Comparative Performance Analysis of KNN (K-Nearest Neighbor) algorithm for Fatigue Detection in vehicular Drivers using pulse oximetry signal

Amit Chauhan[1], Anil Saroliya[2], Varun Sharma[3]
[1,2,3]Amity School of Engineering & Technology
Amity University Rajasthan, India
[1]amit.amity.edu@gmail.com,[2]asaroliya@jpr.amity.edu,[3]vsharma@jpr.amity.edu

Abstract: Vehicular accidents are taking place day by day and people are losing their lives across the world. It is important to detect physiological and Psychological parameter of the vehicular driver so that various levels of physical and mental fatigue can be measured. Research has been initiated at a great extent in India to develop an insidious solution for eliminating such accidents. In this paper the KNN (K-nearest neighbor) algorithm has been designed to organize Post and Pre driving fatigue levels. The performance analysis of K- nearest neighbor algorithm has been discussed here. This paper propose a method to plan and to execute the major element of this computing system which is wearable and the aim is to watch the Fatigue level of a driver using physiological parameters so that a simplified system can be built without disturbing the driver and unaffecting his comfort level. This paper also presents the usage of physiological parameters of the body such as oximetry pulse for classifying the post and pre driving fatigue state of the vehicular drivers.

Introduction: Fatigue is associated with both reduced capacity to perform and motivation to perform. Fatigue is defines a state of the person that results of sustained activity wherein he or she declares being unable to continue the activity further. It is integration of the behavioral physiological and emotional factors that can result in chronic physical or mental states. Some systems have been developed to detect and monitor the drowsiness, alertness and stress level of the vehicular Drivers which is based on some parameters like Electrooculograph (EOG), Electrocardiograph (ECG), Electroencephalograph (EEG) and some set of physiological parameters like Skin Conductance, Oximetry Pulse, Respiration, SPO2, and the current work focuses on the first two parameters to detect and monitor the mental fatigue / drowsiness of a driver. Using KNN (K-nearest neighbor) algorithm has been designed to organize Post and Pre driving fatigue levels. some simple physiological signal with cognitive fatigue in vehicular driver as there exist many tiny and simple physiological sensors that can be embedded in wearable's of the vehicular driver and may not lead to distraction. The physiological sensor signals and their features extracted were used as input to various classification techniques using Neural Networks, SVM, Fuzzy systems, and statistical methods so on.

In this paper we included the details of data collection, design & implementation of two different version of the KNN algorithms and tests results. The comparative analysis of all the algorithms for specific fatigued data has been carried out.

Data Used: The details of data used, features extracted and performance measures used for performance analysis of KNN algorithms used for of pulse oximetry signal for fatigue detection in vehicular drivers. The fatigue detection in vehicular drivers, real time physiological sensory data   related to fatigue state and

normal state, collected and used for further processing. The data recorded of Normal state driver (Pre-driving) and fatigued state (Post-driving) in text file format.

The two state signals of pulse oximetry were processed for extraction of predefined features using MATLAB. Two feature files were extracted from Pulse Oximetry signal. Feature Set I (Biorlet Wavelet with level 4 decomposition features **File -OP_BIOR1.1_L4**), Feature Set II (D'Meyer Wavelet with level 8 decomposition Feature File- **OP_DMey_L8**). Feature set I includes 51 features of Biorlet wavelet at decomposed to level 4 whereas Feature Set II includes total 85 features of D'Meyer wavelet decomposed to level 8. The details of feature vectors and their size are mentioned in **Table 1**.

Table 5.1: Details of Input parameter

Physiological Signal	Feature File Name	No. of Feature Vectors	Size of Training dataset	Size of Testing dataset
PulseOximetry	OP_BIOR1.1_L4	51	400x51	400x51
	OP_DMey_L8	85	400x85	400x85

Proposed Approach:

Basic KNN (Version 1): The basic KNN algorithm is manly used for the classification of the different dataset. In the classification of dataset we use the high dimensional data which is related to the driving condition (pre driving and post driving). In this algorithm first of all we read the data file of wavelet features set which contain two categories dataset (pre and post) from Excel data file.

Algorithm to compute the Basic KNN (Version 1) work as Classifier:

Step 1: Read the Two class data (Pre and Post driving) from Excel data file.

Step 2: Find the number of rows and number of columns.

Step 3: Initially store the dataset i=0 to rows and j=0 to number of columns into 2-D array along with class tag.

Step 4: Assign first two values of each column to corresponding two class reference points.

Step 5: Read and Calculate the minimum distance of neighbors in each column to the corresponding reference point.

Step 6: Assign data neighbors to each class from which the distance is minimum for each reference point.

