تحول بانکداری الکترونیک با استفاده از تکنولوژی‌های بیومتریک CNN-LSTM و بردهای رَسپبِری پای در تشخیص احساسات: رویکردی نوین برای امنیت و هوشمندسازی خدمات مالی

The Evolution of E-Banking Using Biometric Technologies, CNN-LSTM, and Raspberry Pi Boards in Emotion Recognition: A Novel Approach for Security and Smart Financial Services

وحید کبودی

عضو باشگاه پژوهشگران بانک سپه

گروه مهندسی کامپیوتر، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران

[Vahid.kabody@gmail.com](mailto:Vahid.kabody@gmail.com)

نگار طاهری

گروه مهندسی کامپیوتر، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران

[negartaheri@iaut.ac.ir](mailto:negartaheri@iaut.ac.ir)

چکیده:

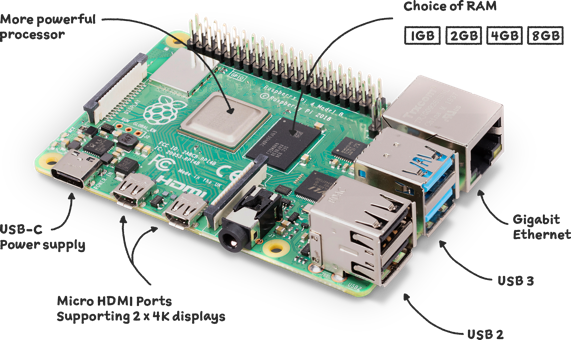
امروزه صنعت بانکداری نقش اساسی در اقتصاد خرد و کلان کشورها ایفا می نماید. توسعه‌ی بانکداری باز و گسترش بازارهای مالی متنوع شرایط جدیدی را در عرصه اقتصاد جهانی شکل داده است.صنعت بانکداري پيشرفته تأثير بسزايي بر اقتصاد و سياست دارد. امروزه بانکداري الکترونيک به يکي از خدمات مالي تبديل شده است که اکثر مردم به طور مکرر از آن استفاده مي­کنند. بنابراين با توجه به رشد اين صنعت، توجه بيشتر به مسائل امنيتي از اهميت بالايي برخوردار است. با استفاده از بردهای رَسپبِری پاي Raspberry Pi و الگوريتم­هاي بيومتريک در صنعت بانکداري شاهد ظهور کیوسک­ها و خودپردازهای نقدی - غیرنقدی هوشمند خواهيم بود. احراز هويت بيومتريک به عنوان يک رويکرد اميدوارکننده براي ايمن­سازي بانکداري الکترونيک مطرح شده است. روش­هاي بيومتريک مانند تشخيص چهره، احساسات، اثرانگشت و صوت برای تایید هویت افراد به کار می­روند. با استفاده از معماری یادگیری عمیق می­توان به شبکه عصبی پیچشی با حافظه کوتاه مدت بلند مدت که به اختصار CNN-LSTM نامیده می­شوند اشاره کرد که می­تواند به دقت بالایی دست یابد. پژوهش‌های قبلی در زمینه تشخیص احساسات انسان توسط حس­گرهایی که در بدن انسان قرار می­گرفت جمع آوری می­شدند پاسخ­های فیزیولوژیکی در انسان تحت تأثیر تغییرات احساسی است. در تشخیص خودکار احساسات بدون اتصال حس­گرها داده­های فیزیولوژیکی استخراج و می‌توان ضربان قلب انسان را از روی ویدیوهای چهره انسان تخمین زد. این تکنیک بر پایه استخراج سیگنال PPG با اندازه‌گیری تغییرات رنگ­های آبی، سبز و قرمز (RGB) چهره یک فرد است. در این راستا، یک چارچوب جدید مبتنی بر سیگنال‌های PPG بدون تماس برای طبقه‌بندی احساسات انسانی در مطالعه حاضر پیشنهاد شده است.

واژه‌های کلیدی: CNN-LSTM، تشخیص احساسات، تشخیص چهره، بانکداری هوشمند، سیگنال PPG

۱. مقدمه

بانک بخشي از صنعت و بخشي از اقتصاد است که به نگهداري دارايي­هاي مالي براي ديگران و سرمايه گذاري آن دارايي­هاي مالي به عنوان روشي اهرمی براي ايجاد ثروت بيشتر اختصاص دارد. صنعت بانکداري پيشرفته تأثير بسزايي بر اقتصاد و سياست دارد. امروزه بانکداري الکترونيکی (E-banking) به يکي از خدمات مالي تبديل شده است که اکثر مردم به طور مکرر از آن استفاده مي کنند. از سوي ديگر[1]، با ذکر اين نکته که 2.5 درصد از توليد ناخالص داخلي (GDP) ايران به صنعت بانکداري وابسته است و اين صنعت دومين ضريب نفوذ در منطقه MENA را به خود اختصاص داده است که 40 درصد از کل دارايي­هاي بانکداري اسلامي جهان را تشکيل مي­دهد. بنابراين با توجه به رشد اين صنعت، توجه بيشتر به مسائل امنيتي از اهميت بالايي برخوردار است. همچنين، همان طور که وانگ و همکاران[2] بحث کرده­اند، مسائل امنيتي مختلف به سرعت در حال افزايش است. با توسعه خدمات و عمليات بانکي، احتمال سوءاستفاده و کلاهبرداری به تدريج در حال افزايش است. با وجود اقدامات پيشگيرانه و امنيتي براي کاهش سوءاستفاده مالي، کلاهبرداران در حال يادگيري و يافتن راه هاي جديدي براي فرار از سيستم هاي پيشگيرانه و شناسايي هستند[3]. بنابراين، روش­هايي براي شناسايي سريع تراکنش­هاي مشکوک و جلوگيري از عواقب جبران ناپذير آنها مورد نياز است. بخش عمده­اي از فعاليت­هاي متقلبانه مربوط به تراکنش با کارت­هاي اعتباري است. بر اين اساس، ايجاد سيستمي که بتواند بر عملکرد سيستم­هاي پرداخت نظارت کند و براي کشف تقلب در تراکنش­ها از طريق کارت­هاي اعتباري ضروري به نظر مي رسد[4].

رَسپي‌بري پاي يک کامپيوتر تک‌بردي است که توسط بنياد رَسپي‌بري پاي در همکاري با شرکت Broadcom توسعه يافته است. اين برد، که از نظر اندازه بسيار کوچک است، با استفاده از يک زبان برنامه‌نويسي ساده قابليت توسعه پروژه‌هاي مختلف را فراهم مي‌آورد. رَسپي‌بري پاي داراي راه‌اندازي سخت‌افزاري ساده‌اي است و مي‌تواند به‌عنوان يک کامپيوتر کوچک عمل کند. اين برد در حوزه‌هاي مختلفي کاربرد دارد که پردازش تصوير يکي از اين حوزه‌ها است و کاربرد زيادي در توسعه برنامه‌هاي مختلف دارد. پردازش تصوير، يکي از انواع واحدهاي پردازش سيگنال است که در آن ورودي، يک تصوير مانند عکس يا فريم ويدئو است و نوعي پردازش بر روي تصوير انجام مي‌شود تا خروجي توليد گردد. همانطور که در شکل 1 نشان داده شده است[5].

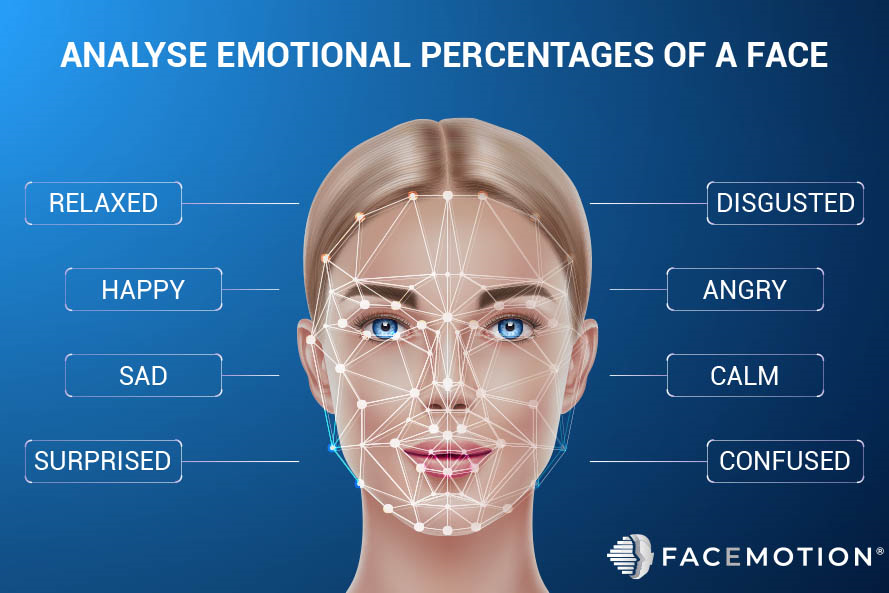


*شکل 1. برد رَسپي‌بري پاي 4*

بيومتريک يکي از برجسته‌ترين روش‌هاي حفاظت از داده‌هاي شخصي است و احراز هويت يک شخص عمدتاً بر اساس تأييد بيومتريک انجام مي‌شود که در برنامه‌هاي مختلف مانند حمل و نقل هوايي، بانکي، تراکنش‌هاي مالي و غيره، به طور فزاينده‌اي محبوب مي‌شود. بنابراين، اين پروژه طراحي يک سيستم بانکي با استفاده از بردهاي رَسپي‌بري پاي و فناوري‌هاي بيومتريک را پيشنهاد مي‌کند. با استفاده از اين سيستم، مي‌توان ابزارهاي تقلبي براي سرقت اطلاعات اعتباري مهم مانند رمز عبور و شماره حساب را از بين برد.

