

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Ekspresi Wajah

Ekspresi wajah merupakan suatu bentuk raut wajah yang dapat digunakan untuk berkomunikasi *nonverbal* yang menggambarkan perasaan yang dapat dirasakan oleh seseorang. Selain itu, ekspresi wajah dapat digunakan untuk mengungkapkan segala macam emosi baik yang negatif maupun emosi positif (Prasetyawan et al., 2020). Ekspresi wajah memiliki tujuh ekspresi dasar diantaranya, yaitu marah, jijik, takut, bahagia, sedih, netral dan terkejut (Zia, et al., 2015)

Seseorang memiliki ekspresi wajah yang dapat dikontrol oleh dirinya sendiri namun, pada umumnya ekspresi wajah timbul secara alami yang disebabkan oleh perasaan atau emosi yang dialami saat itu. Selain itu, seseorang terkadang menginginkan untuk menyembunyikan perasaan atau emosinya akan, tetapi biasanya hal tersebut sulit untuk dilakukan karena pada umumnya wajah mereka biasanya menunjukkan suatu ekspresi wajah yang sebenarnya. Misalnya seseorang yang ingin menyembunyikan rasa bencinya terhadap orang lain namun, pada saat tertentu tanpa disadari akan menunjukkan perasaan di wajahnya walaupun orang tersebut pandai dalam menyembunyikan perasaannya. Maka dari itu, terkadang banyak orang yang salah dalam mendeteksi perasaan seseorang karena hanya melihat saat itu saja. Hal tersebut tentunya, memerlukan suatu pengamatan secara

terus menerus dalam melakukan perubahan ekspresi wajah seseorang (Prasetyawan et al., 2020).

2.1.2 FER (Facial Expression Recognition)

FER merupakan singkatan dari *Facial Expression Recognition* atau pengenalan ekspresi wajah. *FER* ini salah satu penerapan dalam pemanfaatan *Computer Vision* yang dapat menganalisis ekspresi wajah baik dari gambar statis maupun video untuk mengungkapkan suatu informasi tentang keadaan perasaan seseorang.

Ada beberapa langkah dalam analisis FER diantaranya, yaitu *face detection* atau deteksi wajah, *facial expression detection* atau deteksi ekspresi wajah, dan *expression classification to an emotional state* atau klasifikasi ekspresi ke keadaan emosional (Nafis et al., 2020). Deteksi ekspresi didasarkan pada analisis posisi landmark wajah misalnya pada ujung hidung, alis (Kusumastuti et al., 2021).

2.1.3 Deep Learning

Kecerdasan buatan atau *Artificial Intelligence (AI)* merupakan salah satu teknik yang dapat membuat komputer meniru seperti yang dilakukan oleh manusia. *AI* atau kecerdasan buatan memiliki metode yang dikembangkan salah satunya adalah metode machine learning (Ahmad Hania, A., 2017).

Machine Learning (ML) merupakan pendekatan dalam *AI* yang banyak digunakan untuk menirukan perilaku manusia dalam menyelesaikan suatu masalah atau melakukan otomatisasi. Ada beberapa teknik dalam *ML* salah satunya *Artificial Neural Network (ANN)* atau Jaringan Saraf Tiruan. *ANN*, yaitu salah satu teknik

yang menirukan saraf manusia yang merupakan bagian fundamental dari otak. Pada jaringan saraf tiruan terdiri dari *input layer*, *hidden layer*, dan *output layer*. Setiap lapis terdiri dari beberapa unit *neuron* yang memiliki fungsi aktivasi yang dapat menentukan keluaran dari unit tersebut (Ahmad Hania, A., 2017).

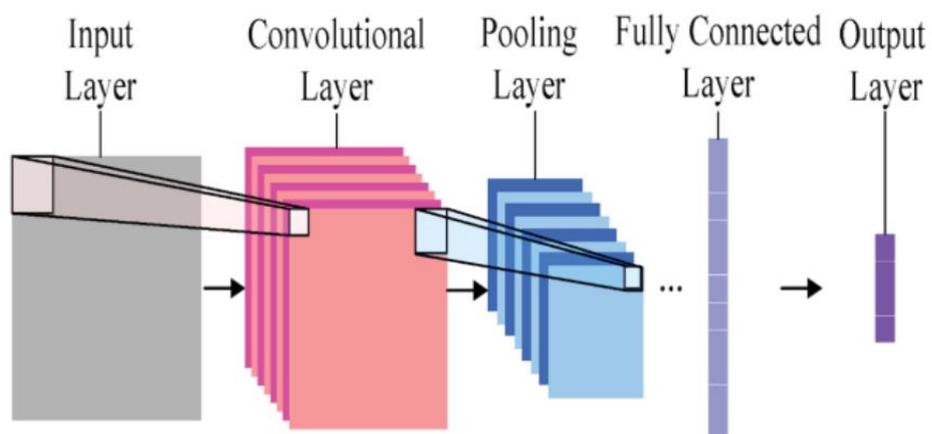
Maka dari itu pengertian dari *Deep Learning* merupakan salah satu bagian atau teknik dari *Artificial Neural Network (ANN)* yang dapat mempercepat proses pembelajaran dalam *Neural Network* pada penggunaan lapis atau *layer* yang banyak atau lebih dari tujuh lapis / *layer* (Ahmad Hania, A., 2017). Penggunaan metode *deep learning* sudah sangat populer dan sering digunakan saat ini, ada beberapa *library* yang digunakan dalam implementasi *deep learning*, yaitu penggunaan *library tensorflow* yang memiliki fitur untuk melatih model menggunakan *Graphics Processing Unit (GPU)*. Selain itu, *library* yang dipakai untuk implementasi *deep learning*, yaitu *Keras* yang menggunakan beberapa fitur dari *library tensorflow* dalam pembuatan jaringan saraf tiruan dikarenakan pada *library keras* telah terdapat beberapa model dasar CNN yang telah dioptimasi untuk mempermudah *Deep Learning* (Arsal et al., 2020).

2.1.4 Algoritma CNN

CNN merupakan salah satu algoritma dari *Deep Learning* yang memiliki kemampuan kecepatan dalam proses *training data* serta tersedia pada *library Keras* sehingga tidak memerlukan pembuatan fitur ekstraksi secara manual (Kusdiananggalih et al., 2021). Terdapat dua operasi di dalam CNN, yaitu operasi *basic convolution* dan operasi *pooling*. Operasi konvolusi digunakan untuk menyimulasikan sel-sel sederhana, sedangkan operasi *pooling* digunakan untuk

menyimulasikan sel-sel kompleks. Pada *CNN*, subblok atau area penerimaan lokal merupakan input yang berhubungan dengan lapisan atau *layer* terendah dari struktur hierarkis dalam gambar serta informasi ditransmisikan ke lapisan yang berbeda melalui lapisan demi lapisan (Xu et al., 2019).

