program Studi Sistem Komputer, Universitas Serang Raya, Indonesia.

---

### Artikel Info

#### Kata Kunci:

CNN-BiLSTM;  
Data Augmentation;  
Facial Expression Recognition;  
Performa.

#### Keywords:

CNN-BiLSTM;  
Data Augmentation;  
Facial Expression Recognition;  
Performance.

---

#### Riwayat Artikel:

Submitted: 14 Januari 2025

Accepted: 10 Maret 2025

Published: 18 Maret 2025

**Abstrak:** Teknologi *computer vision* terus berkembang dan diterapkan dalam berbagai bidang, termasuk interaksi manusia dan komputer (*human-computer interaction*), khususnya dalam pengenalan emosi dari ekspresi wajah atau *facial expression recognition* (FER). Penelitian ini bertujuan untuk mengembangkan model CNN-BiLSTM yang dilengkapi dengan *data augmentation* untuk mendeteksi emosi seperti marah, bahagia, takut, sedih, terkejut, jijik, dan netral. *Dataset* yang digunakan dalam penelitian ini adalah *dataset* publik FER2013, yang terdiri dari 35.887 gambar wajah *grayscale* berukuran 48x48 piksel. Teknik augmentasi data meliputi rotasi, *flipping*, dan pergeseran gambar untuk meningkatkan keragaman data dan mencegah *overfitting*. Model dilatih menggunakan optimizer Adam, dengan *hyperparameter* berupa batch size sebesar 32, jumlah *epoch* sebanyak 100, dan *learning rate* 0.001. Hasil pengujian menunjukkan bahwa model CNN-BiLSTM mencapai akurasi validasi sebesar 95%, lebih tinggi dibandingkan CNN yang hanya mencapai akurasi 70%. Selain itu, *validation loss* pada CNN-BiLSTM lebih stabil, menunjukkan kemampuan generalisasi yang lebih baik dibandingkan CNN, yang mengalami fluktuasi signifikan akibat kemungkinan *overfitting* dan kurang optimalnya parameter seperti *learning rate*. Analisis menggunakan *confusion matrix* menunjukkan bahwa CNN-BiLSTM menghasilkan klasifikasi yang lebih akurat untuk setiap emosi. Model CNN-BiLSTM memberikan kinerja lebih unggul dalam mendeteksi emosi pada FER, terutama dalam menghadapi tantangan seperti pencahayaan, posisi wajah, dan kompleksitas ekspresi.

**Abstract:** *Computer vision technology continues to evolve and is applied in various fields, including human-computer interaction (HCI), particularly in recognizing emotions from facial expressions (facial expression recognition, FER). This study aims to develop a CNN-BiLSTM model enhanced with data augmentation to detect emotions such as anger, happiness, fear, sadness, surprise, disgust, and neutral. The dataset used in this research is the publicly available FER2013 dataset, consisting of 35,887 grayscale facial images with a resolution of 48x48 pixels. The data augmentation techniques employed include rotation, flipping, and shifting to increase data diversity and prevent overfitting. The model was trained using the Adam optimizer, with hyperparameters including a batch size of 32, 100 epochs, and a learning rate of 0.001. The test results indicate that the CNN-BiLSTM model achieved a validation accuracy of 95%, which is higher than the CNN model's 70% accuracy. Additionally, the validation loss of the CNN-BiLSTM model is more stable, demonstrating better generalization capability compared to the CNN model, which exhibited significant fluctuations due to potential overfitting and suboptimal parameters such as the learning rate. Analysis using a confusion matrix reveals that the CNN-BiLSTM model produces more accurate classification results for each emotion. The CNN-BiLSTM model outperforms the CNN model in detecting emotions in FER, particularly when dealing with*

---

*challenges such as lighting conditions, facial positions, and expression complexity.*

---

**Corresponding Author:**

Sutarti

Email: sutarti86@gmail.com

---

## PENDAHULUAN

Teknologi yang menerapkan *computer vision* semakin berkembang, salah satunya pada interaksi manusia dan komputer (*human-computer interaction*), khususnya pada pengenalan emosi dari ekspresi wajah (Canedo & Neves, 2019). Ekspresi wajah merupakan salah satu bentuk komunikasi nonverbal yang paling umum digunakan oleh manusia. Ekspresi wajah dapat digunakan untuk mengungkapkan berbagai emosi, seperti bahagia, sedih, marah, terkejut, dan takut. Pengenalan emosi berdasarkan ekspresi wajah merupakan salah satu bidang penelitian yang berkembang pesat dalam beberapa tahun terakhir (Almeida, Vilaça, Teixeira, & Viana, 2021). Penelitian ini memiliki berbagai aplikasi, seperti pemrosesan bahasa alami (NLP) dimana ekspresi wajah dapat digunakan untuk meningkatkan akurasi NLP, seperti pengenalan emosi dalam teks (Bursic, Boccignone, Ferrara, & Amelio, 2020). Pada kecerdasan buatan (AI), ekspresi wajah dapat digunakan untuk mengembangkan sistem AI yang lebih cerdas dan berempati, seperti chatbot dan asisten virtual. Pada bidang kesehatan ekspresi wajah dapat digunakan untuk mendeteksi gangguan mental, seperti depresi dan autisme (Meyer-Lindenberg et al., 2022). Model CNN-BiLSTM dengan *attention mechanism* menunjukkan peningkatan akurasi yang signifikan dibandingkan metode pengenalan wajah yang hanya menggunakan CNN ((Qi et al., 2023). Akurasi dan keandalan model ditingkatkan menggunakan *attention mechanism* yang memprioritaskan fitur penting untuk prediksi.

Metode dan algoritma pengenalan wajah sangat beragam dan sulit untuk menentukan algoritma terbaik untuk dipilih. Perbandingan akurasi dari beberapa algoritma dilakukan untuk mengetahui kinerja yang lebih baik. CNN mempunyai akurasi yang lebih baik dibandingkan LBPH (Budiman, Fabian, Yaputera, Achmad, & Kurniawan, 2022). MobilenetV2 mempunyai kemampuan mengenali wajah dengan akurasi 78% (Adiatma, Utami, & Hartanto, 2021). Untuk meningkatkan akurasi CNN dapat dilakukan dengan cara menggabungkan dua jenis input (Avtar & Dhingra, 2023). Selain menggunakan algoritma tunggal, solusi lain yang bisa diterapkan adalah menggabungkan beberapa algoritma. Penggabungan *machine learning* (KNN dan SVM) dan *deep learning* (CNN dan VGG16) dapat mengenali wajah sekaligus jenis kelamin serta umur (Teja Chavali, Tej Kandavalli, Sugash, & Subramani, 2022). Model deep learning VGG16 dan MTCNN mampu mengenali emosi wajah (Widodo, Setiawan, Ridwan, & Ambari, 2022) demikian juga dengan BAM-DCNN (G & Krishnaveni, 2023). Meningkatkan akurasi CNN bisa menerapkan proses ekstraksi fitur (Gautam & Seeja, 2023). Hasil penelitian menunjukkan bahwa untuk CKplus dataset. Sedangkan pada Jaffe dataset akurasi HOG-CNN sebesar 91.43%, SIFT-CNN sebesar 82.85%.

