Facial-Expression Analysis for Predicting Unsafe Driving Behavior

A system for tracking driver facial features aims to enhance the predictive accuracy of driver-assistance systems. The authors identify key facial features at varying pre-accident intervals and use them to predict minor and major accidents.

Maria E. Jabon, Jeremy N. Bailenson, Emmanuel Pontikakis, Leila Takayama, and Clifford Nass
Stanford University

Every year in the US alone, more than 42,000 Americans die as a result of 6.8 million automobile accidents.[1] Consequently, driver-safety technology is an active research area in both industry and academia. Pervasive computing environments, with integrated sensors and networking, can provide an ideal platform for developing such technology. Taking effective countermeasures to enhance safe vehicle operation requires merging information from diverse sets of information.

As a first step, an *active driver safety system* (a system designed to prevent accidents) must monitor vehicle state and surroundings (see the "Related Work in Active Driver Safety Systems" sidebar). However, to fully transform the vehicle into a smart environment,[2] the driver must also be monitored. Human factors researchers have long studied the driver's role in causing and preventing accidents and have found that the driver's physical and emotional state, including fatigue[3] and stress levels,[4] play a role in a significant number of traffic accidents. Thus, many researchers have begun developing active driver-safety systems that monitor the driver as well as the vehicle.

We propose an active driver-safety framework that captures both vehicle dynamics and the driver's face. We then merge the two levels of data to produce an accident prediction and investigate the frameworks performance. Our study differs from previous work in active driver safety in four ways. First, we use a bottom-up approach, analyzing the movement of a comprehensive set of 22 raw facial features, rather than simple metafeatures such as eye gaze or head orientation. Second, we evaluate a range of time and frequency domain statistics to determine the most valuable statistics for driving accident prediction. Third, we predict major and minor accidents directly, not intermediate driver states such as fatigue. Finally, we explore the use of the face and car outputs at varying pre-accident intervals, uncovering important temporal trends in predictive accuracy for each feature subset.

Experimental Testbed

We recruited 49 undergraduate students to drive a 40-minute simulated course in a STiSIM driving simulator. We set up the simulator, developed by Systems Technology,[5] to run on a single PC and project the simulated image of the roadway onto a white wall in the lab (see Figure 1).

During the driving course, we projected the sounds of the car, the road, and events happening

Published by the IEEE CS ■ 1536-1268/11/$26.00 © 2011 IEEE

Related Work in Active Driver Safety Systems

Much work has been done in the area of holistic vehicle sensing for active driver safety, including systems for monitoring the vehicle environment, vehicle state, and more recently, driver state. Systems developed to monitor the vehicle environment include pedestrian and obstacle detectors, lane-guidance systems, rear-bumper proximity sensors, blind-spot car detectors, automatic windshield wipers, and surround imaging systems for parking assistance.[1–6] Systems for monitoring the vehicle state include tracking vehicle location via GPS and accelerometers and other sensors to monitor driving speed, steering wheel angle, braking, and acceleration.[7–8] Systems for monitoring driver state include frameworks that gauge driver fatigue, drowsiness, or stress levels.[9–14]

In this article, we extend previous work by using computer vision algorithms to directly map specific facial features to unsafe driving behavior. We use a comprehensive set of raw facial-feature points that are absent in previous work, including points around the nose. Furthermore, we don't infer any specific mental states such as fatigue, but rather implement a more empirical approach that uses machine learning algorithms to find and use the facial features that are most correlated with accidents. In addition, we identify important temporal trends in predictive accuracy for each feature subset, revealing how to best use the face to improve the predictive accuracy of classifiers up to four seconds before a driving accident.

REFERENCES

1. L. Li et al., "IVS 05: New Developments and Research Trends for Intelligent Vehicles," *IEEE Intelligent Systems,* vol. 20, no. 4, 2005, pp. 10–14.
2. T. Gandhi and M. Trivedi, "Vehicle Surround Capture: Survey of Techniques and a Novel Omni Video-Based Approach for Dynamic Panoramic Surround Maps," *IEEE Trans. Intelligent Transportation Systems,* vol. 7, no. 3, 2006, pp. 293–308.
3. J. McCall and M. Trivedi, "Video-Based Lane Estimation and Tracking for Driver Assistance: Survey, System, and Evaluation," *IEEE Trans. Intelligent Transportation Systems,* vol. 7, no. 1, 2006, pp. 20–37.
4. M. Bertozzi, A. Broggi, and A. Lasagni, "Infared Stereo Vision-Based Pedestrian Detection," *Proc. IEEE Intelligent Vehicles Symp.,* IEEE Press, 2005, pp. 24–29.
5. Y. Suzuki, T. Fujii, and M. Tanimoto, "Parking Assistance Using Multi-Camera Infrastructure," *Proc. IEEE Intelligent Vehicles Symp.,* IEEE Press, 2005, pp. 106–110.
6. H. Kurihata, T. Takahashi, and I. Ide, "Rainy Weather Recognition from In-Vehicle Camera Images for Driver Assistance," *Proc. IEEE Intelligent Vehicles Symp.,* IEEE Press, 2005, pp. 205–210.
7. J. Waldo, "Embedded Computing and Formula One Racing," *IEEE Pervasive Computing,* vol. 4, no. 3, 2005, pp. 18–21.
8. M. Bertozzi et al., *Knowledge-Based Intelligent Information and Engineering Systems*, B. Appolloi et al., eds., Springer, 2007, pp. 704–711.
9. M. Trivedi, T. Gandhi, and J. McCall, "Looking-In and Looking-Out of a Vehicle: Computer-Vision-Based Enhanced Vehicle Safety," *IEEE Trans. Intelligent Transportation Systems,* vol. 8, no. 1, 2007, pp. 108–120.
10. A. Williamson and T. Chamberlain, *Review of On-Road Driver Fatigue Monitoring Devices,* NSW Injury Risk Management Research Centre, Univ. of New South Wales, 2005.
11. J.A. Healy and R.W. Picard, "Detecting Stress During Real-World Driving Tasks Using Physiological Sensors," *IEEE Trans. Intelligent Transportation Systems,* vol. 6, no. 2, 2005, pp. 156–166.
12. A. Doshi and M.M. Trivedi, "On the Roles of Eye Gaze and Head Dynamics in Predicting Driver's Intent to Change Lanes," *IEEE Trans. Intelligent Transportation Systems,* vol. 10, no. 3, 2009, pp. 453–462.
13. L. Fletcher, L. Petersson, and A. Zelinsky, "Road Scene Monotony Detection in a Fatigue Management Driver Assistance System," *Proc. IEEE Intelligent Vehicles Symp.,* IEEE Press, 2005, pp. 484–489.
14. Q. Ji, Z. Zhu, and P. Lan, "Real Time Non-intrusive Monitoring and Prediction of Driver Fatigue," *IEEE Trans. Vehicular Technology,* IEEE Press, 2004, pp. 1052–1068.

around the car into the simulator room via four PC speakers. The course simulated driving in a suburban environment with conditions varying from light to intense traffic. We included such challenging situations as busy intersections, unsafe drivers, construction zones, sharp turns, and jaywalking pedestrians in an effort to increase the drive's complexity (see Table 1).

