

Máquinas de soporte vectorial para inferir el punto de atención de automovilistas vistiendo lentes inteligentes

José M. Ramírez, Marcela D. Rodríguez, Ángel G. Andrade, Antonio Ordorica

Universidad Autónoma de Baja California, Facultad de Ingeniería, Mexicali, B.C.,
México

{jramirez53, marcerod, aandrade, aordorica}@uabc.edu.mx

Resumen. La mayoría de los métodos para inferir la distracción durante la conducción, se basan en características visuales de la postura de la cabeza, ya que es un fuerte indicador de la distracción durante la conducción. En este trabajo, se propone el uso de sensores inerciales empotrados en lentes inteligentes. Para ello, recopilamos datos de cinco participantes y realizamos experimentos para evaluar la viabilidad del uso de máquinas de soporte vectorial (SVM) para generar modelos de los conductores para inferir el punto de atención. Los resultados muestran un desempeño aceptable de la SVM para identificar las posiciones particulares de la cabina del automóvil donde los conductores centran su atención. Hasta el momento hemos obtenido una exactitud promedio de 83,42%.

Palabras clave: distracción, conducción; máquinas de soporte vectorial; lentes inteligentes, sensores inerciales.

Support Vector Machines to Infer the Point of Attention of Motorists Wearing Smart Glasses

Abstract. Most of the methods for inferring distraction while driving are based on visual characteristics of head posture, since it is a strong indicator of distraction while driving. This paper proposes the use of mounted inertial sensors on intelligent lenses. Hence, data was collected from five participants and experiments were conducted to evaluate the feasibility of using support vector machines (SVM) to generate models of drivers to infer the point of attention. Results show an acceptable performance of the SVM to identify specific positions of the car cabin on which drivers focus their attention. Until now, an average accuracy of 83.42% has been achieved.

Keywords: distraction, driving, support vector machines; intelligent lenses, inertial sensors.

1. Introducción

La distracción del conductor durante la conducción o la inatención se refiere a la falta de atención a las tareas de conducción debido a la participación en otras tareas durante la conducción. Estas otras tareas pueden ser cualquier distracción de la atención como distracciones cognitivas, físicas o visuales que conducen a la degradación del rendimiento [1]. Se estima que las distracciones causan el 23% de los accidentes o casi accidentes [2], y que podría reducirse un 10-20% mediante sistemas de monitoreo y predicción de los comportamientos de la conducción [3,4]. Evitar la distracción durante la conducción ha sido de particular interés al utilizar artefactos de navegación [5]. También se han explorado métodos para detectar la distracción del conductor y la somnolencia, que se basan principalmente en el análisis de las características abstraídas de imágenes y videos de las expresiones faciales, posición de la cabeza [4], y el comportamiento de la mirada [6].

En particular, la posición de la cabeza (o la orientación de la cabeza) es un fuerte indicador del campo de visión del conductor y el centro de atención actual [7, 8]; por lo tanto, se ha considerado como un proceso integral para monitorear el nivel de alerta del conductor [8]. Sin embargo, los métodos basados en el procesamiento de imágenes requieren que la cabina del vehículo sea equipada con cámaras [7, 8, 9]. Por ejemplo, faceLAB utiliza dos cámaras para determinar la posición de la cabeza y la dirección de la mirada del ojo [9]. Por otro lado, los lentes inteligentes han surgido como una interesante plataforma de investigación y de productos para una amplia gama de sistemas de apoyo portátiles debido a sus capacidades de detección. La seguridad en relación con el uso de los lentes inteligentes durante la conducción ha sido cuestionada, ya que permiten a los usuarios enviar mensajes de texto o acceder sus redes sociales a través de comandos de voz. Sin embargo, las investigaciones han demostrado que podrían mejorar la seguridad de los conductores que corren el riesgo de sufrir fatiga (por ejemplo, los conductores de camiones); y que quienes usan teléfonos inteligentes se distraen más y con más frecuencia que quienes utilizan Google Glass [10]. En este artículo, proponemos el uso de sensores inerciales incorporados en los lentes inteligentes para obtener características de la postura de la cabeza para inferir el foco de atención actual de los conductores. Para alcanzar este objetivo, se utilizaron los datos recolectados del acelerómetro, magnetómetro y giroscopio de un Google Glass. Estos datos fueron utilizados para generar modelos de la Máquina de Soporte Vectorial (SVM). Esta es una técnica de aprendizaje supervisado que busca un hiperplano óptimo para separar los datos entre clases definidas [11, 12]. El objetivo general de este proyecto es conocer cuál es el desempeño de la SVM y bajo qué condiciones, para determinar el foco de atención del conductor.

