# تشخیص حواس‌پرتی و خواب‌آلودگی راننده از طریق روش‌های مبتنی بر پردازش تصویر و یادگیری عمیق

**مریم هاشمی[۱]، علیرضا میررشید[۲]، سید علی‌اصغر بهشتی شیرازی[۳*]**

۱- دانشجوی کارشناسی ارشد، دانشگاه علم و صنعت ایران- تهران- ایران
maryamhashemi1995@gmail.com

۲- دانشجوی کارشناسی ارشد، دانشگاه علم و صنعت ایران -تهران- ایران
alireza.mirrashid@yahoo.com

۳-دانشیار - دانشکده مهندسی برق- دانشگاه علم و صنعت ایران- تهران- ایران
abeheshti@iust.ac.ir

**چکیده: این تحقیق به بررسی یک رویکرد جدید جهت تشخیص حواس‌پرتی و خواب‌آلودگی راننده جهت هوشمندسازی رانندگی پرداخته‌است. به دلیل عدم وجود یک مجموعه داده دقیق و جامع در حوزه مجموعه داده‌های چشم، نویسندگان یک مجموعه داده نوین جمع‌آوری کرده‌اند، همچنین شبکه عصبی مصنوعی‌ای در جهت تشخیص خواب‌آلودگی راننده به گونه‌ای طراحی شده که دو هدف مهم پردازش‌های بلادرنگ، از جمله دقت بالا و سرعت بالا، همزمان در نظر گرفته شوند. اهداف این مقاله به شرح زیر است: تخمین موقعیت سر راننده جهت تشخیص حواس‌پرتی، معرفی یک مجموعه داده جامع جدید برای تشخیص بسته‌بودن چشم، و همچنین، طراحی سه شبکه عصبی مصنوعی که یکی از آن‌ها یک شبکه عصبی کاملاً طراحی شده (FDNN) است و دو شبکه‌ی دیگر از تکنیک انتقال یادگیری از طریق شبکه‌های VGG16 و VGG19با لایه‌های اضافی استفاده می‌کنند (TLVGG). نتایج نشان می‌دهد دقت شبکه‌های پیشنهادی بالا و پیچیدگی محاسباتی کم است، به طوری که روش پیشنهادی نسبت به کارهای قبلی ۴ برابر سریع‌تر و دارای صحت ۹۸.۱۵٪ است.**

**واژه‌های کلیدی:** شبکه‌ی عصبی عمیق، شبکه‌ی عصبی کانولوشنی، انتقال یادگیری، حواس پرتی، خواب‌آلودگی، رانندگی هوشمند.

تاریخ ارسال مقاله: ۹۹/۰۲/۲۶
تاریخ پذیرش مقاله: ۹۹/۰۴/۰۹
نام نویسنده مسئول: سید علی‌اصغر بهشتی شیرازی*

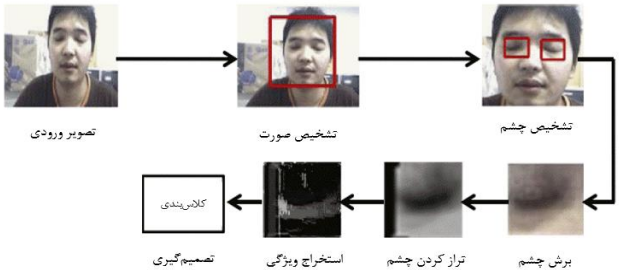
## ۱-مقدمه

با توجه به گزارش‌های منتشر شده از سازمان بهداشت جهانی (WHO)، حوادث جاده‌ای یکی از ۱۰ مورد برتر است که منجر به مرگ در جهان می‌شود. [۱] گزارش‌ها نشان می‌دهد که اولین عامل بروز چنین تصادفاتی رانندگان هستند. بنابراین، تشخیص حواس‌پرتی راننده و خواب‌آلودگی می‌تواند یک روش مناسب برای جلوگیری از تصادفات باشد. همچنین این روش موجب بهبود عملکرد سیستم‌های پیشرفته کمک راننده (ADAS) و سیستم نظارت بر رانندگان (DMS) می‌شود و در نهایت باعث افزایش ایمنی جاده‌ها است.

پن و همکاران [۲] یک روش برای تشخیص چهره‌ی بلادرنگ، با شناختن خودبه‌خودی پلک‌زدن معرفی کرده‌اند. این روش نیازی به سخت افزار اضافی به جز یک وب‌کم عمومی ندارد که یکی از ویژگی‌های بررسی راننده از طریق روش‌های پردازش تصویر است. در این مقاله یک مجموعه داده عمومی در دسترس به نام مجموعه داده ویدئوی پلک‌زن معرفی شده‌است [۳]. این مجموعه داده توسط Logitech pro5000 که یک وب‌کم عمومی است، جمع‌آوری شده‌است. در مجموع ۸۰ ویدئوکلیپ در مجموعه داده ویدیوئی پلک‌زن از ۲۰ نفر، چهار کلیپ برای هر فرد، وجود دارد: یک کلیپ برای مشاهده جلو بدون عینک، یک کلیپ با نمای جلو و پوشیدن عینک قاب نازک، یک کلیپ برای نمای جلو و عینک قاب سیاه و آخرین کلیپ با نمای رو به بالا و بدون عینک. هر فرد ملزم به انجام پلک‌زدن با سرعت عادی با چهار تنظیمات فوق است. یک کلیپ ویدیویی با ۳۰ فریم در ثانیه و اندازه ۳۲۰×۲۴۰ برای هر پیکربندی ضبط می‌شود و حدود ۵ ثانیه طول می‌کشد. تعداد پلک‌زدن در یک کلیپ ویدیوئی از ۱ تا ۶ دفعه متغیر است. در کل ۲۵۵ پلک‌زن در مجموعه داده وجود دارد. همه داده‌ها بدون هیچ‌گونه شرایط خاص در نورپردازی در داخل خانه جمع‌آوری می‌شوند.

در مقاله‌ی [۴] نویسندگان به معرفی رویکردی پرداخته‌اند که بسته یا باز بودن چشم‌ها را تشخیص می‌دهد. این رویکرد ترکیبی از ویژگی‌های قدرتمندی است که برای توصیف اطلاعات قسمت‌های چشم و ساخت مدل حالت چشم لازم است. برای بهبود بیشتر استحکام مدل در برابر نویز تصویر، توصیف‌کننده ویژگی جدیدی به نام هیستوگرام‌های چند مقیاس از گرادیان های اصلی جهت دار پیشنهاد شده‌است. روش مقاله در مجموعه داده‌های چشم در دنیای واقعی از جمله مجموعه داده ZJU آزمایش شده‌است.

اندازه و نوع قلم‌های پارسی مورد استفاده برای هر یک از معماری کلی طرح معرفی شده در شکل ۱ نشان داده شده‌است. روش پیشنهادی به اینصورت است که ابتدا تصویر ورودی دریافت می شود و ناحیه مربوط به صورت شناسایی می‌شود. این کار با استفاده از روش‌های مبتنی بر آشکارسازها صورت گرفته است [۵]. سپس مدل ساختاری تصویری تبعیض‌آمیز برای پیدا کردن نواحی مربوط به چشم و برش آن استفاده می‌شود [۶]. بعد از آن مرحله‌ی اصلاح و هم‌تراز کردن چشم است که با استفاده از روش پیشنهادی هوانگ و همکاران [۷] که با نام عادی‌سازی هندسی شناخته می‌شود، انجام می‌گیرد.



شکل ۱ معماری کلی تشخیص باز یا بسته بودن چشم در مقاله‌ی [۴]

تشخیص وضعیت چشم به طور گسترده در بسیاری از زمینه‌ها، از جمله تشخیص خواب‌آلودگی راننده، طبقه بندی بیان صورت و فناوری رابط انسان و کامپیوتر استفاده می‌شود. تحقیق [۸] یک چارچوب جدید را بر اساس روش یادگیری عمیق برای طبقه بندی حالات چشم در تصاویر ارائه می‌دهد. روش پیشنهادی ترکیبی از یک شبکه‌ی عصبی عمیق و یک شبکه‌ی عصبی کانولوشنی عمیق برای ایجاد یک شبکه‌ی عصبی عمیق یکپارچه برای توصیف اطلاعات مفید در منطقه چشم با استفاده از روش بهینه‌سازی مشترک است. از انتقال یادگیری برای استخراج ویژگی‌های موثر چشم استفاده شده‌است. علاوه بر این، تأثیر روش‌های انتقال یادگیری با مجموعه داده‌های مختلف بررسی شده است. یک مجموعه داده تشخیص خواب‌آلودگی راننده توسط نویسندگان مقاله ساخته‌شده و در یک آزمایش برای ارزیابی اثربخشی روش پیشنهادی در محیط‌های رانندگی در مقایسه با دیگر مجموعه داده‌های موجود مورد استفاده قرار می گیرد.

در این مقاله ابتدا یک شبکه‌ی عصبی عمیق پیشنهادی آموزش داده می‌شود. سپس در عمل ناحیه صورت و چشم استخراج می شود و وارد شبکه عصبی از پیش آموزش داده‌شده می‌شود.

یکی دیگر از مباحثی که در رانندگی مورد توجه قرار می‌گیرد تشخیص حواس‌پرتی راننده است. برای تشخیص حواس‌پرتی تحقیق [۹] مدلی را با استفاده از یک فیلم پایگاه داده جمع‌آوری

شده ارائه داده است. هر فیلم، فریم به فریم پردازش می‌شود تا ویژگی‌های لازم برای تشخیص نشانه‌های حواس‌پرتی را استخراج کند. از طبقه‌بندی دودویی نیز برای ارزیابی اینکه راننده حواس پرت است یا خیر استفاده شده‌است. یعنی کلاس‌بندی تنها به دو طبقه‌ی حواس‌پرت یا هشیار تقسیم می‌شود و حالت بینابینی در نظر گرفته نشده‌است. روش اعتبارسنجی kfold برای تعیین قدرت پیش‌بینی مدل در نظر گرفته شده‌است.