We used a virtual driving simulator instead of real cars in order to safely collect a large sample size of accidents to use in our analyses. This let us generate separate models for major accidents (for example, hitting pedestrians, other vehicles, or off-road objects) and minor accidents (such as unwarranted lane changes, driving off the road, or running a stoplight).

During the experimental sessions, we recorded participants' faces with two Logitech Web cameras at a rate of 15 frames per second. We compressed the videos to AVI format in real time using DirectX and DivX technology. Although many technologies can capture facial movements, we opted for image-based capture because it does not require special markers or user intervention. Thus, our system is less intrusive, increasing transparency. We also recorded the simulator's output during the driving sessions, which was a text-based log file listing road conditions, steering wheel angle, lane tracking

Figure 1. A study participant being monitored in the STISIM driving simulator. The simulator runs on a single PC and projects the image of the roadway onto a white wall in the lab.

information, car speed, longitudinal acceleration (feet/second^2), braking information, and number and type of accidents.

Analysis Procedure

Using the collected videos and driving simulator data, we constructed a dataset to build our computational models. Figure 2 summarizes the data-analysis phases.

Facial-Feature Extraction

The first step in constructing our datasets included extracting key facial features and head movements from the videos we collected. For this processing, we used the Neven Vision library.[6] For each video frame, the Neven Vision library automatically detects, without any preset markers worn on the user's face, the *x* and *y* coordinates of 22 points on the face, eye- and mouth-openness levels, and head movements (for example, yaw, pitch, and roll). It does this at a rate of 30 frames per second. Figure 3 is a screenshot of the Neven face-tracking software.

Data Synchronization

In the next phase of our analysis, we synchronized the video outputs with the driving simulator outputs so we knew when accidents occurred within the video and could extract the pre-accident facial geometry and vehicle information. Our goal was to determine the optimal way to combine the facial movements and vehicle outputs to predict when accidents would occur. Thus, we sampled several pre-accident time intervals beginning between one and four seconds before the accident and ranging from one to 10 seconds long. For each interval, we extracted the data preceeding every major and minor accident in our dataset, along with a random number of nonaccident intervals to use in our analyses.

Time Series Statistics Calculation

After data synchronization, we computed a series of time-domain statistics on the coordinates in each interval to use as inputs to our classifiers. We calculated averages, velocities, maximums, minimums, standard deviations, and ranges for each of the Neven outputs and for the vehicle outputs such as speed, wheel angle, throttle, and braking outputs. For some important facial characteristics, such as eye- and mouth-openness levels, we also created five-bin histograms from 0 to 100 percent to capture their distribution over the time interval.

Frequency Domain Statistics Calculation

We next calculated frequency domain statistics on each facial coordinate

TABLE 1
Simulation parameters.

Miles	Environment	Speed limit (mph)	Intersections	Traffic	Challenges
0–1	Suburban	35	6	Heavy	Many pedestrians
2–6	Highway	65	0	Moderate	Narrow roads, tight curves
7–9	Suburban	35	15	Heavy	Many pedestrians
10–11	Highway	55	1	Light	Construction zone, obstacles
12–18	Suburban	35	24	Heavy	Many pedestrians
19–20	Rural	35	0	Light	Dirt roads, obstacles
21–22	Rural	35	1	Moderate	Narrow roads, tight curves
23–32	Urban	55	6	Moderate	Tight curves

and each car output. We used the Matlab Wavelet toolbox to perform the discrete wavelet transform, in particular the Daubechies wavelet family with orders one, two, and four. For each order, we performed a level-three decomposition of the input signal and collected statistics over the detail coefficients of each level including averages, ranges, histograms, and variances. We calculated these additional statistics because facial signals are dynamic, and we expected that their micromomentary movements could leak information about the internal state of the person making the expression.

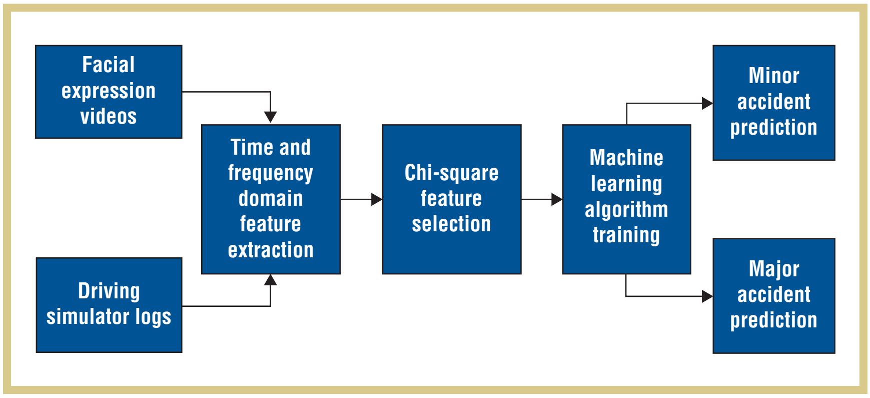


Figure 2. Data-analysis procedure. Facial features were extracted from video recordings and synchronized with simulator output logs. Time and frequency domain statistics were extracted from the resulting datasets, top features were selected, and then used to predict major and minor accidents at varying pre-accident time intervals.

Final Dataset Creation

Finally, for each interval length, we combined all the pre-accident and nonaccident intervals for each of the 49 participants into one large dataset. To improve the reliability of the face measurements, we discarded intervals where the average face-tracking confidence (that is, the measure of how confident the face-tracking software was in its measurement) was lower than 60 percent. Our resulting dataset contained a total of 179 minor accident instances, 131 major accident instances, and 627 nonaccident instances.