CNN memiliki arsitektur yang hampir identik dengan pola komunikasi *neuron* otak manusia. Awalnya arsitektur *CNN* yang memfasilitasi pemrosesan persepsi manusia dengan memanfaatkan bidang reseptif lokal. Metode *CNN* ini memiliki *preprocessing* yang jauh lebih sedikit dibandingkan dengan algoritma klasifikasi lainnya. Tujuan dari arsitektur ini, yaitu menskalakan gambar ke jenis yang dapat diproses dengan mudah tanpa harus mengorbankan karakteristik vital untuk mendapatkan prediksi yang benar. Selain itu, cara kerja dari metode *CNN* ini, yaitu dengan melewati input gambar atau citra ke dalam kumpulan berbagai lapisan atau *layer* seperti *convolution layer*, *relu layer*, *pooling layer*, dan *Fully connected layer* dapat dilihat pada tampilan Gambar 2.1 (Ravi et al., 2020).



Gambar 2.1 Arsitektur *CNN* dengan 5 layer

a. *Convolution layer*

Convolution layer atau lapisan konvolusi berfungsi untuk mengumpulkan fitur – fitur yang bersifat valid dari gambar yang telah diinput serta mentransfernya ke *layer* yang akan datang. Selain itu, *layer* ini dapat mempertahankan hubungan spasial piksel dengan mengamati suatu karakteristik gambar (Ravi et al., 2020).

b. *Relu layer*

Relu merupakan *layer* yang bekerja dengan *output* dari *convolution layer* untuk membuat input menjadi nonlinier dengan bantuan suatu fungsi aktivasi (Ravi et al., 2020).

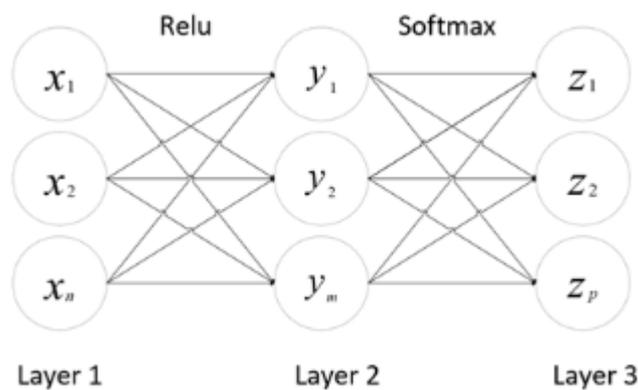
c. *Pooling layer*

Pooling layer memiliki fungsi untuk mengurangi varians, mengurangi kerumitan dalam perhitungan (Mehindra et al., 2020) serta meminimalkan dimensi spasial (Ravi et al., 2020). *Pooling layer* memiliki dua jenis, yaitu *Max pooling* serta *Average pooling*. *Max pooling* memiliki arti, yaitu mengembalikan nilai maksimum dari bagian gambar yang dicakup oleh *kernel*, sedangkan untuk *average pooling* memiliki arti mengembalikan nilai rata rata dari gambar yang dicakup oleh *kernel* (Alwanda et al., 2020). Walaupun kedua jenis *pooling* tersebut memiliki alasan yang sama, yaitu mengembalikan nilai yang berasal dari gambar yang dapat dicakup oleh *kernel* namun, *max pooling* lebih baik dalam hal mengekstrak fitur ekstrem, sedangkan *average pooling* terkadang tidak dapat mengekstrak fitur ekstrem yang baik dikarenakan *average pooling* membutuhkan semua untuk menghitung serta menghasilkan nilai rata-rata yang mungkin atau tidak terlalu penting dalam hal deteksi suatu objek. Sehingga dapat diartikan bahwa *average*

pooling membawa semua nilai ke dalam perhitungan yang digunakan untuk pemetaan dalam fitur serta dalam pembuatan keluaran yang merupakan perhitungan yang bersifat sangat umum. Namun jika tidak diperlukan semua masukan atau input dari *convolution layer*, maka nilai akurasi yang akan didapat menghasilkan akurasi yang buruk untuk *average pooling* (Mehindra et al., 2020).

d. *Fully connected layer*

Fully connected layer memiliki fungsi untuk melakukan suatu transformasi pada dimensi data agar dapat diklasifikasikan secara linier (Alwanda et al., 2020). Selain itu, proses klasifikasi pada *fully connected layer* menggunakan *softmax* yang sesuai yang dapat dilihat pada Gambar 2.2 (Mehindra et al., 2020).



Gambar 2.2 *Fully Connected Layer*

Pada gambar 2.2 *layer* satu dilakukan proses *feed forwarding* ke *layer* dua dengan menggunakan fungsi aktivasi *Relu*. Sehingga *layer* dua akan dilakukan proses klasifikasi dengan menggunakan *softmax*. *Fully connected layer* ini biasanya digunakan pada penggunaan metode *Multilayer perceptron* (MLP) yang memiliki tujuan untuk mengolah suatu data agar dapat diklasifikasikan (Mehindra et al.,

2020). Untuk mendapatkan hasil output atau keluaran dari *fully connected layer* ini tidak memerlukan suatu operasi konvolusi namun, menggunakan komputasi perkalian matriks yang diikuti dengan bias *offset*. Dengan penerapan operasi komputasi perkalian matriks tersebut, setiap *neuron* memiliki koneksi yang penuh ke semua aktivasi pada lapisan sebelumnya (Alwanda et al., 2020).

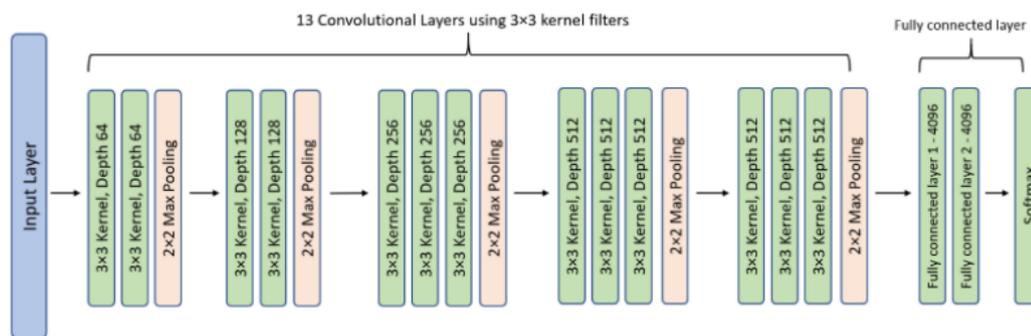
Selain itu, metode *CNN* memiliki beberapa jenis arsitektur yang sering digunakan dalam proses deteksi masker maupun ekspresi wajah diantaranya, yaitu:

a. VGG (*Visual Geometry Group*)

