CSFS-SFS: یک روش انتخاب ویژگی ترکیبی جدید برای تشخیص خواب‌آلودگی حین رانندگی با استفاده از سیگنال‌های EEG، EOG و کیفیت رانندگی

سید محمدرضا نوری[۱]، محمد میکائیلی[۲]، سیده نغمه میری آشتیانی[۳] و زهرا مردی[۴]
[۱، ۲، ۳، ۴] دانشگاه شاهد، smrezanoori@gmail.com

چکیده- طبق آمار، خواب‌آلودگی از دلایل اصلی بسیاری از تصادفات رانندگی است. هدف از این تحقیق ارائه یک روش برای تشخیص خواب‌آلودگی در سیگنال‌های EEG، EOG و سیگنال-های کیفیت رانندگی ثبت شده است. این روش از آنالیز حوزه زمان، طیف فرکانسی و تبدیل موجک استفاده می‌کند. به منظور انتخاب بهترین زیرمجموعه از ویژگی‌های استخراج شده، روش جدیدی با ترکیب دو روش CSFS و SFS معرفی شده است که ضمن کاهش زمان محاسبات، هزینه‌های ساخت یک سیستم عملیاتی و تجاری برای تشخیص خواب‌آلودگی حین رانندگی را کاهش داده و نتایج را نیز، نسبت به روش مرسوم SFS، بدون تغییر نگه می‌دارد. روش پیشنهادی می‌تواند در سیستم‌های تشخیص خواب‌آلودگی خودکار در وسایل نقلیه، به منظور کاهش تصادفات جاده‌ای مورد استفاده قرار بگیرد.

کلمات کلیدی- خواب‌آلودگی حین رانندگی، انتخاب ویژگی، جدایی‌پذیری کلاسی، کلاس‌بندی.

۱- مقدمه

حین خواب‌آلودگی، توانایی فرد در حفظ هوشیاری و تصمیم‌گیری صحیح به طرز قابل توجهی کاهش می‌یابد. این وضعیت برای رانندگان می‌تواند خطرات و مشکلاتی را به همراه داشته باشد. به گزارش سازمان بهداشت جهانی مرگ و میر جاده‌ای در دنیا به ۱/۲۴ میلیون نفر در سال رسیده است که پیش‌بینی می‌شود تا سال ۲۰۳۰ این رقم به ۳/۶ میلیون نفر افزایش یابد [۱]. به همین خاطر تشخیص خواب‌آلودگی حین رانندگی و ساخت دستگاهی خودکار برای این کار اهمیت ویژه‌ای دارد.

روش‌های تشخیص خواب‌آلودگی رابطه نزدیکی با رفتار و عکس-العمل‌های راننده دارد [۲ و ۳]. برخی محققین تکنیک‌هایی را مبتنی بر حرکات چشم و سر برای تشخیص خواب‌آلودگی حین رانندگی پیشنهاد داده‌اند [۴، ۵]. در این روش‌ها به یک دوربین برای ثبت حرکات بدن فرد نیاز است [۲]. اگرچه روش‌های مبتنی بر تصویر روش‌های مناسبی هستند، اما تقریباً همه راننده‌ها تمایلی به ثبت همیشگی تصویر بدنشان با دوربین ندارند. این در حالی است که روش‌های مبتنی بر سیگنال‌های حیاتی با مخالفت‌های کمتری همراه است [۶، ۷].

سیگنال‌های حیاتی به ویژه برای جمع‌آوری اطلاعات مربوط به پاسخ بدن حین خواب‌آلودگی بسیار مفید هستند. به منظور تشخیص خواب-آلودگی، برخی محققین از سیگنال‌های الکتروکاردیوگرام [۸] یا سیگنال-های الکتروأکلوگرام [۹] استفاده کرده‌اند. الکتروانسفالوگرافی (EEG) گسترده‌ترین تکنیک مورد استفاده برای اندازه‌گیری فعالیت الکتریکی مغز است [۱۰]. چندین محقق تجزیه و تحلیل EEG را برای شناسایی حالت خواب‌آلودگی[۱] (DS) پیشنهاد کرده‌اند [۶، ۷].

در کنار استفاده از سیگنال‌های حیاتی برای تشخیص خواب‌آلودگی حین رانندگی، گروهی از محققین تلاش کرده‌اند تا با تحلیل سیگنال‌های به دست آمده از پارامترهای خودرو و سیگنال‌های کیفیت رانندگی[۲] (DQ) فرد این اتفاق را تشخیص بدهند [۱۱]. در این تحقیقات فرض می‌شود که الگوی رفتاری راننده حین هوشیاری و خواب‌آلودگی با هم تفاوت دارد.

هدف این تحقیق پیشنهاد یک روش خودکار برای تشخیص DS در سیگنال‌های EEG، EOG و DQ ثبت شده می‌باشد. در این روش از آنالیز زمانی، طیف فرکانسی و تبدیل موجک استفاده شده است. برای جداسازی DQ از حالت هوشیاری[۳] (AS)، ۷۵۶ ویژگی از کانال‌های به ثبت رسیده محاسبه و آنالیز شده است. بعد از فرآیند انتخاب ویژگی، تعداد متفاوتی ویژگی انتخاب شده و به کلاس‌بند اعمال شد.