Chi-Square Feature Extraction

Initially, our datasets consisted of 7,402 facial features and 1,162 vehicle features for each pre-accident and nonaccident vector. To identify which facial features were the most important indicators of unsafe driving behavior and to more quickly train our algorithms, we performed a chi-square feature selection for each dataset. Tables 2 and 3 list the most predictive 20 car features and most predictive 20 facial features for our major and minor accident predictions.

The facial features most predictive of major and minor accidents differed greatly. Most of the top major accident features were movements of points around the mouth and eyes, whereas top minor accident features centered around the nose. This adds new information to previous works, which used features only around the eyes and mouth to predict dangerous driving states.[3–4] Less of a difference existed between the top major accident car features and top minor accident car features; in both cases, the steering wheel angle and steering inputs made up more than 75 percent of the top 20 features.

We also found that the most useful facial-feature statistics varied across accident type. Wavelets proved to be the most useful statistics for major accident prediction, whereas simple minimums and maximums were more informative for minor accident prediction.

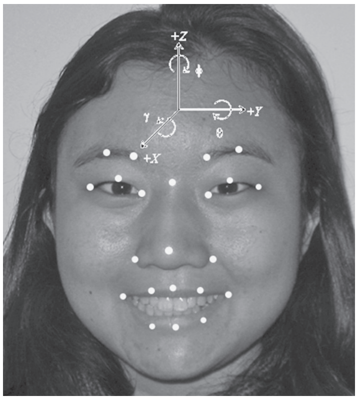


Figure 3. Neven Vision tracking points on a subject's face. We tracked 22 points around the eyes, nose, mouth, and eyebrows at a rate of 30 frames per second.

Classification and Results

We experimented with numerous state-of-the-art classifiers to predict driving accidents, including Bayesian nets, decision tables, decision trees, support vector machines (SVMs), regressions, and LogitBoost simple decision stump classifiers.[7] We evaluated our classifiers according to Cohen's kappa, which corrects for the degree of agreement between a classifier's predictions and reality by considering the proportion of predictions that might occur by chance.[8] This measure has been shown to be more robust than simple measures such as hit rate or overall accuracy. Values range from zero to one, with a score of zero implying completely random classification and a one implying perfect classification.

TABLE 2
Most predictive 20 facial and vehicle features for major accident prediction.

Feature	Statistic	Value
Face		
Lower lip center X	Wavelet 62	33.3
Upper lip center Y	Wavelet 62	26.6
Right lower lip X	Wavelet 62	25.8
Right pupil Y	Wavelet 9	25.3
Right pupil Y	Wavelet 39	25.3
Right upper lip X	Wavelet 62	23.2
Left eye aspect ratio	Wavelet 88	23.0
Left inner eyebrow Y	Wavelet 55	20.9
Lower lip center Y	Wavelet 51	20.2
Right inner eyebrow Y	Wavelet 55	20.2
Left nostril Y	Wavelet 5	18.4
Left mouth corner X	Wavelet 72	17.7
Right eye aspect ratio	Wavelet 78	17.1
Face X	Wavelet 42	16.6
Left eye aspect ratio	Wavelet 53	15.8
Head euler X	Wavelet 89	15.3
Right pupil Y	Wavelet 84	15.1
Left lower lip Y	Wavelet 51	15.1
Face scale	Wavelet 55	15.1
Right lower lip Y	Wavelet 80	14.8
Car		
Velocity	Average	161.0
Velocity	Minimum	154.7
Velocity	Maximum	124.5
Steering wheel angle	Wavelet 40	88.8
Braking input	Wavelet 65	86.9
Steering input	Wavelet 71	83.8
Steering wheel angle	Wavelet 3	82.7
Steering wheel angle	Velocity	78.9
Steering wheel angle	Wavelet 43	78.1
Steering wheel angle	Wavelet 72	76.2
Steering wheel angle	Wavelet 35	76.1
Steering wheel angle	Wavelet 41	76.1
Steering input	Wavelet 44	74.2
Steering wheel angle	Wavelet 2	74.1
Braking input	Wavelet 42	73.9
Steering input	Wavelet 40	73.5
Steering wheel angle	Wavelet 42	73.3
Steering input	Wavelet 37	72.5
Steering wheel angle	Wavelet 11	72.5
Steering wheel angle	Wavelet 37	73.2

Generally, a kappa scores greater than 0.2 are considered statistically significant.[9]

Predicting Minor Accidents

We first attempted to predict only minor accidents (that is, centerline crossings, tickets, and road-edge excursions) to determine the face's role in minor accident predictions. Our minor accident datasets consisted of 806 instances: 179 minor accident instances and 627 nonaccident instances. We trained five classifiers: a SVM classifier with a polykernel, a LogitBoost classifier with the weak classifier of a simple decision stump, a multilayer perceptron neural net, a decision table, and a logistic regression. We built these classifiers using the publicly available Waikato Environment for Knowledge Analysis (WEKA) tool[10] and validated our models using a tenfold cross validation. The LogitBoost classifier provided the highest kappa statistic of the five classifiers. Figure 4 (page 90) shows the results of the LogitBoost classifications across various pre-accident intervals ranging from one to four seconds pre-accident using one to 10 seconds of data.

An interesting trend appeared when analyzing the various pre-accident intervals used in minor accident prediction. Whereas the car features proved more useful in predicting accidents close to the accident time (one to two seconds before), facial features proved more predictive longer before the accident (three to four seconds before). In fact, in all the intervals we analyzed, classifiers using the facial features outperformed classifiers using the car features at four seconds prior to the accident. Furthermore, at four seconds pre-accident, these classifiers outperformed the classifiers using all the features in four of the 10 intervals. This suggests that facial features, especially those around the nose, could significantly improve driver safety systems accuracy a longer time before an accident, allowing drivers more time to react and prevent the accident.

To further analyze the temporal trends in our classifiers' predictive accuracy, we plotted receiver operating characteristic (ROC) curves depicting true versus false positives for the classifiers using all features, only car features, and only facial features, at one to four seconds before accidents occurred (see Figure 5). We used seven seconds of data for each plot given that our highest accuracies occurred in this range.

