

BAB II

KAJIAN PUSTAKA

Bab ini membahas tinjauan pustaka yang relevan dengan permasalahan utama yang menjadi fokus penelitian dalam disertasi ini. Sub-bab 2.1 mengulas tren penelitian terkini dan topik-topik utama yang mendasari motivasi penelitian ini, termasuk metode ekstraksi ciri tekstur dan pengenalan ekspresi wajah yang telah dikembangkan. Keaslian atau kebaruan penelitian ini dijelaskan secara rinci pada sub-bab 2.2, dengan menyoroti kontribusi unik yang diusulkan dibandingkan penelitian sebelumnya. Sub-bab 2.3 memberikan penjelasan menyeluruh tentang landasan teori yang mendukung pendekatan yang diusulkan, mencakup konsep-konsep fundamental yang menjadi dasar pengembangan model atau metode penelitian ini.

2.1 Tinjauan Pustaka

2.1.1 Tren dan Topik Ekstraksi Ciri Tekstur dan Kesenjangan Penelitian

Pengembangan model klasifikasi untuk pengenalan ekspresi wajah melalui ekstraksi ciri adalah tren signifikan dalam penelitian visi komputer. Analisis literatur terbaru menunjukkan bahwa upaya penelitian dalam mengembangkan model pengenalan ekspresi wajah berfokus pada tiga pendekatan utama: ekstraksi ciri geometris (ECG), ekstraksi ciri mendalam (ECM), dan pendekatan hibrid yang menggabungkan kedua teknik.

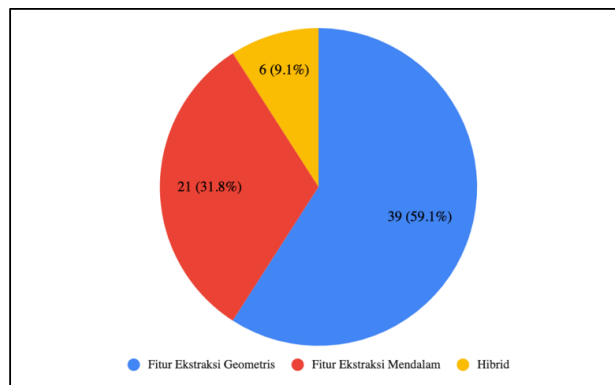
Pendekatan **Ekstraksi Ciri Geometris** (ECG) menggunakan teknik prapengolahan citra yang mengekstrak ciri manual dari fitur wajah. Ciri geometris, seperti posisi, bentuk, dan ukuran elemen wajah, membantu model pembelajaran mesin mengenali ekspresi wajah secara lebih terstruktur dan spesifik. ECG menawarkan kontrol tinggi atas ciri yang digunakan, menjadikannya berguna dalam aplikasi yang menuntut interpretasi yang jelas, seperti diagnosis medis atau

pengambilan keputusan terkait keselamatan. Namun, pendekatan ini memerlukan keterampilan dan waktu yang signifikan, terutama dalam menangani data berukuran besar, dan cenderung kurang efisien dibandingkan pendekatan otomatis.

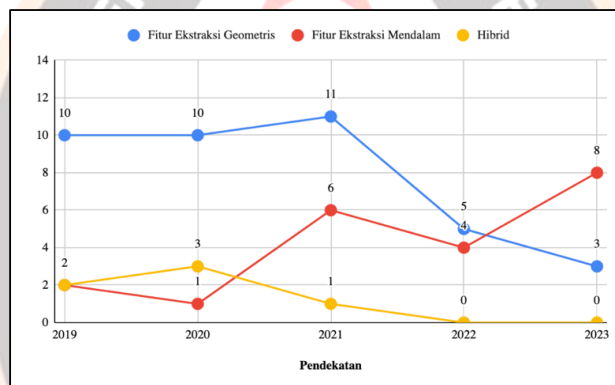
Pendekatan **Ekstraksi Ciri Mendalam** (ECM) memanfaatkan jaringan saraf konvolusional (CNN) untuk menghasilkan representasi ciri yang lebih kompleks dan hierarkis. Jaringan CNN secara bertahap mengidentifikasi ciri yang semakin abstrak dari data, menjadikannya lebih efisien dalam memahami pola non-linear dan ciri rumit pada citra wajah. ECM unggul dalam menghasilkan akurasi yang tinggi, terutama ketika basisdata yang digunakan besar dan variatif, namun rentan terhadap overfitting ketika jumlah data terbatas. Selain itu, ECM biasanya membutuhkan prapengolahan data, seperti normalisasi dan augmentasi, untuk memastikan kualitas ciri yang dihasilkan.

Pendekatan **Hibrid** menggabungkan ECG dan ECM, memanfaatkan keunggulan dari kedua metode untuk menghasilkan representasi ciri yang lebih adaptif dan komprehensif. Gabungan ini memungkinkan model pengenalan ekspresi wajah untuk mengenali ciri geometris spesifik sekaligus menangkap detail yang lebih abstrak dan kompleks, meningkatkan ketepatan serta ketahanan model terhadap berbagai variasi data. Hasil kajian pustaka menunjukkan bahwa pendekatan hibrid memiliki potensi yang signifikan dalam meningkatkan kinerja model pengenalan ekspresi wajah, meskipun penelitian pada pendekatan ini relatif terbatas dibandingkan dua pendekatan lainnya. Distribusi topik penelitian dari 2019-2023 ditunjukkan pada Gambar 2.1 dan 2.2.

SEKOLAH PASCASARJANA



Gambar 2.1 Distribusi Topik Penelitian



Gambar 2.2 Distribusi Trend Topik Penelitian dari Tahun ke Tahun

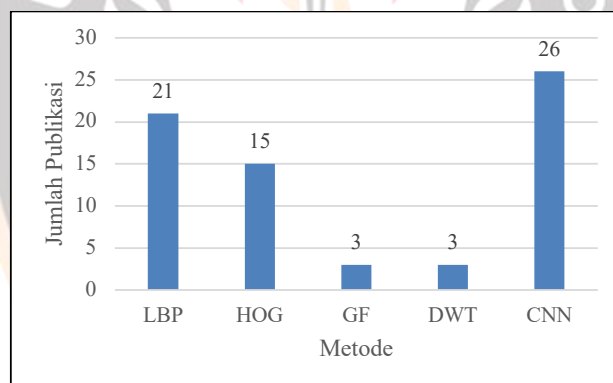
Gambar 2.1 menggambarkan distribusi topik penelitian dari tahun 2019 hingga 2023, menunjukkan bahwa mayoritas penelitian (59,1%) berfokus pada pendekatan ECG, diikuti oleh ECM (31,8%), dan hibrid (9,1%). Dominasi ECG menunjukkan preferensi peneliti untuk metode yang lebih interpretatif dan komputasi ringan, yang sering dibutuhkan oleh industri untuk aplikasi praktis. Meskipun demikian, seperti terlihat pada Gambar 2.2, tren penelitian ECM meningkat tajam antara tahun 2020 hingga 2023, mencerminkan meningkatnya ketertarikan pada model mendalam yang mampu menghasilkan representasi ciri otomatis dengan generalisasi yang lebih baik. Sebaliknya, penelitian terkait ECG mengalami penurunan sejak 2021, mengindikasikan berkurangnya preferensi

Pada Tabel 2.1, ditunjukkan berbagai metode ekstraksi ciri yang sering digunakan dari tahun 2019 hingga 2023. Dari 66 penelitian yang dianalisis, metode Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), Gabor Filter (GF), Discrete Wavelet Transform (DWT), dan Convolutional Neural Network (CNN) menonjol sebagai teknik utama. Gambar 2.3 dan Gambar 2.4 memperlihatkan proporsi adopsi metode tersebut. CNN mendominasi (38%), diikuti oleh LBP (31%) dan HOG (22%), menandakan pergeseran ke arah pendekatan mendalam untuk menangani ciri yang lebih kompleks, sedangkan LBP dan HOG tetap populer dalam pendekatan geometris.

[illegible][illegible]

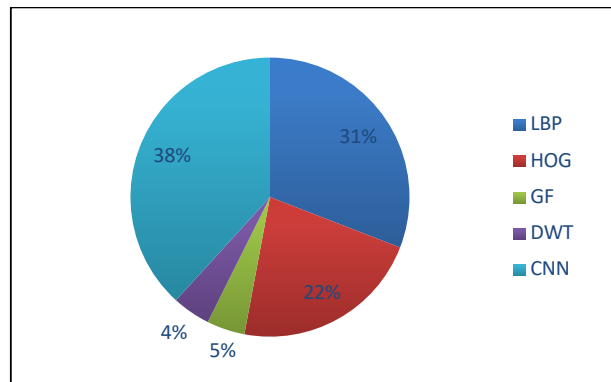
16	MB-LBPUH+HOG	-	√	√													
17	MO-HOG	CNN		√										√			
18	-	CNN												√			
19	HOG+2D-DWT	-		√				√									
20	LBP+LNEP	-	√														
21	SM-SWGT	-						√									
22	-	RNN													√		
23	HOG-LBP	-	√	√													
24	LBP	CNN	√											√			
25	LBP-LPQ-WLD-DCT	-	√														
26	LBP	CNN	√											√			
27	uGMM	-						√									
28	HDG+HDGG	-		√													
29	LDP+LDDRP	-			√												
30	LDN+DGLTP	-											√				
31	-	CNN												√			
32	-	CNN												√			
33	Filter Gabor+LBP+LGBP	-	√			√											
34	OPD-GQMBP	-		√													
35	PHOG+EHD+LBP	CNN	√	√										√			
36	TT+LGC-HD	-							√								
37	LBP	-	√														
38	-	CNN												√			
39	DWT+LL Band+LGC-HD	-						√		√							
40	TPLBP+FPLBP	-	√														
41	HOG+LBP	-	√	√													
42	-	CNN												√			
43	-	RNN													√		
44	SM-SBWT	-						√									
45	-	CNN												√			
46	LBP	CNN	√											√			
47	-	CNN												√			
48	EbITF+EVbWVG	-	√	√													
49	LBP+3D VHOG	-	√	√													
50	LSDP	-	√		√							√					
51	Angle + Magnitude	-	√														
52	-	CNN												√			
53	-	CNN												√			

54	-	GAN															√	
55	-	CNN															√	
56	-	RCNN																√
57	EHOG+FAC+WLBP	-	√	√														
58	PCA	-																
59	MFF-LDTP	-			√													
60	-	CNN															√	
61	-	CNN															√	
62	-	CNN															√	
63	-	CNN															√	
64	-	CNN (AGOA)															√	
65	-	CNN															√	
66	HOG	-		√														
TOTAL			21	15	3	3	1	3	1	2	1	2	1	26	2	1	1	



Gambar 2.3 Metode Ekstraksi Ciri Tekstur dengan Pendekatan Geometriks dan Mendalam yang Banyak Digunakan dalam Pengembangan Model Pengenalan Ekspresi Wajah

SEKOLAH PASCASARJANA



Gambar 2.4 Distribusi Studi Berdasarkan Jenis Metode Ekstraksi Ciri Tekstur

Dengan mempertimbangkan relevansi dan performanya, penelitian ini memilih LBP, HOG, GF, DWT, dan CNN sebagai metode utama dalam pengembangan model. Pemilihan ini didasarkan pada efisiensi dan kemampuannya dalam menangkap ciri yang penting untuk meningkatkan akurasi dan ketahanan sistem pengenalan ekspresi wajah. Hal ini juga didasari dari temuan (Fan dan Tjahjadi, 2019) bahwa gabungan pendekatan ECG dan ECM memungkinkan model klasifikasi memperoleh representasi ciri tekstur yang lebih komprehensif, meningkatkan kinerja, dan memberikan fleksibilitas dalam mengatasi berbagai tipe data dan tugas analisis data yang beragam. Penelitian lain, juga mengungkapkan hal yang serupa, bahwa pendekatan hibrid dapat melakukan pengurangan dimensi dari kompleksitas data yang beragam, memberikan informasi yang lebih bermakna untuk keperluan interpretasi dan pengambilan keputusan (Mame dan Tapamo, 2023).

2.1.2 Metode-Metode Ekstraksi Ciri Tekstur dan Klasifikasi yang Digunakan dalam Pengenalan Ekspresi Wajah

Sejak tahun 2019 ke 2023, ada 66 penelitian yang membahas metode ekstraksi ciri tekstur dan klasifikasi dilaporkan sebagai metode terbaik pada pengenalan ekspresi wajah. Hasil tinjauan metode-metode yang digunakan dalam pengenalan ekspresi wajah disajikan berdasarkan skema: latar belakang, tujuan, metode, dan

hasil. Ringkasan dari metode-metode mutakhir sebagai strategi metode yang diusulkan oleh para peneliti sejak 2019 ke 2023 dalam pengembangan model sistem pengenalan ekspresi wajah ditunjukkan pada Tabel 2.2 (Terlampir).

Tabel 2.2 Ringkasan Metode Mutakhir dalam Pengenalan Ekspresi Wajah

(Terlampir)

Seperti yang ditunjukkan pada Tabel 2.2 Sadeghi dan Raie (2019) memperkenalkan teknik pengenalan ekspresi wajah otomatis menggunakan deteksi landmark, ekstraksi ciri tekstur menggunakan Tapis Gabor (FG), dan pengklasifikasi Support Vector Machine (SVM). Metode ini dievaluasi pada basisdata CK+, SFEW, MMI, dan RAF-DB, dengan akurasi tertinggi sebesar 95,72% di CK+. Meskipun hasil ini menunjukkan keunggulan, penelitian lebih lanjut diperlukan untuk meningkatkan ketahanan terhadap variasi ekspresi wajah, terutama dalam situasi yang dinamis. Liu dkk. (2019) menggabungkan Gabor 2D dan Local Binary Pattern (LBP) dengan Extreme Machine Learning (ELM), menghasilkan akurasi 91,22% pada basisdata CK+. Metode ini efektif dalam menangkap tekstur multiskala, namun performanya terbatas pada basisdata kecil seperti JAFFE.

Hu dkk. (2019) mengusulkan metode pengenalan ekspresi wajah berbasis deskriptor Center-Symmetric Local Octonary Pattern (CS-LOP) yang dievaluasi pada CK+ dan JAFFE. Metode ini menunjukkan peningkatan performa pengenalan, namun memerlukan uji validitas eksternal pada basisdata lain. Lu dkk. (2019) menggunakan Discrete Separable Shearlet Transform (DSST) dan pemilihan ciri tekstur optimal dengan SVM. Meskipun metode ini mengurangi kompleksitas komputasi, ketergantungan pada kualitas data awal masih menjadi batasan.

Sun dkk. (2019) memperkenalkan subruang pembelajaran berbasis ekstraksi ciri tekstur dengan Sparse Representation-Based Classification (SRC), menunjukkan akurasi tinggi pada basisdata CK+, JAFFE, dan MMI. Alphonse dan Starvin (2019) memperkenalkan metode Gabor respons maksimum dan minimum

(MMRG) dengan teknik optimasi Bat Algorithm-Based Gabor Optimization (BAGO). Hasilnya menunjukkan akurasi tinggi pada beberapa basisdata, namun tantangan utama adalah kemampuan model untuk menangani derau dan perubahan posisi.

N. Sun dkk. (2019) mengembangkan Multi-Channel Deep Spatial-Temporal Feature Fusion Network (MDSTFN) yang menggabungkan CNN dan SVM. Eksperimen menunjukkan hasil terbaik dengan akurasi mencapai 99,59% pada basisdata MMI. Sementara itu, X. Sun dan Lv (2019) menggabungkan SIFT dan CNN, menghasilkan peningkatan akurasi hingga 94,82%. Temuan ini menunjukkan potensi besar dalam menggabungkan ciri tekstur geometris dan mendalam.

Penelitian oleh Jain dkk. (2019) memperkenalkan Square-Based Diagonal Pattern (SBDP) dengan Relevance Vector Machine (RVM) untuk ekstraksi ciri tekstur, menghasilkan akurasi yang lebih tinggi dibandingkan metode lain. Kong (2019) memperkenalkan metode CNN berbasis LBP yang mencapai akurasi 91,28%. Studi oleh Du dan Hu (2019) menggunakan Pembelajaran Pasangan Kamus Regularisasi Manifold (MRDPL-IVRF) dan menunjukkan keefektifan dalam pengenalan ekspresi wajah pada basisdata publik.

