

DROWSINESS DETECTION

Dheandra Azzahra Triyasnanda^a, Rijal Akhdan Khairulah^b, Muhammad Hanif Nafis^c, Duaja Kukuh Adiwijaya^d

^a(16201213328) Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

^b(16201213330) Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

^c(16201213333) Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

^d(16201213358) Teknologi Sains Data, Fakultas Teknologi Maju dan Multidisiplin, Universitas Airlangga, Surabaya

Abstrak

Tidur adalah hal yang dibutuhkan oleh setiap manusia. Kebutuhan untuk tidur ditandai dengan munculnya rasa kantuk. Mengantuk merupakan sebuah keadaan yang manusiawi dan tidak dapat dihindari. Namun, rasa kantuk yang dialami dapat menjadi pengganggu, bahkan membahayakan, dalam situasi-situasi tertentu seperti saat pembelajaran, bekerja, hingga saat mengemudi kendaraan. Salah satu penyebab utama kecelakaan lalu lintas adalah rasa kantuk dan kelelahan dari pengendara. Oleh karena itu, dibuatlah sistem yang dapat mendeteksi rasa kantuk tersebut untuk mengantisipasi terjadinya kecelakaan yang disebabkan oleh pengendara yang merasakan kantuk. Pada penelitian ini, dilakukanlah klasifikasi gambar yang dapat mendeteksi apakah seseorang mengalami kantuk. Terdapat tiga metode yang digunakan, yaitu YOLOv5, CNN, dan EAR. Meskipun model YOLOv5 memiliki *recall* 1, namun begitu pula dengan *precision* dan nilai *f1*-nya, sehingga memiliki kemungkinan besar bersifat *overfit*. Disimpulkan bahwa model EAR yang memberikan nilai *recall*, yang penting pada kasus ini karena tujuannya adalah mendeteksi nilai positif sebanyak mungkin, yang tinggi yaitu 0,989 sebagai metode terbaik.

Kata Kunci : *Drowsiness Detection, YOLOv5, CNN, EAR*

1. Pendahuluan

Tidur adalah hal yang dibutuhkan oleh setiap manusia. Suatu penelitian menyimpulkan bahwa durasi rata-rata tidur cukup adalah 7-8 jam (Ferrara dan De Gennaro., 2001). Kebutuhan untuk tidur ditandai dengan munculnya rasa kantuk. Rasa kantuk adalah tahap pertama dari tiga tahap pada non-rapid eye movement sleep (NREM), yang kemudian diikuti oleh tidur ringan kemudian tidur lelap. (Sahayadhas et. al., 2012).

Mengantuk merupakan sebuah keadaan yang manusiawi dan tidak dapat dihindari. Mengantuk dapat dialami oleh semua orang tanpa terkecuali. Namun, terdapat beberapa situasi tertentu yang mampu merugikan seseorang apabila mengantuk pada saat yang tidak tepat, antara lain seperti pada saat mengemudikan kendaraan, mengikuti kelas perkuliahan, serta beberapa kegiatan lain yang memerlukan konsentrasi penuh.

Salah satu penyebab utama kecelakaan lalu lintas adalah rasa kantuk dan kelelahan dari pengendara. Kecelakaan lalu lintas selain membahayakan pengendara yang mengalami kecelakaan, juga dapat membahayakan hal-hal di sekitarnya seperti pengemudi lainnya, fasilitas jalan, serta pejalan kaki. Hal ini dapat dicegah apabila pengendara yang mengantuk diperingatkan tepat waktu. Oleh karena itu, dibuatlah sistem yang dapat mendeteksi rasa kantuk tersebut untuk mengantisipasi terjadinya kecelakaan yang disebabkan oleh pengendara yang merasakan kantuk (Ueno et al., 1994; Fujiwara et.al., 2018 Deng dan Wu, 2019). Penelitian ini bertujuan melakukan deteksi kantuk. Pendekstrian akan dilakukan

menggunakan 3 metode, metode pertama adalah menggunakan YOLOv5, metode kedua adalah menggunakan *convolutional neural network* (CNN), dan metode ketiga menggunakan *eye aspect ratio* (EAR). Lalu ketiga metode tersebut akan dibandingkan hasilnya.

