Driver Drowsiness Detection in Real-time

Daksh Khetan, Arun Nawani, Anshul Aggarwal and Ms. Surinder Kaur*

Bharati Vidyapeeth's College of Engineering, GGSIPU, Delhi, INDIA
Emails; 2012daksh@gmail.com; arunnawani07@gmail.com; anshul.agg1510@gmail.com; kaur.surinder@bharatividyapeeth.edu

* Correspondence: kaur.surinder@bharatividyapeeth.edu

Abstract

In modern life, drowsiness is one of the major causes of road accidents, many of which are fatal. Analyzing statistics, it can be assumed that most road accidents occur as a result of drowsiness leading to serious injury and death. For this reason, various studies have been done on designing programs that can detect driver fatigue and alert them before a serious error occurs. This prevents them from falling asleep and having an accident. Some of the most common methods use automotive-based methods to design their own system. But these traditional measures were strongly influenced by other factors such as road structure, vehicle type and driver-wheel driveability. Some methods use psychological methods of their system that often provide the most accurate and consistent results in the driver's drowsiness monitoring. However, such techniques are very tedious as the electrodes need to be placed on the head and body. In addition, few studies are available where independent measurements are used as system installation, but such methods can confuse the driver and lead to unintended consequences. In this paper, we have proposed a non-disruptive and real-time program. Our proposed system classifies it as sleep deprivation. The model is fed with a large database of closed eyes and open eyes to produce results. The driver is notified by Buzz every time he is found drowsy. In our model, we use a standard forward-looking smartphone camera and use the information we have gained to produce results on our website. This can be more economical than using additional hardware.

Keywords: drowsy driving, facial landmark, image processing, face detection, eye detection, alert

1. Introduction

Among many other things associated with road accidents in this busy world, drowsiness is one of the major problems that need to be addressed. 1,100 people died as a result of sleep deprivation in 2015. It is estimated that some 79,000 road accidents occured annually due to sleep deprivation between 2005 and 2009.

This figure looks at about 886 deaths, 36,998 major injuries, and 45,000 physical injuries as a result of road accidents alone [1]. Drowsy driving often occurs when the driver is tired while driving, which makes them unable to stay alert. This often happens when the driver is not getting enough sleep or taking certain medications. This can also happen due to insomnia or sleep disturbances at work. This disturbs his sense of the road and basic understanding. In the worst-case scenario, the driver may fall asleep driving. Several attempts have been made to determine the driver's drowsiness by looking at different parameters. Current operating measures include sensors installed in various parts of the vehicle. Sensors are usually set on the accelerator and steering wheel, to measure the level of sleep [2]. The process of using car-based steps can also be divided into two categories. This sensor-based assessment can be based on two approaches, namely the Steering Wheel Movement (SWM) [3] and the Standard Deviation of the Route (SDLP) [4]. To measure SWM, an angle sensor is used to measure the driver's drowsiness

Doi : https://doi.org/10.54216/FPA.070203

Received: January 13, 2022 Accepted: April 09, 2022

based on his or her steering functions. Although SDLP uses an external camera that can be used to determine if the driver is drifting in his direction. However, this sensory-based assessment depends on a number of other factors. These factors include road structure, vehicle type and driver's normal driving style. In addition, the same methods are used to identify other sources of road accidents such as alcoholism. As a result, vehicle-based measurements fail to detect the cause of drowsiness. Apart from this, other methods use psychological methods to monitor driver fatigue by recording psychological symptoms using either (a) electroencephalogram (EEG) [5], (b) electrooculography (EOG) [6], electromyography (EMG) [7][9], or (d) electrocardiography (ECG) [8]. But there is one advantage to using psychological scales which is that analysis based on these parameters is able to predict the drowsiness of the driver with better accuracy as psychological signals are able to better measure the active perception of the driver. However, the sensors are not difficult to handle because they can confuse the driver. This may make the driver feel uncomfortable and distract him from driving. There are a few methods available that use independent measurements based on driver rating or questionnaires [10]. However, such tactics can lead to bizarre consequences as the interrogation method can alert the driver unconsciously, lowering the level of drowsiness. In addition, the most significant deviation from the independent measurement would be that a person may not be able to accurately predict his or her sleep level on the basis of self-examination in the real world. Some existing work has provided model-based movement tracking based on visual flow by analyzing eye position and driver position [6]. This method raises a high level of accuracy with generally low errors and false alarms for people of different races and genders. However, this has the disadvantages of requiring high computer proficiency and a side effect of sound sensitivity.