Sistem deteksi emosi juga dapat diimplementasikan menggunakan metode *Attentional Convolutional Network*. Sistem ini menggunakan teknik *attention* untuk meningkatkan akurasi deteksi emosi. *Attention* adalah teknik yang dapat membantu model untuk fokus pada fitur-fitur yang paling penting untuk prediksi. Hasil penelitian menunjukkan bahwa metode yang menggunakan *attention* dapat meningkatkan akurasi pengenalan emosi sebesar 3% dibandingkan dengan metode yang tidak menggunakan *attention*. Penerapan yang dilakukan menggunakan mekanisme *attention* dengan arsitektur CNN-ALSTM dan ACNN-ALSTM pada dataset Fer2013 CK+ (Ming, Qian, & Guangyuan, 2022). Hasil penelitian menunjukkan bahwa model jaringan saraf hibrid ACNN-ALSTM yang diusulkan lebih unggul dibandingkan metode tunggal.

*Ensemble learning* dan *transfer learning* merupakan alternatif lain yang dapat dikombinasikan dengan CNN-BiLSTM. Penggabungan CNN-BiLSTM dengan *ensemble learning* menawarkan beberapa

kelebihan dalam bidang *emotion recognition* seperti peningkatan akurasi, peningkatan ketahanan terhadap gangguan, peningkatan generalisasi, pengurangan *overfitting*, peningkatan dan kemampuan penangkapan fitur yang kompleks. *Ensemble learning* dan *transfer learning* merupakan dua pendekatan yang berbeda untuk meningkatkan akurasi dan ketahanan terhadap gangguan dari sistem pengenalan emosi (Meena Jindal & Khushwant Kaur, 2024). *Ensemble learning* menggabungkan prediksi dari beberapa model yang berbeda, sedangkan *transfer learning* memanfaatkan pengetahuan yang diperoleh dari *dataset* yang sudah ada. Secara umum, *ensemble learning* lebih efektif dalam meningkatkan akurasi, sedangkan *transfer learning* lebih efektif dalam meningkatkan efisiensi dan generalisasi. Namun, kedua pendekatan tersebut dapat digabungkan untuk mendapatkan hasil yang lebih optimal. Teknik *transfer learning* dapat digunakan untuk meningkatkan akurasi deteksi emosi, terutama jika data pelatihan yang tersedia terbatas (Shimamura, 2021).

Pengenalan emosi berbasis wajah. Penelitian tersebut menggunakan dataset FER2013 yang berisi 35.887 gambar wajah dengan 7 kelas emosi. Dataset ini tidak seimbang, karena jumlah gambar per kelas sangat bervariasi. Oleh karena itu diperlukan *data augmentation* untuk memberikan tambahan variasi data. *Data augmentation* menggunakan teknik *rotation* atau *flip* untuk meningkatkan keragaman pose wajah. *Data augmentation* dapat meningkatkan performa model dan untuk mencegah *overfitting* (Paraskevopoulou, Spyrou, & Perantonis, 2022). CNN dan BiLSTM merupakan dua model pembelajaran mendalam yang dapat digunakan untuk pengenalan emosi. CNN model yang kuat untuk mengekstrak fitur visual dari gambar, sedangkan BiLSTM model yang kuat untuk mengolah informasi temporal dari video. Untuk menggabungkan CNN dan BiLSTM untuk meningkatkan akurasi pengenalan emosi, dapat dilakukan dengan cara menggunakan CNN untuk mengekstrak fitur visual dari gambar wajah, kemudian menggunakan BiLSTM untuk mengolah informasi temporal dari fitur visual tersebut atau menggunakan CNN dan BiLSTM secara bersamaan untuk mengekstrak fitur visual dan temporal dari gambar wajah (Qi et al., 2023).

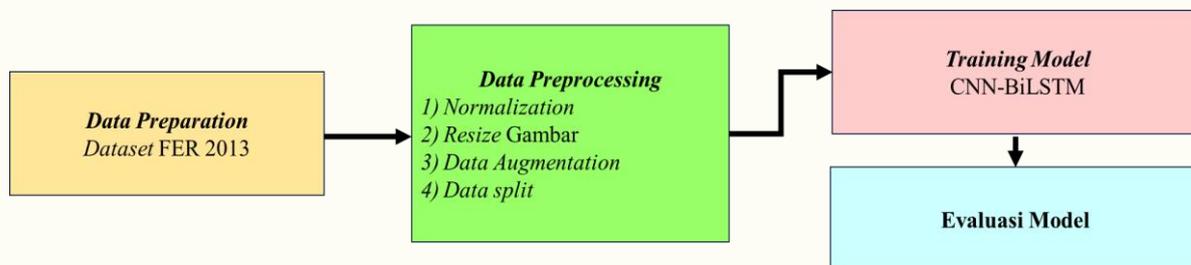
Berbagai algoritma pada *machine learning* dan *deep learning* telah diimplementasikan pada sistem *facial expression recognition* (FER). Kondisi pencahayaan, posisi wajah, ekspresi wajah yang kompleks, kualitas data pelatihan dapat mempengaruhi akurasi deteksi emosi. Untuk mengatasi kekurangan-kekurangan tersebut maka perlu dikembangkan algoritma deteksi emosi yang mempunyai akurasi yang lebih baik. Pada penelitian ini akan menerapkan model CNN-BiLSTM dengan *data augmentation* untuk mendeteksi emosi seperti marah, bahagia, takut, sedih, terkejut, jijik, dan netral. CNN memiliki kemampuan untuk tahan terhadap gangguan (Cahya, Hardi, Riana, & Hadiyanti, 2021), seperti kondisi pencahayaan yang buruk dan pose wajah yang tidak ideal. Hal ini karena CNN dapat mengekstrak fitur visual yang penting untuk deteksi emosi dari gambar wajah, terlepas dari kondisi pencahayaan dan pose wajah. BiLSTM memiliki kemampuan untuk menangani data yang kompleks (Algarni, Saeed, Al-Hadhrani, Ghabban, & Al-Sarem, 2022). Hal ini karena BiLSTM dapat mengolah informasi temporal dari gambar wajah, yang dapat membantu untuk membedakan antara ekspresi wajah yang kompleks.