2. Landasan Teori

2.1 YOLOv5 (*You Only Look Once v5*)

YOLOv5 (You Only Look Once) versi 5 merupakan framework yang didalamnya terdapat algoritma pendekripsi objek. YOLO menggunakan sebuah jaringan saraf tunggal (single neural network) untuk melakukan pendekripsi dan pengenalan objek yang memprediksi secara langsung bounding box dan probabilitas kelas. Untuk pengembangan You Only Look Once ini telah sampai di versi ke-5 yang memiliki akurasi lebih baik dari versi-versi sebelumnya, YOLOv5 dikembangkan oleh peneliti dan CEO dari Ultralytics LLC yaitu Glenn Jocher pada tahun 2020(Baihaqi & Zonyfar.,2022).YOLOv5 menggunakan framework PyTorch yang ditulis menggunakan bahasa pemrograman Python.YOLOv5 memberikan pilihan untuk menyesuaikan dengan hardware yang akan digunakan sehingga dapat berjalan dengan baik sesuai hardware yang telah tersedia. Dibanding versi sebelumnya kinerja YOLOv5 kurang mampu menandingi YOLOv4 akan tetapi jauh lebih fleksibel dan lebih cepat daripada versi YOLOv4 (Yang et al.,2021).

2.2 Convolutional Neural Network (CNN)

Artificial Neural Network (ANN) adalah sistem komputasi yang dirancang agar dapat bekerja menyerupai sistem saraf makhluk hidup, khususnya manusia. ANN terdiri dari banyak *node* yang saling berhubungan, disebut neuron, dimana proses pembelajaran dan optimasi dilakukan (O’Shea dan Nash, 2015). ANN memiliki berbagai macam model, salah satunya adalah *Convolutional Neural Network* (CNN) yang sering digunakan dalam pembelajaran yang berkaitan dengan pengenalan pola pada data gambar. Arsitektur CNN terdiri dari tiga jenis lapisan, yaitu lapisan *convolutional*, lapisan *pooling*, dan lapisan *fully-connected*. Lapisan *convolutional* menentukan hasil dan bobot dari suatu daerah tertentu pada gambar, yang kemudian dihubungkan dengan neuron pada lapisan selanjutnya (Albawi et.al., 2017). Hal ini mengurangi parameter yang terbentuk jika dibandingkan dengan ANN yang akan menghitung setiap bagian pada gambar sebelum meneruskan ke neuron pada lapisan selanjutnya. Lapisan *pooling* menurunkan dimensi gambar sehingga kompleksitasnya berkurang. Metode *pooling* yang umum digunakan adalah *max-pooling* dengan ukuran 2x2. Lapisan *fully-connected* menyambungkan tiap *node* pada lapisannya dengan lapisan sebelum dan lapisan setelahnya (Albawi et. al., 2017). Oleh karena banyaknya koneksi pada lapisan ini, lapisan *fully-connected* menghasilkan banyak parameter yang dapat digunakan, sehingga membutuhkan kemampuan komputasional yang besar.

2.3 Eye Aspect Ratio

Eye Aspect Ratio (EAR) merupakan sebuah nilai yang diperoleh dari mendekripsi wajah dari gambar, lalu dihitung jarak euclidean sesuai letak mata (Soukupova dan Cech, 2016). Dalam kalimat yang lebih mudah, EAR adalah skor untuk mengetahui seberapa terbuka mata manusia (Dewi et. al., 2022).