Sharifnejad dkk. (2020) memperkenalkan Enhanced Local Binary Pattern (ELBP) dan Patterns of Oriented Edge Magnitudes (PHOG), menunjukkan akurasi pengenalan tertinggi pada daerah wajah tertentu. H. Wang dkk. (2020) memperkenalkan gabungan Multi-Orientation HOG (MO-HOG) dan deep-learned features, menghasilkan akurasi pengenalan wajah hingga 97,70% pada CK+. Meskipun demikian, model ini masih perlu ditingkatkan dalam menghadapi kondisi pencahayaan bervariasi.

Tong dkk. (2020) memperkenalkan CNN berbasis arsitektur VGG16 yang dikombinasikan dengan LSTM untuk menangani hubungan temporal antara frame. Alenazy dan Alqahtani (2020) mengusulkan pendekatan semi-supervised Deep Belief Network (DBN) untuk memprediksi ekspresi wajah, namun tantangan generalisasi tetap ada. Shanthi dan Nickolas (2020) menggabungkan LBP dan Local

Neighborhood Encoded Pattern (LNEP), menunjukkan keunggulan dalam menghadapi lingkungan bising.

Kumar dkk. (2020) memperkenalkan Selective Multilevel Stationary Wavelet Gradient Transform (SM-SWGT) untuk ekstraksi ciri tekstur wajah yang menunjukkan hasil menjanjikan dalam pengenalan emosi manusia, namun memiliki keterbatasan dalam pengenalan variasi ekspresi wajah yang sangat kompleks dan membutuhkan basisdata yang representatif untuk hasil yang lebih general. An dan Liu (2020) mengusulkan hybrid CNN dan LSTM yang berhasil mengatasi masalah hilangnya gradien dalam model pembelajaran mendalam, namun LSTM mengalami overfitting ketika jumlah basisdata pelatihan terlalu kecil atau jika model terlalu kompleks. Yadav dan Singha (2020) menggabungkan HOG dan LBP, yang menghasilkan akurasi pengenalan tinggi pada basisdata CK+, akan tetapi Akan LBP mungkin sensitif terhadap perubahan skala citra yang dapat mempengaruhi mengenali pola tekstur.

Li dkk. (2020) memperkenalkan hybrid VGG16 dan LBP, yang menunjukkan performa unggul pada empat basisdata, termasuk CK+ dan JAFFE. Vedantham dan Reddy (2020) mengusulkan DBN-SMO, menunjukkan keunggulan dalam menangani variasi usia dan etnis pada basisdata pengenalan wajah, akan tetapi performa DBN bergantung pada pemilihan parameter yang tepat. Liu dkk. (2020) memperkenalkan CNN dua saluran dengan mekanisme perhatian untuk mendeteksi ekspresi wajah, menunjukkan akurasi 94,24% pada CK+. Namun, kompleksitas model dapat meningkatkan risiko overfitting, terutama pada basisdata kecil, karena model kompleks mungkin tidak dapat menggeneralisasi pada data baru dengan baik.

Penelitian oleh Wang dkk. (2020) memperkenalkan penggabungan ekstraksi ciri tekstur MB-LBPUH dan HOG untuk meningkatkan ketahanan model pengenalan ekspresi wajah, menghasilkan hasil signifikan pada beberapa basisdata publik, namun pengujian dan evaluasi terbatas pada basisdata JAFFE, CK+, dan BU-3DFE sehingga generalisasi ke basisdata lain perlu dicoba. Tripathi dan Jalal (2021) menggabungkan dua deskriptor lokal baru yang mampu mengatasi variasi

usia pada basisdata FGNET. Saeed dkk. (2021) memperkenalkan teknik normalisasi iluminasi, menunjukkan peningkatan hasil pada basisdata YMU dan VMU. Walaupun demikian, penelitian belum menguji pada basisdata dengan variasi ekspresi, pose, dan pencahayaan.

Kas dkk. (2021) memperkenalkan model OPD-GQMBP untuk pengenalan ekspresi wajah dalam platform e-learning, keberhasilan metode ini sangat bergantung pada ukuran dan kualitas basisdata yang digunakan untuk pelatihan dan pengujian di mana ketika berukuran kecil, model cenderung mengalami *overfitting*, yang mengurangi kemampuan generalisasi model. Liu, dkk. (2021) memperkenalkan SACNN dan ALSTM untuk pengenalan ekspresi wajah, yang terbukti efektif pada basisdata CK+ dan JAFFE, meskipun metode ini memberikan hasil yang mengesankan, namun masih bergantung pada pengaturan parameter yang tepat. Chen dkk. (2021) memperkenalkan kernel entropy component analysis network (KECANet) dan DAGSVM yang menghasilkan performa pengenalan lebih baik, akan tetapi cenderung mengalami *overfitting* ketika model mempelajari data latih terlalu mendalam sehingga kinerjanya menurun pada data baru.

Sun et al. (2021) mengembangkan FERFC, sebuah metode penggabungan CNN dan LBP, yang diuji pada basisdata CK+ dan FER2013 dengan hasil akurasi 95,6%. Liang et al. (2021) memperkenalkan Patch Attention Layer (PAL) yang dievaluasi pada basisdata CK+, menunjukkan akurasi tinggi namun rentan terhadap *overfitting*. Krithika dan Priya (2021) memperkenalkan metode GFE-HCA untuk meningkatkan pengenalan ekspresi wajah, menunjukkan akurasi 96,47% pada basisdata MMI.

Sun dkk. (2021) mengembangkan FERFC, sebuah metode penggabungan CNN dan LBP, yang diuji pada basisdata CK+ dan FER2013 dengan hasil akurasi 95,6%, tetapi efektivitasnya berkurang ketika citra wajah berkualitas rendah. Liang dkk. (2021) memperkenalkan Patch Attention Layer (PAL) yang dievaluasi pada basisdata CK+, menunjukkan akurasi tinggi namun rentan terhadap *overfitting*. Krithika dan Priya (2021) memperkenalkan metode GFE-HCA untuk

meningkatkan pengenalan ekspresi wajah, menunjukkan akurasi 96,47% pada basisdata MMI.

Bhatti dkk. (2021) mengembangkan pendekatan feedforward dengan CNN dan ELM untuk pengenalan ekspresi wajah pengajar, menunjukkan peningkatan akurasi pada basisdata CK+, JAFFE, dan FER2013, namun memerlukan sumber daya komputasi tinggi terutama pada video beresolusi tinggi. Kola dan Samayamantula (2021) memperkenalkan LGC-HD dan PCA untuk pengurangan dimensi dan SVM untuk klasifikasi, menunjukkan hasil signifikan pada beberapa basisdata publik, namun performa menurun ketika input basisdata mengandung variasi signifikan dalam kondisi pencahayaan atau posisi wajah. Ullah dkk. (2021) memperkenalkan model hibrida yang menggabungkan ciri tekstur geometris dan deep learning, menunjukkan performa unggul pada basisdata CK+, MMI, dan RaFDB.

Alphonse dan Starvin (2019) memperkenalkan metode Maximum and Minimum Response-Based Gabor (MMRG) untuk mengoptimalkan tapis Gabor dalam ekstraksi ciri tekstur wajah yang lebih diskriminatif. Mereka menggunakan Radial Basis Function-Extreme Learning Machine (RBF-ELM) sebagai pengklasifikasi, yang terbukti efektif pada enam basisdata, termasuk CK+ dan JAFFE, dengan akurasi mencapai 97,4% pada CK+. Namun, metode ini masih rentan terhadap derau acak yang dapat memengaruhi kinerja model.

Penelitian oleh Rajendra Kurup dkk. (2019) memperkenalkan pendekatan semi-pengawasan untuk pengenalan ekspresi wajah dengan menggunakan ciri-ciri tekstur wajah dari data yang sebagian besar tidak berlabel. Metode ini mengombinasikan *contrastive divergence* (CD) dan *backpropagation* (BP), menghasilkan akurasi pengujian yang lebih tinggi dari SVM dan CNN, mencapai 98,57%. Ciri HOG pada bagian mulut terbukti paling relevan dalam klasifikasi, meskipun efek derau belum diperhitungkan secara menyeluruh.

Fan dan Tjahjadi (2019) mengusulkan kerangka kerja yang menggabungkan ciri-ciri tekstur diskriminatif dari jaringan saraf konvolusional untuk meningkatkan

akurasi pengenalan ekspresi wajah. Pengujian pada basisdata CK+ menunjukkan peningkatan akurasi hingga 92,5%. Namun, kerangka ini belum diuji pada beragam kondisi dunia nyata yang lebih dinamis.

Perveen dkk. (2020) mengusulkan *Universal Gaussian Mixture Model* (uGMM) dan *Mean Internal Kernel* (MIK) untuk representasi ekspresi wajah berbasis kernel dinamis. Pendekatan ini efektif dalam menangkap dinamika lokal dan variasi konteks global, namun masih kurang optimal dalam menangani variasi ekspresi wajah yang lebih halus.

Tadalagi dan Joshi (2021) menggunakan *Local Binary Pattern* (LBP) yang tahan terhadap pencahayaan untuk mendeteksi depresi melalui ekspresi wajah. Algoritma ini diterapkan pada video subjek, dengan akurasi yang cukup baik pada basisdata CK+ dan AVEC. Meski begitu, penerapannya pada basisdata yang lebih kompleks dan bervariasi masih menghadapi tantangan dalam generalisasi kinerja model.

Penelitian Wahab dkk. (2021) mengusulkan model hybrid CNN-KNN untuk pengenalan ekspresi wajah di Raspberry Pi-4. CNN digunakan untuk ekstraksi ciri, sementara KNN melakukan klasifikasi. Menggunakan arsitektur EfficientNet-Lite dengan pembelajaran transfer, mereka mengganti lapisan Softmax dengan KNN. Model ini diuji pada basisdata FER2013 dengan akurasi 75,26%, meskipun keterbatasan muncul pada penentuan nilai k yang optimal dalam KNN.

Kashif dkk. (2021) mengembangkan metode ekstraksi ciri tekstur berbasis multiple LBP untuk menganalisis perilaku manusia melalui pengenalan ekspresi wajah. Menggunakan TP-LBP dan FPLBP yang dikombinasikan dengan DCT, vektor ciri diklasifikasikan dengan SVM, k -NN, dan Naive Bayes. Metode ini diuji pada basisdata CK+ dan Oulu-CASIA, menunjukkan ketahanan terhadap variasi sudut dan derau, meski ada sensitivitas terhadap jenis derau tertentu.

Zhou dkk. (2021) memperkenalkan metode Frequency Multiplication Network (FMN) yang bekerja di domain frekuensi untuk pengenalan ekspresi

wajah. Metode ini menggunakan HOG, LBP, dan uniform rectangular features (URF) untuk ekstraksi ciri tekstur yang diklasifikasikan oleh FMN. Uji coba pada basisdata CK+, Oulu-CASIA, dan MMI menunjukkan peningkatan kinerja, meskipun metode ini menghadapi kendala dalam ekstraksi daerah kunci wajah yang penting untuk pengenalan ekspresi optimal.

C. Liu dkk. (2021) memperkenalkan jaringan hibrida yang menggabungkan SACNN dan ALSTM untuk pengenalan ekspresi wajah. SACNN mengekstraksi ciri tekstur dari wajah statis, sementara ALSTM mengeksplorasi potensi landmark wajah. Pengujian pada FER2013, CK+, dan JAFFE menunjukkan akurasi signifikan, meskipun model ini belum mempertimbangkan variasi pencahayaan dan pose yang lebih kompleks.

Jeen Retna Kumar dkk. (2021) menggabungkan ciri dari Subband Wavelet Transform (SBWT) dengan Discrete Cosine Transform (DCT) dan reduksi dimensi melalui PCA untuk meningkatkan pengenalan ekspresi wajah. Pengujian menggunakan lima basisdata menunjukkan peningkatan performa, namun model masih memiliki keterbatasan dalam mengatasi variasi pose dan pencahayaan yang tidak seragam.

Ayeche dan Alti (2021) memperkenalkan metode HDG dan HDGG untuk mengekstraksi ciri tekstur ekspresi wajah yang lebih diskriminatif menggunakan SVM. Pengujian pada basisdata JAFFE dan YALE menunjukkan akurasi 92,12%, dengan waktu komputasi cepat. Namun, pengaturan parameter SVM masih memerlukan optimasi yang hati-hati untuk hasil optimal.

Yang dkk. (2021) memperkenalkan CFPN-NBP, sebuah jaringan piramida ciri bertingkat dengan propagasi non-backward, untuk pengenalan ekspresi wajah. Meskipun meningkatkan akurasi melalui HSIC bottleneck dan lapisan konvolusi, metode ini menghadapi kompleksitas arsitektur dan beban komputasional tinggi, serta terbatas dalam evaluasi pada basisdata FER2013.

Revina dan Emmanuel (2021) mengembangkan metode Enhanced Modified Decision Based Unsymmetric Trimmed Median Filter (EMDBUTMF) untuk menghilangkan derau pada citra wajah serta menggunakan pola Local Directional Number (LDN) dan Dominant Gradient Local Ternary Pattern (DGLTP) untuk ekstraksi ciri tekstur. Meskipun metode ini menghasilkan akurasi 88% pada pengenalan ekspresi wajah, pengujian terbatas pada basisdata JAFFE dan CK+ membatasi generalisasinya dalam kondisi pencahayaan, derau, dan pose wajah yang berbeda.

Bhatti dkk. (2021) memperkenalkan model pengenalan ekspresi wajah pengajar dengan pendekatan feedforward yang mengekstraksi fitur mendalam tanpa pembaruan bobot berulang, sehingga mengurangi kompleksitas komputasi. Meskipun eksperimen menunjukkan peningkatan akurasi, F1-score, dan recall pada basisdata IEV, CK+, JAFFE, dan FER2013, keterbatasan generalisasi ke berbagai lingkungan pengajaran masih menjadi tantangan.

Adjailia dkk. (2022) menggunakan Local Binary Pattern (LBP) dan 3D Voxel Histogram of Oriented Gradients (3DVHOG) untuk ekstraksi ciri tekstur, menghasilkan akurasi 88,5% dan 92,9% pada basisdata MDPa-FACE3D dan Bosphorus. Namun, penggunaan dua basisdata publik membuat generalisasi metode ini terbatas, memerlukan pengujian lebih lanjut pada basisdata yang lebih beragam.

Najmabadi dan Moallem (2022) memperkenalkan Local Symmetric Directional Pattern (LSDP) untuk pengenalan wajah, yang menggunakan gradien kompas simetris dan Sparse Representation-based Classification (SRC) untuk klasifikasi. Dengan akurasi mencapai 97,10% pada Yale-B dan 96,71% pada AR Face, metode ini terbukti efisien namun masih rentan terhadap variasi kondisi pencahayaan dan pose wajah.

Ukwu dan Yurtkan (2022) mengintegrasikan ciri sudut dan ciri profesional dengan teknik seleksi ciri relief dan SVM untuk pengenalan ekspresi wajah. Meskipun menunjukkan performa yang baik pada basisdata BU-4DFE,

generalisasinya terhadap basisdata yang lebih luas dan variasi ekstrem pencahayaan masih terbatas. K. Babu dkk. (2022) menggunakan analisis komponen utama, swarm intelligence, dan Deep Belief Network (DBN) dalam model pengenalan ekspresi wajah, menghasilkan akurasi tinggi pada basisdata RaFDB, AffecteNet, dan CK+. Namun, model ini menghadapi tantangan dalam menangani variasi ekstrem pencahayaan dan pose wajah.

Kim dkk. (2022) menggabungkan CNN dan SVM untuk pengenalan ekspresi wajah dinamis, menghasilkan akurasi tinggi pada basisdata CK+ dan BU4D. Namun, metode ini terbatas pada basisdata barat, sehingga generalisasinya terhadap populasi non-barat belum terbukti. Babu dkk. (2022) mengusulkan metode SIM-LBP dan MCC-CNN untuk ekstraksi ciri tekstur wajah, yang efektif dalam berbagai kondisi pencahayaan. Namun, ketika diterapkan pada basisdata yang lebih beragam, metode ini menunjukkan keterbatasan dalam generalisasi. Dubey dan Jain (2022) menggunakan Extended Wavelet Deep CNN (EWDCNN) dengan prapengolahan citra yang kompleks untuk pengenalan ekspresi wajah. Meskipun akurat pada beberapa basisdata, metode ini memiliki beban komputasi tinggi dan keterbatasan dalam generalisasi untuk berbagai jenis citra.