در این مقاله گزارش شده که میانگین آماری مبتنی بر ماتریس سردرگمی برای معیار صحت ۹۱٪ است.

در تحقیق [۱۰] نویسنده سعی در تشخیص هفت کار مشترک که توسط رانندگان معمولا در زمان رانندگی صورت می‌گیرد، دارد. رانندگی عادی، چک کردن آینه‌های چپ و راست، پاسخ دادن به تلفن همراه، ارسال پیامک با استفاده از تلفن همراه با یک یا هر دو دست و تنظیم دستگاه‌های ویدیویی مانند ضبط صوت در ماشین. در این مقاله از یک دوربین کینکت که شامل اطلاعات رنگی و عمق تصویر از راننده داخل وسیله نقلیه است، استفاده شده‌است. آن‌ها ۴۲ ویژگی ارائه شده توسط سنسور کینکت را جمع‌آوری و ارزیابی کردند و اهمیت ویژگی را با استفاده از ترکیب روش MIC و RF پیش‌بینی کردند و برخی از آنها به عنوان ویژگی موثر انتخاب شده است. در این کار از یک شبکه عصبی روبه جلو (FFNN) به عنوان شبکه یادگیری استفاده می‌شود. همانطور که از نام این شبکه پیداست، از هیچ گونه فیدبکی در داخل شبکه استفاده نشده‌است. این نوع شبکه‌ها معمولا برای حل مسایل غیرخطی کاربرد دارند. تعداد نورون‌ها در هر لایه‌ی پنهان بین ۱۰ الی ۱۰۰ عدد متغیر است. در نهایت برای طبقه بندی با شبکه FFNN دقت ۸۰.۷٪ گزارش شده است.

ماسوز و همکاران در [۱۱] مجموعه داده‌ی جدیدی برای تشخیص خواب‌آلودگی راننده با استفاده از سنسورهای فیزیولوژیکی و تصویری با نام (DROZY) معرفی کرده‌اند. این مجموعه داده مجموعه‌ی ویدیویی از افراد مختلف است، هر شخص در سه حالت فیلم‌برداری شده، حالت نرمال، حالت ۲۰ ساعت بیداری و حالت ۲۸ ساعت بیداری. در هرحالت برای اطمینان بیشتر از وضعیت راننده سیگنال‌های مغزی راننده هم کنترل می‌شود. این فیلم‌برداری در محیط آزمایشگاهی انجام گرفته. گارسا و همکاران [۱۲] از این مجموعه داده و تکنیک‌های بینایی ماشین برای برش دادن چهره از هر فریم و طبقه‌بندی آن استفاده کرده‌اند. در دو کلاس: استراحت یا محروم از خواب تعریف شده و در یک دستگاه اندرویدی کم هزینه اجرا شده‌است که این خود نشان دهنده‌ی قابلیت استفاده‌ی بالای روش‌های مبتنی بر بینایی ماشین است.

مجموعه داده دیگری برای تشخیص خستگی توسط ابطحی و همکاران ایجاد شده‌است [۱۳]. برای طراحی الگوریتم‌های تشخیص خمیازه، دو نوع ویدئو از هر راننده وجود دارد، در یک ویدئو دوربین زیر آینه جلو راننده نصب شده و در ویدئو دوم دوربین روی داشبورد ماشین قرار داده‌شده. سعی شده در این مجموعه داده به تفاوت‌های ظاهری توجه شود و رانندگان با ویژگی‌های مختلفی مانند با و بدون عینک و عینک آفتابی یا نژادهای مختلف انتخاب شوند. آن‌ها همچنین عدد ۶۰٪ را برای تشخیص خمیازه در روش پیشنهادی خود گزارش داده‌اند [۱۳]. البته باید توجه داشت که تقلید یک عمل باعث می‌شود اعتبار یک مجموعه داده تا حدی کاهش یابد چون بهترین داده برای آموزش داده‌ای است که در محیط واقعی و شرایط واقعی جمع آوری شده باشد. هم‌چنین محل قرار دادن دوربین یکی از متغیرهای مهمی است که روی زاویه دید ما از راننده تاثیر می گذارد و تاثیر دادن آن به عنوان یکی از پارمترها، نتیجه‌ی نهایی را بسیار به حالت واقعی نزدیکتر می‌کند.

این مقاله با هدف ارائه یک سیستم هشداردهنده خواب آلودگی راننده، ارائه شده‌است. در این سیستم، اگر چشم راننده به عنوان وضعیت بسته برای فریم‌های پی‌درپی کلاس‌بندی شود، نشانه خواب‌آلودگی است و یک هشدار زود هنگام برای جلوگیری از تصادف به راننده ارسال می‌شود.

کار پیشنهادی این مقاله شامل پنج قسمت است.۱) الگوریتمی برای تخمین میزان چرخش سر راننده معرفی می‌شود که از طریق آن می‌توان حواس‌پرتی را تشخیص داد. ۲) فریم‌های ضبط شده از دوربین برای شناسایی ناحیه چشم‌ها به واحد پیش پردازش وارد می شوند. سپس این واحد عملکرد، تصویر ناحیه چشم را خاکستری می‌کند و سپس هیستوگرام تصویر چشم نرمالیزه می‌شود. نویسندگان از تصاویر با وضوح پایین برای تشخیص آنلاین و از سیستم نرمالیزه کردن هیستوگرام برای غلبه بر وضعیت بد روشنایی استفاده کردند. ۳) نویسندگان سه شبکه عصبی را در رابطه با پارامتر سرعت، دقت بالا و مجموعه داده‌های کوچک پیشنهاد دادند. ۴) یک الگوریتم پیشنهادی برای ارزیابی نتیجه شناسایی شبکه وجود دارد. اگر شبکه تصویر ورودی را به عنوان یک چشم بسته تشخیص دهد، سیستم یک عدد به شمارنده اضافه می کند و اگر تعداد شمارنده به بیش از ۱۲ فریم پی‌درپی برسد، زنگ خطری برای راننده ارسال می‌شود. اگر شبکه تصویر ورودی را به عنوان چشم بسته تشخصیص دهد اما شمارنده عدد کمتر از ۱۲ را نشان دهد، شمارنده را برای فریم بعدی نگه داشته می‌شود، هر زمان که یک تصویر به عنوان چشم باز طبقه‌بندی شود، شمارنده به صفر برمی‌گردد و دوباره شروع به کار می‌کند. به عبارت دیگر، وظیفه این شمارنده دنبال کردن

فریم‌های پی‌درپی برای تمایز پلک‌زدن از خوابیدن است.
۵) جمع‌آوری یک مجموعه داده جدید که موقعیت جدیدی از چشم را در نظر می‌گیرد، به نام نمای مورب از راننده.

برای طبقه‌بندی چشم به کلاس‌های بسته و باز، سه شبکه عصبی مصنوعی در نظر گرفته‌شده‌است. شبکه اول شبکه عصبی کاملاً طراحی شده، شبکه‌ی عصبی دوم، شبکه عصبی عمیق با یادگیری انتقال است و از یک شبکه VGG16 از قبل آموزش دیده استفاده می‌کند که ویژگی‌های سطح پایین ازمجموعه داده‌های ImageNet استخراج می‌شود و ویژگی‌های سطح بالا را از مجموعه داده‌ی موجود یاد می‌گیرد. شبکه سوم مشابه شبکه دوم است، اما از VGG19 استفاده می‌کند. نتایج نشان می‌دهد دقت بالا و زمان محاسبات کوتاه در روش پیشنهادی این مقاله است.

ادامه‌ی این مقاله به شرح زیر است: در بخش ۲ به معرفی سیستم تشخیص حواس‌پرتی می‌پردازیم. بخش ۳ سیستم پیشنهادی برای پیدا کردن ناحیه مورد علاقه را شرح می‌دهد. در بخش ۴ پایگاه‌های داده‌ی موجود را بررسی می‌کنیم. بخش ۵ ساختار شبکه‌های پیشنهادی بررسی می‌شود. بخش ۶ الگوریتم سیستم برای تصمیم‌گیری را معرفی می‌کند. بخش ۷ آزمایش هایی را برای اثبات اثربخشی و استحکام روش و مجموعه داده ارائه می‌دهد. نتیجه‌گیری در مورد سیستم پیشنهادی در بخش ۸ قرار دارد.

## ۲- بررسی روش پیشنهادی تشخیص حواس‌پرتی

هر مقاله باید شامل این بخش‌های اصلی باشد: چکیده، کلمات کلیدی، تشخیص حالت سر کاربردهای گوناگونی دارد، از نظارت بر رفتار راننده تا تشخیص حالت صورت [۱۴] و تشخیص جهت توجه انسان [۱۵] . روش‌های گوناگونی در سال‌های اخیر برای تشخیص حالت یا به عبارت دیگر جهت سر راننده مورد استفاده قرار گرفته که هرکدام مزایا و معایب خاص خود را دارد و متخصص با توجه به کاربردی که مد نظر دارد باید یک روش را انتخاب کند.

ما در این پژوهش از نقاط لندمارک استفاده می‌کنیم. دلیل انتخاب روش مبتنی بر نقاط لندمارک سرعت زیاد این روش است، ضمن اینکه محدوده‌ی زاویه دید افراد در شرایط مختلف جوی و قوای بینایی بسیار متغیر است و تعیین یک زاویه خاص به عنوان آستانه تقریبا غیرممکن است. پس باید در این مرحله ما به دنبال تعیین حدودی محدوده باشیم و آموزش این محدوده توسط شبکه‌های عصبی کاری به مراتب دشوارتر است و به وسیله‌ی نقاط لندمارک راحت‌تر می‌توان یک آستانه‌ی متغیر انتخاب کرد. ضمن اینکه کتابخانه‌های مربوط به تشخیص نقاط لندمارک در سال‌های اخیر پیشرفت چشم‌گیری کرده‌اند و از دقت خوبی برخوردار شدند.

