ISSN 2278-3091
Volume 9 No.2, March - April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse132922020.pdf
https://doi.org/10.30534/ijatcse/2020/132922020

Automated Road Safety Surveillance System using Hybrid CNN-LSTM Approach

Babitha D[1], Mohammed Ismail[2], Subrata Chowdhury[3], Ramya Govindaraj[4], Kolla Bhanu Prakash[5*]

[1,2,5] Department of CSE, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India,
donepudi.babitha@gmail.com, mdismail@kluniversity.in , drkbp@kluniversity.in
[3] Department of Computer Science, VISTAS Chennai, Tamil Nadu, India, subrata895@gmail.com
[4] School of Information Technology and Engineering, VIT-Vellore, Tamil Nadu, India , ramya.g@vit.ac.in

ABSTRACT

Among all the reasons for the occurrence of road accidents, the human state of drowsiness and Underage driving contributes a major share. With the rigid implementation of traffic rules and national schemes, it does not result in decreasing the accidents. Hence, there is a need for automation of surveillance which strictly restricts the teen driving and fatigue driving. In this paper, we introduce a face image descriptor-based combination of deep learning model i.e., convolution neural network (CNN) with ResNet50 architecture to predict age and a recurrent neural network (RNN) with LSTM architecture to detect the drowsiness in driver and alert them when they are drowsy. The algorithm is based on face recognition for age prediction and the blink frequency for detecting the fatigue. Image processing techniques are utilized to obtain the feature-based extraction for prediction. The proposed model developed could give a validation accuracy of 96% thus providing the promising results. This automation model thus could help the road safety authorities in their work and also decreases the occurrence of road accidents.

Key words : Age Prediction, Convolution Neural Networks (CNN), Drowsiness Detection, Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), Residual Network (ResNet50), Deep Learning based Road Safety Surveillance Automation (DRSA)**.**

1. INTRODUCTION

With the tremendous increase in the number of road accidents in the past decade, the implementation of road safety has become a major issue especially in the major cities. Most of the hospitalization cases leading to loss of life are majorly due to the road trauma. Collisions among vehicles and ignoring the safety measures [19] results in road accidents. It is observed that major causes of road accidents are fatigue while driving, minors driving the vehicles without having proper knowledge on road safety rules. Hence, preventive measures must be taken to obstruct both the causes so as to prevent loss of life. In addition to this, surveillance [20] of the road safety has become a major issue in these days. It is highly tedious to check the implementation of rigid rules regarding the road safety. Hence, with the improvement in technology, the combination of advanced fields like Pattern recognition and Image processing [21]-[23] can be implemented to provide automation of road safety. Underage driving with high speed is causing many fatal crashes. It also keeps their family under emotional rollercoaster when their lives are involved in a car crash. The alertness in driving can also be impaired by sleepiness. Loss incurred by drowsy driving is as equal to that of the aggressive driving. Drowsiness can be predicted by features such as frequent blinks in the eye, yawning and deviation in the head posture etc.

In this paper, we applied deep learning techniques, a hybrid model of CNN [1] and LSTM [2] to provide automated road safety surveillance by strictly implementing the rules related to road safety. The deep learning model we proposed is trained such that it allows only the majors (above 18 years old) to drive a vehicle and alerts the driver if he is drowsy so it can prevent the occurrence of road accident. The age and drowsiness are estimated using the age prediction [3] and drowsiness detection phases in the DRSA model.

This paper is organized as follows. Section 2 addresses the associated work related to the algorithms applied and the techniques implemented to detect age and gender. Section 3 outlines the techniques implemented in our research and explains the phases of analysis implemented. Section 4 demonstrates the analysis of the findings produced in our research. Finally, Section 5 summarizes the findings; research conducted in our review, and offers a conclusion.

2. RELATED WORK

J. Xie and C. Pun [1] proposed two ensemble methods in deep learning techniques by applying CNN as the base learner for age prediction and the ordinal relationship is implemented among the weak learners to get the final predicted age. The kind of CNN architecture used in both the models is ResNet101. It is applied with two and three groups classification respectively [17]. Ordinal regression is implemented for estimating the age in both the models from the images of the people's face. Research done by N. Yu et.al., [2] developed a three-stream deep CNN model for estimating the age in non-ideal images. Image processing techniques are applied to pre-process the non-ideal images in three different streams and DCNNs[16] are used to extract the features from those images for accurate age prediction. Ensemble learning is applied based on the models from the three streams as the weak learners.

