International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

AN IN-CAR DETECTION SYSTEM FOR DRINK DRIVING USING MACHINE LEARNING

P.Rajkumar[1],A.Sreehitha[2],B.Saipriya[3],B.Aneeksha[4],B.Vinoothna[5]
[1]Assistant Professor, Department of CSE, Malla Reddy Engineering College for Women
Hyderabad, Telangana, India
[2,3,4,5] UG Scholar, Department of CSE, Malla Reddy Engineering College for Women
Hyderabad, Telangana, India
rajkpatil@gmail.com
srihitha172002@gmail.com, saipriyavarma01@gmail.com, aneekshabairi19@gmail.com,
vinoothnabattu002@gmail.com

ABSTRACT

As one of the biggest contributors to road accidents and fatalities, drink driving is worthy of significant research attention. However, most existing systems on detecting or preventing drink driving either require special hardware or require much effort from the user, making these systems inapplicable to continuous drink driving monitoring in a real driving environment. In this paper, we present a contactless, non- invasive, real-time system that yields a relatively highly accurate drink driving monitoring by combining vital signs (heart rate and respiration rate) extracted from in- car WiFi system and driver's psychomotor coordination through steering wheel operations. The framework consists of a series of signal processing algorithms for extracting clean and informative vital signs and psychomotor coordination, and integrate the two data streams using a self-attention convolutional neural network (i.e., C-Attention). In safe laboratory experiments with 15 participants, This in-car detection system achieves drink driving detection accuracy of 96.6% .These promising results provide ahighly encouraging case for continued development.

Keywords—Drink drive detection, inertial measurement unit

I. INRODUCTION

In the modern day, with the increase in the number of vehicles plying on the roads, traffic accidents have grown significantly in number. One of the primary causes of traffic accidents is drunk driving or driving under influence (DUI). This is particularly an important issue for developing countries, such as India, where 53.4% of unnatural deaths in the year 2014 were due to traffic accidents, with drunk driving being the primary cause. Currently, police inspecting roads sample cars for breath tests to detect alcohol levels. However, this approach is manual and unlikely to detect most cases of driving under influence of alcohol. Alternate and more effective approaches to detect drunk driving may include automatic detection using sensors. Prevention may include reducing he car's speed or alerting people or police via the internet. The network of physical devices embedded with sensors, electronics, computing modules and network connectivity using which they aggregate and share data. In this specific case, sensors to detect alcoholism can be directly or indirectly used. This can be connected to mobile phones which contain communication network for appropriately alerting people. Several models can be found in literature. However, no specific model has been adopted for widespread use. The various methods proposed face problems such as difficulty in implementation, scalability and complexity. Therefore, new models need to maintain cheap cost, easy implementation along with accuracy.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

II. LITERATURE SURVEY

In 2018, The US suffered 10,511 deaths from drunk driving crashes. The WHO reports that, in high-income countries, as many as 20% of fatally injured drivers have excess alcohol in their blood.1 COVID-related deaths may dwarf these numbers, but it is important not to lose our prepandemic perspective. Tens of thousands of deaths per year due to drink driving is a staggering loss of life. At an estimated cost of $44billion2 for just one year in the US alone.

The economic impact is not insignificant either. According to the newly-enacted Halt Act (H.R. 2138) and Ride Act (S.B. 1331) , drunk-driving prevention technology will be asafety standard for all new cars in the future. There is anurgent demand for in-vehicle drunk-driving detection systems to help prevent drunk-related accidents.

Most traditional methods for detecting drunk drivers need tointerrupt the driving process. To administer a breathalyzer,the police must hail the driver to pull over, giving the driver time to implement means of avoiding detection. Blood tests are invasive, and require the driver to stop. Similarly, urinetests and pupil measuring tests require special operations andexpert examiners. While it is desirable to detect whether the driver is drunk before driving and prevent potential risks, it ispossible that alcohol consumption takes time to take effectand the driver may consume alcohol during driving.

Therefore, the most reliable way is to have a continuous monitoring of drunkenness during driving without interfering the driving process.

