

# Sistem Pengenalan Ekspresi Wajah untuk Deteksi Stres dengan Deep Learning

Alexander Louis, Alvin Ignacio, Mario Christopher

*Informatika, Universitas Kristen Petra, Siwalankerto No.121-131, Surabaya, Jawa Timur*

<sup>1</sup>C14210135@john.petra.ac.id

<sup>2</sup>C14210154@john.petra.ac.id

<sup>3</sup>C14210156@john.petra.ac.id

**Abstract**— Stres adalah masalah utama di tempat kerja, dengan hampir satu dari tiga pekerja di Indonesia mengalaminya. Pekerjaan yang menuntut, deadline singkat, tugas berat, dan situasi berbahaya adalah beberapa penyebabnya. Oleh karena itu, dibutuhkan cara yang efektif untuk mengontrol stres. Salah satu pendekatan inovatif adalah dengan membuat alat pendeteksi stres melalui deteksi wajah. Alat ini dapat memonitor secara terus menerus tingkat stres pekerja dan membantu menciptakan lingkungan kerja yang lebih produktif dan sehat. Penelitian ini menggunakan transfer learning dengan beberapa model deep learning, yaitu VGG16, VGG19, dan ResNet50V2, serta fine tuning untuk mengenali ekspresi wajah manusia. Ketiga model ini menunjukkan hasil yang cukup baik untuk trainingnya, dengan akurasi training rata-rata sebesar 85%. Model dengan akurasi terbaik adalah VGG16, yang mencapai 92% dan validasinya sebesar 53% dalam 350 epoch dengan 31 step per epoch. Hasil penelitian ini menunjukkan potensi besar dalam penerapan teknologi pengenalan wajah untuk deteksi dan manajemen stres di tempat kerja.

**Keywords**— stres, pengenalan ekspresi wajah, deep learning, transfer learning, fine-tuning

## I. INTRODUCTION

Stres di tempat kerja memang menjadi masalah serius, terutama dengan tingkat prevalensi yang tinggi di Indonesia. Pekerjaan yang menuntut, tenggat waktu yang ketat, beban kerja yang berat, dan situasi-situasi berbahaya dapat menjadi penyebab utama stres bagi pekerja. Dampaknya tidak hanya pada kesejahteraan individu, tetapi juga dapat mempengaruhi produktivitas dan kepuasan kerja secara keseluruhan. Stres, depresi, dan kecemasan terkait pekerjaan dapat mengakibatkan penurunan performa dan kehadiran pekerja dan menelan biaya berkisaran 3% sampai 4% dari pendapatan kotor produk nasional [1]. Terdapat sebuah bukti bahwa stres adalah kondisi yang dapat dicegah dan diobati pada tempat kerja dan pekerja yang menerima perawatan akan memiliki kemungkinan untuk lebih produktif [2]. Oleh karena itu, sebuah alat pendeteksi stress yang tidak mengganggu kehidupan sehari-hari pekerja dan dapat memonitor secara terus menerus tingkat stres mereka dapat digunakan untuk intervensi pengurangan stres. Pada situasi pekerjaan yang melelahkan dan membuat stres, aplikasi ini bukan saja dapat memberi intervensi, tetapi juga dapat menciptakan lingkungan kerja yang lebih produktif dimana pekerja dapat mengatur workload mereka.

Pada penelitian ini, digunakan metode transfer learning, fine tuning untuk tiga model yang digunakan, yaitu VGG16,

VGG19, dan ResNet50V2. Model VGG16, VGG19, dan ResNet50V2 dipilih karena memiliki arsitektur yang cenderung simpel dan sering digunakan untuk transfer learning, dan ketiga model tersebut terbukti handal dalam pelatihan image. Oleh karena itu, ketiga model tersebut digunakan dalam penelitian ini.

