

Deteksi Indikasi Kelelahan Menggunakan Deep Learning

Dhomas Hatta Fudholi*, Royan Abida N. Nayoan, Maghfirah Suyuti, Ridho Rahmadi
Jurusan Informatika, Universitas Islam Indonesia
Jalan Kaliurang Km 14,5 Yogyakarta, Indonesia
*hatta.fudholi@uii.ac.id

Abstract

Many students experience fatigue due to lack of sleep which can be caused by a psychological conditions or bad habits. Lack of sleep can affect student's performance academically and causes many illnesses, stress and depression. Students with fatigue causes students to not study well, increasing risk of academic failure and will lead to having low GPA. In this research, fatigue detection is carried out to find out which students are experiencing fatigue. In this study, an annotated video dataset was used with a total of 18 subjects acted drowsy and alert. Fatigue detection is based on mouth movements, therefore mouth annotation is used. Mouth annotation has 2 categories, namely annotation 0 which indicates a closed mouth and annotation 1 which indicates the mouth is yawning. Previous study proves ResNet50 has better performance than other pre-trained models such as AlexNet, Clarifia, VGG-16, and GoogLeNet-19. We also applied image augmentation which is useful for providing new image variations to the model in each epoch by changing the rotation, random shift, and random zoom. ResNet50 model is used to perform binary classification which has two outputs, namely mouth stillness and yawning. The results of the frame classification are evaluated using precision, recall and f1-score. By using ResNet model, the results of the classification of frames labeled 0 or mouth stillness, obtained a precision of 0.72, a recall of 0.88, and an f1-score of 0.79. Meanwhile, the frame classification labeled 1 or yawning has a precision value of 0.85, a recall of 0.65, and an f1-score of 0.74.

Keywords: CNN, Deep Learning, Fatigue Detection

Abstrak

Banyak siswa yang mengalami kelelahan akibat kurang tidur yang disebabkan oleh keadaan psikologis atau karena kebiasaan buruk. Hal tersebut dapat mempengaruhi kinerja siswa secara akademik dan dapat menyebabkan timbulnya banyak penyakit, stress dan depresi. Hal tersebut juga menyebabkan siswa dengan kelelahan bisa mengganggu siswa untuk belajar sehingga mempengaruhi IPK dan meningkatkan risiko kegagalan akademik. Pada penelitian ini dilakukan deteksi kelelahan pada siswa untuk mengetahui siswa yang mengalami kelelahan. Pada penelitian ini digunakan video dataset yang sudah dianotasi dengan jumlah 18 subjek yang memperagakan ekspresi mengantuk dan terjaga. Deteksi kelelahan dilakukan berdasarkan pergerakan mulut, oleh karena itu anotasi mulut digunakan. Anotasi mulut memiliki 2 kategori, yaitu anotasi 0 yang menunjukkan mulut diam dan anotasi 1 untuk menunjukkan mulut menguap. Deteksi kelelahan menggunakan model ResNet CNN. Studi sebelumnya telah membuktikan ResNet50 memiliki performa yang lebih baik dibandingkan model pre-trained lainnya seperti AlexNet, Clarifia, VGG-16, dan GoogLeNet-19. Model juga mengaplikasikan augmentasi gambar yang berguna untuk memberikan model variasi gambar baru di setiap epoch dengan mengubah rotasi, pergeseran pixel, dan pembesaran atau pengecilan gambar dari gambar yang telah dibersihkan. Pada penelitian ini, model digunakan untuk melakukan klasifikasi biner yang memiliki keluaran dua kelas yakni mulut diam dan mulut menguap. Hasil dari klasifikasi frame dengan mulut diam dan membuka dievaluasi dengan menggunakan presisi, recall dan f1-score. Dari penggunaan model ResNet50, hasil klasifikasi frame dengan label 0 atau mulut diam didapatkan presisi sebesar 0.72, recall 0.88,

dan *f1-score* 0.79. Sedangkan klasifikasi frame dengan label 1 atau mulut menguap memiliki nilai presisi sebesar 0.85, *recall* 0.65, dan *f1-score* 0.74.