En la siguiente sección, se explican los métodos utilizados para recolectar datos y procesarlos para experimentar con la SVM. La sección 4 presenta el diseño del experimento, el cuál consistió de 3 ensayos en los que probamos diferentes condiciones relacionadas a las técnicas de entrenamiento y a los sensores utilizados. La sección 4 muestra los resultados obtenidos, y la sección 5 concluye y presenta las posibles direcciones futuras de este trabajo.

2. Recolección y procesamiento de datos

Se recolectaron datos de 5 sujetos (media de edad $M=32$; $DS= 11.2$ años), mediante un experimento controlado en condiciones naturalistas. Cada sujeto participó en una sesión en la que realizó un conjunto de tareas (ver Tabla 1), en un automóvil mientras usaba los Google Glass, y conduciendo en un estacionamiento privado. Los datos recogidos consistieron de vectores de 9 tuplas, correspondiente a los valores x , y y z del acelerómetro, giroscopios y magnetómetro.

Tabla 1. Tareas realizadas por los sujetos asociadas a la distracción causantes de riesgos [13].

Tarea Secundaria	Tareas realizadas durante el experimento
Uso del teléfono celular	Marcar (teclas rápidas), hablar / escuchar, enviar mensajes de texto
Interacción con un objeto	Mirar el objeto, mover un objeto, alcanzar un objeto
Bebiendo	Quitar tapadera al recipiente
Uso de la consola de confort e información	Encender / Apagar el clima, ajustar la temperatura, insertar / retirar CD, ajuste de radio



Fig. 1. Posiciones de la cabina usadas como clases (Posición 0-10).

Tabla 2. Instancias capturadas por cada posición de todos los sujetos.

S	Muestra		Posiciones										
	D	DE	0	1	2	3	4	5	6	7	8	9	10
S1	3,443	1,877	1,064	122	434	17	16	89	20	8	97	7	3
S2	3,138	1,203	466	281	238	87	7	51	49	8	6	5	5
S3	2768	1351	933	2	256	26	56	30	24	13	0	4	7
S4	2000	898	463	14	308	23	33	15	17	7	8	3	7
S5	2042	1074	489	37	384	38	0	36	21	18	23	9	19
TOTAL	13,391	6,403	3,415	456	1,620	191	112	221	131	54	134	28	41

La frecuencia de muestreo del acelerómetro se estableció en 15.5 Hz, la más rápida a través de la API de Android, lo que nos permitió obtener un promedio de 15.5 lecturas por segundo. Todas las sesiones fueron grabadas en video para posteriormente etiquetar los datos para poder utilizar técnicas de aprendizaje supervisado como la SVM [11, 12]. Etiquetamos los datos usando diez clases de posiciones relacionadas con las ubicaciones de la cabina del vehículo donde los participantes centraron su atención conforme realizaban las tareas solicitadas. Como se muestra en la Figura 1, estas clases corresponden a diferentes posiciones de la cabina del vehículo. Por ejemplo, un conductor podría centrar su atención en el punto 5 cuando cambia la estación de radio. Por lo tanto, se evaluó la viabilidad de identificar puntos particulares donde los conductores enfocaron su atención, lo que ayudaría a identificar si está llevando a cabo una tarea secundaria.