III. METHODOLOGY

Drunk driving detection using driving pattern uses mobile phone as the platform for drunk detection as they combine the detection and communication functions . As a self-contained device, mobile phone presents a mature hardware and software environment for the development of active drunk driving monitoring system. The system based on mobile phone can function effectively on its own because mobile phones are highly portable, all necessary components are already integrated, and their communication services have vast coverage. The minimum requirement for such a mobile phone platform is the presence of simple sensors like accelerometer, orientation sensor, etc. The communication module and speaker are good enough for alerting. Drunk driving related behavior are classified into three categories. The first and second category focus on driving behaviors related to vehicle movement; the third category is aboutdriving behavior related to subjective judgment and vigilanceof the driver..

IV. TYPES OF CLASSIFICATION ALGORITHMS Many machine learning algorithms are being used in variousfields of research to help in solving the real-world problems.Mostly used machine learning classification algorithms arediscussed below:

A. DECISION TREE CLASSIFIERS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

B. GRADIENT BOOSTING

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees. When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest. A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

B. KNN CLASSIFIER

Simple, but a very powerful classification algorithm

- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not "learn" until the test example is given

Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examplesin feature space
- Feature space means, space with categorizationvariables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset.

C. LOGISTIC REGRESSION CLASSIFIER

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

D. NAIVE BAYES

The naive bayes approach is a supervised learning methodwhich is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class isunrelated to the presence (or absence) of any other feature . Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation basedon the representation bias. The naive bayes classifier is alinear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine).The difference lies on the method of estimating the parameters of the classifier (the learning bias).
While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest ofsuch a technique.

Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (**Weka 3.6.0**, **R 2.9.2**, **Knime 2.1.1**, **Orange 2.0b** and **RapidMiner 4.6.0)**. We try above all to understand the obtained results.

E. RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the

random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.
The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.
An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance.
Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

F. SUPPORT VECTOR MACHINE

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point *x* and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

equation for a multidimensional surface that best separates the different classes in the feature space.
SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate intoseveral hyperplanes' meeting this requirement.

V. EXPERIMENTAL RESULTS

In the fig below ,the classification accuracy, precision, recall and f1 score of all these classifiers are shown. We have achieved approximately 62% classification accuracy (highest) for Random Forest SVM. We have analyzed f1 score also to check if the model works well at both false positive and false negative samples. The equations of the measured parameters are given below:

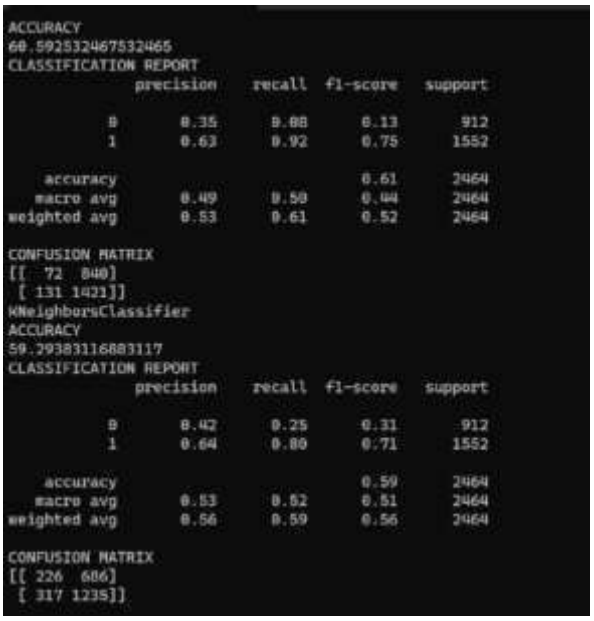


Fig1 Report on accuracy, precision and f1 score.