VGG (Visual Geometry Group) merupakan salah satu jenis arsitektur *CNN* yang terdiri dari lapisan konvolusi dengan *relu*, *max pooling* dan diakhiri dengan *fully connected layer* yang didalamnya berisikan 1000 kelas menggunakan fungsi *softmax* (Kusumastuti et al., 2021). Fungsi tersebut digunakan untuk mendapatkan rentang probabilitas keluaran diantara nilai 0 hingga 1, dan jumlah semua probabilitas akan sama dengan 1 (Kusumastuti et al., 2021). Filter yang digunakan pada arsitektur ini adalah 3x3 (Kusumastuti et al., 2021). Selain itu, jenis arsitektur *VGG* memiliki beberapa versi seperti *VGG16* yang di kembangkan oleh A. Zisserman dan K. Simonyan dari Universitas Oxford.

VGG16 memiliki jenis *layer* yang sama dengan versi *VGG* yang lain. Namun, yang membedakannya, yaitu hanya mendukung 16 *layer*. Jenis arsitektur ini pernah digunakan untuk penelitian “*Very Deep Convolutional Networks for Large-Scale Image Recognition*” oleh Zisserman dan K. Simonyan pada tahun 2014. Model *VGG16* tersebut memiliki akurasi mencapai 92,7% dalam penelitian tersebut. Serta dapat mengklasifikasikan gambar ke dalam 1000 kategori objek dan memiliki

ukuran input gambar 224 x 224 (Simonyan et al., 2015). Selain itu, pada penelitian tersebut menunjukkan nilai *error* pada *VGG16* memiliki nilai *error* terkecil dibandingkan dengan versi *VGG* yang lainnya (Achmad et al., 2019). Adapun bentuk arsitektur dari *VGG16* dapat dilihat pada gambar 2.3



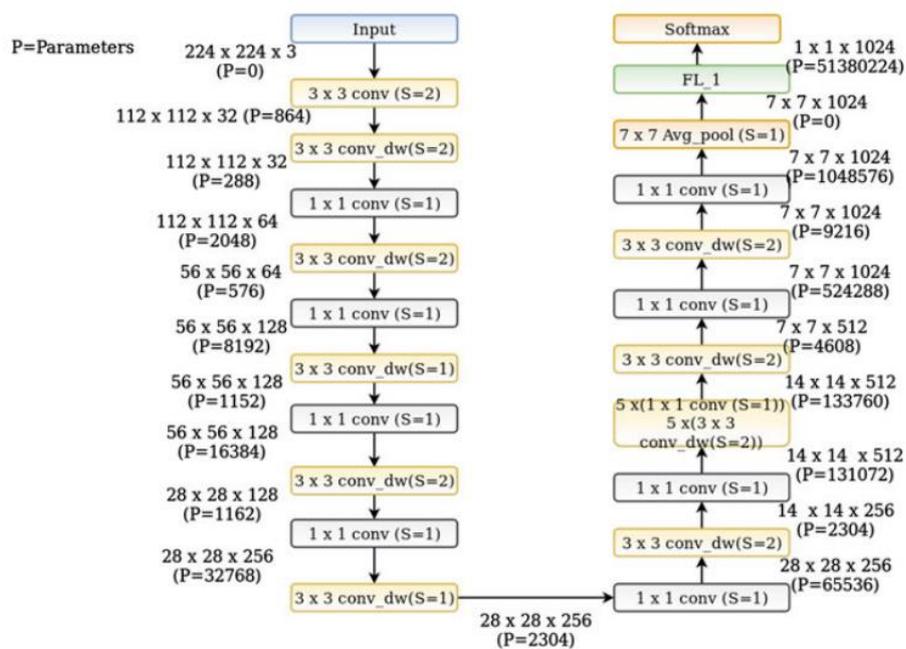
Gambar 2.3 Arsitektur *VGG16* (Rismiyati et al., 2021)

Pada Gambar 2.3 semua lapisan konvolusi memiliki ukuran *kernel* 3x3. Perbedaan utama tiap lapisan konvolusi terletak pada jumlah filter di masing masing lapisan. Dua lapisan konvolusi pertama memiliki jumlah filter 64. Selain itu, lapisan tiga dan empat memiliki jumlah kernel 128. Begitu juga dengan lapisan konvolusi lain mempunyai jumlah filter berbeda-beda, yaitu sebanyak 256 (lapisan 4,5,6) dan 512 (lapisan 7,8,9,10,11,12). *2x2 max pooling* dilakukan setelah lapisan konvolusi 2, 4, 7, 10 dan 13. Keluaran dari *pooling* terakhir dihubungkan ke *fully connected layer*, dan pada akhirnya akan terhubung ke *classifier* untuk menentukan kelas dari citra (Rismiyati et al., 2021)

b. Mobilnet

Mobilnet merupakan salah satu jenis arsitektur dari *CNN* atau bisa dikatakan salah satu *feature extraction* yang memiliki kinerja baik serta ringan untuk digunakan pada perangkat ponsel (Supriadi et al., 2021) dan dapat digunakan untuk

mengatasi kebutuhan computing data dengan jumlah yang besar (Hendriyana et al., 2020). Arsitektur dari *MobileNet* ini memiliki *depthwise convolution layer* serta *pointwise convolution layer* dalam mengekstrak *feature*. Arsitektur *mobilenet* memiliki perbedaan dengan arsitektur *CNN* lainnya, yaitu pada *mobilenet* menggunakan *convolution layer* dengan ketebalan filter sesuai dengan ketebalan dari *input image* (Supriadi et al., 2021). Adapun contoh bentuk dari arsitektur *MobileNet* dapat dilihat pada gambar 2.4



Gambar 2.4 Contoh Arsitektur *MobileNet* (Radhika et al., 2020)

MobileNet secara eksklusif menggunakan filter yang dapat dipisahkan, yang merupakan kombinasi dari *depthwise convolution* *point wise convolution*. Selain itu, *MobileNet* menggunakan filter 1×1 untuk mengurangi *overhead* komputasi dari operasi konvolusi normal. Hal tersebut membuat jaringan lebih ringan (seperti yang ditunjukkan pada Gambar 2.4) dalam hal kompleksitas komputasi dan juga

ukuran. MobileNet-224 mempunyai sekitar 5 juta parameter pada klasifikasi ImageNet. MobileNet juga mengambil *input* ukuran $224 \times 224 \times 3$ dan jumlah filter dapat bervariasi dari 32 hingga 1024 seperti pada Gambar 2.4 (Radhika et al., 2020).