در اولین گام روش پیشنهادی موقعیت سر راننده را تخمین می زنیم. جهت حرکت راننده با حرکت دادن موضوع (در اینجا سر) نسبت به دوربین می‌تواند تغییر کند. معمولا یک نقطه به عنوان مرجع در نظر گرفته می‌شود و سپس به دنبال تغییرات نسبت به مرجع هستیم. هدف اصلی تخمین حالت سر پیدا کردن جهت سر است هنگامی که مکان‌های n نقاط سه بعدی روی جسم، n یک عدد صحیح، در تصویر شناخته شده است و دوربین کالیبره شده نیز وجود دارد. به عبارت دیگر از یک تصویر دوبعدی ضبط شده توسط دوربین پیدا کردن نقاط سه بعدی و تعیین جهت آن غیر ممکن است، اما خوشبختانه نیازی به این کار نیست و ما فقط اختلاف و میزان تغییرات را دنبال می‌کنیم. پس می‌توان یک نقطه‌ی سه بعدی فرضی را مرجع در نظر گرفت و سپس از روی مختصات دوبعدی تغییرات این نقطه‌ی سه بعدی را دنبال کرد. می‌توان گفت انتخاب این نقطه‌ی سه بعدی مرجع آزاد است و آنچه که به ما در تصمیم‌گیری کمک می‌کند میزان فاصله نقطه‌ی فعلی (وابسته به مرجع) و نقطه‌ی مرجع است.

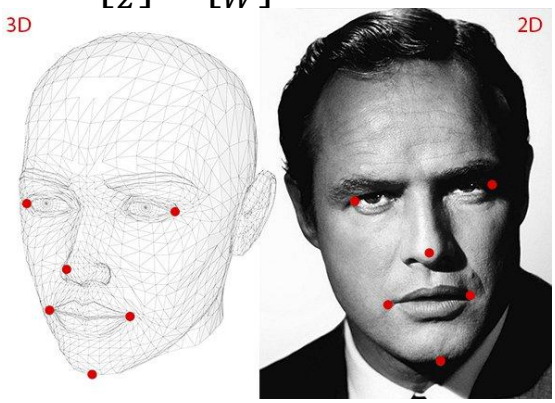
ما در اینجا از یک مرجع استاندارد به نام سیستم مختصات جهانی (WCS) استفاده‌کرده‌ایم. برای محاسبات از تصاویر دوبعدی دوربین استفاده می‌شود و مختصات معادل سه بعدی مشخص می‌شود [۱۶].

WCS از مکان‌های نوک بینی، چانه، گوشه سمت چپ چشم چپ، گوشه سمت راست چشم راست، گوشه سمت چپ دهان و گوشه سمت راست دهان مطابق شکل (۲) استفاده می‌کند. برای توضیح بیشتر می‌توان گفت، الگوریتم ابتدا نقاط نام برده را در تصویر دوبعدی دریافتی پیدا می‌کند، سپس یک سیستم مختصات سه‌بعدی جدید در نظر می‌گیرد که در این سیستم مختصات، شش نقطه‌ی ذکر شده در مختصات مشخص قرار گیرند. لازم به ذکر است که مختصات ذکر شده، مختصاتی است که کتابخانه‌ی OpenCV از آن استفاده می‌کند.

در مواردی که مکان (U,V,W) یک نقطه سه بعدی مانند P در WCS است، چرخش R (یک ماتریس ۳ × ۳) و همچنین انتقال t (یک بردار ۳ × ۱) مشخص و از پیش تعیین شده هستند، مکان (X, Y, Z) نقطه P در سیستم مختصات دوربین را می توان با استفاده از معادله‌ی (۱) با توجه به مختصات دوربین می‌توان محاسبه کرد. در واقع (U,V,W) همان نقاط مرجع هستند و (X, Y, Z) قابل محاسبه است.

به صورت گسترش یافته، معادله (۱) را می‌توان به صورت معادله ی (۲) نوشت.

$$\begin{bmatrix} X \\ Y \\ z \end{bmatrix} = R \begin{bmatrix} U \\ V \\ W \end{bmatrix} + t \qquad (۱)$$




شکل ۲ نقاط مورد استفاده در سیستم WCS در صورت [17].

$$\begin{bmatrix} X \\ Y \\ z \end{bmatrix} = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_x \\ r_{10} & r_{11} & r_{12} & t_y \\ r_{20} & r_{21} & r_{22} & t_z \end{bmatrix} \begin{bmatrix} U \\ v \\ w \\ 1 \end{bmatrix} \qquad (۲)$$

پارامترهای (tx, ty, tz) و ri,j ناشناخته هستند. شایان ذکر است که معادله فوق یک سیستم خطی معادلات است که روش حل مشخص دارد. برای حل این معادله کافی است (X, Y, Z) را وارد کنیم.

مختصات سه بعدی (X,Y,X) با استفاده از مختصات دو بعدی (x,y) با استفاده از رابطه (۳) به دست می آید.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = s \begin{bmatrix} fx & 0 & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (۳)$$

که در این معادله fx و fy فاصله‌ی کانونی (فاصله دوربین تا صورت) در جهات x و y، هستند و (cx , cy) مرکز نور و s عامل مقیاس است. برای حل معادله‌ی (۳) که یک عامل ناشناخته s را دارد، تبدیل مستقیم خطی (DLT) استفاده می‌شود.

پس از محاسبه ماتریس‌های R و t، انتقال سر راننده و چرخش مشخص می‌شود. حال با داشتن این مقادیر می‌توان در مورد میزان حرکت و چرخش سر اظهار نظر کرد. رویکرد کلی به این صورت است که پس از محاسبه‌ی این دو مقدار در مواردی که چرخش و انتقال از محدوده خاصی خارج است، یک هشدار برای راننده ارسال می‌شود. در غیر این صورت، برای بریدن ناحیه چشم به مرحله بعدی می‌رود. خارج از محدوده بودن سر شامل چهار شرط ناامن است: چرخش سر به سمت چپ، راست، پایین و بالا به حدی که سرعت واکنش نشان دادن راننده به میزان قابل توجهی کاهش پیدا کند یا دید کامل برای تسلط به جاده را از وی بگیرد. همانطور که پیش‌تر اشاره شد این محدوده به عوامل متعدی از جمله شرایط جوی، قدرت بینایی راننده، شکل صورت راننده و ... وابسته است. ما در اینجا تلاش داشتیم تا ویژگی‌های ظاهری راننده نیز در نظر گرفته شود. محدوده مجاز برای چرخش با توجه به عرض صورت و چشم محاسبه می‌شود. به این‌صورت که رنج مجاز برای چرخش هر شخص نسبتی از عرض صورت و فاصله‌ی شش نقطه‌ی نام برده از یکدیگر است و سیستم شخصی‌سازی می‌شود. این عرض صورت و فاصله‌ی شش نقطه از طریق همان نقاط لندمارک محاسبه می‌شود. در شکل (۳)، ۶۸ نقطه‌ی لندمارک که الگوریتم به دنبال یافتن آنهاست دیده می‌شود. در این شکل عرض صورت که برای تعیین میزان رنج مجاز چرخش استفاده می‌شود، نشان داده شده‌است. در معادله‌ی (۴) و (۵) نحوه‌ی محاسبه‌ی آستانه‌ی چرخش به راست و چپ، بالا و پایین به ترتیب دیده می‌شود.

$$LR = offset_x / |shape\ [1, x] shape\ [17, x]| \qquad (۴)$$

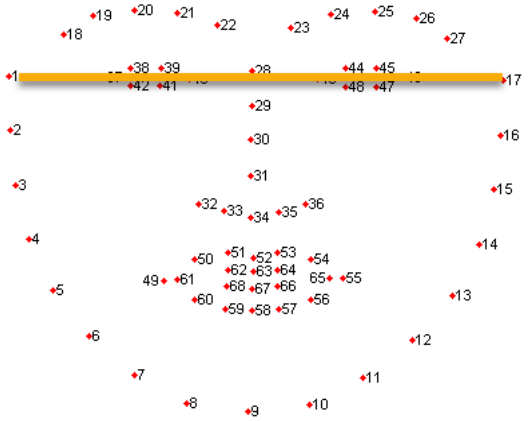
$$UD = (offset_y \times 2) / |shape\ [1, y] shape\ [17, y]| \qquad (۵)$$

که در آن آفست نشان‌دهنده‌ی میزان جابه‌جایی در راستای x و y است، shape [i, j] بیان کننده‌ی مختصات نقطه‌ی i در جهت j است.

$$I = \{1, \ldots, i, \ldots, 68\}$$

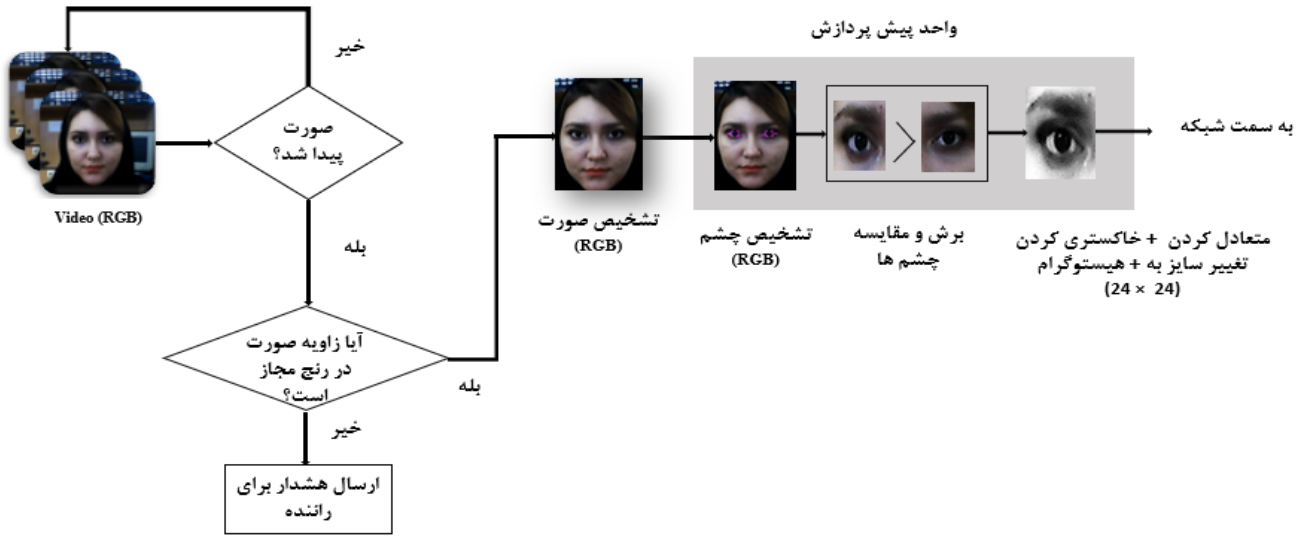
$$J = \{x, y\}$$

پس ازآن مقدار انحراف سر راننده نرمالیزه شده و سپس با آستانه که از طریق LR و UD محاسبه می‌شود، مقایسه می‌شود.