## II. STATE-OF-THE-ART

Terdapat beberapa penelitian yang telah berusaha menyelesaikan masalah deteksi stres melalui pengenalan ekspresi wajah. Salah satu contohnya adalah penelitian yang dilakukan oleh Almeida dan Rodrigues di tahun 2021, dimana penelitian ini mengembangkan aplikasi sistem pengenalan ekspresi wajah untuk deteksi stres dengan menggunakan model-model deep learning seperti VGG16, VGG19, dan Inception-ResNet V2 [3]. Penelitian ini memanfaatkan tujuh emosi dasar untuk mendeteksi stres, yaitu Anger, Disgust, Fear, Happiness, Sadness, Surprise, dan Neutral, dengan fokus klasifikasi stres pada emosi Anger, Disgust, dan Fear. Dataset yang digunakan meliputi KDEF, CK+, dan dataset peneliti, yang digabungkan menjadi satu dengan total 700 gambar per emosi. Hasil penelitian menunjukkan bahwa model VGG16 memberikan akurasi terbaik dengan 85% untuk klasifikasi semua emosi dan 92% untuk klasifikasi stres. Namun, penelitian ini memiliki keterbatasan pada ukuran dataset yang kecil dan homogenitas dataset.

Penelitian lain yang juga berkontribusi dalam bidang ini adalah penelitian oleh Chandika et al. tahun 2024, yang bertujuan untuk membuat aplikasi berbasis web yang dapat mendeteksi wajah pengguna secara real-time dan memvisualisasikan laporan analisis stres yang dihasilkan dari emosi yang terdeteksi dari waktu ke waktu [4]. Penelitian ini menggunakan model convolutional neural network (CNN) yang dilatih dengan dataset FER-2013, yang terdiri dari 21.000 gambar dengan berbagai emosi seperti Bursted, Irritated, Anxious, Relaxed, Neutral, Broked, dan Shocked. Model CNN ini menunjukkan akurasi 98% dalam mendeteksi emosi dalam dataset tersebut. Namun, penelitian ini menemukan bahwa model yang digunakan masih kurang optimal dalam mendeteksi emosi yang lebih kompleks.

Selain itu, penelitian oleh Chew et al. di tahun 2022 mengusulkan metode yang ringan dan baik untuk mendeteksi stres menggunakan pengenalan ekspresi wajah dengan model Enhanced Stress Convolutional Neural Network (ESCNN) [5]. Model ESCNN ini memanfaatkan struktur MobileNet V2 sebagai model pre-train dan menggunakan dataset FER2013

dengan tujuh emosi untuk klasifikasi stres. Hasil penelitian menunjukkan bahwa model ESCNN mencapai akurasi 95% dalam mendeteksi stres dan 85,54% dalam pengenalan ekspresi wajah dalam 15 epoch, mengungguli ResNet50 dan model sekuensial lainnya. Kelebihan dari penelitian ini adalah berhasil mengembangkan model yang ringan dengan akurasi tinggi, meskipun masih memiliki kelemahan dalam akurasi deteksi emosi yang kurang optimal.

### III. PROPOSED METHOD

#### A. Dataset

Penelitian ini menggunakan 3 dataset berbeda dengan perbedaan jumlah gambar, atribut wajah, dan fitur wajah. Ketiga dataset tersebut adalah sebagai berikut:

##### 1) KDEF (Karolinska Directed Emotional Faces)

Dataset KDEF memiliki 2940 gambar (420/ekspresi), homogen (tidak pakai kacamata, tidak ada kumis, ras yang sama, dan range umur kecil, dll).

##### 2) CK+

Dataset CK+ yang diambil memiliki total sebesar 920 gambar dengan ukuran 48x48 piksel. Gambar bersifat heterogen (pakai kacamata, ada kumis, ras yang beda, dan range umur besar).

##### 3) FER-2013

FER-2013 adalah dataset yang berisi total 35.887 gambar wajah. Dataset ini heterogen (pakai kacamata, ada kumis, ras yang beda, dan range umur besar).