3. Sumber Data dan Metodologi

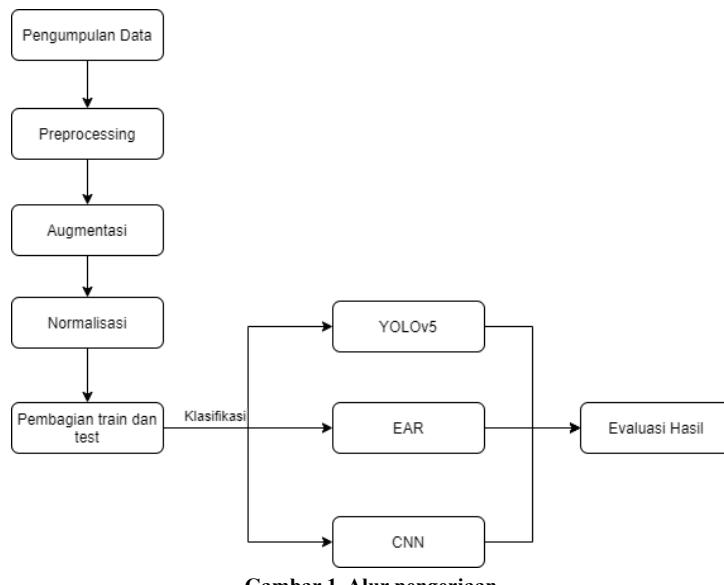
3.1 Sumber Data

Sumber data yang digunakan adalah data primer yang diambil secara pribadi, pada penelitian ini data gambar berasal dari sampel foto foto wajah dari mahasiswa dan mahasiswi jurusan Teknologi Sains Data Universitas Airlangga Surabaya. Data berupa gambar dengan format JPG. Dataset berisi 91 gambar berlabel “awake” dan 94 gambar berlabel “drowsy”. Jumlah total data yang digunakan dalam penelitian ini sebanyak 185 gambar.

3.2 Metodologi

Metode penelitian adalah cara ilmiah untuk mendapatkan data dengan tujuan dapat dideskripsikan, dibuktikan, dikembangkan dan ditemukan pengetahuan, teori, untuk memahami, memecahkan, dan mengantisipasi masalah dalam kehidupan manusia (Sugiyono: 2012). Ilustrasi alur penggerjaan disajikan dalam bentuk diagram alur pada gambar 1. Metode penelitian yang digunakan dalam laporan ini antara lain sebagai berikut.

1. Pengumpulan Dataset gambar
2. Pemberian label untuk tiap gambar
3. Preprocessing berupa cropping
4. Augmentasi untuk menambah variasi data
5. Normalisasi
Normalisasi yang dilakukan adalah dengan membagi nilai dengan 255 agar nilai akhir berkisar antara 0 hingga 1.
6. Pembagian *train* dan *test*
Pembagian data *train* dan *test* dilakukan secara acak dengan rasio data *train*:data *test* adalah 8:2.
7. Melakukan klasifikasi menggunakan tiga metode berbeda yaitu sebagai berikut.
 - a) YOLOv5
 - b) CNN
 - c) EAR
8. Evaluasi Hasil