Step 8: Update value of each reference point for corresponding columns.

Step 9: Recalculate the distance between previously assigned neighbors to the new reference point of each column.

Step10: If no neighbor of each column was reassigned then stop otherwise go to step5 for corresponding column.

Step 11: Prepare confusion matrix of test data for dataset & write to output file.

Modified KNN(Version 2): The basic difference in the concept from Version 1 is that here one problem with our regression application. That is how to select and we want to find the first three nearest neighbors to X. That is node 1, node 3, and node 2. But node 2 is too far away from node X. In order to alleviate this problem, First of all, we found the minimal distance from all the training set to the unknown node. In this case, we see that the distance from node 1 to node X is this value. If we just locate the neighbors that are of 1 time of minimal distance from this node, then we take this unknown node as the center and draw a circle by taking minimal distance as the radius. If we want to find the neighbors within 1.3 times of minimal distance from the unknown point, then node 3 is added. We can call our strategy clustering as well and we choose the same distance metric as in Naive KNN method. The method makes sense because only the closest neighbors join the prediction computation. We can call the proportion parameter as learning rate. Usually, we set this value from 1.0 to 2.0.

Algorithm to compute the Modified KNN (Version 2) work as Classifier:

Step 1: Read the Two class data (Pre and Post driving) from Excel data file.

Step 2: Find the number of rows and number of columns.

Step 3: Initially store the dataset i=0 to rows and j=0 to number of columns into 2-D array along with class tag.

Step 4: Read and Calculate the minimum distance between the each data point in each column to the corresponding data point.

Step 5: Set the value of learning parameter ἡ=1.0 in each column.

Step 6: Store the dataset i=0 to rows and j=0 to number of columns into 2-D array along with class tag.

Step 7: Assign data neighbors to each class from which the distance is minimum from the ἡ * calculated minimum distance each data points of each column.

Step 8: Update value of each reference point for corresponding columns.

Step 9: Recalculate the distance between previously assigned neighbors to the new reference point of each column.

Step 10: Update the value of ἡ of each column.

Step11: If no neighbor of each column was reassigned then stop otherwise go to step6 & step10 for corresponding column.

Step 12: Prepare confusion matrix of test data for dataset & write to output file.

Comparative Analysis of Performance of KNN algorithms for Pulse Oximetery Parameters:

1. **For Feature Set-I of Oximetry Pulse OP_BIOR1.1_L4:** The algorithm when tested for training and test data could perform up to 100 % classification accuracy for MAX CA4 and MAX A4.
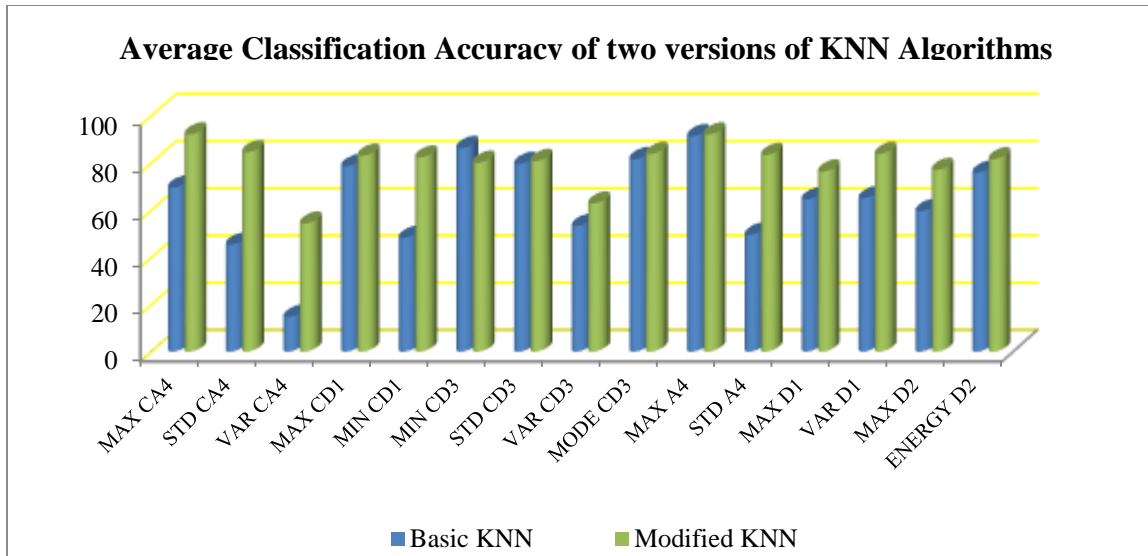