Accuracy = TP+TN/TP+FP+FN+TNPrecision = TP/TP+FP
Recall = TP/TP+FN
F1 Score = 2*(Recall * Precision) / (Recall + Precision) (TP= TruePositive, TN= True Negative, FP= False Positive, FN= False Negative)

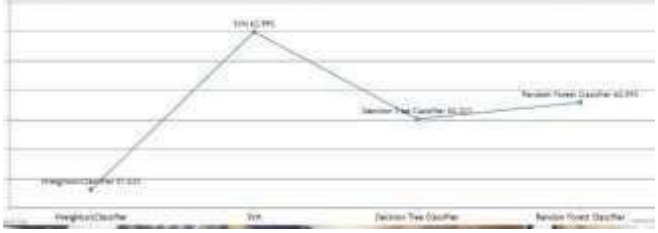


Fig2.Barchart representing the accuracy of the algorithms.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

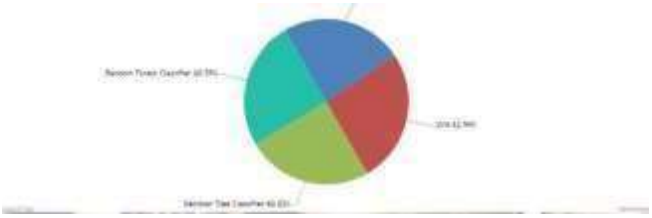


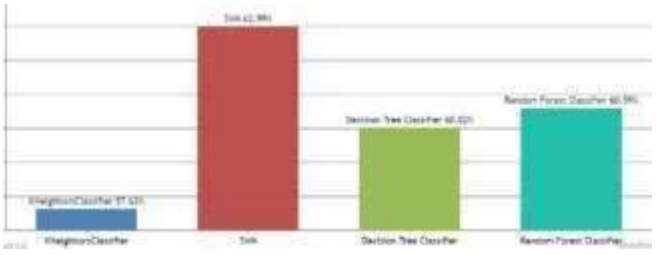
Fig3.Piechart showing the accuracy of the algorithms

View Drink Driving Detection Found Ratio Details

Detection Type	Ratio
No Drink Driving Detection	10.526315789473683
Drink Driving Detection	89.47368421052632

Fig4.The ratio of drink driving of the trained data.

The values of both precision and recall are also good for the trained model. In fig1. The confusion matrix is given. Most of the test data are positioned diagonally.



Trained and Tested Drink Driving Datasets Results

Model Type	Accuracy
KNeighborsClassifier	58.40097402597403
SVM	62.98701298701299
Decision Tree Classifier	60.024350649350644
KNeighborsClassifier	55.1948051948052
Random Forest Classifier	60.592532467532465
KNeighborsClassifier	59.29383116883117

Fig. 4. Accuracy of the algorithms used in the detection.

VI. CONCLUSION AND FUTURE WORK

In this paper, we presented *Detect DUI*, a non-intrusive, contactless, and continuous system of measuring and monitoring

the side effects of alcohol on drivers. To develop *Detect DUI* to this stage, we have overcome two main challenges. The first is to eliminate interference in the WiFi signals caused by the motions of a moving vehicle. This problem was solved with a series of signal processing algorithms. The second is determining which specific features of alcohol's side effects best reflect driving under the influence of alcohol. We have addressed this challenge with a C-Attention network. The results of extensive experiments confirm that *Detect DUI* provides highly accurate drink driving detection.

Apart from drinking alcohols, other factors may also affect vital signs and psychomotor coordination, e.g., catching a cold or other respiratory diseases. Respiratory diseases will change breathing patterns, which are expected to be different from the breathing patterns of drinking. However, it is difficult to collect training samples to help differentiate the breathing patterns under the two conditions. In the future, we intend to refine our drink driving detection model by considering other impact factors.

VII. REFERENCE

[1] Alcohol Impaired Driving: 2018 Data (Traffic Safety Facts. Report No.DOT HS 812 864). Accessed: Sep. 2019. [Online].Available:https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812864

[2] H.R. 2138: Honoring Abbas Family Legacy to Terminate Drunk Driving Act of 2021. Accessed: Mar. 2021. [Online]. Available:https://www.govtrack.us/congress/bills/117/hr2138

[3] S.1331—RIDE Act of 2021. Accessed: 2021. [Online].Available:https://www.congress.gov/bill/117th- congress/senate-bill/1331/text

[4] C.-W. You et al., "Enabling personal alcohol tracking using transdermal sensing wristbands: Benefits and challenges," in Proc. 21st Int. Conf.Hum.-Comput. Interact. Mobile Devices Services, Oct. 2019, pp. 1–6.

