

EyeNet: An Improved Eye States Classification System using Convolutional Neural Network

Md. Moklesur Rahman*, Md. Shafiqul Islam*, Mir Kanon Ara Jannat*, Md. Hafizur Rahman[†], Md. Arifuzzaman[‡]

Roberto Sassi[§], and Md Aktaruzzaman[¶]

*Dept. of Computer Science and Eng., The People's University of Bangladesh, Dhaka, Bangladesh.

[†]Dept. of Electrical and Electronic Eng., Islamic University, Kushtia, Bangladesh.

[‡]Department of Civil and Environmental Eng., King Faisal University, Saudi Arabia.

[§]Dipartimento di Informatica, Università degli Studi di Milano, Italy.

[¶]Dept. of Computer Science and Eng., Islamic University, Kushtia, Bangladesh.

{moklesur.ai, msislam.iu, kanon.iucse, hafizur.iueee}@gmail.com, marifuzzaman@kfu.edu.sa, roberto.sassi@unimi.it, and aktaruzzaman@iu.ac.bd

Abstract—The classification of eye states (open or closed) has numerous potential applications such as fatigue detection, psychological state analysis, smart home devices controlling, etc. Due to its importance, there are a number of works already reported in the literature using traditional shallow neural networks or support vector machines, which have reported good accuracy (about 96%). However, there is still enough space to improve the accuracy of existing systems using proper classification methods. The major problem with traditional classifiers is that they depend on manual selection of features and that is very challenging to select meaningful features for such classifiers. Convolutional neural networks (CNNs) have become popular for computer vision and pattern recognition problems with better performance than traditional methods. In this study, we proposed a model of CNN (EyeNet) for eye states classification and tested it on three datasets (CEW, ZJU, and MRL Eye). The quality (or diversity) of a recently proposed larger dataset (MRL Eye) has been compared with two other existing datasets with respect to the sufficient training of the model. The model shows very high performance (about 99% accuracy for classification of eye states on the test set of data when it is trained by the training samples of the same dataset. The proposed model improves the accuracy of the best existing method by about 3%. The performance of the model for classification of samples coming from different datasets is reduced when it is trained with the MRL Eye dataset. This concludes that even though, the MRL Eye has a large number of samples compared to other datasets, but diversity still lacks in the MRL Eye samples to sufficiently train the model.

Index Terms—Eye State, Classification, Deep Learning, Convolutional Neural Network.

I. INTRODUCTION

Eye states classification (or recognition) deals with recognizing the status of an eye if it is open or closed. The interest in the research of eye states classification is reflected by a large number of publications related to its numerous potential applications in biomedical systems [1], emotion tracking [2], epileptic seizure detection [3], and so on. According to the report provided by the National Highway Traffic Safety Administration (NHTSA) in 2014, about 2.6% of total casualties were due to drowsy, sleepy, asleep, or exhausted drivers.

Almost (2-2.3)% fatal accidents of total road accidents from 2005 to 2009 were occurred due to drowsy driving [4].

Drowsy driver detection system utilizes eye states to assess the driver's attentiveness [5]. The eye blinking information of an individual can be used as an alternative mean of communication with computers [6]. The concept of controlling an external device with one's thought is highly promising in assisting motion or speech impaired people. Eye-blink switch technology allows motion or speech impaired people to utilize and control domestic devices, speech synthesizers, lights, wheelchair, computers, and other assistive deployments.

Considering its importance, there are many studies already published in the literature on eye states classification. The eye states classification methods can be categorized into non-image based [7] and image-based methods [8]. Non-image based methods classify eye states based on the data acquired through Electroencephalograms (EEG) [7], Electrooculography (EOG) [9] and Electromyography (EMG) [10]. Although data acquisition is faster and provides reasonable accuracy with lower error rate in non-image based methods, they suffer from some limitations such as the necessity of wearing sensor devices which might lead to discomfort. Also, physiological signals (EEG, EOG, EMG, etc.) are very weak and can easily be corrupted by artifacts that may be caused by muscle movements of different body parts.

To overcome the limitations of non-image based methods for eye states classification, some image-based methods [6], [11], [12] have been proposed. These methods do not require wearing sensor devices; the users comfort can also be enhanced by allowing the natural movement of the head or body. Image-based methods can be categorized into single image-based and video-based methods. Mohanakrishnan *et al.* [12] have proposed a method based on the eyelid and face movement.