3. Descripción del experimento

La plataforma utilizada para generar un modelo SVM fue WEKA (Waikato Environment for Knowledge Analysis), ya que ofrece una colección de algoritmos de aprendizaje automático para tareas de minería de datos. WEKA contiene herramientas para pre procesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización. [14] Adicionalmente se utilizó la funcionalidad de WEKA para normalizar los datos.

Este experimento consistió de 3 ensayos para identificar bajo qué condiciones la SVM logra un mejor desempeño de clasificación.

Para cada ensayo se reportan la Exactitud (*Accuracy*), la Precisión (*Precision*) y el Recuerdo (*Recall*) como las métricas del desempeño de la clasificación [11,12]. Precisión es la proporción de elementos clasificados como positivos que son realmente verdaderos positivos [16]:

$$Precisión = \frac{vp}{vp+fp}, \quad (1)$$

donde vp son los verdaderos positivos y fp los falsos positivos.

Recuerdo (*Recall*) es la proporción de elementos verdaderos positivos que son correctamente clasificados como positivos [16]:

$$Recuerdo = \frac{vp}{vp+fn}, \quad (2)$$

donde fn son los falsos negativos.

Exactitud (*Accuracy*): que es una medida de eficiencia generalizada para evaluar el desempeño de un clasificador, y se refiere a las instancias correctamente clasificadas [16]:

$$Exactitud = \frac{vp+vn}{vp+vn+fp+fn}, \quad (3)$$

donde vn son los verdaderos negativos.

Cabe mencionar que se generó un modelo de clasificación para cada sujeto, ya que en experimentos previos se identificó que este logra mayor precisión en comparación con utilizar un modelo generalizado, obtenido con datos de todos los sujetos [15].

3.1. Ensayo 1

El objetivo fue determinar el desempeño de la SVM generando modelos entrenados con las técnicas "Split-Percent" y "Cross-Validation" las cuáles son las más utilizadas. Para el entrenamiento "Split Percent" se utilizó un 66% de datos para aprender y el 33% restante para clasificar. Para el entrenamiento "Cross Validation", el cual consiste en dividir los datos suministrados a la SVM en n particiones y por cada una de ellas, construir el clasificador con las $n-1$ partes restantes, e ir probando con cada una de las particiones, donde el valor de n de este primer ensayo fue 10. Dado que por defecto WEKA utiliza el Polikernel, decidimos utilizarlo para el primer experimento.

3.2. Ensayo 2

El objetivo de este ensayo fue identificar con cuál de los kernels comúnmente utilizados (Polinomial, Normalized, RBFKernel y PUK) se logra un mejor desempeño de clasificación.

3.3. Ensayo 3

Utilizar Google Glass para inferir inatención del conductor, representa un reto en cuanto al buen uso de la batería, de manera similar que en otros dispositivos vestibles. Similar a otros trabajos [17,18], esto lleva a preguntarnos ¿cuál es el desempeño de la SVM si se limita la cantidad de sensores? Es por esto que el objetivo de este ensayo fue determinar cuál sensor o combinación de sensores entre los disponibles en Google Glass, mejora el desempeño de la SVM para identificar el punto de atención del conductor. En este ensayo se realizaron pruebas de clasificación con datos de un sensor, parejas de sensores y la combinación de todos ellos.

4. Resultados

Dado que el objetivo de realizar estos tres ensayos fue conocer cuáles condiciones permiten obtener una mejor clasificación de los datos; se decidió que las condiciones del modelo que ofreciera un mejor desempeño serían las que se utilizarían en el posterior ensayo. Es decir, la condición del ensayo 1 que arrojó el mejor desempeño, se usó para el ensayo 2, y así sucesivamente.