Penelitian pengenalan emosi berbasis wajah menghadapi tantangan lain seperti ketidakseimbangan data dalam *dataset*, variasi kondisi pencahayaan, posisi wajah, dan kompleksitas ekspresi. *Data augmentation* telah terbukti menjadi solusi efektif untuk meningkatkan variasi data dan mencegah *overfitting* (Shorten & Khoshgoftaar, 2019). Teknik augmentasi seperti rotasi, *flipping*, dan pergeseran gambar dapat membantu model belajar dari pola yang lebih beragam (Paraskevopoulou et al., 2022). Namun, meskipun kombinasi CNN dan BiLSTM telah digunakan dalam penelitian sebelumnya, tantangan seperti *overfitting*, ketidakseimbangan data, serta stabilitas performa model dalam berbagai kondisi masih menjadi masalah. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model CNN-BiLSTM yang dilengkapi dengan *data augmentation* guna meningkatkan akurasi dan kemampuan generalisasi dalam mendeteksi emosi berbasis ekspresi wajah.

Penelitian ini diharapkan dapat mengatasi keterbatasan metode sebelumnya dengan memanfaatkan keunggulan CNN dalam mengekstraksi fitur visual serta BiLSTM dalam menangkap informasi temporal. Dengan demikian, model yang diusulkan diharapkan lebih unggul dalam menghadapi tantangan seperti variasi pencahayaan, posisi wajah yang beragam, dan ekspresi wajah yang kompleks.

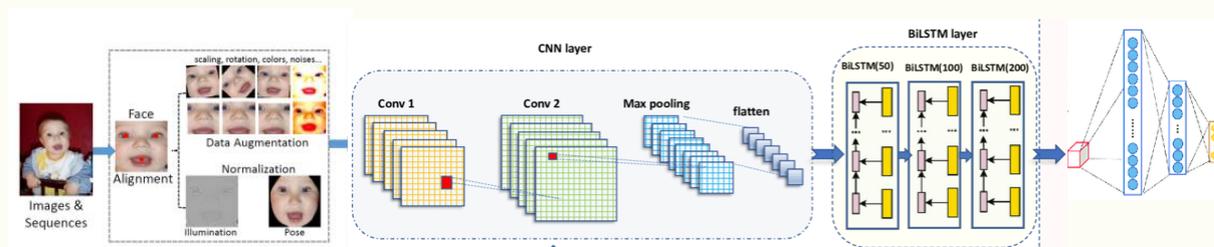
## METODE

Penelitian ini terdiri dari empat tahapan utama yaitu penyiapan *dataset*, *data preprocessing* menggunakan *data augmentation*, *training model* menggunakan CNN-BiLSTM, dan evaluasi model seperti terlihat pada gambar 1. Dataset yang digunakan adalah FER 2013. *Data preprocessing* yang dilakukan pada FER 2013 meliputi beberapa langkah yaitu ukuran gambar wajah dikonversi menjadi 48x48 piksel.



Gambar 1. Tahapan Penelitian

*Data augmentation* menggunakan library Keras ImageDataGenerator. Gambar 2 menunjukkan arsitektur CNN-BiLSTM dan *data augmentation*.



Gambar 2. Sistem Pengenalan Wajah CNN-BiLSTM dengan *Data Augmentation*

Penelitian ini dilakukan dalam beberapa tahapan, yaitu:

### 1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah dataset FER 2013 yang merupakan *dataset* publik. Dataset ini terdiri dari 35.887 gambar wajah berukuran 48x48 piksel dalam skala abu-abu, yang terbagi menjadi tujuh kelas emosi yaitu bahagia, marah, takut, sedih, terkejut, jijik, dan netral. *Dataset* ini dipilih karena merupakan standar dalam penelitian pengenalan emosi berbasis ekspresi wajah.

### 2. *Preprocessing* Data

*Preprocessing* bertujuan untuk meningkatkan kualitas data sebelum digunakan dalam pelatihan model. Langkah-langkah *preprocessing* meliputi:

- Normalisasi, yaitu semua piksel gambar dinormalisasi ke rentang [0, 1] untuk memastikan bahwa data memiliki skala yang seragam. Normalisasi dilakukan untuk mempercepat konvergensi model selama pelatihan.
- Resize gambar dilakukan agar semua gambar berukuran 48x48 piksel. Hal ini dilakukan agar sesuai dengan input model CNN dan CNN-BiLSTM.
- Data augmentation* dilakukan dengan cara rotasi, *flipping* horizontal, dan pergeseran gambar. Hal ini dilakukan untuk meningkatkan variasi data dan mencegah *overfitting*.
- Data split* dengan cara *dataset* dibagi menjadi tiga yaitu *training* sebesar 80% yang digunakan untuk melatih model. *Validation* sebesar 10% digunakan untuk mengevaluasi performa model selama pelatihan. Sedangkan *testing* sebesar 10% yang digunakan untuk menguji performa akhir model.

### 3. Pengembangan dan Pelatihan Model

Penelitian ini mengembangkan dua arsitektur model yaitu CNN dan CNN-BiLSTM.

#### a) Arsitektur CNN

CNN digunakan untuk mengekstraksi fitur visual dari gambar wajah. Model terdiri dari beberapa lapisan *convolutional* dan *pooling*, diakhiri dengan lapisan *fully connected* dan fungsi aktivasi *softmax*.

*Hyperparameter* yang digunakan:

- i. *Batch size*: 32, dipilih untuk menjaga keseimbangan antara konsumsi memori dan kecepatan pelatihan.
- ii. *Epochs*: 100, dipilih berdasarkan hasil eksperimen awal yang menunjukkan bahwa model cenderung mencapai konvergensi pada jumlah tersebut.
- iii. *Learning rate*: 0.001, dipilih untuk menghindari fluktuasi yang signifikan selama proses pelatihan.

#### b) Arsitektur CNN-BiLSTM

Model ini menggabungkan CNN untuk ekstraksi fitur visual dengan BiLSTM untuk menangkap pola temporal dari data yang diekstraksi. BiLSTM memiliki keunggulan dalam mempelajari hubungan temporal dari perubahan ekspresi wajah.