To obtain high accuracy, video-based methods [12] deal with occlusion and clutter from several images in the video. But they require complicated mathematical operations and

hence lead to heavy computational costs. In addition, these methods are very sensitive to the illumination conditions so that they have poor performance with lighting changes. The image-based methods classify eye states from captured images of the eyes. These methods depend on the features extracted from eye images which are used to train a classification model. Most of the studies reported in the literature have used support vector machines (SVM) [13], artificial neural network (ANN) [14], linear discriminant analyses (LDA) [15], local binary patterns (LBP) [16], Gabor wavelets [17], etc for this classification problems. However, the performance of these models can vary depending on the image quality, the robustness of extracted features, and the proposer selection of the classification model. In addition, manual features extraction methods are too cumbersome and demand a high level of experience and expertise.

Recently, deep learning (DL) based classification methods [18]–[20] have been proposed to overcome the limitations of manual features extraction based classifiers. The latest advancements in computational capabilities and the development of high computing power Graphics Processing Units (GPUs) have made possible the implementation of these DL methods. There are several DL architectures such as deep belief networks [21], stacked autoencoder [22], and convolutional neural networks [23] that have been proved effective to provide state-of-the-art accuracy in the field of computer vision, image processing, and natural language processing. There are a few studies reported in the literature [11], [24], [25] that have used CNNs for eye states classification.

The performance of a model may vary for the problems in one domain with respect to another domain [26]. The objectives of this study are: i) to propose a model (architecture) of a convolutional neural network, ii) to evaluate the performance of the proposed model for eye states classification, and iii) to compare the performance of the proposed model with some existing methods on the same dataset.

The remainder of this paper is structured as follows. Section II describes briefly the dataset considered for this study. Section III explains the steps involved in the proposed method. Section IV represents the experimental results obtained from this study, and finally section V provides the conclusions and future research directions.

II. DATASET

We considered three publicly available datasets: CEW [27], ZJU [28] and MRL Eye [29] for this study. The images quality and sources of these datasets are quite different from each other. The statistics of each dataset have been presented in Table I and some samples of these datasets are also shown in Figure 1. A brief description of each dataset is given in the following:

A. CEW dataset

The CEW dataset was developed for eye states detection in the wild by Song *et al.* [27]. The samples of this dataset were collected from 2423 subjects, among which 1231 subjects

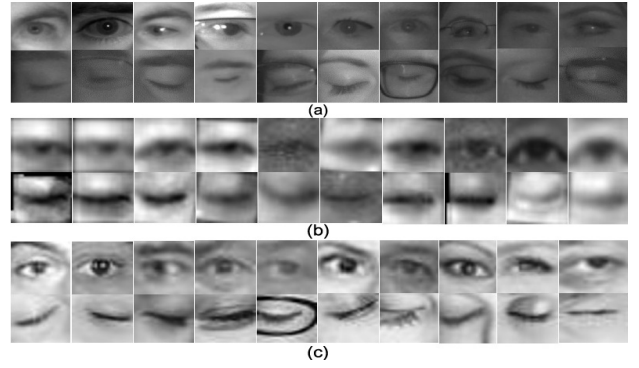


Fig. 1. Sample eye patches in (a) CEW, (b) ZJU, and (c) MRL Eye datasets.

TABLE I
STATISTICS OF THE DIFFERENT DATA SET USED IN OUR EXPERIMENT

Dataset	Subjects	Eye Open	Eye Closed	Total
CEW	2423	2462	2384	4846
ZJU	20	7000	1984	8984
MRL Eye	37	42951	41947	84,898

with eyes open are chosen from the Labeled Face in the Wild (LFW) [30] database, while 1192 subjects with both eyes closed were accumulated from the Internet. The resolution of all samples of this dataset is 24×24 pixels.

B. ZJU dataset

Images in the ZJU dataset were collected from the ZJU eye blink dataset [28]. Song *et al.* [27] cropped the images in the ZJU dataset to extract eye patches and expanded the initial set of eye images by rotation, blurring, the addition of Gaussian noise and contrast adjustment.