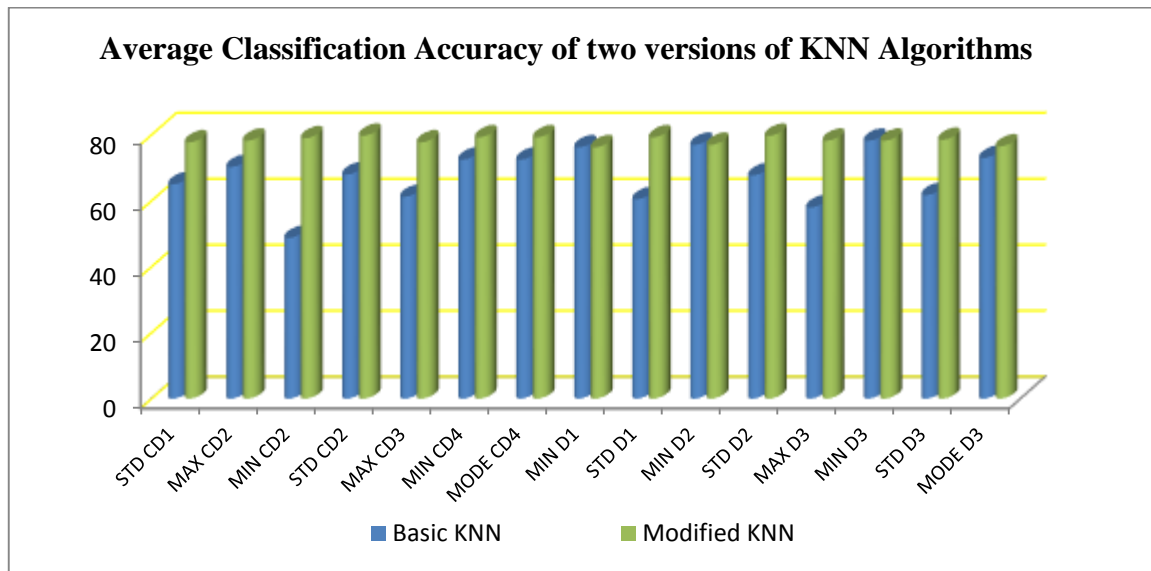Figure 1(a): PO Feature Set I versus PCLA for K-NN 1 to 2

Figure 1 (b): PO Feature Set I versus PCLA for K-NN 1 to 2

2. **Feature Set II: File Name : PO_DMey_L8 (Total 85 features):** Here the 85 features were extracted from Oximetry Pulse signal using DMeyer wavelet decomposed to level 8 with feature vector size (800x85) including 400x85 for training and 400x85 as test dataset. The MAX, MEAN of CA8, VAR, MODE of CD1, MAX, MEAN, MODE of A8, VAR CD2, MODE D8 and entropy only could give classification accuracy more than 90 % for each class while tested with training and test datasets. Some of the features could fetch good results in either of the classes.
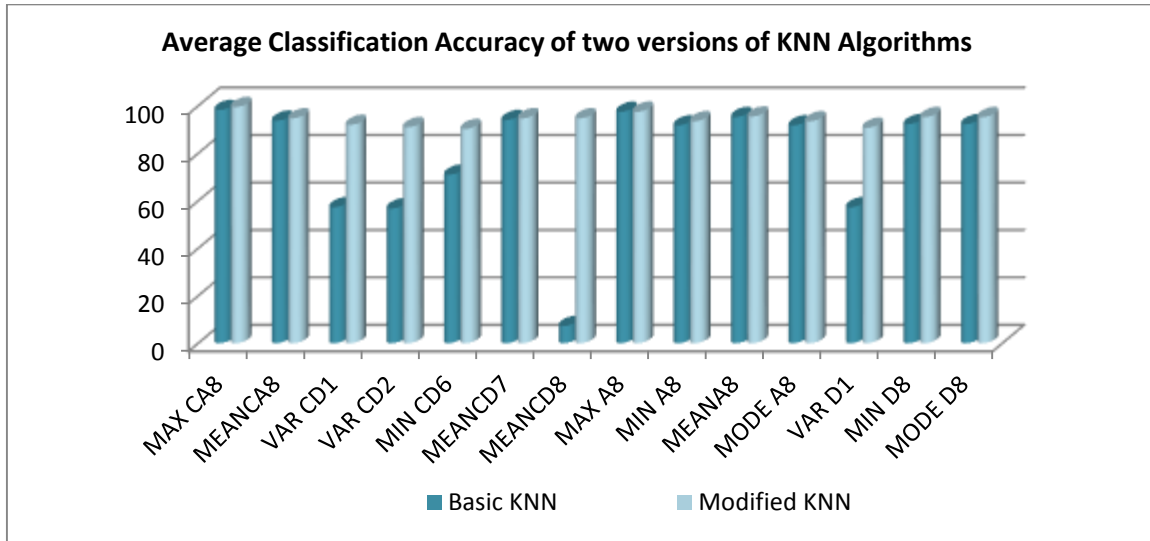