Kata kunci: CNN, Deep Learning, Deteksi Kelelahan

1. PENDAHULUAN

Banyak factor yang membuat siswa kelelahan di kelas, misalnya banyak siswa yang tidak mendapatkan waktu tidur yang cukup akibat tidur terlalu terlambat dan bangun terlalu pagi sehingga mendapatkan tidur kurang dari 8 jam [1]. Hal ini bisa saja terjadi akibat kebiasaan yang tidak baik seperti minuman berenergi atau teknologi yang membuat siswa terbangun hingga malam. Kurangnya tidur dapat mengurangi kinerja siswa secara akademik. Studi menyatakan bahwa kebanyakan siswa yang memiliki IPK rendah memiliki kualitas tidur yang buruk, dengan durasi tidur kurang dari 5 hingga 7 jam, sehingga mengalami disfungsi pada siang hari setiap hari [2]. Banyak siswa yang memiliki waktu kurang untuk tidur memiliki lebih banyak penyakit, mengalami stress dan depresi, dan lebih mungkin untuk menggunakan obat stimulant untuk tetap terjaga

Pada penelitian ini, akan dilakukan deteksi untuk mengetahui siswa yang mengalami kelelahan. Kelelahan pada siswa dapat ditunjukkan melalui ekspresi wajah, yakni mata sering berkedip, mata diam, dan menguap. Pada penelitian ini, dilakukan deteksi kelelahan dengan menggunakan pergerakan mulut. Data yang digunakan yakni data video dari 18 subjek. Setiap video telah memiliki anotasi untuk setiap frame yang ada pada video. Untuk mendeteksi kelelahan siswa melalui pergerakan mulut, maka digunakan anotasi mulut yakni anotasi mulut diam atau mulut terbuka. Mulut diam maka terjaga dan mulut membuka menandakan mengantuk.

Deteksi kelelahan pada siswa menggunakan CNN based metode yakni ResNet50. ResNet50 merupakan model yang memiliki 50 layer Convolutional Neural Network (CNN). Pada studi sebelumnya, ResNet50 memiliki performa yang lebih baik dibandingkan model pre-trained lainnya seperti AlexNet, Clarifia, VGG-16, dan GoogLeNet-19 [3]. Pada penelitian ini, model digunakan untuk melakukan klasifikasi biner yang memiliki keluaran dua kelas yakni mulut diam dan mulut menguap, sehingga fully connected layer (FC layer) yang pada model sebelumnya memiliki 1000 output diubah menjadi 2 output [4].

Penelitian untuk mendeteksi kantuk lain juga menggunakan ResNet dan menghasilkan model yang memiliki nilai di atas 85% [5]–[7]. Masih terkait dengan deteksi kelelahan ketika sekolah, penelitian [7] mendeteksi jika siswa kursus online mengantuk. Penelitian ini berdasarkan akibat tidak adanya pengawasan khusus oleh guru ketika kursus online. Penelitian ini menggunakan fitur aspek rasio mata dan mulut sebagai step pertama sebelum memberikan fitur mata dan mulut yang berhasil didapatkan ke dalam model CNN.

Penelitian [8], [9] menggunakan fitur mata dengan metode Viola-Jones dan mampu memiliki hasil di atas 94%. Pada penelitian [9] bisa bekerja



dengan baik secara real-time, namun subjek harus memperlihatkan mata secara jelas, jika subjek menggunakan kacamata hitam atau pencahayaan yang kurang bagus, model gagal untuk mendeteksi drowsiness. Selain itu ada beberapa penelitian yang menggunakan fitur ketika menguap [10], [11]. Metode dan pengklasifikasi yang digunakan oleh [10] adalah cascade classifier, SVM, dan Viola-Jones untuk melatih model pada gambar mulut dan mulut menguap. Sedangkan [11] menggunakan metode SVM dan CHT. Pada penelitian ini SVM digunakan untuk mendeteksi mulut yang terbuka. Hasil yang didapatkan model yaitu 98%.

2. METODOLOGI PENELITIAN

2.1. Dataset

Dataset yang digunakan untuk mendeteksi kantuk adalah data NTHU Drowsy Driver Detection (NTHU-DDD) Video Dataset [16]. Video yang disediakan berukuran 640x480, 15/30 frame per second (fps), dan direkam menggunakan iluminasi infrared (IR) dalam keadaan pencahayaan yang bervariasi. Video berformat AVI dan tidak memiliki suara. Dataset terdiri dari laki-laki dan perempuan yang memiliki ciri wajah yang berbeda, suku yang berbeda, serta masing-masing subjek memiliki 5 skenario yang berbeda.