شکل ۳ ۶۸ نقطه‌ی لندمارک صورت و نقاطی که به عنوان عرض صورت مشخص شده‌اند (فاصله‌ی نقاط ۱ تا ۱۷).

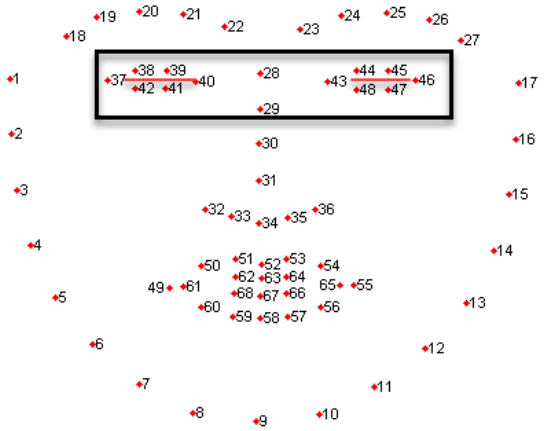
الگوریتم معرفی شده توسط ویولا و جوناس [5] برای تشخیص صورت و کتابخانه‌ی تشخیص نقاط برجسته‌ی صورت Dlib's برای دستیابی به نقاط لندمارک استفاده شده‌است.




شکل۴ – الگوریتم پیشنهادی برای تشخیص ناحیه‌ی چشم.

## ۳-پیدا کردن ناحیه مورد علاقه[1]

در این مرحله سعی داریم منطقه چشم را پیدا کرده و آماده کنیم و سپس آن را به شبکه عصبی مصنوعی جهت کلاس‌بندی وارد کنیم. چارچوب پیشنهادی این قسمت در شکل (۴) نشان داده شده‌است. برای یافتن چشم، ابتدا ناحیه سر از طریق الگوریتم ویولا و جونز تخمین زده می‌شود [5] . تشخیص نقاط لندمارک صورت، اجرای کار کاظمی و سالیوان [18] است. مزیت این روش برای دیگران در دقت بالای تشخیص در موقعیت‌های مختلف سر است. پس از رسیدن به نقطه لندمارک چشم، ناحیه مورد علاقه بریده می‌شود. صورت تصویری متقارن است، از این رو برای تشخیص خواب‌آلودگی، فقط بررسی یک چشم کافی است. این روش زمان محاسباتی را برای آشکارسازی کاهش می‌دهد. از بین دو چشم، چشمی باید انتخاب شود که جلوی دوربین است زیرا حاوی اطلاعات بیشتری است و میزان خطا را کاهش می‌دهد. برای این هدف، الگوریتم مسافت فاصله از راست‌ترین و چپ‌ترین نقطه از چشم راست و چپ را محاسبه می‌کند و آنها را برای انتخاب فاصله بیشتر برای برداشت محاسبه می‌کند. فاصله مورد نظر درشکل (۵) نشان داده شده است. الگوریتم فاصله مطلق بین نقاط ۳۷ و ۴۰، ۴۳ و ۴۶ را اندازه گیری می‌کند و فاصله بزرگتر را به عنوان چشم جلوی چشم دوربین انتخاب می‌کند. برای غلبه بر چالش وضعیت روشنایی، نویسندگان از یک اکولایزر هیستوگرام برای برابری کنتراست چشم استفاده می‌کنند. پس از تشخیص و انتخاب چشم، به شبکه جهت کلاس‌بندی ارسال می‌شود.



شکل ۵ فاصله‌ی مورد نظر برای انتخاب چشم جلوی دوربین (نقاط ۳۷ تا ۴۰ و ۴۳ تا ۴۶).

## ۴-پایگاه داده

برای بررسی عملکرد سیستم پیشنهادی دو مجموعه داده مورد بررسی قرار گرفته‌است. اولین مورد، مجموعه داده ZJU است، و دسته دوم مجموعه داده ترکیبی از ZJU و پایگاه داده ایجاد شده توسط نویسندگان است. هر یک از این پایگاه‌های داده در زیر شرح داده‌شده‌اند.

### ۴-۱-مجموعه داده‌ی ZJU

اولین مجموعه داده، گالری ZJU از پایگاه داده ZJU Eyeblink است [۱۹]. این مجموعه داده مجموعه تصاویری است که از ۸۰ کلیپ ویدیویی در پایگاه داده ویدیویی پلک‌زن جمع‌آوری شده است. برای هر نفر ۴ عدد ویدیو ضبط شده و تعداد شرکت کنندگان ۲۰ نفر است، چهار کلیپ برای هر فرد: یک کلیپ برای

[1] Region of Interest (ROI)

نمای جلو بدون عینک، یک کلیپ با نمای جلو و پوشیدن عینک رینگ نازک، یک کلیپ برای نمای جلو و عینک قاب سیاه و آخرین کلیپ نمای رو به بالا بدون عینک. همچنین، تصاویر از چشم چپ و راست به طور جداگانه جمع‌آوری می‌شود. قابل ذکر است که دراینجا تنها قاب عینک سیاه است و همچنان چشم قابل تشخیص است.

از آنجا که صورت متقارن است، بسیاری از محققین از یک رویکرد مبتنی بر تقارن صورت استفاده می‌کنند. شایان ذکر است که استفاده از نسخه‌های زیر نمونه‌برداری شده و خاکستری تصاویر کافی است [۲۰]. در این تحقیق فقط یک چشم برای تشخیص خواب‌آلودگی در نظر گرفته می‌شود و آن هم چشم بزرگتر است، اما هر دو چشم راست و چپ از طریق الگوریتم آموزش داده می شوند و بعد الگوریتم، چشمی که مساحت بیشتری از تصویر را دارد برای پردازش بیشتر به شبکه می‌فرستد و چشم کوچکتر حذف می‌شود، این رویکرد به این دلیل است که در مواقعی که صورت مایل است تصمیم‌گیری بر اساس طرفی از صورت باشد که رو به دوربین است و طرف دیگر که اطلاعات کمتری در اختیار شبکه می‌گذارد و باعث تصمیم‌گیری غلط شبکه می‌شود، حذف گردد. پس اولین مرحله پس از برش ناحیه مربوط به چشم آن است که اندازه‌ی دو چشم راست و چپ با یکدیگر مقایسه شوند و چشم بزرگ‌تر برای پردازش بیشتر انتخاب شود، اما در مرحله‌ی آموزش، مجموعه داده‌هایی شامل چشم راست و چپ برای آموزش انتخاب می‌شوند تا شبکه توانایی تشخیص برای هر دو چشم را دارا باشد. برخی از نمونه‌های مجموعه داده‌های ZJU در شکل (۶) نشان داده‌شده‌است. این مجموعه داده دارای ۴۸۴۱ تصویر، ۲۳۸۳ چشم بسته و ۲۴۵۸ چشم باز است. همه این تصاویر از نظر هندسی به ابعاد ۲۴ × ۲۴ پیکسل تغییر می‌یابند.

### ۲-۴-مجموعه داده پیشنهادی

در این تحقیق، مجموعه داده ZJU با مجموعه داده‌های شامل ۴۱۸۵ تصویر (۲۱۰۶ تصویر با چشم باز و ۲۰۷۹ تصویر با چشم بسته) گسترش یافته‌است. این مجموعه داده با یک دوربین وب کم با کیفیت p HD۷۲۰ تهیه‌شده و این تصاویر از یک ویدئو جمع‌آوری می‌شود. نرخ ۶ فریم/ثانیه برای نمونه‌برداری در نظر گرفته‌شده. نقطه قوت این مجموعه داده این است که وضعیت متفاوتی از چشم‌ها را در نظر می‌گیرد. داده‌ها از مسافت‌های مختلف، چرخش و زاویه‌های متفاوت جمع‌آوری می‌شوند که باعث افزایش آزادی عمل رانندگان می‌شود که به وضعیت واقعی نزدیک‌تر است.

داده‌ها همچنین از رانندگان با و بدون عینک جمع‌آوری شده. در حالت کلی می‌توان این مجموعه را به دو گروه اصلی تقسیم کرد، دسته اول شامل داده‌هایی است که سر رانندگان به طور مستقیم به جلو نگاه میکند و چشم‌ها چرخش‌های متفاوتی دارند، به این معنی که زاویه دید به طور مستقیم است اما چشم‌ها با زاویه‌های متفاوت قرار دارند. دسته دوم داده‌هایی را شامل می‌شود که راننده سر خود را می چرخاند اما در محدوده مجاز، در این حالت دیگر زاویه دید چشم‌ها از جلو نیست و با زاویه دید متمایل (چشم در حالت نیمرخ) مواجه هستیم. دسته دوم یک رویکرد جدید برای مجموعه داده‌های چشم است که قبلا توسط محققان به منظور آموزش شبکه‌ی عصبی بررسی نشده. در گذشته تحقیقاتی در خصوص نوشتن یک کتابخانه مخصوص تشخیص چشم متمایل انجام شده‌است [۲۱] اما پایگاه داده‌ای مربوط به چشم متمایل و آموزش با شبکه‌ی عصبی وجود ندارد. شبکه‌های پیشنهاد شده با هر دو دسته داده دید مستقیم و متمایل به طور همزمان آموزش داده می‌شوند. از دیگر نقاط قوت این مجموعه داده‌ی پیشنهادی می‌توان گفت که مجموعه داده‌ی ZJU فقط دارای ملیت چینی است، اما داده‌های موجود در اینجا قومیت دیگری را شامل می‌شود، که انواع داده‌های مورد استفاده برای آموزش را گسترش می‌دهد.