Ma dan Ma (2022) memperkenalkan Relation and Context Augmentation Network (RCANet) untuk meningkatkan pengenalan ekspresi wajah dengan modul relasi dan konteks. Meskipun akurasinya tinggi pada basisdata RAF-DB dan AffectNet, modul konteks masih belum sepenuhnya menangkap informasi multi-skala, sehingga performanya belum optimal dalam situasi kompleks. Yan dkk. (2023) memperkenalkan metode MFF-LDTP yang menggabungkan ciri global dan lokal untuk meningkatkan ketangguhan pengenalan ekspresi wajah dalam lingkungan liar. Meskipun akurat pada basisdata JAFFE, kompleksitas metode ini membuatnya sulit diimplementasikan secara praktis.

Muthamilselvan dkk. (2023) memperkenalkan metode pembelajaran pengenalan ekspresi wajah OFELBW menggunakan optimasi paus biner. Metode ini melalui tiga tahap, yaitu pemrosesan citra dengan CNN dan penghilangan derau,

ekstraksi ciri tekstur wajah, dan seleksi ciri menggunakan optimasi paus biner. Hasil evaluasi pada basisdata SFEW, CK+, JAFFE, dan FERG menunjukkan peningkatan akurasi yang signifikan, meskipun masih terdapat keterbatasan terkait kualitas data dan sensitivitas terhadap variasi basisdata.

Khan dkk. (2023) mengusulkan pendekatan hibrid yang menggabungkan SIFT dan CNN dengan pembelajaran transfer menggunakan AlexNet untuk pengenalan ekspresi wajah. Meskipun akurasi tinggi dicapai pada basisdata CK+ dan FERG, metode ini mengalami penurunan kinerja pada basisdata realistik seperti FER2013, khususnya dengan masalah oklusi, skala, dan variasi pencahayaan.

Sumithra dan Rajkumar (2023) memperkenalkan model pembelajaran mendalam efisien untuk pengenalan ekspresi wajah menggunakan ADFD dan WDCM yang dioptimalkan dengan AVOA. Model ini menunjukkan peningkatan akurasi yang signifikan hingga 99% pada basisdata JAFFE dan CK+, namun memerlukan validasi lebih lanjut dengan basisdata yang lebih beragam untuk memastikan generalisasinya.

Subudhiray dkk. (2023) menggunakan pembelajaran transfer untuk ekstraksi ciri tekstur wajah melalui metode SIFT dan CNN dengan arsitektur VGG16. Meskipun strategi ini menunjukkan performa unggul dibandingkan metode lain pada basisdata JAFFE, CK+, KDEF, dan FER2013, kompleksitas komputasional meningkat karena dimensi tinggi dari matriks ciri dan proses analisis korelasi yang intensif.

Ch dan P (2023) mengembangkan algoritma AGOA yang dikombinasikan dengan LSTM untuk meningkatkan pengenalan ekspresi wajah. Evaluasi menggunakan basisdata YALE menunjukkan akurasi pengenalan sebesar 93,90%, dengan AGOA-LSTM terbukti lebih unggul dibandingkan model tanpa AGOA, baik dalam hal akurasi maupun tingkat kesalahan.

Prasad dan Chandana (2023) mengusulkan model pengenalan wajah termal menggunakan YOLOv4 dan DenseNet. Meskipun model ini mencapai akurasi

95,97% pada basisdata RGB-D-T, teknik pra-pemrosesan yang kompleks menambah beban komputasi dan membutuhkan sumber daya yang tinggi.

Mame dan Tapamo (2023) menggunakan teknik ekstraksi ciri tekstur lokal yang dimodifikasi dengan SVM untuk meningkatkan pengenalan ekspresi wajah. Meskipun teknik ini meningkatkan akurasi secara signifikan pada basisdata CK+, tidak ada analisis lebih lanjut mengenai potensi bias akibat penggunaan basisdata yang sama untuk pelatihan dan pengujian.

Albraikan dkk. (2022) memperkenalkan model IFER-DTFL yang menggunakan teknik mask RCNN dan Densely Connected Networks untuk ekstraksi ciri tekstur serta mesin pembelajaran ekstrim untuk klasifikasi ekspresi. Hasil evaluasi menunjukkan performa yang sangat baik, namun kompleksitas model memerlukan daya komputasi tinggi.

Ilham dkk. (2023) menggabungkan DWT dan PCA dengan JST yang dimodifikasi menggunakan Lyapunov Stability Theory untuk mempercepat pelatihan dan meningkatkan akurasi pengenalan wajah. Model ini menunjukkan peningkatan akurasi dan kecepatan pelatihan pada basisdata JAFFE, namun membutuhkan validasi lebih lanjut dengan basisdata yang lebih beragam untuk memastikan keandalannya.

2.1.3 Model-Model Klasifikasi Paling Banyak Digunakan

Seperti yang ditunjukkan pada Tabel 2.3, terdapat empat belas algoritma pembelajaran klasifikasi yang dapat dikelompokkan ke dalam kategori algoritma pembelajaran mesin (PM) dan pembelajaran mendalam (PMD) telah diimplementasikan oleh para peneliti. Dari keempat belas algoritma tersebut, khususnya pada kategori pembelajaran mesin ada tiga algoritma yang paling banyak diterapkan di bidang pengenalan ekspresi wajah diidentifikasi sebagai berikut.

1. Support vector machine (SVM)

2. Ekstrim learning machine (ELM)
3. Jaringan saraf tiruan (JST)

Tabel 2.3 Distribui Model Pembelajaran untuk Pengenalan ekspresi wajah

KT	ALGORITMA	KATEGORI	SUB KATEGORI	FREKUENSI
PEMBELAJARAN MESIN	SVM	PM	Support Vector Machine	25
	ELM	PM	Jaringan Saraf Tiruan	4
	JST	PM	Jaringan Saraf Tiruan	3
	KNN	PM	Tetangga Terdekat	2
	FSVM	PM	Support Vector Machine	1
	RVM	PM	Support Vector Machine	1
	SVM+KNN+NB	PM	-	1
PEMBELAJARAN MENDALAM	CNN-Softmax	PMD	JSK	16
	LSTM	PMD	Jaringan Saraf Tiruan Berulang	4
	DBN	PMD	JSK	4
	SRC	PMD	-	2
	EWDCNN	PMD	JSK	1
	FMN	PMD	JSK	1
-	IOCS	-	-	1
TOTAL				66

Catatan: (PM) Pembelajaran Mesin; (PMD) Pembelajaran Mendalam; (KT) Kategori

Namun, prosedur penelitian yang lebih andal perlu dikembangkan, sebelum kesimpulan yang meyakinkan dari studi perbandingan model pengenalan ekspresi wajah dapat dibuat. Kerangka kerja perbandingan baru untuk pengenalan ekspresi wajah diusulkan dalam penelitian ini untuk memenuhi kebutuhan akan metode yang lebih sistematis dan tidak bias untuk membandingkan kinerja pengenalan ekspresi wajah berbasis pembelajaran mesin. Penelitian ini mengadopsi dan mengimplementasikan dua algoritma yang paling banyak diterapkan termasuk SVM dan JST ditambah dengan KNN, NB, RF, dan algoritma CART sebagai model klasifikasi ekspresi wajah.

Perlu diketahui dalam penelitian ini kami mengecualikan pengklasifikasi dengan pendekatan pembelajaran mendalam (PMD) karena didasari dari beberapa temuan, seperti (Fan dan Tjahjadi, 2019; Revina dan Emmanuel, 2021; Tsalera dkk., 2022) yang mengungkapkan bahwa memisahkan jaringan saraf konvolusional (JSK) dan SVM memungkinkan optimasi yang lebih baik pada masing-masing tahap. JSK dapat dioptimalkan untuk tugas ekstraksi ciri tekstur yang spesifik, sedangkan SVM dapat diatur dengan parameter yang sesuai untuk tugas klasifikasi. Menurut (Tsalera et al., 2022), hal ini juga memungkinkan untuk mengganti atau memperbarui salah satu model tanpa harus mempengaruhi atau mengganggu model lainnya, memberikan modularitas pada sistem.

2.2 Keaslian Penelitian

Penelitian ini mengusulkan pendekatan hibrid untuk pengenalan ekspresi wajah dengan menggabungkan histogram of oriented gradients (HOG) dan jaringan saraf konvolusional (CNN) berbasis VGG16 sebagai metode utama untuk ekstraksi ciri tekstur citra. Strategi ini diawali dengan tahap prapengolahan yang memanfaatkan tapis median dan kaskade Haar, di mana tapis median digunakan untuk mengurangi derau impulsif tanpa menghilangkan detail penting pada citra, sementara kaskade Haar berfungsi mempercepat deteksi wajah melalui identifikasi perbedaan intensitas antara area terang dan gelap.

Inovasi utama dari penelitian ini adalah kombinasi metode HOG dan VGG16 untuk mengekstraksi fitur tekstur secara lebih mendalam dan kompleks, menggabungkan keunggulan HOG dalam menangkap pola geometris dengan kekuatan VGG16 untuk mengekstraksi ciri tekstur yang lebih abstrak dan informatif. Untuk memastikan akurasi dan ketahanan model dalam kondisi nyata, sistem ini diuji pada empat basisdata publik (CK+, JAFFE, FER2013, dan RaFDB) dan juga pada data nyata dari Ahmad Ilham Simple Face Database (AIsFDB), yang mencakup delapan sudut pandang untuk menguji ketahanan model terhadap variasi pose.

Evaluasi performa mencakup akurasi, precision, recall, F1-score, standar deviasi, dan waktu komputasi, dengan analisis perbandingan antara klasifikasi menggunakan Support Vector Machine (SVM) dan JST. Hasil pengujian menunjukkan bahwa pendekatan hibrid ini dapat meningkatkan akurasi dan ketahanan model terhadap perubahan kondisi lingkungan seperti pencahayaan, gangguan derau, dan pose wajah. Dengan kombinasi ciri geometris dan mendalam, penelitian ini memberikan kontribusi signifikan pada pengembangan teknologi pengenalan ekspresi wajah yang lebih adaptif untuk aplikasi dunia nyata, termasuk bidang keamanan dan interaksi manusia-komputer.

2.3 Landasan Teori

2.3.1 Tantangan Pengenalan ekspresi wajah

Mendefinisikan ekspresi sebagai representasi emosi tertentu bisa jadi sulit bahkan bagi manusia. Penelitian telah menunjukkan bahwa orang yang berbeda mengenali jenis emosi yang berbeda dalam ekspresi wajah yang sama. Sistem pengenalan ekspresi wajah melibatkan banyak tantangan, seperti kenyataan bahwa diperlukan data latih yang beragam, serta citra dengan latar belakang yang beragam, citra yang bersih dari derau dll.

1. Kelangkaan Kumpulan data Sistem Klasifikasi Ekspresi Wajah

Kumpulan data yang tersedia untuk umum tidak cukup untuk SPEW yang efektif, dan juga tidak cukup beragam. Permasalahan tersebut memerlukan solusi yang efektif, seperti augmentasi data, kombinasi beberapa basisdata, modifikasi data yang sudah ada, atau pembuatan basisdata baru (Ullah dkk., 2022). Biasanya, model pembelajaran yang kompleks sangat “haus data”, dan memerlukan data dalam berbagai bentuk untuk pelatihan yang lebih efektif dan lebih mudah. Solusi ini menghindari masalah overfitting dalam pelatihan model. Oleh karena itu, PEW memerlukan metode yang dapat menangkap semua ekspresi dari semua sudut yang memungkinkan untuk mendapatkan hasil yang efektif.

a. Pencahayaan

Penerangan (*illumination*) mengacu pada variasi cahaya dari sudut yang berbeda atau tunggal. Sedikit perubahan dalam kondisi cahaya merupakan tantangan yang signifikan untuk pengenalan emosi dan secara signifikan memengaruhi hasilnya. Perubahan pencahayaan dapat mengubah penampilan wajah secara drastis. Oleh karena itu, perbedaan antara dua wajah yang ditangkap di bawah penerangan yang berbeda, akan lebih tinggi daripada dua wajah yang berbeda, yang ditangkap di bawah penerangan yang sama. Masalah ini membuat SPEW menjadi sangat menantang dan telah menarik perhatian selama beberapa dekade terakhir. Sejumlah algoritme telah diusulkan untuk menangani penyinaran, dan secara garis besar, algoritme ini melibatkan tiga perbedaan. Pendekatan pertama berkaitan dengan metode pemrosesan gambar yang berguna untuk normalisasi wajah dengan efek pencahayaan yang berbeda. Untuk tujuan ini, penyetaraan histogram (PH) (Pizer dkk., 1987; Shiguang Shan dkk., 2003), transformasi logaritma (Savvides dan Vijaya Kumar, 2003), atau korelasi intensitas gamma (Shiguang Shan dkk., 2003) telah dipertimbangkan. Pendekatan lain adalah pemodelan wajah 3D. Para peneliti dalam (Georghiades dkk., 2001, 2002) menyarankan bahwa wajah yang dilihat dari depan dengan pencahayaan yang berbeda akan menciptakan kerucut yang dikenal sebagai kerucut pencahayaan. Demikian pula, pada pendekatan ketiga, ciri wajah diekstraksi di tempat yang disinari, dan ciri-ciri tersebut kemudian diteruskan untuk dikenali.

b. Pose Wajah

Pose wajah adalah tantangan utama lainnya; sistem PEW sangat sensitif terhadap perubahan kecil dalam pose. Pose wajah bervariasi dengan gerakan kepala dan perubahan sudut pandang. Gerakan kepala atau variasi dalam sudut pandang kamera dapat menyebabkan perubahan pada tampilan wajah, sehingga menciptakan variasi intra-kelas dan sangat menurunkan kinerja metode PEW (J. Liu, Feng, dkk., 2021). Namun, meskipun tingkat pengenalan model jaringan saraf konvolusional (JSK) sangat kuat untuk mengekstrak ciri, tingkat pengenalannya menurun secara

signifikan dengan diperkenalkannya pose wajah (W. Wu dkk., 2018). Wajah manusia secara kasar berbentuk seperti bola cembung, dan pose menyebabkan fenomena oklusi diri dan mengurangi akurasi PEW. Oleh karena itu, melakukan PEW secara andal untuk berbagai pose kepala tetap menjadi tantangan yang signifikan.

2. Derau pada citra untuk PEW

Tantangan derau pada data citra untuk PEW merupakan salah satu isu utama dalam bidang pengenalan pola dan pengolahan citra. Derau dalam data citra dapat disebabkan oleh berbagai faktor, seperti variasi pencahayaan, perubahan latar belakang, resolusi gambar yang rendah, dan gerakan wajah yang tidak terduga. Hal ini dapat mengakibatkan ketidakakuratan dalam ekstraksi ciri tekstur dan pengenalan pola, yang pada akhirnya mempengaruhi performa model pengenalan ekspresi wajah.

Salah satu pendekatan untuk mengatasi tantangan ini adalah dengan menggunakan teknik prapengolahan citra untuk meningkatkan kualitas data. Teknik-teknik seperti normalisasi histogram, peningkatan kontras, dan tapis median yang adaptif dapat membantu mengurangi derau dan meningkatkan kejelasan ciri wajah. Selain itu, penggunaan model pembelajaran mendalam seperti jaringan saraf konvolusi (JSK) telah terbukti efektif dalam menangani derau dengan secara otomatis belajar representasi ciri yang lebih robust dari data mentah. Teknik augmentasi data juga dapat diterapkan untuk memperkaya set data pelatihan dengan variasi yang lebih besar, sehingga model dapat lebih tahan terhadap derau.

Penelitian terbaru menunjukkan berbagai metode untuk mengatasi tantangan ini. Misalnya, (Dubey dan Jain, 2022) memperkenalkan metode hybrid yang menggabungkan JSK dengan prapengolahan berbasis wavelet untuk meningkatkan ketahanan terhadap derau pada pengenalan ekspresi wajah. Begitu pula, (Shin dkk., 2024) mengembangkan teknik augmentasi data yang lebih canggih untuk meningkatkan performa model pada basisdata dengan derau tinggi.

Dengan pendekatan-pendekatan ini, tantangan derau pada data citra untuk pengenalan ekspresi wajah dapat diatasi lebih efektif, meskipun penelitian lebih lanjut masih diperlukan untuk mencapai hasil yang optimal dalam berbagai kondisi nyata.