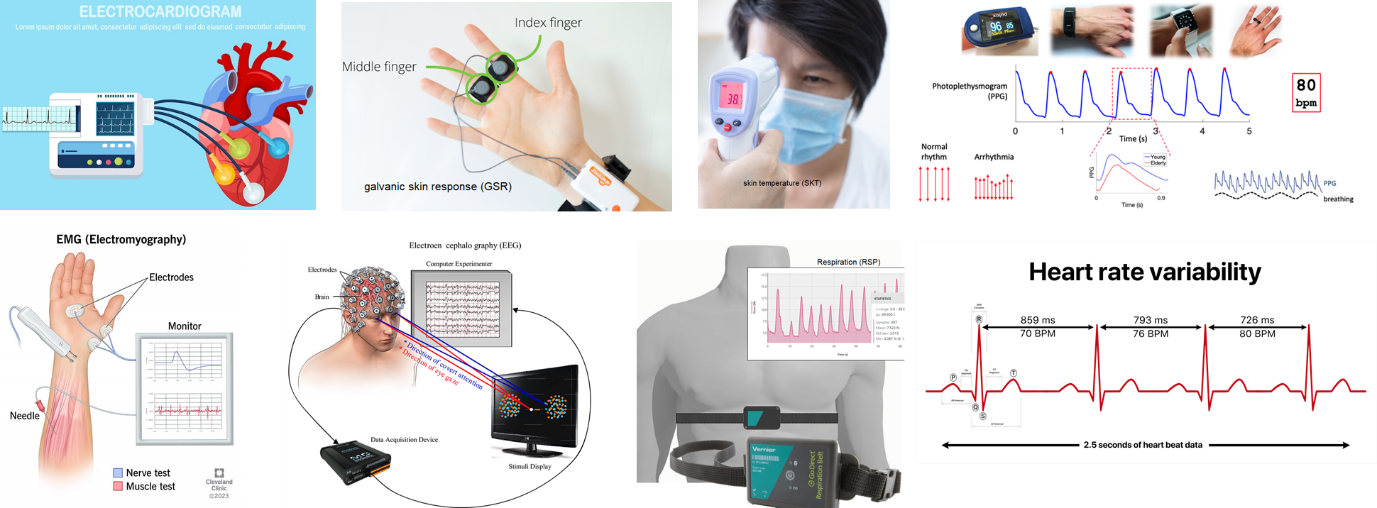
تشخیص احساسات یک حوزه تحقیقاتی نسبتاً جدید و مهم است. این فرآیند شناسایی احساسات انسان از طریق شناخت وضعیت روانی و عاطفی آنهاست. احساساتی که افراد می­توانند در زندگی روزمره خود داشته باشند می­توانند مثبت یا منفی باشند[6].

امروزه، تعداد زیادی از محققان به دنبال بهره‌برداری از پیشرفت‌های تکنولوژیکی هستند که در سراسر جهان در حال رخ دادن است تا از تشخیص احساسات در ماشین‌ها و روبات‌ها برای دستیابی به تعامل طبیعی بین ماشین‌ها و انسان‌ها استفاده کنند. در این راستا مشخص شده است که وقتی ماشین‌ها وضعیت عاطفی انسان را بشناسند، می‌توانند خدمات عملی بیشتری را در زمینه‌های مختلف مانند بازاریابی[7]، آموزش [8]به‌ویژه در دوره همه‌گیری COVID 19 [9]به آنها ارائه دهند. در شکل 2 تمام احساسات یک چهره نشان داده شده است.



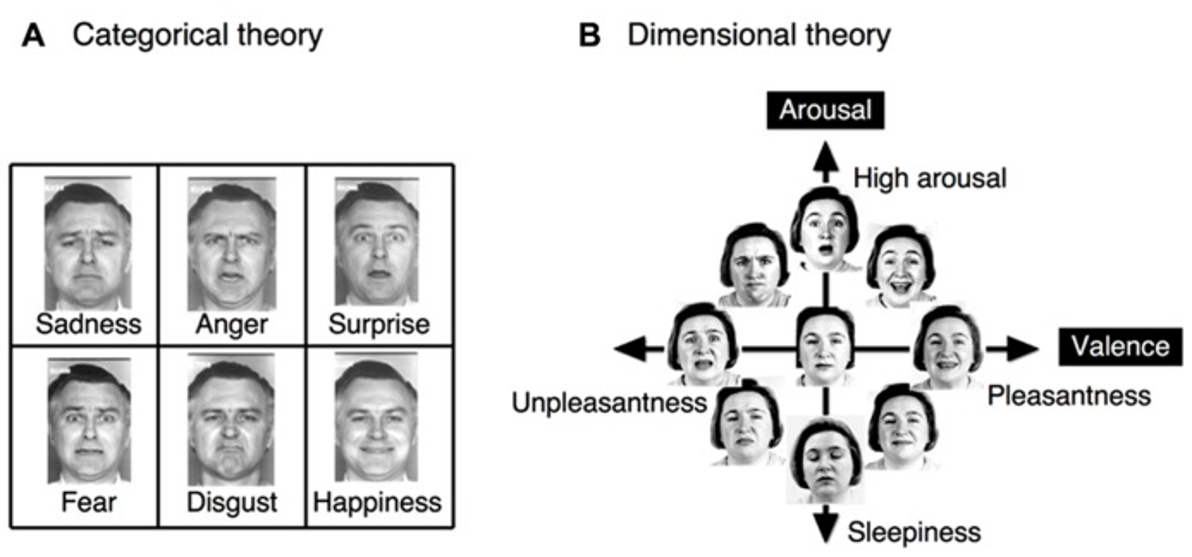
*شکل 2. تحلیل احساسات یک چهره*

بسیاری از محققان در این زمینه نشان داده اند که احساسات زمانی فعال می شوند که افراد در معرض موقعیت­های خاص و غیرعادی قرار گیرند. این احساسات عموماً از طریق تغییرات فیزیکی مانند حالات چهره، حالت‌ها و لحن صدا بیان می‌شوند[10]. آنها همچنین می­توانند از طریق تغییرات فیزیولوژیکی که با استفاده از ابزارهای تشخیصی بالینی مانند الکتروکاردیوگرام (ECG)، الکترومیوگرام (EMG)، الکتروانسفالوگرام (EEG)، پاسخ گالوانیکی پوست (GSR)، تنفس (RSP)، دمای پوست (SKT)، فتوپلتیسموگرام (PPG) و تغییرپذیری ضربان قلب (HRV) که می­تواند از سیگنال­های ECG و PPG استخراج شود [11]قابل تشخیص هستند. که در شکل 3 نشان داده شده است[12] [7].



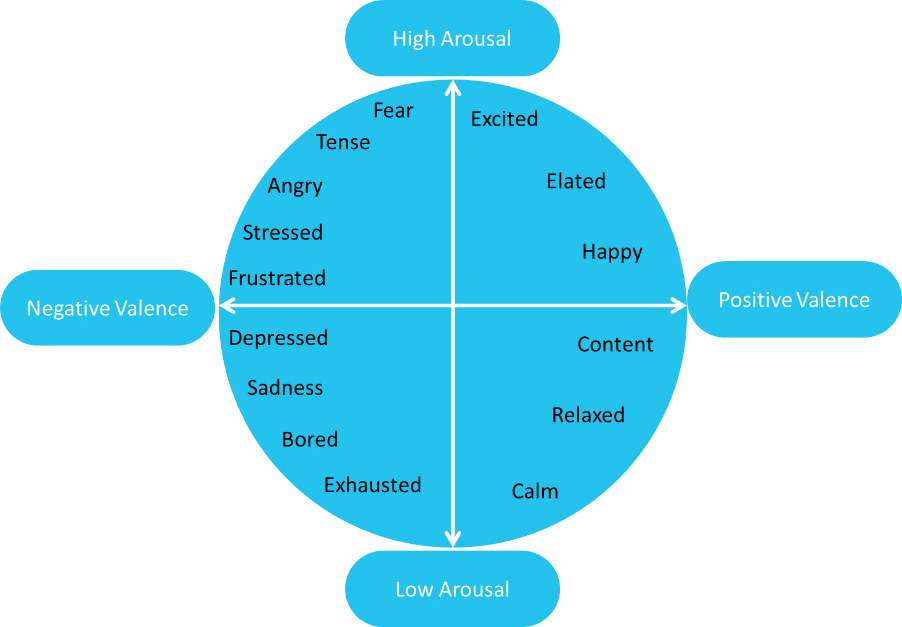
*شکل 3. ابزارهای ارزیابی فعالیت‌های فیزیولوژیکی بدن*

با توجه به مقالات، عواطف عموماً بر اساس دو رویکرد تقسیم می‌شوند: رویکرد مقوله‌ای و رویکرد بعدی ( شکل 4) [10]. رویکرد مقوله‌ای به دنبال شناسایی، جداسازی و طبقه‌بندی احساسات اساسی به منظور ساده‌سازی آنها می‌پردازد. در این راستا، بسیاری از محققان روان‌شناسی، مانند ایزارد، پلوچیک و بسیاری دیگر، هر یک به شیوه‌ای به بررسی و طبقه‌بندی احساسات پرداخته‌اند. اکمن[10] نیز احساسات را در شش دسته مختلف طبقه بندی کرد که بیشتر در حوزه تشخیص خودکار هیجان استفاده می­شود: خشم، تعجب، انزجار، شادی، ترس و اندوه[13].



*شکل 4. دو نظریه رویکرد مقوله‌ای و رویکرد بعدی در مورد ادراک احساسات از چهره*

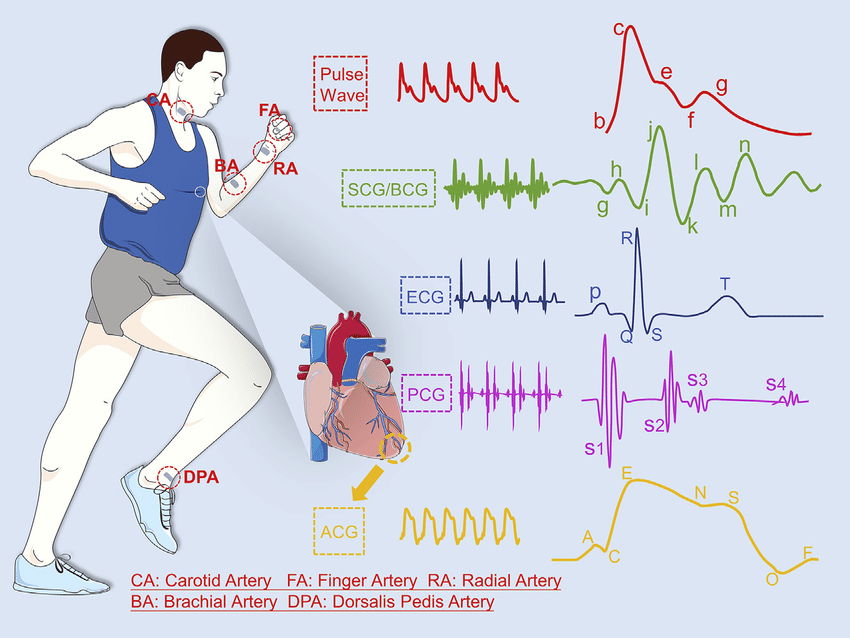
رویکرد ابعادی امکان طبقه‌بندی و نمایش گرافیکی احساسات را بر روی محورهای ظرفیت و برانگیختگی فراهم می‌کند. محور ظرفیت برای طبقه بندی حالت هیجانی مثبت یا منفی طراحی شده است در حالی که محور برانگیختگی برای احساسات هیجان زده یا بی تفاوت است. شایان ذکر است که ظرفیت از منفی به مثبت گسترش می­بابد و به کیفیت مطلوب یا نامطلوب یک محرک اشاره دارد، در حالیکه برانگیختگی که از کسل‌کننده تا تحریک‌کننده متغیر است و مربوط به شدت یک محرک است. شکل 5 توزیع احساسات را در امتداد دو محور ظرفیت و برانگیختگی نشان می دهد.