روش پیشنهادی این مقاله شامل شرایط مختلف نورپردازی است و همه تصاویر از یک متعادل‌کننده‌ی هیستوگرام عبور داده می شوند. برخی از نمونه‌های نگاه مستقیم و متمایل به‌ترتیب در شکل (۷) و شکل (۸) نشان داده‌شده‌است. تصاویر نمایش داده شده پس از متعادل‌کردن هیستوگرام هستند. مجموعه داده‌های پیشنهادشده با ZJU ترکیب شده‌اند تا یک بانک اطلاعاتی جامع‌تر برای آموزش مدل ایجاد کنند.

شکل ۶ چند نمونه از مجموعه داده‌ی ZJU. ردیف اول: چشم بسته‌ی چپ، ردیف دوم: چشم بسته‌ی راست، ردیف سوم: چشم باز چپ و ردیف چهارم: چشم باز راست.

شکل ۷ چند نمونه از مجموعه داده‌ی پیشنهادی با نگاه مستقیم. ردیف اول: چشم بسته، ردیف دوم: چشم باز.

شکل ۸ چند نمونه از مجموعه داده‌ی پیشنهادی با نگاه متمایل. ردیف اول: چشم بسته، ردیف دوم: چشم باز.

## ۵-ساختار شبکه

در روش پیشنهادی خود سه شبکه عصبی مختلف را طراحی کرده‌ایم که هر یک ویژگی‌های خود را دارند، این شبکه‌ها برای هر دو مجموعه داده نام برده استفاده می‌شوند ونتایج آن با یکدیگر مقایسه می‌شود. در ادامه به توضیح تئوری شبکه‌ها می پردازیم.

### ۵-۱-شبکه عصبی تمام طراحی شده

اولین شبکه عصبی به عنوان یک شبکه عصبی کاملاً طراحی شده(FDNN[1]) است. معماری کلی شبکه و لایه‌های مدل در شکل (۹) نشان داده شده‌است. این شبکه یک شبکه‌ی عصبی کانولوشنی است، بدان معنا که در لایه‌های خود از عمل کانولوشن استفاده می‌کند. در لایه‌ی اول که به عنوان لایه‌ی ورودی است یک تابع کانولوشنی استفاده‌شده‌است، ابعاد ورودی این تابع برابر است با (تعداد کانال‌های تصویر× تعداد ستون ماتریس تصویر × تعداد سطر ماتریس تصویر). تابع فعال‌سازی Relu[2] برای افزایش ویژگی‌های غیرخطی وجود دارد. تابع Relu مطابق رابطه‌ی (۶) تعریف می‌شود که در آن x ورودی نورون است.

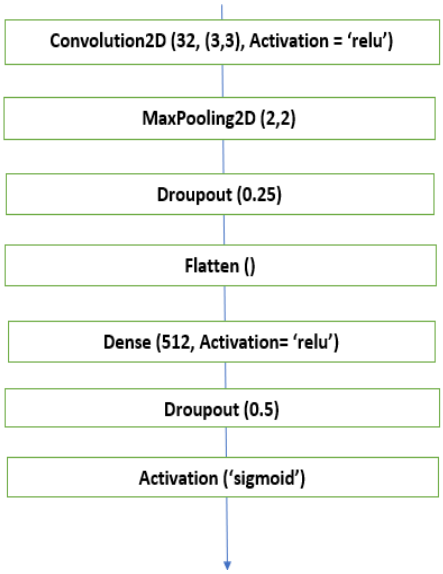
$$S(x)=\frac{1}{1+\exp\{-x\}} \qquad (۶)$$

در حالت کلی با توجه به شکل (۹) می‌توان گفت یک لایه‌ی کانولوشنی دو بعدی در این ساختار با فیلتری به ابعاد ۳ × ۳ وجود دارد، برای تصویر که معمولا دو بعد سطر و ستون را دارند از کانولوشن دوبعدی استفاده می‌شود. لایه‌ی Maxpooling به منظور حذف ویژگی‌های کم‌اثر و تقویت ویژگی‌ها مهمتر استفاده می‌شود. در لایه‌ی Dropout ارتباط بعضی نورن‌ها با لایه‌های بعدی قطع می‌شود در این حالت می‌توان انتظار داشت مدت زمان آموزش کوتاه‌تر شود. اندازه فیلتر Maxpooling ۲×۲ است، به این معنی که از هر ۴ پیکسل تصویر بیشنه را انتخاب و جایگزین می‌کند و نسبت Dropout به ترتیب ۰.۲۵ و ۰.۵ است، بدان معنی که ۲۵٪ و ۵۰٪ از ارتباط‌های تمام متصل، قطع می‌شوند. برای آخرین لایه با توجه به اینکه طبقه‌بندی باینری داریم، ازتابع Sigmoid در لایه خروجی استفاده می‌کنیم. این تابع مطابق با فرمول (۷) تعریف می‌شود که در آن x ورودی نورون است. چنانچه پیش‌بینی شبکه بزرگتر از سطح آستانه (در اینجا ۰.۵) باشد، به عنوان کلاس ۱ طبقه‌بندی می‌کند و چنانچه کمتر از مقدار آستانه باشد، به عنوان کلاس ۰ می‌شناسد. سایرتوابع فعال‌سازی به جز آخرین لایه Relu است. هایپرپارامترهای انتخاب‌شده برای هر شبکه نیز در بخش مربوط به آنها ذکر شده‌است.

$$f(x) = \max(0, x) \qquad (۷)$$

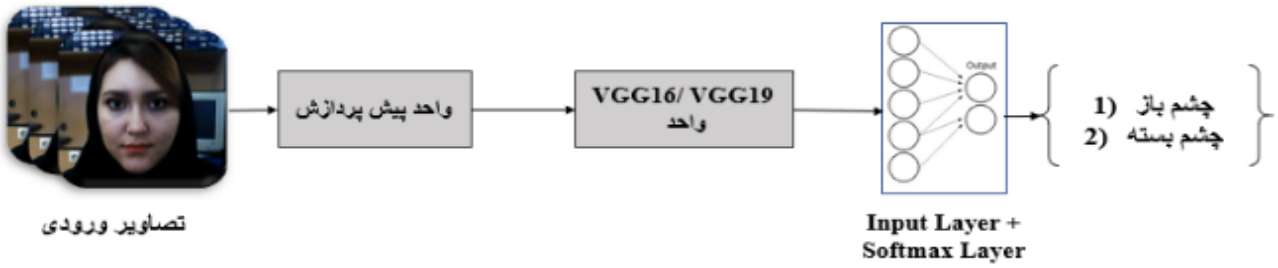
برتری FDNN به شبکه‌های موجود عدم پیچیدگی و سرعت آن است، ضمن اینکه صحت قابل قبولی را هم از خود نشان می‌دهد. این شبکه برای ZJU و مجموعه داده‌های گسترش‌یافته‌ی ما اعمال شده است.

در شبکه‌های دوم و سوم پیشنهادشده، از مفهوم انتقال‌یادگیری و از استفاده از شبکه‌های پیش آموزش داده‌شده، به خصوص شبکه های عصبی کانولوشنی استفاده شده‌است. با توجه به ویژگی‌های مجموعه داده که مخصوص چهره است، شبکه‌ی VGG16و VGG19انتخاب می‌شوند.




شکل ۹ معماری شبکه‌ی پیشنهادی FDNN

[1] Fully Designed Neural Network
[2] Rectified linear unit




شکل ۱۰- ساختار شبکه‌ی انتقال یادگیری **VGG16** و **VGG19**

## ۵-۲-شبکه انتقال یادگیری VGG16

VGG16 یک مدل شبکه عصبی کانولوشنی است که توسط سیمونیان و زیرسمن ارائه شده است [۲۲]. این مدل در مجموعه داده‌ی ImageNet به دقت ۹۲.۷٪ دست می‌یابد. ورودی لایه‌ی اول این شبکه تصاویر RGB با اندازه‌ی ثابت ۲۲۴×۲۲۴ است. تصویر از میان لایه‌های کانولوشنی عبور می‌کند، جایی که فیلترها با کوچکترین اندازه‌ی ممکن که ۳×۳ است استفاده شده اند (که کوچکترین اندازه ممکن برای درک مفهوم چپ / راست، بالا / پایین و مرکز تصویر است). پدینگ برای فیلترهای کانولوشنی ۳×۳ یک پیکسل است معماری شبکه دارای پنج لایه ی Max pooling است که روی یک پنجره‌ای با ابعاد ۲×۲ پیکسل اعمال می‌شود. سه لایه کاملاً متصل وجود دارد. دو لایه اول هر کدام دارای ۴۰۹۶ کانال هستند، لایه‌ی سوم مرتباً در حال طبقه‌بندی به ۱۰۰۰ کلاس است و بنابراین شامل ۱۰۰۰ کانال (یک کانال برای هر کلاس است). پیکربندی لایه‌های کاملاً متصل در همه جای شبکه یکسان است. لایه‌ی نهایی شبکه Softmax است. شبکه‌های پیشنهادی بدین صورت هستند که ویژگی‌های سطح پایین با وزن‌های از پیش مشخص‌شده‌ی مجموعه داده ImageNet مشخص می‌شوند و ویژگی‌های سطح بالا را با سه لایه کاملاً متصل طراحی شده، تعلیم داده می‌شود. سه لایه آخر در هر دو شبکه کاملا مشابه بوده. لایه اول به عنوان لایه‌ی ورودی با ابعاد ۷×۷×۵۱۲ است، در لایه دوم یک لایه‌ی Dense تابع فعالسازی Relu و اندازه‌ی ۱۰۲۴ را داریم و در لایه آخر از تابع Softmax به عنوان تابع تصمیم‌گیرنده برای خروجی استفاده‌شده. استفاده از این تابع برای حالتی که دو کلاس داریم بلامانع است. معماری

کلی شبکه‌ی TLVGG16 در شکل (۱۰) نشان داده‌شده‌است. همان‌طور که مشاهده می‌شود خروجی شبکه‌ی VGG16 1000 کلاسی است که برای آن طراحی شده. حال سه لایه‌ی آخر سعی بر آموزش ویژگی جدید چشم با استفاده از دانشی هستند که وزن‌های شبکه‌ی VGG16 در اختیار آن‌ها قرار می‌دهد. این شبکه‌ی جدید را که از ترکیب شبکه‌ی عصبی عمیق VGG16 و سه لایه کاملا متصل ایجاد شده را شبکه‌ی انتقال یادگیری VGG16 یا به اختصار TLVGG16 می‌نامیم.