C. MRL Eye dataset

The MRL Eye dataset [29] were collected from 37 different persons (33 men and 4 women). The samples of the MRL Eye images were collected from subjects, some of them were wearing glasses. The images were captured under three reflection modes based on the size of reflections (none, small, and high), bad and good lighting using different sensors like RealSense, IDS, and Aptina.

III. PROPOSED METHOD

The major steps involved in the proposed method for eye states classification have been summarized in the block diagram in Figure 2.

A. Data Preprocessing

256 level gray color images were resized at 24×24 pixels. The image pixels were normalized by using Minmax normalizer (1) which transforms each gray level pixel in the range [0,1]. To diminish the impact of the illumination differences

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}, \quad (1)$$

where x is the original intensity and x' is its normalized value.

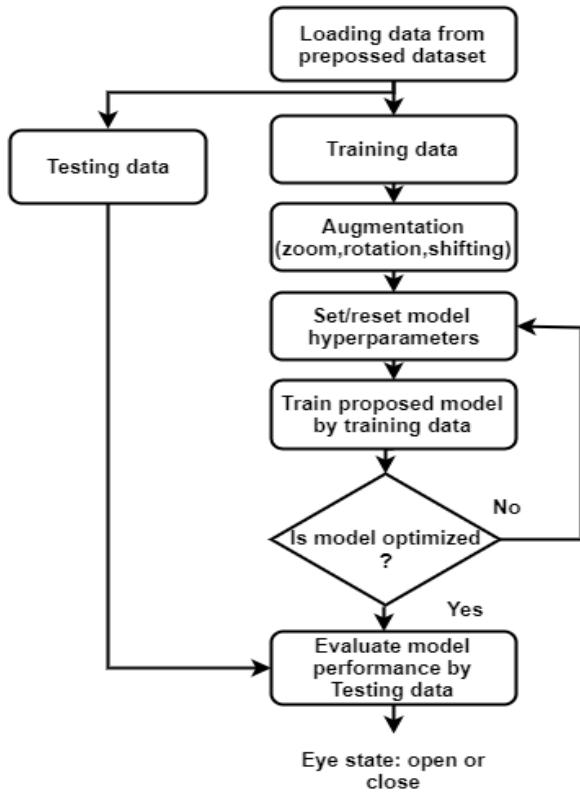


Fig. 2. The methodological steps of the proposed method.

B. Data Augmentation

Deep learning-based classification methods require large datasets for proper training. However, it is very difficult to collect this huge number of samples manually. Alternatively, the dataset size can be increased by adding diversities in the samples of some small or moderately large datasets. This way of increasing the number of samples of a dataset is called data augmentation. The images were augmented by randomly varying rotations, shifting, and zooming.

C. Convolutional Neural Network Model

The concept of CNN was introduced first by [23] inspired by the visual cortex of the brain which processes visual information. The visual cortex comprises multiple cells which are called receptive fields and accountable for determining the light in small, overlapping sub-regions of the visual field. Broadly, a CNN is a deep network topology that generally assembles convolution layers in conjunction with a classification network which is considered as a fully connected (FC) neural network. Usually, the convolution layers of a CNN are followed by a subsampling layer and a nonlinear layer. A large set of features from each pixel of an image is extracted by the network and then construct the features maps of the input image. Convolution layers are devoted as a feature extractor. The final FC layer represents the output layer which decided the final decision. The convolution layer is trained to adopt striking feature information from the input

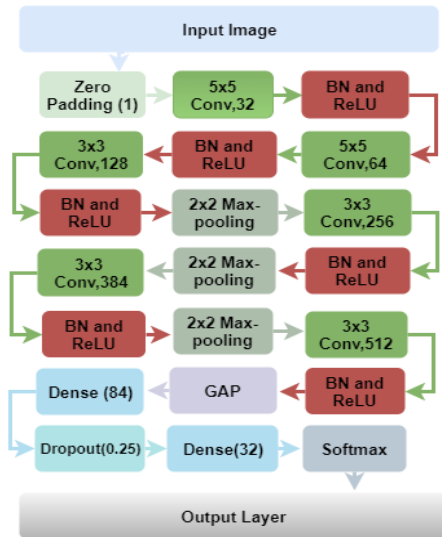


Fig. 3. The architecture of the proposed EyeNet model.

data through the back-propagation training algorithm. Here, we manifest that these features mainly represent those features present within data that characterize the classes (eye states) from each other. The architecture takes these features and passes them to the classification process. In this study, we have proposed a model of CNN, here we called it EyeNet, as the most promising approach to deal with lighting condition (difference in illumination), distortion, reflection mode, and noise on the problem of eye states classification.