Ketiga dataset tersebut berisikan 6-7 emosi dan hanya dipilih 7 untuk emosi, yaitu Marah (anger), Jijik (disgust), Takut (fear), Bahagia (happy), Sedih (sad), Terkejut (surprised), Netral (neutral). Gambar akan dipre-process menjadi ukuran 224x224. Kemudian, dataset dibagi menjadi 80% training dan 20% testing.

#### B. Model

Pada penelitian ini, digunakan 3 model deep learning yaitu VGG16, VGG19, dan ResNet50V2. Ketiga model tersebut digunakan dengan mengambil model transfer learning yang kemudian difine-tuning.

##### 1) Transfer Learning

Transfer learning menggunakan pengetahuan yang dipelajari dari menyelesaikan masalah sebelumnya sebagai titik awal untuk menyelesaikan masalah lain, dimana ini memungkinkan penyelesaian masalah dengan menggunakan source lebih sedikit daripada sebelumnya [3]. Model transfer learning yang digunakan untuk penelitian ini berasal dari library TensorFlow keras application, dengan mengambil pre-trained model VGG16, VGG19, dan ResNetV2 yang sudah ada. Untuk pengubahan yang ada yaitu di dense layer, yang diubah menjadi 7 class saja.

##### 2) VGG16

VGG16 merupakan model berbasis CNN (Convolutional Neural Network). CNN bekerja dengan memproses gambar melalui serangkaian lapisan konvolusi untuk mengekstraksi fitur-fitur penting, diikuti oleh lapisan pooling untuk pengurangan dimensi, dan fully connected layer untuk klasifikasi atau prediksi. VGG16 yang digunakan memiliki total sebanyak 19 layer, dimana terdiri dari 13 lapisan konvolusi, 5 lapisan pooling, dan 3 lapisan fully connected, dengan total parameter sebesar 14,714,688. Studi menunjukkan bahwa model VGG16 sering kali memberikan akurasi yang lebih baik dibandingkan VGG19 dan ResNet50 dalam klasifikasi gambar [6]. Di sisi lain, VGG16 sendiri memiliki kelemahan, dimana model ini rawan akan terjadinya overfitting saat training, sehingga dibutuhkan dataset yang beragam dan cukup besar untuk train model VGG16.

##### 3) VGG19

VGG19 memiliki prinsip yang sama dengan VGG16, namun VGG19 memiliki lebih banyak lapisan, dengan total 22 lapisan yaitu 16 lapisan konvolusi, 5 lapisan pooling, dan 1 lapisan softmax, dengan total parameter sebesar 20,024,384. Tambahan lapisan ini memungkinkan VGG19 untuk mengekstraksi fitur yang lebih kompleks [7].

##### 4) ResNet50V2

ResNet50V2 adalah versi modifikasi dari ResNet50 yang memiliki performa lebih baik dibandingkan ResNet50. Pada ResNet50V2, dilakukan modifikasi dalam formulasi propagasi koneksi antar blok [8]. ResNet50V2 yang digunakan memiliki jumlah layer sebesar 192 secara keseluruhan. ResNet50V2 didesain untuk mengurangi masalah vanishing gradient, dimana hal ini membuat model ini cocok untuk deteksi fitur-fitur kompleks dari gambar-gambar yang lebih besar dan beragam.

##### 5) Data Augmentation

Data augmentation dilakukan, agar dataset yang digunakan memiliki variasi yang lebih banyak. Hal ini dilakukan untuk mengurangi resiko terjadinya overfitting pada saat training model. Augmentasi yang dilakukan pada penelitian ini yaitu:

- Rotasi hingga 20 derajat
- Translasi 10% dan 15% untuk lebar dan tinggi (masing-masing)
- Perubahan kecerahan antara 0,2 dan 1
- Zoom-out hingga 10% dan zoom-in hingga 20%
- Flip horizontal

Augmentasi yang dilakukan hanya untuk data training, tidak untuk gambar yang akan diuji coba.