Skenario terdiri dari wajah tanpa kacamata (NoGlasses), menggunakan kacamata (Glasses), kacamata hitam (Sunglasses), wajah tanpa kacamata pada malam hari (Night-NoGlasses), dan kacamata pada malam hari (Night-Glasses). Setiap video memiliki status mengantuk, dan tidak mengantuk. Setiap skenario memiliki situasi yang berbeda beserta anotasinya, antara lain video yang mengandung tingkah laku menguap, mata mengantuk dan menunduk, video kombinasi tertawa, berbicara, dan melihat ke samping, serta video kombinasi mengantuk seperti mata mengantuk, menguap, dan menunduk. Setiap video memiliki beberapa anotasi seperti, anotasi mengantuk, anotasi pergerakan kepala, anotasi pergerakan mulut, dan anotasi pergerakan mata. Pada penelitian ini, digunakan video dimana subjek menggunakan kacamata dan tidak menggunakan kacamata pada siang hari, serta menggunakan anotasi mulut diam yang memiliki label 0 dan menguap yang memiliki label 1. Pada Tabel 1 menunjukkan keterangan anotasi mulut pada video.

Tabel 1. Keterangan Anotasi

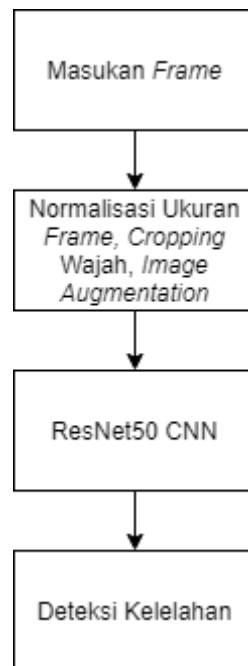
Kondisi	Anotasi	Keterangan
Mulut	0	Diam
	1	Menguap

2.2. Ekstraksi Frame

Dengan dataset yang setiap videonya memiliki label untuk setiap framenya, mempermudah dalam ekstraksi fitur. Pada tahap ini akan dilakukan pengambilan setiap frame yang ada pada video dengan menyesuaikan label videonya. Label yang digunakan yaitu 2 label, yakni mulut diam yang memiliki label 0 dan menguap yang memiliki label 1.

Ekstraksi frame diperlukan untuk mendapatkan frame yang memiliki label yang sama yang kemudian dapat digunakan untuk melakukan deteksi kantuk. Digunakan opencv untuk melakukan ekstraksi frame.