To summarize, there were a number of challenges that needed to be addressed:

i. Accurately identify the time of alarm sound

A major challenge for driver drowsiness is to accurately identify the driver's drowsiness while driving and sounding an alarm or warning at the right time because part of the delay in sounding the alarm or warning will increase the chances of an accident or failure.

ii. Featured image quality as input

Sometimes a portable webcam cannot detect and detect drowsiness due to its low-quality video capture capabilities and hence the accuracy of the model is compromised. As a result, there is a need for more and as a result, there is a need for cheaper equipment that can change and have better image capture capabilities as an additional webcam. The model is cost-effective while still maintaining a high-quality visual input for the neural network since the input can be taken directly from the driver's smartphone.

iii. Significant change or decreased accuracy with or without eyeglasses

Apart from that, detecting drowsiness in and out of the mirrors is another challenge that must be addressed, as it is clear that there is a significant loss of accuracy both outside and outside the mirrors during the driver's drowsiness.

iv. The frame decreases during image feeding

The feed from the webcam sometimes indicates a drop in frames or self-pressing which means that the output frame is sent to the model and the delay of parts of seconds and part of the alarm beep or warning delay will increase the chances of accident or failure.

2. Proposed System

The proposed plan aims to fill the gaps left by past transport and driver systems and to reduce the number of daily accidents around the world. The main goal of our system is to create a smart system that will be completely autonomous and would provide for the least hindrance to the driver owing to its independence from physical

Doi : https://doi.org/10.54216/FPA.070203

Received: January 13, 2022 Accepted: April 09, 2022

sensors. This is done by alerting the driver using a sound alert system when he is found drowsy and alerting the car owner to take immediate action. The model is trained with a data set that includes more than 50000 image samples of open and closed eyes to get the most accurate results. The model is very capable of distinguishing between closed eyes and open eyes in various light conditions. To avoid disturbing the drowsiness with normal blinking the system will keep a timer to track how long the eyes are kept close. The threshold value is set to exceed when the buzzer is sent to notify the driver. Smartphone Camera Module is used for continuous eye recording so that separation can be done in real-time. In addition, the counter is used to detect the number of times he or she has been drowsy. In our model, the counter limit was set to 15. If the counter exceeds this limit, the car owner will be notified to inform the driver that the driver is drowsy and it is no longer safe to continue driving.

3. Technique Used

The model is based on the principles of Transfer Learning. Transfer learning is a machine learning method in which a model designed for a task in the past is also used as the starting point for a second activity model. It is a popular method for in-depth learning where previously trained models are used as a starting point for computer vision and natural language processing activities. Transferring learning makes sense because of the great computing power and time resources needed to develop neural network models for such problems. They also produce very accurate results on related problems. Transfer learning is closely related to issues related to multi-task learning and mental retardation and is not the only learning area in deep learning. However, transfer learning is a popular form of in-depth learning considering the special computer capabilities required to train in-depth learning models or large and challenging data sets in which in-depth learning models are trained. This type of transfer learning used in deep learning is commonly known as Inductive Transfer. This is where the range of possible models (model bias) is reduced in a profitable way by using a model that is equal to another but related function. In our model, we will use the MRL eye database to be trained. The MRL eye data set contains more than 37,000 eye samples.

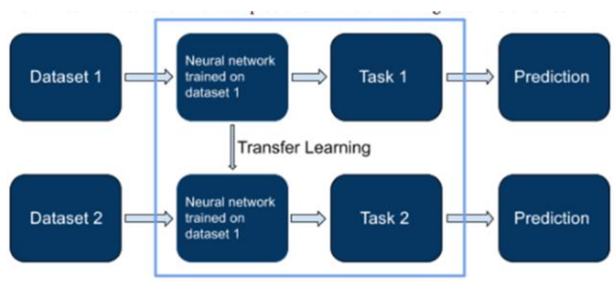

Fig. 1: Transfer Learning process flowchart

Architecture—In recent years, there has been a notable improvement in the field of study of representation with the introduction of multiple models such as Deep Belief Networks (DBN) [14], Deep Boltzmann Machine (DBM) [13], Restricted Boltzmann Machine (RBM) [16][17], Convolutional Neural Networks (CNN) [15], Recurrent Neural Networks (RNN) [18][19], Autoencoder [20] and others. The main reason for the success of the above-mentioned models is to learn to represent the features that can capture the most intelligent data in a non-labelled input database. Most of them use a wide range of subtle layers of learning complex, informal, high-quality presentations that also categorize high-level tasks.

