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:

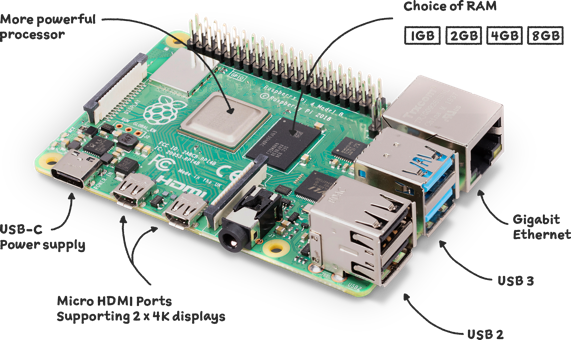
امروزه صنعت بانکداری نقش اساسی در اقتصاد خرد و کلان کشورها ایفا می نماید. توسعه‌ی بانکداری باز و گسترش بازارهای مالی متنوع شرایط جدیدی را در عرصه اقتصاد جهانی شکل داده است.صنعت بانکداري پيشرفته تأثير بسزايي بر اقتصاد و سياست دارد. امروزه بانکداري الکترونيک به يکي از خدمات مالي تبديل شده است که اکثر مردم 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، 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 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. 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.

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.

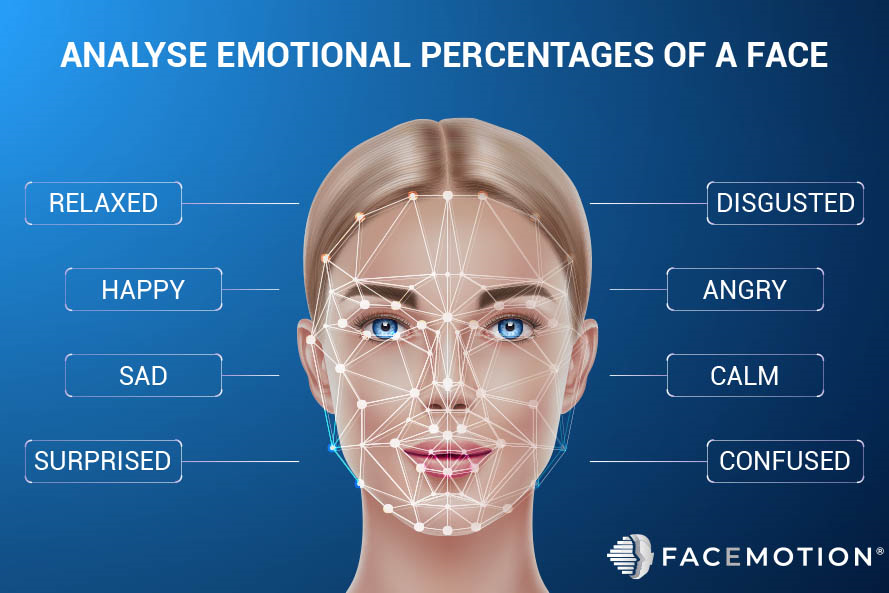


*شکل 1. برد رَسپي‌بري پاي 4*

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.