2.3.2 Model Pengenalan ekspresi wajah dan Penerapannya di Berbagai Bidang

Model pengenalan ekspresi wajah telah menjadi area penelitian yang sangat aktif dalam kecerdasan buatan dan pengolahan citra, dengan aplikasi yang luas di berbagai bidang. Manfaat dari sistem ini sangat signifikan, karena mampu menangkap dan menginterpretasikan emosi manusia melalui analisis ekspresi wajah, yang merupakan salah satu bentuk komunikasi non-verbal yang paling penting. Salah satu manfaat utama dari sistem pengenalan ekspresi wajah adalah peningkatan interaksi antara manusia dan mesin. Dengan kemampuan untuk mengenali dan merespons ekspresi wajah, mesin dan perangkat lunak dapat memberikan respons yang lebih intuitif dan alami. Misalnya, asisten virtual yang dapat mendeteksi kekecewaan pengguna dapat menawarkan bantuan yang lebih relevan dan cepat. Dalam bidang keamanan dan pengawasan, sistem pengenalan ekspresi wajah dapat digunakan untuk mendeteksi perilaku mencurigakan atau emosi negatif seperti kemarahan atau stres, yang bisa menjadi indikator potensi ancaman. Penggunaan teknologi ini dapat meningkatkan respons keamanan di tempat-tempat publik, seperti bandara atau stadion.

Di bidang kesehatan, terutama dalam terapi dan diagnosa, sistem pengenalan ekspresi wajah dapat membantu dalam memantau kondisi emosional pasien, seperti penderita depresi atau gangguan kecemasan. Teknologi ini juga dapat digunakan dalam terapi anak-anak dengan autisme untuk membantu mereka mengenali dan memahami ekspresi wajah orang lain. Dalam bidang pemasaran, memahami reaksi emosional konsumen terhadap produk atau iklan sangat penting. Sistem pengenalan ekspresi wajah memungkinkan perusahaan untuk melakukan riset pasar yang lebih

mendalam dengan mengamati reaksi spontan konsumen terhadap berbagai stimulus, sehingga dapat mengoptimalkan strategi pemasaran. Dalam bidang pendidikan, teknologi ini dapat digunakan untuk mengukur keterlibatan dan respons emosional siswa terhadap materi pelajaran atau metode pengajaran tertentu. Hal ini memungkinkan pendidik untuk menyesuaikan strategi pengajaran mereka guna meningkatkan efektivitas pembelajaran.

Pentingnya sistem pengenalan ekspresi wajah terletak pada kemampuannya untuk mendeteksi dan menginterpretasikan emosi manusia secara akurat dan cepat. Emosi merupakan bagian integral dari komunikasi manusia, dan kemampuan untuk mengenali emosi dapat meningkatkan interaksi dalam berbagai konteks. Sistem pengenalan ekspresi wajah memungkinkan komunikasi yang lebih efektif antara manusia dan teknologi, membuat interaksi lebih alami dan humanis. Dalam konteks keamanan dan kesehatan mental, kemampuan untuk mendeteksi emosi negatif dapat membantu dalam pencegahan dan intervensi dini, mengurangi risiko kejadian berbahaya atau memburuknya kondisi kesehatan. Dengan mengetahui emosi pengguna, layanan dapat disesuaikan secara lebih personal dan relevan, meningkatkan kepuasan dan pengalaman pengguna.

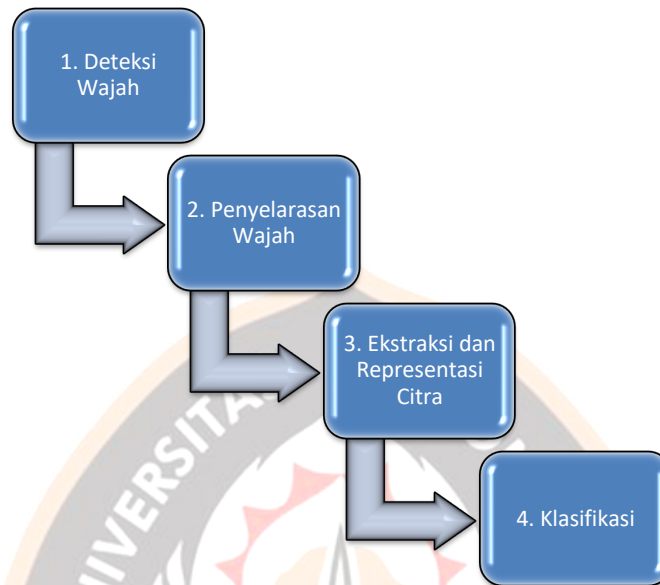
Aplikasi sistem pengenalan ekspresi wajah tersebar di berbagai bidang. Dalam bidang medis, khususnya psikiatri dan psikologi, sistem ini digunakan untuk pemantauan kondisi emosional pasien. Misalnya, pasien dengan gangguan stres pasca-trauma (PTSD) dapat dipantau untuk tanda-tanda stres atau kecemasan yang tidak terdeteksi oleh metode konvensional. Dalam bidang pendidikan, teknologi ini dapat digunakan untuk menilai keterlibatan siswa dan efektivitas metode pengajaran. Misalnya, reaksi wajah siswa dapat digunakan untuk menentukan apakah mereka memahami materi pelajaran atau apakah mereka merasa bosan, sehingga pendidik dapat menyesuaikan metode pengajaran secara real-time. Perusahaan menggunakan sistem pengenalan ekspresi wajah untuk mengukur respons emosional konsumen terhadap produk atau kampanye iklan mereka. Ini membantu dalam mengembangkan strategi pemasaran yang lebih efektif dan menargetkan konsumen dengan lebih tepat. Dalam keamanan publik, teknologi ini

digunakan untuk mendeteksi perilaku mencurigakan di tempat-tempat umum seperti bandara, stasiun kereta api, dan pusat perbelanjaan. Sistem ini dapat membantu dalam mendeteksi potensi ancaman dan merespons dengan cepat untuk mencegah kejadian yang tidak diinginkan. Dalam industri hiburan, terutama dalam permainan video dan realitas virtual, sistem pengenalan ekspresi wajah dapat meningkatkan pengalaman pengguna dengan membuat karakter virtual yang dapat merespons emosi pemain secara real-time, menciptakan pengalaman yang lebih imersif dan interaktif.

Secara keseluruhan, sistem pengenalan ekspresi wajah menawarkan banyak manfaat dan aplikasi di berbagai bidang, yang tidak hanya meningkatkan efektivitas dan efisiensi proses tetapi juga membuka peluang baru untuk inovasi dalam interaksi manusia-mesin. Dengan kemajuan teknologi yang terus berlanjut, potensi aplikasi sistem ini di masa depan akan semakin berkembang, memberikan kontribusi yang lebih besar bagi berbagai aspek kehidupan manusia.

2.3.3 Gambaran Umum Sistem Pengenalan ekspresi wajah

Sistem pengenalan ekspresi wajah terdiri dari beberapa tahapan utama yang bertujuan untuk mendeteksi dan mengenali ekspresi wajah secara akurat. Skema dasar ini dapat dibagi menjadi empat tahapan utama: 1) deteksi wajah, 2) penyelarasan wajah, 3) ekstraksi dan representasi ciri wajah, dan 4) klasifikasi ekspresi. Setiap tahapan memiliki peran dan teknik spesifik yang digunakan untuk mencapai hasil yang optimal yang diilustrasikan pada Gambar 2.5.



Gambar 2.5 Kerangka Dasar Sistem Pengenalan ekspresi wajah

Pertama, **deteksi wajah** merupakan tahap awal yang penting dalam sistem pengenalan ekspresi wajah. Deteksi wajah bertujuan untuk menemukan dan mengisolasi wajah dari latar belakang dalam sebuah citra dan video. Teknik deteksi wajah yang umum digunakan termasuk metode berbasis ciri seperti haar kaskade dan metode berbasis model seperti jaringan saraf konvolusional (Huang & Huang, 2023). Deteksi wajah yang akurat sangat penting untuk memastikan bahwa area wajah yang relevan diekstraksi untuk analisis lebih lanjut.

Kedua, setelah wajah terdeteksi, tahap berikutnya adalah **penyelarasan wajah**. Penyelarasan wajah bertujuan untuk menormalkan orientasi dan posisi wajah sehingga ciri-ciri wajah dapat diekstraksi secara konsisten. Teknik yang sering digunakan dalam penyalarsan wajah melibatkan pendeteksian titik-titik kunci wajah, seperti mata, hidung, dan mulut, yang kemudian digunakan untuk merotasi dan menyesuaikan wajah ke dalam posisi standar (Jabberi dkk., 2023). Penyelarasan yang baik membantu meningkatkan akurasi ekstraksi ciri tekstur dan klasifikasi ekspresi.

Ketiga, **ekstraksi dan representasi ciri** wajah merupakan tahap di mana informasi penting dari wajah diidentifikasi dan diubah menjadi representasi numerik yang dapat dianalisis oleh algoritma klasifikasi. Teknik ekstraksi ciri tekstur dapat dibagi menjadi dua kategori utama: ciri geometris dan ciri tekstur. Ciri geometris melibatkan pengukuran jarak dan sudut antara titik-titik kunci wajah, sedangkan ciri tekstur melibatkan analisis pola intensitas piksel di sekitar titik-titik kunci tersebut. Metode ekstraksi ciri tekstur modern sering kali menggunakan teknik berbasis pembelajaran mendalam, seperti jaringan saraf konvolusional (JSK), untuk menangkap representasi ciri yang lebih kompleks dan informatif (Goulart dkk., 2019).

Keempat adalah **klasifikasi** ekspresi merupakan tahap terakhir di mana ciri-ciri wajah yang telah diekstraksi digunakan untuk mengidentifikasi ekspresi wajah. Algoritma klasifikasi yang umum digunakan termasuk support vector machine (SVM), random forest (RF), dan jaringan saraf tiruan (JST). Algoritma ini dilatih menggunakan basisdata yang telah diberi label untuk mengenali berbagai ekspresi wajah, seperti bahagia, sedih, marah, dan terkejut. Hasil klasifikasi kemudian dapat digunakan dalam berbagai aplikasi, termasuk interaksi manusia-komputer, pengawasan keamanan, dan analisis perilaku (Kashif dkk., 2021; M. Li dkk., 2021).

Setiap tahapan dalam skema dasar sistem pengenalan ekspresi wajah memainkan peran penting dalam memastikan akurasi dan efisiensi sistem. Dengan kemajuan dalam teknologi pembelajaran mesin dan visi komputer, sistem pengenalan ekspresi wajah terus mengalami peningkatan kinerja dan aplikasi yang lebih luas di berbagai bidang.

SEKOLAH PASCASARJANA

2.3.4 Prapengolahan Citra

Prapengolahan data citra merupakan langkah kritis dalam analisis citra yang bertujuan untuk meningkatkan kualitas dan kegunaan data sebelum diterapkan pada algoritma lebih lanjut. Salah satu masalah yang sering muncul pada citra adalah derau impulsif atau *salt-and-pepper noise*. Derau impulsif sering ditemukan pada

citra karena berbagai alasan termasuk gangguan teknis seperti dari alat penangkap dan pengiriman data atau kurangnya pencahayaan yang menyebabkan objek citra menjadi kabur, dan lain-lain, oleh karena itu derau impulsif perlu dihilangkan karena menghalangi pengambilan informasi dari citra asli. Derau impulsif disebut juga sebagai lonjakan intensitas karena merusak citra di mana piksel yang mengandung derau mengambil nilai maksimum dan minimum dalam rentang dinamis (0, 255) yang artinya mengaburkan informasi penting pada citra asli. Hal yang sama juga dilaporkan oleh Deivalakshmi dan Palanisamy (2016) mengungkapkan temuan bahwa jumlah derau impulsif yang lebih sedikit secara signifikan dapat menurunkan kualitas citra, dengan demikian menghilangkan derau impulsif merupakan aspek fundamental yang perlu dilakukan sebelum dilakukan pemodelan. Derau impulsif pada citra dapat didefinisikan dengan Persamaan 2.1.

$$y_{i,j} = \begin{cases} P_{max}, & \text{untuk } p_{i,j} = 2^k - 1 \\ P_{min}, & \text{untuk } q_{i,j} = 0 \\ x_{i,j}, & \text{untuk } i,j = 1 - (P_s + P_p) \end{cases} \quad (2.1)$$

di mana $y_{i,j}$, $x_{i,j}$ menunjukkan citra asli dan citra derau yang sesuai di lokasi piksel i, j . Nilai dari P_s , P_p adalah nilai abu-abu maksimum $p_{i,j} = 2^k - 1$ dan minimum $q_{i,j} = 0$ dari citra dengan nilai k piksel yang terpengaruh oleh derau. Nilai $q_{i,j}$ pada probabilitas derau garam (*salt*) P_{min} ditetapkan dengan nilai 0 untuk semua area piksel i, j . Demikian pula, kami menetapkan $p_{i,j} = 2^k - 1$ ke semua area lokasi i, j di mana nilai itu muncul sebagai derau. Kesimpulan dari persamaan di atas adalah $p = q$ jika tingkat deraunya merata.

Tapis median (TM) merupakan salah satu metode yang paling menonjol dalam prapengolahan data citra. Metode ini sangat efektif dalam mengatasi derau, terutama derau impulsif atau salt-and-pepper noise, yang sering muncul dalam data citra (J. Wu dkk., 2022). Metode TM berkerja secara nonlinier dari operator lokal citra yang bertujuan mengurangi derau yang memiliki karakteristik unik (Erkan

dkk., 2020; Fan dan Tjahjadi, 2019). Metode ini pertama kali diusulkan oleh Petrou dan Petrou (2010).

1. Definisi

Didefinisikan sebuah citra $C_{(i,j)}$ yang terdiri dari nilai-nilai piksel $p(m,n)$ dengan ukuran kernel matriks 3×3 $\{H(k,l)\}$ di mana $\{H(k,l)\} \in C_{(i,j)}$. Larik matriks ini akan mengklasifikasikan sembilan piksel dengan memindai seluruh citra dari kombinasi bawah-atas dan kiri-kanan. Mendeteksi derau impuls dari citra ekspresi wajah, dengan contoh matriks 3×3 diberikan dalam Persamaan (2.2), sedangkan menghapus derau dan mengganti nilai median pada citra mengikuti Persamaan (2.1).

$$H(k,l) = \begin{bmatrix} p(m-1,n-1) & p(m-1,n) & p(m-1,n+1) \\ p(m,n-1) & p(m,n) & p(m,n+1) \\ p(m+1,n-1) & p(m+1,n) & p(m+1,n+1) \end{bmatrix} \quad (2.2)$$

Di sini, $p(m,n)$ adalah nilai piksel pada koordinat (m,n) .

2. Proses Tapis median

a. Extraksi nilai pixel

Untuk setiap piksel $p(m,n)$ dalam citra C , ekstrak nilai piksel dari lingkungan kernel 3×3 , dengan mengikuti Persamaan 2.3.

$$\{p(m-1,n-1), p(m-1,n), p(m-1,n+1), p(m,n-1), p(m,n), p(m,n+1), p(m+1,n-1), p(m+1,n), p(m+1,n+1)\} \quad (2.3)$$

b. Hitung nilai median

Urutkan nilai-nilai piksel yang diekstrak. Misalnya, jika nilai-nilai yang diekstrak adalah

{100, 200, 50, 150, 175, 25, 125, 75, 225}.

Setelah diurutkan, menjadi

{25, 50, 75, 100, **125**, 150, 175, 200, 225}.

Nilai median merupakan nilai tengah dalam deret yang diurutkan. Untuk kernel 3×3 , nilai median adalah nilai kelima (karena ada sembilan nilai).

Nilai median yang terpilih adalah *median* = 125.