## ۵-۳-شبکه انتقال یادگیری VGG19

یک شبکه‌ی کانولوشنی است که به عنوان یک نسخه‌ی عمیقتر از VGG16 آموزش داده شده‌است. این شبکه دارای ۱۹ لایه است و ابعاد تصویر ورودی به شبکه همانند شبکه‌ی قبلی ۲۲۴×۲۲۴ پیکسل است. ساختار این شبکه همانند VGG16 است که پیشتر توضیح داده شده است. سه لایه‌ی پیشنهادی در هر دو شبکه‌یVGG16 و VGG19 کاملا یکسان است. از ترکیب شبکه‌ی عصبی عمیق VGG19 و سه لایه کاملا متصل ایجاد شده را شبکه‌ی انتقال یادگیری VGG19 یا به اختصار TL VGG19 می‌نامیم

روش‌های مبتنی بر انتقال یادگیری صحت را افزایش می‌دهند و هدف استفاده از آن‌ها، داشتن یک شبکه عمیق‌تر و با دقت بیشتر به ویژه در مواقعی که مجموعه داده‌های آموزش کوچک است و یا قصد کاهش زمان آموزش شبکه را داریم، زیرا قسمتی زیادی از وزن‌ها که ویژگی‌های سطح پایین را تعریف می‌کنند از قبل آموزش داده شدند و تلاش شبکه برای یادگیری ویژگی‌های سطح بالا است که درنتیجه نیاز به میزان کمتری وزن و زمان و در نتیجه داده‌ی کمتری برای آموزش دارد.

## ۶-تصمیم‌گیری

به عنوان آخرین مرحله تشخیص، از راننده ۶ تصویر در ثانیه گرفته می‌شود. اگر شبکه احتمال بسته بودن چشم‌ها را برای بیش از ۲۰ تصویر پیاپی بیش از ۵۰٪ تخمین بزند، به عنوان نشانه‌ای از خواب‌آلودگی در نظر گرفته می‌شود. نتایج تجربی نشان می‌دهد که برای حالت پلک‌زدن معمولی، این درصد باید کمتر از ۲۰ تصویر پیاپی باشد، در واقع مدت زمان مورد نیاز برای پلک‌زدن در هرشخص متفاوت و به عوامل محیطی و فیزیکی زیادی وابسته است، اما می‌توان گفت پلک‌زدن به طور

میانگین کمتر از ۲۰ فریم پیاپی خواهد بود. بنابراین درصد بیشتر از ۵۰ درصد به مدت بیشتر از ۲۰ فریم پیاپی نشانه خستگی چشم است. پس از این تشخیص، زنگ خطری برای بیداری به راننده ارسال می‌شود.

## ۷-بررسی نتایج و تفسیر آن‌ها

در این بخش ابتدا به مقایسه شبکه‌های پیشنهاد شده با دیگر شبکه‌های ارائه شده که از مجموعه داده‌ی ZJU استفاده کرده‌اند، می‌پردازیم.

### ۷-۱-مشخصات پیاده‌سازی

در این تحقیق، ما از سیستمی با پردازنده Intel Core i76700K CPU @ 4.00GH با ۱۶ گیگابایت رم و NVIDIA GeForce GTX 1070 استفاده کرده‌ایم. مجموعه داده ZJU و مجموعه داده گسترش‌یافته ما برای تشخیص خواب‌آلودگی استفاده شده است. در هر تکرار، به ترتیب ۷۰٪ و ۳۰٪ از تصاویر مجموعه داده ZJU برای آموزش و ارزیابی استفاده می‌شوند. در مجموعه داده‌ی گسترش یافته در هر تکرار ۸۰٪ از داده‌ها برای آموزش ۱۰٪ برای آزمون و ۱۰٪ هم برای ارزیابی مورد استفاده قرار گرفته‌اند.

### ۷-۲-ارزیابی صحت

مجموعه داده‌هاZJU، که برای آزمایش سه شبکه‌ی پیشنهادی استفاده می‌شود. شامل تصاویر ثابت از چشم در شرایط مختلف نورپردازی، جهت‌های مختلف مردمک چشم و دیگر ویژگی‌ها است. چشم در فرایند آموزش به دو دسته طبقه‌بندی می شود، چشم باز یا بسته، این تصاویر بدون توجه به اینکه مربوط به چشم راست یا چپ است آموزش داده می‌شوند. این نواحی مورد علاقه که توسط نقاط لندمارک انتخاب شده‌است، با استفاده از شبکه انتقال یادگیری (TLVGG16)، شبکه (TLVGG19) و شبکه عصبی کاملاً طراحی شده(FDNN) به عنوان ورودی سیستم تشخیص خواب آلودگی راننده در نظر گرفته می‌شود.

مجموعه داده‌ZJU ، که برای آزمایش مدل استفاده می‌شود، شامل تصاویری از چشم در شرایط مختلف روشنایی، جهت‌های مختلف شرکت کنندگان و ویژگی‌های مختلف چشم است. این نواحی مورد علاقه که توسط نقاط لندمارک انتخاب شده است، با استفاده از شبکه TLVGG16، شبکه TLVGG19 و همچنین FDNN، آموزش داده می‌شوند. نویسندگان سه شبکه را با مجموعه داده ZJU آموزش دادند. تعداد تصاویر در مجموعه داده های‌ZJU در جدول (۱) قرار دارد.

جدول ۱ اندازه مجموعه داده ZJU

| چشم باز | چشم بسته | تعداد کلی |
|---|---|---|
| ۲۰۵۷ | ۲۱۰۰ | ۴۱۵۷ |

در مجموعه داده ZJU، برای اینکه بتوانیم شبکه‌های طراحی شده در این تحقیق را با سایر آثار پیشنهادی مقایسه کنیم، مطابق منابع مرجع فقط از داده‌های آموزش و اعتبار سنجی استفاده کردیم. نزول گرادیان تصادفی و آنتروپی متقاطع به ترتیب به عنوان عملکرد بهینه‌ساز و تابع هزینه انتخاب شده است. ۰.۰۱ به عنوان نرخ یادگیری و ۷۰٪ داده‌ها برای آموزش و بقیه برای اعتبارسنجی انتخاب شدند. نتایج مربوط به صحت شبکه‌های موجود در مجموعه داده‌ZJU، درجدول (۲) گزارش شده است، لازم به ذکر است که صحت مطابق با رابطه (۸) محاسبه می‌شود. که در آن tp نشان دهنده‌ی مثبت صحیح، tn نماینده‌ی منفی صحیح، fp نشان دهنده‌ی مثبت اشتباه و fn به معنی منفی اشتباه است. FDNN، TLVGG16 و TLVGG19 به ترتیب به صحت ۹۸.۱۵٪، ۹۵.۴۵٪ و ۹۴.۹۶٪ در تصاویر ROI دست میابند. در جدول (۳) فراخوان و دقت برای شبکه‌ی FDNN گزارش شده است.

جدول ۲ نتایج سه شبکه‌ی پیشنهادی برای مجموعه داده‌ی ZJU

| شبکه | صحت | سطح زیر منحی | تکرار |
|---|---|---|---|
| FDNN | 98.15% | 99.8% | 50 |
| TLVGG16 | 95.45% | 99.0% | 100 |
| TLVGG19 | 94.96% | 99.0% | 100 |

فراخوان و دقت به ترتیب مطابق روابط (۹) و (۱۰) محاسبه شده اند. نویسندگان همچنین روی مجموعه داده‌های گستردش داده شده که حاوی تصاویر ZJU و تصاویر پیشنهادی است، آزمایش می‌کنند. تصاویر پیشنهادی چهار دسته دارند که شامل ۱) چشم بسته و نگاه به جلو، ۲) چشم باز و نگاه به جلو، ۳) چشم بسته و سر چرخان، و ۴) چشم باز و سر چرخانده است. تعداد تصاویر در هر گروه در جدول (۴) ثبت شده است. در مجموعه داده گسترش یافته با در نظر گرفتن بهینه ساز SGD و ۰.۰۱ برای نرخ یادگیری سه شبکه پیشنهادی آموزش داده می‌شود و تابع هزینه آنتروپی متقاطع است.

جدول ۳ نتایج دقت و فراخوانی برای شبکهFDNNدر مجموعه دادهیZJU.

| فراخوانی | دقت | FN[1] | TN[2] | FP[3] | TP[4] | تعداد کل تصاویر |
|---|---|---|---|---|---|---|
| ٪۸۶.۷ | ٪۹۹.۸ | ۸۵ | ۶۳ | ۱ | ۵۵۸ | ۱۲۴۷ |

۸۰٪ از داده برای آموزش، اعتبار سنجی ۱۰٪و آزمون ٪۱۰ مورد استفاده قرار می‌گیرد. از جدول (۵) می‌توان دریافت که برای دادههای اعتبارسنجی، شبکه‌هایFDNN، TL،TLVGG16 VGG19 و به ترتیب به صحت۹۷.۰۱٪، ۹۸.۵۳٪ و ۹۶.۴۲٪ می رسند. همانطور که مشاهده می‌شود بیشترین دقت متعلق TL VGG 16 است. در بخش آزمون، صحت به ترتیب ۹۶.۷۹٪، ۹۷.۵۴٪ و ۹۶.۰۹٪ است. دقت و فراخوان همچنین در جدول (۶) گزارش شده‌است. در اینجا هم بیشترین دقت متعلق به TL VGG 16 است. مقایسه سه نتیجه شبکه پیشنهادی از مجموعه داده‌های گستردش‌یافته با تنها مجموعه داده ZJU دشوار است زیرا در مجموعه داده‌های گستردش‌یافته ما از آزمایش و اعتبار سنجی استفاده می‌کنیم، این درحالی است که در ZJU فقط اعتبارسنجی گزارش شده‌است و امکان مقایسه با مقالات مرجعی که از داده‌ی اعتبارسنجی استفاده نکرده‌اند فراهم نیست. علت تاکید بر استفاده از داده‌های اعتبارسنجی در مجموعه‌داده گسترش یافته آن است که ارزیابی برروی صحت شبکه روی داده های تازه اضافه‌شده داشته باشیم. منحصر به‌فرد بودن مجموعه داده‌های گسترش‌یافته در قسمت نمای مایل است، یعنی شبکه با چشم در حالت دید مایل هم آموزش داده‌شده.