2.1.5 Data Augmentation (Augmentasi data)

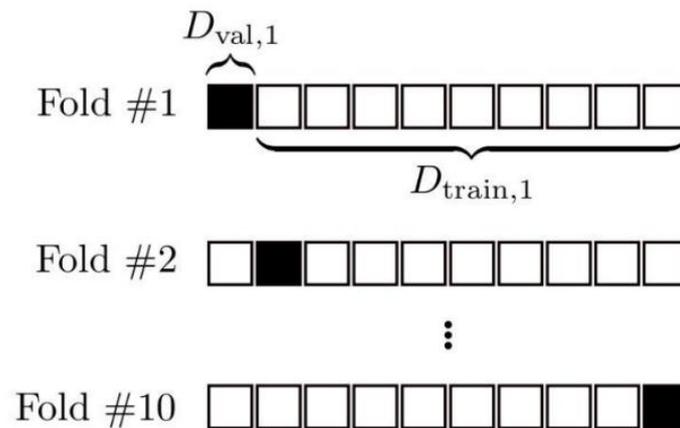
Augmentasi data merupakan suatu proses dalam pengolahan data gambar atau merupakan suatu proses memodifikasi gambar sedemikian rupa sehingga komputer akan mendeteksi bahwa gambar yang diubah merupakan gambar yang berbeda namun, masih dapat diketahui oleh manusia bahwa gambar tersebut adalah gambar yang sama. Selain itu, augmentasi data dapat meningkatkan akurasi dari model *CNN* yang dilatih karena dengan augmentasi data model mendapatkan data-data tambahan yang dapat berguna untuk membuat model yang dapat melakukan generalisasi dengan lebih baik (Kamal Hasan et al., 2019).

Teknik augmentasi data dapat diklasifikasikan menjadi dua kategori, yaitu *Geometric Augmentation* dan *Synthetic Augmentation*. *Geometric Augmentation* merupakan salah satu kelas teknik augmentasi yang melibatkan *cropping*, *zooming* dan *shifting* gambar asli. Sementara itu, *Synthetic Augmentation* adalah teknik yang memperkenalkan perubahan yang dibuat secara artifisial pada gambar asli (Kandel et al., 2022).

Brightness atau kecerahan adalah metode augmentasi yang tidak dapat dengan mudah ditugaskan ke kategori sebelumnya karena ini bukan transformasi geometris atau sintetik. Perubahan kecerahan pada gambar akan mengubah struktur dasarnya (Kandel et al., 2022).

2.1.6 Cross Validation

Validasi silang atau *cross validation* merupakan salah satu metode resampling data yang paling banyak digunakan untuk memperkirakan kesalahan prediksi sebenarnya dari suatu model serta salah satu metode yang digunakan untuk mencegah terjadinya kondisi *overfitting*. Selain itu, teknik *cross validation* yang sering digunakan adalah *k-fold cross validation*. Dalam *k-fold cross validation* set pembelajaran yang tersedia di partisi menjadi k memisahkan himpunan bagian dengan ukuran yang kira-kira sama. Sementara itu *fold* mengacu terhadap jumlah himpunan bagian yang dihasilkan. Adapun contoh ilustrasi proses *k-fold cross validation* dapat dilihat pada tampilan gambar 2.5 dengan menerapkan nilai $k=10$ yang merupakan proses *cross validation* dilakukan sebanyak 10 kali. Kumpulan data dibagi secara acak menjadi sepuluh himpunan bagian terpisah masing masing berisi kira-kira 10% dari data. Kemudian, pada *fold* pertama di *subset* pertama berfungsi sebagai data validasi ($D_{val,1}$) dan sisa sembilan *subset* berfungsi sebagai data latih (D_{train}). Selanjutnya, pada *fold* kedua, himpunan bagian kedua adalah data validasi dan subset yang tersisa adalah data latih dan seterusnya (Berrar, 2018).



Gambar 2.5 K-Fold 10 Cross Validation

2.1.7 Overfitting dan Underfitting

Overfitting dan *underfitting* merupakan salah satu masalah pada *neural network* yang menyebabkan hasil prediksi tidak akurat dan kurang optimal. Kondisi *overfitting* terjadi ketika perbedaan antara *training error* dan *test error* terlalu besar. Di sisi lain, kondisi *overfitting* dapat terjadi saat *neural network (NN)* yang terlalu bergantung pada set pelatihan dan mempelajari pemetaan yang salah yang bekerja dengan baik di set pelatihan, tetapi berkinerja buruk di validasi atau set pengujian. Ada beberapa kemungkinan penyebab untuk masalah ini. Jika arsitektur *NN* terlalu rumit dalam rentang tertentu sementara jumlah data pelatihan tidak mencukupi, *NN* dapat mempelajari karakteristik data pelatihan terlalu baik sehingga mereka terlalu fokus pada fitur spesifik dari set pelatihan dan mengabaikan fitur umum. Jika distribusi karakteristik set pelatihan berbeda dari set validasi atau pengujian, *NN* juga tidak dapat mempelajari pemetaan yang dapat diadaptasi. Kedua situasi tersebut dapat menyebabkan masalah *overfitting* (Zhang et al., 2019). Selain itu, model yang dilatih dengan kumpulan data yang tidak seimbang, model tersebut

dapat disesuaikan dengan sampel pelatihan dari kelas yang kurang terwakili dan mungkin tidak dapat digeneralisasi dengan baik selama waktu pengujian sehingga hal tersebut dapat menjadi salah satu efek terjadinya kondisi *overfitting* (Li et al., 2021).