2.3. Model CNN Based

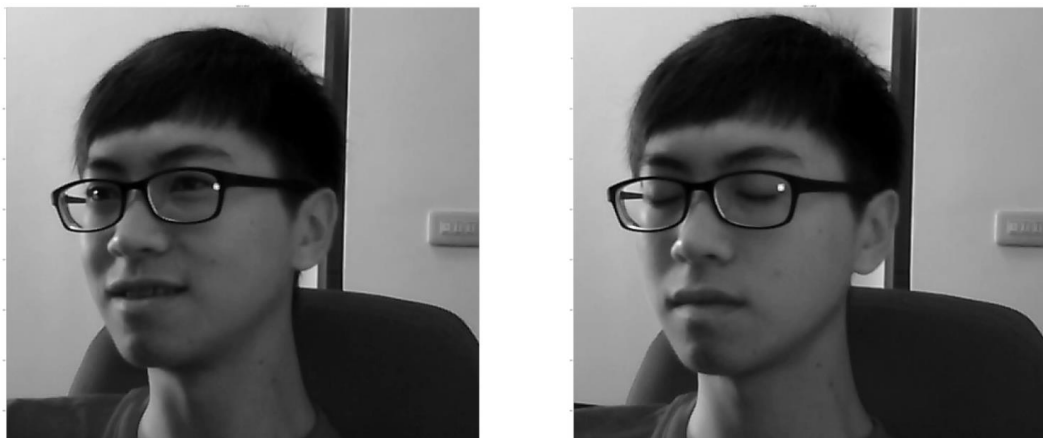


Gambar 1. Tahapan klasifikasi mulut dengan model ResNet50

Gambar 1 menunjukkan flowchart yang digunakan mengklasifikasi mulut yang terbuka ataupun diam menggunakan metode yang akan digunakan. Pada tahap pertama, frame yang telah didapatkan pada tahap sebelumnya akan digunakan sebagai input, yang kemudian akan dilakukan perubahan ukuran gambar yang berukuran 640 x 480 pixel menjadi berukuran 100 x 100 pixel dan memotong bagian frame agar mendapatkan wajah pada tahap kedua. Gambar 2, 3 menunjukkan perbedaan frame sebelum dan sesudah dilakukan *cropping* wajah. Hasil dari gambar-gambar tersebut, selanjutnya akan dilakukan *image augmentation* sehingga dapat memperbanyak variasi gambar pada model. Gambar yang sudah diaugmentasi kemudian akan digunakan sebagai masukan pada CNN based untuk melakukan klasifikasi mulut diam dan menguap.



Gambar 2. Hasil ekstraksi frame sebelum dilakukan *cropping* wajah



Gambar 3. Hasil ekstraksi frame sesudah dilakukan *cropping* wajah

Metode yang digunakan yaitu model ResNet-50 [12]. Struktur CNN yang digunakan menggunakan jumlah node, filter, ukuran filter dan layer yang sama sesuai dengan ResNet-50 CNN, kecuali jumlah output pada fully connected layer (FC layer). Pada penelitian ini, model digunakan untuk melakukan klasifikasi biner yang memiliki keluaran dua kelas yakni mulut diam dan mulut menguap, sehingga layer FC diubah menjadi 2.

2.4. Evaluasi

Performa dari deteksi kantuk diukur dengan presisi, *recall*, dan $F\beta=1$ yang dinyatakan pada Persamaan 1, 2. Sebagai ilustrasi, Tabel 2 menampilkan *confusion matrix* X yang digunakan. Label l merupakan label yang dimiliki oleh objek, sedangkan z merupakan label yang dihasilkan oleh mesin. TP merupakan true positive dimana frame terklasifikasi dengan benar, FP merupakan *false positive* dimana frame memiliki label, dan FN merupakan *false negative* [13]. Presisi mengukur frame yang ada pada korpus terklasifikasi benar dengan menggunakan mesin pembelajaran. *Recall* mengukur frame yang relevan yang didapatkan oleh model NER. Sedangkan F1 score merupakan rata-rata terimbang dari presisi dan recall [14].

Tabel 2. Tabel *confusion matrix* X

		Assignment z	
		+	-
Label ℓ	+	TP	FN
	-	FP	TN

$$F_{\beta} = \frac{(\beta^2+1)*presisi*recall}{(\beta^2*presisi+recall)} \tag{1}$$

$$p = \frac{TP}{TP+FP} \quad r = \frac{TP}{TP+FN} \tag{2}$$

3. HASIL DAN PEMBAHASAN

Dataset yang digunakan merupakan dataset video dimana subjek menggunakan kacamata dan tidak menggunakan kacamata pada siang hari. Pada video dilakukan Ekstraksi frame yang terbagi menjadi 2 label, yakni mulut diam dan menguap. Jumlah frame dengan label mulut diam didapatkan sejumlah 252657 frame, sedangkan label menguap didapatkan sebanyak 109283 frame. Pada setiap label, dipilih 4500 gambar pada label 0 atau mulut diam, dan 4500 gambar pada label 1 atau mulut terbuka. Gambar 4, 5 menunjukkan frame dengan label 0 mulut diam dan label 1 mulut mulai menguap.



Gambar 4. Frame dengan label 0 (mulut diam)



Gambar 5. Frame dengan label 1 (mulut mulai menguap)

Penelitian ini dibagi menjadi 3 dataset yakni training set, validation set dan testing set. Training set menggunakan 83% dari total data dan 17% dari total data digunakan sebagai testing set. Sedangkan untuk validation set, digunakan 226 gambar hasil ekstraksi dari video yang memiliki 2 label mulut diam dan mulut menguap sekaligus.