جدول۴ اندازه مجموعه دادهی گسترش یافته

| چشم باز و سر چرخیده | چشم بسته و سر چرخیده | چشم باز و سر مستقیم | چشم بسته و سر مستقیم | تعداد کلی |
|---|---|---|---|---|
| ۶۶۱ | ۵۵۸ | ۱۴۴۵ | ۱۵۲۱ | ۴۱۸۵ |

همان‌طور که در بخش‌های قبلی ذکر شد، صورت متقارن است. بنابراین، فقط یک چشم در فرآیند تشخیص خواب‌آلودگی شرکت می‌کند که این رویکرد موجب کاهش زمان محاسبه و کاهش احتمال تشخیص اشتباه است و همواره از چشمی استفاده می‌شود که اطلاعات بیشتری را ارائه می‌کند.

جدول ۵ نتایج سه شبکه‌ی پیشنهادی برای مجموعه داده‌ی گسترش یافته.

| شبکه | صحت برای داده‌های آزمون | سطح زیر منحنی برای داده‌های آزمون | صحت برای داده‌های ارزیابی | سطح زیر منحنی برای داده‌های ارزیابی |
|---|---|---|---|---|
| FDDNN | ٪۹۷.۰۱ | ٪۹۹.۴ | ٪۹۶.۷۹ | ٪۹۹.۳ |
| TL VGG16 | ٪۹۸.۵۳ | ٪۹۹.۸ | ٪۹۷.۵۴ | ٪۹۹.۵ |
| TL VGG19 | ٪۹۶.۴۲ | ٪۹۹.۴ | ٪۹۶.۰۹ | ٪۹۹.۳ |

جدول ۶ نتایج دقت و فراخوانی برای شبکه FDNN در مجموعه دادهی گسترش‌یافته

| فراخوانی | دقت | FN | TN | FP | TP | تعداد کل تصاویر |
|---|---|---|---|---|---|---|
| ٪۹۸.۱۷ | ٪۹۸.۸ | ۸ | ۳۵ | ۵ | ۴۲۶ | ۸۳۴ |

.

$$\text{صحت} = (tp + tn) / (tp + tn + fn + fp) \quad (۸)$$

$$\text{فراخوان} = tp / (tp + fn) \quad (۹)$$

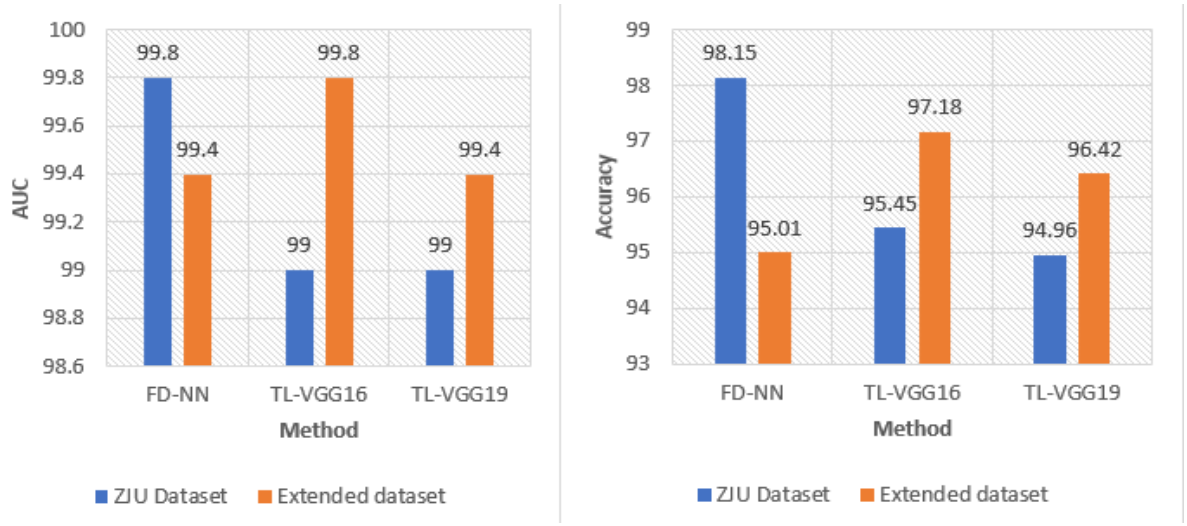
$$\text{دقت} = tp / (tp + fp) \quad (۱۰)$$

در جدول (۷)، صحت و ناحیه‌ی زیر منحنی هر روش‌های قبلی و پیشنهادی به تفصیل ارائه شده‌است. نتایج مجموعه داده‌های گسترش یافته و مجموعه داده‌های ZJU نیز در شکل (۱۱) مقایسه شده‌است. شکل نماینگر آن است که مجموعه داده‌های گسترش‌یافته در VGG16 و VGG19 منجر به دقت و ناحیه زیر منجنی بیشتر می‌شوند، زیرا مجموعه داده‌ی گسترش یافته نسبت به مجموعه داده‌ی ZJU اندازه‌ی بزرگتری دارد و برای آموزش آن الگوریتم‌های یادگیری عمیق‌تر (نسبت به شبکه‌ی FDNN) که نیاز به یک پایگاه داده بزرگ دارند مناسب‌تر هستند، مانند انتقال یادگیری اما در مقابل آن مجموعه داده‌ی ZJU به تنهایی بهتر است توسط شبکه‌هایی آموزش ببینند که ساده‌تر هستند و برای مجموعه داده‌های کوچکتر مناسبند.

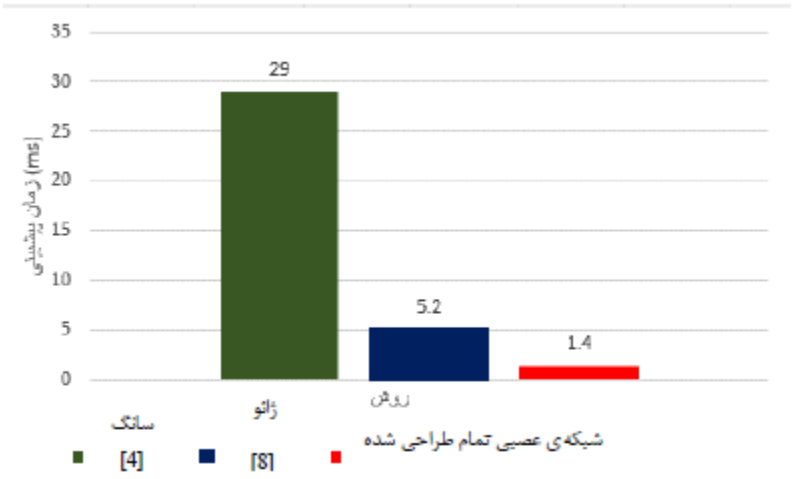
جدول ۷ مقایسه صحت و AUC بین روش پیشنهادی و دیگر تحقیقات روی مجموعه دادهی ZJU

| تحقیق | روش | صحت(٪) | سطح زیر منحنی (%) |
|---|---|---|---|
| پان[19] | Cas-Adaboost (W=3) | ۸۸.۸ | - |
| سانگ[4] | HPOG+ LTP+ Gabor(s) | ۹۵.۹۱ | ۸۹.۲۲ |
| سانگ[4] | MultiHPOG+ LTP+ Gabor | ۹۶.۸۳ | ۹۹.۲۷ |
| سانگ[4] | MultiHPOG+ LTP+ Gabor(s) | ۹۶.۴۰ | ۹۶.۶۷ |
| ژائو[8] | DNN | ۹۴.۴۵ | ۹۷.۹۱ |
| ژائو[8] | DCNN | ۹۵.۷۹ | ۹۸.۱۵ |
| ژائو[8] | DINN | ۹۷.۲۰ | ۹۹.۲۹ |
| روش پیشنهادی | FD-DNN | ۹۸.۱۵ | **۹۹.۸۰** |
| روش پیشنهادی | TL-VGG16 | ۹۵.۴۵ | ۹۹.۰۰ |
| روش پیشنهادی | TL-VGG19 | ۹۴.۶۴ | ۹۹.۰۰ |

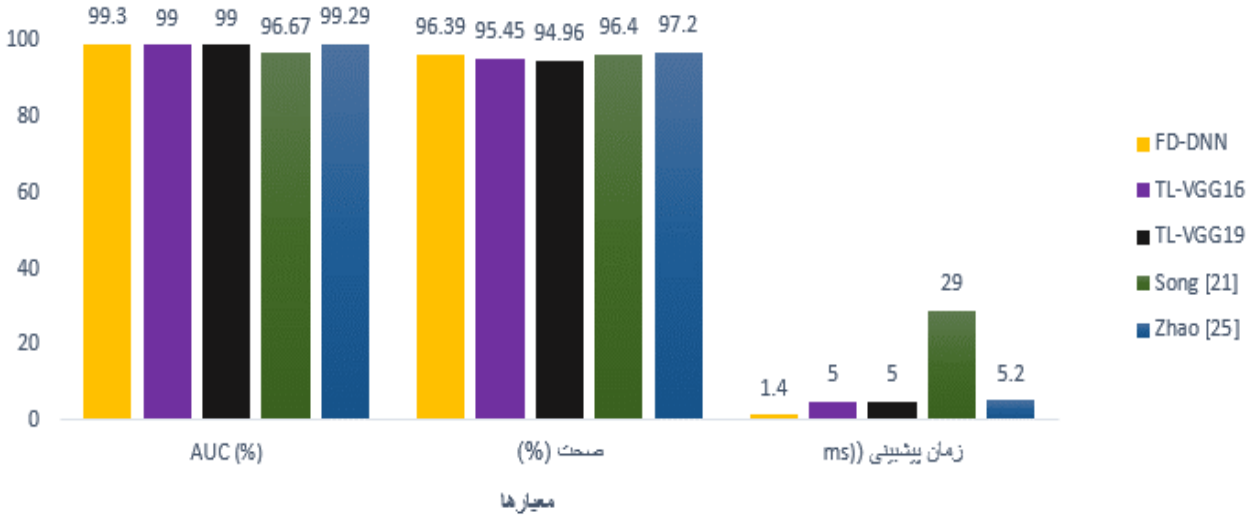
[1] False Negative
[2] True Negative
[3] False Positive
[4] True Positive




شکل ۱۱- مقایسه‌ی صحت و **AUC** بین سه روش پیشنهادی روی مجموعه داده‌های مختلف.