*شکل 5. توزیع احساسات در محورهای ظرفیت (valence) و برانگیختگی (arousal)*

علاوه بر این، باید توجه داشت که تغییرات فیزیکی همیشه وضعیت روانی دقیق یک فرد را بیان نمی­کند، زیرا انسان می­تواند احساسات یا عواطف واقعی خود را در زمانی که واقعاً غمگین یا واقعاً خوشحال است پنهان کند. حتی ممکن است حالات چهره واقعی خود را پنهان کنند. به عنوان مثال، فردی می­تواند در یک رویداد خانوادگی لبخند بزند، اما در واقع احساس غم و اندوه عمیقی می­کند. شایان ذکر است که سیگنال­های فیزیولوژیکی می­توانند به طور خود به خود در طول احساسات فعال شوند و به راحتی قابل کنترل نیستند.

مشخص شد که سیگنال­های فیزیولوژیکی یک فرد مطمئناً تحت تأثیر وضعیت عاطفی او قرار می­گیرد. این موضوع علاقه زیادی را در بین محققان برانگیخته است زیرا امکان جمع آوری و استفاده از اطلاعات واقعی و دقیق احساسی را فراهم می­کند. در واقع بسیاری از مطالعات تحقیقاتی در این زمینه با استفاده از روش‌های یادگیری ماشین سنتی (ML) مانند K نزدیکترین همسایه (KNN)[14]، تجزیه و تحلیل تشخیص خطی (LDA)[15]، ماشین بردار پشتیبان [16](SVM) و جنگل تصادفی RF [17]انجام شده است. با پیشرفت­های قابل توجه در تکنیک­های یادگیری عمیق (یادگیری عمیق) در داده­های متوالی، تحقیقات اخیر نشان داده است که استخراج بیشتر ویژگی­های احساسات و طبقه­بندی خودکار آنها امکان­پذیر است. در واقع، تعداد زیادی از محققان این تکنیک­ها را برای طبقه بندی احساسات بر اساس سیگنال­های فیزیولوژیکی اتخاذ کرده اند. (شکل 6) با استفاده از این روش ها می­توان به نتایج موفق تری نسبت به روش­های سنتی دست یافت.[18, 19]

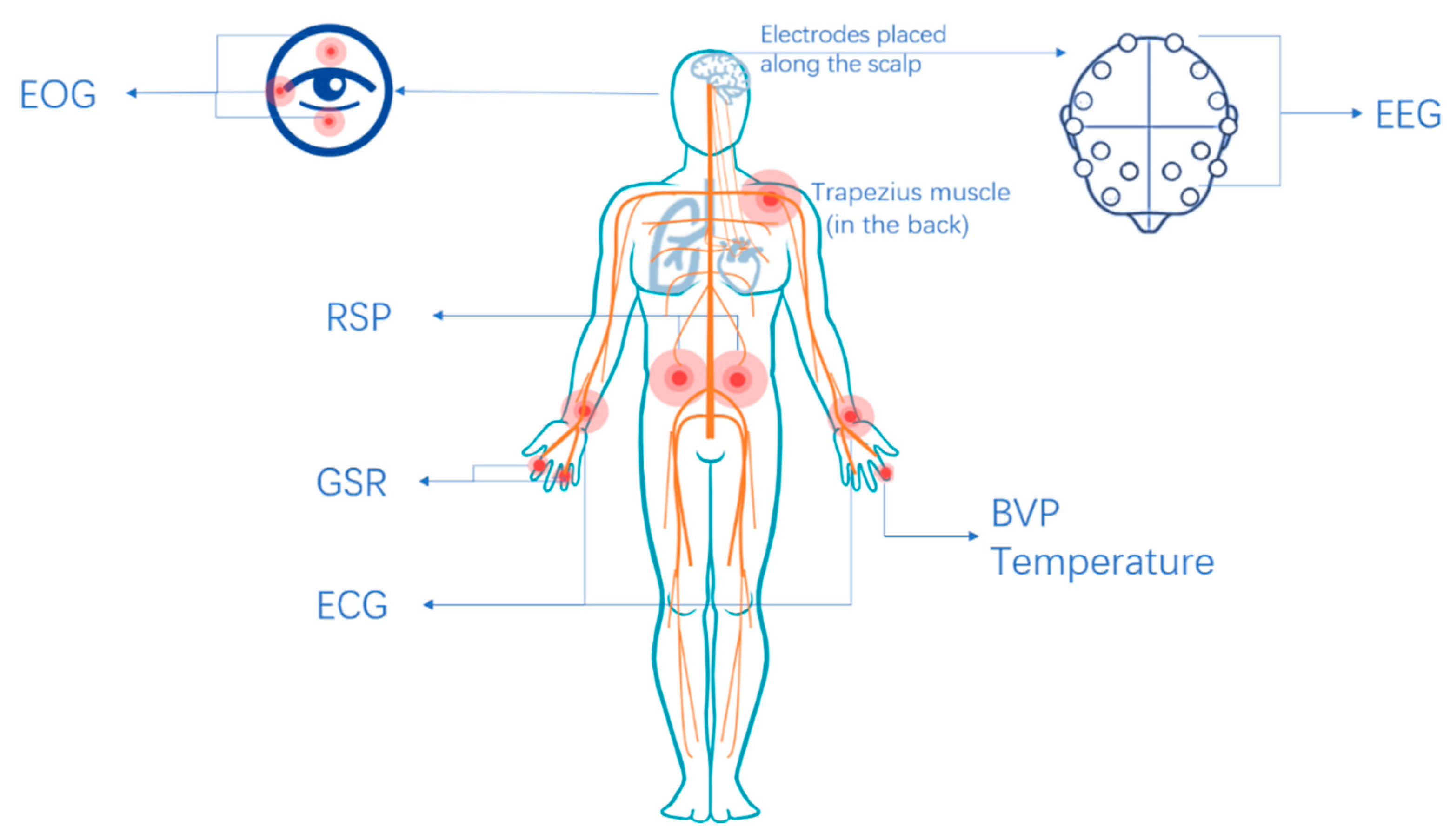


*شکل 6. شماتیک سیگنال‌های فیزیولوژیکی با استفاده از ابزارهای ارزیابی فعالیت‌های فیزیولوژیکی بدن*

چندین محقق نشان داده­اند که PPG یکی از سیگنال­های فیزیولوژیکی است که می­تواند اطلاعات زیادی در مورد وضعیت عاطفی یک فرد ارائه دهد. این یک تکنیک نوری ساده است که می­تواند به تشخیص تغییرات حجم خون در سیستم عروقی و همچنین کنترل ضربان قلب کمک کند. روش‌های سنتی معمولاً از سنسورهای تماسی برای اندازه‌گیری این سیگنال استفاده می‌کردند. با این حال، در سال‌های اخیر محققان برای استخراج سیگنال PPG از ویدیوهای صورت با اندازه‌گیری تغییراتی که در رنگ پوست فرد در طول چرخه قلبی رخ می‌دهد، تکیه کرده‌اند.

تکنیک ذکر شده در بالا بر اساس پنج مرحله ضروری است، یعنی تشخیص چهره، شناسایی منطقه مورد نظر (ROI)، استخراج سیگنال های RGB، استفاده از یک الگوریتم استخراج سیگنال خام PPG و در نهایت، تخمین ضربان قلب (HR)

شایان ذکر است که اخیراً روش غیر تماسی برای تخمین و پایش ضربان قلب (HR) [20]توجه تعداد زیادی از محققان را به خود جلب کرده است. در این حالت، شخص مجبور نیست از حس­گر روی بدن خود استفاده کند. ( شکل 7 ) این تکنیک ساده است و مزایای زیادی دارد. به عنوان مثال می­توان آن را به طور خاص برای افرادی که در طول همه­گیری COVID-19 به ویروس کرونا مبتلا شده­اند استفاده کرد تا از هر گونه آلودگی از طریق تماس مستقیم جلوگیری شود.

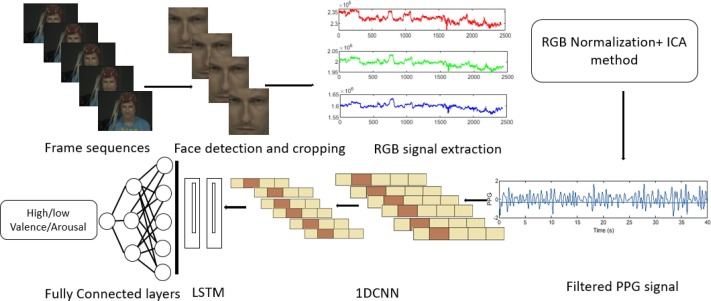


*شکل 7. موقعیت حس­گرهای ارزیابی فعالیت‌های فیزیولوژیکی بدن*

در سال 2008، ورکرویس و همکاران [21]پیشنهاد کردند که می­توان از تکنیک PPG برای تخمین ضربان قلب انسان (HR) از یک ویدیوی صورت استفاده کرد. آنها نشان دادند که کانال سبز برای سیگنال­های BPV در باند رنگی RGB بهتر است. لازم به ذکر است که در این روش از سیگنال سبز رنگی استفاده می­شود که از ROI استخراج می­شود. پس از آن، چندین روش و تکنیک به منظور تشخیص دقیق ضربان قلب انسان (HR) و از این رو غلبه بر چندین چالش مانند نورهای متغیر، نوع دوربین مورد استفاده و حرکات فرد ایجاد شد.

در حال حاضر پذیرفته شده است که تشخیص خودکار احساسات و استخراج ضربان قلب از راه دور دو زمینه تحقیقاتی مجزا هستند که بسیار جالب هستند. تحقیقات نشان داده است که احساسات می­توانند تا حد زیادی بر واکنش­های قلبی عروقی تأثیر بگذارند. براساس یافته‌های فوق، تصمیم بر آن شد تا با استفاده از فضای هیجانی دوبعدی، مطالعه‌ای در زمینه بازشناسی عواطف انسانی انجام شود. در این حالت داده­ها در امتداد محورهای ظرفیت و برانگیختگی توزیع می­شوند. برای انجام این کار، تصمیم گرفته شد از سیگنال های PPG بدون تماس استفاده شود. سپس از مدل یادگیری عمیق (یادگیری عمیق) برای طبقه‌بندی این سیگنال‌ها استفاده شد.