The ROC curves for the face become almost equal to the curves using all the data by four seconds prior to the minor accidents. This implies that by four seconds prior to the accident the signal that provides the bulk of predictive power comes from the face. In all cases, the ROC curves using all the features provide the best overall tradeoff between true and false positives. From this we conclude that the face provides a signal that isn't in the vehicular features and that this signal occurs before the signal provided by the car features.

Predicting Major Accidents

We next attempted to predict major accidents (that is, hitting objects or pedestrians) to determine the face's role in major accident predictions. Our major accident datasets consisted of 758 instances: 131 major accident instances and 627 non-accident instances. We again trained five classifiers, a SVM classifier with a polykernel, a LogitBoost classifier with the weak classifier of a simple decision stump, a multilayer perceptron neural net, a simple decision table, and a logistic regression. As with minor accidents, we built all classifiers using WEKA and validated the models using a tenfold cross validation. Again, the LogitBoost classifier gave the highest kappa statistics.

TABLE 3. Most predictive 20 facial and vehicle features for minor accident prediction.

Feature	Statistic	Value
Face		
Right outer eye corner Y	Minimum	41.6
Mouth aspect ratio	Wavelet 57	33.6
Left outer eye corner Y	Maximum	32.6
Left pupil Y	Average	31.5
Left outer eye corner Y	Average	28.0
Left pupil Y	Maximum	28.0
Left inner eyebrow X	Minimum	27.1
Left pupil X	Minimum	26.4
Nose tip Y	Minimum	26.2
Right nostril X	Maximum	25.9
Nose root X	Minimum	25.3
Nose tip X	Maximum	25.3
Right nostril X	Average	25.0
Face X	Velocity	24.7
Left eyebrow center X	Minimum	23.9
Left outer eye corner X	Minimum	23.5
Left inner eye corner X	Wavelet 28	23.3
Right nostril Y	Minimum	23.1
Left inner eyebrow X	Average	23.0
Left inner eye corner X	Minimum	22.4
Car		
Steering wheel angle	Maximum	143.9
Steering wheel angle	Velocity	136.7
Steering input	Velocity	128.4
Steering wheel angle	Wavelet 3	122.0
Steering input	Maximum	119.7
Steering input	Wavelet 3	110.3
Steering wheel angle	Variance	102.2
Steering wheel angle	Wavelet 4	98.4
Steering wheel angle	Wavelet 69	98.3
Steering wheel angle	Wavelet 70	96.1
Steering wheel angle	Average	96.0
Steering wheel angle	Wavelet 2	93.5
Steering wheel angle	Wavelet 72	93.4
Steering input	Variance	93.4
Steering input	Wavelet 4	89.9
Steering wheel angle	Wavelet 34	89.4
Steering wheel angle	Wavelet 44	89.4
Steering input	Range	87.8
Steering wheel angle	Wavelet 42	87.3
Steering wheel angle	Wavelet 41	85.4
Steering wheel angle	Wavelet 12	85.2

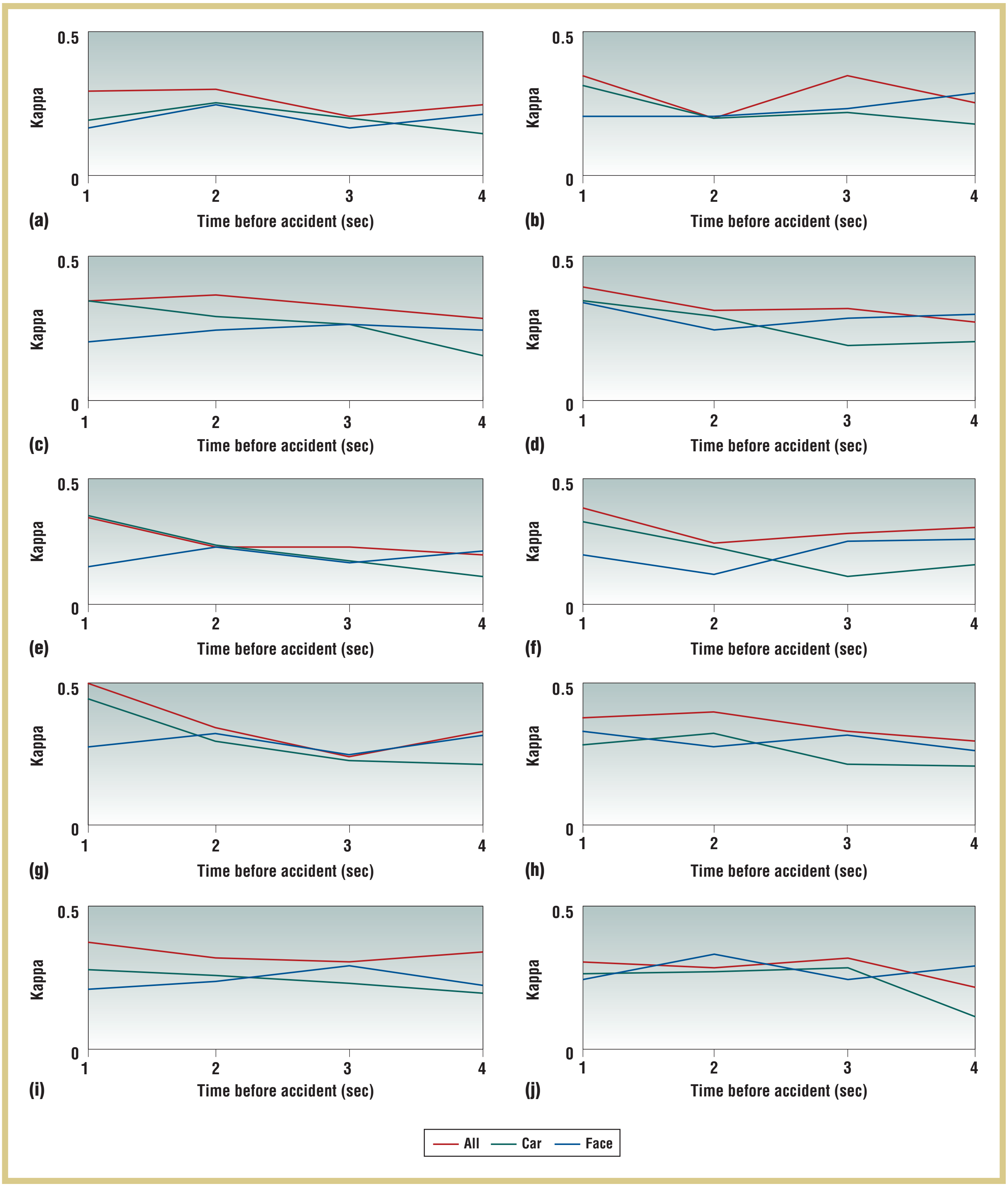


Figure 4. Minor accident classifier performance. We used face, car, and all inputs to predict minor accidents one to four seconds before accidents occurred using statistics calculated over a range of intervals of data: (a) one, (b) two, (c) three, (d) four, (e) five, (f) six, (g) seven, (h) eight, (i) nine, and (j) 10 seconds. Kappa values greater than 0.2 are considered statistically significant.

