

Pengenalan Ekspresi Wajah dengan LBP dan Multi-Level CNN

Muhammad Bagas Septyono¹, Fetty Tri Anggraeny², Retno Mumpuni³

^{1,2,3}Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia

¹21081010049@student.upnjatim.ac.id, ²fettyanggraeny.if@upnjatim.ac.id, ³retnomumpuni.if@upnjatim.ac.id

Abstrak

Pengenalan ekspresi wajah menjadi salah satu komponen penting dalam sistem interaksi manusia-komputer, khususnya untuk memahami emosi pengguna. Tantangan utama dalam pengenalan ekspresi wajah meliputi variasi pencahayaan, pose wajah, oklusi, dan kemiripan antar ekspresi. Penelitian ini mengusulkan metode klasifikasi ekspresi wajah berbasis kombinasi antara ekstraksi fitur tekstur dari Local Binary Pattern (LBP) dan representasi spasial dari arsitektur Multi-Level Convolutional Neural Network (MLCNN). Dataset yang digunakan adalah FER2013 yang terdiri dari 35.887 gambar grayscale berukuran 48×48 piksel dan terbagi dalam tujuh kategori ekspresi dasar. Tahapan penelitian mencakup prapemrosesan data (deteksi wajah, augmentasi, dan normalisasi), ekstraksi fitur melalui LBP dan MLCNN, serta penggabungan fitur (feature fusion) sebelum klasifikasi. Tiga jenis koneksi fitur diuji dalam MLCNN, yaitu dari lapisan max pooling, lapisan konvolusi tengah, dan lapisan konvolusi terakhir sebelum pooling. Model dilatih menggunakan optimizer Adam, early stopping, dan penyesuaian learning rate otomatis (ReduceLROnPlateau). Hasil eksperimen menunjukkan bahwa kombinasi LBP dan MLCNN dengan koneksi tipe 3 memberikan hasil terbaik dengan akurasi pengujian sebesar 69,48%. Pendekatan ini menunjukkan peningkatan akurasi dan kestabilan terutama dalam mengenali ekspresi seperti happy dan angry. Penelitian ini menunjukkan bahwa integrasi fitur tekstur dari LBP dan spasial dari berbagai level abstraksi MLCNN dapat meningkatkan performa sistem pengenalan ekspresi wajah dalam kondisi dunia nyata.

Kata kunci: pengenalan ekspresi wajah, LBP, CNN, multi-level CNN, FER2013.

1. Pendahuluan

Ekspresi wajah merupakan bentuk komunikasi non-verbal yang penting karena dapat menggambarkan perasaan atau emosi seseorang. Dalam berbagai bidang seperti interaksi manusia-komputer, sistem autopilot, pelayanan publik, hingga perawatan medis, kemampuan mengenali ekspresi wajah memainkan peran yang semakin besar. Melalui analisis ekspresi wajah, sebuah sistem dapat memahami kondisi fisik dan mental manusia dalam konteks komunikasi (Li et al., 2020). Menurut Ekman, terdapat tujuh ekspresi wajah dasar yang dapat dikenali secara universal, yaitu bahagia, sedih, marah, terkejut, jijik, takut, dan netral. Ekspresi wajah ini tidak hanya menggambarkan emosi yang sedang dirasakan, tetapi juga dapat memberikan gambaran mengenai kondisi psikologis seseorang (Ekman, 1971). Beberapa studi menunjukkan bahwa pengenalan ekspresi wajah mampu meningkatkan kepercayaan dalam interaksi dan mendukung komunikasi verbal, terutama dalam layanan berbasis kecerdasan buatan.

Namun, tantangan utama dalam pengenalan ekspresi wajah terletak pada keragaman kondisi nyata seperti variasi pencahayaan, oklusi oleh masker atau rambut, perbedaan pose kepala, serta latar belakang yang kompleks. Pendekatan tradisional seperti

deteksi titik kunci wajah atau model berbasis geometri sering kali tidak memberikan hasil yang memadai dalam kondisi semacam ini (Nguyen et al., 2022).

Dengan pesatnya perkembangan teknologi kecerdasan buatan, khususnya pembelajaran mesin (machine learning) dan pembelajaran mendalam (deep learning), metode berbasis Convolutional Neural Network (CNN) menjadi andalan dalam tugas pengenalan citra. CNN terbukti mampu mengekstraksi fitur visual yang relevan dan memiliki kinerja tinggi dalam klasifikasi gambar karena meniru cara kerja visual cortex pada manusia (Nugroho et al., 2020).

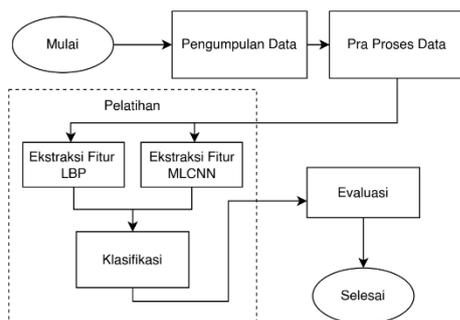
Penelitian sebelumnya oleh Shao dan Qian mengusulkan tiga model CNN, salah satunya adalah dual-branch CNN. Model ini dirancang untuk menggabungkan fitur global dari citra asli dan fitur tekstur lokal dari citra Local Binary Pattern (LBP). Namun menambahkan fitur LBP ke dalam jaringan yang dalam tidak meningkatkan kinerja klasifikasi secara signifikan. Ini menunjukkan bahwa CNN yang dalam dapat mengekstrak informasi yang cukup dari piksel mentah saja dan akurasi yang didapat hanya 54% (Shao & Qian, 2019).