Ada beberapa alternatif untuk menangani model yang mengalami kondisi *overfitting* diantaranya, yaitu:

1. Menambahkan penerapan regulasi eksplisit pada *loss* yang dapat menghambat kompleksitas parameter model (Rice et al., 2020).
2. Menambahkan penerapan augmentasi data yang telah terbukti secara empiris dapat mengurangi *overfitting* dengan data berdimensi sangat tinggi dengan meningkatkan jumlah dan variasi data pelatihan (Rice et al., 2020).
3. Penerapan *dropout* dengan membuang informasi secara acak membidik setiap *node* tersembunyi dari jaringan saraf selama fase pelatihan. Dengan cara tersebut, jaringan dapat menikmati efek *ensemble* dari *subnetwork* kecil sehingga mencapai efek regularisasi yang baik. Namun, tidak seperti *fully connected layers*, menerapkan *dropout* ke *convolutional feature map* tidak efektif. Salah satu alasannya adalah bahwa piksel yang berdekatan secara spasial dalam *convolutional feature map* sangat berkorelasi karena dapat berbagi informasi kontekstual yang berlebihan. Oleh karena itu, *convolutional pixel-based dropout* tidak dapat sepenuhnya membuang informasi pada *convolutional feature map* (Choe et al., 2019)

Sementara itu, kondisi *underfitting* terjadi ketika model tidak mampu memperoleh nilai *error* yang cukup rendah pada training set. Kondisi *underfitting* terjadi ketika NN tidak dapat menangkap karakteristik set pelatihan dan tidak dapat menyesuaikan pemetaan target dengan baik, yang akan menyebabkan tingkat akurasi yang rendah dan nilai kerugian yang tinggi untuk ketiga set data tersebut. Faktor yang mungkin menyebabkan masalah *underfitting* adalah arsitektur *NN* terlalu sederhana dan memiliki terlalu sedikit lapisan tersembunyi atau parameter yang dapat dilatih, yang membuat *NN* tidak cukup kuat untuk menangkap karakteristik data yang kompleks. Kemungkinan lain adalah *NN* belum cukup terlatih untuk mempelajari karakteristik atau distribusi data. Untuk menangani suatu model yang mengalami kondisi *underfitting* salah satunya, yaitu menerapkan metode regulasi yang biasanya memodifikasi *loss function* (Zhang et al., 2019).

2.1.8 Confusion Matrix

Confusion matrix adalah salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem atau model dengan hasil klasifikasi yang seharusnya. Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat empat istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)*. Nilai *True Negative (TN)* merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive (FP)* merupakan data negatif namun, terdeteksi sebagai data positif

(Karsito et al., 2019). Adapun perhitungan nilai *accuracy*, *precision* dan *recall* berdasarkan *confusion matrix* dapat dilihat pada persamaan 2.1, 2.2 dan 2.3

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2.1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2.2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (2.3)$$

2.1.9 F1-Score

F1-Score menggambarkan perbandingan rata-rata *precision* dan *recall* yang dibobotkan. *Accuracy* tepat digunakan sebagai acuan performansi algoritma jika dataset memiliki jumlah data *False Negatif* dan *False Positif* yang sangat mendekati (*symmetric*). Namun jika jumlahnya tidak mendekati, maka menggunakan *F1-Score* sebagai acuan. Adapun persamaan untuk menghitung nilai *F1-score* dapat dilihat pada persamaan 2.4

$$F1 - Score = \frac{(2 \times Recall \times Precision)}{(Recall + Precision)} \quad (2.4)$$

2.1.10 Python

Python merupakan salah satu bahasa pemrograman tingkat tinggi serta mudah untuk dipelajari karena memiliki sintaks yang jelas dan dapat dikombinasikan dengan penggunaan modul yang siap pakai serta struktur data tingkat tinggi yang efisien. Bahasa pemrograman *python* memiliki beberapa sifat diantaranya, yaitu interpreter, *interactive*, *object-oriented* dan dapat beroperasi hampir di semua *platform* seperti *Mac*, *Linux* dan *Windows* (Zulkhaidi et al., 2020).

2.1.11 Google Colab

Google Colab merupakan suatu layanan bersifat *open source* yang disediakan oleh *google* untuk setiap orang yang memiliki akun *Gmail*. Selain itu, *google colab* menyediakan 12,72 GB RAM dan 358,27 GB ruang hard disk dalam satu waktu proses. Setiap proses *runtime* berlangsung selama 12 jam setelah *runtime* diatur ulang dan pengguna harus membuat koneksi lagi. Hal tersebut bertujuan agar layanan GPU yang dimiliki *google colab* tidak digunakan oleh orang yang melakukan proses penambangan mata uang kripto serta tujuan illegal lainnya. Saat pengguna membuka *file google colab*, mereka diharuskan untuk memilih jenis waktu proses diantaranya, yaitu (Kanani et al., 2019) :

- a. *None* yang berarti bahwa yang akan menggunakan *CPU* komputer yang digunakan pengguna.
- b. *GPU*
- c. *TPU* digunakan untuk pemrosesan *tensor*

2.2 Penelitian Terkait dan Keterbaruan Penelitian

Penelitian terkait akan menjawab pertanyaan yang berhubungan dengan deteksi penggunaan masker, ekspresi wajah serta deteksi ekspresi wajah bagi orang yang menggunakan masker dapat dilihat pada tabel 2.1 *state of the art*.

Tabel 2.1 *State of The art*

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
1.	Wikarta dkk.(2020)	Deteksi pemakaian masker pengemudi kendaraan.	<ol style="list-style-type: none"> a. <i>CNN</i> b. <i>Optimizer ADAM</i> c. <i>Optimizer SGD</i> 	Hasil dari penelitian paper ini, yaitu hasil <i>training CNN</i> yang dilakukan menunjukkan bahwa pemilihan suatu jenis <i>optimizer</i> akan memberikan tingkat akurasi serta losses yang

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
			d. <i>Optimizer RMSprop</i>	berbeda-beda. Nilai akurasi untuk <i>CNN</i> dengan menggunakan <i>optimizer SGD</i> mendapatkan nilai 0.75777. Kemudian, untuk nilai akurasi <i>CNN</i> dengan menggunakan <i>optimizer RMSprop</i> mendapatkan nilai 0.9577, sedangkan untuk penerapan <i>optimizer ADAM</i> pada <i>CNN</i> mendapatkan nilai 0.9654 yang menunjukkan bahwa <i>optimizer SGD</i> memiliki tingkat akurasi rendah. Selain itu, <i>optimizer</i> yang memiliki tingkat akurasi tertinggi pada <i>CNN</i> adalah <i>ADAM</i> .
2.	Ramdhani dkk. (2021)	Deteksi masker wajah.	a. <i>CNN</i> b. <i>MobilNet</i> c. <i>Single shot multibox detector</i>	Hasil penelitian pada paper ini, yaitu pada penelitian ini terdapat beberapa tahapan diantaranya tahapan pengumpulan data, <i>preprocessing</i> , <i>training model</i> , <i>testing model</i> , dan implementasi model. Model yang dibangun dalam penelitian ini mendapatkan akurasi sebesar 99% baik untuk tahapan <i>testing</i> maupun pada tahapan <i>training</i> .
3.	Nyoman dkk. (2021)	Deteksi jenis masker yang umum digunakan masyarakat.	a. <i>CNN</i> b. <i>MobileNet</i> c. <i>Optimizer ADAM</i> d. <i>Optimizer gradient descent</i>	Hasil pada penelitian ini menghasilkan suatu nilai akurasi sebesar 90 %, untuk optimasi <i>ADAM</i> dan 80% untuk optimasi <i>Gradient Descent</i> . Berdasarkan hasil dari pendeteksian jenis masker dengan metode <i>Convolutional Neural Network</i> menggunakan optimasi <i>ADAM</i> dan <i>gradient descent</i> dapat dinilai bekerja dengan baik.
4.	Hermawati dkk. (2021)	Pemakaian masker medis dan masker non medis serta pemakaian masker yang tidak benar dan benar serta orang	a. <i>CNN</i> b. <i>Viola Jones</i>	Pada penelitian ini menghasilkan suatu sistem deteksi pemakaian masker yang dapat mengenali pemakaian masker yang benar dan yang tidak benar serta membedakan masker <i>medis</i> dan <i>nonmedis</i>