*Hyperparameter* yang digunakan:

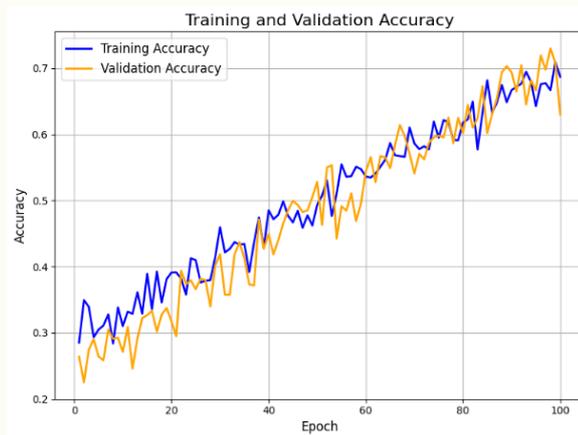
- i. *Batch size*: 32, untuk menjaga konsistensi dengan model CNN dan memastikan stabilitas proses pelatihan.
- ii. *Epochs*: 100, untuk memastikan kedua model diuji dengan kondisi pelatihan yang sama.
- iii. *Learning rate*: 0.001, untuk memastikan model mampu belajar pola tanpa mengalami overshooting.

### 4. Evaluasi Model

Evaluasi model dilakukan beberapa metrik seperti menggunakan akurasi, *loss function* dan *confusion matrix*. Akurasi bertujuan untuk mengukur persentase prediksi yang benar. *Loss function* digunakan untuk mengukur *error* yang terjadi selama pelatihan dan pengujian. *Confusion Matrix* berfungsi untuk mengevaluasi performa model pada masing-masing kelas emosi. Model CNN dan CNN-BiLSTM kemudian dibandingkan berdasarkan akurasi, *loss function*, dan hasil dari *confusion matrix* untuk mengetahui keunggulan dan kekurangan masing-masing arsitektur.

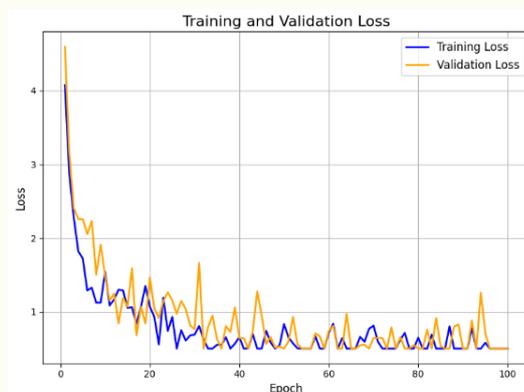
## HASIL DAN PEMBAHASAN

Penelitian ini berfokus pada implementasi arsitektur CNN-BiLSTM untuk deteksi emosi berdasarkan gambar wajah. Pendekatan ini mengkaji kemampuan model dalam mengenali tujuh kategori emosi melalui analisis ekspresi wajah. Penelitian ini membandingkan kinerja model CNN dan CNN-BiLSTM dengan penerapan *data augmentation*, menggunakan berbagai metrik evaluasi, seperti akurasi, *loss function* dan *confusion matrix*, untuk menilai efektivitas dan generalisasi dari kedua model tersebut. Gambar 3, 4 dan 5 merupakan hasil dari evaluasi model CNN dengan menerapkan *data augmentation*.



Gambar 3. Akurasi CNN Menggunakan *Data Augmentation*

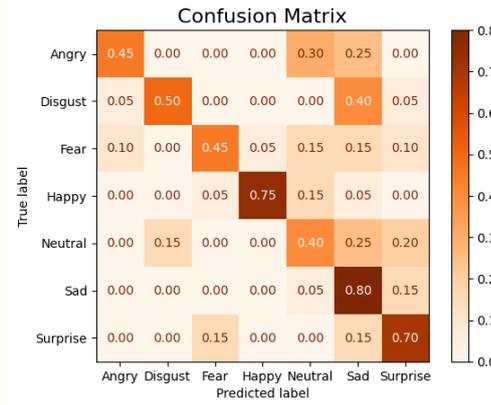
Gambar 3 menunjukkan grafik akurasi *training* dan akurasi validasi model selama proses pelatihan, dengan sumbu X mewakili jumlah *epoch* (iterasi pelatihan) sebanyak 100 dan sumbu Y mewakili nilai akurasi. *Training accuracy* berupa garis biru menggambarkan akurasi model pada *data training* selama proses pelatihan. Dari grafik terlihat bahwa akurasi *training* meningkat seiring bertambahnya jumlah *epoch*, yang berarti model semakin baik mempelajari pola dari *data training*. *Validation accuracy* berupa garis oranye menggambarkan akurasi model pada data validasi, yang digunakan untuk mengevaluasi kemampuan generalisasi model. Akurasi validasi juga meningkat selama *epoch* awal, tetapi ada fluktuasi yang lebih tinggi dibandingkan akurasi *training*, karena data validasi tidak dilatih oleh model. Kedua garis menunjukkan peningkatan yang stabil, dengan perbedaan kecil antara akurasi *training* dan validasi pada akhir proses. Pada *epoch* terakhir, akurasi validasi mendekati akurasi *training* sebesar 70%, yang merupakan indikasi kinerja model yang cukup baik.



Gambar 4. *Loss Function* CNN Menggunakan *Data Augmentation*

Gambar 4 menunjukkan grafik *loss function* selama proses pelatihan model, baik pada data *training* maupun validasi. Sumbu X menunjukkan jumlah iterasi pelatihan (*epoch*), yaitu seberapa sering model mempelajari seluruh dataset. Sumbu Y menunjukkan nilai fungsi *loss* (kerugian), yaitu ukuran seberapa jauh prediksi model dari label yang benar. Semakin kecil nilai *loss*, semakin baik performa model. Garis Biru menggambarkan *loss* model pada data *training*. *Loss* ini menunjukkan seberapa baik model mempelajari pola dari data *training*. Garis Oranye menunjukkan *loss* model pada data validasi. *Loss* ini digunakan untuk mengevaluasi kemampuan generalisasi model pada data yang tidak dilatih.

Pada *epoch* awal, baik *training loss* maupun *validation loss* turun dengan cepat dari nilai awal yang tinggi yaitu 4 menjadi mendekati 1. Setelah sekitar *epoch* 20, *training loss* dan *validation loss* stabil pada nilai yang relatif rendah yaitu 0.5 hingga 1. Pada sebagian besar *epoch*, *training loss* sedikit lebih rendah daripada *validation loss*. *Validation loss* memiliki fluktuasi lebih besar dibandingkan *training loss*. Penurunan yang signifikan di awal dan stabilitas di akhir menunjukkan bahwa model berhasil mempelajari pola dari *data training*.