تمرکز اصلی ما در این مقاله، در بخش پردازش، به کارگیری یک شیوه جدید برای انتخاب بهترین ویژگی‌ها بوده است. این روش که از ترکیب دو روش انتخاب ویژگی ترتیبی جلوسو[۴] (SFS) و روش انتخاب ویژگی مبتنی

---

۱ Drowsy State
۲ Driving Quality
۳ Alert State
۴ Sequential Forward Selection

بر جدایی‌پذیری کلاسی[۵] (CSFS) تشکیل شده، با نام CSFS-SFS نام گذاری شده است. در بخش‌های بعدی مقاله به معرفی مجموعه داده‌ها، نحوه استخراج و انتخاب ویژگی می‌پردازیم. سپس نتایج و بحث‌های مربوطه آورده شده است.

۲-جمع‌آوری داده‌ها

داده‌های این مقاله در آزمایشگاه پردازش سیگنال‌های بیولوژیکی، واقع در دانشگاه شاهد، و با استفاده از شبیه‌ساز رانندگی ثبت شده است. برای ثبت داده‌های EEG و EOG از دستگاه ۱۶ کاناله g.USBamp شرکت gtec اتریش و برای ثبت داده‌های DQ از یک نرم‌افزار شبیه‌ساز رانندگی[۶] استفاده شده است. ثبت سیگنال از ۶ نفر از دانشجویان مرد به صورت داوطلبانه انجام شد. برای تشخیص بهتر الگوهای خواب‌آلودگی از هوشیاری، داوطلبین قبل از انجام آزمایش، به طور میانگین، حدود ۲۰ ساعت محرومیت از خواب را تجربه کردند.

در حین انجام آزمایش، به منظور ارزیابی هر چه بهتر میزان خواب-آلودگی از آزمون مقیاس خواب‌آلودگی کارولینسکا[۷] (KSS) استفاده شد [۱۲]. این مقیاس تمام بازه بین بیداری تا ابتدای خواب را در ۹ مرحله (عددی بین ۱ تا ۹) تقسیم‌بندی می‌کند. بر این اساس، عدد ۱ معرف وضعیت به شدت هوشیار و عدد ۹ به معنی خیلی خواب‌آلوده است. هر ۱۰ دقیقه، از ابتدای شروع رانندگی، با ظاهر شدن یک مستطیل در پایین صفحه، راننده موظف به اعلام یک عدد به عنوان آزمون KSS بود.

داده‌های EEG و EOG با فرکانس ۲۵۶ Hz و براساس استاندارد الکترودگذاری ۲۰-۱۰ ثبت شد. تعداد ۱۵ کانال برای ثبت EEG و یک کانال برای ثبت سیگنال EOG از چشم چپ مورد استفاده قرار گرفت. کانال‌های ثبت EEG عبارتند از: CZ، PZ، O2، P4، F4، FPZ، C4، T6، F8، O1، O3، FP1، T5، T3 و F7. محل FZ به عنوان زمین دستگاه و A2 به عنوان مرجع ثبت سیگنال به کار گرفته شد. فیلترهای داخلی دستگاه برای ثبت سیگنال شامل یک فیلتر میانگذر با فرکانس قطع ۰/۵ Hz تا ۳۰ Hz و یک فیلتر ناچ ۵۰ Hz بود. فیلتر ناچ ۵۰ Hz برای تضعیف هر چه بیشتر سیگنال‌ها در محدوده فرکانس برق شهر به کار گرفته شد. مدت زمان ثبت سیگنال برای هر فرد حدود ۵۰ دقیقه بود.

سیگنال‌های کیفیت رانندگی که توسط نرم‌افزار شبیه‌ساز رانندگی ثبت می‌شود عبارتند از: سیگنال موقعیت شامل سه مؤلفه X، Y و Z، سیگنال جهت خودرو شامل چهار مؤلفه X، Y، Z و W، سرعت، زاویه فرمان، موقعیت پدال گاز و موقعیت پدال ترمز. فرکانس ثبت سیگنال‌های کیفیت رانندگی در حدود ۱۷ Hz می‌باشد که برای همزمان‌سازی با سیگنال‌های حیاتی به فرکانس ۲۵۶ Hz نرخ‌افزایی می‌شوند.

به منظور بهبود ذخیره‌سازی رخدادها حین ثبت، تصویر چهره فرد با استفاده از یک دوربین وب‌کم ضبط شد. همچنین فیلم مربوط به رانندگی توسط کامپیوتری که نرم‌افزار شبیه‌ساز بر روی آن نصب شده بود، به ثبت رسید. به این ترتیب برای لیبل‌گذاری داده‌های به دست آمده از نتایج مربوط به تست KSS، دو فیلم ذکر شده و نیز گزارش‌های به ثبت رسیده حین انجام آزمایش استفاده شد. این گزارش‌ها شامل مواردی از قبیل حرکت‌های ناگهانی فرد و وضعیت خاص بدن می‌باشد. بعد از لیبل‌گذاری داده‌های ثبت شده، از مجموعه داده‌ای با ۴۹۰ مشاهده، شامل ۲۴۵ مشاهده در هر کلاس، برای تجزیه و تحلیل استفاده شد.