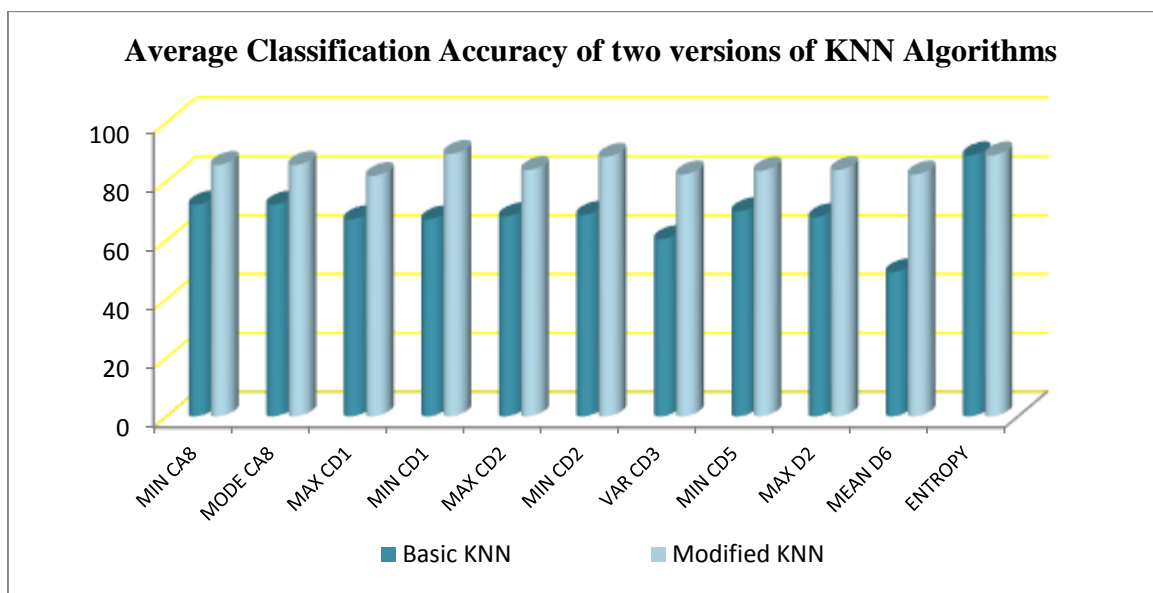**Figure 2(a): PO Feature Set II versus PCLA for K-NN 1 to 2**

Figure 2 (b): PO Feature Set II versus PCLA for KNN 1 to 2

Optimal Performing Parameters and Two Versions of KNN: It shows the summary of various features those performed the best in each class for training as well as test datasets in corresponding algorithms.

Table 2: Optimal Performing Physiological Parameters / Features & Algorithms

Physiological Parameters / Algorithms	Feature File	Total Features	K-NN Version 1	K-NN Version 2
Oximetry Pulse	Feature Set I OP_BIOR1.1_L4	51	MAX A4	MAX CA4 and MAX A4
	Feature Set II OP_DMey_L8	100	MAX, MEAN of CA8, VAR, MODE of CD1, MAX, MEAN, MODE of A8, VAR CD2, MODE D8 and entropy **(10 Features)**	MAX, MEAN of CA8, VAR of CD1,CD2 MIN of CD6, MEAN of CD7, CD8 MAX, MIN, MODE, MEAN of A8 VAR D1, MODE& MIN D8 **(14 Features)**

Results: That KNN version 2 (modified KNN) could perform very well for maximum features of pulse oximetry parameters Although KNN1 and KNN2 performed well for some of the features, the test has been on individual features rather than in combination and hence these features only can be used for detection of cognitive fatigue as the classification accuracies for both the classes has been up to 100 %. For Oximetry Pulse signal, Bior's wavelet features only the approximate coefficient of Decomposed and Reconstructed signals i.e. CA4, A4 has performed well in KNN Version 1 and 2. The PO DMeyer' s feature set Approximate coefficient of decomposed and reconstructed wave along with some detail coefficients has performed very well in Version 1 & 2.