Secara umum, nilai median $n(l)$ dari deret nilai $\{y_{-n}(l), \dots, y_{-1}(l), y_0(l), y_1(l), \dots, y_n(l)\}$ dinyatakan pada Persamaan 2.4.

$$n(l) = \text{median}\{y_{-n}(l), \dots, y_{-1}(l), y_0(l), y_1(l), \dots, y_n(l)\} \quad (2.4)$$

c. Ganti nilai piksel yang hilang dengan nilai median

Gantikan nilai piksel pusat $p(m, n)$ dengan nilai median yang telah dihitung:

$$p_{\text{output}}(m, n) = \text{median}\{p(m-1, n-1), p(m-1, n), p(m-1, n+1), p(m, n-1), p(m, n), p(m, n+1), p(m+1, n-1), p(m+1, n), p(m+1, n+1)\}$$

d. Iterasi untuk seluruh citra

Pindahkan kernel ke setiap piksel dalam citra dan ulangi proses penghitungan median serta penggantian nilai piksel. Secara matematis, untuk setiap piksel (i, j) dalam citra C , lakukan:

$$p_{output}(i,j) = median\{p(i+k, j+l) \mid k, l \in \{-1,0,1\}\}$$

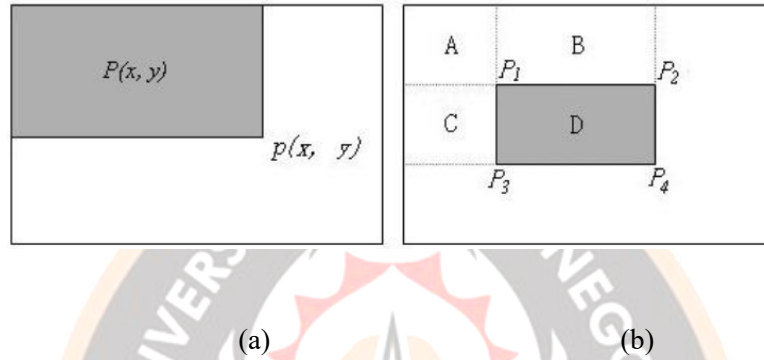
Ciri citra yang lebih kecil dari setengah ukuran tapis median sepenuhnya dihapus oleh tapis median. Kelebihan dari teknik ini adalah intensitas tingkat keabuan oleh tapis median tidak mempengaruhi garis tepi objek citra secara signifikan, meskipun posisinya dapat bergeser beberapa piksel. Oleh karena itu, operasi nonlinier dari tapis median mampu mengurangi derau garam-merica dari citra tanpa menghapus tepi secara signifikan.

2.3.5 Deteksi dan Penyelarasan Wajah

Masalah yang sering muncul pada PEW adalah adanya latar belakang objek wajah yang beragam yang dapat mempengaruhi lambatnya komputasi model dan dapat menyebabkan overfitting. Kaskade Haar (KH) telah sukses menjadi solusi yang banyak digunakan untuk mendeteksi objek wajah pada sebuah citra yang mengandung beberapa wajah dengan latar belakang yang beragam. Kelebihan dari FCH adalah komputasi yang lebih cepat dan akurat, namun generalisasi klasifikasi ciri wajah dapat menyebabkan overfitting selama proses pelatihan karena derau pada citra sulit dihilangkan (Mustafa dkk., 2014). Dalam penelitian ini, kami menggunakan penggabungan antara metode tapis median dan FCH, dimana kedua metode tersebut saling mengatasi kelemahan masing-masing.

Metode FCH adalah fungsi matematika yang berbentuk kotak dan memiliki prinsip yang sama dengan fungsi Fourier dengan bentuk ciri persegi panjang (*rectangular features*). Metode ini mengenali objek berdasarkan nilai sederhana dari ciri wajah seperti mata, hidung, dan bibir, tetapi bukan berdasarkan nilai piksel dari keseluruhan objek citra. Selanjutnya, operasi integral citra digunakan untuk mengetahui ada tidaknya ciri wajah di setiap lokasi citra asli. Ilustrasinya dapat dilihat pada Gambar 2.6. Kelebihan metode ini adalah komputasi yang sangat cepat karena hanya bergantung pada jumlah piksel dalam persegi panjang, bukan pada

setiap nilai piksel dari seluruh area citra (Murugappan dan Mutawa, 2021). Citra integral ditunjukkan pada Gambar 2.6.



Gambar 2.6 Konsep Citra Integral

Seperti yang ditunjukkan pada Gambar 2.6 (a), lokasi piksel (x, y) berisi jumlah nilai piksel. Menghitung setiap nilai piksel dapat didefinisikan dengan Persamaan (2.5).

$$IC(x, y) = \sum_{x'=0}^x \sum_{y'=0}^y N(x', y') \quad (2.5)$$

di mana dapat dihitung secara rekursif (lihat Persamaan (2.6)):

$$P(x, y) = IC(x, y - 1) + P(x - 1, y) + N(x, y) - P(x - 1, y - 1) \quad (2.6)$$

dengan kondisi sesuai Persamaan (2.7),

$$P(-1, y) = P(x, -1) = P(-1, -1) = 0 \quad (2.7)$$

sehingga hanya membutuhkan satu kali pemindaian pada citra masukan. Berdasarkan definisi "citra integral" di atas yang diilustrasikan pada Gambar 2.6.(a), jumlah nilai piksel dalam area D pada Gambar 2.6.(b) dapat dihitung dengan Persamaan (2.8).

$$D = P_4 + P_3 - P_2 - P_1 \quad (2.8)$$

di mana, $P_1 = A$, $P_2 = A + B$, $P_3 = A + C$, dan $P_4 = A + B + C + D$.

Untuk mengenali objek yang menarik, citra dipindai oleh sub-jendela yang berisi ciri persegi panjang tertentu. Hasil pengenalan wajah menggunakan Ciri Haar terdiri dari dua atau tiga persegi panjang "hitam" dan "putih" yang terhubung, ilustrasinya dapat dilihat pada Gambar 2.7. Berdasarkan ciri persegi panjang yang terbentuk dari ciri Haar, pengklasifikasi koresponden $h_j(x)$ didefinisikan dengan Persamaan (2.9).

$$h_j(x) = \begin{cases} 1, & \text{jika } p_j f_j(x) < p_j \theta_j \\ 0 & \text{lainnya} \end{cases} \quad (2.9)$$

di mana x adalah sub-jendela, dan θ_j adalah ambang batas. Seperti yang ditunjukkan pada Persamaan (2.9), p_j menunjukkan arah tanda pertidaksamaan.



Gambar 2.7 Pelacakan objek wajah menggunakan FCH (Q. Chen et al., 2007)

Tahapan metode FCH adalah sebagai berikut:

Tahap 1 : Inisialisasi

diketahui $(x_1, y_1) \dots (x_n, y_n)$ adalah citra masukan dimana $y_i \in \{-1, 1\}$ adalah contoh negatif dan positif sesuai;

Tahap 2 : Inisialisasi bobot

$$w_{1,i} = \begin{cases} \frac{1}{2p}, & \text{jika } y_i = -1 \\ \frac{1}{2q}, & \text{jika } y_i = 1 \end{cases}$$

Pada langkah pertama inisialisasi $w_{1,i}$ dilakukan. $w_{1,i}$ adalah bobot awal untuk sampel ke- i . Terdapat dua kondisi yang menentukan nilai dari $w_{1,i}$ yaitu jika $y_i = -1$, maka bobot $w_{1,i}$ diatur menjadi $\frac{1}{2p}$ di mana p adalah jumlah contoh negative dalam basisdata. Jika $y_i = 1$, maka bobot $w_{1,i}$ diatur menjadi $\frac{1}{2q}$ di mana q adalah jumlah contoh positif dalam basisdata.

Tahap 3 : Iterasi

for $t = 1 \dots T$, lakukan:

- a) **Normalisasi bobot** dengan mengikuti Persamaan (2.10):

$$w_{t,i} = \frac{w_{t,i}}{\sum_{i=1}^N w_{t,i}} \quad (2.10)$$

- b) **Pelatihan dan pengklasifikasi** Untuk setiap ciri j , latih pengklasifikasi h_j yang terbatas hanya pada satu ciri (ciri wajah). Evaluasi galat (*error*) dengan Persamaan (2.11).

$$\epsilon_j = \sum_i w_i |h_j(x_i) - (y_i)| \quad (2.11)$$

c) **Pilih pengklasifikasi terbaik:** pilih pengklasifikasi h_t dengan tingkat kesalahan minimum (*minimum error rate*) ϵ_t .

d) Lakukan **pembaruan bobot**

$$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

Untuk memperbarui bobot $w_{t+1,i}$ pada iterasi ke- $(t + 1)$ berdasarkan bobot sebelumnya $w_{t,i}$ dengan mempertimbangkan nilai β_t dan kesalahan e_i . Nilai t adalah indeks iterasi algoritma.

- $w_{t+1,i}$: Bobot untuk sampel ke- i pada iterasi ke- $(t + 1)$
- $w_{t,i}$: Bobot untuk sampel ke- i pada iterasi ke- t
- β_t : Faktor pembaruan bobot pada iterasi ke- t . Nilai β_t ditentukan berdasarkan tingkat kesalahan ϵ_t pada iterasi ke- t

$$\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}$$

- e_i : Indikator kesalahan untuk sampel ke- i . Nilai e_i digunakan untuk menentukan apakah sampel ke- i diklasifikasikan benar atau salah pada indeks iterasi ke- t . Bernilai 0 jika sampel diklasifikasikan dengan benar, dan bernilai 1 jika sampel diklasifikasikan dengan salah

e) Pembentukan klasifikasi yang kuat (lihat Persamaan (2.12)):

$$h(x) = \begin{cases} 1, & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ -1, & \text{Sebaliknya} \end{cases} \quad (2.12)$$

di mana $h(x)$ adalah pengklasifikasi kuat yang dibentuk dari kombinasi beberapa pengklasifikasi lemah $h_t(x)$. Nilai $h(x)$ ditentukan berdasarkan jumlah tertimbang dari pengklasifikasi lemah $h_t(x)$. Jika jumlah tertimbang dari pengklasifikasi lemah $\sum_{t=1}^T \alpha_t h_t(x)$ lebih besar atau sama dengan setengah dari jumlah total bobot $\frac{1}{2} \sum_{t=1}^T \alpha_t$, maka $h(x)$ akan diklasifikasikan sebagai 1. Jika tidak maka $h(x)$ akan diklasifikasikan sebagai -1. α_t adalah bobot dari pengklasifikasi lemah ke- t . α_t dihitung sebagai $\log\left(\frac{1}{\beta_t}\right)$, di mana β_t adalah tingkat kesalahan dari pengklasifikasi lemah ke- t . Kesimpulannya:

- $h(x)$ adalah hasil akhir dari penggabungan beberapa pengklasifikasi lemah $h_t(x)$ yang bekerja bersama-sama untuk membuat keputusan akhir.
- $h_t(x)$ adalah pengklasifikasi sederhana yang memberikan prediksi awal (misalnya, -1 atau 1) berdasarkan ciri tertentu.
- Bobot α_t menunjukkan seberapa penting pengklasifikasi lemah $h_t(x)$ dalam membuat keputusan akhir. Semakin kecil tingkat kesalahan β_t , semakin besar bobot α_t .

2.3.6 Ekstraksi Ciri Tekstur

Ekstraksi ciri dalam pengenalan ekspresi wajah merupakan proses penting yang bertujuan untuk mengekstrak informasi yang relevan dan bermakna dari gambar wajah sehingga dapat mengidentifikasi ekspresi yang sedang ditampilkan. Dalam konteks ini, ekstraksi ciri mengacu pada proses memilih dan mengekstrak atribut-atribut yang relevan dari wajah manusia, seperti bentuk bibir, mata, hidung, dan area wajah lainnya yang memiliki peran penting dalam mengungkapkan emosi.

Metode ekstraksi ciri untuk pengenalan ekspresi wajah dapat bervariasi, mulai dari teknik-teknik sederhana hingga pendekatan yang lebih kompleks. Salah satu pendekatan yang umum digunakan adalah menggunakan algoritma deteksi wajah untuk menentukan wilayah-wilayah penting pada gambar, seperti titik-titik landmark wajah. Setelah wilayah-wilayah ini teridentifikasi, ciri-ciri seperti jarak antara mata, perubahan intensitas piksel di sekitar mata, hidung, atau mulut dapat diekstrak, pendekatan ini juga sering disebut pendekatan geometris atau *hancarfetd features*. Pendekatan geometris merupakan ciri-ciri yang secara manual dirancang atau dihasilkan oleh manusia berdasarkan pengetahuan domain spesifik atau aturan-aturan yang telah ditetapkan sebelumnya. Ekstraksi ciri tekstur yang didasarkan pada pengetahuan manusia tentang masalah yang dihadapi dan karakteristik yang dianggap penting untuk dianalisis. Kelebihan dari pendekatan geometris adalah interpretabilitas yang tinggi, di mana manusia dapat memahami ciri yang diekstrak dan bagaimana ciri-ciri tersebut berkontribusi dalam proses analisis. Akan tetapi memiliki keterbatasan dalam mengekstrak ciri-ciri yang mungkin penting tetapi sulit diwakili secara manual, terutama dalam kasus data yang kompleks atau memiliki varian yang tinggi.

Teknik machine learning modern, seperti deep learning, cenderung mengurangi ketergantungan pada handcrafted features dengan memungkinkan pembelajaran ciri dari data mentah (*end-to-end learning*). Melalui pendekatan ini, jaringan saraf dapat belajar sendiri untuk mengekstrak ciri-ciri yang relevan dari

data, seringkali menghasilkan representasi ciri yang lebih abstrak dan kompleks daripada yang bisa dirancang secara manual dalam pendekatan geometris.

Proses ekstraksi ciri tekstur bertujuan untuk mengurangi dimensi data wajah dan menyoroti informasi yang paling penting untuk mengidentifikasi ekspresi yang ditampilkan. Dengan kata lain, ekstraksi ciri tekstur memungkinkan sistem untuk fokus pada aspek-aspek kunci dari wajah yang diperlukan dalam pengenalan ekspresi, yang kemudian dapat digunakan sebagai input untuk proses klasifikasi ekspresi. Pentingnya ekstraksi ciri tekstur ini terletak pada kemampuannya untuk mereduksi kompleksitas data wajah menjadi representasi yang lebih sederhana dan informatif, memfasilitasi proses pengenalan ekspresi dengan tingkat akurasi yang lebih tinggi.

1. Ekstraksi Ciri Geometris

a. HOG

Algoritma *Histogram of Oriented Gradients* (HOG) adalah teknik ekstraksi ciri tekstur yang digunakan dalam pengolahan citra untuk mengidentifikasi pola atau objek berdasarkan distribusi gradien orientasi dalam gambar. Tahapan-tahapan utama dari algoritma HOG adalah sebagai berikut:

- Tahap 1. Konversi Gambar ke Skala Abu-Abu: Langkah pertama adalah mengubah gambar berwarna atau RGB menjadi citra skala abu-abu untuk mempermudah perhitungan dan mengurangi kompleksitas.
- Tahap 2. Perhitungan Gradien: Menggunakan operator Sobel atau Prewitt untuk menghitung gradien (magnitudo dan arah gradien) dari gambar. Ini membantu dalam menemukan perubahan intensitas piksel yang menandakan tepi atau batas objek.
- Tahap 3. Pembagian menjadi Sel (Cells): Gambar dibagi menjadi sel-sel kecil (cell) yang bertetangga. Gradien piksel dalam setiap sel dihitung.

- Tahap 4. Perhitungan Histogram Gradien: Untuk setiap sel, histogram gradien orientasi dihitung. Gradien diarahkan ke arah bins (kotak) berbeda yang merepresentasikan rentang orientasi tertentu. Histogram ini merefleksikan distribusi gradien orientasi dalam sel.
- Tahap 5. Normalisasi Blok (Block Normalization): Setelah perhitungan histogram dalam setiap sel, blok-blok sel yang tumpang tindih (overlap) dibuat. Normalisasi dilakukan pada blok-blok ini dengan menghitung nilai vektor gradien normalisasi untuk menekan efek variasi cahaya dan kontras yang mungkin ada dalam gambar.
- Tahap 6. Deskriptor HOG: Deskriptor HOG akhir dihasilkan dengan menggabungkan histogram dari setiap sel ke dalam satu vektor yang merepresentasikan ciri-ciri dari seluruh gambar. Vektor ini mengandung informasi tentang distribusi gradien orientasi dalam gambar.