We used MobileNet architecture to separate images. MobileNet is a model of CNN's architecture for image segmentation and Mobile Vision. But let us take a closer look at CNNs.

Doi : https://doi.org/10.54216/FPA.070203

Received: January 13, 2022 Accepted: April 09, 2022

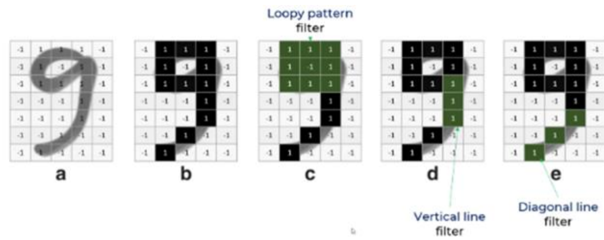

Fig. 2: Feature detection with filters in CNN.
Adapted from [21]

CNNs have been revolutionary in the field of Image related learning owing to their working principle. Let us have a look at the figure 2 to better understand CNNs. Here, we have a number 9, which we have divided into a grid filled with 1 and -1. We see that 1s in the grid make up 9 and empty spaces are filled with -1 in the grid. We identify three characteristic features of '9'. The loop, the vertical line in the middle and the short diagonal line that makes up the tail. So we have 3 filters, one for each feature. We will superimpose all these filters on an image if we want to detect it has a '9'. We took the loop filter (figure 2) we identified in number 9 and moved it over the grid of 1 and -1 of the image we wanted to check for. We map, multiply and then take the average of the numbers generated by superimposing the filter on the grid image. The filter is superimposed on the entire grid one by one. Think of the image as the parent matrix and the filter as a submatrix. Look at the image below. We multiply every cell in the filter to the corresponding submatrix in the grid image and then take the average of the sum of the numbers generated. In the case below, we do the following math:

$$1*(-1) + 1*1 + 1*1 + 1*(-1) + (-1)*1 + 1*(-1) + 1*(-1) + 1*1 + 1*1 = -1$$

Taking the average we get, -1/9 or - 0.11.

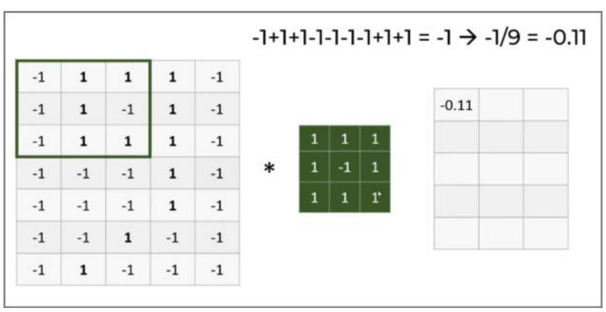

Fig. 3: Superimposing a filter on the image for feature detection

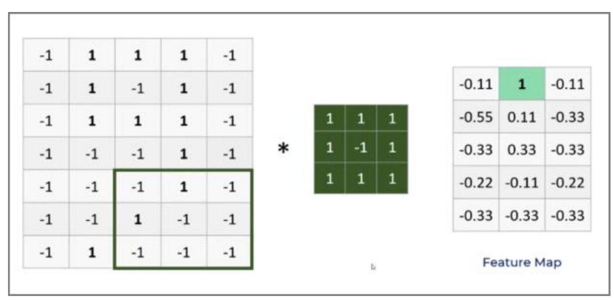

Fig. 4: Creating a feature map for each filter

So, for the submatrix that exactly matches the filter, the resultant is 1. This implies that there was a feature detection and the location of the detected feature can be identified in the resultant matrix. This resultant matrix is known as a feature map. A feature map is created for every filter we have. After this, we may have feature map aggregation to form another feature map that would essentially represent all the details of its component feature maps. Look at the example given for a better understand g.in

Doi : https://doi.org/10.54216/FPA.070203

Received: January 13, 2022 Accepted: April 09, 2022

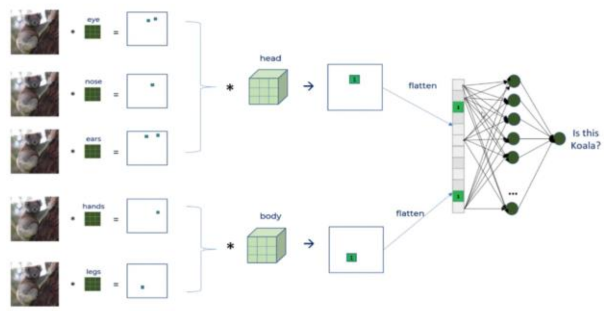

Fig. 5: Image detection by CNN, step by step.
Adapted from [22]