۳-استخراج ویژگی

سیگنال‌های ثبت شده به بخش‌های۳ ثانیه‌ای بخش‌بندی شد. این بخش‌بندی زمانی برای حصول اطمینان از ایستان بودن سیگنال، به لحاظ آماری، نیاز است. تعداد ۷۵۶ ویژگی، شامل ۴ ویژگی زمانی، ۹ ویژگی از حوزه فرکانس و ۱۵ ویژگی از حوزه تبدیل موجک سیگنال، برای همه ۲۷ کانال EEG، EOG و DQ استخراج شد. در حوزه زمان این ویژگی‌ها شامل ماکزیمم، مینیمم، انحراف معیار و میانگین می‌باشد.

برای استخراج ویژگی‌های مربوط به حوزه فرکانس، ابتدا مقدار میانگین هر بخش ۳ ثانیه‌ای حذف شد، سپس PSD هر بخش با استفاده از روش Burg مرتبه ۲۰ تخمین زده شد [۱۳]. ویژگی‌های استخراج شده از PSD عبارتند از: فرکانس مرکزی (CF)، فرکانس قله (PF)، نسبت H به L (RH/L)، که در آن H و L، به ترتیب، انتگرال قدرمطلق سیگنال تا فرکانس ۳۰ Hz و ۷ Hz است، فرکانس چارک اول و سوم (Q1F و Q3F)، انحراف معیار (SSD)، محدوده میان‌چارکی (IR)، ضریب چولگی[۸] (SC) و درجه اوج[۹] (KC).

۱۵ ویژگی از هر بخش با استفاده از تبدیل موجک استخراج شد. استفاده از تبدیل موجک امکان جداسازی سیگنال‌های غیرایستان با مؤلفه-های فرکانسی متفاوت را فراهم می‌کند. بخش‌ها با استفاده از تبدیل موجک گسسته دابچیز مرتبه ۲ تجزیه شدند. این نوع تبدیل موجک کلاس‌بندی بهتری را نسبت به دیگر موجک‌های معمول ارائه می‌کند. در شکل (۱) نحوه تجزیه سیگنال با استفاده از تبدیل موجک نشان داده شده است. ویژگی‌های به دست آمده از سیگنال‌های D_5، D_6 و A_6 عبارتند از: عبور از صفر (ZC)، انتگرال قدر مطلق داده[۱۰] (ID)، میانگین، انحراف معیار (SD) و آنتروپی

---

[۵] Class Seperability Feature Selection

[۶] قابل دسترسی در: http://opends.de/

[۷] Karolinska Sleepiness Scale

[۸] Skewness Coefficient

[۹] Kurtosis

[۱۰] Integral of the Absolute Data

زیرباند موجک[11] (SWE). ID و ZC به صورت زیر، طبق فرمول‌های (۱) و (۲) تعریف می‌شود:

$$ID = \sum_{k=1}^{N} |d_i(k)| \qquad (۱)$$

$$ZC = \sum_{k=1}^{N-1} [\mathrm{sgn}(-d_i(k+1)d_i(k))] \qquad (۲)$$

که در آن sgn تابع علامت است و مقدار آن برای $d_i < 0$، صفر و برای $d_i \geq 0$، ۱ است. k شماره نمونه، $d_i(k)$ نمونه‌های مربوط به D_5، D_6 و A_6، و N تعداد کل نمونه‌ها در هر دوره است.

آنتروپی زیر باند موجک برحسب انرژی نسبی ضرایب موجک تعریف شده است. بدین ترتیب که انرژی در هر سطح تجزیه (j=1,…,N) محاسبه می‌شود. نحوه این محاسبه در فرمول‌های (۶-۳) آورده شده است [۱۳].

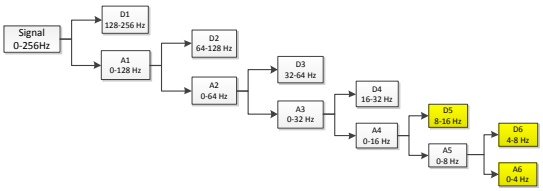
$$E_j = \sum_k |d_j(k)|^2 \qquad (۳)$$

$$E_{total} = \sum_j E_j \qquad (۴)$$

$$P_j = \frac{E_j}{E_{total}} \qquad (۵)$$

و در نهایت مقدار SWE برای هر سطح تجزیه j محاسبه می‌شود:

$$H_j = -P_j \log P_j \qquad (۶)$$



شکل ۱: نحوه انجام تبدیل موجک و انتخاب سطوح مختلف مربوط به تقریب و جزئیات سیگنال (بخش‌های زرد رنگ برای استخراج ویژگی انتخاب شده است).

۴-روش CSFS-SFS برای انتخاب ویژگی

یکی از روش‌های بسیار مرسوم برای انتخاب ویژگی روش SFS است. الگوریتم SFS [۱۴] با یک زیرمجموعه تهی آغاز می‌شود و ویژگی‌ها به صورت تک‌تک به زیرمجموعه مورد نظر اضافه شده و زیرمجموعه جدید ارزیابی می‌شود. انتخاب یک ویژگی مشروط به حداکثر کردن صحت کلاس‌بندی است. در روش به کار رفته در این مقاله حداکثر شدن مجموع فواصل ماهالانوبیس داده‌های یک کلاس از مرکز کلاس دیگر، عنوان تابع ارزیابی، از نوع رَپِر، در نظر گرفته شد.