برای رسیدن به اهدف­مان چندین مرحله طی شد. در ابتدا، ما یک پایگاه داده عاطفی را انتخاب کردیم که احساسات واقعی را نشان می­دهد. برای این منظور، استفاده از پایگاه داده MAHNOB-HCI که اغلب توسط محققان به منظور اعتبارسنجی الگوریتم‌های استخراج ضربان قلب غیر تماسی استفاده می‌شود، مناسب ارزیابی شد. دوم، یک روش استخراج سیگنال PPG کارآمد در پایگاه داده MAHNOB-HCI [22]اعمال شد. توجه داشته باشید که سیگنال­های PPG ثبت شده با استفاده از تکنیک­های یادگیری عمیق سنتی (یادگیری عمیق) طبقه بندی شدند. شکل 8 یک نمای کلی از روش های مورد استفاده در مطالعه را نشان می دهد.



*شکل 8. نمای کلی از روش پیشنهادی*

روش های استخراج از راه دور PPG و HR

از سال 2008، محققان رویکرد جدیدی را اتخاذ کرده‌اند که حول توسعه روش‌های جدید برای ثبت سیگنال‌های فیزیولوژیکی بدون تماس با بدن انسان، با استفاده از دوربین‌های ساده و کم‌هزینه است. این سیگنال‌های فیزیولوژیکی ممکن است با استفاده از حسگرهای فوتوپلتیسموگرافی که می‌توانند تغییرات حجمی خون را که در طول چرخه قلبی رخ می‌دهند، ارزیابی کنند.

همانطور که پوه و همکاران [23] یک روش تجزیه و تحلیل مؤلفه مستقل (ICA) را توسعه دادند که یک تکنیک یادگیری ماشینی است که اجازه می­دهد منابع مستقل را از یک سیگنال مختلط جدا کنید. هدف اصلی این روش تجزیه سیگنال­های RGB خام به اجزای مستقل است. این روش در از بین بردن نویز ایجاد شده توسط حرکات سر کاملاً مؤثر است. علاوه بر این، هان و جین [24] روش CHROM را پیشنهاد کردند که امکان حذف نویز موجود در سیگنال فوتوپلتیسموگرافی (PPG) را فراهم می کند. هدف این روش در درجه اول از بین بردن مولفه‌ی مشکوک است که در واقع مسئول حضور رنگ روشن در سیگنال PPG است. در همین زمینه، وانگ و همکاران [25]روش صفحه متعامد به پوست (POS) را توسعه داد که سطح متعامد به رنگ پوست را در فضای معمولی RGB برای استخراج سیگنال PPG تعریف و استفاده می‌کند.

در چند سال گذشته، روش‌ها و تکنیک‌های مختلفی برای توسعه روش‌های سنتی استخراج منابع انسانی پیشنهاد شده‌اند. در این رابطه، سونگ و همکاران [26] شبکه PulseGAN را برای کاهش یا حذف نویز از سیگنال‌های PPG خام به‌دست‌آمده با روش CHROM پیشنهاد کردند. این رویکرد با استفاده از پایگاه داده UBFC-RPPG پیاده سازی شد و نتایج به دست آمده امیدوار کننده­تر از نتایج گزارش شده توسط سایر محققان بود. علاوه بر این، به منظور دستیابی به تشخیص بهتر صورت و غلبه بر مشکلات حرکت سر ، Subramaniam K. Rajitha [27]از تکنیک­های ردیاب ویژگی سریع­تر RCNN و Kanade-Lukas-Tomasi (KLT) استفاده کرد. با این حال، آنها از فیلتر حداقل مربعات بازگشتی (RLS) برای کاهش نویز ناشی از روشنایی استفاده کردند. بعداً از روش ICA برای تعیین HR استفاده شد.

شایان ذکر است که انتخاب منطقه مناسب برای استخراج سیگنال RGB یک مرحله بسیار مهم است. در این رابطه، لامبا و ویرمانی [28] یک مطالعه مقایسه­ای بین دو ناحیه صورت، یعنی فقط گونه­ها و گونه­ها با بینی انجام دادند.

علاوه بر این، غنادیان و همکاران [29] به منظور بهبود روش ICA [23] یک تکنیک یادگیری ماشین را برای شناسایی بهترین جزء اطلاعاتی سیگنال PPG پس از اعمال تکنیک ICA اتخاذ کردند. همچنین، کوریهارا و همکاران [30]روش جدیدی را توسعه و پیشنهاد کردند که مبتنی بر ادغام سیگنال‌های RGB و سیگنال‌های مادون قرمز نزدیک (NIR) برای به دست آوردن ویژگی‌های بیشتر تحت تغییرات روشنایی است. ژنگ و همکاران [31] الگوریتمی را توسعه دادند که می­تواند ضربان قلب را از بخشی از ناحیه برجسته صورت محاسبه کند تا مشکل مربوط به عدم تشخیص کل صورت به دلیل حرکت سر یا ماسک­های انسداد صورت را حل کند. مطالعه آنها بر روی هشت ناحیه صورت با استفاده از روش ICA انجام شد.

تشخیص احساسات از داده های فیزیولوژیکی بدون تماس

علیرغم تمام پیشرفت‌هایی که در زمینه‌های PPG بدون تماس و استخراج منابع انسانی غیر تماسی انجام شده است، تعداد کمی از آثار از این سیگنال‌ها در کاربردهایی مانند تشخیص احساسات انسانی استفاده کرده‌اند. تعدادی از محققان از دوربین­های مادون قرمز حرارتی برای استخراج سیگنال­های فیزیولوژیکی مانند ضربان تنفس و تغییرات ضربان قلب برای دستیابی به طبقه بندی بهتر احساسات استفاده کردند[32] . با توجه به تمام پیشرفت‌هایی که تاکنون در زمینه استخراج فیزیولوژیکی بدون تماس حاصل شده است، با استفاده از دوربین‌های ساده و ارزان، می‌توان گفت که همچنان می‌توان از همین تجهیزات در طبقه‌بندی خودکار احساسات انسانی استفاده کرد.

در این رابطه، کسلر و همکاران [33] از سیگنال‌های خام PPG برای طبقه‌بندی درد در چهار سطح استفاده کردند. آنها با استفاده از طبقه‌بندی‌کننده جنگل تصادفی، که یک الگوریتم یادگیری ماشینی بسیار محبوب است، به دقت 65.79% دست یافتند. همان‌طور که برای [34]، آنها از یک CNN سه‌بعدی برای استخراج سیگنال PPG دقیق، با همبستگی پیرسون 0.8، از ویدئوهای پایگاه داده BioVid استفاده کردند. نویسندگان از معماری DCNN1 استفاده کردند و دقت طبقه‌بندی 58.92٪ را به دست آوردند.

در مورد مائوئی و همکاران [35]، آنها از طبقه‌بندی‌کننده‌های SVM و LDA برای طبقه‌بندی سطح استرس با استفاده از سیگنال‌های PPG بدون تماس استفاده کردند. اخیراً بنزث و همکاران [36]یک مطالعه مقایسه‌ای بین تغییرات ضربان قلب بدون تماس (HR) و حالات عاطفی انجام دادند. برای این کار، از روش CHROM برای استخراج دقیق سیگنال PPG استفاده کردند. با توجه به ژنگ و همکاران [31]، آنها تأیید کردند که روش تخمین HR غیر تماسی آنها می‌تواند اطلاعات احساسی بسیار جالبی را با تجسم تغییرات HR که در طول تحریک عاطفی شرکت‌کنندگان رخ می‌دهد، ارائه دهد.

علاوه بر این، در سال 2021، مزیاتی صبور و همکاران [37] یک پایگاه داده جدید، یعنی پایگاه داده UFBC-Phys، ایجاد کردند که برای طبقه‌بندی استرس انسان در نظر گرفته شده بود. این پایگاه داده حاوی سیگنال‌های EDA و PPG است که توسط سنسورها گرفته می‌شود. این محققان نشان دادند که روش POS می‌تواند عملکرد بالایی در اندازه‌گیری سیگنال PPG بدون تماس داشته باشد. علاوه بر این، چندین محقق از طبقه‌بندی‌کننده ماشین بردار پشتیبان (SVM) برای طبقه‌بندی احساسات بر اساس تغییرپذیری ضربان قلب (HRV) استفاده کرده‌اند و نرخ شناخت 85.48٪ را به دست آوردند.

برای اولین بار، اوزار و همکاران [38]با در نظر گرفتن ادغام حالات چهره و سیگنال PPG غیر تماسی و تغییرپذیری HR، یک مدل تشخیص احساسات چندوجهی را پیشنهاد کردند که بر اساس کلیپ‌های ویدیویی چهره جلویی است. محققان در این کار از یک شبکه توجه کانولوشنال تغییر موقتی چندوظیفه‌ای (MTTS-CAN) برای استخراج سیگنال‌های PPG بدون تماس از پایگاه داده احساسات خود به خودی چندوجهی (BP4D+) استفاده کردند. آنها به ترتیب با استفاده از تکنیک‌های PPG و HRV به دقت 55.33٪ و 53.59٪ رسیدند. با توجه به لَمپیِر و همکاران [39]، آنها با استفاده از تشخیص غیر تماسی ویژگی‌ها از سیگنال‌های PPG، به نرخ دقت 42 درصدی در مقیاس ظرفیت دست یافتند.

در این مقاله، مطالعه‌ای با هدف طبقه‌بندی هیجان‌های ظرفیت-برانگیختگی بر اساس سیگنال‌های PPG بدون تماس انجام شد. برای این کار، برخی از تکنیک‌های اخیر یادگیری عمیق (یادگیری عمیق) برای طبقه‌بندی احساسات استفاده شد، زیرا در این زمینه بسیار موفق بودند. ما معتقدیم که این مطالعه کاملاً جامع است، زیرا تغییرات فیزیولوژیکی از یک سو می‌توانند ویژگی‌های واقعی را برای حالت‌های عاطفی فراهم کنند و از سوی دیگر، احتیاجی به اتصال حس­گرهایی که بر روی بدن قرار می‌دهند نیست.

جمع‌آوری سیگنال PPG بدون تماس

پایگاه داده MAHNOB-HCI برای مطالعه‌ی حاضر استفاده شد[22]. در واقع، این یک پایگاه داده برای تشخیص احساسات است؛ با این حال، بیشتر محققان از آن برای ارزیابی روش‌های استخراج سیگنال‌های انسانی غیرتماسی خود استفاده می‌کنند[17, 35, 40] .