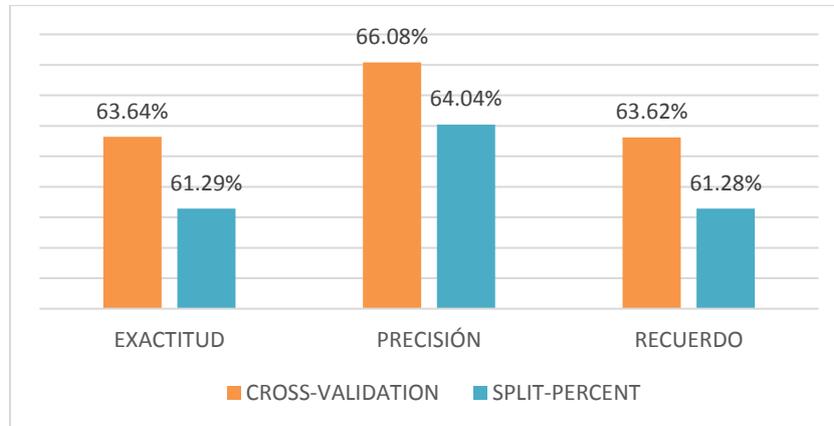


Fig. 2. Comparación del desempeño de la SVM utilizando dos técnicas de entrenamiento.

Tabla 3. Matriz de confusión del sujeto 5 correspondiente al ensayo 1 con la técnica Cross-Validation.

Instancias Correctamente Clasificadas 52.1701 %, Instancias Incorrectamente Clasificadas 47.8299 %, Precisión 0.526, Recuerdo 0.522

a	b	c	d	e	f	g	h	i	<-- classified as
224	43	19	0	46	39	0	13	0	a=Pos2
85	136	21	0	43	39	20	40	0	b=Pos5
18	22	120	10	44	84	20	66	0	c=Pos3
0	0	20	240	0	0	80	0	44	d=Pos10
31	11	7	0	288	36	11	0	0	e=Pos1
36	75	36	0	0	179	4	54	0	f=Pos6
0	0	42	118	0	0	213	0	11	g=Pos9
15	73	17	0	0	161	0	118	0	h=Pos8
0	0	0	78	0	6	15	0	285	i=Pos7

4.1. Ensayo 1

Los datos que se muestran en la figura 2 corresponden a los promedios de los resultados de los 5 modelos SVM, correspondientes a los 5 sujetos. Se observa que *Cross-Validation* permitió obtener el mejor desempeño de clasificación (*Accuracy*=63.64%) comparado con *Split-Percent* (*Accuracy*=61.29%), En consecuencia, para los ensayos posteriores se utilizó esta técnica de entrenamiento.

A manera de ejemplo se muestra la matriz de confusión del sujeto 5 (tabla 3), donde se aprecia que la SVM confundió las posiciones 5 que tiene a confundirla con la posición 6 y la 9 que tiene a confundirla con la posición 10, esto se debe a que la posición 5 y 6 se encuentran muy cerca al igual que las posiciones 9 y 10 (ver figura 1).

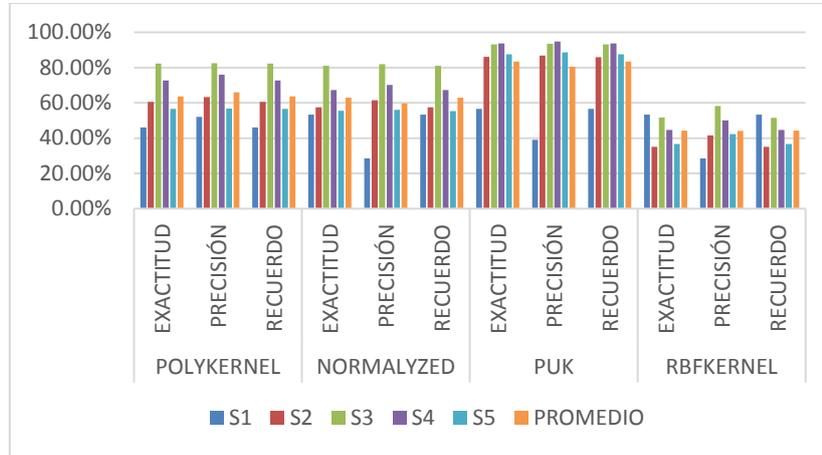


Fig. 3. Comparación de los 4 kernels más comunes.