Conclusion: The data collected for Pulse Oximetry signal of vehicular drivers for two states while driving, its analysis and feature extraction using wavelet function of varying order and level of decomposition could lead to make it feasible to correlate with fatigue level in vehicular drivers using K-Nearest Neighbor Classifier. Percentage classification accuracy has been used as performance measures to test the performance of the classifiers. The various solutions provided by the researchers to detect and monitor the drowsiness / alertness / fatigue in vehicular drivers. The review outcome could lead to understand the problem in-depth and the solution methodologies with their pros and cons. It also directed to attempt for a simple solution that may be acceptable to the vehicular drivers and at the same time fetch an accurate and early detection methodology to avoid vehicular accidents and the loss of lives or health injuries. The test results proved that some of the features of all the parameters individually could be classified to the extent of 100 % by KNN 1 and 2.

References:

[1] AbdulnasirYildiz, Mehmet Akin, Mustafa Poyraz, GokhanKirbas, 'Application of adaptive neuro-fuzzy inference system for vigilance level estimation by using wavelet-entropy feature extraction', Journal of Elsvier Expert Systems and Applications, 2009 pp.7390-7399 www.elsevier.com/locate/eswa

[2] AbdulhamitSubasi, 'Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients', Expert Systems with Applications 28 (2005) pp.701–711, Elsevier, www.elsevier.com/locate/eswa

[3] Bundele Mahesh M., and Banerjee Rahul, 2009, " An SVM Classifier for Fatigue-Detection using Skin Conductance for Use in BITS Life Guard Wearable Computing System," International Conference Technically Co-sponsored by IEEE Systems, Man & Cybernetics Society Bombay Section, G H Raisoni College of Engineering (Nagpur, MS, India, scheduled on December 16-18, 2009) pp.936-941.

 [4] Charniya, Nadir N., Dudul, S. V., Neural Network based Sensor for Classification of Material Type and its Surface Properties. Proceeding of International Joint Conference on Neural Networks, (August 2007), 424-429.

[5] Chong Zhang, Chong-XunZheng, Xiao-Lin Yu, 'Automatic recognition of cognitive fatigue from physiological indices by using wavelet packet transform and kernel learning algorithms' Journal of Elsevier Expert Systems and Applications, 2009 pp.4664-4671 www.elsevier.com/locate/eswa

[6] 'Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and the Basis for Alertness Management', Published by NHTSA U. S. Department of Transportation published in April 1998. www.itsdocs.fhwa.dot.gov/jpodocs/edlbrow/

[7] Fatigue Report Safety Net (2009), Project co-financed by the European Commission, Directorate-General Transport and Energy.

[8] He Zifen, Zhan Zhaolin , Zhang Yinhui "WLSMB Halftoning Based on Improved K-means Cluster Algorithm Using Direct Binary Search" 2013 Fifth Conference on Measuring Technology and Mechatronics Automation. PP. 978-0-7695-4932-3.

[9] Hong J. Eoh, Min K. Chung_, Seong-Han Kim, 'Electroencephalographic study of drowsiness in simulated driving with sleep deprivation', International Journal of Industrial Ergonomics 35 (2005) pp.307–320, Elsevier, www.elsevier.com/locate/ergon

[10] Hu Shuyan, ZhengGangtie, 'Driver drowsiness detection with eyelid related parameters by Support Vector Machine', Expert Systems with Applications, Elsevier 36 (2009) pp.7651–7658 www.elsevier.com/locate/eswa

[11] Mahesh M. Bundele, Rahul Banerjee" *Design of Early Fatigue Detection Elements of a Wearable Computing System for the Prevention of Road Accidents*" published in [Intelligent Systems and Applications (ISA), 2010 2nd International Workshop on 22-23 may 2010 at wuhan, pp.](#) 11357345.

[12] Mahesh M. Bundele, 'Identification of body Parameters for changes in reflexes of a vehicular driver under drowsiness / fatigue / stress conditions',Published in the Proceeding of FRONTIER 2008, A National Conference organized by BabasahebNaik College of Engineering Pusad, MS India.

[13] Mahesh M Bundele, Rahul Banerjee, "Detection of Fatigue of Vehicular driver using Skin Conductance and Oximetry Pulse: A Neural Network Approach" presented in MDC2009 at AeU Malaysia, 14-16 Dec. 2009.

[14] Wang Shu-Bin, Ling Ping, You Xiang-Yang, Xu Ming, Rong Xiang-Sheng, "Classification Algorithm Based on Weighted SVMs and Locally Tuning kNN", 2008 IEEE International Conference on BioMedical Engineering and Informatics.

[15] ZHANG Xiao-fei HUANG He-yan, ZHANG Ke-liang, "KNN Text Categorization Algorithm Based on Semantic Centre", 2009 International Conference on Information Technology and Computer Science.