Berangkat dari kekurangan dan potensi yang terlihat pada penelitian sebelumnya, penelitian ini mengusulkan pengembangan arsitektur klasifikasi

ekspresi wajah dengan menggabungkan metode Local Binary Pattern (LBP) dan Multi-Level CNN. Multi-Level CNN diusulkan oleh Zhao dkk. untuk menangkap fitur dari berbagai tingkat abstraksi pada dataset FER2013 (Zhai et al., 2022). Tujuan utama dari penelitian ini adalah untuk mengekstraksi fitur wajah dari berbagai tingkat (low-level, mid-level, high-level) serta fitur tekstur lokal dari LBP, guna memperoleh akurasi klasifikasi yang lebih baik dan stabil terhadap beragam kondisi dunia nyata. Penelitian ini juga membandingkan pengaruh hasil kombinasi LBP dan Multi-Level CNN dibanding CNN biasa. Ruang lingkup penelitian ini difokuskan pada pengenalan ekspresi wajah berbasis gambar statis dari dataset FER2013, yang mencakup tujuh kategori emosi dasar. Evaluasi dilakukan menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Penelitian ini diharapkan memberikan kontribusi nyata dalam pengembangan sistem pengenalan ekspresi wajah yang lebih akurat dan aplikatif, serta menjadi referensi bagi penelitian lanjutan dalam bidang analisis emosi berbasis kecerdasan buatan..

2. Metode

Penelitian ini bertujuan mengembangkan sistem pengenalan ekspresi wajah dengan menggabungkan metode Local Binary Pattern (LBP) dan Multi-Level Convolutional Neural Network (MLCNN). Alur penelitian dilakukan melalui beberapa tahap utama, yaitu pengumpulan data, prapemrosesan, perancangan model, pelatihan, dan evaluasi. Diagram alur proses penelitian secara umum ditunjukkan pada Gambar 1.



Gambar 1 Metode Penelitian

2.1 Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini adalah Facial Expression Recognition 2013 (FER2013), yang terdiri dari 35.887 citra grayscale berukuran 48x48 piksel. Citra grayscale menampilkan gradasi hitam ke putih, di mana tiap piksel ditentukan oleh nilai intensitas antara 0 (hitam) hingga 255 (putih). Nilai ini direpresentasikan dalam 8 bit (1 byte), sehingga citra jenis ini sangat sesuai untuk pengolahan gambar digital (Marpaung et al., 2022)(Kusumanto & Tomponu, 2011). Dataset telah

dipisahkan menjadi tiga bagian, yaitu data latih (28.709), data validasi (3.589), dan data uji (3.589). Terdapat tujuh kelas ekspresi yaitu: marah, jijik, takut, bahagia, netral, sedih, dan terkejut. Gambar 2 merupakan kumpulan gambar dataset



Gambar 2 Kumpulan data FER2013

2.2 Prapemrosesan Data

Langkah pertama dalam prapemrosesan adalah deteksi wajah untuk memastikan validitas data. Penghapusan data manual dilakukan untuk menghilangkan gambar yang tidak mengandung wajah, atau wajah yang tidak terdeteksi dengan baik. Contohnya adalah pada Gambar 3, terdapat kumpulan gambar yang tidak valid dan perlu dihapus karena nantinya akan mengganggu proses pelatihan data.



Gambar 3 Gambar Tidak Valid

Kemudian, semua citra diubah ukurannya menjadi 48x48 piksel agar sesuai dengan input standar CNN. Selanjutnya dilakukan augmentasi data dengan rotasi (hingga ±20°), zoom (hingga 20%), dan horizontal flip untuk memperkaya keragaman data pelatihan serta mengurangi risiko overfitting, sebagaimana juga dilakukan dalam penelitian oleh (Li et al., 2020) dan (Bahmanabady et al., 2023). Setelah augmentasi, data dinormalisasi ke skala [0,1] dengan membagi setiap nilai piksel dengan 255.

2.3 LBP

Local Binary Pattern (LBP) merupakan salah satu metode ekstraksi fitur tekstur yang digunakan dalam penelitian ini untuk memperkuat representasi spasial citra wajah. LBP pertama kali diperkenalkan oleh (Ojala et al., 1996) sebagai metode deskripsi tekstur lokal yang sederhana namun sangat efektif dalam menangkap pola tekstur, khususnya dalam kondisi pencahayaan yang bervariasi. Metode ini terbukti efektif dalam merepresentasikan tekstur, memiliki tingkat kemampuan diskriminasi yang tinggi, serta cukup tangguh terhadap perubahan nilai grayscale yang bersifat seragam atau monoton (Aminarto et al., 2018)

LBP bekerja dengan cara membandingkan nilai intensitas piksel pusat dengan piksel-piksel tetangganya di sekelilingnya, dalam jendela kernel (biasanya 3×3). Setiap piksel tetangga dikodekan sebagai 1 bila intensitasnya lebih besar atau sama dengan piksel pusat, dan 0 jika lebih kecil. Nilai-nilai biner tersebut disusun dalam urutan tertentu, membentuk pola biner 8-bit yang kemudian dikonversi ke bilangan desimal. Secara matematis, nilai LBP dari sebuah piksel pusat (x, y) didefinisikan sebagai berikut:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (1)$$

dengan fungsi ambang $s(x)$ adalah:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

Keterangan:

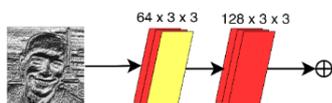
P : jumlah piksel tetangga

R : nilai radius

G_c : nilai piksel x,y

G_p : nilai piksel tetangga

Local Binary Pattern (LBP) tidak hanya digunakan sebagai deskriptor tekstur, tetapi juga dirancang sebagai cabang tersendiri (LBP branch) dalam model klasifikasi ekspresi wajah. Cabang ini bertujuan untuk mengekstraksi fitur tekstur lokal secara paralel dengan cabang utama Multi-Level CNN (MLCNN), kemudian hasil ekstraksi digabungkan dalam proses feature fusion untuk klasifikasi akhir.