##### 6) Fine-Tuning

Untuk proses fine-tuning, penting menentukan seberapa banyak bagian dasar konvolusional yang akan dilatih, dengan mempertimbangkan kesamaan antara domain target dan sumber serta ukuran dataset target. Karena domain tidak mencakup ekspresi wajah manusia dan dataset pelatihan kecil, sebagian besar lapisan jaringan pre-trained perlu

di-fine-tune [3]. Fine tuning yang digunakan pada penelitian ini adalah berupa freeze layer untuk ketiga model yang digunakan. Freeze layer yang dilakukan di penelitian digunakan untuk membuat tiap model dapat mewarisi pengetahuan model untuk mendeteksi secara general. Jumlah layer yang difreeze untuk tiap model berbeda bergantung dengan jumlah total layer-nya. VGG16 menggunakan layer ke-11 (stage 5 convolution), VGG19 menggunakan layer ke-12 (stage 5 convolution), dan ResNet50V2 menggunakan layer ke-170 (stage 5 convolution) sebagai pembatas freeze layer-nya.

#### IV. EXPERIMENTATION

Penelitian ini melakukan eksperimen dengan menggunakan 3 model deep learning yaitu VGG16, VGG19, dan ResNet50V2. Menggunakan 3 dataset yang diambil dari kaggle, yaitu KDEF, CK+, dan FER2013 dengan total dataset sebesar 39,798 gambar ekspresi wajah manusia. Dari ketiga dataset, semua gambar disatukan dan diacak (shuffle), kemudian hanya diambil 50% dari seluruh data, sehingga hanya ada sekitar 19,899 total gambar ekspresi wajah manusia. Pembagian 80/20 dilakukan menghasilkan, jumlah data training sebanyak 15,923 gambar dan testing sebanyak 3,976 gambar. Hal ini dilakukan karena keterbatasan kemampuan dari RAM device peneliti.

Setelah melakukan pembagian data, fine-tuning dilakukan, dengan pembagian jumlah parameter untuk tiap modelnya sebagai berikut:

TABLE I  
LAYER FREEZE > (PEMBATAS FREEZE LAYER MODEL)

	VGG16	VGG19	ResNet50V2
Trainable Parameter	2,915,648 (11.12 MB)	3,505,728 (13.37 MB)	15,646,687 (59.69 MB)
Non Trainable Parameter	11,799,040 (45.01 MB)	16,518,656 (63.01 MB)	9,974,120 (38.05 MB)

TABLE II  
LAYER FREEZE < (PEMBATAS FREEZE LAYER MODEL)

	VGG16	VGG19	ResNet50V2
Trainable Parameter	12,979,200 (49.51 MB)	17,698,816 (67.52 MB)	9,935,687 (37.90 MB)
Non Trainable Parameter	1,735,488 (6.62 MB)	2,325,568 (8.87 MB)	15,685,120 (59.83 MB)

TABLE III  
FREEZE LAYER

	VGG16	VGG19	ResNet50V2
--	-------	-------	------------

Trainable Parameter	57,351	57,351 (224.03 KB)	7,007 (27.37 KB)
Non Trainable Parameter	14,714,688 (56.13 MB)	20,024,384 (76.39 MB)	25,613,800 (97.71 MB)

Kemudian, untuk metrik evaluasinya, menggunakan akurasi dan loss yang didapat dari training dan testing tiap model yang sudah difine-tuning. Ketiga model menggunakan batch size yang sama, tetapi step per epoch yang berbeda, kemudian, tiap model juga menggunakan jumlah epoch yang berbeda, bergantung dari akurasi yang sudah didapat selama proses. Selain itu, juga membedakan klasifikasi untuk stress dan non-stress berdasarkan emosi sebagai berikut:

- Stress: marah (anger) dan takut (fear), jijik (disgust)
- Non-stress: senang (happy), sedih (sad), netral, dan kaget (surprised)

Berikut adalah tabel hasil dari percobaan yang dilakukan:

TABLE IV  
AKURASI TRAINING, LAYER FREEZE > (PEMBATAS FREEZE LAYER MODEL)