Model juga menggunakan Image Data Generator dari Keras yang berguna untuk memberikan model variasi gambar baru di setiap epoch. Image Data Generator akan mengubah gambar yang ada dan menambahkan ke model tanpa menyimpan gambar. Adapun beberapa perubahan pada gambar yang bisa dilakukan yaitu, rotasi secara acak, pada model ini digunakan 30 derajat. Ada pula *random shifts* untuk menggeser pixel gambar secara horizontal maupun vertikal, serta *random zoom* untuk memperbesar atau memperkecil gambar. Pada penelitian ini *random shifts* dan *random zoom* ditetapkan nilai 0.5.

Model diawali dengan layer masukan yang menerima gambar berukuran 100x100. Lalu diikuti oleh layer ResNet50 CNN yang berguna untuk melakukan klasifikasi gambar. Selanjutnya diikuti oleh dense layer dengan neuron 1024 dan aktivasi relu agar model dapat mempelajari hal kompleks sehingga mampu menghasilkan model yang bagus. Kemudian digunakan dropout dengan nilai 0.1 untuk mengurangi *overfitting* [15]. Kemudian diikuti oleh *dense layer* kedua dengan neuron 1024 dan aktivasi relu. Setelah itu digunakan *batch normalization* agar mesin bekerja lebih cepat dan lebih stabil karena adanya normalisasi nilai masukan ke dalam layer. Ditambahkan dropout yang kedua dengan nilai dropout sebesar 0.5. Diikuti dense layer ketiga dengan neuron sebesar 512 dan aktivasi relu. Layer terakhir yaitu fully connected layer yang memiliki 2 output menggunakan aktivasi sigmoid untuk mengklasifikasikan frame menjadi dua kelas berlabel 0 dan 1. Model dijalankan sebanyak epoch 40, dan memiliki *batch size* sebesar 64.

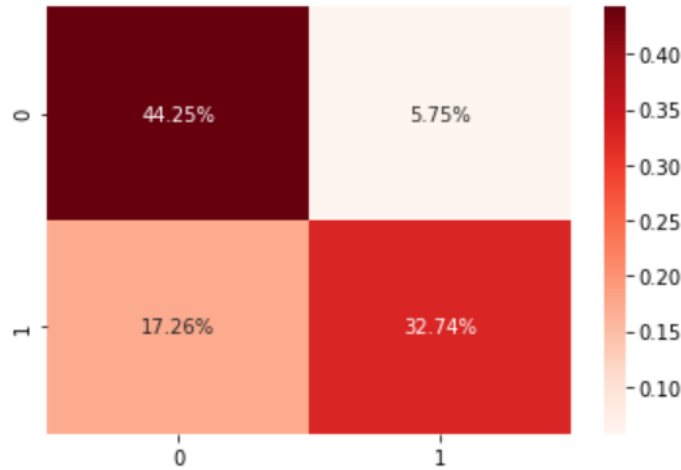
Hasil dari klasifikasi frame dengan mulut diam dan membuka dievaluasi dengan menggunakan presisi, *recall* dan f1-score. Pada table X, hasil klasifikasi frame dengan label 0 atau mulut diam didapatkan presisi sebesar 0.72, *recall* 0.88, dan f1-score 0.79. Sedangkan klasifikasi frame dengan label 1 atau mulut menguap memiliki nilai presisi sebesar 0.85, *recall* 0.65, dan f1-score 0.74.

Tabel 3. Nilai presisi, recall, f1-score

	Precision	Recall	F1-Score
0	0.72	0.88	0.79
1	0.85	0.65	0.74

Model yang diaplikasikan pada validation set didapatkan 100 gambar atau sebanyak 44.25% dari total validation set yang berjumlah 226 yang terklasifikasi benar dengan label 0 atau mulut diam, sedangkan 74 gambar atau 32.74% terklasifikasi benar dengan label 1 atau mulut membuka.

Gambar 6 menunjukkan kemampuan model dalam mengklasifikasikan gambar yang divisualisasikan menggunakan *heatmap*.