Gambar 4 Arsitektur LBP

Setelah transformasi LBP, hasilnya tidak langsung digunakan untuk klasifikasi, melainkan dimasukkan ke dalam arsitektur jaringan saraf konvolusional mini yang terdiri dari tiga blok konvolusi dan pooling. Tujuannya adalah untuk melakukan reduksi dimensi serta menangkap pola-pola tekstur pada skala yang lebih tinggi seperti pada Gambar 4. Proses konvolusi dilakukan dengan memanfaatkan kernel dan stride, di mana konvolusi merupakan operasi yang menggabungkan dua matriks berbeda untuk menghasilkan matriks baru yang nilainya dihitung berdasarkan interaksi antara elemen-elemen dari kedua matriks tersebut (Azmi et al., 2023).

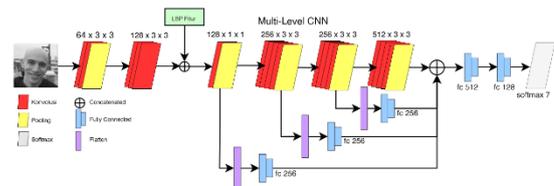
2.4 Multi Level CNN

Selain cabang LBP, arsitektur utama dalam penelitian ini menggunakan pendekatan Multi-Level Convolutional Neural Network (MLCNN) sebagai fondasi utama ekstraksi fitur spasial dan semantik dari citra wajah. Pendekatan ini bertujuan untuk menangkap fitur pada berbagai tingkat kedalaman

(low-level, mid-level, dan high-level) yang sangat penting dalam proses klasifikasi ekspresi wajah.

MLCNN merupakan pengembangan dari Convolutional Neural Network (CNN) yang berfungsi untuk mendeteksi dan mengenali objek dalam gambar. Secara garis besar, CNN bekerja dengan melakukan proses konvolusi, yaitu menggeser sebuah filter atau kernel berukuran tertentu melintasi gambar. Proses ini menghasilkan informasi representatif baru yang diperoleh dari hasil perkalian antara area gambar dan filter tersebut (Lina, 2019).

Secara umum, CNN standar hanya mengambil fitur dari blok konvolusional terakhir sebelum klasifikasi. Namun dalam penelitian ini, fitur dari beberapa blok konvolusi dikumpulkan secara bersamaan untuk memperoleh representasi yang lebih lengkap. Strategi ini dikenal sebagai multi-level feature extraction, yang terbukti efektif dalam meningkatkan performa klasifikasi pada citra wajah sebagaimana ditunjukkan dalam studi oleh (Nguyen et al., 2022)

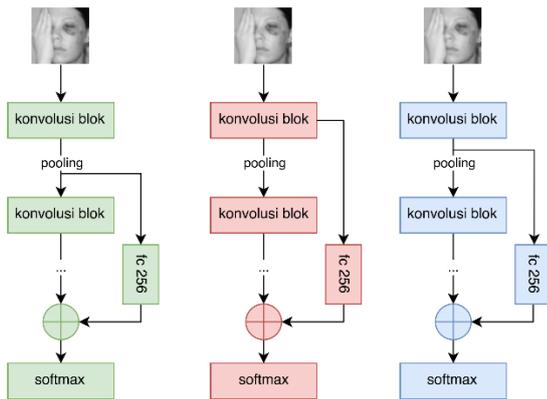


Gambar 5 Arsitektur MultiLevel CNN

Model MLCNN dalam penelitian ini terdiri atas lima blok utama. Setiap blok terdiri dari dua hingga empat lapisan konvolusi dengan kernel 3×3 dan diikuti oleh max pooling 2×2. Jumlah filter meningkat secara progresif dari 64 hingga 512 seperti yang ada pada Gambar 5.

Fitur yang diambil dari blok-blok tersebut kemudian dikombinasikan berdasarkan tiga tipe Koneksi seperti pada Gambar 6. Koneksi Tipe 1 – Mengambil output dari max-pooling setiap blok. Koneksi Tipe 2 – Mengambil fitur dari lapisan konvolusi tengah di setiap blok. Koneksi Tipe 3 – Mengambil output dari lapisan konvolusi terakhir sebelum pooling, untuk mempertahankan informasi spasial penuh sebelum disederhanakan. Pada tahap klasifikasi, lapisan fully connected mengubah fitur hasil ekstraksi menjadi vektor melalui proses flattening. Vektor ini kemudian diproses oleh hidden layer dan akhirnya menghasilkan prediksi pada output layer. Lapisan ini berperan dalam menghubungkan fitur yang telah dipelajari dengan kelas yang sesuai (Intyanto, 2021).

Fitur dari blok 2 hingga blok 5 diekstraksi dan kemudian di-flatten menjadi vektor satu dimensi. Setiap vektor fitur tersebut merepresentasikan tingkat abstraksi yang berbeda. Seluruh vektor dari blok-blok tersebut kemudian digabungkan menggunakan operasi concatenation menjadi satu vektor besar yang mewakili citra input secara menyeluruh. Seperti pada Gambar 6 tipe Koneksi MLCNN terbagi menjadi 3.