[5] Y. Jung, J. Kim, O. Awofeso, H. Kim, F. Regnier, and E. Bae,"Smartphone-based colorimetric analysis for detection of saliva alcohol concentration," Appl. Opt., vol. 54, no. 31, pp. 9183–9189, 2015.

[6] H.-L. Kao, B.-J. Ho, A. C. Lin, and H.-H. Chu, "Phone- based gait analysis to detect alcohol usage," in Proc. ACM Conf. Ubiquitous Comput. (UbiComp), 2012, pp. 661–662.

[7] S. Bae et al., "Detecting drinking episodes in young adults using smartphone-based sensors," Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol., vol. 1, no. 2, pp. 1– 36, Jun. 2017.

[8] A. Mariakakis, S. Parsi, S. N. Patel, and J. O. Wobbrock,"Drunk user interfaces: Determining blood alcohol level through everyday smartphone tasks," in Proc. CHI Conf.Hum. Factors Comput. Syst.,Apr. 2018, pp. 1–13. [9]Alcohol. Accessed: 2021. [Online]. Available:https://en.wikipedia.org/wiki/Alcohol

[10] Alcohol's Effects on the Body. Accessed: 2021. [Online]. Available:https://www.niaaa.nih.gov/alcohols- effects-health/alcohols-effects-body

[11]Hangovers. Accessed: 2021. [Online]. Available: https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/hangovers

[12] What Alcohol Really Does To Your Body. Accessed:2005. [Online]. Available: https://www.independent.co.uk/life-style/health-and-families/health-news/what-alcohol-really-does-your- body-5349097.html

[13] Alcohol and Cardiovascular Disease. Accessed: 2020.[Online]. Available: https://alcoholthinkagain.com.au/alcohol-your- health/alcohol-andlong-term-health/alcohol-and- cardiovascular-disease/

[14] C. Wu, K. Tsang, H. Chi, and F. Hung, "A precise drunk driving detection using weighted kernel based on electrocardiogram," Sensors,vol. 16, no. 5, p. 659, May 2016.

[15] S. Brunner et al., "Alcohol consumption, sinus tachycardia, and cardiac arrhythmias at the munich octoberfest: Results from the munich beer related electrocardiogram workup study (MunichBREW)," Eur.Heart J., vol. 38, no. 27, pp. 2100–2106, Jul. 2017.

[16] First Aider's Guide to Alcohol by Rick Curtis, Director,Outdoor Action. [Online]. Available:https://www.princeton.edu/~oa/safety/alcohol.sht ml

[17] Drink Driving UK Law History. [Online]. Available:https://www.drinkdriving.org/drink_driving_information_uk lawhistory.php

[18]Breathmeter. [Online]. Available:https://www.breathometer.com/
[19]BACtrack Mobile Pro. [Online]. Available:https://www.bactrack.com/
products/bactrack-mobile-smartphone-breathalyzer [20]Breathalyzer. [Online]. Available:
https://en.wikipedia.org/wiki/Breathalyzer

[21] SCRAM CAM Textrm. [Online]. Available: https://adlmonitoring.com/cam/
[22] I. Chatterjee et al., "Driving fitness detection: A holistic approach for prevention of drowsy and drunk driving using computer vision techniques," in Proc. South-Eastern Eur. Design Autom., Comput. Eng.,Comput. Netw. Soc. Media Conf. (SEEDA CECNSM), 2018, pp. 1–6.
[23] V. Mehta, S. S. Katta, D. P. Yadav, and A. Dhall, "DIF: Dataset of perceived intoxicated faces for drunk person identification," in Proc.Int. Conf. Multimodal Interact., Oct. 2019, pp. 367–374.
[24]S. Szymkowski. (2021). Tesla Interior Cameras Threaten Driver Privacy, Consumer Reports Says. [Online]. Available: https://www.cnet.com/roadshow/news/tesla-interior
cameras-driver-privacyconsumer-reports/.