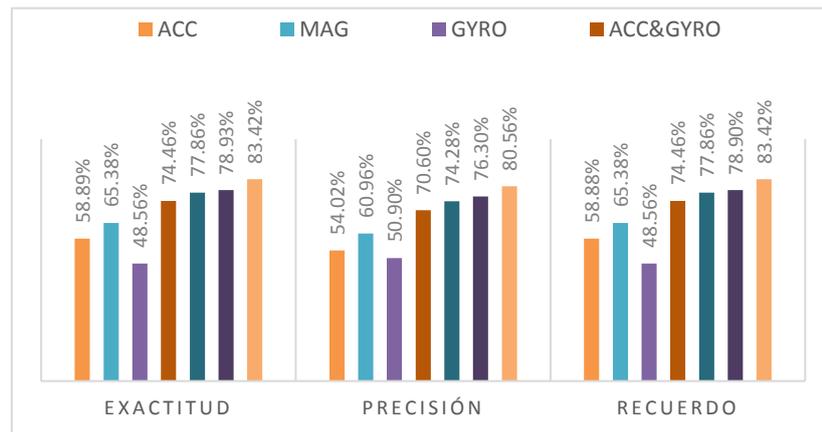


Fig. 4. Desempeño de la SVM usando datos de cada sensor y sus combinaciones.

4.2. Ensayo 2

De la comparación entre los cuatro kernels comúnmente utilizados, obtenemos los siguientes resultados, como podemos ver en la figura 3 el kernel que proporciona una mejor clasificación fue el PUK con un promedio en la exactitud de 83.42%

4.3. Ensayo 3

Los resultados que se muestran a continuación se obtuvieron de la técnica de clasificación Cross-Validation, y el kernel PUK. Como se observa en la figura 4, la mejor

Tabla 4. Matriz de confusión del sujeto 5 correspondiente al ensayo 3 utilizando todos los sensores.

Instancias Correctamente Clasificadas 87.5579 %, Instancias Incorrectamente Clasificadas 12.4421 % ,
Precisión 0.887 Recuerdo 0.876

a	b	c	d	e	f	g	h	i	<-clasificado como
234	30	26	1	35	28	0	30	0	a=Pos2
16	284	9	0	11	32	0	32	0	b=Pos5
4	20	280	0	0	40	0	40	0	c=Pos3
0	0	0	384	0	0	0	0	0	d=Pos10
0	0	0	0	384	0	0	0	0	e=Pos1
0	0	0	0	0	367	0	17	0	f=Pos6
0	0	0	0	0	0	384	0	0	g=Pos9
0	5	0	0	0	33	0	346	0	h=Pos8
0	0	0	0	0	21	0	0	363	i=Pos7

Tabla 5. Comparación con el trabajo más similar al nuestro.

	Nuestro trabajo	[21]
Zonas de la cabina	10 clases	6 clases
Técnicas de predicción	SVM	Random Forest
Entrenamiento	10 folds, cross-validation	10 folds, cross-validation
Sujetos	5	7
Desempeño de clasificación	Exactitud=71.1132 %,	Exactitud= 79.4%

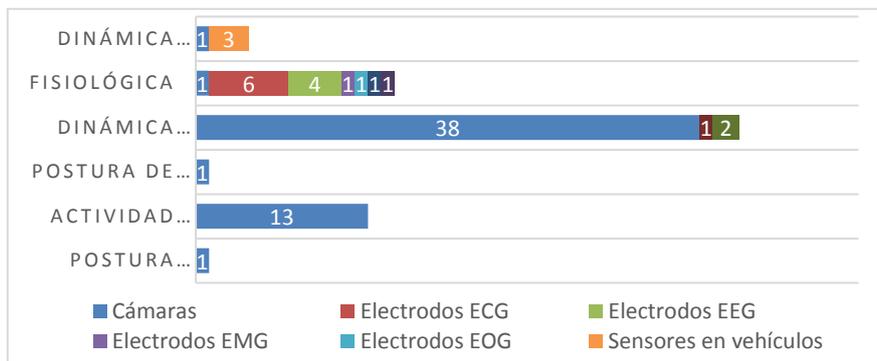


Fig. 5. Tipos de sensores v Enfoque utilizados en los trabajos similares al nuestro.