شکل ۱۲- مقایسه‌ی زمان تصمیم‌گیری برای شبکه‌های مختلف روی مجموعه داده **ZJU**.




شکل ۱۳- مقایسه‌ی معیار زمان تصمیم‌گیری، صحت و سطح زیر منحنی بین شبکه‌های مختلف برای مجموعه داده **ZJU**.

## ۸-ارزیابی پیچیدگی

برای مقایسه‌ی بهتر، پیچیدگی محاسباتی را باید به عنوان یک معیار مهم بین روش‌های مختلف در نظر گرفت. در کاربردهای بلادرنگ، روش‌های سریع‌تر قابل اطمینان‌تر و پرطرفدارتر هستند. در شکل (۱۲) زمان لازم برای تصمیم‌گیری در مورد بسته‌بودن یا باز بودن چشم برای هر روش در مجموعه داده ZJU ارائه شده‌است. FDNN تقریبا ۴ برابر و ۲۰ برابر سریعتر به ترتیب از مدل‌های ژائو و همکاران [۸] و سانگ و همکاران [۴] است. دلیل این امر این است که FDNN از پارامترهای کمتری نسبت به شبکه‌های اشاره‌شده استفاده می‌کند و پیچیدگی کمتری نسبت به بقیه دارد. در مقاله‌ی [۸] نویسندگان از شبکه ی عمیق با استفاده از انتقال یادگیری استفاده کرده‌اند و این شبکه جدید خود از ترکیب دو شبکه‌ی عمیق دیگر به دست‌آمده که این امر باعث پیچیدگی بیش از حد می‌شود.

در بحث طراحی شبکه‌ها همواره یک مصالحه‌ای وجود دارد که هرچه شبکه عمیق‌تر باشد، با صحت بالاتر کار می‌کند اما نیاز به سخت‌افزار قوی‌تر و مجموعه‌داده بزرگتر و زمان آموزش و تصمیم گیری بیشتر دارد اما هرچه شبکه پارامترهای کمتری داشته باشد، نیاز به مجموعه‌داده کوچکتر، حافظه‌ی کمتر برای آموزش و زمان آموزش و تصمیم‌گیری کمتر دارد (بدون در نظر گرفتن وضعیت شایستگی بیش از حد). با توجه به این مصالحه نتایج ما نشان می‌دهد که شبکه‌ی طراحی شده‌ی FDNN پیچیدگی محاسباتی کمتر نسبت به بقیه شبکه‌های معرفی شده را داراست و در بحث صحت و دقت توانایی خود را نسبت به سایر کارها نشان می‌دهد. به عنوان یک نتیجه‌گیری نهایی، سه معیار صحت، سطح زیر منحنی و زمان پیش‌بینی بین روش‌های پیشنهادی این مقاله و سایر آثار در مجموعه داده‌های ZJU مورد بررسی قرار گرفته است که می‌توان در شکل (۱۳) مشاهده کرد.

## ۹-نتیجه‌گیری و پیشنهادها

خواب آلودگی نقش مهمی در رانندگی ایمن بازی می‌کند بنابراین، در این مقاله چندین شبکه برای دستیابی به دقت بهتر و زمان محاسباتی کمتری برای تشخیص خواب آلودگی بر اساس وضعیت چشم مورد مطالعه قرار گرفته است.

در مرحله اول سیستم، نقاط لندمارک برای دسترسی به ناحیه مورد علاقه استفاده می‌شود، سپس ناحیه چشم انتخاب می‌شود و کنتراست نرمالیزه می‌شود، از خروجی این مرحله به عنوان ورودی به شبکه برای طبقه‌بندی حالت چشم استفاده کرده‌ایم. اگر شبکه تشخیص دهد که چشم بیش از ۱۲ تصویر پی‌درپی بسته است، زنگ خطر برای راننده ارسال می‌شود.

نتایج تجربی نشان می‌دهد که شبکه FDNN از مدل‌های موجود دقیق‌تر و سریع‌تر است.

در این تحقیق ما به دنبال معرفی یک مجموعه داده‌ی جامع برای تشخیص خواب‌آلودگی نیز بوده‌ایم. از همین رو، مجموعه داده‌ی جامعی را معرفی کردیم که مزیت مجموعه داده پیشنهادی نسبت به کارهای قبلی در نمای مورب است که باعث می شود سیستم در شرایط متنوع‌تری کار کند.

در آینده، نویسندگان قصد دارند تا بر روی تحلیل خمیازه برای تشخیص خستگی توسط نقاط عطف لب تمرکز کنند.

همچنین، سطح دیگری از خواب‌آلودگی بدون محدودیت در خواب و بیدار می‌تواند مورد بررسی قرار گیرد زیرا سطوح مختلف خواب‌آلودگی بسیار باریک است و نمی‌توان آن را در دو سطح خلاصه کرد [۲۳].

## مراجع


[1] World Health Organization The top ten causes of death. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs310/en/.

[2] G. Pan, L. Sun, Z. Wu, S. Lao, Eyeblinkbased anti spoofing in face recognition from a generic web camera, in: Proceedings of IEEE International Conference on Computer Vision, 2007, pp. 18.

[3] ZJU Eyeblink Database, http://www.cs.zju.edu.cn/~gpan or http://www.stat.ucla.edu/~gpan.

[4] Song, Fengyi & Tan, Xiaoyang & Liu, Xue & Chen, Songcan. (2014). Eyes closeness detection from still images with multiscale histograms of principal oriented gradients. Pattern Recognition. 47. 28252838. 10.1016/j.patcog.2014.03.024.

[5] P. Viola, M. Jones, Robust realtime face detection, Int. J. Comput. Vis. 57 (2007) 137–154.

[6] X. Tan, F. Song, Z. Zhou, S. Chen, Enhanced pictorial structures for precise eye localization under uncontrolled conditions, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1621–1628.

[7] G. Huang, V. Jain, E. LearnedMiller, Unsupervised joint alignment of complex images, in: Proceedings of IEEE International Conference on Computer Vision, 2007, pp. 1–8.

[8] Zhao, L., Wang, Z., Zhang, G. et al. Multimed Tools Appl (2018) 77: 19415.

[9] Michael Jay C. De Castro, Joel C. De Goma, Madhavi Devaraj, John Paul G. Lopez, and Joshua Rodgregor E. Medina. 2018. Distraction Detection through Facial Attributes of Transport Network Vehicle Service Drivers. In Proceedings of the 2018 International Conference on Information Hiding and Image Processing (IHIP 2018). ACM, New York, NY, USA, 112118


.

[10] Y. Xing et al., "Identification and Analysis of Driver Postures for InVehicle Driving Activities and Secondary Tasks Recognition," in IEEE Transactions on Computational Social Systems, vol. 5, no. 1, pp. 95108, March 2018.
[11] Q. Massoz, T. Langohr, C. Franois and J. G. Verly," The ULg multimodality drowsiness database (called DROZY) and examples of use," 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, 2016
[12] GarcaGarca, Miguel & Caplier, Alice & Rombaut, Michle. Sleep Deprivation Detection for RealTime Driver Monitoring Using Deep Learning, 2018.
[13] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, YawDD: A Yawning Detection Dataset, Proc. ACM Multimedia Systems, Singapore, March 19 21 2014.
[14] Feifei Zhang, Tianzhu Zhang, Qirong Mao, and Changsheng Xu. Joint pose and expression modeling for facial expression recognition. In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[15] Eunji Chong, Nataniel Ruiz, Yongxin Wang, Yun Zhang, Agata Rozga, and James M. Rehg. Connecting gaze, scene, and attention: Generalized attention estimation via joint modeling of gaze and scene saliency. In Proceedings of European Conference on Computer Vision (ECCV), 2018.
[16] Karoui M.F. and Kuebler T. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2019. Pages 182197.
[17] LLICK, SATYA, Head Pose Estimation using OpenCV and Dlib, 2016.
[18] Kazemi,V & Sullivan,J. One milli second face alignment within ensemble of regression trees .In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 18671874 (2014).
[19] G. Pan, L. Sun, Z. Wu, S. Lao, Eyeblinkbased anti spoofing in face recognition from a generic web camera, in: Proceedings of IEEE International Conference on Computer Vision, 2007, pp. 18.
[20] Parmar, Singh Himani, Mehul Jajal, and Yadav Priyanka Brijbhan." Drowsy Driver Warning System Using Image Processing 1." (2014)
[21] B. Mandal, L. Li, G. S. Wang and J. Lin, "Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State," in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 3, pp. 545557, March 2017
[22] Simonyan, Karen and Andrew Zisserman. Very Deep Convolutional Networks for LargeScale Image Recognition. CoRR abs/1409.1556 (2014)
[23] Samiee, S., Azadi, S., Kazemi, R., Nahvi, A., & Eichberger, A.Data fusion to develop a driver drowsiness detection system with robustness to signal loss. Sensors. 14(9), 1783217847 (2014).