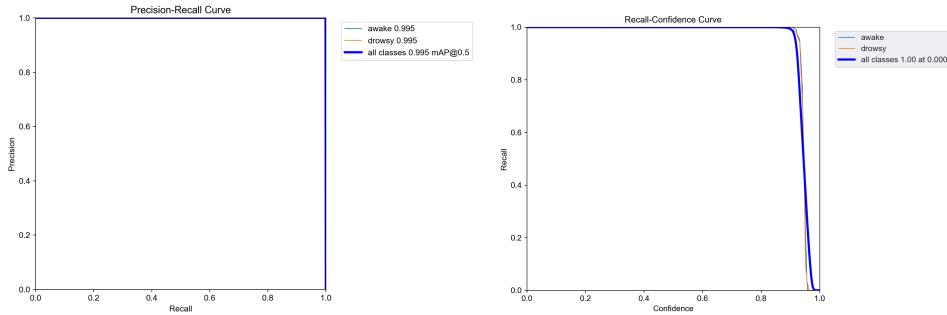


Gambar 1. Alur penggerjaan

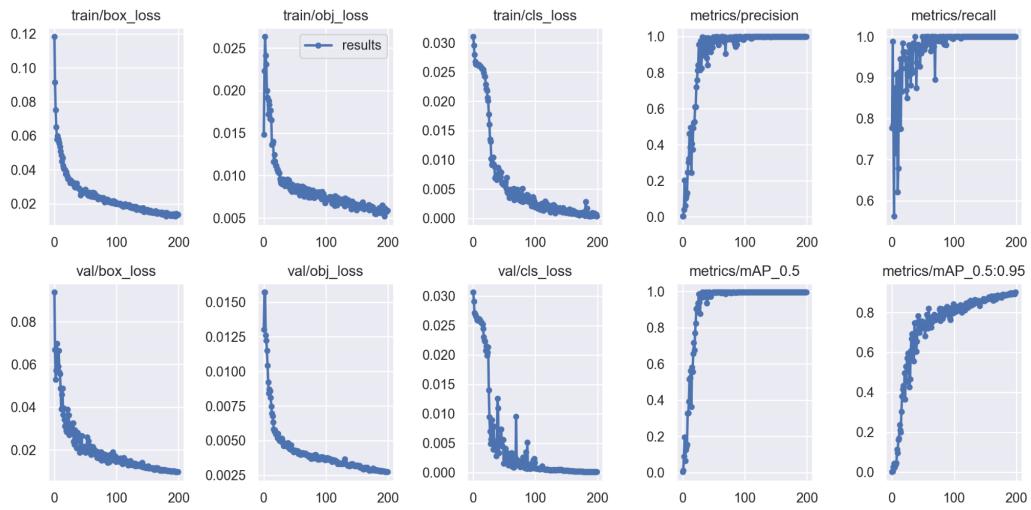
4. Analisis dan Pembahasan

4.1 Metode YOLOv5

Pengujian dilakukan menggunakan nilai batch dan nilai epoch sebesar 16 dan 200. Hasil training data peneliti menggunakan model deteksi objek YOLOv5 mendapatkan nilai yang cukup tinggi, dengan nilai precision mendapatkan rata-rata nilai 0.995 terhadap nilai recall. Nilai puncak rata-rata recall mendapatkan nilai 1,00 pada nilai confidence 0,00. Dibawah ini adalah nilai recall:

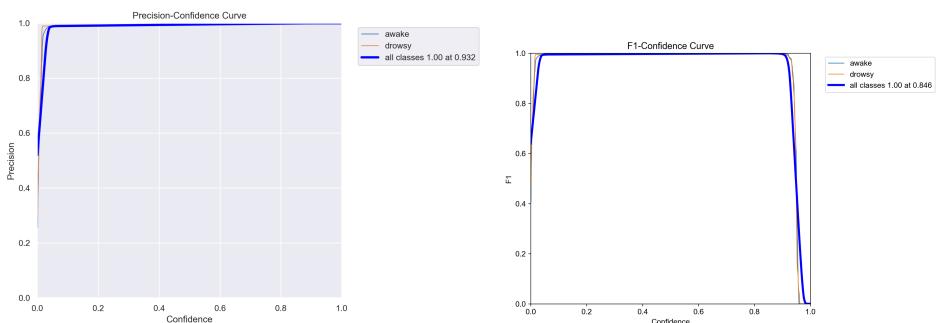


Gambar 2. Kurva nilai *precision-recall* dan *recall* terhadap nilai *confidence*



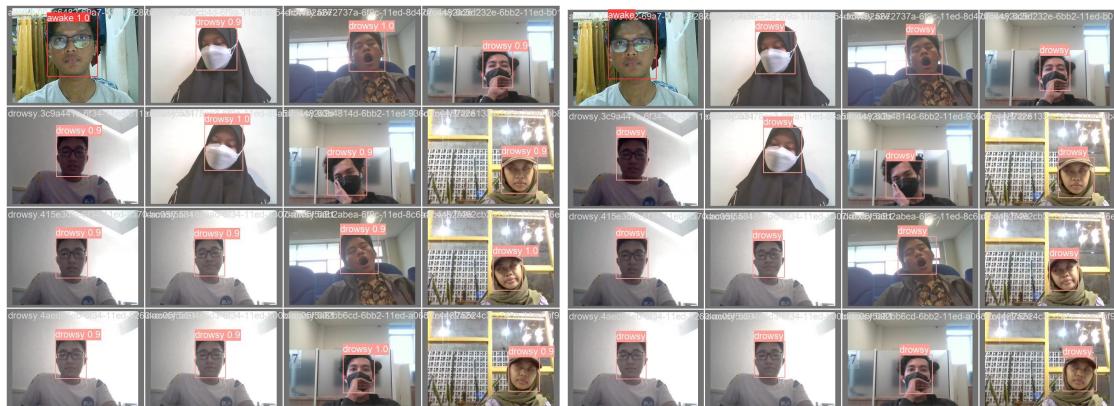
Gambar 3. Hasil evaluasi data training

Berdasarkan data yang telah didapat di atas, berkurang nilainya dari epoch pertama hingga epoch ke-200 dengan nilai 0.02, begitu pula nilai validation box loss yang dapat mencapai nilai kurang dari 0.02. Nilai Objectness loss dapat tereduksi dari nilai 0.025 hingga 0.005, begitu juga dengan validation objectness loss yang nilai tereduksi mendekati 0.025. Classification loss mendapatkan hasil yang cukup baik dengan nilai 0.00 dan nilai validation classification loss memiliki nilai yang baik dengan hasil 0.000. Nilai Precision mendapatkan nilai terbaik hingga 1.0 dan nilai Recall 1.0. Selain itu hasil dari mean Average Precision 0,5 sebesar 1.0 dan nilai mAP 0.95 mendapatkan hasil melebihi 0.8