clasificación se obtiene al utilizar la combinación de los 3 sensores, con una Exactitud promedio de 83.42%. Sin embargo, utilizar solamente los datos del giroscopio y magnetómetro, el porcentaje de instancias correctamente clasificadas es de 78.93%, lo cual sugiere que es una combinación de sensores que podría ser apropiado a utilizar.

Adicionalmente podemos observar en la tabla 4 que se disminuye la confusión de las posiciones 5 con 6 y 9 con 10 en comparación con la tabla 3

5. Comparación con trabajos relacionados

Con el propósito de identificar trabajos similares a nuestro enfoque, recientemente realizamos un mapeo de literatura a partir de la cual obtuvimos 75 trabajos que están enfocados en la predicción de la postura de la cabeza durante la conducción, la figura 5 muestra la cantidad de artículos y los diferentes enfoques que se utilizan para esta predicción. Como se observa en la tabla, son 41 de un total de 75, los trabajos que detectan la postura de la cabeza. De estos 41, solamente se identificaron 2 que exploran el uso de sensores inerciales. El primero presenta resultados preliminares del desempeño del HMM para determinar solamente 7 posturas de la cabeza, con base a la guiñada (yaw), cabeceo (pitch) y alabeo(roll) que no están asociadas a zonas de enfoque particulares de la cabina [19]. El segundo, se limita a proponer detectar los ángulos asociados al giro de la cabeza utilizando solamente un sensor inercial (giroscopio) [20]. Sin embargo, de los 38 artículos que exploran el uso de cámaras para detectar la postura de la cabeza, solo uno propone detectar la orientación de la cabeza con base a las zonas de la cabina [21]. Dada la similitud en cuanto a las variables que se proponen inferir (orientación de la cabeza hacia zonas de la cabina), a continuación, presentamos una comparación de nuestros resultados con respecto a este.

Para poder compararnos realizamos un entrenamiento generalizado, esto es, generar un modelo con base a los datos de todos los usuarios, similar al realizado en [21]. Se puede observar en la tabla que nuestra exactitud es ligeramente menor que la reportada en [21]. Esto podría deberse a que es una menor cantidad de clases y a la poca variabilidad de nuestros datos debido al desbalanceo obtenido durante la recolección.

5.1. Conclusiones

Nuestros resultados muestran que inferir la distracción durante la conducción, basándonos en datos recolectados de sensores inerciales de lentes inteligentes es una solución viable.

Similar a otros trabajos [18], podemos observar que no existe compromiso entre la precisión y el recuerdo al variar las diferentes condiciones de los ensayos. Esto sugiere que en nuestro caso la disposición fija de los sensores (montados en la cabeza), además del movimiento del vehículo no influye en la tendencia de las tres métricas de desempeño.

Los resultados demuestran que el uso de datos crudos (vectores de 9-tuplas) es una opción adecuada para evitar realizar un procesamiento adicional en el Google Glass que detrimento el uso de la batería.

Adicionalmente observamos que el uso de la combinación de dos sensores (gyro y mag), podría ser una opción viable a explorar en futuros experimentos

Se planea realizar experimentos adicionales para obtener una recolección de datos balanceada, para lo cual estamos considerando diseñar un experimento semi-controlado en lugar de realizarlo en condiciones naturalistas.

Agradecimientos. Se agradece a Luis Castro y Jessica Beltrán por su retroalimentación para conducir estos experimentos. Y a los voluntarios que participaron en el estudio.