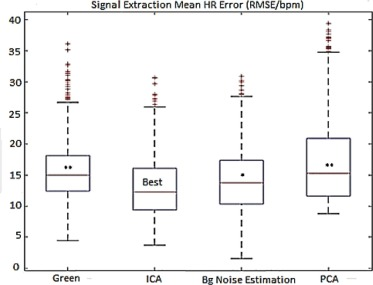
به طور گسترده‌ای پذیرفته شده است که MAHNOB-HCI یک پایگاه داده چندوجهی (شامل داده‌های صوتی، تصویری و فیزیولوژیکی) است که شامل 30 شرکت‌کننده از جنسیت‌ها و قومیت‌های مختلف است. در این حالت، هر یک از شرکت‌کنندگان برای برانگیختن احساسات خود، تعدادی ویدیوی انتخابی را با دقت تماشا می‌کنند و در این حین، از آنها فیلم گرفته می‌شود تا عکس‌العمل‌های آنها ضبط و نظارت شود. شرکت‌کنندگان برای گرفتن سیگنال‌های فیزیولوژیکی مانند سیگنال‌های GSR، EEG، ECG، SKT و RSP به حس­گرهایی متصل هستند. سپس فیلم‌های ضبط‌شده صورت با وضوح (780\*580) پیکسل و 61 فریم در ثانیه فشرده می‌شوند.

در پایان هر آزمون، از شرکت‌کنندگان خواسته می‌شود تا با استفاده از دو کلاس «ظرفیت» و «برانگیختگی»، بازخورد خود را در مورد احساساتشان در مقیاس 1 تا 9 ارائه دهند. در این مقیاس، 1 به کمترین تحریک و منفی‌ترین ظرفیت و 9 به بالاترین تحریک و مثبت‌ترین ظرفیت اشاره دارد.

برای انجام کارآمد این مطالعه، ایجاد یک پایگاه داده جدید که از سیگنال‌های PPG بدون تماس استخراج‌شده از ویدیوهای RGB صورت تشکیل شده باشد، ضروری تشخیص داده شد. برای دستیابی به این هدف، ما به مطالعات مقایسه‌ای که قبلاً انجام شده است، برای به‌دست‌آوردن یک روش کارآمد در پایگاه داده MAHNOB-HCI تکیه می‌کنیم.

از این منظر، Boccignone و همکاران [17]یک مطالعه مقایسه‌ای انجام دادند و دریافتند که وقتی روش‌های ICA و PCA در پایگاه داده MAHNOB-HCI استفاده می‌شوند، می‌توانند به طور قابل‌توجهی کارآمدتر از روش‌های Green، CHROM، POS، SSR، LGI و PBV باشند. به همین ترتیب، وانگ و همکاران مطالعه تطبیقی سه روش ICA، PCA و Green را بر روی پایگاه داده MAHNOB-HCI انجام دادند. مشخص شد که روش ICA که بر روی کل صورت یا صورت بدون دهان و چشم اعمال می‌شود، نتایج بهتری نسبت به سایر نواحی صورت مانند پیشانی، چانه و گونه‌ها به‌دست می‌آورد. با توجه به موارد فوق، تصمیم بر این شد که از روش ICA برای استخراج سیگنال‌های PPG استفاده شود.[40]

شکل 9 ارزیابی عملکرد هر روش استخراج HR غیر تماسی را در پایگاه داده MAHNOB-HCI ارائه می‌دهد.



*شکل 9. ریشه میانگین مربع خطا (RMSE) به‌دست‌آمده توسط Wang و همکاران*

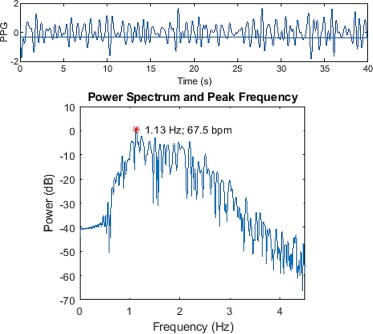
لازم به ذکر است که چندین مرحله برای ایجاد پایگاه داده جدید ما که شامل سیگنال‌های PPG بدون تماس است دنبال شد. ابتدا، ده شرکت‌کننده از پایگاه داده MAHNOB-HCI بارگذاری شد. هر سوژه شامل 20 ویدیوی RGB از چهره جلویی است. سپس، تکنیک ICA بر روی توالی‌های ویدیویی انتخابی با طول 40 ثانیه اعمال شد. پس از آن، نرخ HR به‌دست‌آمده با استفاده از روش ICA با نرخ محاسبه‌شده از سیگنال ECG موجود در پایگاه داده MAHNOB-HCI مقایسه شد. شایان ذکر است که تنها سیگنال‌های PPG که HR دقیقی را ارائه می‌دهند، جمع‌آوری شده‌اند.

بخش زیر فرآیند استخراج سیگنال‌های PPG بدون تماس را ارائه می‌دهد. سپس روش مورد استفاده با محاسبه HR و مقایسه آن با HR محاسبه‌شده از سیگنال ECG تأیید شد.

روش استخراج سیگنال PPG بدون تماس

به منظور دستیابی به هدف خود در طبقه‌بندی احساسات انسانی با استفاده از سیگنال‌های PPG از راه دور، لازم بود ابتدا یک روش کارآمد برای استخراج این نوع سیگنال ایجاد کنیم. این امر مطمئناً به دستیابی به ضربان قلب دقیق (HR) کمک می‌کند. جستجو در مطالعات نشان داد که مرحله تشخیص ناحیه صورت نقش مهمی در استخراج سیگنال‌های حامل اطلاعات قلبی دارد. در آزمایشی که در این مطالعه انجام شد، الگوریتم Viola-Jones به منظور تشخیص چهره در فریم‌های توالی ویدیو به کار رفت و به دنبال آن، از تکنیک برش برای حذف پس‌زمینه و حفظ تنها چهره استفاده شد. پس از آن، پوست قطعه‌بندی شد و سیگنال‌های RGB با استفاده از مقدار پیکسل پوست میانه (RGB) استخراج شدند. سپس، آن‌ها برش داده شده و نرمالیزه شدند.

علاوه بر این، روش تجزیه و تحلیل مؤلفه‌های مستقل (ICA) با استفاده از تکنیک JADE (تقریبی ماتریس‌های ویژه مشترک) برای به‌دست‌آوردن سیگنال‌های مستقل از سیگنال‌های RGB نرمال‌شده استفاده شد [41]. همچنین توجه داشته باشید که یک فیلتر تبدیل فوریه سریع (FFT) برای انتخاب سیگنال PPG که بالاترین انرژی را حمل می‌کند، به کار رفته است. سپس، این سیگنال PPG در محدوده [0.7-2.5] هرتز، که معادل [40-150] bpm است، با استفاده از فیلتر باند گذر مرتبه سوم Butterworth فیلتر شد. بیشترین فرکانسی که با چگالی طیفی توان (PSD) به‌دست آمد، ضربان قلب (HR) در نظر گرفته می‌شود [41]. شکل 10 سیگنال PPG و تخمین HR به‌دست‌آمده از سیگنال چگالی طیفی توان را نشان می‌دهد.



*شکل 10. سیگنال PPG بدون تماس و ضربان قلب (HR) تخمینی.*

ارزیابی سیگنال PPG بدون تماس

قبل از استفاده از سیگنال‌های PPG استخراج‌شده از فیلم‌ها، ابتدا باید آنها را ارزیابی کرد. ضربان قلب تخمین‌زده‌شده با استفاده از روش PPG بدون تماس محاسبه می‌شود و سپس با ضربان قلب واقعی که از نوار قلب (ECG) به دست می‌آید، مقایسه می‌شود.

در پایگاه داده MAHNOB-HCI، سیگنال‌های ECG از قسمت‌های مختلف بدن انسان گرفته شده است. در مطالعه ما، از داده‌های کانال EXG2 مربوط به گوشه سمت چپ بالای قفسه سینه، در زیر استخوان ترقوه، استفاده شده است. ( شکل 11 )



LA

*شکل 11. محل قرارگیری حسگر EXG2 در زیر استخوان ترقوه برای جمع‌آوری سیگنال‌های فیزیولوژیکی*

شایان ذکر است که سیگنال های ECG در فرمت BDF با فرکانس 256 هرتز در دسترس هستند. سپس از دو روش مختلف برای محاسبه استفاده شد. روش اول از تشخیص اوج برای محاسبه میانگین اختلاف زمانی بین دو پیک استفاده کرد در حالی که روش دوم از جعبه ابزار BioSPPy که به طور گسترده در دسترس است استفاده کرد [42]. مشخص شد که هر دو روش نتایج مشابهی را ارائه می دهند.

به منظور ارزیابی نتایج به‌دست‌آمده، تصمیم گرفته شد که سه پارامتر، یعنی میانگین خطای مطلق (MAE)، ریشه میانگین مربع خطا (RMSE) و نسبت سیگنال به نویز (SNR) به عنوان تابعی از و محاسبه شود. این سه پارامتر به صورت زیر محاسبه شد:

تعداد ویدیوها است. SNR با استفاده از روش Haan et Jeanne [32] محاسبه شد. این مرحله از مطالعه با استفاده از کد متلب موجود در GitHub که توسط McDufft و Blackford ارائه شده است [41]اجرا شد. علاوه بر این، تعداد 150 سیگنال بسیار دقیق از 193 ویدیو استخراج شد. این سیگنال‌ها ضربان قلب دقیق (HR) را با مقادیر متوسط برای MAE = 1.55bpm، RMSE = 2.18bpm و SNR = - 3.70dB ارائه می‌دهند.

طبقه بندی احساسات با استفاده از تکنیک های یادگیری عمیق

شایان ذکر است که هدف اصلی این مطالعه طبقه بندی احساسات انسانی با استفاده از سیگنال های PPG بدون تماس است. تکنیک­های یادگیری عمیق برای انجام طبقه بندی به کار گرفته شد. کلاس­های احساسات، مراحل پیش پردازش PPG و همچنین معماری یادگیری عمیق پیشنهادی، همگی در این بخش ارائه شده­اند. شکل 12 نمای کلی از مراحل دنبال شده در مطالعه پیشنهادی را ارائه می­دهد.

*شکل 12. مروری بر مراحل دنبال شده در مطالعه پیشنهادی*

تعریف کلاس های احساسات و پیش پردازش سیگنال های PPG بدون تماس

در پایگاه داده MAHNOB-HCI، پس از هر آزمایش، از شرکت کنندگان خواسته شد تا حالات عاطفی گسسته خود را به دو روش مختلف ثبت کنند. ابتدا این کار را بر حسب برچسب‌های مجزا مانند شادی، سرگرمی، ترس، خشم، اضطراب، غم، ملال، تعجب و خنثی انجام دادند و سپس یک نمایش ابعادی از نظر ظرفیت و برانگیختگی شکل 13 با توزیع در مقیاس‌های مختلف از 1 تا 9 ارائه کردند.

*شکل 13.* *حالات عاطفی گسسته با برچسب نمایش ابعادی از نظر ظرفیت و برانگیختگی*

در مقیاس عاطفی شکل 14، دامنه از 1 تا 4 به ظرفیت منفی و برانگیختگی کم اشاره دارد و از 5 تا 9 مربوط به ظرفیت مثبت و برانگیختگی زیاد است [43].