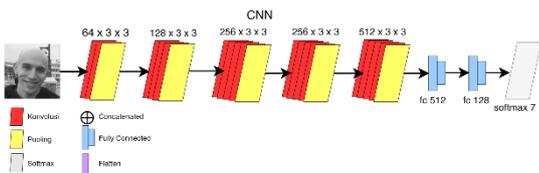


Gambar 6 Tipe Koneksi MLCNN

2.5 CNN

Arsitektur Convolutional Neural Network (CNN) pada penelitian ini digunakan sebagai baseline, tanpa integrasi fitur tambahan seperti LBP maupun penggabungan fitur dari jalur lain. Model CNN murni ini bertujuan untuk mengevaluasi performa dasar dalam tugas klasifikasi ekspresi wajah, sehingga dapat menjadi tolak ukur perbandingan dengan model-model gabungan seperti LBP-MLCNN.

Struktur CNN yang digunakan terdiri dari beberapa blok konvolusi dan pooling yang disusun secara bertingkat. Setiap blok konvolusi terdiri dari beberapa lapisan Conv2D yang diikuti oleh fungsi aktivasi ReLU dan Batch Normalization. Setelah melalui beberapa blok konvolusi dan pooling, output dari CNN dilanjutkan ke lapisan fully connected (FC) untuk klasifikasi akhir.



Gambar 7 Arsitektur CNN

Arsitektur CNN ini adalah arsitektur dasar dari LBP dan MLCNN dimana tidak ada penambahan fitur LBP dan juga tidak mengambil fitur dari setiap tingkat level. Berdasarkan Gambar 7, arsitektur dimulai dari citra grayscale ukuran 48x48 piksel, kemudian melalui lima blok konvolusi yang masing-masing terdiri dari beberapa filter konvolusi dan diikuti oleh operasi max pooling. Setelah blok konvolusi terakhir, fitur dilewatkan ke dua lapisan fully connected, dan akhirnya menuju lapisan output untuk menghasilkan prediksi kelas ekspresi wajah. Model CNN ini tidak hanya berperan sebagai baseline eksperimen, tetapi juga berfungsi untuk menunjukkan kontribusi langsung dari fitur tekstur (LBP) dan multi-level CNN.

2.6 Pelatihan

Proses pelatihan model dilakukan dengan menggunakan optimasi Adam dan fungsi loss categorical cross-entropy, yang umum digunakan pada masalah klasifikasi multikelas. Selama proses pelatihan, dilakukan pengamatan terhadap metrik akurasi validasi untuk menghindari overfitting. Konfigurasi pelatihan yang digunakan adalah Epoch 75, Batch size 32, Optimizer Adam, dan Learning rate 1e-4 hingga 1e-6.

2.7 Evaluasi

Nilai prediksi model dipetakan menjadi 4 bagian, yakni True Positive (TP) terjadi saat model benar memprediksi kelas positif. False Positive (FP) adalah ketika model salah mengangap data negatif sebagai positif. True Negative (TN) menunjukkan prediksi benar untuk kelas negatif, sedangkan False Negative (FN) adalah kesalahan saat data positif diprediksi sebagai negatif (Sitompul et al., 2022).

Untuk mengevaluasi performa model secara menyeluruh, digunakan beberapa metrik evaluasi, yaitu

1. Akurasi

Mengukur proporsi prediksi yang benar dari keseluruhan data. Dihitung dengan rumus:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2. Presisi

Mengukur ketepatan prediksi positif yang dilakukan model. Rumus:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

3. Recall

Mengukur seberapa banyak data positif yang berhasil dikenali oleh model. Rumus:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

4. F1-Score

Rata-rata harmonik antara precision dan recall, berguna saat data tidak seimbang. Rumus:

$$F1 - Score = \frac{2(recall*precision)}{(recall+precision)} \quad (6)$$

Confusion Matrix juga digunakan untuk melihat distribusi prediksi model pada masing-masing kelas ekspresi. Matriks ini menunjukkan jumlah true positive, false positive, true negative, dan false negative untuk masing-masing kelas, sehingga membantu mengidentifikasi kelas mana yang paling sulit dikenali.

3. Hasil dan Pembahasan

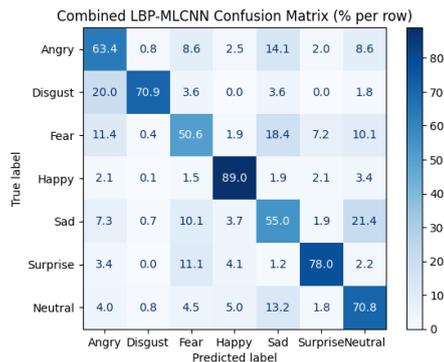
3.1 Hasil

Penelitian ini telah berhasil mengimplementasikan model klasifikasi ekspresi wajah berbasis kombinasi antara Local Binary Pattern

(LBP) dan Multi-Level CNN (MLCNN). Evaluasi dilakukan menggunakan dataset FER2013 yang terdiri dari tujuh kelas ekspresi wajah. Model diuji melalui delapan skenario pengujian dengan kombinasi berbagai tipe koneksi dan arsitektur.

1. LBP + MLCNN Tipe Koneksi 1

Pada koneksi tipe 1 Gambar 6(a), fitur dari cabang MLCNN diambil setelah proses max pooling, kemudian digabungkan dengan fitur dari cabang LBP melalui proses concatenation, dan diteruskan ke fully connected layer berukuran 256 neuron.