The core components of the proposed EyeNet model have been depicted in Figure 3. A brief description of each layer of the model is given here:

- Zero-padding layer: In the zero-padding layer, the rows and columns of extra zeros are added on all sides of the input image to deal with boundary issues in the convolution operations.
- Convolutional layer: Convolutional operation is performed by sliding convolutional filter of size 3×3 and 5×5 over the input data matrix by stride 1. We tried the number of kernels starting at 64 to 1024 with different step sizes and selected the combination of kernels which maximized the validation accuracy.
- Activation function: We utilized ReLU [31] nonlinear activation function (2), which transfers the negative element into 0 and can be formulated as:

$$ReLU(x) = \max(0, x), \quad (2)$$

where x denotes the input to a neuron.

- Batch normalization: Batch normalization layer is followed by every activation function. It enables accelerating learning process [32] and helps to reduce sensitivity to variations within the input data.
- Dropout: Over-fitting occurs when the model learns and performing well on the training dataset, but does not perform well on the test set. The dropout layer is responsible

to prevent the over-fitting problem and in enhancing the performance [33] by setting activation to zero in a random manner.

- Max-pooling layer: It down-samples every depth slice of the input images with filters of size 2×2 , with a stride of size 2.
- Global average pooling layer: The global average pooling (GAP) layer calculates the average output of each feature map in the previous layer.
- Fully-connected layer: In a fully connected layer, every neuron in one layer is linked to every neuron in another layer.
- Output layer: This layer is responsible for yielding the occurrence probability of each class given the input image. It results in a vector of length 2 with softmax activation function (3), where each scalar is the probability of belonging to one of the eye state (open or closed). The softmax activation function can be expressed mathematically as:

$$\sigma(X_j) = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}}; \text{ for } j = 1, 2, 3, \dots, K. \quad (3)$$

where, x_j are the inputs from the previous fully-connected layer used to each Softmax layer node and K is the number of final FC layer nodes.

D. Training and Evaluation

The training and evaluation of the model has been divided into two steps: i. model hyper-parameters selection and ii. training and evaluation using those parameters.

1) *Model Hyper-Parameters Selection*: The performance of the proposed model has been evaluated on three different datasets of eye states images, which have been introduced in section II. To select the optimum set of values of the hyperparameters of the model, we used the ZJU dataset only. This dataset has separate train and test subsets, where 80% of the total data belongs to the training subset and the remaining 20% belongs to the test subset. In this study, we considered 15% of the training set for validation.

For data augmentation, we varied the rotation from -10 degrees to 10 degrees, shifting by $\pm 5\%$ to 20% at step 5, and zooming by 5% to 20% at step 5. The number of kernels was varied from 64 to 1024 with step size 32. In our model, the Adam optimizer [34] has been used to minimize the error with initial learning rate: $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1 \times 10^{-8}$. To calculate the error for optimizing the algorithm, we have used categorical cross-entropy(4).

$$C = \frac{1}{m} \sum [y \ln a + (1 - y) \ln(1 - a)], \quad (4)$$

where m is the batch size, y is an expected value, and a is an actual value from the output layer.

Recent studies assert that cross-entropy performs better than other functions like classification error and mean squared error [35]. We trained our model for up to 200 epochs with batch size 128 and 64 steps per epoch. An early stopping callback on

the validation loss with the patience of 5 epochs was applied to halt the training process when no improvements were detected.

The learning rate is one of the most significant hyper-parameters used to tune the training of CNNs. A small learning rate helps to classify more accurately but the optimizer consumes more time to reach the global optima or by reducing the loss. However, a large learning rate may lead the model not to converge. So, choosing the best learning rate is more cumbersome. In this study, the learning rate was initialized randomly, and then it was updated to 75% of its previous value if our model validation accuracy did not improve for six consecutive epochs. Weights were initialized from a standard normal distribution with mean $\mu = 0$, and standard deviation $\sigma = 1$. Moreover, we allowed weight decay at the rate of 1×10^{-6} . Then, we selected the combination of those sets of values of the parameters which maximized the accuracy on the validation set of the ZJU dataset.