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
		yang tidak memakai masker.		yang merupakan sebuah inovasi dalam meningkatkan ketertiban pemakaian masker di kondisi wabah virus <i>Covid -19</i> . Percobaan dilakukan menggunakan 1300 data <i>training</i> dengan akurasi sebesar 84.23%.
5.	Darmatasia (2020)	Deteksi penggunaan masker.	<ul style="list-style-type: none"> a. <i>CNN</i> b. <i>Transfer learning</i> c. <i>VGG16</i> d. <i>ResNet101 V2</i> e. <i>Inception V3</i> f. <i>MobilNet</i> g. <i>DenseNet 201</i> h. <i>NASNet Large</i> 	Hasil penelitian ini, yaitu secara keseluruhan <i>InceptionV3</i> memiliki performa yang paling baik berdasarkan nilai akurasi, <i>presisi</i> , <i>recall</i> , dan <i>F1-score</i> pada dataset yang digunakan. Pengukuran performa waktu komputasi menunjukkan bahwa <i>MobileNet</i> memberikan waktu komputasi yang paling cepat dengan akurasi yang terbilang cukup baik. <i>MobileNet</i> memberikan waktu komputasi dua kali lebih cepat jika dibandingkan dengan <i>Inception V3</i> . Selain itu, <i>MobileNet</i> juga memiliki keunggulan lain dibandingkan dengan <i>Inception V3</i> , <i>MobileNet</i> dapat diimplementasikan pada perangkat dengan kemampuan komputasi yang rendah seperti perangkat <i>mobile</i> .
6.	Ihsan dkk. (2021)	Deteksi ekspresi wajah untuk klasifikasi tujuh ekspresi dasar.	a. <i>CNN</i>	Hasil dari penelitian ini, yaitu klasifikasi ekspresi wajah menggunakan metode Convolutional Neural Network (<i>CNN</i>) dengan menggunakan library TensorFlow berhasil diimplementasikan untuk mendeteksi ketujuh ekspresi, yakni marah, senang, sedih, jijik, terkejut takut dan neutral. Dihasilkan nilai parameter optimal batch 32, epoch 100, dan dropout 0,6; dengan akurasi training 61,15%; akurasi validasi 64,54%; training loss 4,29%; dan validasi loss 3,96%; dari hasil akurasi tersebut dapat disimpulkan bahwa sistem tidak

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
				mengalami overfit karena nilai akurasi validasi lebih besar dari nilai akurasi training. Parameter Optimal tersebut diujikan kepada data asing yang belum pernah dideteksi sebelumnya dan menghasilkan nilai akurasi deteksi 67%, presisi 67% dan recall 66%.
7.	Tanuwijaya dkk. (2021)	Deteksi ekspresi wajah murid yang ada pada aplikasi Zoom atau Google Meet.	a. <i>Pipeline YOLO-face</i> b. <i>CNN</i>	Hasil pada penelitian ini, yaitu didapatkan akurasi sebesar 0.94 saat latih dan 0,69 saat validasi. Kedua akurasi ini melebihi akurasi dari model <i>CNN Alexnet</i> . Selain itu, juga <i>precision</i> dan <i>recall</i> dari model modifikasi <i>CNN</i> adalah sebagai berikut secara berurutan 0.92, dan 0.96 pada fase latih dan 0,66 dan 0,71 pada fase validasi. Sehingga peneliti menggunakan modifikasi <i>CNN</i> ini sebagai model untuk klasifikasi hasil dari deteksi wajah <i>YOLO face</i> .
8.	Kusumastuti dkk. (2021)	Deteksi ekspresi wajah dan gestur pada saat siswa memperhatikan video pembelajaran yang diajarkan oleh gurunya.	a. <i>CNN</i> b. <i>VGG16</i> c. <i>ResNet50</i> d. <i>Xception</i>	Hasil dari penelitian ini, yaitu klasifikasi ketertarikan dengan menggunakan <i>CNN</i> yang memanfaatkan transfer learning pada arsitektur <i>VGG16</i> , <i>ResNet 50</i> dan <i>Xception</i> hasil terbaik, yaitu pada saat menggunakan arsitektur <i>VGG16</i> . Hal tersebut telah terlihat sejak proses pelatihan, dengan hasil validasi akurasi area mulut dan gestur lebih baik dibandingkan kedua model lainnya, sedangkan pada area wajah, mendapatkan hasil validasi akurasi yang seimbang, dan validasi akurasi 86% pada area mata. Hal tersebut dibuktikan pada skenario 4 area potong. Hasil pengujian klasifikasi ketertarikan belajar memperoleh persentase validasi minimum 70%. Nilai persentase tersebut didapatkan melalui skenario tiga kelas ketertarikan empat area potong