S. Chen et.al.[3] proposed a deep novel framework to estimate the age from images. A new algorithm ranking-CNN is implemented which contains a set of basic CNNs which are pre-trained. Aggregation is applied on the binary output from the basic CNNs to get the final outcome. The proposed model outperformed all the other standard algorithms applied on several benchmark datasets. Work done by C. Li et.al., [4] developed a new feature-selection technique for age estimation. Information regarding each feature, such as the local structure and ordinal information, is collected from the facial images. Then correlation [15] among the features is minimized and the semi-supervised prediction algorithm is implemented and the results are compared with the state-of-the-art algorithms.

Survey done by M. Ramzan et.al., [5] made a comparative study of various state-of-the-art algorithms in drowsiness detection. The results show that a good combination of EEG and ECG features gives highly accurate performance. A Study done by M. A. Tanveer et.al., [6] applied deep learning algorithms and near-infrared spectroscopy to apply on the brain-computer interface to estimate the fatigue. CNNs were applied on the brain colormap images to find the best channels for fatigue detection. F. Lin et.al., [7] applied neural fuzzy method on the EEGs, as they are correlated to the behavioural attributes of the driver, to detect the fatigue in the driver. The self-organised [14] fuzzy system is then compared with the other neural network models developed.

3. METHODOLOGY

The DRSA consists of a camera component which is placed in the vehicle in front of the steering at a position where it can capture the video of the face of the person who is in driving seat. The continuous stream of video is monitored by the technology embedded in DRSA. The video is captured and every frame is processed and face detection algorithms such as Haar Cascade Classifier are applied to detect the face region in every frame. It is then passed to the next phase for DRSA. The entire process can be described using the representation of a flowchart.

The Figure 1 shows the implementation methodology of the DRSA system. The entire process can be divided into 2 phases. First phase is regarding the age of the driver and whether to allow a person to drive or not. Second phase deals with fatigue of the driver and calculating the drowsiness index. In both the phases, alerts raise when the violations occur. The main component is the DRSA which automates the predictions in both phases by applying deep learning models. CNN architecture ResNet50 and RNN [24] architecture LSTM are implemented to serve this purpose. The entire process is implemented using the tensorflow background.

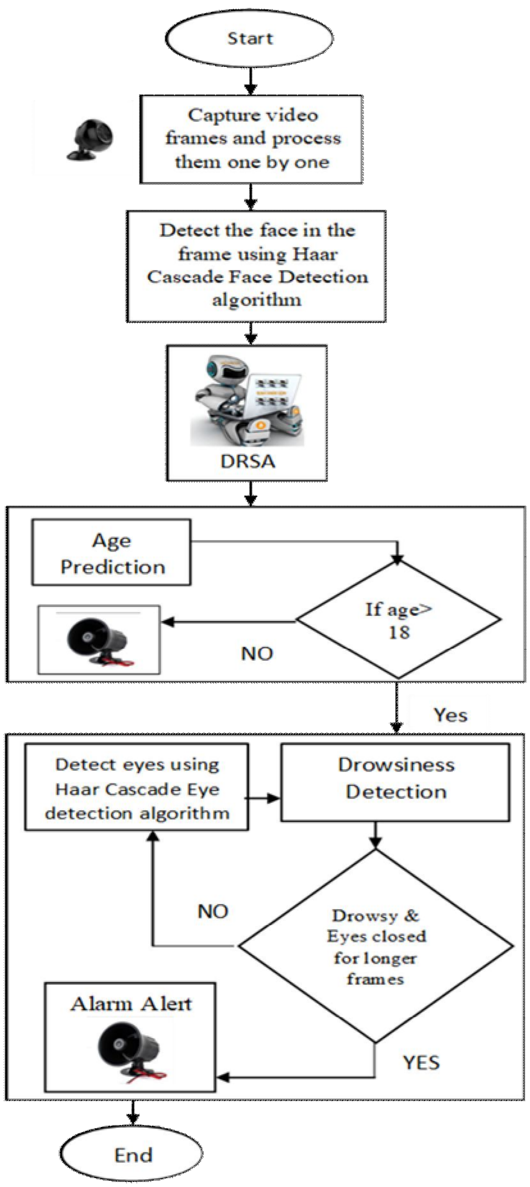

Figure 1: Flowchart of DRSA implementation

3.1 Experimentation

The experimentation process in our study is carried out by fetching or creating the dataset, pre-processing the data, generating the trained data, training the network, modeling and then evaluating the models developed. The entire process can be represented diagrammatically in Figure 2.