2) *Training and evaluation with optimum hyper-parameters*: After training the model with training samples and selecting the model with hyper-parameters which provided maximum accuracy on the validation set, we evaluated the performance of the model on the test subset of the ZJU dataset. Some existing methods [27], [36] have reported high accuracy (about 94%) on the CEW dataset using 10 fold cross-validation. In order to compare the performance of the proposed model with the existing works reported in the literature, we followed the same techniques and used 10 fold validation. To the best of our knowledge, no significant work (neither traditional nor deep learning-based techniques) exists on the MRL Eye dataset for eye states recognition. This dataset was proposed for pupil localization. Since there is no work reported on the MRL Eye dataset, we applied 10-fold cross-validation for training and evaluating the proposed model on this dataset also. In 10-fold cross-validation, the dataset was divided into 10 random folds, 9-folds among them were used for training the model with the values of hyper-parameters selected in section III-D1, and the remaining fold was used to evaluate the performance of it. This train-test procedure was repeated 10 times, and the average accuracy of these 10 repetitions was reported as the accuracy.

The experiments were carried out on a Desktop computer with Intel Core i7 3.90 GHz CPU and NVIDIA Titan XP Pro 12 GB GPU, 1TB HDD and 8 GB RAM. The code was implemented using Keras with a Tensorflow backend.

IV. RESULTS

Experimental results show that the model provided maximum accuracy when data were augmented by varying rotations between -5 to 5 degrees, zooming by 5%, and shifting by 5%. The value of α and number of epochs for which the model provided maximum accuracy on the validation set of ZJU dataset were 0.0003164 and 134, respectively.

The performance of the proposed model for eye states classification has been summarized in Tables II to V.

TABLE II

COMPARISON OF THE PROPOSED METHOD WITH THE EXISTING METHODS ON THE CEW DATASET

Paper	Method	Accuracy (%)	SENS (%)	SPEC (%)
Dong [36]	HOG + Random Forest	94.57	-	-
Song [27]	MultiHPOG + LTP + Gabor + SVM	94.72	-	-
Lei [11]	DINN	97.19	-	-
Proposed	CNN	99.03	99.19	98.90

The - refers to empty value i.e., the measure was not reported using the corresponding method.

TABLE III

PERFORMANCE COMPARISON OF DIFFERENT METHODS ON ZJU DATASET

Paper	Method	Accuracy (%)	SENS (%)	SPEC (%)
Wu [37]	LBP + SVM	90.37	-	-
Dong [36]	HOG + Random Ferns	94.76	-	-
Eddine [38]	Multi-TPLBP + MLP	95.18	-	-
Liu [13]	Gabor + LBP + HOG + SVM	95.42	-	-
Song [27]	MultiHPOG + LTP + Gabor + SVM	96.83	-	-
Lei [11]	DINN	97.20	-	-
Proposed	CNN	98.80	99.70	98.79

A. Experimental results on the CEW dataset

A comparison of the performance of the proposed EyeNet model with other existing methods with high accuracy on the CEW dataset has been reported in Table II. It is observed from Table II that the highest performance (accuracy: 97.19%) has been reported by Lei et al., [11] using Deep Integrated Neural Network (DINN) among the existing methods. However, the proposed EyeNet CNN model outperformed the existing methods with accuracy: 99%, specificity (SPEC): 98.90%, and sensitivity (SENS): 99.19%. Thus, improving the accuracy of [11] by about 2%.

B. Experimental results on ZJU dataset

There are six works with high accuracy (more than 94%) on the ZJU dataset. Among them, the DINN model proposed by Lei et al. [11] reported the best accuracy (97.20%). However, the proposed EyeNet CNN model improves significantly the recognition accuracy from 97.20% to 98.80% with SENS: 99.70%, and SPEC: 98.79%. Observing the results in Table II and Table III, it is seen that the performance of the proposed model on both CEW and ZJU datasets remains similar which is also true for DINN model of [11].