*شکل 14. دامنه ابعادی از نظر ظرفیت و برانگیختگی حالات عاطفی گسسته*

سیگنال PPG فیلتر شده نشان داده شده در شکل 8 در آزمایش انجام شده در این کار مورد استفاده قرار گرفت. سپس نرمال سازی شد و قبل از وارد شدن به معماری DL به بخش هایی تقسیم شد.

علاوه بر این، لازم است که مرحله مقیاس‌بندی ویژگی را اعمال کنیم تا همه داده‌هایمان را به یک مقیاس برسانیم، زیرا مجموعه داده‌ها شامل سیگنال‌های PPG از افراد مختلف است. برای این منظور، از روش نرمال‌سازی داده‌ها به صورت زیر استفاده شد:

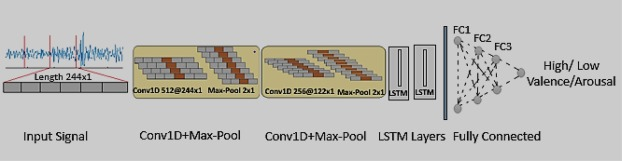
در اینجا X سیگنال PPG و سیگنال PPG عادی‌شده است. توجه داشته باشید که هدف اصلی ما حذف کوچک‌ترین مقدار از بردار سیگنال و سپس تقسیم آن بر وسعت برداری است. به این ترتیب، تمام مقادیر بین 0 و 1 قرار می‌گیرند. کوچک‌ترین مقدار صفر و بزرگ‌ترین مقدار یک است. این گام مهمی است که به طور گسترده توسط چندین محقق، مانند لی و همکاران [11]و [44]به کار گرفته شده است.

در این مطالعه، سیگنال PPG از یک کلیپ ویدیویی 40 ثانیه‌ای استخراج شد که معادل 2440 مقدار است. پس از مرحله نرمال‌سازی، سیگنال تحت فرآیند تقسیم‌بندی به اندازه‌های مختلف 2، 4 و 10 ثانیه قرار گرفت. هدف مطالعه ما ارزیابی عملکرد استراتژی تقسیم‌بندی سیگنال در فواصل زمانی کوتاه در طبقه‌بندی احساسات است. این عمل اجازه می‌دهد تا در برنامه‌های عاطفی بلادرنگ در آینده استفاده شود [45, 46] .

پیشنهاد یک معماری جدید یادگیری عمیق 1DCNN-LSTM

به منظور طبقه‌بندی، یک معماری یادگیری عمیق ترکیبی شامل 1DCNN و LSTM پیشنهاد شد. در چند سال گذشته، رویکرد یادگیری عمیق در طبقه‌بندی احساسات بسیار مؤثر بوده است، به‌ویژه شبکه عصبی کانولوشن (CNN) که ساختار آن مبتنی بر فیلترهای پیچشی است و قادر به استخراج ویژگی‌های فضایی به طور خودکار هستند، اما نمی‌توانند این ویژگی‌ها را به صورت داده‌های متوالی حفظ کنند. لایه‌های حافظه کوتاه‌مدت (LSTM) شامل ساختاری با دروازه‌های ورودی، خروجی و فراموشی هستند که برای حفظ توالی ویژگی‌ها در طول زمان به کار می‌روند. این ساختار به‌طور مؤثری امکان جمع‌آوری، انتقال و حذف اطلاعات را در سلول‌ها با دقت بالا فراهم می‌کند [47].

با توجه به داده‌های متوالی، مانند سیگنال‌ها، روش 1DCNN یکی از مناسب‌ترین روش‌های یادگیری عمیق است که می‌تواند برای طبقه‌بندی استفاده شود، همان‌طور که قبلاً توسط برخی از محققان اثبات و تأیید شده است[11, 18] .مشخص شده که وقتی لایه‌های پیچشی به بخش‌های متوالی سیگنال اعمال می‌شوند، بردارهای ویژگی یک‌بعدی تولید می‌شوند. پس از این مرحله، لایه‌های ادغام برای کاهش ویژگی‌های قابل یادگیری اعمال می‌شوند. در این مطالعه، از لایه‌های حافظه کوتاه‌مدت بلندمدت (LSTM) برای حفظ توالی ویژگی‌ها در طول زمان استفاده می‌شود. در پایان، از لایه‌های کاملاً متصل برای فرآیند طبقه‌بندی استفاده می‌شود. نمای کلی معماری پیشنهادی در شکل 15 ارائه شده است.



*شکل 15. شبکه 1DCNN-LSTM پیشنهادی*

در چند سال گذشته، ترکیب 1DCNN-LSTM به طور گسترده‌ای در زمینه طبقه‌بندی خودکار احساسات انسان از سیگنال‌های فیزیولوژیکی استفاده شده است [48, 49].در این راستا، ساختارهای جدیدی توسط محققان مختلف ارائه شده است. این ساختارها از نظر تعداد لایه‌های به‌کار رفته و تعداد پارامترهای در نظر گرفته‌شده متفاوت بودند. در مطالعه ما، سیگنال PPG یک‌بعدی به عنوان ورودی شبکه معرفی شده است. سپس دو لایه کانولوشن اعمال می‌شود. پس از هر لایه، یک لایه ادغام حداکثر با اندازه 2\*1 قرار دارد. سپس دو لایه LSTM اضافه شده و برای طبقه‌بندی، چهار لایه کاملاً متصل اعمال می‌شود. لازم به ذکر است که لایه آخر شامل یک نورون است که با یک تابع سیگموئید فعال می‌شود تا احساسات با قطبیت منفی یا مثبت و برانگیختگی بالا یا پایین را طبقه‌بندی کند. برای کاهش یا جلوگیری از بیش‌برازش، لایه‌های حذف و نرمال‌سازی دسته‌ای در معماری ما استفاده شده است. جزئیات شبکه پیشنهادی در جدول 1 خلاصه شده است.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type layer | Size filter | Activation | Output size | Parameters |
| Conv1D | 512,50 | Relu | 244,512 | 26,112 |
| Max Pooling | 2,1 | / | 122,512 | 0 |
| Batch Normalization | / | / | 122,512 | 2048 |
| Dropout | 0.3 | / | 122,512 | 0 |
| Conv1D | 256,25 | Relu | 122,256 | 3,277,056 |
| Max Pooling | 2,1 | / | 61,256 | 0 |
| Batch Normalization | / | / | 61,256 | 1024 |
| Dropout | 0.3 | / | 61,256 | 0 |
| LSTM | 256 | Hard sigmoid | 61,256 | 525,312 |
| LSTM | 128 | Hard sigmoid | 61,256 | 197,120 |
| Flatten | / | / | 7808 | 0 |
| Fully connected | 256 | Sigmoid | 256 | 1,999,104 |
| Dropout | 0.5 | / | 512 | 0 |
| Fully connected | 128 | Sigmoid | 128 | 32,896 |
| Fully connected | 64 | Sigmoid | 64 | 8256 |
| Output | 1 | Sigmoid | 1 | 65 |

*جدول 1 جزئیات شبکه پیشنهادی 1DCNN-LSTM*

نتیجه گیری

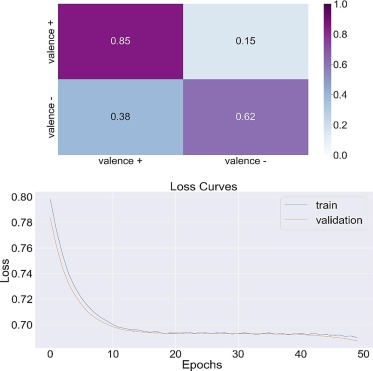
به‌منظور ارزیابی مؤثر شبکه پیشنهادی CNN-LSTM، محاسبه چهار پارامتر، یعنی دقت (AC)، دقت مثبت (PR)، امتیاز F1 و فراخوانی (RE) مناسب تلقی شد. یادآوری این نکته مفید است که از سال 1950، محققان به طور گسترده از این پارامترها برای ارزیابی عملکرد روش‌های خود استفاده کرده‌اند [50] . این پارامترها را می‌توان با استفاده از فرمول‌های زیر ارزیابی کرد:

در اینجا به ترتیب به معنای مثبت واقعی و منفی واقعی هستند، در حالی که به ترتیب به معنای مثبت کاذب و منفی کاذب هستند.   
مدل پیشنهادی بر روی 50 دوره (epoch) آموزش داده شد و از بهینه‌ساز (Adam) با نرخ یادگیری 0.0001 و همچنین تابع باینری cross entropy برای محاسبه‌ی زیان استفاده شد. لازم به ذکر است که 90% پایگاه داده برای آموزش و بقیه برای اعتبارسنجی استفاده شد. سپس، آزمایش ما با استفاده از کتابخانه‌های پایتون بر روی یک کامپیوتر قدرتمند با پردازنده Intel Core i7-7700K با سرعت 4.20 گیگاهرتز و 16 گیگابایت رم اجرا شد.  
ابتدا طبقه‌بندی احساسات انسانی بر روی یک مقیاس قطبیت (valence) انجام شد. یک طبقه‌بندی باینری برای قطبیت مثبت و منفی صورت گرفت و سیگنال PPG در اندازه‌های مختلف تقسیم‌بندی شد. جدول 2 عملکرد به‌دست‌آمده در این تجربه را ارائه می‌دهد، در حالی که شکل 15 ماتریس سردرگمی (confusion matrix) و همچنین منحنی‌های زیان به‌دست‌آمده را نمایش می‌دهد.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Segmented size signal | Accuracy | Precision | Recall | F1 score |
| 2 sec | 60.67% | 61.29% | 48.5% | 55.52% |
| 4 sec | 73.33% | 80.43% | 61.66% | 69.81% |
| 10 sec | 70.83% | 83.78% | 51.66% | 63.91% |

*جدول 2 نتایج به‌دست‌آمده برای طبقه‌بندی برانگیختگی.*

نتایج به‌دست‌آمده نشان می‌دهد که بهترین دقت با تقسیم سیگنال در پنجره‌های 4 ثانیه‌ای به‌دست آمده است، به این معنی که در هر ویدئوی 40 ثانیه‌ای، 10 بخش با بردارهای 244\*1 به‌دست آمده است. علاوه بر این، ماتریس سردرگمی ارائه‌شده در شکل 16 به‌وضوح نشان می‌دهد که مدل پیشنهادی می‌تواند نرخ شناسایی بالاتری را برای احساسات مثبت نسبت به احساسات منفی به‌دست آورد.