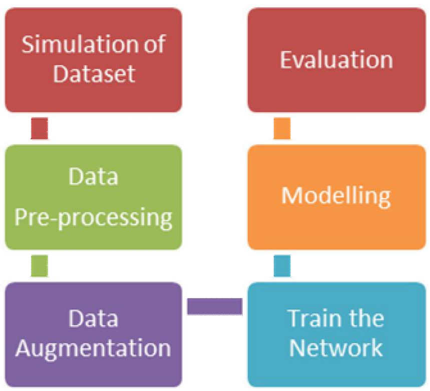

Figure 2: Process flow of the Experimentation

3.2 Dataset

Two different datasets are used in our research, one for training the model to predict age of the people and the other one for detecting the drowsiness.
UTKFace Dataset:

The UTKFace dataset shown in Figure 3 has been used for age prediction. It has around 20,000 images of people of different age groups ranging from 0 to 116 years. All the images are of different resolutions and are posed at different angles and illumination. Each image is labeled using age, ethnicity and gender.

Figure 3: Sample images from UTKFace dataset

3.3 Simulation of Driving Dataset

For the second phase, we created our own dataset of images by capturing 3 videos each of 8 people. This data has a good combination of the people who are either drowsy or awake during driving [8] and the subjects are of different age groups and gender. All the participants involved are above 18 years old. The videos were self-recorded by the subjects using webcam device. Videos were captured at different lighting and background scenarios [9] of which few subjects wore glasses. The data is then divided into 16 videos for training and 8 videos for testing [10]. The frames in all the videos are stored and labeled as Drowsy or Alert as shown in Figure 4.

Figure 4: Sample images from the dataset simulated

3.4 Pre-Processing

Hence the face images are cropped and normalized to a specified ratio to extract better features. Computer vision techniques are implemented to detect the faces in the sequence of frames in the video captured. The frame images are converted to gray scale and are size enhanced for further processing. The face is captured in the entire frame and stored. The images may have noise in them due to several factors like lighting, camera etc. Hence, noise removal has to be applied to get clear images which otherwise would affect the accuracy of the model. Similarly, the eyes features are extracted from the face image only if the predicted age of the person is above the specified limit. Haar Cascade Classifier algorithms [25] are used to detect the respective features in the face image. Then the eye-aspect ratio is calculated to detect the blinking of the eye.

3.5 Data Augmentation

As the training data is small, it may result in overfitting and less performance [11] for unseen test tuples. Hence, we apply the strategy of Data Augmentation which increases the diversity of data to train the neural networks. Several techniques like cropping, flipping can be applied on the original data to train the model. In our experiment, we applied shear angle transformation, width shift and cropping techniques. This helps in increasing the validation accuracy.

3.6 Train Network

The network architecture uses the pre-processed [12] data that is generated as the output of the previous phase. Adam optimization was implemented to adjust the weights of the network and also the learning rate. The hyper parameters are tuned for number of epochs and the number of steps for every epoch using the grid search. The tuning is performed to train the model which gives the least validation loss value as the final model [13]. As it is observed that the network was over fitted, it can be tested again by training the network with the augmented data.

3.7 Modeling

For image data, deep learning architectures works well and gives good results compared to the standard machine learning algorithms. In our study, we applied ResNet50, which is a deep-residual based architecture and LSTM [18], a model of RNN architecture. CNN is widely applied for image classification tasks whereas RNN is preferred for sequential data. As our DRSA system requires face recognition and age prediction, CNN could give accurate outcomes. For detecting the drowsiness, in addition to the current frames, the data of previous frames has also to be considered to calculate the drowsiness, hence LSTM solves the issues in it. In the modelling phase, the 2 tasks are implemented using single architecture and evaluated. It is then compared with applying the best model for each phase of detecting age using ResNet50 and drowsiness using LSTM. After finding whether both the eyes are open or closed, the number of frames of same label of eyes will be counted to calculate the blinking duration of the eyes so as to calculate the fatigue score.