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
				menggunakan <i>VGG16</i> , yaitu sebesar 75%, sedangkan untuk dua area potong menggunakan <i>ResNet50</i> dapat mencapai 71%.
9.	Reynaldo dkk. (2020)	Klasifikasi ekspresi wajah.	a. <i>Viola Jones</i> b. <i>CNN</i>	telah dilakukan, sesuai dengan arsitektur <i>CNN</i> yang diusulkan dengan parameter <i>epoch</i> 10 dan <i>learning rate</i> 0,001 diperoleh akurasi sebesar 96,14%. Terdapat beberapa temuan yang diperoleh, yaitu akurasi yang kecil disebabkan karena banyaknya informasi yang hilang pada <i>feature map</i> hasil dari proses konvolusi kelima. Kemudian penggunaan segmentasi <i>Viola Jones</i> pada saat proses segmentasi dapat menghasilkan peningkatan akurasi yang cukup signifikan.
10.	Kusdianan ggalih dkk. (2021)	Pengenalan ekspresi wajah dengan dataset ras Indonesia	a. <i>CNN</i>	Hasil dari penelitian ini , yaitu, kinerja dari sistem yang dibangun menggunakan <i>Convolutional Neural Network</i> berjumlah delapan lapisan konvolusi menggunakan data augmentation dan diuji menggunakan metode <i>cross dataset</i> citra orang-orang Indonesia didapat nilai akurasi sebesar 91,29%, dengan sensitivity atau recall atau <i>True Positive Rate (TPR)</i> sebesar 91,29%, <i>precision</i> atau <i>Positive Predictive Value (PPV)</i> sebesar 91,29%, dan <i>accuracy</i> sebesar 97,51%. Oleh karena itu, dengan nilai <i>recall</i> dan <i>precision</i> yang tinggi, artinya kelas pada data uji ditangani dengan sempurna oleh model yang dibangun.
11.	Cotter (2020)	Pengenalan ekspresi wajah pada <i>smartphones</i>	b. <i>CNN</i> c. <i>MobilNet</i> d. <i>MobilEx</i> e. <i>ResNet</i>	Hasil dari penelitian ini, yaitu memperkenalkan model <i>deep learning</i> baru yang ringan dalam komputasi, yaitu model <i>MobilExpressNet</i> untuk <i>FER</i> . Ukuran model dan <i>FLOP</i> terbukti lebih dari lima kali lebih kecil

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
				daripada model <i>MobileNet</i> terkecil dan model tersebut memberikan akurasi <i>FER</i> yang sangat baik sebesar 67,96% pada dataset <i>FER2013</i> .
12.	Yang dkk. (2021)	Pengenalan ekspresi wajah (<i>FER</i>) tiga jenis ekspresi pada orang bermasker.	<ul style="list-style-type: none"> a. <i>CNN</i> b. <i>MobilNet V2</i> c. <i>VGG19</i> 	Hasil pada penelitian ini, yaitu penelitian ini mengusulkan model perhatian dua tahap yang terdiri dari pengklasifikasi dalam biner bermasker/terbuka dan pengklasifikasi <i>FER</i> untuk meningkatkan akurasi <i>FER</i> pengenalan wajah. Kemudian, pada penelitian ini lebih memperhatikan daerah yang tidak terhalang daripada bagian wajah bermasker, yang kemudian memperhatikan daerah yang penting untuk klasifikasi ekspresi dari kedua daerah bermasker dan tidak bermasker. Metode yang diusulkan dapat mengungguli metode pengenalan ekspresi wajah lainnya pada dataset ekspresi wajah bermasker. Pada paper ini, hanya mempertimbangkan klasifikasi <i>FER</i> dengan tiga kategori emosi: positif, negatif, dan netral.
13.	Grundman dkk (2021)	Masker wajah mengurangi akurasi pengenalan ekspresi dan kedekatan yang dirasakan.	<ul style="list-style-type: none"> a. Regresi bertingkat (logistik). b. Optimizer <i>nlminb</i>. c. <i>Glmer</i> 	Hasil dari penelitian ini, yaitu akurasi pengenalan emosi atau ekspresi secara keseluruhan menurun dari 69,9 % untuk yang tidak membuka masker serta 48,9 % untuk yang memakai masker. Masker wajah secara efektif mengurangi penyebaran COVID-19, Namun, mereka memiliki konsekuensi tambahan untuk kesimpulan emosional dan penilaian sosial. Masker wajah mengganggu kemampuan orang untuk secara akurat mengklasifikasikan ekspresi emosional yang cenderung

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
				diperburuk untuk orang dewasa yang lebih tua, mengurangi kedekatan yang dirasakan DAN meningkatkan persepsi kedekatan, kepercayaan, dan kesukaan terhadap target yang mengekspresikan emosi negatif. Selain itu, relevansi faktor psikologis dalam konteks virus corona dan menginformasikan pembuatan kebijakan dengan menggambarkan efek (samping) masker pada fungsi sosial.
14.	Castellano dkk (2021).	Pengenalan ekspresi otomatis dari ekspresi wajah saat mengenakan masker.	<ul style="list-style-type: none"> a. <i>CNN</i> b. <i>VGG16</i> c. <i>MobileNetV2</i> d. Optimizer <i>ADAM</i> 	<p>Hasil utama dalam penelitian ini, yaitu bertujuan untuk menilai keakuratan pendekatan yang diusulkan dalam mengenali emosi meskipun oklusi karena masker serta membandingkan hasil ini dengan yang diperoleh dengan pengenalan ekspresi atau emosi dari seluruh wajah. Akurasi global FER hanya berdasarkan ROI mata lebih rendah daripada akurasi FER berdasarkan seluruh wajah. Namun, dapat mengamati bahwa emosi positif (Bahagia dan Kejutan, dengan akurasi rata-rata 79,00%) dikenali lebih baik daripada emosi negatif (Angry, Disgust, Fear dan Sad, dengan akurasi rata-rata 23,00%) meskipun bertopeng. Ekspresi Jijik tidak pernah dikenali saat memakai topeng, yang diekspresikan terutama dengan bagian bawah wajah. Melihat matriks kebingungan FER saat mengenakan topeng dan membandingkannya dengan matriks FER berdasarkan seluruh Wajah ,</p>

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
				<p>kita dapat melihat bahwa perilaku pengenalan emosi dari sistem <i>FER</i> mirip dengan yang manusia. Memang, seperti yang dilakukan manusia, sistem <i>FER</i> memiliki kekurangan dalam mendeteksi kemarahan terutama dengan jijik, ketakutan, dan kesedihan, yang semuanya merupakan emosi dengan valensi negatif. Kebahagiaan diakui lebih baik daripada yang lain meskipun ada masker dan itu terutama dikacaukan dengan ekspresi netral.</p>
15.	Genc dkk (2020).	Desain masker wajah untuk mengurangi oklusi ekspresi wajah.	Pendekatan <i>Wizard of Oz</i>	<p>Hasil dari penelitian ini, yaitu dengan memanfaatkan teknologi tampilan elektrokromik (<i>EC</i>), penelitian ini dapat merancang dan membuat masker wajah pintar yang menghadirkan langkah yang dapat dipercaya menuju solusi yang dapat diproduksi. Sementara dalam penelitian ini tidak mengeksplorasi mekanisme untuk mengontrol perpindahan tampilan. Pendekatan dalam penelitian ini untuk mengevaluasi masker wajah melalui percakapan video dimandatkan oleh aturan isolasi sosial yang berlaku pada saat penelitian. Namun, pendekatan yang dikembangkan berhasil dengan baik dan menyampaikan temuan yang juga berlaku untuk percakapan tatap muka yang dilakukan secara fisik. Penelitian ini mencatat kesepakatan umum antara pendapat peserta uji tentang prototipe masker dan menggunakan konteks</p>