یکی دیگر از روش‌های مناسب برای انتخاب ویژگی، استفاده از معیار اطلاعات متقابل (MI) است [۱۵]. تعریف MI بین دو متغیر تصادفی X و Y در فرمول (۷) آمده است.

$$I(X;Y) = \iint p(x,y)\log\left(\frac{P(x,y)}{p(x)P(y)}\right) dxdy \qquad (۷)$$

که در آن $p(x,y)$ تابع چگالی احتمال توأم، و $p(x)$ و $p(y)$ توابع چگالی احتمال کناری هستند. MI را می‌توان کاهش عدم قطعیت X در نتیجه دانستن Y توصیف کرد. رابطه بین MI و آنتروپی می‌تواند از تعاریف به شکل فرمول (۸) استخراج شود.

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \qquad (۸)$$

در فرمول (۸)، $H(X,Y)$ آنتروپی توأم X و Y است که به صورت فرمول (۹) تعریف می‌شود.

$$H(X;Y) = -\iint p(x,y)\log(p(x,y))\, dxdy \qquad (۹)$$

فرض کنیم $(x,y) \in (R^m \times Y)$، که در آن R^m یک فضای ویژگی با m بُعد و $Y = \{1,2,\dots,C\}$ مجموعه برچسب‌ها است. نماد n_i نشان‌دهنده تعداد نمونه‌های متعلق به کلاس i-ام، و N تعداد کل نمونه‌ها است. اگر فرض x_{ij} نشان‌دهنده j-امین نمونه در کلاس i-ام باشد، u میانگین نمونه‌ها در تمام کلاس‌ها، و u_i میانگین نمونه‌ها در کلاس i است. فرمول‌های (۱۰) و (۱۱) نحوه محاسبه u و u_i را نشان می‌دهد.

$$u = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{n_i} x_{ij} \qquad (۱۰)$$

$$u_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij} \qquad (۱۱)$$

در ادامه ماتریس پراکندگی بین کلاسی، S_b، و ماتریس پراکندگی درون کلاسی، S_w، طبق فرمول‌های (۱۲) و (۱۳) تعریف می‌شود.

$$S_b = \sum_{i=1}^{C} n_i(u_i - u)(u_i - u)^T \qquad (۱۲)$$

$$S_w = \sum_{i=1}^{C}\sum_{j=1}^{n_i}(u_i - x_{ij})(u_i - x_{ij})^T \qquad (۱۳)$$

برای یک زیرمجموعه از ویژگی‌ها (کانال‌ها)، میزان جدایی‌پذیری کلاسی، براساس ماتریس پراکندگی، با توجه به نسبت مقادیر متناظر ماتریس پراکندگی بین کلاسی به ماتریس پراکندگی درون کلاسی ارزیابی می‌شود. زیرمجموعه‌ای از کانال‌ها که جدایی‌پذیری کلاسی بالایی را فراهم کند، به عنوان زیرمجموعه مطلوب شناخته می‌شود. ایده جدایی‌پذیری کلاسی برای انتخاب کانال‌های بهینه برای کلاس‌بندی قابل استفاده است. ماتریس پراکندگی بین متغیری، S_f، برای جلوگیری از انتخاب ویژگی‌های دارای افزونگی تعریف می‌شود. به کمک این ماتریس میزان افزونگی بین متغیری، مطابق فرمول (۱۴)، ارزیابی می‌شود.

$$S_f = \frac{1}{|S|}\sum_{j=1}^{|S|}\sum_{i=1}^{C} n_i(u_i - u_{ji})(u_i - u_{ji})^T \qquad (۱۴)$$

که در آن $|S|$ اندازه S و نشان‌دهنده تعداد کانال‌های انتخاب شده است. u_{ji} میانگین نمونه‌های کانال j-ام در زیرمجموعه S در کلاس i-ام

[11] Subband Wavelet Entropy

است. S_f قابلیت جدایی‌پذیری کلاسی را بین کانال‌های انتخاب نشده و زیرمجموعه مربوط به کانال‌های انتخاب شده بیان می‌کند. هرچه S_f بزرگ‌تر باشد، میزان افزونگی بین کانال‌ها کمتر است. براساس این معیار، یک روش مؤثر انتخاب زیرمجموعه‌ای از ویژگی‌ها (کانال‌ها)، به نام انتخاب ویژگی مبتنی بر جدایی‌پذیری کلاسی (CSFS) پیشنهاد شده است [۱۶]. در ادامه گام‌های این روش به صورت شبه‌کد توضیح آمده است.

گام ۱- استخراج ویژگی: محاسبه MI برای هر جفت کانال و استخراج ماتریس MI (I) برای سیگنال‌ها با سایز $m \times m$. ماتریس I در فرمول (۱۵) نشان داده شده است.

$$I_j = \begin{bmatrix} I_j(X_1,X_1) & I_j(X_1,X_2) & \vdots & I_j(X_2,X_m) \\ I_j(X_2,X_1) & I_j(X_2,X_2) & \vdots & I_j(X_2,X_m) \\ \vdots & \vdots & \ddots & \vdots \\ I_j(X_m,X_1) & I_j(X_m,X_1) & \dots & I_j(X_m,X_m) \end{bmatrix}, \quad j = 1,2,\dots,N \qquad (۱۵)$$

در فرمول (۱۵)، $X_i\ , i = 1,2,\dots,m$ نشان دهنده i-امین کانال است. بردار ستونی $\beta_{ji} = \left[I_j(X_1,X_i)\ I_j(X_2,X_i) \dots I_j(X_m,X_i)\right]^T$، $i = 1,2,\dots,m$ به عنوان i-امین ویژگی از مشاهده j-ام در نظر گرفته می‌شود. میانگین به دست آمده برای هر ویژگی یک بردار خواهد بود.