3.8 ResNet-50

ResNet-50 is the short form for residual neural network which is a type of CNN architecture model used to classify and detect images. It is mainly used for huge datasets with more than 1000 images. It is a deep residual layered network model with around 50 layers and about 23 million training parameters. It applies the concept of skip connection to propagate information to deeper layers. Some of the similar training models are AlexNet, GoogleNet and VGG16. To overcome the vanishing gradient problem, it has identical connections. This model contains 5 stages each having both the convolution block and identity block. Both the blocks have 3 convolution layers each. Because of its high accurate predictions, this model is widely utilized in computer vision applications.

The data is spilt into training and validation sets with the test ratio 0.1. Global Max Pooling followed with 2 dropout and dense layers are also implemented. ModelCheckpoint is applied to save the best model with the highest validation accuracy by applying the EarlyStopping callback with the same metric. A patience value of 50 epochs has been set.

3.9 LSTM

LSTM is a kind of RNN architecture which can handle the problems encountered while training the simple RNN. It can handle exploding gradient and gradient vanishing issues. LSTMs use special units that can store information for long time in memory. They give accurate results when the data has a long sequence. In our experiment, we trained the LSTM-4096 wide layer with batch-size of 4, 3 dense layers and one dropout layer and finally a dense layer with 2 outputs. The model is trained for 10 epochs using the Adam optimizer with a learning rate of 0.0001 to get the optimal weights for the network.

3.10 Evaluation

Evaluation in DRSA has to be carried out for every phase. For the first phase, cross validation accuracy is calculated for the validation set to test whether the model developed can predict the age accurately. The calculated age and the original age are compared to find the residual error. Once the accuracy is high and the error rate is low, the model is considered good. It can be further used to predict unseen faces of real-time scenario.

In the second phase of drowsiness detection, the evaluation can be done by checking the status of eyes and the score of drowsiness. Models are evaluated based on the above metrics to select the better performing one. It also has to be checked whether the alarm sound is playing when the drowsiness has crossed the threshold limit. Thus, in both the cases, the metrics evaluated are accuracy and loss value.

4. RESULTS AND DISCUSSION

As the first pre-condition to satisfy for our proposed system is the age limit of the driver, we have to detect his age and then permit him only if his age matches the range allowed. To detect the age of the person, our DRSA system extracts the features from each frame of the image captured from the device installed in the vehicle. The pre-trained models will then predict the age of the driver. Based on the result, the DRSA will further take decision based on its rule system.

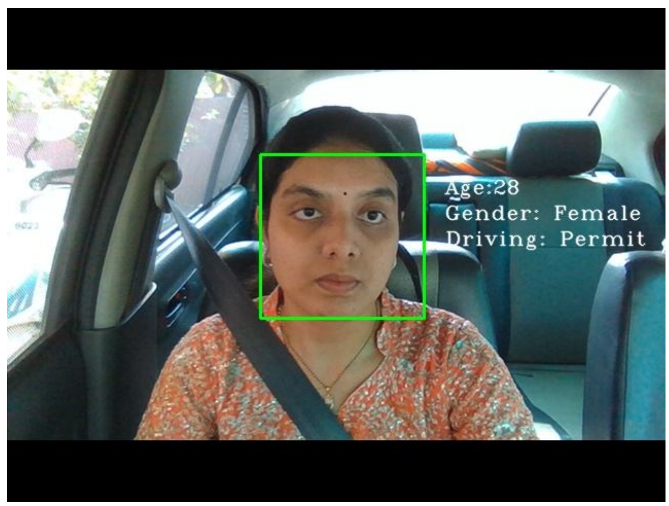

Figure 5: Verification of Phase 1 in DRSA model

The Figure 5-8 shows the results of age and gender predicted by the DRSA and the status whether he/She is permitted to drive or not. As the result in this case is positive, the person can proceed to drive further.

As one of the main objectives of DRSA is to prevent minors from driving the vehicles, the test case above shows the result that a child of 2 years if attempted to drive is not permitted. If he tries to, then an alarm alert raises and an automatic message is reported to the respective authorities associated with road safety. An additional alert message is transmitted to the official contact number linked to the corresponding vehicle unique ID. Thus, DRSA can prevent novice and authorized persons from driving the vehicles resulting in road accidents.