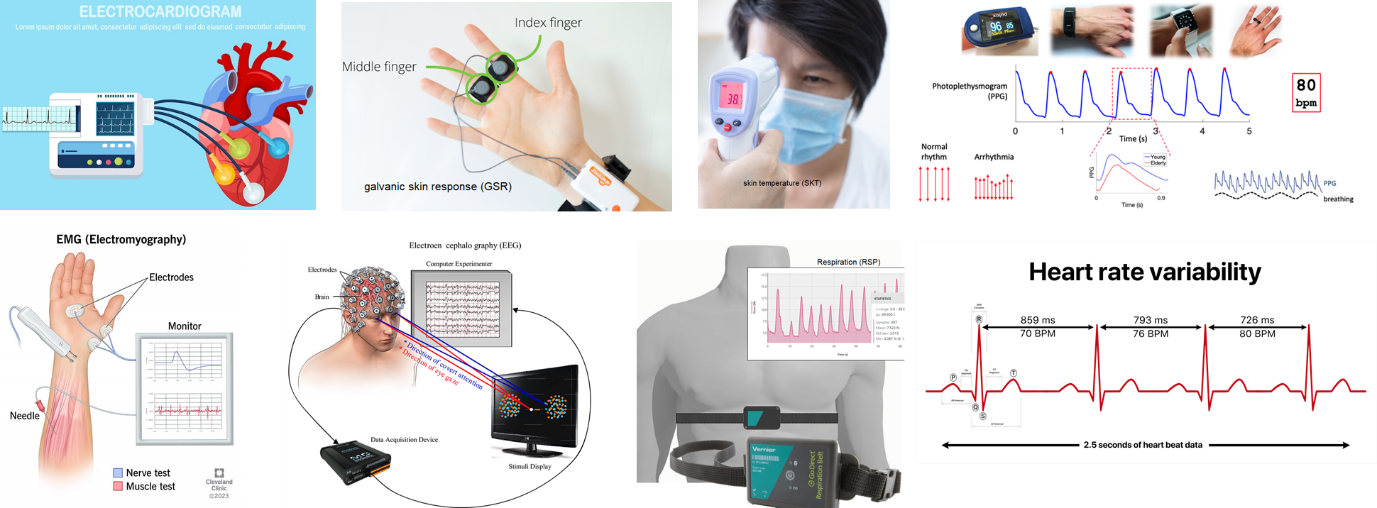
تشخیص احساسات یک حوزه تحقیقاتی نسبتاً جدید و مهم است. این فرآیند شناسایی احساسات انسان از طریق شناخت وضعیت روانی و عاطفی آنهاست. احساساتی که افراد می­توانند در زندگی روزمره خود داشته باشند می­توانند مثبت یا منفی باشند.

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، آموزش به‌ویژه در دوره همه‌گیری COVID 19 به آنها ارائه دهند. در شکل 2 تمام احساسات یک چهره نشان داده شده است.



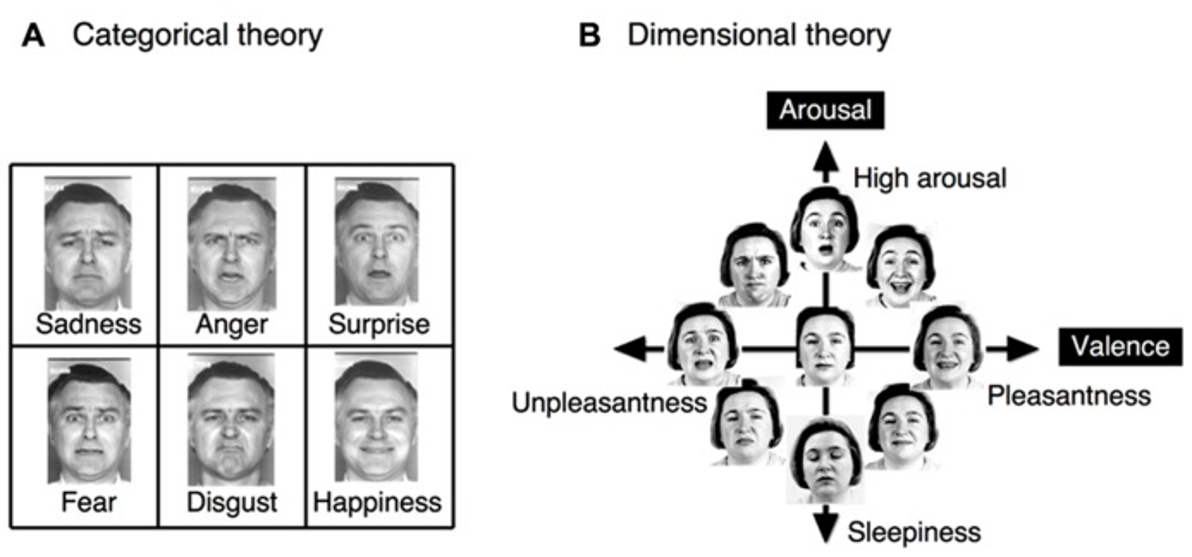
*شکل 2. تحلیل احساسات یک چهره*

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. 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 قابل تشخیص هستند. که در شکل 3 نشان داده شده است.