گام ۲- انتخاب ویژگی: محاسبه S_{bi} و S_{wi} برای هر ویژگی با استفاده از ماتریس MI به دست آمده در گام ۱. سپس تمام کانال‌ها براساس نسبت بین S_{bi} و S_{wi} مرتب می‌شوند.

$$J_{mi} = \frac{S_{bi}}{S_{wi}}\ , i = 1,2,\dots,m \qquad (۱۶)$$

طبق فرمول (۱۶)، هرچه مقادیر J_{mi} بزرگ‌تر باشد، کانال i-ام از اهمیت بیشتری برای کلاس‌بندی برخوردار خواهد بود.

گام ۳- انتخاب اولین ویژگی جلوسو: انتخاب متغیر با بزرگ‌ترین مقدار برای J_{mi} به عنوان اولین عضو از زیرمجموعه ویژگی‌های بهینه S. انتخاب ویژگی i-ام به عنوان یک ویژگی بهینه در حقیقت منتاظر با انتخاب کانال i-ام به عنوان کانال بهینه خواهد بود.

گام ۴- انتخاب ویژگی جلوسوی دوم به بعد: برای انتخاب کانال بعدی، لازم است تا افزونگی بین کانال‌ها ارزیابی شود. میزان افزونگی S_{fi} را می‌توان با فرمول (۱۷) به صورت کمی معرفی کرد.

$$J_{ri} = \frac{S_{bi}+S_{fi}}{S_{wi}}\ , i = 1,2,\dots,m \qquad (۱۷)$$

هر چه J_{ri} بزرگ‌تر باشد، i-امین کانال از اهمیت بیشتری برای کلاس‌بندی برخوردار است. کانالی که دارای بزرگ‌ترین مقدار J_{ri} است، وارد مجموعه S می‌شود.

گام ۵- معیار توقف: اگر $|S| = K$، الگوریتم متوقف می‌شود، در غیر این صورت، الگوریتم به گام ۴ می‌رود.

در این مقاله روشی جدیدی برای انتخاب ویژگی‌های بهینه معرفی شده است. این روش ترکیبی از دو روش توضیح داده شده یعنی SFS و CSFS می‌باشد، لذا روش جدید CSFS-SFS نام‌گذاری شده است. در روش پیشنهادی ابتدا با استفاده از CSFS یک زیرمجموعه از کانال‌های بهینه انتخاب می‌شود، سپس ویژگی‌های مورد نظر (در مسئله ما ۲۸ ویژگی) از کانال‌های منتخب استخراج خواهد شد. در ادامه با استفاده از روش SFS زیرمجموعه‌ای از بهترین ویژگی‌ها انتخاب می‌شود. به این ترتیب بهترین ویژگی‌ها از میان بهترین کانال‌ها به دست می‌آید.

۵-کلاس‌بند

برای کلاس‌بندی از کلاس‌بند با تابع تفکیک‌کننده خطی استفاده شده است. در این روش فرض می‌شود داده‌های هر دو کلاس از توزیع گوسی پیروی می‌کنند، همچنین ماتریس کوواریانس کلاس‌ها برابر و میانگین آن‌ها متفاوت در نظر گرفته می‌شود. در نتیجه میانگین داده‌های هر کلاس به عنوان پارامتر میانگین در نظر گرفته شده و ماتریس کوواریانس، پس از کم کردن میانگین داده‌ها از مشاهدات هر کلاس، برای هر دو کلاس محاسبه می‌شود. برای داده‌های تست، کلاس‌بندی به گونه‌ای صورت می‌گیرد که تابع هزینه کلاس‌بند مینیمم شود:

$$\hat{y} = \underset{y\ =\ 1,\dots,K}{argmin} \sum_{k=1}^{K} \hat{P}(k|x)C(y|k) \qquad (۱۸)$$

در فرمول (۱۸)، $\hat{y}$ نتایج کلاس‌بند، K تعداد کلاس‌ها، $\hat{P}(k|x)$ احتمال پسین مربوط به کلاس k به شرط مشاهده x است. $C(y|k)$ هزینه کلاس‌بندی یک مشاهده به عنوان کلاس y است در حالی که آن مشاهده در حقیقت متعلق به کلاس k بوده است. هزینه کلاس‌بندی صحیح صفر و هزینه کلاس‌بندی غلط ۱ است. برای محاسبه احتمال پسین، احتمال پیشین برای هر کلاس به صورت حاصل تقسیم تعداد نمونه‌های آموزش آن کلاس بر تعداد کل نمونه‌های آموزش، تخمین زده می‌شود. برای آموزش و تست کلاس‌بند، به ترتیب، از ۷۵ و ۲۵ درصد داده‌های ثبت شده استفاده شد.

۶-نتایج

از بخش‌های سیگنال ثبت شده با طول ۳ ثانیه برای کمی‌سازی آنالیز الگوریتم مورد نظر استفاده شد. الگوریتم توسط نرم‌افزار Matlab 2013 پیاده‌سازی شده و سیستمی با این مشخصات مورد استفاده قرار گرفته است: CPU Intel core i5, 2.53GHz و 4GB Ram.