C. Cross dataset performance evaluation

Finally, to evaluate the robustness of the model with respect to completely unknown samples, we trained the model using MRL Eye data (i.e., the largest dataset) and evaluate the performance of the model for classifying samples of CEW and ZJU datasets. The model provided accuracy of 90.94% and 90.12% for the eye states classification of CEW and ZJU datasets, respectively. Thus, it seems that even though there are a large number of samples of the MRL Eye dataset (about

TABLE IV

PERFORMANCE OF THE MODEL ON CROSS-VALIDATION

Training Dataset	Testing dataset	Accuracy (%)
MRL Eye	CEW	90.94
	ZJU	90.12
ZJU	CEW	93.07
CEW	ZJU	67.34

TABLE V

PERFORMANCE THE PROPOSED MODEL ON THE MRL EYE DATASET UNDER DIFFERENT MODES

Dataset	Condition	Accuracy (%)	SENS (%)	SPEC(%)
MRL Eye	With glass	99.16	99.05	99.44
	Without glass	99.55	99.47	99.67
	Good light	99.05	99.16	99.00
	Bad light	99.27	99.44	99.29
	No reflection	99.18	99.24	99.16
	Low reflection	99.02	99.13	98.91
	High reflection	98.90	99.08	98.87
	Sensor RS300	99.58	99.83	99.21
	Sensor IDS	98.42	98.03	98.61
	Sensor aptina	99.36	99.12	99.61
	Average(total)	99.24	99.26	99.18

20 times of the CEW and 6 times of the ZJU datasets), the model fails to capture all information for sufficient training from the MRL Eye data, or in another word, the model is not trained sufficiently to capture the diversity presents in any unknown datasets.

The performance of EyeNet when trained with one dataset and evaluated on other datasets has been presented in Table IV. The maximum accuracy (93.07%) is obtained when the model is trained with the ZJU dataset and evaluated with the CEW dataset. Although the MRL Eye dataset is 6 times larger than ZJU, EyeNet provides more than 2% better accuracy for recognition of CEW when trained with ZJU instead of MRL Eye dataset. On the other hand, the lowest accuracy (67.34%) has been obtained for the recognition of ZJU when trained with CEW dataset. This might be due to its smaller size (samples of CEW might not possess sufficient variety to train the model).

D. Experimental results on the MRL Eye dataset

The performance of the proposed model on the MRL Eye dataset has been evaluated under 10 different modes for eye states classification, and the results have been summarized in Table V. The performance of the model does not vary significantly with respect to the modes of image acquisition. The lowest performances (accuracy: 98.42%, SENS: 98.03%, and SPEC: 98.61%) were obtained for images acquired through the IDS sensor. The overall performance of the EyeNet, computed as the average of the performances for each mode was accuracy: 99.24%, SENS: 99.26%, and SPEC: 99.18%.

V. CONCLUSIONS AND FUTURE WORKS

We presented a novel convolutional neural network model called EyeNet. Its performance was evaluated for eye states classification using three publicly available datasets. A number of good works have already been reported on these datasets.

However, the proposed EyeNet model outperformed all existing models. The three datasets used in this study contains challenging cases with significant illumination, appearance, reflection mode, different devices, and head pose variability. This permitted the evaluation of the model under different conditions (lighting, reflection, appearances, and devices), EyeNet significantly improved the performance of existing methods. Even though the MRL Eye dataset contains a very large number of samples compared to the other two datasets, but it does not capture all diversities to sufficiently train the model. A comparatively smaller ZJU dataset contains more versatility than the others. The purpose of the model can be extended for drowsy driving detection and controlling the electrical environment system through eye states information. In the future, we plan to extend the work real-time eye states classification from video data instead of static images.

ACKNOWLEDGMENTS

We acknowledge the contribution of NVIDIA corporation for providing a GPU card to do this research. In addition, we are also grateful to the King Faisal University for providing partial financial support for this study.