	VGG16	VGG19	ResNet50V2
Akurasi	0.8876	0.8750	0.2812
Loss	0.3343	0.4098	1.8096
Epoch	350 (31 steps per epoch)	100 (124 steps per epoch)	170 (62 steps per epoch)
Waktu/step	13-15s	12-14s	17-20s
Total Waktu	151,900s	161,200s	189,720s

TABLE V  
AKURASI TRAINING, LAYER FREEZE < (PEMBATAS FREEZE LAYER MODEL)

	VGG16	VGG19	ResNet50V2
Akurasi	0.9249	0.7999	0.9062
Loss	0.2083	0.5749	0.3852
Epoch	330 (31 steps per epoch)	87 (124 steps per epoch)	170 (62 steps per epoch)
Waktu/step	23-27s	35-42s	5-6s
Total Waktu	255,750s	409,944s	57,970s

TABLE VI  
AKURASI TRAINING, FREEZE LAYER

	VGG16	VGG19	ResNet50V2
--	-------	-------	------------

Akurasi	0.6895	0.6684	0.2656
Loss	0.8771	0.9425	1.6734
Epoch	350 (31 steps per epoch)	100 (124 steps per epoch)	170 (62 steps per epoch)
Waktu/step	7s	8s	4s
Total Waktu	75,950s	99,200s	42,160s

TABLE VII  
AKURASI DATA TESTING

	VGG16	VGG19	ResNet50 V2	VGG16	VGG19	ResNet50 V2	VGG16	VGG19	ResNet50 V2
	Layer freeze > (pembatas freeze layer model)			Layer freeze < (pembatas freeze layer model)			Freeze Layer		
<b>Akurasi</b>	0.5372	0.5371	0.2331	0.5372	0.5186	0.5451	0.5229	0.5142	0.3206
<b>Loss</b>	2.1680	2.1026	1.8503	2.1680	1.7201	1.5099	1.3171	1.3464	1.7286

TABLE VIII  
AKURASI TIAP EMOSI

	VGG16	VGG19	ResNet50 V2	VGG16	VGG19	ResNet50 V2	VGG16	VGG19	ResNet50 V2
	Layer freeze > (pembatas freeze layer model)			Layer freeze < (pembatas freeze layer model)			Freeze Layer		
<b>angry</b>	0.4088	0.2832	0.0000e+00	0.4649	0.1130	0.3193	0.2460	0.2901	0.0542
<b>disgust</b>	0.2043	0.0657	0.0000e+00	0.1783	0.0000e+00	0.0000e+00	0.0000e+00	0.1692	0.0000e+00
<b>fear</b>	0.2609	0.3114	0.0000e+00	0.2244	0.0965	0.3811	0.1692	0.1824	0.0652
<b>happy</b>	0.2609	0.6506	1.0000	0.7450	0.2146	0.8078	0.6322	0.7726	0.8942
<b>neutral</b>	0.3194	0.4421	0.0000e+00	0.4211	0.3833	0.5343	0.1176	0.2666	0.0945
<b>sad</b>	0.4399	0.2807	0.0000e+00	0.2834	0.5838	0.4442	0.5660	0.3738	0.1767
<b>surprise</b>	0.6068	0.5489	0.0000e+00	0.4575	0.2756	0.4367	0.1784	0.3962	0.1566
<b>stress</b>	0.3208	0.2832	0.0000e+00	0.3262	0.1060	0.3362	0.1870	0.2909	0.0599
<b>non-stress</b>	0.5049	0.5623	0.7059	0.5051	0.2731	0.7071	0.4143	0.6379	0.6679

Penelitian ini juga melakukan percobaan untuk beberapa foto testing yang diambil dari google image dan beberapa video cuplikan yang diambil dari youtube. Berikut adalah beberapa hasil prediksi dari percobaan yang telah dilakukan dengan menggunakan model terbaik, yaitu VGG16:



Tiga gambar paling kiri merupakan gambar yang diambil dari data test (dari gabungan ketiga dataset). Tiga gambar di tengah merupakan foto yang diambil dari google image, dan tiga gambar paling kanan yaitu 3 gambar dari frame video cuplikan film.