Gambar 8 Confusion Matrix Tipe Koneksi 1

```

==== Combined LBP-MLCNN Classification Report (Per Class) ====
precision recall f1-score support
Angry 0.64 0.63 0.64 489
Disgust 0.71 0.71 0.71 55
Fear 0.58 0.51 0.54 526
Happy 0.89 0.89 0.89 878
Sad 0.55 0.55 0.55 593
Surprise 0.79 0.78 0.78 414
Neutral 0.63 0.71 0.67 623

accuracy 0.69 3578
macro avg 0.68 0.68 0.69 3578
weighted avg 0.69 0.69 0.69 3578

==== Combined LBP-MLCNN Overall Performance ====
Accuracy : 0.6948
Precision : 0.6841
Recall : 0.6823
F1-Score : 0.6825
    
```

Gambar 9 Evaluasi Tipe Koneksi 1

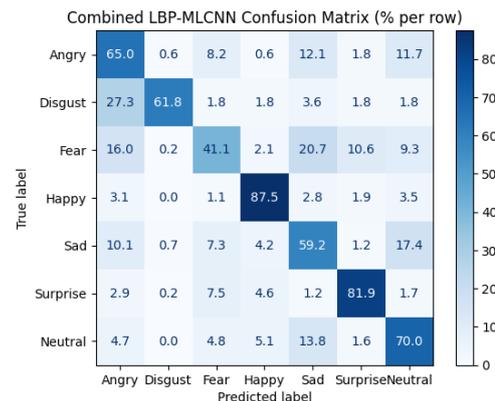
Pada koneksi tipe 1, fitur dari jalur LBP dan MLCNN digabungkan menggunakan metode concatenate. Jalur MLCNN mengambil fitur setelah pooling dan mengalirkannya ke lapisan fully connected 256 unit. Model dilatih selama 75 epoch dan mencapai akurasi pelatihan 80,33% serta akurasi validasi 68,40%. Nilai loss masing-masing sebesar 0,6279 (train) dan 1,0406 (val), tanpa indikasi overfitting signifikan.

Hasil confusion matrix pada Gambar 8 menunjukkan model sangat baik dalam mengenali ekspresi happy dengan precision dan recall sebesar 0,89. Kelas disgust dan surprise juga cukup baik (f1-score 0,71 dan 0,78), namun masih terdapat kesalahan pada kelas fear dan sad. Gambar 9 secara keseluruhan, model mencatatkan akurasi 69,48%, precision 68,41%, recall 68,23%, dan f1-score 68,25%. Gabungan fitur tekstur dan spasial melalui koneksi tipe 1 terbukti meningkatkan performa dibanding CNN standar, serta menawarkan solusi yang stabil untuk klasifikasi ekspresi wajah.

Secara keseluruhan, koneksi tipe 1 menunjukkan performa yang cukup baik, dengan akurasi mendekati performa validasi dan hasil klasifikasi stabil pada kelas dengan ciri spasial dan tekstur yang kuat.

2. LBP + MLCNN Tipe Koneksi 2

Pada konfigurasi koneksi tipe 2, fitur dari jalur LBP digabungkan dengan fitur hasil konvolusi antar-blok dari MLCNN. Tidak seperti koneksi tipe 1 yang menggunakan fitur pasca-pooling, tipe 2 memanfaatkan fitur di tengah proses konvolusi untuk menangkap representasi spasial yang lebih beragam.



Gambar 10 Confusion Matrix Tipe Koneksi 2

```

==== Combined LBP-MLCNN Classification Report (Per Class) ====
precision recall f1-score support
Angry 0.58 0.65 0.62 489
Disgust 0.79 0.62 0.69 55
Fear 0.58 0.41 0.48 526
Happy 0.89 0.87 0.88 878
Sad 0.55 0.59 0.57 593
Surprise 0.77 0.82 0.79 414
Neutral 0.64 0.70 0.67 623

accuracy 0.69 3578
macro avg 0.69 0.67 0.69 3578
weighted avg 0.69 0.69 0.69 3578

==== Combined LBP-MLCNN Overall Performance ====
Accuracy : 0.6881
Precision : 0.6873
Recall : 0.6663
F1-Score : 0.6725
    
```

Gambar 11 Evaluasi Tipe Koneksi 2

Model dilatih selama 75 epoch dengan learning rate awal 1e-4, yang otomatis menurun menjadi 1e-5 karena tidak ada peningkatan pada akurasi validasi. Early stopping terjadi pada epoch ke-70.

Evaluasi terhadap data uji menunjukkan akurasi total sebesar 68,81%, dengan precision 68,73%, recall 66,63%, dan f1-score sebesar 67,25% Gambar 10. Berdasarkan confusion matrix Gambar 10, performa terbaik terdapat pada kelas Happy dengan f1-score 0,88 dan akurasi klasifikasi 87,5%. Kelas Surprise juga menunjukkan hasil baik (precision 0,77, recall 0,82). Namun, kesalahan klasifikasi masih banyak ditemukan pada kelas Fear (recall hanya 41%), yang sering tertukar dengan Sad dan Neutral. Kelas Angry juga menunjukkan ambiguitas, dengan kesalahan klasifikasi ke Sad sebesar 12,1% dan ke Neutral sebesar 11,7%.

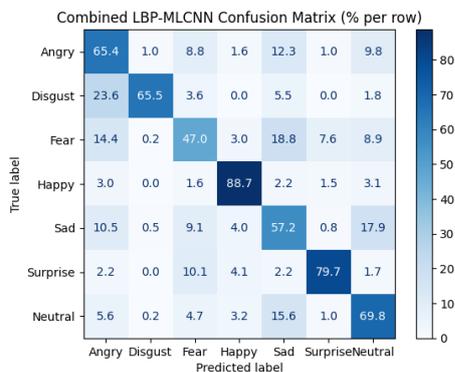
Kelas Disgust yang memiliki jumlah data relatif sedikit tetap berhasil dikenali dengan baik (f1-score 0,69), menandakan bahwa kontribusi fitur tekstur dari

LBP efektif dalam membedakan pola-pola ekspresi minor. Meskipun performa koneksi tipe 2 sedikit lebih rendah dibanding tipe 1, metode ini tetap memberikan hasil yang kompetitif. Kompleksitas tambahan dari fitur antar-blok memberikan variasi, tetapi juga dapat menimbulkan noise yang memengaruhi konsistensi model terutama pada kelas-kelas yang sulit dibedakan.