Proses ini menghasilkan representasi ciri yang relatif sederhana namun informatif dari citra, yang kemudian dapat digunakan sebagai input untuk berbagai tugas pengenalan objek atau deteksi pola dalam citra. HOG memiliki keunggulan dalam mendeteksi struktur seperti tepi, batas objek, dan tekstur dalam citra, sehingga sering digunakan dalam aplikasi deteksi wajah, pengenalan objek, dan pemrosesan citra lainnya.

b. LBP

Algoritma local binary patterns (LBP) adalah teknik ekstraksi ciri yang digunakan untuk mengekstrak ciri tekstur dari gambar dengan cara membandingkan intensitas piksel dengan tetangganya dalam suatu area tertentu. Berikut adalah tahapan dari algoritma LBP:

- Tahap 1. Pemilihan Piksel Pusat: Pilih piksel pusat pada gambar grayscale yang akan diekstrak ciri teksturnya.
- Tahap 2. Pembentukan Pola Biner:

- a. Tentukan area tetangga sekitar piksel pusat yang akan digunakan untuk perhitungan. Biasanya, area tetangga berbentuk lingkaran atau persegi dengan radius tertentu.
- b. Bandingkan nilai intensitas piksel pusat dengan nilai intensitas tetangga sekitarnya. Jika nilai intensitas tetangga lebih besar atau sama dengan nilai intensitas piksel pusat, tandai sebagai 1; jika kurang, tandai sebagai 0.
- c. Misalkan ada 8 tetangga di sekeliling piksel pusat, akan dihasilkan pola biner 8-bit berdasarkan perbandingan nilai intensitas ini.

Tahap 3. Konversi Pola Biner ke Desimal: Konversikan pola biner yang dihasilkan menjadi bilangan desimal. Misalnya, pola biner '10010110' akan menjadi bilangan desimal '150'.

Tahap 4. Membuat Histogram dari Pola yang Ditemukan:

- a. Lakukan langkah-langkah di atas untuk setiap piksel pada gambar untuk menghasilkan pola LBP.
- b. Buat histogram dari pola LBP yang dihasilkan dari seluruh gambar. Histogram ini merepresentasikan distribusi pola LBP dalam gambar.

Algoritma LBP menghasilkan representasi ciri berupa histogram dari pola biner yang digunakan untuk menggambarkan tekstur dalam gambar. Histogram ini mencerminkan sebaran variasi pola tekstur yang ditemukan pada gambar. Kelebihannya, kecepatan komputasinya yang relatif tinggi dan kemampuannya dalam menangkap informasi tentang tekstur lokal dalam gambar tanpa bergantung pada perubahan intensitas global.

c. FG

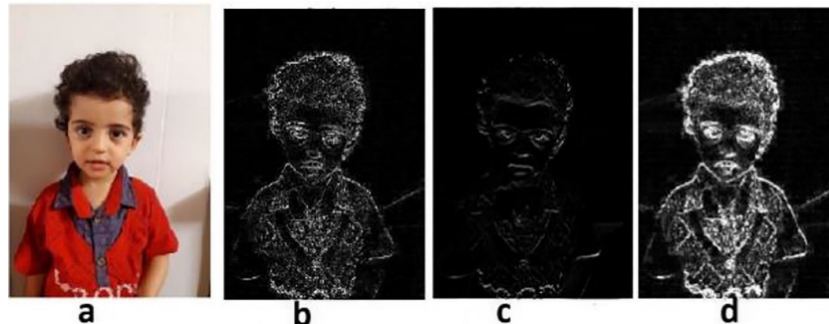
Tapis Gabor (FG) pertama kali diperkenalkan oleh Gabor (1947) dan dipopulerkan oleh Mehrotra dkk. (1992) merupakan teknik pengolahan citra yang digunakan untuk analisis frekuensi dan orientasi dalam domain spasial. Metode ini menggabungkan kernel Gaussian dengan gelombang sinusoidal untuk mengekstrak informasi frekuensi dan orientasi dari suatu citra, seperti yang didefinisikan pada Persamaan (2.13).

$$G(GF, \alpha, \beta) = \frac{GF \cdot \gamma}{2 \times \pi \times GF \cdot \sigma^2} \times \underbrace{\exp\left(-\frac{GF \cdot \gamma^2 \times \beta'^2}{2 \times GF \cdot \sigma^2}\right)}_{\text{Fungsi Gaussian}} \times \underbrace{\exp(i(2 \times \pi \times GF \cdot f \times \alpha' + GF \cdot \varphi))}_{\text{Fungsi Sinusoid}} \quad (2.13)$$

di mana α , dan β adalah koordinat dari 2D-mesh persegi, $\alpha' = \alpha \cos GF \cdot \theta + \beta \sin GF \cdot \theta$, $\beta' = -\alpha \sin GF \cdot \theta + \beta \cos GF \cdot \theta$, dan GF adalah struktur data yang berisi nilai empat parameter yang memengaruhi pembuatan tapis Gabor yaitu f, θ, σ , dan γ (Mohammadian dkk., 2020).

Tapis Gabor dapat menciptakan tiga jenis ciri, *gabor real* (GR), *gabor imaginery* (GI), dan ciri *energy gabor* (EG). GR dibuat dengan menerapkan mask *real* pada citra, dan GI dibuat dengan menerapkan mask *imaginary* pada citra, dalam penelitian ini menggunakan CEG, didefinisikan dengan Persamaan (2.14). Gambar 2.8 menunjukkan ciri nyata Gabor, ciri imajiner Gabor, dan ciri energi Gabor dari sebuah gambar sampel. Seperti yang terlihat jelas dari gambar, ciri energi Gabor dapat menunjukkan komponen wajah dengan lebih baik daripada dua ciri lainnya.

$$EG(\text{nilai piksel}, GF) = \sqrt{GR(\text{nilai piksel}, GF)^2 + GI(\text{nilai piksel}, GF)^2} \quad (2.14)$$



Gambar 2.8 Bagian-bagian dari Keluaran Tapis Gabor: **a.** Citra asli; **b.** GR; **c.** GI; dan **d.** EG. Ciri Ekstraksi Geometris dapat Menunjukkan Komponen Wajah dengan Lebih baik Daripada Dua Ciri Lainnya

d. DWT

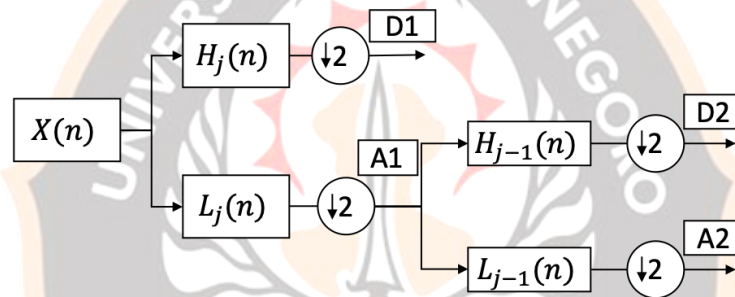
DWT populer digunakan sebagai ekstraksi ciri tekstur citra untuk menangkap variasi temporal dari gerakan wajah dan ekspresi dari nilai koordinat 2D yang bervariasi terhadap waktu dan tidak stasioner yang bersifat dinamis pada pengenalan ekspresi wajah (Othman dan Zeebaree, 2020). Metode ini berfungsi sebagai filter ketika menguraikan sinyal mentah menjadi koefisien rinci, di mana sinyal pita frekuensi tinggi mengelompok dalam koefisien rinci (Satiyan dkk., 2010). Karena sinyal frekuensi tinggi terkontaminasi dengan derau, koefisien aproksimasi digunakan untuk analisis lebih lanjut. Persamaan untuk fungsi wavelet induk $\Psi(t)$ didefinisikan pada Persamaan (2.15).

$$\Psi(t) = a^{-\frac{1}{2}} \Psi\left(\frac{t-b}{a}\right) \quad (2.15)$$

di mana, $a, b \in R (a > 0)$, R adalah ruang wavelet. Parameter ' a ' adalah faktor penskalaan dan ' b ' adalah faktor pergeseran.

DWT menguraikan sinyal input menjadi sinyal aproksimasi dan sinyal detail. Sinyal aproksimasi kemudian dibagi menjadi sinyal aproksimasi dan detail yang

baru. Proses ini dilakukan secara iteratif menghasilkan satu set sinyal aproksimasi pada tingkat detail (skala) yang berbeda ke sinyal aproksimasi kotor akhir (Murugappan dkk., 2009). Struktur bank filter dari transformasi wavelet ditunjukkan pada Gambar 2.9, di mana $X(n)$ adalah input yang diberikan, $H(n)$ dan $L(n)$ adalah koefisien bank filter low pass dan high pass, 'j' menunjukkan jumlah level dekomposisi dan A dan D masing-masing menunjukkan ruang aproksimasi dan ruang detail.

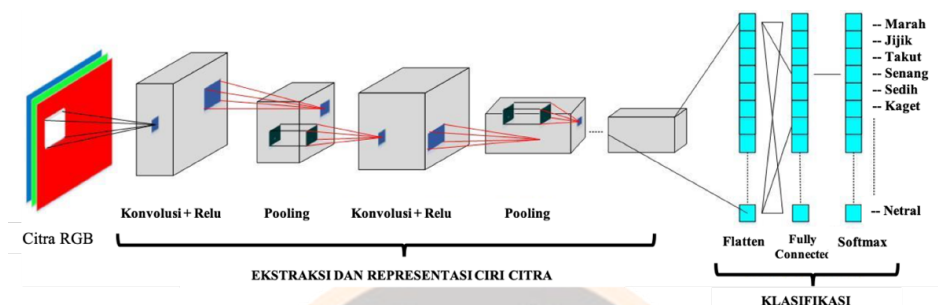


Gambar 2.9 Struktur Blok Filter DWT

2. Ekstraksi ciri Mendalam

a. JSK untuk ekstraksi ciri tekstur

Jaringan saraf konvolusional (JSK) untuk ekstraksi ciri yang mendalam pertama kali diusulkan oleh (Marblestone dkk., 2016) diberi nama *deep neural networks* (DNNs) di mana bertujuan melakukan ekstraksi ciri data secara otomatis. Walaupun JSK menghasilkan kinerja yang luar biasa akan tetapi memiliki keterbatasan seperti ukuran kernel konvolusi hanya fokus pada pola lokal. Jika citra masukan mengandung derau, maka JSK mungkin tidak cukup efektif karena akan dideteksi sebagai piksel. Arsitektur JSK umumnya terdiri dari tiga kerangka utama yaitu masukan data, lapisan ekstraksi dan representasi ciri, dan lapisan klasifikasi jaringan. Struktur arsitektur JSK dapat ditunjukkan pada Gambar 2.10.



Gambar 2.10 Arsitektur JSK (Gambar yang dibuat termotivasi dari (Islam et al., 2018))

b. Pra-latih (*transfer learning*)

Meskipun deep learning memiliki kapasitas penyesuaian yang kuat, namun masih membutuhkan waktu dan proses yang sangat lama untuk mempelajari data berjumlah banyak dan berlabel multidimensi. Pra-latih (*pre-trained*) menjawab masalah ini dengan mempelajari pengetahuan sebelumnya dari basisdata yang diberikan kemudian menggunakan kembali parameter tersebut untuk tugas-tugas selanjutnya. Akan tetapi, cara ini belum dapat digeneralisasi penggunaannya ke domain yang distribusinya berbeda dari data pelatihan. Alasannya terletak pada asumsi utama algoritma pembelajaran bahwa data latih dan uji di kasus yang berbeda memiliki distribusi basisdata yang sama. Dalam banyak skenario di dunia nyata, asumsi ini tidak berlaku. Hal ini dapat diatasi dengan adaptasi domain yang bertujuan untuk mentransfer pengetahuan dari domain sumber yang memiliki label ke domain target yang minim label atau tidak berlabel (Anzalone dkk., 2019). Ilustrasinya dinyatakan pada Persamaan (2.16),

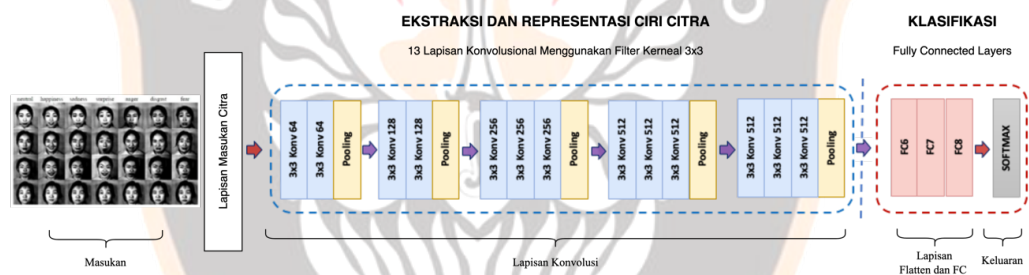
$$p(x) \neq q(x) \quad (2.16)$$

dimana p sebagai sumber distribusi pelatihan dan q sebagai distribusi target. Untuk tugas yang sama, asumsi adaptasi domain adalah bahwa distribusi output bersyarat

adalah invarian, yaitu $p(y|x) = q(y|x)$, di mana x , y masing-masing adalah data masukan dan label keluaran. Sehingga dengan cara ini, distribusi data pelatihan dan pengujian pada kasus yang berbeda dapat dilakukan.

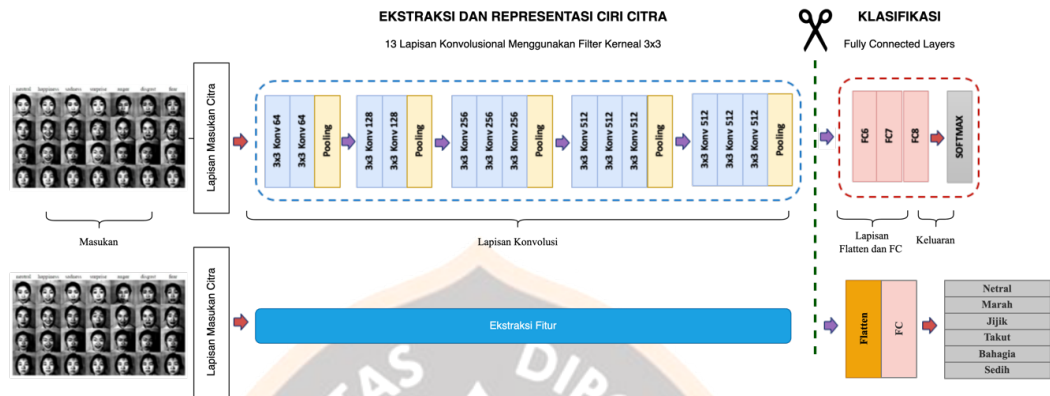
c. VGG16

VGG16 adalah arsitektur pada jaringan saraf konvolusional (JSK) yang dikembangkan dan dipopulerkan pertama kali oleh (Simonyan dan Zisserman, 2015) dari Universitas Oxford. Arsitektur VGG16 terdiri dari 16 lapisan, termasuk 13 lapisan konvolusional dan 3 lapisan fully connected, arsitektur lengkap ditunjukkan pada Gambar 2.11. Dalam kasus ekstraksi ciri wajah, 3 lapisan fully connected ditiadakan. Arsitektur VGG16 untuk ekstraksi ciri ditunjukkan pada Gambar 2.12.



Gambar 2.11 Arsitektur VGG16 (digambar ulang berdasarkan penelitian (Simonyan & Zisserman, 2015))

SEKOLAH PASCASARJANA



Gambar 2.12 Arsitektur VGG16 untuk Ekstraksi ciri

Seperti yang diilustrasikan pada Gambar 2.12, setelah menghapus lapisan-lapisan fully connected (FC) dari VGG16 dan menggunakan lapisan-lapisan konvolusi atau max pooling sebagai ekstraktor ciri, keluaran akan berbentuk tensor tiga dimensi. Tensor tiga dimensi merupakan struktur data multidimensional yang memiliki tiga tingkat dimensi (merah, hijau, dan biru) di mana akan digunakan untuk merepresentasikan data spasial citra.

Ukuran dari keluaran ciri yang diekstrak bergantung pada jumlah filter atau channel pada lapisan terakhir sebelum lapisan FC. Misalnya, jika model VGG16 dihentikan sebelum lapisan FC, dan lapisan terakhir sebelumnya adalah lapisan konvolusi dengan 512 filter, bentuk data keluarannya akan menjadi (ukuran_batch, tinggi, lebar, 512).

Sebagai contoh, jika menggunakan VGG16 untuk mengolah satu citra dengan ukuran 224x224 piksel, lapisan konvolusi terakhir sebelum lapisan FC pada VGG16 memiliki 512 filter. Oleh karena itu, keluaran ciri yang diekstrak dari citra ini akan berbentuk (1, 7, 7, 512). Di sini, 1 menunjukkan jumlah sampel (karena hanya satu citra yang diproses), dan 7x7x512 menunjukkan dimensi representasi ciri: tinggi x lebar x jumlah filter. Ciri-ciri ini kemudian digunakan sebagai

masukannya pada model klasifikasi dengan tujuan mempelajari pola-pola yang lebih spesifik dalam memprediksi ekspresi wajah pada citra masukan.