Figure 6 presents the results of the LogitBoost classifiers across pre-accident intervals ranging from one to four seconds pre-accident and using one to 10 seconds of data.

Unlike in the case of minor accidents, where we observed a marked increase in the predictive power of facial features as the time before the accident increased, we see the performances remain relatively stable across time for major accidents. However, as with minor accident classifiers, the classifiers that use facial features in combination with the car features consistently provided a higher classification kappa than the classifiers that used either feature set alone. This suggests that the vehicle features might be a better signal than facial features in major accidents, but that facial features can still improve the classifiers' overall performance.

Predicting Minor and Major Accidents

As a final step, we predicted minor and major accidents together by combining all major, minor, and nonaccident instances into one large dataset and creating classifiers on this comprehensive dataset. We trained the same five classifiers andvalidated our models using a tenfold cross validation. Once again the LogitBoost classifier provided the highest kappa statistic of the five sampled classifiers.

Figure 7 presents the results of the LogitBoost classifications across pre-accident intervals ranging from one to four seconds pre-accident and using one to 10 seconds of data.

As in the case of both minor and major accident predictions, the classifiers using all the features performed best across sampled intervals. In addition, the performance of the classifiers using the facial features alone tended to remain steady even out to four seconds pre-accident, whereas the performance of the classifiers using only the car features tended to fall off.

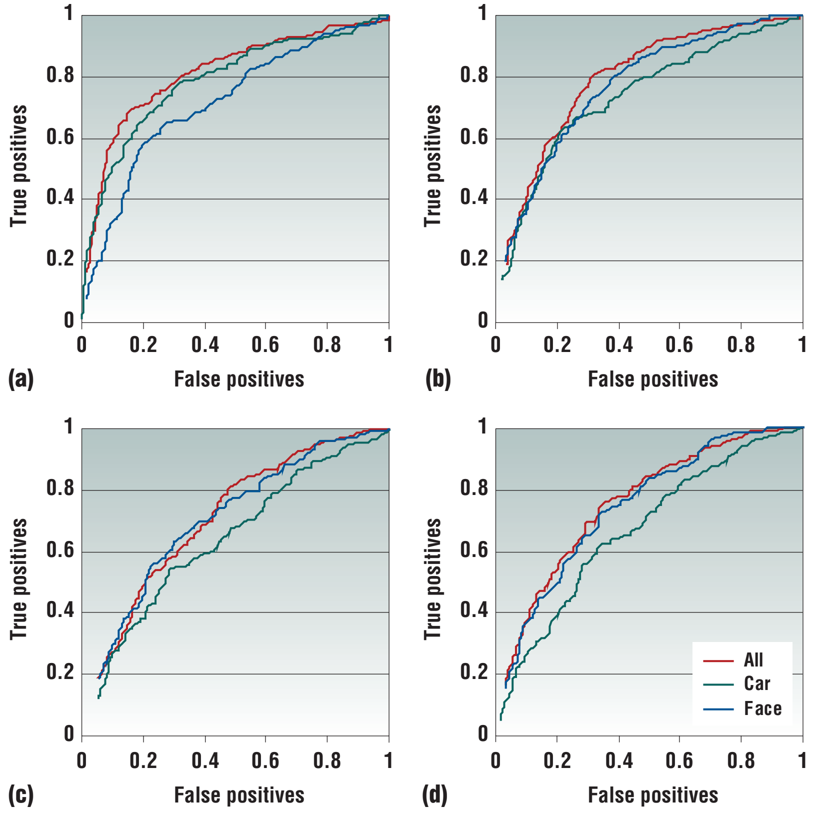


Figure 5. Receiver operating characteristic (ROC) plots of classifiers. We used car features, facial features, and all features to predict minor accidents one to four seconds before accidents occur: (a) one, (b) two, (c) three, and (d) four seconds.

For a closer view of these trends, we plotted ROC curves depicting true versus false positives for the classifiers using all features, only car features, and only facial features, at one to four seconds before accidents occurred. We did this for each class (major and minor) in isolation. This let us examine whether the facial signals were more predictive of major or minor accidents across the pre-accident time intervals. We used seven seconds of data for each plot given that our highest accuracies occurred using seven seconds of data (see Figures 8 and 9).

When viewing the ROC curves for the minor accident predictions, it becomes apparent that the face accounts for most of the predictive accuracy of the classifiers after three seconds prior to the accidents; the classifiers using only the facial features perform essentially the same as the classifiers using all the features in combination. However, the predictive accuracy for major accidents appears to come primarily from the vehicle features. This confirms the results we saw in our binary classifiers, where the facial features proved more helpful in predicting minor accidents than major accidents. Overall, this suggests that important signals for accident prediction exist in drivers' faces up to four seconds prior to accidents and that these signals are strongest for minor accidents.