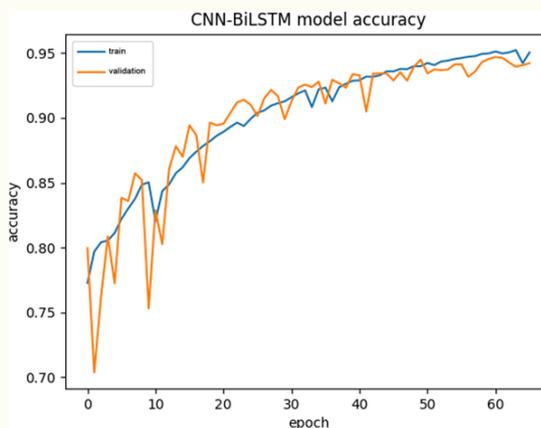


Gambar 5. Confusion Matrix CNN Menggunakan Data Augmentation

Gambar 5 adalah *confusion matrix* yang merupakan representasi visual dari kinerja model klasifikasi dalam memprediksi data. Setiap sel dalam matriks menunjukkan proporsi atau jumlah prediksi model untuk masing-masing kelas dibandingkan dengan label asli. Sumbu X merupakan kelas yang diprediksi oleh model yaitu *Angry*, *Disgust*, *Fear*, *Happy*, *Neutral*, *Sad*, dan *Surprise*. Sumbu Y merupakan kelas yang sebenarnya dari data. Nilai pada diagonal utama dari kiri atas ke kanan bawah menunjukkan jumlah prediksi yang benar untuk setiap kelas. Misalnya untuk kelas *Angry*, model memprediksi dengan benar 45% dari semua data kelas ini untuk kelas *Happy*, model memprediksi dengan benar 75% dan untuk kelas *Sad*, model memiliki akurasi tertinggi sebesar 80% untuk prediksi benar.

Nilai di luar diagonal utama menunjukkan prediksi yang salah, yaitu kelas yang diprediksi tidak sesuai dengan label aslinya, yaitu untuk data sebenarnya kelas *Neutral*, 15% diprediksi sebagai *Disgust*, 40% benar (*Neutral*), 25% sebagai *Sad*, dan 20% sebagai *Surprise*, data kelas *Fear* ada 10% yang salah diprediksi sebagai *Angry*. Dari matriks terlihat bahwa model cukup baik untuk kelas seperti *Happy* (75%) dan *Sad* (80%) namun kinerja kurang optimal untuk kelas seperti *Angry* (45%) dan *Fear* (45%), karena banyak prediksi yang salah ke kelas lain seperti *Neutral* atau *Sad*. Model ini memiliki kekeliruan yang signifikan dalam membedakan emosi yang serupa misalnya, *Fear* sering diprediksi sebagai *Neutral*.

Tahap selanjutnya adalah mengetahui performa dari CNN-BiLSTM dengan menggunakan *data augmentation*. Gambar 9, 10 dan 11 menunjukkan performa model CNN-BiLSTM dengan menggunakan *data augmentation*.

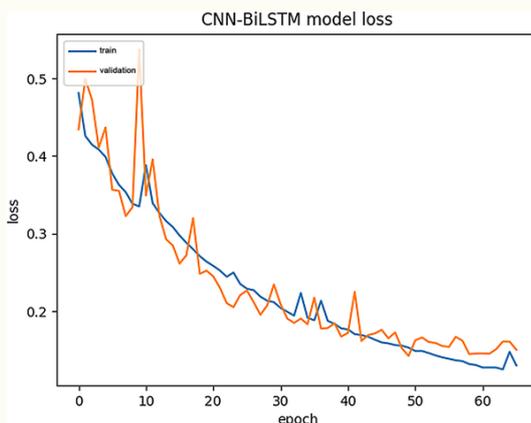


Gambar 6. Akurasi CNN-BiLSTM Menggunakan Data Augmentation

Gambar 6 menunjukkan grafik akurasi model CNN-BiLSTM menggunakan *data augmentation* untuk *data training* dan validasi selama proses pelatihan. Sumbu X menunjukkan jumlah iterasi pelatihan model (*epoch*), yaitu seberapa banyak model telah belajar dari dataset. Sumbu Y menunjukkan akurasi model, yaitu seberapa baik model memprediksi data dengan benar. Garis Biru (*train*) menunjukkan akurasi model pada *data training*. Garis ini menunjukkan bagaimana model belajar dan

menyesuaikan dengan data selama pelatihan. Secara garis besar akurasi pada *data training* meningkat secara stabil dan mendekati 0.95 pada akhir *epoch*. Garis Oranye (*validation*) menunjukkan akurasi model pada data validasi, yang digunakan untuk mengevaluasi kemampuan generalisasi model.

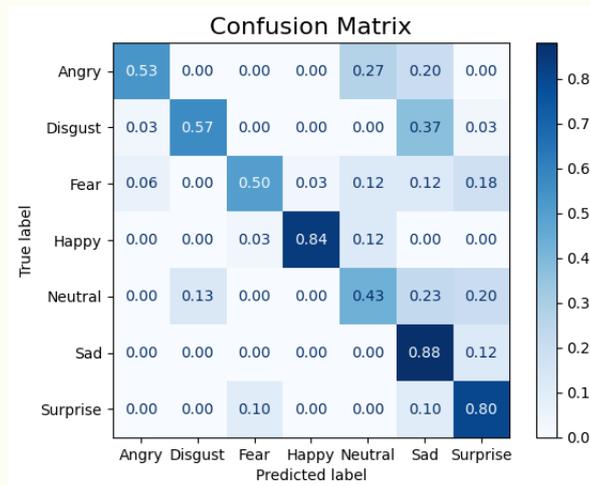
Dari grafik terlihat bahwa akurasi *training* maupun validasi meningkat secara stabil seiring bertambahnya *epoch*, yang menunjukkan model belajar dengan baik. Fluktuasi di awal pelatihan (*epoch* 0-10) menandakan model sedang menyesuaikan diri dengan dataset, namun fluktuasi ini berkurang seiring waktu. Pada *epoch* terakhir, akurasi validasi mendekati akurasi *training*, menunjukkan bahwa model tidak mengalami *overfitting*. Pada akhir *epoch*, baik *training* maupun *validation accuracy* mencapai 95%, yang menunjukkan kinerja model sangat baik. Karena akurasi validasi dan *training* hampir sama, ini menunjukkan model dapat mempelajari pola tanpa *overfitting* atau *underfitting*.