3. LBP + MLCNN Tipe Koneksi 3

Pada skenario ini digunakan koneksi tipe 3, yaitu penggabungan fitur LBP dengan hasil konvolusi dari tiap blok MLCNN sebelum memasuki proses pooling. Pendekatan ini dirancang untuk menangkap detail spasial awal yang mungkin hilang setelah proses pooling, guna memperkuat representasi fitur.

Model dilatih selama 75 epoch dengan learning rate awal sebesar 1e-4 yang secara otomatis turun menjadi 1e-5. Akurasi validasi tertinggi sebesar 68,34% tercapai pada epoch ke-73 dan bertahan hingga akhir pelatihan.



Gambar 12 Confusion Matrix Tipe Koneksi 3

```

==== Combined LBP-MLCNN Classification Report (Per Class) ====
precision    recall  f1-score   support

 Angry      0.59     0.65     0.62     489
 Disgust    0.78     0.65     0.71     55
 Fear       0.57     0.47     0.52     526
 Happy      0.90     0.89     0.89     878
 Sad        0.54     0.57     0.56     593
 Surprise   0.83     0.80     0.81     414
 Neutral    0.65     0.70     0.67     623

 accuracy   0.69     0.69     0.69     3578
 macro avg  0.70     0.68     0.68     3578
 weighted avg 0.70     0.69     0.69     3578

==== Combined LBP-MLCNN Overall Performance ====
Accuracy : 0.6948
Precision : 0.6951
Recall : 0.6761
F1-Score : 0.6836
    
```

Gambar 13 Evaluasi Tipe Koneksi 3

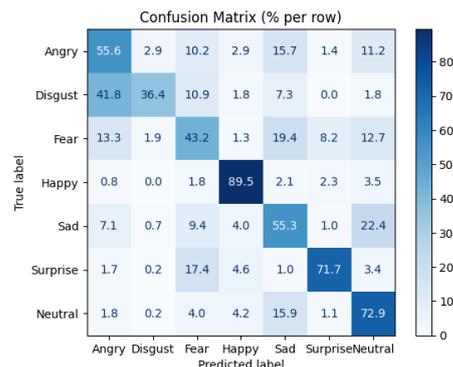
Evaluasi performa model pada data uji Gambar 13 mencatatkan akurasi 69,48%, precision 69,51%, recall 67,61%, dan f1-score 68,36%—menjadi yang tertinggi dibandingkan koneksi tipe 1 dan 2. Confusion matrix Gambar 12 menunjukkan bahwa model paling akurat dalam mengenali ekspresi Happy dengan f1-score sebesar 0,89 dan prediksi benar sebesar 88,7%. Ekspresi Surprise juga dikenali dengan baik (f1-score 0,81), menunjukkan efektivitas model dalam menangkap ciri khas visual di area mata dan mulut.

Kelas Disgust, meskipun merupakan kelas minor, berhasil diklasifikasikan dengan f1-score 0,71, menandakan bahwa fitur tekstur dari LBP membantu mengenali pola ekspresi yang khas. Sementara itu, Fear masih menjadi kelas dengan performa terendah, dengan recall hanya 0,47, karena kemiripan visualnya dengan Sad dan Neutral. Meski demikian, model menunjukkan distribusi performa yang cukup seimbang antar kelas, dibuktikan dengan macro average f1-score sebesar 0,68.

4. CNN

Sebagai baseline, dilakukan pengujian menggunakan arsitektur CNN standar tanpa integrasi fitur tekstur LBP dan fitur Multi-Level CNN. Tujuan dari skenario ini adalah untuk mengevaluasi performa dasar CNN murni dalam mengklasifikasikan ekspresi wajah, sehingga dapat dibandingkan secara objektif dengan model-model gabungan yang lebih kompleks.

Model dilatih selama 75 epoch dengan learning rate awal sebesar 1e-4. Selama pelatihan, akurasi pelatihan mengalami peningkatan bertahap dan mencapai 69,50% pada epoch terakhir. Namun, akurasi validasi tertinggi hanya sebesar 65,94%.



Gambar 14 Confusion Matrix CNN

```

==== Classification Report (Per Class) ====
precision    recall  f1-score   support

 Angry      0.63     0.56     0.59     489
 Disgust    0.40     0.36     0.38     55
 Fear       0.50     0.43     0.46     526
 Happy      0.90     0.90     0.90     878
 Sad        0.52     0.55     0.54     593
 Surprise   0.78     0.72     0.75     414
 Neutral    0.60     0.73     0.66     623

 accuracy   0.62     0.61     0.61     3578
 macro avg  0.67     0.67     0.66     3578

==== Overall Performance ====
Accuracy : 0.6663
Precision : 0.6186
Recall : 0.6066
F1-Score : 0.6106
    
```

Gambar 15 Evaluasi CNN

Evaluasi pada data uji menunjukkan bahwa model mencatatkan akurasi sebesar 66,63%, precision 61,86%, recall 60,66%, dan f1-score 61,06%. Dapat dilihat pada Gambar 15 dan Gambar 14 menunjukkan model CNN sebagai skenario dengan performa terendah dibandingkan dengan model yang menggunakan tambahan fitur LBP dan koneksi multi-level. Meski demikian, model masih mampu mengenali ekspresi Happy dengan sangat baik, terbukti dari f1-score sebesar 0,90 dan tingkat

klasifikasi benar sebesar 89,5%. Ekspresi Surprise juga diklasifikasikan cukup baik dengan f1-score 0,75 dan akurasi prediksi 71,7%.