Although our study proves encouraging for the prospect of using facial features to aid in driver accident

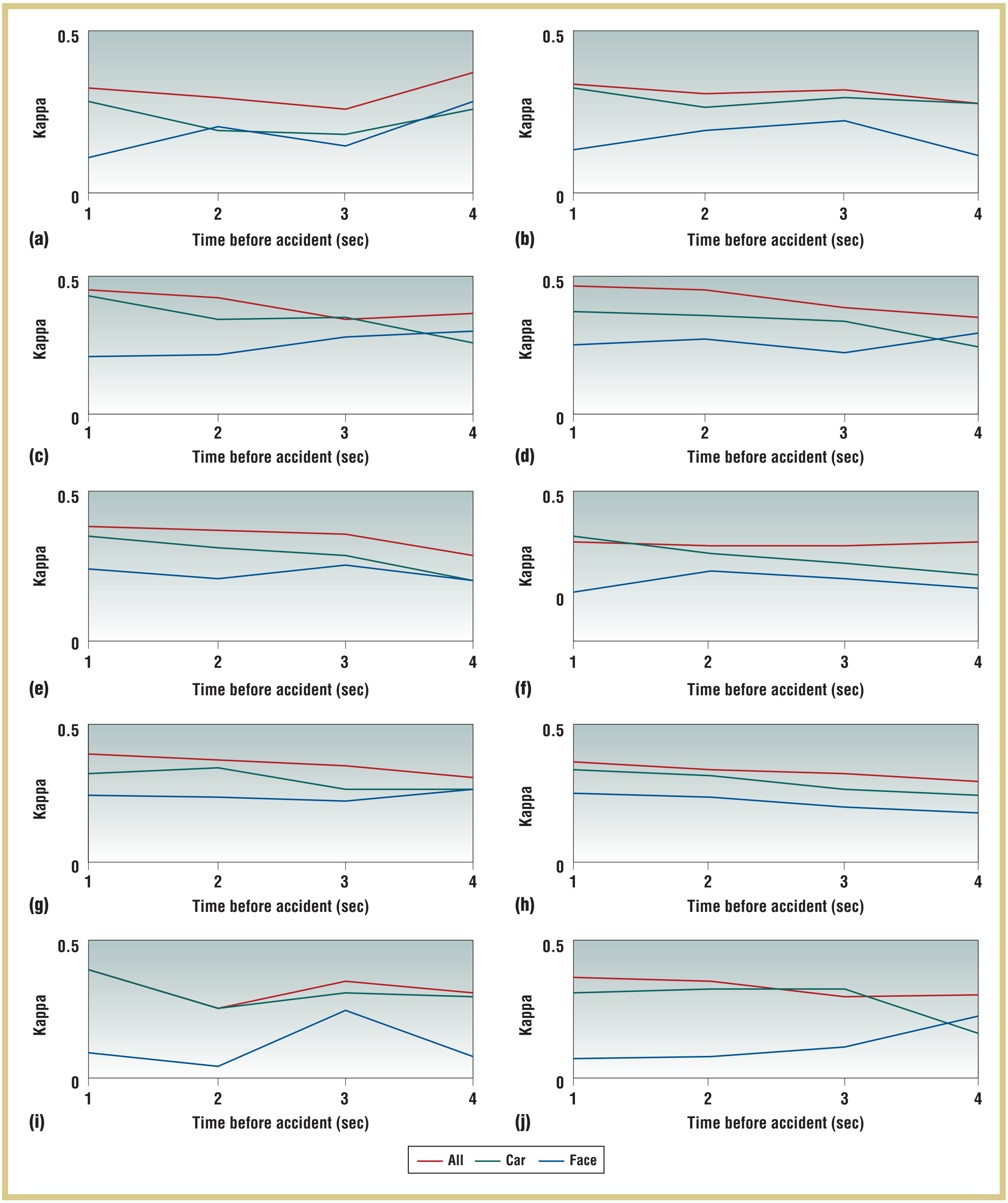


Figure 6. Major accident classifier performance. We used face, car, and all inputs to predict major accidents one to four seconds before accidents occur using statistics calculated over a range of intervals of data: (a) one, (b) two, (c) three, (d) four, (e) five, (f) six, (g) seven, (h) eight, (i) nine, and (j) 10 seconds. Kappa values over 0.2 are considered statistically significant.