Gambar 7. Loss Function CNN-BiLSTM Menggunakan Data Augmentation

Gambar 7 menunjukkan grafik *loss* model CNN-BiLSTM selama proses pelatihan, baik pada *data training* maupun validasi. Sumbu X menunjukkan jumlah iterasi pelatihan (*epoch*), yaitu berapa kali seluruh *dataset* dilalui oleh model. Sumbu Y menunjukkan nilai fungsi *loss*, yaitu seberapa besar perbedaan antara prediksi model dengan label yang sebenarnya. Semakin kecil nilai *loss*, semakin baik performa model. Garis biru menunjukkan nilai *loss* pada *data training*, yang menunjukkan seberapa baik model mempelajari data selama pelatihan. Garis oranye merupakan nilai *loss* pada data validasi, yang menunjukkan seberapa baik model mampu melakukan generalisasi pada data yang tidak dilatih. *Train loss* dan *validation loss* menunjukkan penurunan secara konsisten sepanjang *epoch*, dari sekitar 0.5 di awal pelatihan menjadi di bawah 0.2 pada akhir pelatihan. Hal ini menunjukkan bahwa model belajar pola dengan baik dari data. *Validation loss* berfluktuasi cukup besar pada *epoch* awal, tetapi stabil setelah *epoch* ke-20. *Train loss* dan *validation loss* saling mengikuti dengan nilai yang hampir sejajar di *epoch* akhir. Hal ini menunjukkan tidak ada *overfitting*, karena model mampu melakukan generalisasi dengan baik pada data validasi. Pada akhir pelatihan, baik *train loss* maupun *validation loss* stabil, yang berarti model telah mencapai konvergensi.

Penurunan *train loss* dan *validation loss* menunjukkan bahwa model CNN-BiLSTM dengan *data augmentation* mampu mempelajari pola dalam *dataset* dengan efektif. Nilai *train loss* dan *validation loss* yang hampir sama menunjukkan bahwa model mampu melakukan generalisasi dengan baik pada data yang tidak dilatih. Fluktuasi awal menunjukkan model sedang beradaptasi dengan *dataset*. Namun, fluktuasi ini berkurang seiring waktu, yang menunjukkan pelatihan yang stabil. *Loss* akhir yang rendah yaitu sebesar 0.2 menunjukkan bahwa prediksi model sudah mendekati label yang sebenarnya.



Gambar 8. Confusion Matrix CNN-BiLSTM Menggunakan Data Augmentation

Gambar 8 merupakan *confusion matrix* yang menunjukkan kinerja model klasifikasi, tetapi dengan hasil yang lebih baik dibandingkan matriks sebelumnya. Model ini menunjukkan peningkatan akurasi pada beberapa kelas dibandingkan matriks sebelumnya. Misalnya akurasi *Angry* meningkat dari 0.45 menjadi 0.53, akurasi *Happy* meningkat dari 0.75 menjadi 0.84, akurasi *Sad* meningkat dari 0.80 menjadi 0.88, dan akurasi *Surprise* meningkat dari 0.70 menjadi 0.80. Sebagian besar nilai pada diagonal utama memiliki intensitas warna lebih gelap, menunjukkan prediksi yang lebih akurat dibandingkan matriks sebelumnya. Nilai pada sel *off-diagonal* (kesalahan prediksi) lebih kecil, menunjukkan model ini lebih baik dalam membedakan kelas-kelas emosi.

### Perbandingan performa CNN dan CNN-BiLSTM yang menerapkan *data augmentation*

#### 1. Akurasi

Pada kedua grafik, akurasi *training* dan validasi meningkat seiring bertambahnya *epoch* yang menunjukkan bahwa model belajar dengan baik. Pada awal pelatihan, kedua model menunjukkan fluktuasi akurasi validasi, tetapi stabil di *epoch* akhir. Tabel 1 merupakan perbandingan akurasi CNN dan CNN-BiLSTM

Tabel 1. Perbandingan Akurasi CNN dan CNN-BiLSTM

Aspek	CNN dengan Data Augmentation	CNN-BiLSTM dengan Data Augmentation
Akurasi Akhir	70% untuk <i>testing</i> dan validasi	95% untuk <i>testing</i> dan validasi
Kinerja Model	kinerjanya lebih rendah	kinerja yang lebih baik
Fluktuasi Validasi	Fluktuasi tinggi hingga <i>epoch akhir</i>	Fluktuasi lebih rendah di <i>epoch</i> akhir yang menunjukkan stabilitas
Keselaran Akurasi	Ada kesenjangan antara akurasi <i>training</i> dan validasi sebesar 5-10%	Akurasi <i>testing</i> dan validasi hampir sama yaitu 95%
Stabilitas Peningkatan	Akurasi validasi tidak selalu meningkat secara konsisten	Akurasi validasi meningkat lebih konsisten
Potensi <i>Overfitting</i>	Ada indikasi potensi <i>overfitting</i> karena kesenjangan akurasi	Tidak ada indikasi <i>overfitting</i> , validasi dan <i>training</i> sangat dekat

CNN-BiLSTM menggunakan *data augmentation* lebih sesuai untuk untuk tugas klasifikasi emosi wajah. Model ini lebih unggul dengan akurasi 95%, sedangkan model tanpa *data augmentation* hanya mencapai 70%. Ini menunjukkan arsitektur model menggunakan *data augmentation* lebih baik dalam memahami data. CNN memiliki fluktuasi validasi yang lebih besar di semua *epoch*, yang bisa disebabkan oleh *dataset* yang lebih kecil. CNN-BiLSTM lebih stabil di akhir *epoch*, menunjukkan pelatihan yang lebih baik. CNN-BiLSTM menunjukkan kemampuan generalisasi yang lebih baik, karena akurasi validasi sangat dekat dengan akurasi training yaitu 95%. CNN memiliki kesenjangan antara *training* dan validasi, yang mengindikasikan model kurang optimal. CNN-BiLSTM lebih baik dalam hal akurasi akhir 95% dan stabilitas akurasi validasi. CNN memiliki kinerja lebih rendah 70% dengan fluktuasi validasi yang lebih besar.

## 2. Loss Function

Kedua grafik *loss* menunjukkan proses pelatihan model, tetapi memiliki perbedaan dalam kinerja, stabilitas, dan kemampuan generalisasi. Pada kedua grafik, *train loss* dan *validation loss* menurun secara konsisten, menunjukkan bahwa kedua model belajar pola data dengan baik. *Validation loss* pada kedua grafik memiliki fluktuasi yang lebih besar dibandingkan *train loss*, terutama di awal pelatihan. Kedua grafik menunjukkan stabilitas *loss* pada akhir pelatihan, yang berarti model telah mencapai konvergensi. Tabel 2 menunjukkan perbandingan *loss function* CNN dan CNN-BiLSTM.