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
				<p>serta percaya bahwa kumpulan data yang digunakan mendekati saturasi. Sementara beberapa gerakan interaksi sosial memerlukan interpretasi budaya, fokus pekerjaan kami, senyum, dipahami secara universal. Masker wajah pintar, termasuk tampilan yang menampilkan pemakainya senyum tertutup dianggap berguna dalam konteks publik untuk mendukung ritual sanksi sosial positif informal yang diharapkan secara sosial. Dalam interaksi yang begitu singkat penyajian 2 keadaan, netral dan tersenyum dianggap memadai. Untuk percakapan yang lebih lama dan saat berbicara dengan keluarga dan teman dekat, masker dianggap tidak berguna. Umumnya visualisasi yang secara langsung mewakili mulut pemakainya lebih disukai daripada visualisasi gaya emoji wajah penuh.</p>
16.	Pei dkk. (2019)	Pengenalan Wajah melalui Pembelajaran Mendalam Menggunakan Augmentasi Data Berdasarkan Eksperimen Ortogonal	<p>a. <i>CNN</i> b. <i>VGG16</i></p>	<p>Menurut penelitian yang dilakukan oleh Pei dkk. (2019) yang menerapkan teknik data augmentation dalam pengenalan wajah menggunakan model <i>CNN</i> dengan arsitektur <i>VGG16</i> untuk menangani kurangnya sampel data adalah dengan menerapkannya <i>transformasi geometris</i>, <i>brightness augmentation</i>, <i>image translation</i>, <i>image rotation</i>, <i>image zoom</i> dan <i>filter operation</i>. Hasil penelitiannya menunjukkan bahwa penerapan model <i>CNN</i> dengan data augmentation dapat mencapai akurasi 86,3% lebih</p>

No.	Peneliti	Objek Penelitian	Metode	Hasil dan Akurasi
				tinggi dari metode <i>PCA</i> atau <i>LBPH</i> .
17.	Kandel dkk. (2022)	Kecerahan sebagai Teknik Augmentasi untuk Klasifikasi Citra pada Dataset Histopathology yaitu pada Invasive Ductal Carcinoma Dataset.	<ul style="list-style-type: none"> a. <i>CNN</i> b. <i>Resnet50</i> c. <i>DenseNet121</i> d. <i>Inception V3</i> e. <i>Xception</i> 	<p>menurut penelitian yang dilakukan oleh Kandel dkk. (2022) menerapkan teknik augmentasi data untuk klasifikasi tentang <i>histopathology</i> khususnya pada <i>Invasive Ductal Carcinoma Dataset</i> yang menggunakan dua jenis data augmentasi yaitu <i>geometric augmentation</i> dan <i>brightness augmentation</i> yang menerapkan delapan skala kecerahan dengan menggunakan empat model <i>CNN</i> yaitu <i>Resnet50</i>, <i>DenseNet121</i>, <i>InceptionV3</i> dan <i>Xception</i>. Adapun hasil penelitiannya adalah dengan adanya penerapan <i>geometric augmentation</i> memberikan nilai akurasi yang lebih baik dibandingkan dengan penerapan <i>brightness augmentation</i>. Selain itu, dengan model <i>CNN</i> tanpa menerapkan teknik augmentasi data memberikan hasil yang lebih baik dibandingkan dengan adanya penerapan <i>brightness augmentation</i>. Sehingga dapat dikatakan bahwa pada penelitian yang dilakukan oleh Kandel dkk. penerapan <i>brightness augmentation</i> dapat menurunkan kinerja model secara signifikan saat diterapkan dengan nilai ekstrim.</p>

2.2.1 Matriks Penelitian

Beberapa jurnal terkait berhubungan dengan penggunaan arsitektur komunikasi, teknologi, tujuan dan objek penelitian dengan penelitian yang sedang dilakukan. Tabel 2.2 menggambarkan perbedaan penelitian yang diusulkan dengan penelitian-penelitian terkait.

2.2.2 Relevansi Penelitian

Tabel 2.3 Relevansi Penelitian

Peneliti	Yang dkk, 2021	Desty Mustika Ramadhan, 2023
Judul	<i>Face mask aware robust facial expression recognition during the covid-19 pandemic</i>	Penerapan Brightness Augmentation Untuk Mengetahui Pengaruh Terhadap Akurasi Klasifikasi Ekspresi Wajah Bermasker Dengan Model CNN
Masalah Penelitian	Penyebaran <i>Covid-19</i> di seluruh dunia, mengenakan masker wajah saat berinteraksi di depan umum menjadi perilaku umum untuk melindungi diri dari infeksi. Maka dari itu, bagaimana meningkatkan efektivitas teknologi pengenalan ekspresi wajah (<i>FER</i>) yang ada pada wajah bermasker.	Penelitian Yang dkk hanya memiliki 3 kategori, yaitu <i>positif</i> , <i>negative</i> dan netral. Sistem yang dirancang menggunakan arsitektur CNN Yang dkk menggunakan <i>VGG19</i> dan <i>MobileNetV2</i> . Dari penggunaan jenis arsitektur yang dilakukan oleh penelitian tersebut menjadi suatu acuan dalam penerapan <i>transfer learning</i> dengan penerapan data augmentasi seperti <i>geometric augmentation</i> dan <i>brightness augmentation</i> . Maka dari itu masalah dalam penelitiannya ini bagaimana perbedaaan performa nilai akurasi antara arsitektur <i>VGG16</i> dan <i>MobileNet</i> dalam mengenali dan mengklasifikasikan ekspresi wajah bermasker? Serta bagaimana nilai <i>brightness range</i> yang optimum mempengaruhi performa nilai akurasi klasifikasi ekspresi wajah bermasker menggunakan arsitektur <i>VGG16</i> dan <i>MobileNet</i> ?
Objek Penelitian	Pengenalan ekspresi wajah bagi orang bermasker dengan tiga jenis ekspresi dasar , yaitu netral, positif dan negative.	Klasifikasi ekspresi wajah bagi orang bermasker dengan tujuh ekspresi dasar , yaitu jijik, marah, netral, senang, sedih, takut, terkejut.
Algoritma / Metode	<i>CNN (MobilNeV2, VGG19)</i>	<i>CNN (MobilNet, VGG16)</i> dengan <i>Transfer Learning</i> , <i>Data Augmentation (Geometric Augmentation, Brightness Augmentation)</i> dan <i>Cross Validation</i>
Dataset	<i>M-LFW-FER</i> dan <i>M-KDDI-FER</i>	<i>MaskedDatasetFER</i>