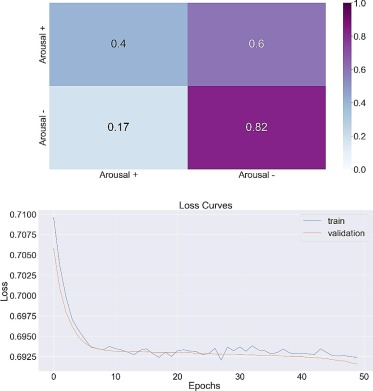
*شکل 16. ماتریس سردرگمی و منحنی‌های زیان به‌دست‌آمده بر روی مقیاس برانگیختگی.*

ما مراحل مشابه آزمایش را که بر روی مقیاس قطبیت به‌دست آورده بودیم، در مقیاس برانگیختگی تکرار کردیم. جدول 3 نتایج به‌دست‌آمده را ارائه می‌دهد.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Segmented size signal | Accuracy | Precision | Recall | F1 score |
| 2 sec | 57.5% | 56.83% | 62.5% | 59.53% |
| 4 sec | 60.00% | 58.33% | 70.5% | 63.63% |
| 10 sec | 50.63 | 50.74% | 42.5% | 46.5% |

*جدول 3 نتایج به‌دست‌آمده برای طبقه‌بندی برانگیختگی.*

شایان ذکر است که شبکه پیشنهادی ما عملکرد طبقه‌بندی پایین‌تری در مقیاس برانگیختگی نسبت به مقیاس قطبیت دارد، به‌طوری که دقت (AC) = 60%، دقت مثبت (PR) = 56.89%، فراخوانی (RE) = 82.5% و امتیاز F1 = 67.34% می‌باشد. مشابه با قطبیت، شبکه پیشنهادی ما با تقسیم سیگنال به مدت 4 ثانیه نرخ شناسایی بهتری را به‌دست می‌آورد. شکل 17 ماتریس سردرگمی و منحنی‌های زیان به‌دست‌آمده را ارائه می‌دهد. علاوه بر این، در مقیاس برانگیختگی، مدل ما در برانگیختگی منفی عملکرد بهتری نسبت به برانگیختگی مثبت دارد.



*شکل 17. ماتریس سردرگمی و منحنی‌های زیان به‌دست‌آمده بر روی مقیاس برانگیختگی.*

بحث‌ها

مشارکت حاضر عمدتاً به شناسایی و طبقه‌بندی احساسات انسانی با استفاده از سیگنال‌های فیزیولوژیکی بدون تماس متمرکز است که در مورد ما، سیگنال‌های PPG استخراج‌شده از ویدیوهای چهره RGB هستند. همان‌طور که قبلاً ذکر شد، این موضوع تحقیقاتی در پردازش سیگنال و طبقه‌بندی احساسات انسانی از اهمیت بالایی برخوردار است. در واقع، در سال‌های اخیر، تعداد زیادی از محققان در این زمینه به این رویکرد روی آورده‌اند [34] .   
با این حال، لازم است ذکر شود که در این نوع تحقیقات، به شدت توصیه می‌شود که از الگوریتم استخراج سیگنال PPG استفاده شود که قادر به شناسایی و ثبت اطلاعات دقیق قلبی عروقی باشد. برای این منظور، الگوریتم ICA برای استفاده در این مطالعه انتخاب شد، زیرا طبق بسیاری از مطالعات مقایسه‌ای موجود در مقالات، این الگوریتم مناسب‌ترین گزینه برای پایگاه داده MAHNOB-HCI است.   
از تجربه ما، قبلاً ذکر کردیم که برخی ویدیوها از پایگاه داده MAHNOB-HCI حذف شده‌اند، زیرا روش ICA در این ویدیوها نتوانسته است سیگنال PPG دقیقی که اطلاعات قلبی صحیح را حمل کند، ارائه دهد. این اولین مانع است که ما در این نوع مطالعه با آن مواجه هستیم. در کاربردهای دنیای واقعی، نمی‌توانیم به حس­گرهای تماسی برای اندازه‌گیری ضربان قلب واقعی انسان تکیه کنیم تا سیگنال‌های PPG استخراج‌شده نادرست را حذف کنیم. برای این منظور، باید از تکنیک‌های استخراج سیگنال PPG استفاده کنیم که در بیشتر ویدیوها توانمندتر و کارآمدتر باشند و همچنین الگوریتم‌هایی را پیشنهاد دهیم که بتوانند سیگنال‌های نامناسب را کنار بگذارند.   
اخیراً، با پیشرفت‌های تحقیقاتی در زمینه نظارت بر ضربان قلب (HR) بدون تماس، محققان هنگام استفاده از تکنیک‌های یادگیری عمیق برای استخراج سیگنال PPG نتایج بهتری به‌دست آورده‌اند. در این راستا، یو و همکاران [51] یک شبکه مکانی-زمانی به نام PhysNet128-3DCNN برای استخراج سیگنال‌های PPG پیشنهاد کردند. همان گروه تحقیقاتی [52] با اعمال یک معماری جدید یادگیری عمیق شامل دو بخش، یعنی یک شبکه تقویت ویدیو مکانی-زمانی (STVEN) و معماری rPPGNet، موفق به دستیابی به دقت بالایی در استخراج سیگنال‌های PPG از ویدیوهای فشرده‌شده شدند. در مورد سانگ و همکاران [53]، آن‌ها روشی را پیشنهاد کردند که از ویژگی‌های مکانی-زمانی استخراج‌شده از تصاویر به‌دست‌آمده از روش CHROM سنتی استفاده می‌کند؛ این ویژگی‌ها سپس به معماری ResNet-18 وارد می‌شوند تا ضربان قلب (HR) به‌دست آید. از آنجا که این روش‌ها نتایج بهتری نسبت به روش‌های سنتی به‌دست آورده‌اند، بنابراین به شدت توصیه می‌شود که از آن‌ها برای شناسایی و طبقه‌بندی احساسات انسانی استفاده شود و در نتیجه به دستیابی به عملکرد بالای نرخ شناسایی کمک کند.   
برخی از کارهای اخیر در زمینه طبقه‌بندی احساسات انسانی، با استفاده از سیگنال‌های فیزیولوژیکی بدون تماس استخراج‌شده از ویدیوهای RGB، در جدول 4 زیر ارائه شده است.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Authors | Methods | Non\_contact physiological Signals | Emotion | Database | Accuracy |
| Maaoui et al. [45] | SVM\_LDA | PPG | Stress | / | 94.40%–91.10% |
| Meziati Sabour et al. [47] | SVM | HRV | Stress | UBFC\_Phys | 85.48% |
| Yang et al. [44] | 1DCNN | PPG | Pain | BioVid | 58.92% |
| Ouzar et al. [48] | Feed\_forward NN | HRV\_PPG | Happiness\_fear\_pain\_Embarrassment | BP4D+ | 53.59%–55.33% |
| Lampier et al. [49] | SVM\_K\_NN | PPG | Valence | / | 42%–38% |
| Yu et al. [57] | SVM | HRV | Valence\_Arousal\_(3 classes) | MAHNOB\_HCI | 46.86%–44.02% |
| Our proposed method | 1DCNN\_LSTM | PPG | Valence\_Arousal\_(2 classes) | MAHNOB\_HCI | 73.33%–60% |

*جدول 4 نتایج مقایسه‌ای در طبقه‌بندی احساسات با استفاده از سیگنال‌های فیزیولوژیکی بدون تماس.*

علاوه بر این، مطالعه مقایسه‌ای ارائه‌شده در جدول 4 نشان می‌دهد که روش ما می‌تواند نتایج به مراتب بهتری (دقت 73.33% در مقیاس قطبیت و 60% در مقیاس برانگیختگی) نسبت به نتایج گزارش‌شده قبلی توسط Lumpier et al [39]، Ouzar et al. [38]، Yu et al. [51] و Yang et al. [34] به‌دست آورد.   
در مورد طبقه‌بندی استرس انسانی، بسیاری از محققان به نرخ شناسایی بالایی دست یافته‌اند، زیرا تغییرات شدیدی در واکنش‌های قلب و عروق خونی انسان هنگام تجربه استرس ایجاد می‌شود. علاوه بر این، لازم به ذکر است که مطالعه ما با تحقیقات قبلاً انجام‌شده توسط [35] و Meziati Sabour et al. [37] تفاوت دارد.   
علاوه بر این، به محققان در این حوزه به شدت توصیه می‌شود که یک پایگاه داده جدید و مدرن با کیفیت تصویر بالاتر و شرایط نورپردازی یکنواخت مانند UFBC-Phys توسعه دهند که استخراج سیگنال PPG را بهبود می‌بخشد. علاوه بر این، به محققان در این حوزه به شدت توصیه می‌شود که یک پایگاه داده جدید و مدرن‌تر مانند UFBC-Phys با شرایط نورپردازی یکنواخت توسعه دهند.   
اگرچه نتایج ما در ابتدا می‌توانند به‌عنوان نتایج قابل قبول در نظر گرفته شوند، اما تحقیقات بیشتری برای بهبود آن‌ها از طریق استفاده از روش‌های استخراج سیگنال PPG دقیق‌تر مانند تکنیک‌های یادگیری عمیق هنوز مورد نیاز است. Ouzar et al. [38]یکی از تکنیک‌های یادگیری عمیق معتبر و قدرتمند را با هدف استخراج سیگنال‌های PPG دقیق‌تر از ویدیوهای صورت (MTTS-CAN) استفاده کردند. همچنین، Yu et al. [51] شبکه PhysNet128-3DCNN-ED را پیشنهاد کردند. با این حال، باید تأکید شود که روش ما به نرخ شناسایی بالاتری دست یافته است که عمدتاً به طبقه‌بندی مبتنی بر شبکه یادگیری عمیق پیشنهادی در این کار مربوط می‌شود. بنابراین می‌توان نتیجه‌گیری کرد که معماری 1DCNN-LSTM پیشنهادی که تنها از بخش‌های 4 ثانیه‌ای استفاده می‌کند، می‌تواند دقت طبقه‌بندی بسیار بهتری نسبت به نتایج به‌دست‌آمده با روش‌های سنتی یادگیری ماشین مانند SVM که توسط Yu et al. [51] پیشنهاد شده و شبکه پیش‌خور که توسط [38]پیشنهاد شده، به‌دست آورد.

نتیجه گیری و کارهای آینده

هدف اصلی این مطالعه توسعه یک رویکرد جدید در زمینه شناسایی خودکار احساسات انسانی است. این روش شامل استفاده از تکنیک فوتوپلتیسموگرافی (PPG) بدون تماس می‌باشد. برای طبقه‌بندی، از تکنیک‌های یادگیری عمیق استفاده می‌شود.

در مرحله استخراج سیگنال، از پایگاه داده واقعی احساسات MAHNOB-HCI استفاده شد و تکنیک ICA به‌کار گرفته شد. همچنین، چارچوب شناسایی اشیاء Viola–Jones فقط برای شناسایی چهره مورد استفاده قرار گرفت.