برای گزارش نتایج، ابتدا با روش SFS از میان ۷۵۶ ویژگی استخراج شده، از میان همه ۲۷ کانال، زیرمجموعه‌ای از ویژگی‌ها انتخاب شده و نتایج حاصل از کلاس‌بندی به دست آمد. در ادامه با استفاده از روش CSFS-SFS ابتدا یک زیرمجموعه ۱۳ عضوی از کانال‌های بهینه انتخاب شد. سپس از میان کانال‌های منتخب، ۲۸ ویژگی معرفی شده استخراج شد. در ادامه از میان ۳۶۴ ویژگی یک زیرمجموعه بهینه از ویژگی‌ها

انتخاب شد. بعد از بررسی تعداد کانال‌های مختلف انتخاب شده در روش CSFS-SFS، بهترین نتایج با انتخاب ۱۳ کانال به دست آمد، لذا نتایج مربوط به این تعداد کانال منتخب گزارش شده است.

در شکل ۲ نتایج مربوط به صحت کلاس‌بند با تابع جدایی خطی آمده است. نمودار سمت راست نتایج مربوط به صحت به دست آمده به روش SFS و نمودار سمت چپ نتایج مربوط به صحت به دست آمده به روش CSFS-SFS را نشان می‌دهد. همان‌طور که در نمودارها مشخص است، در هر دو روش، بهترین نتایج به ازای یک زیرمجموعه ۱۰۰ عضوی از ویژگی‌ها به دست می‌آید. به ازای این تعداد ویژگی، میانه برای روش SFS و CSFS-SFS، به ترتیب ۹۳/۹۰ % و ۹۲/۵۴ %، تعداد نقاط پَرت، در هر دو روش، ۳ نقطه به دست آمد. همچنین مقدار انحراف معیار استاندار برای دو روش، به ترتیب، ۳/۸۳ و ۳/۹۵ محاسبه شد. با توجه به مقادیر گزارش شده، در روش پیشنهادی میزان صحت در حدود ۱/۴۶ % کاهش یافته، همچنین تعداد نقاط پرت بدون تغییر و انحراف معیار استاندارد تقریباً بدون تغییر است. مقادیر مربوط به ۲۰ بار آموزش و تست کلاس‌بند است.

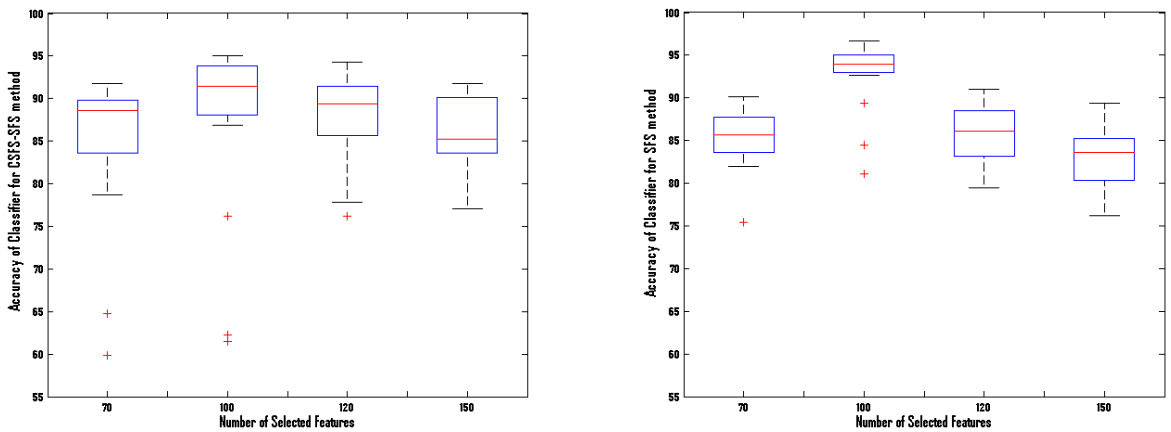
برای مقایسه بهتر نتایج به دست آمده از دو روش انتخاب ویژگی باید توجه داشت که یکی از پارامترهای بسیار مهم و ضروری زمان مورد نیاز برای انجام محاسبات است. زمان مورد نیاز برای انتخاب ۱۰۰ ویژگی از میان ۷۵۶ ویژگی با استفاده از روش SFS درحدود ۴۸۲۵ ثانیه است. این در حالی است که زمان مورد نیاز برای انتخاب ویژگی در روش CSFS-SFS در حدود ۱۳۴۱ ثانیه است. ملاحظه می‌شود که در روش پیشنهادی برای انتخاب ویژگی، زمان مورد نیاز نسبت به روش مرسوم SFS حدود ۷۳% کاهش یافته است. این کاهش قابل ملاحظه مؤید مزیت روش پیشنهادی خواهد بود، چراکه یکی از اهداف روش‌های انتخاب ویژگی کاهش زمان انجام محاسبات است. زمان محاسبات در پردازنده‌ها، به طور معمول، یک پارامتر متناسب با توان تلفاتی است. روش پیشنهادی با کاهش زمان انجام محاسبات میزان تلفات توان و مصرف انرژی در پردازنده‌ها را نیز کاهش می‌دهد. به این ترتیب می‌توان ادعا کرد که در سیستم‌های عملی روش پیشنهادی عملکرد بهتری خواهد داشت.