REFERENCES

- [1] O. Ossmy, O. Tam, R. Puzis, L. Rokach, O. Inbar, and Y. Elovici, "MindDesktop-Computer accessibility for severely handicapped," in *International Conference on Enterprise Information Systems*, 2011, pp. 316–320.
- [2] T. D. Pham and D. Tran, "Emotion recognition using the emotiv epoc device," in *International Conference on Neural Information Processing*, 2012.
- [3] R. Hussein, H. Palangi, R. Ward, and Z. J. Wang, "Epileptic seizure detection: A deep learning approach," *arXiv:1803.09848*, 2018.
- [4] TRAFFIC SAFETY FACTS by National Center for Statistics and Analysis (NHTSA). [Online]. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811449>
- [5] M. V. Yeo, X. Li, K. Shen, and E. P. Wilder-Smith, "Can SVM be used for automatic EEG detection of drowsiness during car driving?" *Safety Science*, vol. 77, pp. 115–124, 2009.
- [6] A. Królak and P. Strumiłło, "Eye-blink detection system for human-computer interaction," *Universal Access in the Information Society*, vol. 11, pp. 409–419, 2012.
- [7] W.-D. Chang, J.-H. Lim, and C.-H. Im, "An unsupervised eye blink artifact detection method for real-time electroencephalogram processing," *Physiological measurement*, vol. 37, p. 401, 2016.
- [8] W. O. Lee, E. C. Lee, and K. R. Park, "Blink detection robust to various facial poses," *Journal of neuroscience methods*, vol. 193, pp. 356–372, 2010.
- [9] C.-S. Hsieh and C.-C. Tai, "An improved and portable eye-blink duration detection system to warn of driver fatigue," *Instrumentation Science & Technology*, vol. 41, pp. 429–444, 2013.
- [10] B. Champaty, K. Pal, and A. Dash, "Functional electrical stimulation using voluntary eyeblink for foot drop correction," in *International Conference on Microelectronics, Communications and Renewable Energy*, 2013, pp. 1–4.
- [11] L. Zhao, Z. Wang, G. Zhang, Y. Qi, and X. Wang, "Eye state recognition based on deep integrated neural network and transfer learning," *Multimedia Tools and Applications*, vol. 77, pp. 19415–19438, 2018.
- [12] J. Mohanakrishnan, S. Nakashima, J. Odagiri, and S. Yu, "A novel blink detection system for user monitoring," in *Workshop on User-Centered Computer Vision*, vol. 1, 2013, pp. 37–42.
- [13] X. Liu, X. Tan, and S. Chen, "Eyes closeness detection using appearance based methods," in *International Conference on Intelligent Information Processing*, vol. 385, 2012, pp. 398–408.
- [14] B. Chambayil, R. Singla, and R. Jha, "EEG eye blink classification using neural network," in *World Congress on Engineering*, vol. 1, London, U.K., 2010, pp. 2–5.
- [15] H. Mandelkow, J. A. de Zwart, and J. H. Duyn, "Linear discriminant analysis achieves high classification accuracy for the bold fmri response to naturalistic movie stimuli," *Frontiers in human neuroscience*, vol. 10, p. 128, 2016.
- [16] R. Sun and Z. Ma, "Robust and efficient eye location and its state detection," in *International Symposium on Intelligence Computation and Applications*, vol. 5821, 2009, pp. 318–326.
- [17] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Eye-state action unit detection by gabor wavelets," in *International Conference on Multimodal Interfaces*, vol. 1948, 2000, pp. 143–150.
- [18] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps*, 2018, vol. 26, pp. 323–350.
- [19] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Science and Information Conference*, 2019, pp. 128–144.
- [20] S. Liu, B. Tang, Q. Chen, and X. Wang, "Drug-drug interaction extraction via convolutional neural networks," *Computational and mathematical methods in medicine*, vol. 2016, 2016.
- [21] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, pp. 1527–1554, 2006.
- [22] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *International Conference on Machine Learning*, 2008, pp. 1096–1103.
- [23] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner et al., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, 1998.
- [24] B. Li and H. Fu, "Real time eye detector with cascaded convolutional neural networks," *Applied Computational Intelligence and Soft Computing*, vol. 2018, 2018.
- [25] K. Kim, H. Hong, G. Nam, and K. Park, "A study of deep CNN-based classification of open and closed eyes using a visible light camera sensor," *Sensors*, vol. 17, p. 1534, 2017.
- [26] M. M. Rahman, M. S. Islam, R. Sassi, and M. Aktaruzzaman, "Convolutional neural networks performance comparison for handwritten bengali numerals recognition," *SN Applied Sciences*, vol. 1, no. 12, p. 1660, 2019.
- [27] F. Song, X. Tan, X. Liu, and S. Chen, "Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients," *Pattern Recognition*, vol. 47, no. 9, pp. 2825–2838, 2014.
- [28] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcam," in *International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007, pp. 1–8.
- [29] R. Fused, "Pupil localization using geodesic distance," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11241 LNCS, pp. 433–444, 2018.
- [30] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," 2008.
- [31] Hao Zheng, Zhanlei Yang, Wenju Liu, Jizhong Liang, and Yanpeng Li, "Improving deep neural networks using softplus units," in *International Joint Conference on Neural Networks*, Killarney, Ireland, July 2015, pp. 1–4.
- [32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Welcome to the Computing Research Repository*, vol. abs/1502.03167, 2015.
- [33] S. Park and N. Kwak, "Analysis on the dropout effect in convolutional neural networks," in *Asian Conference on Computer Vision*, vol. 10112, Seoul, Korea, 2016, pp. 189–204.
- [34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*, 2014.
- [35] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," *arXiv:1702.05659*, 2017.
- [36] Y. Dong, Y. Zhang, J. Yue, and Z. Hu, "Comparison of random forest, random ferns and support vector machine for eye state classification," *Multimedia Tools and Applications*, vol. 75, pp. 11763–11783, 2016.
- [37] Y.-S. Wu, T.-W. Lee, Q.-Z. Wu, and H.-S. Liu, "An eye state recognition method for drowsiness detection," in *Vehicular Technology Conference*, 2010, pp. 1–5.
- [38] B. D. Eddine, F. N. Dos Santos, B. Boulebtateche, and S. Bensaoula, "Eyelsd a robust approach for eye localization and state detection," *Journal of Signal Processing Systems*, vol. 90, no. 1, pp. 99–125, 2018.