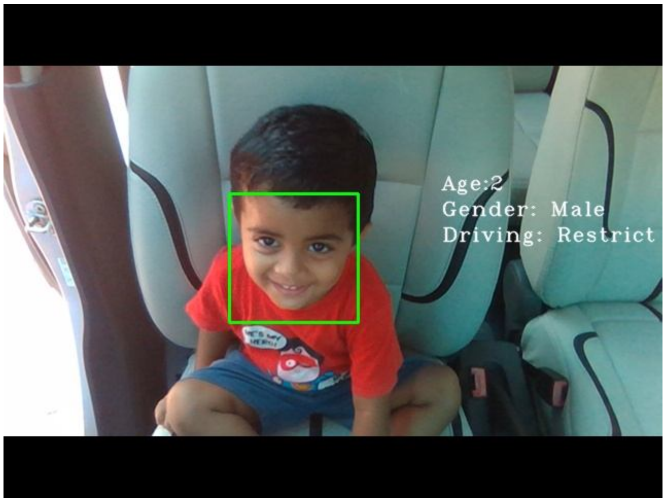

Figure 6: DRSA alert for Minors

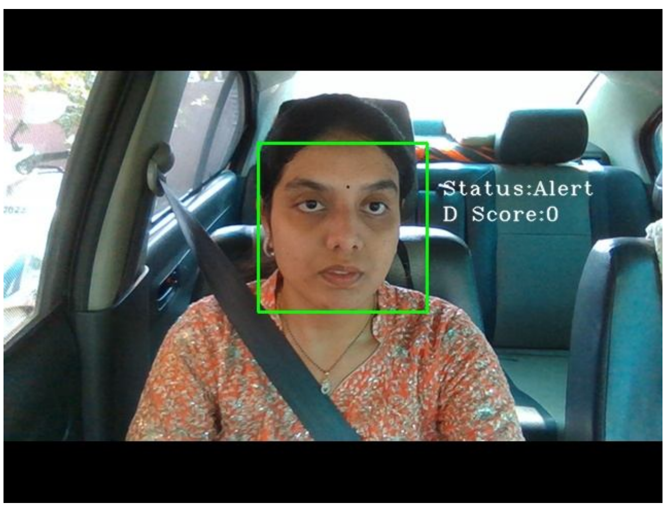

Figure 7: Evaluation of Phase 2 in DRSA

The scenarios where the criteria in phase 1 are successfully satisfied, then the phase 2 conditions will be evaluated. It can be seen from the above picture that when the person is fully conscious and is driving safe with his both eyes open and with drowsiness less than the specified threshold, then the DRSA predicts the case as completely awake and the results are displayed as above. The same procedure is repeated for every sample of the live stream video data and if the conditions fail, then the other set of rules fires out.

It is common for the persons who drive to get drowsy during long drives and thereby causing major accidents which can result even to loss of life. Hence, to prevent such road accidents, DRSA alerts the driver when the drowsiness score increases beyond the limit. An alarm sound plays and automatic switch on of light in the car is made. This can make the person awake and get normal. If still, he is drowsy, automated alerts were sent similar to the alert system in phase 1.

The Figure 9 shows the curve of accuracy of both training and validation during different epochs. It can be observed that the training accuracy is high compared to the validation accuracy. The model can thus predict both the age and drowsiness with a good accuracy of 96%. Thus, deep learning gave reliable predictions with good accuracy.