Referencias

1. Sajan, S., Ray, G. G.: Human Factors in Safe Driving - A Review of Literature on Systems Perspective. Distractions and Errors, Proceedings IEEE Global Humanitarian Technology Conference, pp. 83–88 (2012)
2. Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., Ramsey, D. J.: The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. National Highway Traffic Safety Administration, USDOT (2006)
3. Bayly, M., Fildes, B., Regan, M., Young, K.: Review of crash effectiveness of intelligent transport system. Traffic Accident Causation in Europe (TRACE) (2007)
4. Kang, H. B.: Various Approaches for Driver and Driving Behavior Monitoring: A Review. Proceedings IEEE International Conference on Computer Vision Workshops (ICCVW '13), IEEE Computer Society, pp. 616–623 (2013)
5. Tarqui, G., Castro, L. A., Favela, J.: Reducing Drivers' Distractions in Phone-Based Navigation Assistants Using Landmarks. Ubiquitous Computing and Ambient Intelligence, (UCAMI LNCS) 8276, Springer, pp. 342–349 (2013)
6. Chamberlin, J.: Smart glasses: Driver distraction or safety tool?. 45(3), pp. 12, Note available at: <http://www.apa.org/monitor/2014/03/smart-glasses.aspx> (2014)
7. Zhang, L., Liu, F., Tang, J.: Real-Time System for Driver Fatigue Detection by RGB-D Camera. (ACM), Trans. Intell. Syst. Technol., 6(2), pp. 22 (2015)
8. Murphy-Chutorian, E., Manubhai, M.: Head Pose Estimation and Augmented Reality Tracking: An Integrated System and Evaluation for Monitoring Driver Awareness. IEEE Translate on Intelligent Transportation Systems, 11(2), pp. 300–311(2010)
9. Seeing machines Driver State Sensor: <https://www.seeingmachines.com/solutions/>
10. Zhang, Y. F., Gao, X. Y., Zhu, J. Y., Zheng, W. L., Lu, B. L.: A novel approach to driving fatigue detection using forehead EOG. Neural Engineering (NER), 7th International IEEE/EMBS Conference, pp. 707–710 (2015)
11. Mohri, M., Rostamizadeh, A., Talwalkar, A.: Foundations of Machine Learning. The MIT Press (2012)
12. Mitchell, T. M.: Machine Learning, 1st. Edition, Mc Graw Hill Higher Education (1997)
13. NHTSA: National Motor Vehicle Crash Causation Study Report to Congress. DOT HS 811 059, National Highway Traffic Safety Administration (2008)
14. WEKA: <http://www.cs.waikato.ac.nz/ml/weka/index.html>
15. Ordorica, A., Rodríguez, M. D., Castro, L. A., Beltran, J.: Support Vector Machines for Inferring Distracted Behavior of Drivers Wearing Smart Glasses. To be in the Springer Proceedings of Ubiquitous Computing and Ambient Intelligence, (UCAMI) (2016)
16. Michie, D., Spiegelhalter, D. J., Taylor, C. C.: Machine learning, neural and statistical classification (1994)

17. Bulling, A., Blanke, U., Schiele, B.: A tutorial on human activity recognition using body-worn inertial sensors. (ACM) Comput. Surv., 46(3), art. 33, pp. 1–33, DOI:10.1145/249962133 (2014)
18. Ozlem-Durmaz, I.: Analysis of movement, orientation and rotation-based sensing for phone placement recognition. Sensors 15(10), pp. 25474–25506, DOI:10.3390/s151025474 (2015)
19. Chuang, C. F., Yang, C. H., Lin, Y. H.: HMM-based driving behavior recognition for in-car control service. IEEE International Conference on Consumer Electronics - Taiwan, pp. 258–259, DOI:10.1109/ICCE-TW.2015.7216886 (2015)
20. Wei-Yao, C., Chung-Hsien, Y., Hsiao-Chien, T., Yi-Chun, L., Chun-Fu, C., Kao-Hung C.: Driver distraction recognition based on dual compass motion sensing. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 1375–1380. DOI: 10.1109/ITSC.2014.6957879 (2014)
21. Tawari, A., Trivedi, M. M.: Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos. IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, pp. 344–349. DOI:10.1109/IVS.2014.6856607 (2014)