۷- بحث

در این تحقیق، آنالیزی از ریتم‌ها و الگوهای موجود در سیگنال‌های ثبت شده EEG، EOG و DQ در حوزه زمان، طیف و موجک ارائه شده است. پروتکل و فرآیند ثبت داده‌ها در این مطالعه به گونه‌ای طراحی شد که در عین سادگی، شرایط واقعی رانندگی را به خوبی شبیه‌سازی کند. نرم‌افزار به کار گرفته شده به عنوان شبیه‌ساز رانندگی بسیار قدرتمند و منطبق با شرایط واقعی است. مسیر انتخاب شده برای رانندگی یک مسیر ساده، یکنواخت و تکراری است تا بتواند شرایط واقعی خواب‌آلودگی را به خوبی القا کند. مزیت دیگر سیستم، نسبت به سایر سیستم‌های شبیه‌ساز رانندگی، کم هزینه بودن فرآیند ثبت است، به گونه‌ای که تجهیزات مربوط به سیستم شبیه‌ساز راه‌اندازی شده ارزان قیمت و در دسترس هستند.

برای دست‌یابی به بهترین ویژگی‌های استخراج شده، روش ترکیبی جدیدی با نام CSFS-SFS پیشنهاد شده است. با توجه به افزایش سرعت در این روش، تقریباً بدون تغییر ماندن نتایج و نیز کاهش زمان محاسبات به میزان ۷۳%، می‌توان روش پیشنهادی را به عنوان یک روش بهتر و با کاربرد گسترده‌تر، نسبت به روش مرسوم SFS، معرفی نمود. از آنجا که در سیستم‌های عملی حجم محاسبات و به دنبال آن میزان اشغال پردازنده و تلفات توان بسیار ارزشمند و در انتخاب روش انجام محاسبات تعیین کننده است، لذا پیاده‌سازی روش پیشنهادی و استفاده از آن در سیستم‌های کاربردی راحت‌تر خواهد بود.

با در نظر گرفتن مفاهیم به کار گرفته شده در روش CSFS-SFS مشخص می‌شود که در این روش روابط خطی و غیر خطی بین ویژگی‌ها، همچنین میزان افزونگی بین کانال‌ها در بخش CSFS در نظر گرفته می‌شود. در مرحله انتخاب بهترین کانال‌ها، از میان ۱۳ کانال انتخاب شده، ۵ کانال متعلق به سیگنال‌های EEG، یک کانال EOG و ۷ کانال به DQ تعلق دارد. در مطالعات قبلی اغلب از یک نوع سیگنال برای تشخیص خواب‌آلودگی حین رانندگی استفاده شده است. لذا استفاده از ترکیب این سه نوع



شکل ۲: نتایج به دست آمده برای صحت کلاس‌بند با انتخاب ویژگی به روش SFS (نمودار سمت راست) و CSFS-SFS (نمودار سمت چپ)

سیگنال یکی از مزیت‌های این مطالعه نسبت به سایر مطالعات مشابه بوده و نتایج نشان می‌دهد هر سه نوع سیگنال حاوی اطلاعات مفیدی برای تفکیک الگوهای AS و DS هستند.

انتخاب تعداد محدودی از کانال‌ها در گام نخست از روش پیشنهادی در این مقاله یک مزیت دیگر نیز خواهد داشت. در سیستم‌های عملی و کاربردی بسیار مهم است که سیستم با کمترین تعداد کانال‌های ثبت داده بتواند به تشخیص صحیح برسد. به این ترتیب باید کانال‌های مهم و دارای بیشترین اثر تشخیصی شناسایی شده و ثبت داده تنها توسط آن‌ها صورت گیرد. در روش پیشنهادی ابتدا تعداد کانال‌ها کاهش یافته و کانال‌های بهینه شناسایی می‌شود که این موضوع به لحاظ کاربردی یک مزیت خواهد بود، چراکه به این ترتیب می‌توان تعداد سنسورهای مورد نیاز برای ساخت تجهیز را کاهش داده و علاوه بر صرفه جویی در زمان محاسبات، هزینه‌های ساخت را نیز کاهش داد.

در روش پیشنهادی CSFS-SFS سعی شده تا با ترکیب تابع ارزیابی از نوع رَپر و فیلتر از مزیت‌های هر دو روش بهره برده و در عین حال اثر معایب آن‌ها به حداقل برسد. ویژگی‌های منتخب توسط فیلترها عمومیت بیشتری داشته و سرعت انتخاب آن‌ها بالاتر است. اما در مقابل، فیلترها تمایل به انتخاب زیرمجموعه‌های بزرگتر (در حدود زیرمجموعه کامل) دارند. همچنین رَپِرها نسبت به فیلترها میزان صحت بالاتری داشته و مکانیسم آن‌ها به گونه‌ای است که دچار تطبیق بیش از حد نمی‌شوند. اما رَپِرها سرعت کمی در انتخاب ویژگی داشته و ویژگی‌های منتخب آن‌ها عمومیت پایین‌تری دارند. ما در روش پیشنهادی سعی کردیم تا در ابتدا با استفاده از روش انتخاب ویژگی CSFS که تابع ارزیابی آن از نوع فیلتر است، با سرعت بالا یک زیرمجموعه محدود از بهترین کانال‌ها را انتخاب کرده و در گام بعد با استفاده از یک روش انتخاب ویژگی با تابع ارزیابی از نوع رَپِر، بهترین زیرمجموعه از ویژگی‌ها که منجر به بالاترین صحت می‌شوند، انتخاب شود. به این ترتیب با محدود کردن تعداد کانال‌ها در گام نخست، سعی شده تا اثر معایب روش‌های مبتنی بر رَپِر و فیلتر به حداقل برسد.