2.3.7 Pengenalan ekspresi wajah

1. SVM

Support vector machine (SVM) merupakan salah satu algoritma pembelajaran mesin yang digunakan untuk masalah klasifikasi dan regresi. Dalam kasus yang paling sederhana dari klasifikasi dua kelas, SVM menemukan hyperplane yang memisahkan dua kelas data dengan margin selebar mungkin (Sammot dan Webb 2011). Hal ini menghasilkan akurasi generalisasi yang baik pada data yang tidak terlihat, dan mendukung metode pengoptimalan khusus yang memungkinkan SVM untuk belajar dari sejumlah besar data.

SVM memiliki dasar matematika yang kuat dan terkait erat dengan beberapa teori yang sudah mapan dalam statistik. SVM tidak hanya mencoba mengklasifikasikan data pelatihan dengan benar, tetapi juga memaksimalkan margin untuk kinerja generalisasi yang lebih baik. Formulasi ini mengarah pada hyperplane pemisah yang hanya bergantung pada titik-titik data yang berada pada margin, yang disebut vektor pendukung. Oleh karena itu, seluruh algoritma ini disebut mesin vektor pendukung. Selain itu, karena masalah analisis data di dunia nyata sering kali melibatkan ketergantungan nonlinier, SVM dapat dengan mudah diperluas untuk memodelkan nonlinieritas tersebut dengan menggunakan kernel semi-definit positif. Selain itu, SVM dapat dilatih melalui pemrograman kuadratik, yang (1) membuat analisis teoritis lebih mudah, dan (2) memberikan banyak kemudahan dalam mendesain pemecah masalah yang efisien yang berskala untuk set data besar. Terakhir (3), ketika diterapkan pada data dunia nyata, SVM sering kali memberikan kinerja yang canggih dalam hal akurasi, fleksibilitas, ketahanan, dan efisiensi.

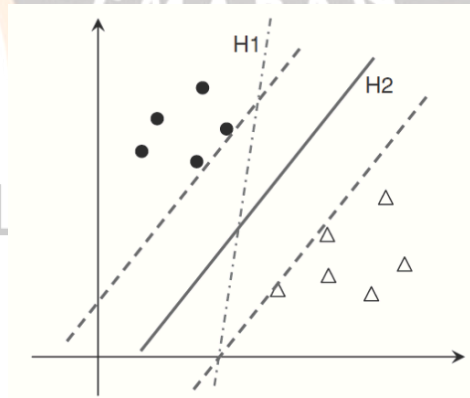
a. Hyperplane optimal untuk contoh yang dapat dipisahkan secara linier

Pertimbangkan set pelatihan $\{(x_i, y_i)\}_{i=1}^n$ di mana $x_i \in \mathbb{R}^p$ adalah vektor ciri input untuk contoh ke i , dan $y_i \in \{1, -1\}$ adalah label yang sesuai yang menunjukkan apakah contoh tersebut positif ($y_i = +1$) atau negatif ($y_i = -1$). Pertama-tama, kita asumsikan bahwa himpunan contoh positif dan negatif dapat dipisahkan secara linear. Ini berarti, ada sebuah fungsi $f(x) = \langle w, x \rangle + b$ dimana $w \in \mathbb{R}^p$ (disebut vektor bobot) dan $b \in \mathbb{R}$ (disebut bias) sehingga

$$\langle w, x_i \rangle + b > 0 \text{ untuk } y_i = +1 \quad (2.17)$$

$$\langle w, x_i \rangle + b < 0 \text{ untuk } y_i = -1 \quad (2.18)$$

Kita menyebut $\langle w, x_i \rangle + b = 0$ sebagai hyperplane keputusan dan pada kenyataannya, dapat terdapat beberapa hyperplane yang memisahkan contoh positif dan negatif, seperti yang ditunjukkan pada Gambar 2.13. Akan tetapi, hyperplane tersebut tidak dibuat sama. Terkait dengan setiap hyperplane tersebut adalah sebuah gagasan yang disebut margin, yang didefinisikan sebagai jarak antara hyperplane dan contoh terdekat. SVM bertujuan untuk menemukan hyperplane tertentu yang memaksimalkan margin.



Gambar 2.13 Pemisah Margin Maksimum

Secara matematis, mudah untuk memeriksa bahwa jarak dari titik x_i ke hyperplane $\langle w, x \rangle + b = 0$ adalah $\|w\|^{-1} |\langle w, x_i \rangle + b|$. Oleh karena itu, SVM mencari w, b yang optimal dari masalah optimasi yang didefinisikan pada Persamaan (2.19)

$$\begin{aligned} & \underset{w \in \mathbb{R}^p}{\text{maximize}} \quad \min_{1 \leq i \leq n} \frac{|\langle w, x_i \rangle + b|}{\|w\|}, \\ & \text{s.t.} \quad \begin{cases} \langle w, x_i \rangle + b > 0 & \text{if } y_i = +1 \\ \langle w, x_i \rangle + b < 0 & \text{if } y_i = -1 \end{cases} \quad \forall i \end{aligned} \quad (2.19)$$

Jelas bahwa jika (w, b) adalah solusi optimal, maka $(\alpha w, \alpha b)$ juga merupakan solusi optimal untuk setiap $\alpha > 0$.

Oleh karena itu, untuk memperbaiki skala, kita dapat secara ekuivalen mengatur pembilang dari $\min_{1 \leq i \leq n} |\langle w, x_i \rangle + b|$ ke 1, dan meminimalkan penyebut $\|w\|$ yang didefinisikan pada Persamaan (2.20).

$$\begin{aligned} & \underset{w \in \mathbb{R}^p, b \in \mathbb{R}}{\text{minimize}} \quad \|w\|^2, \\ & \text{s.t.} \quad \begin{cases} \langle w, x_i \rangle + b \geq 0 & \text{if } y_i = +1 \\ \langle w, x_i \rangle + b \leq 0 & \text{if } y_i = -1 \end{cases} \quad \forall i \end{aligned} \quad (2.20)$$

Persamaan di atas merupakan program kuadratik yang dibatasi secara linier, yang dapat diselesaikan secara efisien. Oleh karena itu, ini menjadi bentuk SVM (primal) yang paling umum digunakan untuk kasus yang dapat dipisahkan secara linear.

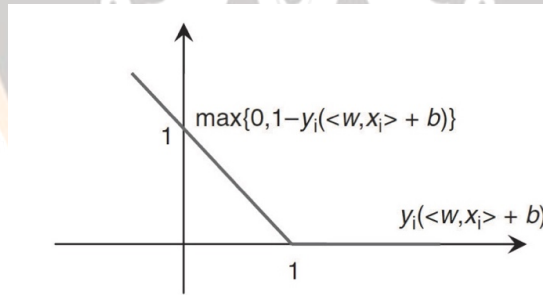
b. Margin lunak

Dalam praktiknya, sebagian besar, jika tidak semua, kumpulan data tidak dapat dipisahkan secara linier, yaitu tidak ada w dan b yang dapat memenuhi

batasan masalah optimasi. Dalam kasus ini, kita akan membiarkan beberapa titik data melanggar kondisi margin, dan menghukumnya dengan tepat. Secara matematis, perhatikan bahwa kendala-kendala pada Persamaan (2.20) dapat dituliskan secara ekuivalen sebagai $y_1(\langle w, x_i \rangle + b) \geq 1$. Sekarang, kita masukkan satu set baru variabel slack nonnegatif ξ_i ke dalam kendala: $y_1(\langle w, x_i \rangle + b) \geq 1 - \xi_i$ dan memasukkan penalti ke dalam tujuan awal untuk mendapatkan soft margin SVM, seperti yang didefinisikan pada (2.21).

$$\begin{aligned} \underset{w, b, \xi_i}{\text{minimize}} \quad & \lambda \|w\|^2 + \frac{1}{n} \sum_{i=1}^n \xi_i, \\ \text{s.t. } & y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i \text{ and } \xi_i > 0 \forall i \end{aligned} \quad (2.21)$$

$\lambda > 0$ adalah faktor *trade-off*. Penting untuk dicatat bahwa ξ_i dapat ditulis sebagai $\xi_i = \max \{0, 1 - y_i(\langle w, x_i \rangle + b)\}$, yang disebut dengan kehilangan engsel dan digambarkan pada Gambar 2.14.



Gambar 2.14 Grafik Kehilangan Engsel (Hinge Loss)

SEKOLAH PASCASARJANA

Dengan cara ini, masalah optimasi dapat dirumuskan kembali menjadi masalah non-lancar (*non-smooth problem*) yang tidak dibatasi:

$$\underset{w \in \mathbb{R}^p, b \in \mathbb{R}}{\text{minimize}} \quad \frac{\lambda}{2} \|w\|^2 + \frac{1}{n} \sum_{i=1}^n \max\{0, 1 - y_i(\langle w, x_i \rangle + b)\} \quad (2.22)$$

Perhatikan bahwa $\max\{0, 1 - y_i(\langle w, x_i \rangle + b)\}$ juga merupakan batas atas cembung dari $\delta(y_i(\langle w, x_i \rangle + b))$, di mana $\delta(x) = 1$ jika x benar dan 0 jika tidak. Oleh karena itu, penalti yang sering digunakan adalah batas atas cembung dari rata-rata kesalahan pelatihan. Ketika training set benar-benar dapat dipisahkan, masalah soft margin (Persamaan (2.21)) secara otomatis memulihkan masalah hard margin (Persamaan (2.20)) ketika λ cukup besar.

c. Bentuk ganda dan kernelisasi

Karena kendala dalam bentuk primal tidak mudah untuk ditangani, para peneliti secara konvensional menggunakan masalah dual dari Persamaan (2.21). Dengan mengikuti prosedur standar, seseorang dapat menurunkan dual Lagrangian, seperti yang didefinisikan pada Persamaan (2.23).

$$\begin{aligned} \min_a \quad & \frac{1}{2\lambda} \sum_{i,j} y_i y_j a_i a_j \langle x_i, x_j \rangle - \sum_i a_i, \\ \text{s. t. } & a_i \in [0, n^{-1}] \text{ and } \sum_i y_i a_i = 0 \end{aligned} \quad (2.23)$$

yang sekali lagi merupakan program kuadratik tetapi dengan kendala yang jauh lebih sederhana yaitu kendala kotak ditambah dengan satu kendala persamaan linier. Untuk mendapatkan solusi primal w^* dari solusi dual a_i^* , kita memiliki, seperti yang didefinisikan pada Persamaan (2.24).

$$w^* = \sum_{i=1}^n a_i^* y_i x_i \quad (2.24)$$

Contoh latih dapat dibagi menjadi tiga kategori berdasarkan nilai a_i^* . Jika $a_i^* = 0$, itu berarti contoh pelatihan yang sesuai tidak mempengaruhi keputusan batas, dan pada kenyataannya terletak di luar batas, yaitu, $y_i(\langle w, x_i \rangle + b) > 1$. Jika $a_i^* \in (0, n^{-1})$, maka contoh pelatihan terletak di dalam margin, yaitu $y_i(\langle w, x_i \rangle + b) = 1$. Jika $a_i^* = n^{-1}$ berarti contoh pelatihan melanggar margin, yaitu $y_i(\langle w, x_i \rangle + b) < 1$. Pada dua kasus terakhir di mana $a_i^* > 0$, contoh pelatihan ke- i disebut sebagai vektor pendukung.

Dalam banyak aplikasi, sebagian besar a_i^* dalam solusi optimal adalah 0, yang memberikan solusi yang jarang. Karena pengklasifikasi akhir hanya bergantung pada vektor-vektor pendukung tersebut, maka seluruh algoritme ini dinamakan mesin vektor pendukung (*support vector machines*).

Dari masalah ganda (Persamaan (2.24)), pengamatan utama yang dapat ditarik adalah bahwa ciri dari contoh pelatihan $\{x_i\}$ mempengaruhi pelatihan hanya melalui *inner product* $\langle x_i, x_j \rangle$. Oleh karena itu, kita dapat mendefinisikan ulang ciri dengan memetakan x_i ke ruang ciri yang lebih kaya melalui $\phi(x_i)$ dan kemudian menghitung inner product di sana ($x_i, x_j := \langle \phi(x_i), \phi(x_j) \rangle$). Lebih jauh lagi, kita bahkan dapat secara langsung mendefinisikan k tanpa secara eksplisit menentukan ϕ . Hal ini memungkinkan kita untuk (a) secara implisit menggunakan ruang ciri yang kaya yang dimensinya bisa sangat tinggi, dan (b) menerapkan SVM pada ruang non-Euclidean selama kernel $k(x_i, x_j)$ dapat didefinisikan di atasnya. Secara matematis, tujuan (Persamaan (2.25)) dapat di-kernel-kan menjadi:

$$\frac{1}{2\lambda} \sum_{i,j} y_i y_j a_i a_j k \langle x_i, x_j \rangle - \sum_i a_i, \quad (2.25)$$

$$s. t. a_i \in [0, n^{-1}] \text{ and } \sum_i y_i a_i = 0$$

Namun, sekarang w tidak dapat diekspresikan hanya dalam bentuk kernel karena $w^* = \sum_{i=1}^n a_i^* y_i \phi(x_i)$. Untungnya, ketika memprediksi pada contoh baru x kita hanya membutuhkan inner inner product dan karenanya hanya menggunakan kernel saja (didefinisikan pada Persamaan (2.26)).

$$\langle w^*, x \rangle = \sum_{i=1}^n a_i^* y_i \langle \phi(x_i), \phi(x) \rangle = \sum_{i=1}^n a_i^* y_i k(x_i, x) \quad (2.26)$$

Kernel yang umum digunakan pada $\mathbb{R}^n$ termasuk kernel polynomial $(1 + \langle x_i, x_j \rangle)^d$, kernel RBF Gaussian $\exp(-\gamma \|x_i - x_j\|^2)$, kernel RBF Laplace $(-\gamma \|x_i - x_j\|)$, dll. Kernel pada string dan pohon biasanya didasarkan pada konvolusi, yang membutuhkan algoritma yang terlibat untuk evaluasi yang efisien.

SVM terbukti kuat akan tetapi model ini hanya terbatas pada dua kelas saja sehingga menjadi perhatian serius bagi para peneliti belakangan ini mengembangkan SVM pada masalah klasifikasi multi-kelas (Ali & Mallaiah, 2021). Dalam penelitian ini strategi *winner-takes-all* (WTA) akan diterapkan untuk mengatasi klasifikasi multi-kelas, penggunaan startegi ini termotivasi dari (Zhang & Wu, 2012).

Strategi WTA pada SVM untuk klasifikasi multi-kelas melibatkan penggunaan beberapa model SVM biner. Setiap model SVM dilatih untuk membedakan satu kelas tertentu dari semua kelas lainnya. Secara matematis, ini dapat dirumuskan sebagai berikut:

Misalkan kita memiliki data latih (x_i, y_i) di mana $x_i \in \mathbb{R}^n$ adalah vektor ciri $y_i \in \{1, 2, \dots, K\}$ adalah label kelas dengan K sebagai jumlah total kelas. Untuk setiap kelas k (dengan $k = 1, 2, \dots, K$), kita melatih model SVM biner SVM_k yang memisahkan kelas k dari semua kelas lainnya. Model SVM biner akan mencari hyperplane yang memaksimalkan margin antara kelas k dan kelas lainnya. Fungsi keputusan untuk model SVM_k dapat didefenisikan dengan Persamaan (2.27)

$$f_k(x) = w_k x + b_k \quad (2.27)$$

di mana w_k adalah vektor bobot dan b_k bias.

Selama model dilatih, model SVM_k memecahkan masalah optimasi yang didefinisikan pada Persamaan (2.28).

$$\begin{aligned} \min_{\mathbf{w}_k, b_k, \xi_i} \quad & \frac{1}{2} \|\mathbf{w}_k\|^2 + C \sum_{i=1}^m \xi_i \\ \text{dengan syarat} \quad & y_i(\mathbf{w}_k^\top \mathbf{x}_i + b_k) \geq 1 - \xi_i, \quad \forall i \\ & \xi_i \geq 0, \quad \forall i \end{aligned} \quad (2.28)$$

Di mana ξ_i adalah variabel slack yang memungkinkan pelanggaran margin, dan C adalah parameter regularisasi yang mengontrol trade-off antara margin maksimal dan pelanggaran margin.