Gambar 4. Kurva nilai F1 dan *precision* terhadap nilai *confidence*

Pada gambar diatas nilai F1 mendapatkan nilai puncak rata-rata 1,00 terhadap nilai confidence 0,846. Nilai precision mendapatkan nilai rata-rata 1,00 pada nilai confidence 0,932 yang berarti dapat memprediksi data yang akan dideteksi dengan baik melalui data yang telah dilatih. Untuk mengevaluasi model deteksi objek YOLO ini nilai *mean average precision* (mAP) digunakan untuk melihat seberapa baik sebuah model. mAP membandingkan kotak pembatas ground-truth dengan kotak yang terdeteksi dan mengembalikan skor. Semakin tinggi skor, semakin akurat model dalam pendeksiannya.Untuk model pada penelitian ini nilai mAP yang didapat sebesar 0,90098 atau 90%.



Gambar 5. Pendeksiyan kantuk menggunakan YOLOv5

Kebaikan model dipengaruhi dari berbagai macam hal diantaranya kualitas kamera, kualitas dataset, dan pengambilan gambar pada berbagai sudut. Metode YOLOv5 ini membutuhkan kamera atau kualitas gambar yang bagus untuk dapat mendeksi wajah. Dalam hal klasifikasi kantuk metode ini dapat terbilang sangat bagus dalam pengklasifikasianya dimana dengan melihat nilai confidence yang ditampilkan di sebelah label klasifikasi menunjukkan keyakinan model ini dalam mengklasifikasikan gambar tersebut. Selain kualitas kamera, perlunya dataset training gambar yang jelas dan mempunyai banyak sudut pandang agar model ini dapat bekerja dengan baik dalam hal pendeksiyan kantuk

4.2 Metode CNN

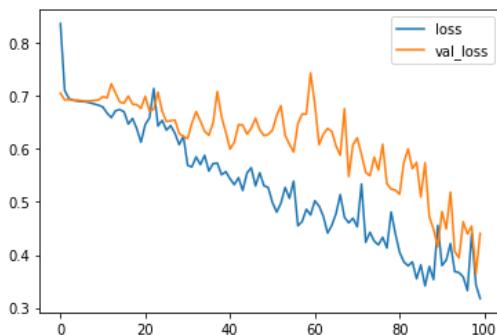
CNN sering digunakan dalam klasifikasi gambar (sumber1;sumber2;sumber3). Penggunaan CNN dalam klasifikasi gambar dikarenakan kemampuannya dalam mendeksi fitur-fitur spasial. CNN yang digunakan memiliki 2 lapisan *convolution* dan 2 lapisan *pooling*, kemudian lapisan *flatten* dan 3 lapisan *fully connected* di akhir. Gambaran arsitektur CNN secara jelas dapat dilihat pada tabel 1 dibawah.

Layer	Output Shape	Parameter
Convolution 2D	(None,64,64,1)	160
Max Pooling 2D	(None,32,32,16)	0
Convolution 2D	(None,32,32,32)	4640
Max Pooling 2D	(None,16,16,32)	0
Flatten	(None,8192)	0

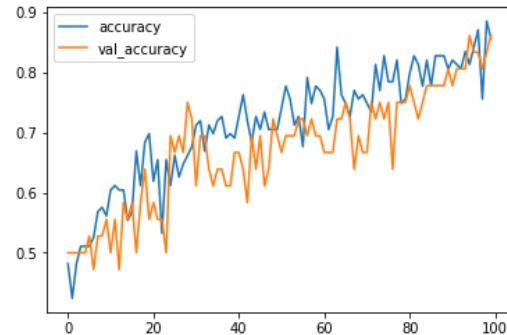
Dense	(None,512)	4194816
Dense	(None,108)	55404
Dense	(None,1)	109

Tabel 1. Lapisan CNN

Arsitektur CNN tersebut kemudian dilatih dengan menggunakan *batch size* 8 dan *epoch* 100. Jika dibandingkan *epoch* pertama dan *epoch* terakhir, pelatihan tersebut memberikan hasil bahwa nilai *loss* untuk data *train* berkurang sebesar 0,6 dan untuk data *validation* berkurang sebesar 0,5. Sedangkan nilai akurasi untuk data *train* bertambah 0,45 dan data *validation* bertambah sebesar 0,44. Hasil perubahan nilai *loss* untuk data *train* dan data *validation* dapat dilihat pada gambar 5, sedangkan hasil perubahan nilai akurasi untuk data *train* dan data *validation* dapat dilihat pada gambar 6.