Now, say we want to train an ANN to detect a Koala. We will detect all the prominent features using the filters that we decided, namely - nose, eyes, and ears. These feature maps could be further aggregated to confirm the presence of a Koala Head. Similarly, legs and hands feature maps can be aggregated to derive the presence of a koala body. This results in two new feature maps, depicting the head and the body of the koala. The derived feature maps are 2D arrays which are converted to a single-dimensional array for the deep neural network to process the image. This process is known as 'Flattening'. This forms a fully connected neural network. The neural network can then classify the image based on the input it received from the feature map aggregation.

MobileNet architecture—MobileNet for vision and embedded mobile applications is proposed, based on a structured architecture that utilizes divisively intelligent convolutions to build deep neural networks. Introduced two simple global parameters that trade well between delays and accuracy. MobileNet's main layer is a highly sophisticated separate filter, named Depthwise Separable Convolution. Network design is another aspect of improving performance. Finally, the width and adjustment can be processed to trade between delays and accuracy.

Depthwise separable convolutions is a type of factorized complexity that resolve a standard to a depthwise convolution and a 1×11×1 convolution called a pointwise convolution. In MobileNet, the depthwise convolution apply one filter to each to each channel of input. The pointwise convolution then uses a 1×11×1 convolution to combine output with the depthwise convolution. The following figure shows the difference between standard convolution and depthwise separable convulsions:

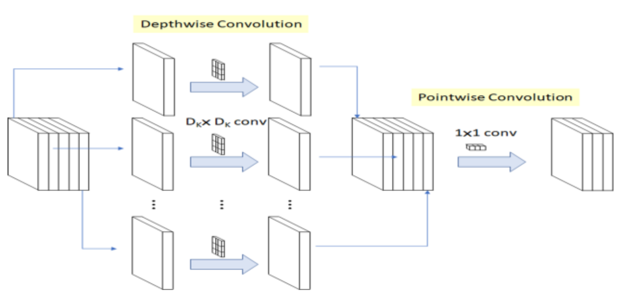

Fig. 6: Depthwise Separable Convolution.
Adapted from [22]

Doi : https://doi.org/10.54216/FPA.070203

Received: January 13, 2022 Accepted: April 09, 2022

4. Working Principle

i. Training the model using various random open and closed eyes images of people.

ii. Using OpenCV and Haar Cascade Classifier, detect the region of interest on the driver's face.

iii. We will then feed this to our model (based on Transfer Learning which is built using MobileNet architecture) for a binary classification of the input data ("Open eye" or "Closed eye").

iv. 0 will be the threshold value. A value greater than 0 will be considered an "open eye". If not, the "closed eye" may sound an alarm if the eyes remain closed for a specified period of time.

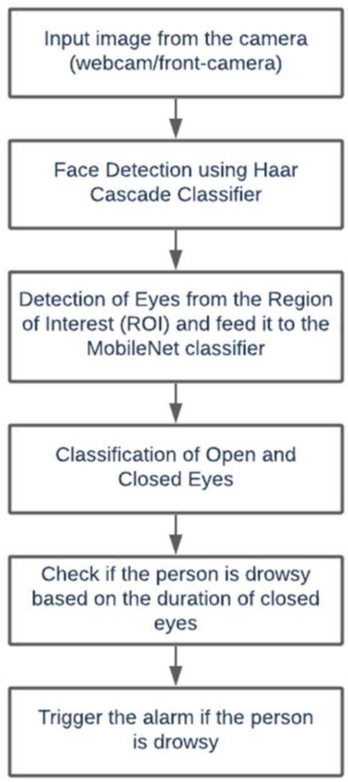

Fig. 7: Project flowchart

5. Result

The model can distinguish between closed eyes and open eyes with considerable accuracy.