Sebaliknya, performa pada kelas minor seperti Disgust dan Fear tergolong rendah. Kelas Disgust hanya dikenali dengan benar sebesar 36,4%, dan sering tertukar dengan Angry (41,8%) maupun Fear (10,9%). Sementara itu, kelas Fear hanya dikenali sebesar 43,2%, dan sering salah diklasifikasikan sebagai Sad (19,4%) atau Angry (13,3%). Kesalahan ini mengindikasikan bahwa CNN standar mengalami kesulitan dalam membedakan fitur visual yang mirip antar kelas. Ekspresi Sad dan Neutral menampilkan performa sedang dengan akurasi masing-masing sebesar 55,3% dan 72,9%, meskipun sering tertukar satu sama lain.

Secara keseluruhan, koneksi tipe 3 memberikan hasil terbaik dari semua skenario. Penggabungan fitur sebelum pooling terbukti meningkatkan kualitas representasi spasial awal yang penting untuk klasifikasi ekspresi wajah. Skema ini memberikan kontribusi signifikan terhadap akurasi dan konsistensi model.

3.2 Pembahasan

Penelitian ini mengevaluasi tiga pendekatan berbeda dalam menggabungkan fitur dari Local Binary Pattern (LBP) dan Multi Level Convolutional Neural Network (MLCNN), yang masing-masing dirancang untuk mengeksplorasi titik ekstraksi fitur terbaik dari jalur CNN. Ketiga pendekatan tersebut dibedakan berdasarkan lokasi pengambilan fitur dari MLCNN, yaitu setelah proses pooling (koneksi tipe 1), di tengah blok konvolusi (koneksi tipe 2), dan sebelum pooling (koneksi tipe 3). Evaluasi dilakukan terhadap metrik klasifikasi, yaitu akurasi, precision, recall, dan F1-score. Ringkasan hasil pengujian dapat dilihat pada Tabel 1.

Tabel 1 Hasil Model

Model	Acc	Pre	Recall	FIScore
LBP MLCNN 1	69,48	68,41	68,23	68,25
LBP MLCNN 2	68,81	68,73	66,63	67,25
LBP MLCNN 3	69,48	69,51	67,61	68,36
CNN (Baseline)	66,63	61,86	60,66	61,06

Pada koneksi tipe 1, fitur dari LBP digabungkan dengan fitur MLCNN yang telah melalui proses pooling. Metode ini menunjukkan performa yang stabil dan seimbang, dengan akurasi 69,48% dan f1-score sebesar 68,25%. Model mampu mengenali ekspresi Happy dan Surprise dengan baik, meskipun masih mengalami kesalahan pada kelas-kelas yang memiliki kemiripan visual seperti Fear dan Sad.

Koneksi tipe 2 menggabungkan fitur LBP dengan hasil konvolusi di tengah blok CNN. Meskipun akurasi pelatihan cukup tinggi, performa validasi dan generalisasi menurun dengan f1-score 67,25%. Nilai recall yang lebih rendah pada kelas-kelas tertentu menunjukkan bahwa struktur koneksi ini menghasilkan representasi yang lebih kompleks

namun kurang efisien dalam membedakan ekspresi dengan pola visual serupa.

Sementara itu, koneksi tipe 3 memanfaatkan fitur hasil konvolusi sebelum proses pooling, dengan tujuan menangkap detail spasial awal yang lebih halus. Pendekatan ini memberikan hasil terbaik secara keseluruhan. Model mencatatkan akurasi sebesar 69,48%, precision 69,51%, recall 67,61%, dan f1-score tertinggi sebesar 68,36%. Selain performa yang tinggi pada kelas mayoritas seperti Happy, model juga mampu mengenali kelas minor seperti Disgust dengan baik, yang menunjukkan efektivitas kontribusi fitur tekstur dari LBP terhadap pola ekspresi yang khas namun jarang.

Sebagai pembandingan dasar, model CNN standar tanpa integrasi fitur LBP dan multi-level digunakan sebagai baseline. Model ini menunjukkan performa terendah dibanding ketiga konfigurasi lainnya, dengan akurasi 66,63% dan f1-score sebesar 61,06%. Meskipun model mampu mengenali ekspresi mayoritas seperti Happy dan Surprise dengan cukup baik, performa pada kelas minor seperti Disgust dan Fear cenderung rendah. Hal ini menunjukkan bahwa CNN konvensional masih terbatas dalam mengenali pola tekstur dan detail spasial halus yang penting dalam klasifikasi ekspresi wajah. Oleh karena itu, hasil dari CNN baseline ini menjadi tolok ukur untuk menilai efektivitas penambahan fitur tekstur dan strategi penggabungan fitur pada model-model LBP-MLCNN.

Berdasarkan hasil tersebut, dapat disimpulkan bahwa koneksi tipe 3 memberikan performa paling optimal di antara ketiga konfigurasi. Penggabungan fitur sebelum pooling memungkinkan model untuk menangkap informasi spasial dan tekstur yang lebih relevan sejak tahap awal, sehingga meningkatkan kemampuan diskriminatif terhadap berbagai ekspresi wajah. Sementara itu, koneksi tipe 1 tetap menjadi alternatif yang stabil, sedangkan koneksi tipe 2 memerlukan perhatian lebih terhadap efek kompleksitas representasi terhadap performa klasifikasi.