Gambar 6. Loss train dan validation



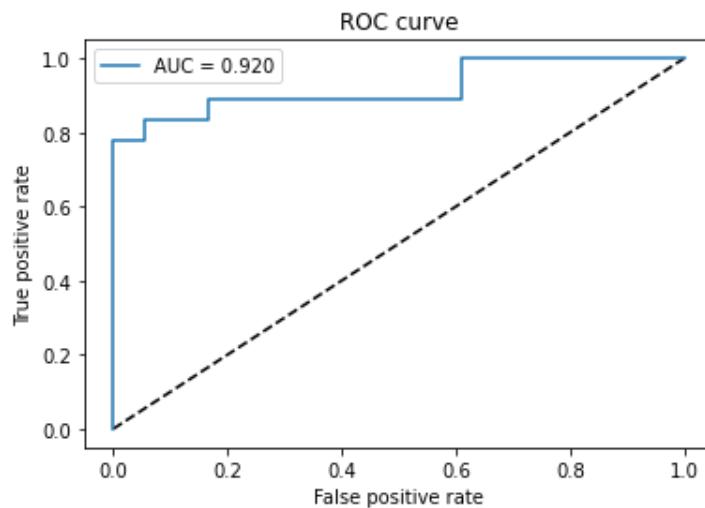
Gambar 7. Akurasi train dan validation

Selain melihat perubahan nilai *loss* dan akurasi pada data *train* dan *validation* selama proses pelatihan, diperhatikan pula *true positive*, *true negative*, *false positive*, dan *false negative* yang pada akhirnya digunakan untuk menghitung *precision*, *recall*, nilai *f1*, dan kurva ROC. Hasil *precision*, *recall*, dan nilai *f1* yang didapat disajikan dalam tabel 2. Dapat diketahui bahwa rata-rata *recall*, dan nilai *f1* adalah 0,86, sedangkan rata-rata *precision* adalah 0,89. Dikarenakan kelas ‘*drowsy*’ lebih penting dibanding kelas ‘*awake*’ dalam kasus ini, maka parameter evaluasi *recall* lebih diperhatikan.

	Precision	Recall	Nilai F1	Support
awake	0,78	1	0,88	18
drowsy	1	0,72	0,84	18
Akurasi			0,86	36
Rata-rata Makro	0,89	0,86	0,86	36
Rata-rata Beban	0,89	0,86	0,86	36

Tabel 2. Precision, Recall, nilai F1 dari CNN

Selanjutnya, gambar 7 menampilkan kurva *Receiver Operator Characteristics* (ROC). Kurva ROC menampilkan kemampuan prediksi model klasifikasi biner. Untuk mengetahui nilainya, dihitunglah *Area Under Curve* (AUC), yaitu area dibawah kurva ROC. Nilai tersebut berkisar antara 0 hingga 1, dimana nilai mendekati 1 menunjukkan kemampuan prediksi yang semakin baik, begitu pula sebaliknya. Nilai AUC dari kurva ROC yang ditampilkan pada gambar 8 memiliki nilai 0,92 yang mendekati 1. Hal ini berarti kemampuan prediksi model yang dibuat relatif baik.