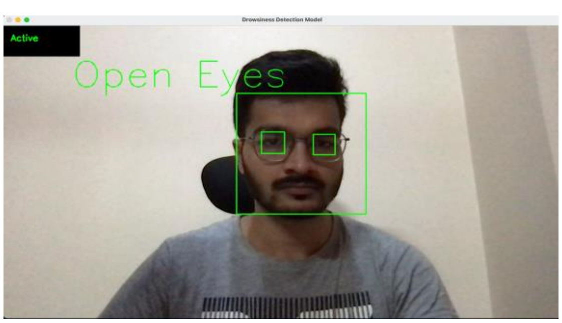

Fig. 8: Open eyes prediction sample

Fig. 9: Closed eyes prediction sample

Doi : https://doi.org/10.54216/FPA.070203

Received: January 13, 2022 Accepted: April 09, 2022

The model successfully raises a "Sleep Alert" whenever eyes are kept close for more than threshold time and the buzzer goes off to alert the drowsy driver.

Fig. 10: Sleep alert raised

The model we used for transfer learning is Tensorflow's MobileNet architecture. The model uses Haar Cascade classifier for facial detection and region of interest (i.e. eyes in this case). The model was trained with over 6000 images provided by the MRL eye dataset. The dataset has images of 37 different people (33 men and 4 women) of varying ages and ethnicity to provide versatility and better learning for the model. The final accuracy score achieved was over 98% which is in line with the principles of transfer learning.

6. Comparison with other popular techniques

S. No.	Model Name	Technique Used	Accuracy	Drawbacks
1	**Authors**: H. Varun Chand, J. Karthikeyan. **Research Paper**: CNN Based Driver Drowsiness Detection System Using Emotion Analysis **Publication**: Tech Science Press	Driver Behaviour and Emotion Detection using CNN.	93%	The model relies on human emotions rather than actual drowsiness indicators such as eyes or mouth. Also, the model has a relatively lower accuracy than ours.
2	**Authors**: Mohit Dua, Shakshi, Ritu Singla. **Research Paper**: Deep CNN models-based ensemble approach to driver drowsiness detection **Publication**: Springer	Detecting driver drowsiness using AlexNet, VGG-FaceNet, FlowImageNet, ResNet. The model studies various behavioural features such as hand and face gestures.	85%	The model takes into consideration a lot of factors before generating the results. However, the accuracy is still considerably low.

Doi : https://doi.org/10.54216/FPA.070203

Received: January 13, 2022 Accepted: April 09, 2022

3	**Authors**: Pradhan, Tapan & Bagaria, Ashutosh & Routray, Aurobinda. **Research Paper**: Measurement of PERCLOS using eigen-eyes **Publication**: Intelligent Human Computer Interaction	Detecting driver drowsiness studying the subject face and region of interest, which is eyes and mouth in this case and giving a prediction based on percentage eye closure - PERCLOS.	98%	Percentage eye closure are vulnerable to noisy image samples and requires ideal conditions to work optimally.

7. Future Scope

i. The current model takes into consideration a single factor to classify if a person is drowsy, i.e. if their eyes are closed or not. But the model can be further improved by monitoring yawning and nodding. Although monitoring yawning or nodding alone doesn't generate reliable and consistent results. But coupled with eye monitoring it could yield far more accurate results.

ii. Infrared cameras could be used to improve results in low light.

iii. The model can be further refined by giving input more and more data. More the data fed to the model, the more accurate and consistent the results would be. Currently trained for about 6000 images, the model has room for improvement.

iv. At the moment, there are zoom adjustments or camera orientation. But to yield better results, cameras could be programmed to zoom in on the region of interest once localized.

8. Conclusion

Average Eye Aspect Ratio (EAR) isn't consistent across the globe. People of Asian origin have relatively lower EAR than those in the West. The model doesn't need to be specifically optimized for different ethnicities, unlike EAR. Hence, models based on EARs aren't reliable enough for use all over the world. Deep Learning models provide more versatility by taking into account a variety of training data. Our model is trained on images of 37 different people in varying conditions like with and without glasses, lighting, and reflections to make it more suitable for real-world use cases. With a final accuracy score of over 98%, it's safe to say the model is highly reliable in varying testing conditions.

Funding: "This research received no external funding."

Conflicts of Interest: "The authors declare no conflict of interest."