علاوه بر این، یک مدل شبکه عصبی کانولوشنی یک بعدی (1DCNN) به همراه حافظه کوتاه‌مدت بلندمدت (LSTM) برای هدف طبقه‌بندی احساسات بر روی مقیاس ابعادی، پس از مراحل پیش‌پردازش نرمال‌سازی و تقسیم‌بندی، پیشنهاد شد. یافته‌ها نشان داد که مدل پیشنهادی می‌تواند دقت طبقه‌بندی قابل توجهی معادل 73.33% برای قطبیت و 60% برای برانگیختگی با یک سیگنال کوتاه 4 ثانیه‌ای به‌دست آورد.

تحقیقات آینده باید به بهبود نتایج به‌دست‌آمده در این مطالعه با استفاده از روش‌های استخراج سیگنال PPG جدیدتر و قوی‌تر اختصاص یابند. همچنین، همکاری با سایر محققان متخصص در این زمینه بسیار مطلوب خواهد بود. علاوه بر این، برای شناسایی دقیق‌تر احساسات در آینده، توصیه می‌شود از طبقه‌بندی چهار بعدی استفاده شود. چهار بعد شامل: قطبیت مثبت با برانگیختگی بالا، قطبیت مثبت با برانگیختگی پایین، قطبیت منفی با برانگیختگی بالا و قطبیت منفی با برانگیختگی پایین است.

در شناسایی خودکار احساسات انسانی، محققان به نرخ شناسایی نسبتاً بالایی با استفاده از مدالیته ابراز چهره دست یافته‌اند که از 90% فراتر می‌رود [54] .این نتایج بسیار امیدوارکننده هستند، زیرا ما را ترغیب می‌کنند تا از این مدالیته در مطالعات آینده استفاده کنیم و یک سیستم چندمدیاته با نرخ شناسایی بالا آموزش دهیم.

1. Sofiyabadi, J. and C. Valmohammadi, *Impact of knowledge management practices on innovation performance.* IEEE Transactions on Engineering Management, 2020. **69**(6): p. 3225-3239.

2. Wang, Y., L. Wang, and J. Yang. *Egonet based anomaly detection in E-bank transaction networks*. in *IOP Conference Series: Materials Science and Engineering*. 2020. IOP Publishing.

3. Błaszczyński, J., et al., *Auto loan fraud detection using dominance-based rough set approach versus machine learning methods.* Expert Systems with Applications, 2021. **163**: p. 113740.

4. Olatunji, S.O., et al., *Early diagnosis of thyroid cancer diseases using computational intelligence techniques: A case study of a Saudi Arabian dataset.* Computers in biology and medicine, 2021. **131**: p. 104267.

5. Azpiazu, J.I.R., et al., *Out-of-hospital cardiac arrest following the COVID-19 pandemic.* JAMA Network Open, 2024. **7**(1): p. e2352377-e2352377.

6. Zafar, M.S., et al., *Biomimetic aspects of restorative dentistry biomaterials.* Biomimetics, 2020. **5**(3): p. 34.

7. Shu, L., et al., *A review of emotion recognition using physiological signals.* Sensors, 2018. **18**(7): p. 2074.

8. Denervaud, S., et al., *Emotion recognition development: Preliminary evidence for an effect of school pedagogical practices.* Learning and Instruction, 2020. **69**: p. 101353.

9. Bani, M., et al., *Behind the mask: Emotion recognition in healthcare students.* Medical science educator, 2021. **31**(4): p. 1273-1277.

10. Meléndez, J.C., et al., *Emotion recognition changes in a confinement situation due to COVID-19.* Journal of Environmental Psychology, 2020. **72**: p. 101518.

11. Lee, M.S., et al., *Fast emotion recognition based on single pulse PPG signal with convolutional neural network.* Applied Sciences, 2019. **9**(16): p. 3355.

12. Ziccardi, S., F. Crescenzo, and M. Calabrese, *“What is hidden behind the mask?” Facial emotion recognition at the time of COVID-19 pandemic in cognitively normal multiple sclerosis patients.* Diagnostics, 2021. **12**(1): p. 47.

13. Ekman, P., *Facial expression and emotion.* American psychologist, 1993. **48**(4): p. 384.

14. Li, M., et al., *Emotion recognition from multichannel EEG signals using K-nearest neighbor classification.* Technology and health care, 2018. **26**(S1): p. 509-519.

15. Amjadzadeh, M. and K. Ansari-Asl, *An innovative emotion assessment using physiological signals based on the combination mechanism.* Scientia Iranica, 2017. **24**(6): p. 3157-3170.

16. Hakim, L., et al. *Emotion recognition in elderly based on SpO 2 and pulse rate signals using support vector machine*. in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*. 2018. IEEE.

17. Ayata, D., Y. Yaslan, and M.E. Kamasak, *Emotion recognition from multimodal physiological signals for emotion aware healthcare systems.* Journal of Medical and Biological Engineering, 2020. **40**: p. 149-157.

18. Santamaria-Granados, L., et al., *Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS).* IEEE Access, 2018. **7**: p. 57-67.

19. Domínguez-Jiménez, J.A., et al., *A machine learning model for emotion recognition from physiological signals.* Biomedical signal processing and control, 2020. **55**: p. 101646.

20. Boccignone, G., et al., *An open framework for remote-PPG methods and their assessment.* IEEE Access, 2020. **8**: p. 216083-216103.

21. Verkruysse, W., L.O. Svaasand, and J.S. Nelson, *Remote plethysmographic imaging using ambient light.* Optics express, 2008. **16**(26): p. 21434-21445.

22. Soleymani, M., et al., *A multimodal database for affect recognition and implicit tagging.* IEEE transactions on affective computing, 2011. **3**(1): p. 42-55.

23. Poh, M.-Z., D.J. McDuff, and R.W. Picard, *Non-contact, automated cardiac pulse measurements using video imaging and blind source separation.* Optics express, 2010. **18**(10): p. 10762-10774.

24. De Haan, G. and V. Jeanne, *Robust pulse rate from chrominance-based rPPG.* IEEE transactions on biomedical engineering, 2013. **60**(10): p. 2878-2886.

25. Wang, W., et al., *Algorithmic principles of remote PPG.* IEEE Transactions on Biomedical Engineering, 2016. **64**(7): p. 1479-1491.

26. Song, R., et al., *PulseGAN: Learning to generate realistic pulse waveforms in remote photoplethysmography.* IEEE Journal of Biomedical and Health Informatics, 2021. **25**(5): p. 1373-1384.

27. Subramaniam, A. and K. Rajitha, *Estimation of the Cardiac Pulse from Facial Video in Realistic Conditions.* ICAART (2), 2019. **318**: p. 319.

28. Lamba, P.S. and D. Virmani, *Contactless heart rate estimation from face videos.* Journal of Statistics and Management Systems, 2020. **23**(7): p. 1275-1284.

29. Ghanadian, H., M. Ghodratigohar, and H. Al Osman, *A machine learning method to improve non-contact heart rate monitoring using an RGB camera.* IEEE Access, 2018. **6**: p. 57085-57094.

30. Kurihara, K., D. Sugimura, and T. Hamamoto, *Non-contact heart rate estimation via adaptive rgb/nir signal fusion.* IEEE Transactions on Image Processing, 2021. **30**: p. 6528-6543.

31. Zheng, K., et al., *Non-contact heart rate detection when face information is missing during online learning.* Sensors, 2020. **20**(24): p. 7021.

32. Filippini, C., et al., *Thermal infrared imaging-based affective computing and its application to facilitate human robot interaction: A review.* Applied Sciences, 2020. **10**(8): p. 2924.

33. Kessler, V., et al. *Pain recognition with camera photoplethysmography*. in *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. 2017. IEEE.

34. Yang, R., et al., *Non-contact pain recognition from video sequences with remote physiological measurements prediction.* arXiv preprint arXiv:2105.08822, 2021.

35. Maaoui, C., F. Bousefsaf, and A. Pruski, *Automatic human stress detection based on webcam photoplethysmographic signals.* Journal of Mechanics in Medicine and Biology, 2016. **16**(04): p. 1650039.

36. Benezeth, Y., et al. *Remote heart rate variability for emotional state monitoring*. in *2018 IEEE EMBS international conference on biomedical & health informatics (BHI)*. 2018. IEEE.

37. Sabour, R.M., et al., *Ubfc-phys: A multimodal database for psychophysiological studies of social stress.* IEEE Transactions on Affective Computing, 2021. **14**(1): p. 622-636.

38. Ouzar, Y., et al. *Video-based multimodal spontaneous emotion recognition using facial expressions and physiological signals*. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

39. Lampier, L., et al. *A Preliminary Approach to Identify Arousal and Valence Using Remote Photoplethysmography*. in *Brazilian Congress on Biomedical Engineering*. 2020. Springer.

40. Wang, C., T. Pun, and G. Chanel, *A comparative survey of methods for remote heart rate detection from frontal face videos.* Frontiers in bioengineering and biotechnology, 2018. **6**: p. 33.

41. McDuff, D. and E. Blackford. *iphys: An open non-contact imaging-based physiological measurement toolbox*. in *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. 2019. IEEE.

42. Carreiras, C., et al., *BioSPPy: Biosignal Processing in Python. 2015.* URL <https://github>. com/PIA-Group/BioSPPy, 2018.

43. Wiem, M.B.H. and Z. Lachiri, *Emotion classification in arousal valence model using MAHNOB-HCI database.* International Journal of Advanced Computer Science and Applications, 2017. **8**(3).

44. Al Machot, F., et al., *A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors.* Sensors, 2019. **19**(7): p. 1659.

45. Nakisa, B., et al., *Automatic emotion recognition using temporal multimodal deep learning.* IEEE Access, 2020. **8**: p. 225463-225474.

46. Shen, J., et al. *A novel depression detection method based on pervasive EEG and EEG splitting criterion*. in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2017. IEEE.

47. Alom, M.Z., et al., *A state-of-the-art survey on deep learning theory and architectures.* electronics, 2019. **8**(3): p. 292.

48. Dar, M.N., et al., *CNN and LSTM-based emotion charting using physiological signals.* Sensors, 2020. **20**(16): p. 4551.

49. Nahar, N., et al. *A Hybrid CNN-LSTM-Based Emotional Status Determination using Physiological Signals*. in *Proceedings of the Third International Conference on Trends in Computational and Cognitive Engineering: TCCE 2021*. 2022. Springer.

50. Mellouk, W. and W. Handouzi, *CNN-LSTM for automatic emotion recognition using contactless photoplythesmographic signals.* Biomedical Signal Processing and Control, 2023. **85**: p. 104907.

51. Yu, Z., X. Li, and G. Zhao, *Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks.* arXiv preprint arXiv:1905.02419, 2019.

52. Yu, Z., et al. *Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement*. in *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.

53. Song, R., et al., *Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks.* IEEE Transactions on Instrumentation and Measurement, 2020. **69**(10): p. 7411-7421.

54. Mellouk, W. and W. Handouzi, *Facial emotion recognition using deep learning: review and insights.* Procedia Computer Science, 2020. **175**: p. 689-694.