Untuk hasilnya sendiri, model yang digunakan masih belum bisa memprediksi gambar ekspresi wajah manusia dengan baik, dimana hanya sebagian kecil prediksi yang benar.

## V. CONCLUSION

Dalam paper ini kita sudah membuat sebuah model yang dapat mendeteksi facial expression dan mengklasifikasikan emosi stress dan non-stress. Untuk pembuatan model klasifikasi ini kita menggunakan model pre-train VGG16, VGG19, dan ResNet50V2.

Model VGG16 menjadi model dengan akurasi terbaik yaitu mencapai 92% dalam 350 epoch. Tetapi meskipun dapat mencapai akurasi rata-rata 85% untuk semua model, Validasi kita hanya mencapai rata-rata 50% akurasi. Hal ini dikarenakan terdapat permasalahan dalam dataset yang kita gunakan. Seperti contohnya ada dataset yang memiliki warna RGB sedangkan kebanyakan dataset yang lain hanya memiliki warna grayscale. Karena itu, sebagai pekerjaan di masa depan meningkatkan akurasi dari validation dan memperbaiki dataset menjadi pekerjaan yang sangat diperlukan.

## REFERENCES

- [1] C. S. Dewa and J. S. Hoch, "Barriers to mental health service use among workers with depression and work productivity," *Journal of Occupational and Environmental Medicine*, vol. 57, no. 7, pp. 726–731, Jul. 2015, doi: 10.1097/jom.0000000000000472.
- [2] S. Carolan, P. R. Harris, and K. Cavanagh, "Improving Employee Well-Being and Effectiveness: Systematic review and Meta-Analysis of Web-Based psychological interventions delivered in the workplace," *JMIR. Journal of Medical Internet Research/Journal of Medical Internet Research*, vol. 19, no. 7, p. e271, Jul. 2017, doi: 10.2196/jmir.7583.
- [3] J. Almeida dan F. Rodrigues, "Facial Expression Recognition System for Stress Detection with Deep Learning," dalam *Proc. 23rd Int. Conf. Enterprise Information Systems (ICEIS)*, SciTePress, 2021, vol. 1, pp. 256–263, doi: 10.5220/0010474202560263.
- [4] H. P. Chandika, B. Soumya, B. N. E. Reddy, and B. M. S. SaiManideep, "Real-Time stress Detection and analysis using facial emotion recognition," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 13, no. 3, Mar. 2024, doi: 10.17148/ijarccce.2024.13324.
- [5] W.-T. Chew, S.-C. Chong, T.-S. Ong, and L.-Y. Chong, "Facial Expression Recognition via Enhanced Stress Convolution Neural Network for Stress," *IAENG International Journal of Computer Science | EBSCOHost*, Sep. 01, 2022, [https://www.iaeng.org/IJCS/issues\\_v49/issue\\_3/IJCS\\_49\\_3\\_20.pdf](https://www.iaeng.org/IJCS/issues_v49/issue_3/IJCS_49_3_20.pdf)
- [6] K. Kamal and H. Ez-Zahraouy, "A comparison between the VGG16, VGG19 and ResNet50 architecture frameworks for classification of normal and CLAHE processed medical images," *Research Square (Research Square)*, Apr. 2023, doi: 10.21203/rs.3.rs-2863523/v1.
- [7] N. W. I. Kusumawati and N. A. Z. Noorizki, "Perbandingan performa algoritma VGG16 dan VGG19 melalui metode CNN untuk klasifikasi varietas beras," *Journal of Computer, Electronic, and Telecommunication*, vol. 4, no. 2, Dec. 2023, doi: 10.52435/complete.v4i2.387.
- [8] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," *Informatics in Medicine Unlocked*, vol. 19, p. 100360, Jan. 2020, doi: 10.1016/j.imu.2020.100360.