Doi : https://doi.org/10.54216/FPA.070203

Received: January 13, 2022 Accepted: April 09, 2022

References

1. K. Dwivedi, K. Biswaranjan, and A. Sethi, "Drowsy driver detection using representation learning", in *2014 IEEE International Advance Computing Conference (IACC)*, 2014, pp. 995–999.
2. S. Sangle, B. Rathore, R. Rathod, A. Yadav, and A. Yadav, "Real Time Drowsiness Detection System", *IOSR Journal of Computer Engineering (IOSR-JCE),* pp. 87–92, 2018.
3. M. Shakeel, N. Bajwa, A. Anwaar, A. Sohail, A. Khan, and H. U. Rashid, *Detecting Driver Drowsiness in Real Time Through Deep Learning Based Object Detection*, 05 2019.
4. S. Mehta, S. Dadhich, S. Gumber, and A. Bhatt, "Real-Time Driver Drowsiness Detection System Using Eye Aspect Ratio and Eye Closure Ratio," *SSRN Electronic Journal*, 01 2019.
5. J. D. Fuletra and D. Bosamiya, "A Survey on Driver's Drowsiness Detection Techniques", *International Journal on Recent and Innovation Trends in Computing and Communication*, pp. 816–819, 11 2013.
6. R. Behera and K. Das, "A Survey on Machine Learning: Concept, Algorithms and Applications," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, 02 2017.
7. S. Podder and S. Roy, "Driver's drowsiness detection using eye status to improve the road safety", *International Journal of Innovation Research in Computer and Communication Engineering*, vol. 1, no. 7, 2013.
8. S. Junawane, S. Jagtap, P. Deshpande, and L. Soni, "Driver Drowsiness Detection Techniques: A Survey", vol. 6, no. 11, pp. 2015–2017.
9. T. Soukupova and J. Cech, "Real-time eye blink detection using facial landmarks", *Computer Vision Winter Workshop*, 2016.
10. F. Omidi and G. Nasleseraji, "Non-intrusive Methods used to Determine the Driver Drowsiness: Narrative Review Articles", *International Journal of Occupational Hygiene*, vol. 8, no. 3, pp. 186–191, 2016.
11. S. S. Nagargoje and D. S. Shilvant, "Drowsiness Detection System for Car Assisted Driver Using Image Processing", *International Journal of Electrical and Electronics Research ISSN*, vol. 3, no. 4, pp. 175–179, 2015.
12. Tejasweenimusale, Prof B, H. Pansambal, "Real Time Driver Drowsiness Detection System using Image Processing", *IJREAM*, vol. 2, no. 8, 2016.
13. R. Salakhutdinov and G. E. Hinton, "An efficient learning procedure for deep Boltzmann machines", *Neural Computation August 2012*, vol. 24, no. 8: 1967–2006.
14. G. E. Hinton, S. Osindero, and Y. Teh. "A fast learning algorithm for deep belief nets", *Neural Computation 18*:1527-1554, 2006.
15. Y. LeCun, Y. Bengio. "Convolutional networks for images, speech, and time series", *Handbook of Brain Theory and Neural Networks*, MIT Press. pp. 3361, 1995.
16. H. Chen, A. F. Murray, "Continuous restricted Boltzmann machine with an implementable training algorithm", Vision, Image and Signal Processing, *IEEE Proceedings*, vol. 150, no. 3, pp. 153-158, IET, 06 2003.
17. R. Salakhutdinov, A. Mnih, G. Hinton, "Restricted Boltzmann machines for collaborative filtering", *In Proc. of the 24th international conference on Machine learning*, pp. 791-798, ACM, 06 2007.
18. P. J. Angeline, G. M. Saunders and J. B. Pollack, "An evolutionary algorithm that constructs recurrent neural networks," in *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 54-65, 01 1994.
19. D. P. Mandic, J. Chambers, "Recurrent Neural Networks for prediction: learning algorithms, architectures and stability", *John Wiley & Sons, Inc*, 09 2001.
20. G. E. Hinton, R. S. Zemel, "Autoencoders, Minimum Description Length, and Helmholtz Free Energy", *Advances in Neural Information Processing Systems*, pp. 3-3, 1994.
21. D. Patel, "Explanation of Convolutional Neural Network - Deep Learning Tutorial (Tensorflow & Python)," Accessed on: Sep. 27, 2021. [Online]. Available: https://youtu.be/zfiSAzpy9NM
22. S. Tsang, "MobileNetV1 — Depthwise Separable Convolution (Light Weight Model)," Accessed on: Sep. 30, 2021. [Online]. Available: https://towardsdatascience.com/a382df364b69

Doi : https://doi.org/10.54216/FPA.070203

Received: January 13, 2022 Accepted: April 09, 2022