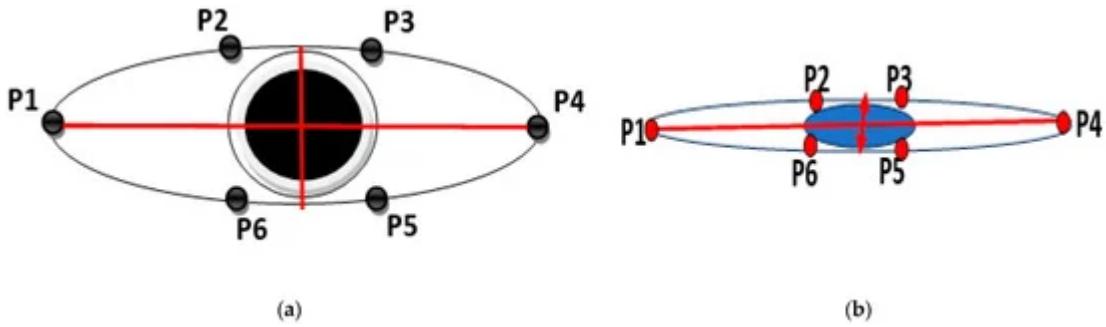
Gambar 8. Kurva ROC dan Nilai AUC

4.3 Eye Aspect Ratio

EAR adalah nilai yang didapatkan dari keterbukaan mata manusia (Dewi et. al., 2022). Langkah pertama dalam perhitungan EAR adalah dengan menentukan 68 titik di wajah, yang kemudian hanya diambil 6 titik masing masing mata, seperti yang terlihat pada gambar 9.



Gambar 9. Pendekatan Titik Mata



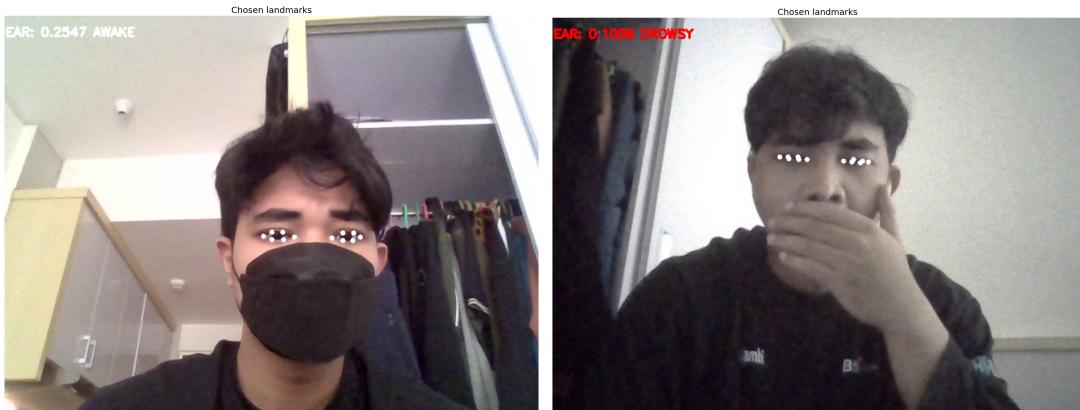
Gambar 10. Ilustrasi Titik Mata

Kemudian dari 6 titik tersebut dihitung nilai EAR tiap mata dengan menghitung jarak euclideanya dengan rumus di gambar 11.

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

Gambar 11. Rumus Jarak Euclidean

Nilai EAR rata-rata didapatkan dari rata-rata nilai EAR kedua mata. Jika EAR score kurang dari *threshold*, maka akan dideteksi mengantuk, seperti yang terlihat pada gambar 12.



Gambar 12. Deteksi Kantuk Metode EAR

Pengujian EAR menggunakan *threshold* 0.18, 0.19, 0.2, 0.21, dan 0.22 memberikan hasil klasifikasi yang dapat dilihat pada gambar 13. Dengan menggunakan *threshold* 0.21, klasifikasi menggunakan EAR mendapatkan akurasi yang baik sebesar 79.3%. Sehingga EAR dapat digunakan untuk klasifikasi kantuk.

	metode	akurasi	recall	precision	f1-score
0	thresh_021	0.793478	0.912088	0.734513	0.813725
1	thresh_022	0.793478	0.868132	0.752381	0.806122
2	thresh_018	0.788043	0.989011	0.703125	0.821918
3	thresh_019	0.788043	0.978022	0.706349	0.820276
4	thresh_020	0.777174	0.945055	0.704918	0.807512

Gambar 13. Hasil Klasifikasi Metode EAR

Kekurangan EAR untuk klasifikasi kantuk adalah kurang bisa mengklasifikasi kantuk bagi manusia yang memiliki mata yang sipit. Hal ini karena EAR menghitung keterbukaan mata, sedangkan mata sipit pasti akan mendapatkan nilai EAR yang lebih rendah daripada mata normal. Sehingga penentuan threshold akan menjadi bagian paling susah. jika menggunakan threshold yang normal, maka orang dengan mata sipit meskipun tersadar akan dideteksi sebagai kantuk. pun dengan jika threshold diturunkan, orang dengan mata normal yang mengantuk akan dideteksi sebagai tersadar.