Gambar 6. *Heatmap* hasil prediksi

4. SIMPULAN

Deteksi kelelahan pada siswa dilakukan menggunakan ResNet50 CNN berdasarkan pergerakan mulut. Pada penelitian ini digunakan video dataset yang sudah dianotasi dengan jumlah 18 subjek yang memperagakan ekspresi mengantuk dan terjaga. Dilakukan deteksi kelelahan pada siswa dengan melihat pergerakan mulut yang kemudian diklasifikasikan menjadi 2 keluaran yakni mulut diam dan mulut menguap. Hasil klasifikasi frame dengan model ResNet50 dengan label 0 atau mulut diam didapatkan presisi sebesar 0.72, *recall* 0.88, dan f1-score 0.79. Sedangkan klasifikasi frame dengan label 1 atau mulut menguap memiliki nilai presisi sebesar 0.85, *recall* 0.65, dan f1-score 0.74.

Penelitian ini dilakukan berdasarkan pergerakan mulut dan anotasinya untuk mendeteksi kelelahan. Untuk penelitian selanjutnya, akan lebih akurat jika tidak hanya menggunakan pergerakan mulut, namun juga pergerakan mata seperti mata sering berkedip atau mata menutup, dan kepala yang menunduk untuk mendapatkan model yang lebih baik dalam mendeteksi kelelahan.

DAFTAR PUSTAKA

- [1] M. A. Alsaggaf, S. O. Wali, R. A. Merdad, and L. A. Merdad, "Sleep quantity, quality, and insomnia symptoms of medical students during clinical years: Relationship with stress and academic performance," *Saudi Med. J.*, vol. 37, no. 2, pp. 173–182, 2016, doi: 10.15537/smj.2016.2.14288.
- [2] G. Maheshwari and F. Shaukat, "Impact of Poor Sleep Quality on the Academic Performance of Medical Students," *Cureus*, vol. 11, no. 4, pp.

- 3-8; 2019, doi: 10.7759/cureus.4357.
- [3] M. Z. Alom *et al.*, "A state-of-the-art survey on deep learning theory and architectures," *Electron.*, vol. 8, no. 3; pp. 1-67, 2019, doi: 10.3390/electronics8030292.
 - [4] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, vol. 20, no. 11, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.
 - [5] K. W. Kim, H. G. Hong, G. P. Nam, and K. R. Park, "A study of deep CNN-based classification of open and closed eyes using a visible light camera sensor," *Sensors (Switzerland)*, vol. 17, no. 7, 2017, doi: 10.3390/s17071534.
 - [6] M. Ngxande, J. R. Tapamo, and M. Burke, "Detecting inter-sectional accuracy differences in driver drowsiness detection algorithms," *arXiv*, 2019.
 - [7] U. Lahoti, R. Joshi, N. Vyas, K. Deshpande, and S. Jain, "Drowsiness detection system for online courses," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 1930-1934, 2020, doi: 10.30534/ijatcse/2020/158922020.
 - [8] T. P. Nguyen, "Eye Tracking System to Detect Driver Drowsiness," pp. 472-477, 2015.
 - [9] R. Ahmad and P. J. N. Borole, "Drowsy Driver Identification Using Eye Blink detection," vol. 6, no. 1, pp. 270-274, 2015.
 - [10] M. S. Devi and P. Bajaj, "Driver Fatigue Detection Using Mouth and Yawning Analysis," no. January 2008, 2015.
 - [11] C. Yan, F. Coenen, Y. Yue, X. Yang, and B. Zhang, "Video-Based Classification of Driving Behavior Using a Hierarchical Classification System with Multiple Features," vol. 30, no. 5, pp. 1-33, 2016, doi: 10.1142/S0218001416500105.
 - [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 770-778, 2016, doi: 10.1109/CVPR.2016.90.
 - [13] C. Goutte and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation," in *Lecture Notes in Computer Science*, vol. 3408, no. April, 2005, pp. 345-359.
 - [14] K. Tjong, E. F. Sang, and F. De Meulder, "Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition," 2003, doi: 10.3115/1118853.1118877.
 - [15] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," 2014.
 - [16] C. Weng, Y. Lai, S. Lai, "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network", *In Asian Conference on Computer Vision Workshop on Driver Drowsiness Detection from Video*, 2016.