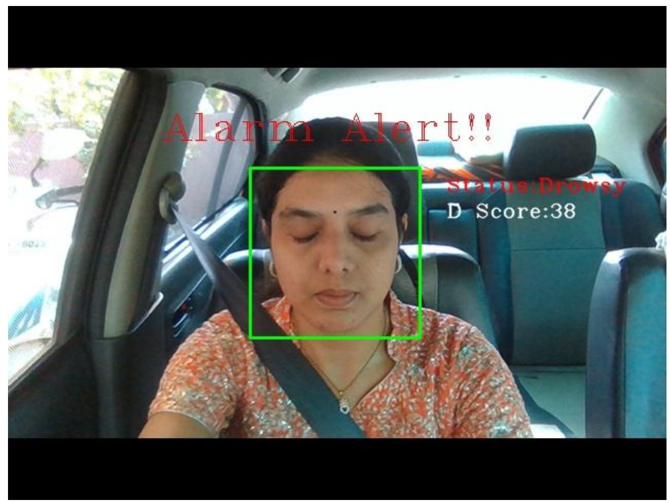

Figure 8: Alert from DRSA to prevent road accident

Figure 9: Accuracy Percentage Vs Number of epochs

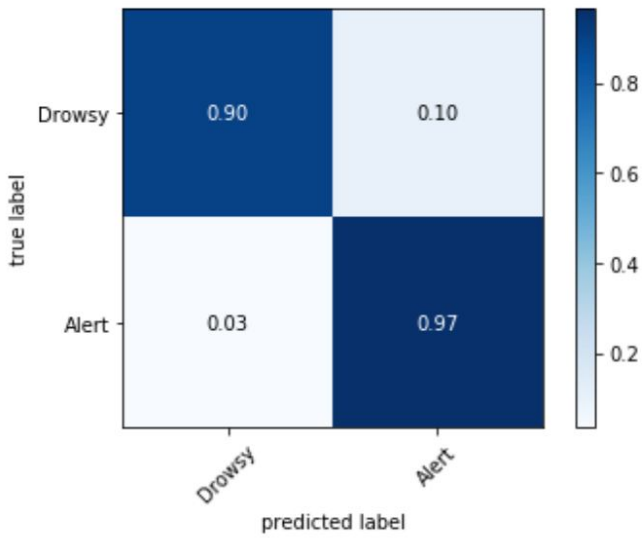

Figure 10: Confusion Matrix of CNN-LSTM model

The Figure 10 confusion matrix shows that the model could detect the alert cases with good accuracy compared to the drowsy scenarios. It failed to detect few cases which are alert by treating them as drowsy because of some arbitrary conditions.

5. CONCLUSION

In this paper, we have thus implemented the proposed method of applying deep automated road safety surveillance efficiently. The hybrid model of ResNet50 and LSTM can detect the age and fatigue of the driver with an accuracy of 96% by applying the computer vision algorithms. It can thus alert the drivers from causing road accidents. The accuracy can be further improved by implementing more deeper networks which can extract features to give reliable predictions.

REFERENCES

1. X. Wang, L. Gao, J. Song and H. Shen, **Beyond Frame-level CNN: Saliency-Aware 3-D CNN With LSTM for Video Action Recognition**, IEEE Signal Processing,Vol. 24, no. 4, pp. 510-514.2017. https://doi.org/10.1109/LSP.2016.2611485
2. T. Li, M. Hua and X. Wu, **A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5),**IEEE Access, Vol. 8, pp. 26933-26940,2020.
3. S. Wang, D. Tao and J. Yang, **Relative Attribute SVM+ Learning for Age Estimation**, IEEE Transactions on Cybernetics,Vol. 46, no. 3, pp. 827-839,2016. https://doi.org/10.1109/TCYB.2015.2416321
4. J. Xie and C. Pun, **Deep and Ordinal Ensemble Learning for Human Age Estimation from Facial Images**,IEEE Transactions on Information Forensics and Security, Vol. 15, pp. 2361-2374,2020. https://doi.org/10.1109/TIFS.2020.2965298
5. N. Yu, L. Qian, Y. Huang and Y. Wu, **Ensemble Learning for Facial Age Estimation Within Non-Ideal Facial Imagery**, in IEEE Access, 2019,Vol. 7, pp. 97938-97948.
6. S. Chen, C. Zhang and M. Dong, **Deep Age Estimation: From Classification to Ranking**, IEEE Transactions on Multimedia, Vol. 20, no. 8, pp. 2209-2222,2018.
7. C. Li, Q. Liu, W. Dong, X. Zhu, J. Liu and H. Lu, **Human Age Estimation Based on Locality and Ordinal Information**, IEEE Transactions on Cybernetics,2015 , Vol. 45, no. 11, pp. 2522-2534. https://doi.org/10.1109/TCYB.2014.2376517
8. M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas and A. Mahmood, **A Survey on State-of-the-Art Drowsiness Detection Techniques**, IEEE Access,Vol. 7, pp. 61904-61919,2019
9. M. A. Tanveer, M. J. Khan, M. J. Qureshi, N. Naseer and K. Hong,**Enhanced Drowsiness Detection Using Deep Learning: An fNIRS Study**, IEEE Access, Vol. 7, pp. 137920-137929,2019
10. F. Lin, L. Ko, C. Chuang, T. Su and C. Lin, **Generalized EEG-Based Drowsiness Prediction System by Using a Self-Organizing Neural Fuzzy System,** IEEE Transactions on Circuits and Systems I: Regular Papers, Vol. 59, no. 9, pp. 2044-2055,2012 https://doi.org/10.1109/TCSI.2012.2185290
11. K.B., Prakash. **Information extraction in current Indian web documents**. International Journal of and Technology(UAE), 2018,pp. 68-71.
12. Kolla B.P., Dorairangaswamy M.A., Rajaraman A. **A neuron model for documents containing multilingual Indian texts**. *International Conference on Computer and Communication Technology*, *ICCCT-2010*, 2010 pp. 451-454.
13. Prakash K.B., Dorai Rangaswamy M.A., Raman A.R. **Text studies towards multi-lingual content mining for web communication**. *Proceedings of the 2nd International Conference on Trendz in Information Sciences and Computing*, *TISC-2010*, 2010 pp. 28-31. https://doi.org/10.1109/TISC.2010.5714601
14. Kolla B.P., Raman A.R. **Data Engineered Content Extraction Studies for Indian Web Pages**.,Advances in Intelligent Systems and Computing, 2019: 505-512.
15. Prakash K.B., Dorai Rangaswamy M.A. **Content extraction studies using neural network and attributegeneration.** Indian Journal of Science and Technology, pp. 1-10,2016.