5. Kesimpulan

Pada tabel 3 disajikan perbandingan hasil yang didapat dari tiga metode yang digunakan untuk klasifikasi, yaitu YOLOv5, CNN, dan EAR. Dikarenakan pentingnya menangkap kelas ‘drowsy’, maka pengukuran evaluasi yang paling diperhatikan adalah *recall*. Hal ini karena model dengan *recall* yang tinggi cenderung lebih banyak mengklasifikasikan gambar sebagai kelas positif, dalam kasus ini ‘drowsy’. Jika melihat *recall*, maka model YOLOv5 adalah yang tertinggi yaitu 1, namun nilai *precision* dan nilai f1 juga adalah 1. Hal ini menunjukkan adanya kemungkinan *overfitting* yang tinggi. Selanjutnya, model EAR dengan *threshold* 0,18 memiliki *recall* tertinggi kedua yaitu 0,989, begitu pula dengan nilai f1-nya dengan nilai 0,921. Meskipun begitu, nilai *precision* dan akurasi lebih rendah daripada model CNN.

Dari hasil yang didapat disimpulkan bahwa dalam kasus ini, model yang terbaik adalah model EAR dengan *threshold* 0,18. Hal ini dikarenakan tingginya *recall* yang dimiliki. Meskipun model YOLOv5 memiliki *recall* 1, dikarenakan *precision* dan nilai f1 yang juga 1, maka model YOLOv5 memiliki kemungkinan besar bersifat *overfit*. Sedangkan model CNN memiliki *recall* yang paling buruk yaitu 0,86.

Metode	Precision	Recall	Akurasi	Nilai F1
YOLOv5	1	1	0,900	1
CNN	0,890	0,860	0,860	0,860
EAR (thresh_018)	0,703	0,989	0,788	0,921

Tabel 3. Perbandingan Hasil YOLOv5, CNN, dan EAR

Daftar Pustaka

1. O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.
2. Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)* (pp. 1-6). Ieee.
3. Fujiwara, K., Abe, E., Kamata, K., Nakayama, C., Suzuki, Y., Yamakawa, T., & Kadotani, H. (2018). Heart rate variability-based driver drowsiness detection and its validation with EEG. *IEEE Transactions on Biomedical Engineering*, 66(6), 1769-1778.
4. Deng, W., & Wu, R. (2019). Real-time driver-drowsiness detection system using facial features. *Ieee Access*, 7, 118727-118738.
5. Ueno, H., Kaneda, M., & Tsukino, M. (1994, August). Development of drowsiness detection system. In *Proceedings of VNIS'94-1994 Vehicle Navigation and Information Systems Conference* (pp. 15-20). IEEE.
6. Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors*, 12(12), 16937-16953.
7. Aguilar, S., FM, R. A., ML, M. A., Jiménez, G., Martínez, R., & AI, M. N. (2005). Sleeping habits and sleep disorders during adolescence: relation to school performance. *Atención Primaria*, 35(8), 408-414.
8. Ferrara, M., & De Gennaro, L. (2001). How much sleep do we need?. *Sleep medicine reviews*, 5(2), 155-179.
9. Soukupova, T., & Cech, J. (2016, February). Eye blink detection using facial landmarks. In *21st computer vision winter workshop, Rimske Toplice, Slovenia*.
10. Dewi, C., Chen, R.-C., Chang, C.-W., Wu, S.-H., Jiang, X., & Yu, H. (2022). Eye aspect ratio for real-time drowsiness detection to improve driver safety. *Electronics*, 11(19), 3183. <https://doi.org/10.3390/electronics11193183>
11. Maulana, M. A. (2022). Deteksi Kantuk Pada Pengendara Roda Empat Melalui Citra Wajah Menggunakan Metode Facial Landmark.
12. Yang, D., Cui, Y., Yu, Z., & Yuan, H. (2021). Deep learning based steel pipe weld defect detection. *Applied Artificial Intelligence*, 35(15), 1237-1249.
13. Baihaqi, K. A., & Zonyfar, C. (2022). Deteksi Lahan Pertanian Yang Terdampak Hama Tikus Menggunakan Yolo v5: Indonesia. *Syntax: Jurnal Informatika*, 11(02), 01-11.