در روش پیشنهادی CSFS-SFS، نسبت به زمانی که تنها از روش SFS استفاده می‌شود، ممکن است برخی از ویژگی‌های بهینه از دست برود، چراکه در روش SFS ویژگی‌ها در تک تک کانال‌ها مورد ارزیابی قرار می‌گیرد ولی در روش پیشنهادی در این مقاله SFS فقط به ویژگی‌ها در کانال‌های منتخب دسترسی دارد. براساس نتایج به دست آمده، مجموعه‌ای با ۱۰۰ ویژگی انتخاب شده بیشترین صحت (۹۲/۵۴٪) و کمترین میزان انحراف معیار (۳/۹۵٪) را نتیجه می‌دهد.

در مجموع، با توجه به مزیت‌های پروتکل و فرآیند تعریف شده برای ثبت داده‌ها نسبت به سایر مطالعات، مانند انطباق با واقعیت، سادگی، کم هزینه بودن و در دسترس بودن، و همچنین با توجه به مزیت‌های روش پردازشی پیشنهاد شده، از جمله سرعت بالا و کاهش چشم‌گیر زمان انجام محاسبات، می‌توان نتایج به دست آمده را قابل اتکا فرض کرده و به عنوان مبنایی برای مطالعات بعدی در نظر گرفت. قطعاً استفاده از روش‌های دیگری برای استخراج ویژگی و کلاس‌بندی داده‌ها می‌تواند ما را در رسیدن به یک سیستم واقعی تشخیص خواب‌آلودگی حین رانندگی یاری کند.

مراجع

[1] Word Health Organization, Global health observatory data repository, 2013. http://apps.who.int/gho/data/node.main.A997?lang=en

[2] T.H. Chang, C.S. Hsu, C. Wang and L.K. Yang, "Onboard measurement and warning module for irregular vehicle behavior" IEEE Transactions on Intelligent Trans-portation Systems, 9(3):501–13, 2008.

[3] D. Sandberg, T. Åkerstedt, A. Anund, G, Kecklund and M. Wahde, "Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators" IEEE Transactions on Intelligent Transportation Systems, 12(1): 97-108, 2011.

[4] P. Smith, M. Shah and N. Da-Vitoria-Lobo, "Determining driver visual attention with one camera" IEEE Transactions on Intelligent Transportation Systems, 4(4):205–18, 2003.

[5] H.J. Dikkers, M. Spaans, D. Dactu, M. Novak and L.J.M. Rothkrantz, "Facial recognition system for driver vigilance monitoring" Proceeding IEEE SMC, 4:3787–92, 2004.

[6] S. N. Miri Ashtiani, Z. Mardi and M. Mikaili, "Spectral Characteristics Assessment in Recognition of Drivers' Drowsiness Using Statistical Tests", 18th Iranian conference on Biomedical Engineering, 2011.

[7] Z. Mardi, S. N. Miri Ashtiani and M. Mikaili, "EEG-Based Drowsiness Detection for Safe Driving Using Chaotic Features and Statistical Tests" Journal of Medical Signals & Sensors,Vol1, 2011.

[8] M. Tasaki, M. Sakai, M. Watanabe, H. Wang and D. Wei, "Evaluation of drowsiness during driving using electrocardiogram: a driving simulation study" IEEE Inter-national 351 Conference on CIT, 1480-1485, 2010.

[9] M. Fabbri, F. Provini, E. Magosso, A. Zaniboni, A. Bisulli, G. Plazzi and M. Ursino, "Detection of sleep onset by analysis of slow eye movements: a preliminary study of MSLT recordings" Sleep Medicine,10:637–40, 2009.

[10] A. Crespel, P. Gélisse, M. Bureau and P. Genton, In: J. Libbey, editor, "Atlas of Electroen-cephalography" Paris: Eurotext, 2005.

[11] J. Krajewski, D. Sommer, U. Trutschel, D. Edwards and M. Golz, "Steering Wheel Behavior Based Estimation Of Fatigue", Proc. Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp 118-124, 2009.

[12] T. Akerstedt, and M. Gillberg, "Subjective and objective sleepiness in the active individual" International Journal of Neuroscience, 52 29-37, 1990.

[13] H. A. Al-Nashash, J. S. Paul and N. V.Thakor, "Wavelet Entropy Method for EEG Analysis: Application to Global Brain Injury" 1st International IEEE EMBS Conference on Neural Engineering, pp. 348-351, 2003.

[14] J. Reunanen "Overfitting in making comparisons between variable selection methods" Journal of Machine Learning Research, 3:1371–82, 2003.

[15] G. Doquire and M. Verleysen, "Feature selection with missing data using mutual information estimators" Neurocomputing 90, 3–11, 2012.

[16] M. Han and X. Liu, "Feature selection techniques with class separa bility for multivariate time series" Journal of Neurocomputing 110 29–34, 2013.