16. Prakash K.B., Rangaswamy M.A.D. "**Content extraction of biological datasets using soft computing techniques**, Journal of Medical Imaging and Health Informatics, pp. 932-936,2016. https://doi.org/10.1166/jmihi.2016.1931
17. Prakash, K.B., Rajaraman, A., Perumal, T. & Kolla, P. 2016, **Foundations to frontiers of big data analytics**, *Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, IC3I 2016*, 2016,pp. 242
18. Kolla, B.P. & Raman, A.R. 2019, **Data Engineered Content Extraction Studies for Indian Web Pages,**Advances in Intelligent Systems and Computing, pp. 505-512. https://doi.org/10.1007/978-981-10-8055-5_45
19. Prakash, K.B. 2017, **Content extraction studies using total distance algorithm**, *Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016*, 2016, pp. 673.
20. Prakash, K.B., Kumar, K.S. & Rao, S.U.M. 2017, **Content extraction issues in online web education**, *Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016*,2016, pp. 680.
21. Ismail, M., Prakash, K.B. & Rao, M.N. 2018, **Collaborative filtering-based recommendation of online social voting**, International Journal of Engineering and Technology(UAE), vol. 7, no. 3, pp. 1504-1507. https://doi.org/10.14419/ijet.v7i3.11630
22. Prakash, K.B., Rajaraman, A. & Lakshmi, M. 2017, **Complexities in developing multilingual on-line courses in the Indian context**, *Proceedings of the 2017 International Conference On Big Data Analytics and Computational Intelligence, ICBDACI 2017*, 2017 ,pp. 339. https://doi.org/10.1109/ICBDACI.2017.8070860
23. H.I. Okagbue, N.J. Peter, A.O. Akinola, C.O. Iroham, A.P. Opoko, "Smart Review of the Application of Genetic Algorithm in Construction and Housing", **International Journal of Advanced Trends in Computer Science and Engineering**, Vol 9(1) (2020), pp: 266-273, DoI: https://doi.org/10.30534/ijatcse/2020/ 40912020.
24. Bandara, Ravimal. “Nature Inspired Dimensional Reduction Technique for Fast and Invariant Visual Feature Extraction.” **International Journal of Advanced Trends in Computer Science and Engineering**, Vol 8(3) (2019), pp: 696–706. DoI: 10.30534/ijatcse/2019/57832019
25. Bharat, Vinod & Malik, Dr. Study of Detection of Various types of Cancers by using Deep Learning: A Survey. **International Journal of Advanced Trends in Computer Science and Engineering**, Vol 8, ,2019, pp: 1228-1233, DoI: 10.30534/ ijatcse/2019/ 31842019.