Setelah model dilatih, untuk melakukan klasifikasi data baru x , selanjutnya menghitung nilai fungsi keputusan untuk semua model SVM_k seperti didefinisikan pada Persamaan (2.29).

$$\hat{y} = \arg \max_{k \in \{1, 2, \dots, K\}} f_k(x) \quad (2.29)$$

di mana kelas k yang menghasilkan nilai fungsi keputusan tertinggi dipilih sebagai kelas prediksi.

Penelitian yang dilakukan oleh Ali dan Mallaiah (2021) telah mengkonfirmasi penerapan dari strategi WTA efektif karena memungkinkan penggunaan SVM biner yang sudah ada tanpa perubahan besar. Namun, penting untuk diingat bahwa setiap model SVM_k harus dilatih secara terpisah, sehingga jumlah total model yang dilatih adalah K . Selain itu, skor yang dihasilkan oleh

setiap model harus dibandingkan untuk menentukan prediksi akhir, yang dapat meningkatkan kompleksitas komputasi.

Strategi WTA memberikan kerangka kerja yang sederhana namun kuat untuk memperluas SVM ke masalah klasifikasi multi-kelas, memungkinkan penerapan SVM dalam berbagai domain aplikasi yang memerlukan klasifikasi lebih dari dua kelas.

2. JST

Jaringan saraf tiruan (JST) merupakan sebuah model komputasi yang terinspirasi oleh struktur dan fungsi jaringan saraf biologis dalam otak manusia (Savchenko dkk., 2022). JST memiliki kemampuan untuk memodelkan dan mengekstraksi pola dari data, serta melakukan tugas-tugas kompleks seperti klasifikasi, regresi, dan prediksi. Struktur dari algoritma ini terdiri dari tiga bagian utama, yaitu neuron, bobot, dan fungsi aktivasi.

Tahap pertama adalah tahap feedforward, di mana data input disampaikan melalui jaringan untuk menghasilkan output. Setiap neuron pada hidden layer menghitung nilai inputnya dengan menggunakan bobot dan bias, diwakili oleh Persamaan (2.30).

$$z_j = \sum_{i=1}^n w_{ij} * x_i + b_j \quad (2.30)$$

di mana w_{ij} adalah bobot antara neuron i dan neuron hidden j , x_i adalah input dari neuron i , sedangkan b_j adalah bias neuron j .

Selanjutnya, fungsi aktivasi diterapkan pada hasil tersebut, diwakili oleh Persamaan (2.31).

$$a_i = f(z_i), \quad (2.31)$$

di mana $f(.)$ adalah fungsi aktivasi seperti Sigmoid.

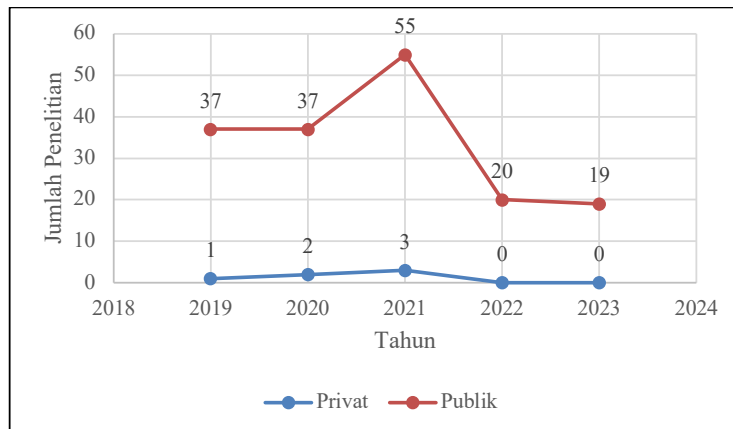
Tahap berikutnya adalah propagasi ke depan pada layer output dengan Persamaan (2.32).

$$y_k = \sum_{j=1}^m v_{jk} * a_j + c_k \quad (2.32)$$

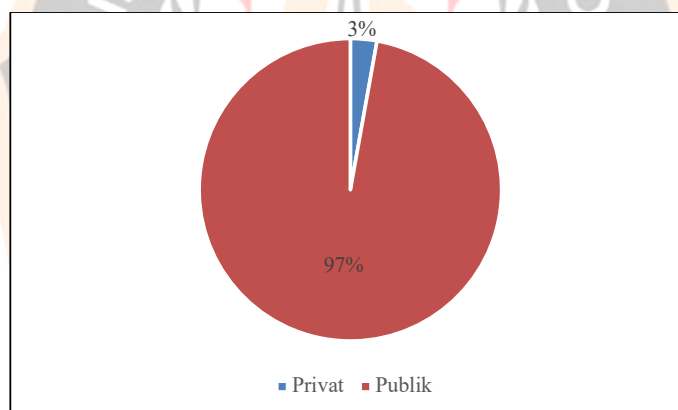
di mana v_{jk} adalah bobot antara neuron hidden j dan neuron output k , a_j adalah output dari neuron hidden j , dan c_k adalah bias untuk neuron output k . Setelah itu, fungsi aktivasi pada layer output diaplikasikan, diwakili oleh rumus $\hat{y}_k = g(y_k)$, dengan $g(.)$ sebagai fungsi aktivasi pada layer output.

2.3.8 Basisdata yang Digunakan dalam Menguji Sistem Pengenalan Ekspresi Wajah

Salah satu masalah yang paling penting dalam studi pengenalan ekspresi wajah adalah penggunaan basisdata non-publik (Canal dkk., 2022; Fang dkk., 2023). Beberapa peneliti mengembangkan metode ekstraksi ciri dan model klasifikasi pengenalan ekspresi wajah menggunakan data milik sendiri termasuk yang dilakukan oleh (Happy dkk., 2021; Jain dkk., 2019; J. Li dkk., 2020; Perveen dkk., 2020; Tadalagi dan Joshi, 2021). Namun, dalam penelitian ini tidak mungkin membandingkan hasil studi tersebut dengan hasil model yang diusulkan, karena basisdata mereka tidak dapat dinilai dan mahal. Seperti yang dilustrasikan pada Gambar 2.15, para peneliti pembelajaran mesin dan pembelajaran mendalam sejak tahun 2019 telah memiliki kesadaran akan pentingnya menggunakan repositori basisdata sumber terbuka (*open access*) karena dapat diamati dan divalidasi secara langsung tanpa ada syarat yang membebaskan peneliti. Total persentase basisdata privat dan publik ditunjukkan pada Gambar 2.16.



Gambar 2.15 Distribusi Basisdata Privat dan Publik



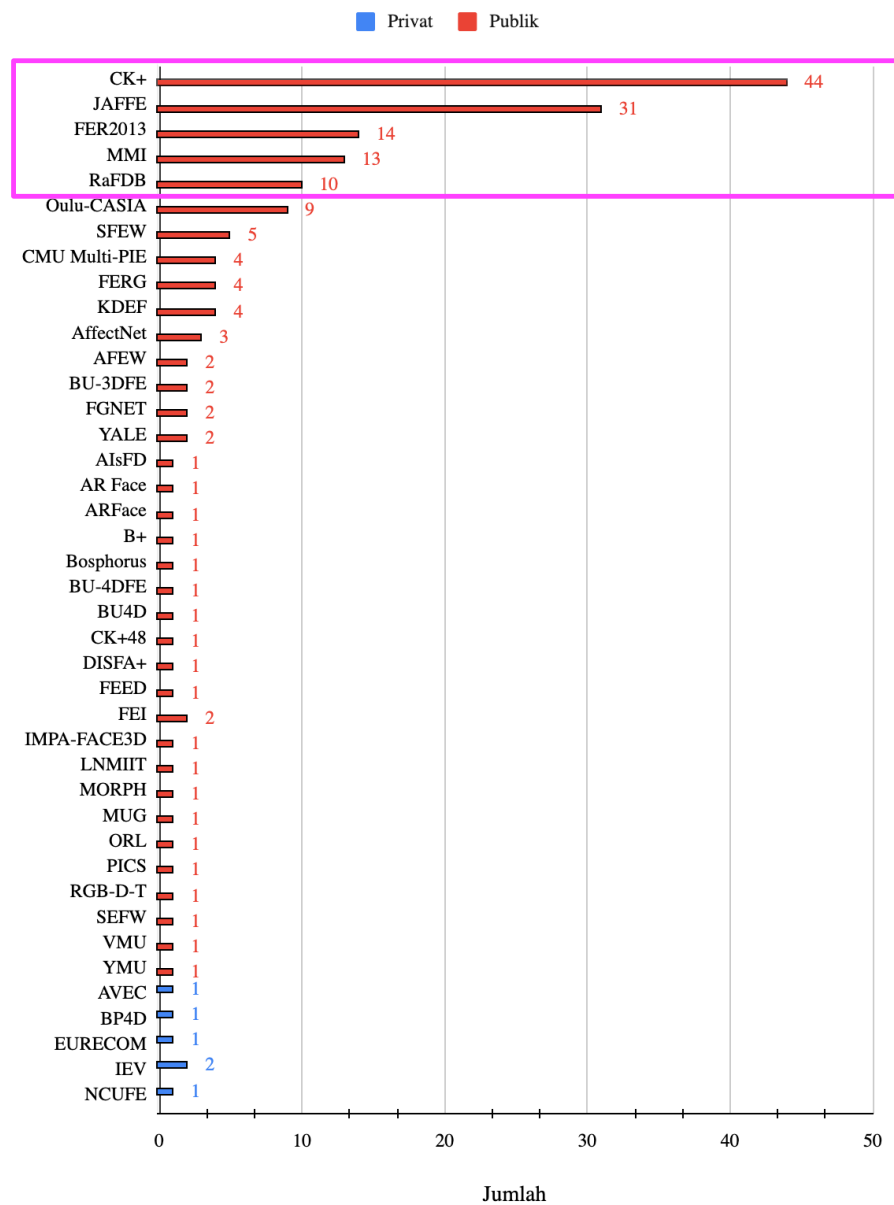
Gambar 2.16 Total Distribusi Basisdata

Seperti yang ditunjukkan pada Gambar 2.15, terlihat perbedaan jumlah yang signifikan di mana sebagian besar penelitian sejak tahun 2019-2023 total persentase penggunaan basisdata publik mendominasi sebesar 97% dibandingkan total persentase penggunaan data privat hanya 3%. Dengan demikian, tidak mungkin untuk membandingkan hasil penelitian tersebut dengan hasil model yang diusulkan, karena basisdata mereka tidak dapat digunakan. Hal ini menjelaskan mengapa penelitian-penelitian tersebut tidak dapat disanggah, diverifikasi, dan diulang.

Berbagai basisdata yang digunakan selama lima tahun terakhir dan basisdata yang paling berpengaruh dalam menguji dan mengevaluasi metode ekstraksi ciri dan model klasifikasi pengenalan ekspresi wajah ditunjukkan pada Gambar 2.17 dan Tabel 2.4.



SEKOLAH PASCASARJANA



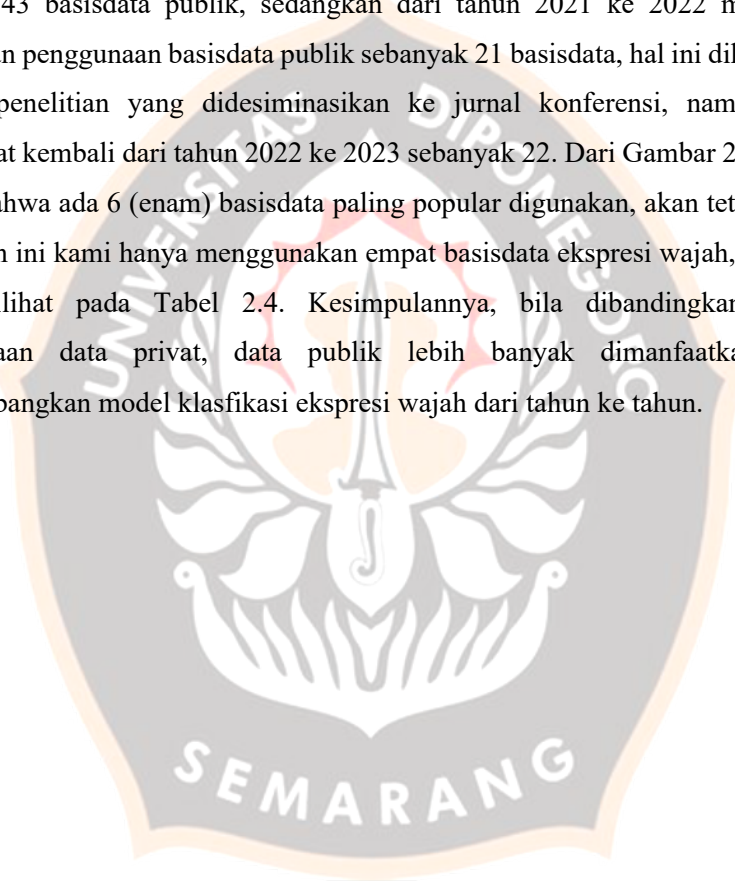
Gambar 2.17 Basisdata yang Berpengaruh dalam Pengembangan Model Sistem Pengenalan ekspresi wajah

Tabel 2.4 Basisdata Paling Berpengaruh dalam Studi Pengembangan Model Sistem Pengenalan ekspresi wajah

BASISDATA	SAMPEL	SUBJEK	PENGAMBILAN DATA*	EKSPRESI	LINK
CK+	593 citra wajah	123	Lab	6 ekspresi dasar ditambah netral	https://www.kaggle.com/datasets/shuvoalok/ck-dataset
JAFFE	213 citra wajah	10	Lab	6 ekspresi dasar ditambah netral	http://www.kasrl.org/jaffe.html
FER2013	35,887 citra wajah	Tidak Diketahui	Lab	6 ekspresi dasar ditambah netral	https://www.kaggle.com/datasets/rkuo2000/Fer2013
RaFDB	29,672 citra wajah	67	Lab	6 ekspresi dasar ditambah netral	http://www.whdeng.cn/raf/model1.html

Hasil ini mengkonfirmasi dari hasil penelitian tinjauan pustaka yang dipublikasikan oleh (Canal dkk., 2022). Mereka mengulas 51 artikel ilmiah tentang pengenalan ekspresi wajah dari berbagai jurnal dan konferensi. Menurut Canal dkk. (2022), basisdata dikategorikan ke dalam empat jenis: 1) pribadi, 2) parsial, 3) publik, dan 4) tidak diketahui. Basisdata publik diunggah di berbagai repositori dan didistribusikan secara bebas. Basisdata privat adalah milik perusahaan swasta dan tidak didistribusikan sebagai basisdata publik. Basisdata parsial adalah basisdata yang dibuat dengan menggunakan data dari proyek sumber terbuka dan belum didistribusikan ke komunitas. Basisdata parsial ini tidak disimpan di repositori publik. Jika tidak ada informasi tentang basisdata, jenis basisdata disebut tidak diketahui. Berbeda dari temuan kami semua basisdata disebutkan sumbernya, hal ini juga yang membedakan penelitian kami yang hanya menyaring artikel dari jurnal, sedangkan penelitian lainnya menggunakan keduanya, jurnal dan prosiding

konferensi. Terlepas dari perbedaan ini, hampir semuanya memiliki fakta yang serupa bahwa hanya satu hasil penelitian dari penelitian lain yang dapat dibandingkan dan dianalisis, di mana 172 (97%) basisdata publik, 5 (3%) basisdata privat yang ada dalam penelitian. Distribusi basisdata setelah tahun 2019 ke 2021 menunjukkan bahwa penggunaan basisdata publik meningkat dari 6 basisdata menjadi 43 basisdata publik, sedangkan dari tahun 2021 ke 2022 mengalami penurunan penggunaan basisdata publik sebanyak 21 basisdata, hal ini dikarenakan banyak penelitian yang didesiminasikan ke jurnal konferensi, namun mulai meningkat kembali dari tahun 2022 ke 2023 sebanyak 22. Dari Gambar 2.16, dapat dilihat bahwa ada 6 (enam) basisdata paling populer digunakan, akan tetapi dalam penelitian ini kami hanya menggunakan empat basisdata ekspresi wajah, detailnya dapat dilihat pada Tabel 2.4. Kesimpulannya, bila dibandingkan dengan penggunaan data privat, data publik lebih banyak dimanfaatkan untuk mengembangkan model klasifikasi ekspresi wajah dari tahun ke tahun.



SEKOLAH PASCASARJANA