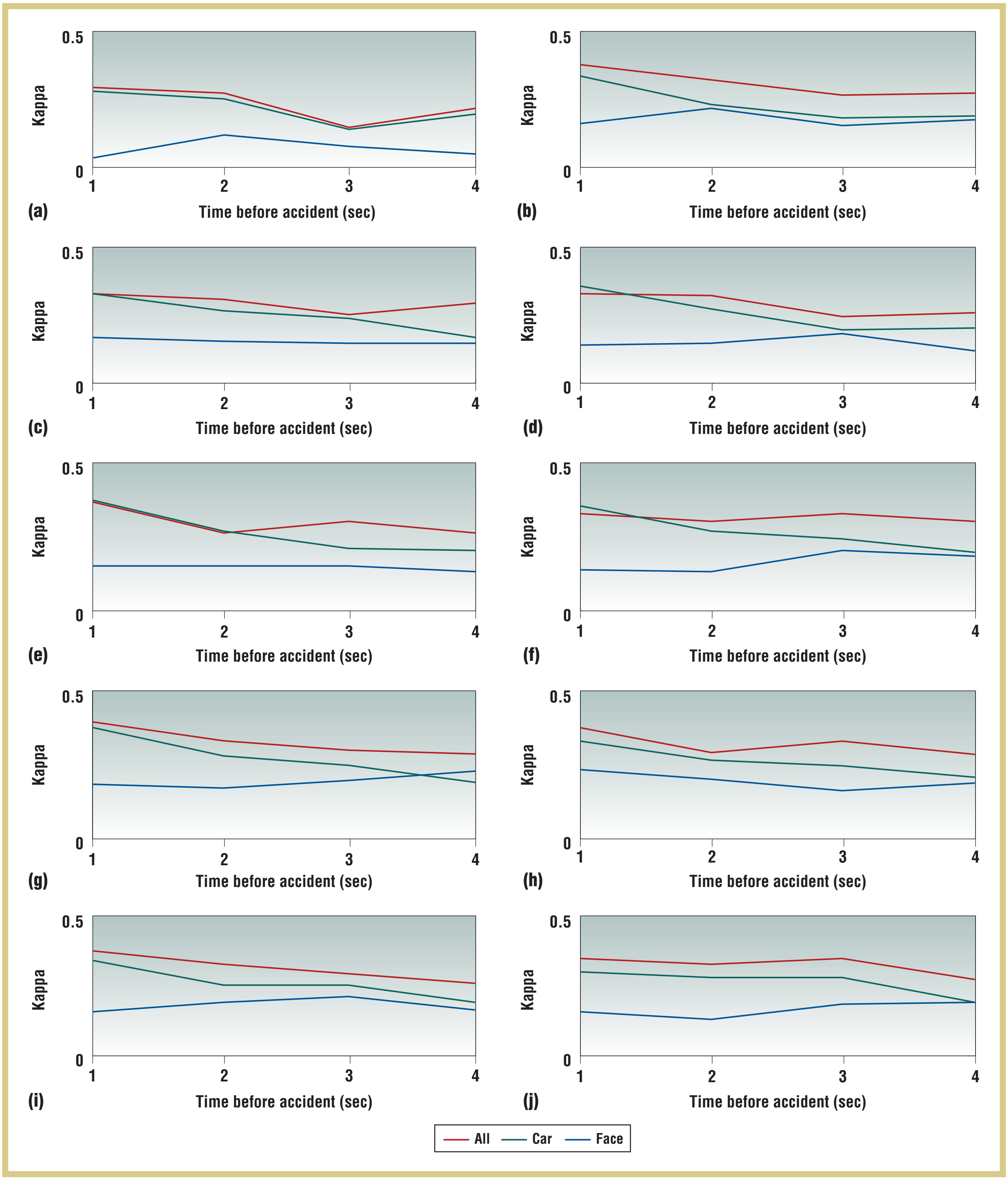


Figure 7. Minor and major accident classifier performance. We used face, car, and all inputs to predict major and minor accidents one to four seconds before accidents occur using statistics calculated over a range of intervals of data: (a) one, (b) two, (c) three, (d) four, (e) five, (f) six, (g) seven, (h) eight, (i) nine, and (j) 10 seconds. Kappa values over 0.2 are considered statistically significant.

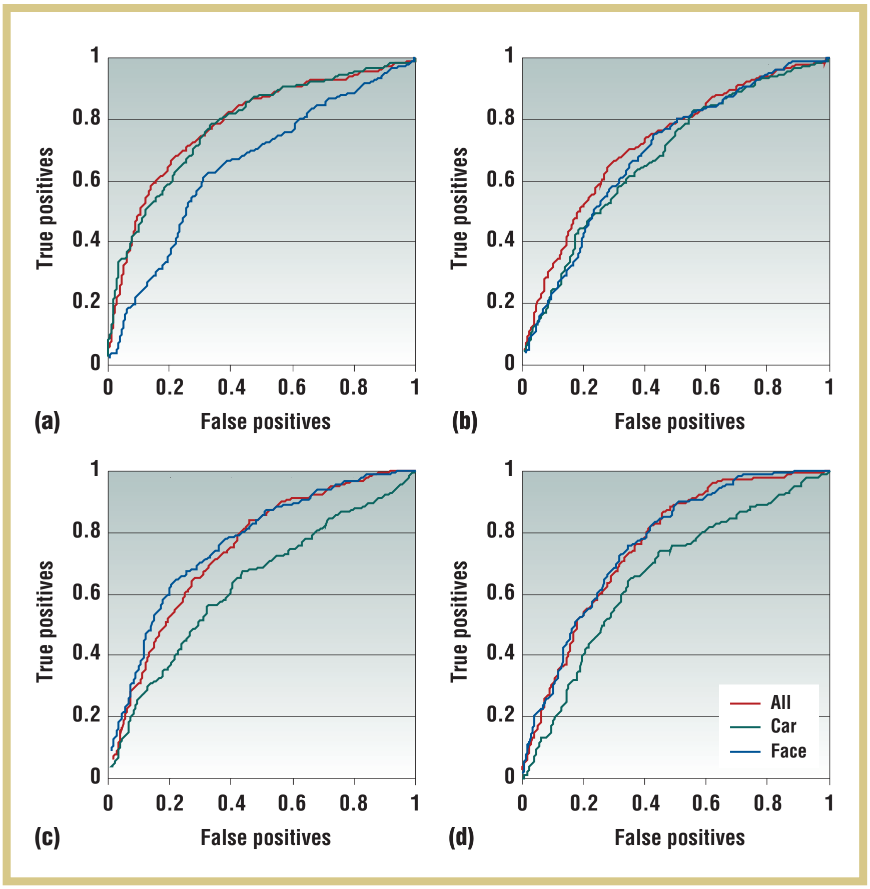


Figure 8. Minor accident ROC curves. We used car, face, and all features at pre-accident time intervals: (a) one, (b) two, (c) three, and (d) four seconds before the minor accident to predict the incident.

prediction, it does have several limitations. First, we did not run the study on the road because the study design was deemed too hazardous to run in the physical world until we had run the study in a simulator. Thus, our study ignores important parameters, such as vehicle motion and surround effects, that affect driver's perception and reaction to situations. This limits the generalizability of the study's findings, and we must do further work to determine the impact of these parameters.

Second, although this study thoroughly investigated ways of sensing impending accidents on the road, it didn't investigate exactly what a pervasive system could do to prevent that accident. The system could notify the driver, take action on its own, or do some combination of the two. These possible actions are open for future work.

Third, because we didn't implement our system in real time, we couldn't analyze the classifiers' false-alarm rate over time. However, given that Neven Vision can process streaming facial video at a rate of 30 frames per second, and that our feature calculations and classifier predictions can be made in under 500 milliseconds using Matlab on a standard 2.5-GHz Intel processor, we're confident that our system can be run in real time.

Finally, as with any statistical model, these results are limited to the specific features included in the original models and to these particular datasets. If we had sensed other aspects of the driver (for example, heart rate) or any other part of the driver-environment system, we might have generated very different models for predicting impending driver accidents. Thus, future work would benefit from including a wider range of sensor data to improve the accuracy of such driver safety support systems. Generating more datasets from other populations of participants and other driving contexts would also improve the study's generalizability. ■

ACKNOWLEDGMENTS

We thank Dan Merget for software development; Rabindra Ratan for helping to set up the driving simulator; Alexia Nielsen, Karmen Morales, Julia Saladino, and Anna Ho for running this experiment; and David Jabon, Grace Ahn, Doug Blumeyer, Steffi Paepcke, Jeff Wear, Kathryn Segovia, and Jesse Fox for helpful comments on earlier drafts of this article. This work was funded in part by US National Science Foundation grant SES-0835601 and a grant from Omron Corporation.