Tabel 2. Perbandingan *Loss Function* CNN dan CNN-BiLSTM

Aspek	CNN dengan <i>Data Augmentation</i>	CNN-BiLSTM dengan <i>Data Augmentation</i>
Loss Akhir	<i>Train loss</i> dan <i>validation loss</i> mendekati 1.0 di akhir <i>epoch</i>	<i>Train loss</i> dan <i>validation loss</i> mendekati 0.2, jauh lebih rendah.
Fluktuasi <i>Validation Loss</i>	<i>Validation loss</i> menunjukkan fluktuasi signifikan sepanjang pelatihan	<i>Validation loss</i> lebih stabil setelah epoch ke-20
Keselarasan <i>Loss</i>	<i>Validation loss</i> sedikit lebih tinggi daripada <i>train loss</i> , ada indikasi model kurang optimal pada generalisasi	<i>Train loss</i> dan <i>validation loss</i> hampir sama, menunjukkan generalisasi yang sangat baik
Kecepatan Penurunan	<i>Loss</i> menurun lebih lambat dan stabil di angka yang lebih tinggi	<i>Loss</i> menurun lebih cepat dengan nilai akhir yang jauh lebih rendah
Generalisasi	Model memiliki kemampuan generalisasi yang kurang optimal	Model memiliki kemampuan generalisasi yang baik

CNN-BiLSTM memiliki kinerja yang lebih baik, karena *Loss* lebih rendah dibandingkan CNN, dan generalisasi lebih baik. CNN memiliki *loss* yang lebih tinggi, menunjukkan bahwa model masih memiliki ruang untuk perbaikan. *Validation loss* pada CNN-BiLSTM lebih stabil, menunjukkan bahwa model lebih baik dalam menangani variasi data validasi. Fluktuasi *validation loss* pada CNN menunjukkan model kurang stabil. Model CNN-BiLSTM lebih baik dibandingkan model CNN dalam hal *loss* akhir, stabilitas, dan generalisasi.

## 3. Confusion Matrix

Kedua matriks menunjukkan distribusi prediksi model terhadap tujuh kategori emosi yaitu *angry*, *disgust*, *fear*, *happy*, *neutral*, *sad*, dan *surprise*. Sumbu Y mewakili label sebenarnya, sedangkan sumbu X mewakili label prediksi. Kedua matriks menunjukkan bahwa emosi *happy*, *sad*, dan *surprise* memiliki performa yang baik, dengan nilai diagonal yang tinggi, menunjukkan banyak prediksi yang benar. *Confusion Matrix* CNN-BiLSTM menunjukkan kinerja yang lebih baik secara keseluruhan dibandingkan *Confusion Matrix* CNN, terlihat dari peningkatan akurasi pada hampir semua kelas. Kelas *Happy* dan *Sad* menunjukkan peningkatan akurasi yang paling besar, dengan akurasi masing-masing meningkat dari 75% ke 84% dan dari 80% ke 88%.

Grafik *loss function* menunjukkan bahwa model CNN mengalami fluktuasi *validation loss* yang lebih besar dibandingkan CNN-BiLSTM. Hal ini mengindikasikan bahwa CNN lebih rentan terhadap *overfitting*. *Overfitting* pada CNN kemungkinan disebabkan oleh ketidakseimbangan data dalam *dataset* FER2013, di mana kelas emosi tertentu memiliki jumlah gambar yang jauh lebih sedikit dibandingkan kelas lainnya. Teknik data augmentation yang diterapkan pada model CNN-BiLSTM membantu meningkatkan kemampuan generalisasi model dengan memperkenalkan variasi data, sehingga model lebih tahan terhadap *overfitting*.

*Dataset* FER2013 memiliki ketidakseimbangan jumlah gambar antar kelas emosi, yang dapat memengaruhi performa model. Model CNN cenderung bias terhadap kelas emosi dengan jumlah data yang lebih besar, seperti kelas "happy", yang terlihat dari confusion matrix yang menunjukkan bahwa prediksi untuk kelas ini lebih akurat dibandingkan kelas lain. Sementara itu, CNN-BiLSTM yang dilengkapi dengan *data augmentation* mampu meningkatkan performa pada kelas-kelas dengan jumlah data yang lebih sedikit, seperti "disgust" dan "fear". Teknik *data augmentation* seperti rotasi, *flipping*, dan pergeseran gambar meningkatkan keragaman data dan membantu model belajar dari pola yang lebih luas.

CNN-BiLSTM menunjukkan kemampuan generalisasi yang lebih baik dibandingkan CNN, yang ditunjukkan oleh perbedaan yang lebih kecil antara *train loss* dan *validation loss*. Ini menandakan bahwa model CNN-BiLSTM dapat menangkap pola yang lebih representatif dari *data training* dan mampu mempertahankan performanya saat diuji pada data yang belum pernah dilihat sebelumnya. Sebaliknya, perbedaan signifikan antara *train loss* dan *validation loss* pada model CNN menunjukkan bahwa model tersebut cenderung hanya menghafal pola dari data training tanpa mampu menggeneralisasi dengan baik pada data validasi. Berdasarkan hasil analisis, model CNN-BiLSTM lebih unggul dibandingkan CNN dalam hal akurasi, stabilitas *validation loss*, dan kemampuan generalisasi. Hal ini terutama disebabkan oleh kombinasi arsitektur CNN untuk ekstraksi fitur visual dan BiLSTM untuk menangkap pola temporal, serta penerapan *data augmentation* yang membantu mengatasi ketidakseimbangan data dan mengurangi risiko *overfitting*.

## KESIMPULAN

Berdasarkan hasil pengujian dan analisis, dapat disimpulkan bahwa model CNN-BiLSTM memiliki performa yang lebih baik dibandingkan dengan model CNN dalam mendeteksi emosi pada wajah. Hal ini dibuktikan dari akurasi validasi model CNN-BiLSTM yang lebih tinggi, yaitu sebesar 95%, dibandingkan dengan model CNN yang hanya mencapai akurasi sebesar 70%. Selain itu, *validation loss* pada CNN-BiLSTM lebih stabil, menunjukkan kemampuan generalisasi yang lebih baik, sedangkan model CNN mengalami fluktuasi *validation loss* yang lebih besar, yang mengindikasikan potensi *overfitting*. Berdasarkan hasil analisis *confusion matrix*, model CNN-BiLSTM juga menghasilkan klasifikasi yang lebih akurat untuk setiap kategori emosi, terutama pada kelas-kelas dengan jumlah data yang lebih sedikit. Kemungkinan perbaikan penelitian selanjutnya yaitu penggunaan teknik *balancing* data untuk mengatasi ketidakseimbangan *dataset* dengan teknik *oversampling* atau *undersampling*, atau dengan menggunakan metode seperti *Synthetic Minority Over-sampling Technique* (SMOTE) untuk meningkatkan performa model pada kelas dengan data yang lebih sedikit. Selain itu dapat dilakukan optimasi *hyperparameter* menggunakan algoritma seperti *grid search* atau *random search* untuk meningkatkan performa model.