Md. Moklesur Rahman received the B.Sc. and M.Sc degree in information and communication engineering from Islamic University, Bangladesh in 2015, and 2017, respectively. He is currently working as a lecturer at the Department of Computer Science and Engineering, The People's University of Bangladesh. His current research interests include machine learning, human-computer interaction, deep learning, pattern recognition, big data, data mining, natural language processing, computer vision, signal processing, and health informatics.



Md Shafiqul Islam has completed his both B.Sc. and M.Sc. in Computer Science and Engineering from Islamic University, Kushtia, Bangladesh. Currently, he is working as a lecturer at The People's University of Bangladesh. His research interests mainly include computer vision, machine learning and deep learning. He is also interested in Biomedical signal processing and wireless communication.



Mir Kanon Ara Jannat has completed both B.Sc. and M.Sc. in Computer Science and Engineering from Islamic University, Kushtia, Bangladesh. Currently, he is working as a lecturer at The People's University of Bangladesh. His research interests mainly include computer vision, machine learning, and deep learning.



Md. Hafizur Rahman is currently working toward the B.Sc degree with the Department of Electrical and Electronic Engineering, Islamic University, Kushtia, Bangladesh. His current research interests include Internet of Things, computer vision and machine learning.



Md. Arifuzzaman obtained his Ph.D. in Civil Engineering from the University of New Mexico (UNM), USA (2010), MS Engineering in Geotechnical and Geo-environmental Engineering from Asian Institute of Technology (AIT), Thailand and B.Sc in Civil Engineering from Bangladesh University of Engg and Technology (BUET). He is currently working as an assistant professor of civil Engg at KFUPM, Saudi Arabia. He is licensed with PE (professional engineer) in the state of NM, USA.



Roberto Sassi received the Laurea degree (B.Sc. and M.Sc.) in electronic engineering in 1996, and the Ph.D. degree in biomedical engineering in 2001, both from Politecnico di Milano, Italy. After graduating, he worked as Postdoctoral Researcher at the University of California, Santa Cruz, in mathematical modeling of non-Newtonian fluids and at Politecnico di Milano on ECG signals and atrial fibrillation. He is currently full Professor at the Università degli Studi di Milano, Italy. He taught courses of digital signal and image processing, statistics and programming. His research interests include biomedical signal processing, linear and nonlinear time series analysis, biometrics, with special emphasis on privacy protection issues, and applied mathematics. He is a senior member of IEEE.



Md Aktaruzzaman has completed his M.Sc. in Computer Science and Engineering from Islamic University, Kushtia, Bangladesh. In 2015, he achieved his PhD in computer science from the Dipartimento di Informatica, Università degli Studi di Milano, Italy. Currently, he is Professor at the department of Computer Science and Engineering in Islamic University, Kushtia, Bangladesh. His research interests mainly include regularity (or complexity) analysis of time series especially heart rate variability, feature extraction, and artificial intelligence. He is also interested in pattern recognition and computer vision.