REFERENCES

1. U.S. Dept. of Transportation, "Traffic Safety Facts 2006: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System," tech. report DOTHS 810 818, Nat'l Highway Traffic Safety Administration, 2006.
2. M. Satyanarayanan, "Pervasive Computing: Vision and Challenges," *IEEE Personal Comm.*, vol, 8, no. 4, 2001, pp. 10–17.
3. Z. Zu, Q. Ji, and P. Lan, "Real-Time Nonintrusive Driver Fatigue Monitoring," *IEEE Trans. Vehicle Technology*, vol. 53, no. 4, 2004, pp. 1052–1068.
4. J.A. Healy and R.W. Picard, "Detecting Stress During Real-World Driving Tasks Using Physiological Sensors," *IEEE Trans. Intelligent Transportation Systems*, vol. 6, no. 2, 2005, pp. 156–166.
5. R.W. Allen et al., "A Low Cost PC Based Driving Simulator for Prototyping and Hardware-in-the-Loop Applications,"

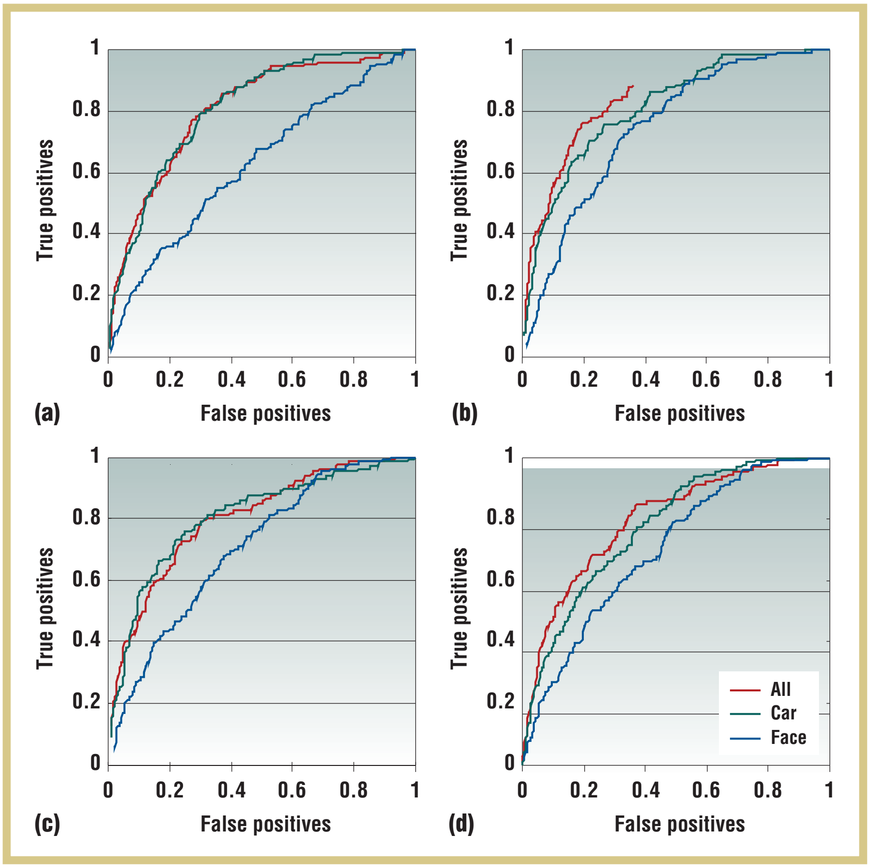
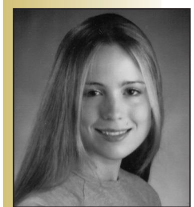


Figure 9. Major accident ROC curves. We used car, face, and all features at pre-accident time intervals: (a) one, (b) two, (c) three, and (d) four seconds before the major accident to predict the incident.

Proc. SAE Int'l Congress & Exposition, SAE Int'l, 1998, paper no. 98-0222.

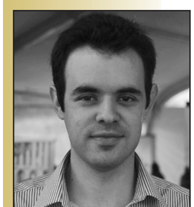
6. H. Neven, Neven Vision [computer software], Neven Vision, Inc., 2003.
7. J. Friedman, T. Hastie, and R. Tibshirani, "Additive Logistic Regression: A Statistical View of Boosting," *Annals of Statistics*, vol. 28, no. 2, 2000, pp. 337–407.
8. A. Ben-David, "What's Wrong with Hit Ratio?," *IEEE Intelligent Systems*, vol. 21, no. 6, 2006, pp. 68–70.
9. J. Landis and G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, 1977, pp. 159–174.
10. H.I. Witten and E. Fank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2005.

the AUTHORS



Maria Jabon is a software developer at LinkedIn. Her research interests include machine learning and Web interfaces for visualizing high-dimensional datasets. Jabon has a MS in electrical engineering from Stanford University. Contact her at mjabon@stanfordalumni.org.

Jeremy Bailenson is the founding director of the Virtual Human Interaction Lab, coauthor of the book Infinite Reality: Avatars, New Worlds, Eternal Life, and the Dawn of the Virtual Revolution, and an associate professor in the Department of Communication at Stanford University. His research interests include digital human representation, especially in the context of immersive virtual reality. Bailenson has a PhD in cognitive psychology from Northwestern University. Contact him at bailenso@stanford.edu.



Emmanuel (Manos) Pontikakis is a quantitative analyst at Citadel Investment Group. His research interests include privacy-preserving data mining, machine learning, human-computer interaction, and quantitative finance. Pontikakis has an MS in computer science from Stanford University. Contact him at pontikakis@gmail.com.

Leila Takayama is a research scientist at Willow Garage, studying human-robot interaction. Her research interests include interactions with and through personal robots, focusing on issues of embodied cognition and perceived agency. Takayama has a PhD from Stanford University's Communication between Humans and Interactive Media (CHIMe) lab, with a minor in psychology. She is a member of the Association for Psychological Science, Cognitive Science Society, and ACM Special Interest Group on Computer-Human Interaction. Contact her at takayama@willowgarage.com.

Clifford Nass is the Thomas M. Storke Professor at Stanford University, founder and director of the Communication between Humans and Interactive Media (CHIMe) Lab, and codirector of the Kozmetsky Global Collaboratory and Auto-X at Stanford University. His research focuses on (laboratory and field) experimental studies of social-psychological aspects of human-interactive media interaction, and his research interests include automobile interfaces, mobile interfaces, human-robot interaction, cognitive and social effects of multitasking, adaptive and personalized systems, and voice interfaces. Nass has a PhD in sociology from Princeton University. Contact him at nass@stanford.edu.