## DAFTAR PUSTAKA

- Akhand, M.A.H., Roy, S., Siddique, N., Kamal, M.A.S., and Shimamura, T. (2021). Facial Emotion Recognition Using Transfer Learning in the Deep CNN. *Electronics* 2021, 10(9), 1036. <https://doi.org/10.3390/electronics10091036>.
- Adiatma, B. C. L., Utami, E., & Hartanto, A. D. (2021). Pengenalan Ekspresi Wajah Menggunakan Deep Convolutional Neural Network. *Explore*, 11(2), 75. <https://doi.org/10.35200/explore.v11i2.478>

- Algarni, M., Saeed, F., Al-Hadhrami, T., Ghabban, F., & Al-Sarem, M. (2022). Deep Learning-Based Approach for Emotion Recognition Using Electroencephalography (EEG) Signals Using Bi-Directional Long Short-Term Memory (Bi-LSTM). *Sensors*, 22(8). <https://doi.org/10.3390/s22082976>
- Almeida, J., Vilaça, L., Teixeira, I. N., & Viana, P. (2021). applied sciences Emotion Identification in Movies through Facial Expression Recognition. *Appl. Sci*, 11(15), 6827. <https://doi.org/10.3390/app11156827>
- Avtar, R., & Dhingra, S. (2023). Measurement : Sensors Empirical analysis of multiple modalities for emotion recognition using convolutional neural network. *Measurement: Sensors*, 26(January), 100716. <https://doi.org/10.1016/j.measen.2023.100716>
- Białek, C., Mاتیolański, A., and Grega, M. (2023). An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks, *Electronics* 2023, 12(12), 1–21. <https://doi.org/10.3390/electronics12122707>.
- Budiman, A., Fabian, Yupitera, R. A., Achmad, S., & Kurniawan, A. (2022). Student attendance with face recognition (LBPH or CNN): Systematic literature review. *Procedia Computer Science*, 216, 31–38. <https://doi.org/10.1016/j.procs.2022.12.108>
- Bursic, S., Boccignone, G., Ferrara, A., & Amelio, A. D. (2020). Improving the Accuracy of Automatic Facial Expression Recognition in Speaking Subjects with Deep Learning, *applied sciences* 1–15. <https://doi.org/10.3390/app10114002>
- Cahya, F. N., Hardi, N., Riana, D., & Hadiyanti, S. (2021). Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network (CNN). *Sistemasi*, 10(3), 618. <https://doi.org/10.32520/stmsi.v10i3.1248>
- Canedo, D., & Neves, A. J. R. (2019). Facial expression recognition using computer vision: A systematic review. *Applied Sciences (Switzerland)*, 9(21), 1–31. <https://doi.org/10.3390/app9214678>
- Gautam, C., & Seeja, K. R. (2023). Emotion recognition using Handcrafted features and CNN Conference on Chahak using Facial emotion recognition Handcrafted features and CNN. *Procedia Computer Science*, 218(2022), 1295–1303. <https://doi.org/10.1016/j.procs.2023.01.108>
- Krishnaveni, K. (2023). A novel framework using binary attention mechanism based deep convolution neural network for face emotion recognition. *Measurement: Sensors*, 30, 100881. <https://doi.org/10.1016/j.measen.2023.100881>
- Meena Jindal, & Khushwant Kaur. (2024). Enhancing Emotion Recognition through Multimodal Systems and Advanced Deep Learning Techniques. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 10(3), 656–661. <https://doi.org/10.32628/cseit24103216>
- Mellouk, W., & Handouzi, W. (2020). Facial emotion recognition using deep learning : review and insights Facial emotion recognition using deep learning : review and insights the Future. *Procedia Computer Science*, 175, 689–694. <https://doi.org/10.1016/j.procs.2020.07.101>
- Meyer-Lindenberg, H., Moessnang, C., Oakley, B., Ahmad, J., Mason, L., Jones, E. J. H., ... Loth, E. (2022). Facial expression recognition is linked to clinical and neurofunctional differences in autism. *Molecular Autism*, 13(1), 1–18. <https://doi.org/10.1186/s13229-022-00520-7>
- Ming, Y., Qian, H., & Guangyuan, L. (2022). CNN-LSTM Facial Expression Recognition Method Fused with Two-Layer Attention Mechanism,. *Computational Intelligence and Neuroscience* Volume 2022, 1-9. <https://doi.org/10.1155/2022/7450637>
- Paraskevopoulou, G., Spyrou, E., & Perantonis, S. (2022). A Data Augmentation Approach for Improving the Performance of Speech Emotion Recognition, *Sigmap*, 61–69.

<https://doi.org/10.5220/0011148000003289>

- Qi, X., Wu, C., Shi, Y., Qi, H., Duan, K., & Wang, X. (2023). A Convolutional Neural Network Face Recognition Method Based on BiLSTM and Attention Mechanism. *Computational Intelligence and Neuroscience*, 2023(1). <https://doi.org/10.1155/2023/2501022>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0197-0>
- Suartika E. P, I Wayan, Wijaya Arya Yudhi, S. R. (2016). Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) Pada Caltech 101. *Jurnal Teknik ITS*, 5(1), 76. Retrieved from <http://repository.its.ac.id/48842/>
- Teja Chavali, S., Tej Kandavalli, C., Sugash, T. M., & Subramani, R. (2022). Smart Facial Emotion Recognition with Gender and Age Factor Estimation. *Procedia Computer Science*, 218(2022), 113–123. <https://doi.org/10.1016/j.procs.2022.12.407>
- Widodo, S., Setiawan, D., Ridwan, T., & Ambari, R. (2022). Perancangan Deteksi Emosi Manusia berdasarkan Ekspresi Wajah Menggunakan Algoritma VGG16. *Syntax: Jurnal Informatika*, 11(01), 01–12. <https://doi.org/10.35706/syji.v11i01.6594>