4. Kesimpulan

Penelitian ini berhasil mengembangkan sistem klasifikasi ekspresi wajah berbasis kombinasi Local Binary Pattern (LBP) dan Multi-Level Convolutional Neural Network (MLCNN), yang dievaluasi menggunakan dataset FER2013. Arsitektur model diuji melalui tiga variasi koneksi untuk menentukan titik ekstraksi fitur terbaik dari cabang MLCNN. Hasil evaluasi menunjukkan bahwa pendekatan dengan koneksi tipe 3 yang menggabungkan fitur LBP dengan hasil konvolusi sebelum pooling memberikan performa terbaik dengan akurasi sebesar 69,48% dan F1-score sebesar 68,36%. Pendekatan ini terbukti mampu mempertahankan informasi spasial yang penting untuk klasifikasi ekspresi wajah, terutama pada kelas-kelas minor seperti Disgust.

Model juga menunjukkan performa tinggi pada kelas mayoritas seperti Happy dan Surprise, serta memberikan peningkatan signifikan dibandingkan CNN konvensional.

Integrasi dua jenis fitur dari tekstur dan spasial memberikan hasil yang lebih stabil, terutama pada kelas ekspresi dengan ciri visual yang kuat seperti happy dan angry. Namun, performa masih terbatas pada kelas dengan jumlah data rendah dan ekspresi yang mirip secara visual, seperti disgust dan fear. Penelitian ini menegaskan bahwa pemilihan strategi koneksi dan titik ekstraksi fitur sangat berpengaruh terhadap efektivitas sistem pengenalan ekspresi wajah.

Dengan demikian, metode yang diusulkan efektif dalam meningkatkan akurasi dan generalisasi sistem klasifikasi ekspresi wajah. Penelitian selanjutnya dapat mengembangkan model ini ke data video atau real-time, serta mempertimbangkan integrasi teknik attention dan balancing data untuk peningkatan performa lebih lanjut.

Daftar Pustaka:

- Amyntarto, N., Sari, Y. A., & Wihandika, R. C. (2018). Pengenalan emosi berdasarkan ekspresi mikro menggunakan metode local binary pattern. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3230–3238. <https://doi.org/10.1109/TAFFC.2019.2946540>
- Azmi, K., Defit, S., & Sumijan, S. (2023). Implementasi Convolutional Neural Network (CNN) Untuk Klasifikasi Batik Tanah Liat Sumatera Barat. *Jurnal Unitek*, 16(1), 28–40. <https://doi.org/10.52072/unitek.v16i1.504>
- Bahmanabady, E., Imani, M., & Ghassemian, H. (2023). Fusion of Multi-Level CNN With LBP Features For Facial Emotion Recognition. *2023 31st International Conference on Electrical Engineering, ICEE 2023*, 463–467. <https://doi.org/10.1109/ICEE59167.2023.10334654>
- Ekman, P. (1971). Consants Across Cultures In the Face and Emotion. *Journal Of Personality and Social Psychology*, 17(2), 124–129. <http://psycnet.apa.org/record/1971-07999-001>
- Intyanto, G. W. (2021). Klasifikasi Citra Bunga dengan Menggunakan Deep Learning: CNN (Convolution Neural Network). *Jurnal Arus Elektro Indonesia*, 7(3), 80. <https://doi.org/10.19184/jaei.v7i3.28141>
- Kusumanto, R., & Tompunu, A. N. (2011). MENDETEKSI OBYEK MENGGUNAKAN PENGOLAHAN WARNA MODEL NORMALISASI. *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2011 (Semantik 2011)*. [https://doi.org/10.1016/S0166-1116\(08\)71924-1](https://doi.org/10.1016/S0166-1116(08)71924-1)
- Li, J., Jin, K., Zhou, D., Kubota, N., & Ju, Z. (2020). Attention mechanism-based CNN for facial expression recognition. *Neurocomputing*, 411, 340–350. <https://doi.org/10.1016/j.neucom.2020.06.014>
- Lina, Q. (2019). *Apa itu Convolutional Neural Network?* Medium. <https://medium.com/@16611110/apa-itu-convolutional-neural-network-836f70b193a4>
- Marpaung, F., Aulia, F., & Nabila, R. C. (2022). *Computer Vision Dan Pengolahan Citra Digital*. www.pustakaaksara.co.id
- Nguyen, H. D., Kim, S. H., Lee, G. S., Yang, H. J., Na, I. S., & Kim, S. H. (2022). Facial Expression Recognition Using a Temporal Ensemble of Multi-Level Convolutional Neural Networks. *IEEE Transactions on Affective Computing*, 13(1), 226–237. <https://doi.org/10.1109/TAFFC.2019.2946540>
- Nugroho, P. A., Fenriana, I., & Arijanto, R. (2020). Implementasi Deep Learning Menggunakan Convolutional Neural Network (CNN) Pada Ekspresi Manusia. *Algor*, 2(1), 12–21.
- Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1), 51–59. [https://doi.org/10.1016/0031-3203\(95\)00067-4](https://doi.org/10.1016/0031-3203(95)00067-4)
- Shao, J., & Qian, Y. (2019). Three convolutional neural network models for facial expression recognition in the wild. *Neurocomputing*, 355, 82–92. <https://doi.org/10.1016/j.neucom.2019.05.005>
- Sitompul, P., Okprana, H., Prasetyo, A., & Artikel, G. (2022). Identifikasi Penyakit Tanaman Padi Melalui Citra Daun Menggunakan DenseNet 201. *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, 1(2), 143–150. <https://doi.org/10.55123/jomlai.v1i2.889>
- Zhai, H., Zhao, J., & Zhang, H. (2022). Double Attention Based Multilevel One-Dimensional Convolution Neural Network for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15(DI), 3771–3787. <https://doi.org/10.1109/JSTARS.2022.3162423>