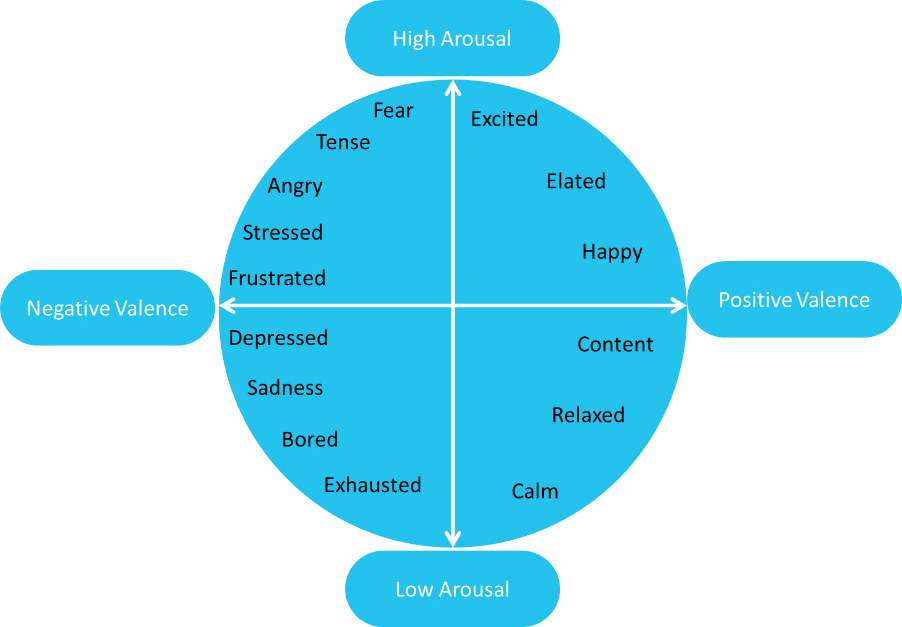
*شکل 3. ابزارهای ارزیابی فعالیت‌های فیزیولوژیکی بدن*

با توجه به مقالات، عواطف عموماً بر اساس دو رویکرد تقسیم می‌شوند: رویکرد مقوله‌ای و رویکرد بعدی ( شکل 4) . 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 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.



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

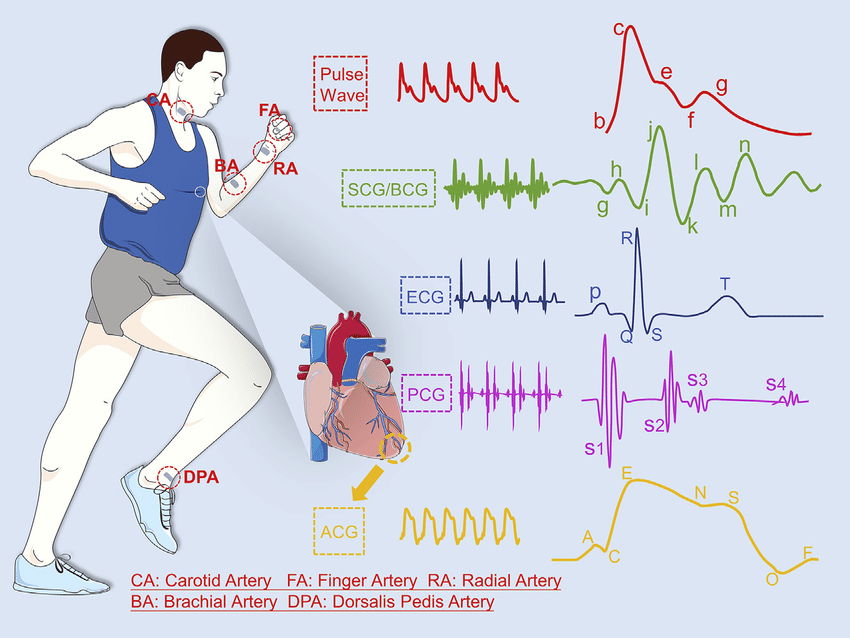
رویکرد ابعادی امکان طبقه‌بندی و نمایش گرافیکی احساسات را بر روی محورهای ظرفیت و برانگیختگی فراهم می‌کند. محور ظرفیت برای طبقه بندی حالت هیجانی مثبت یا منفی طراحی شده است در حالی که محور برانگیختگی برای احساسات هیجان زده یا بی تفاوت است. شایان ذکر است که ظرفیت از منفی به مثبت گسترش می­بابد و به کیفیت 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)، تجزیه و تحلیل تشخیص خطی (LDA)، ماشین بردار پشتیبان (SVM) و جنگل تصادفی RF 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. با استفاده از این روش ها می­توان به نتایج موفق تری نسبت به روش­های سنتی دست یافت.

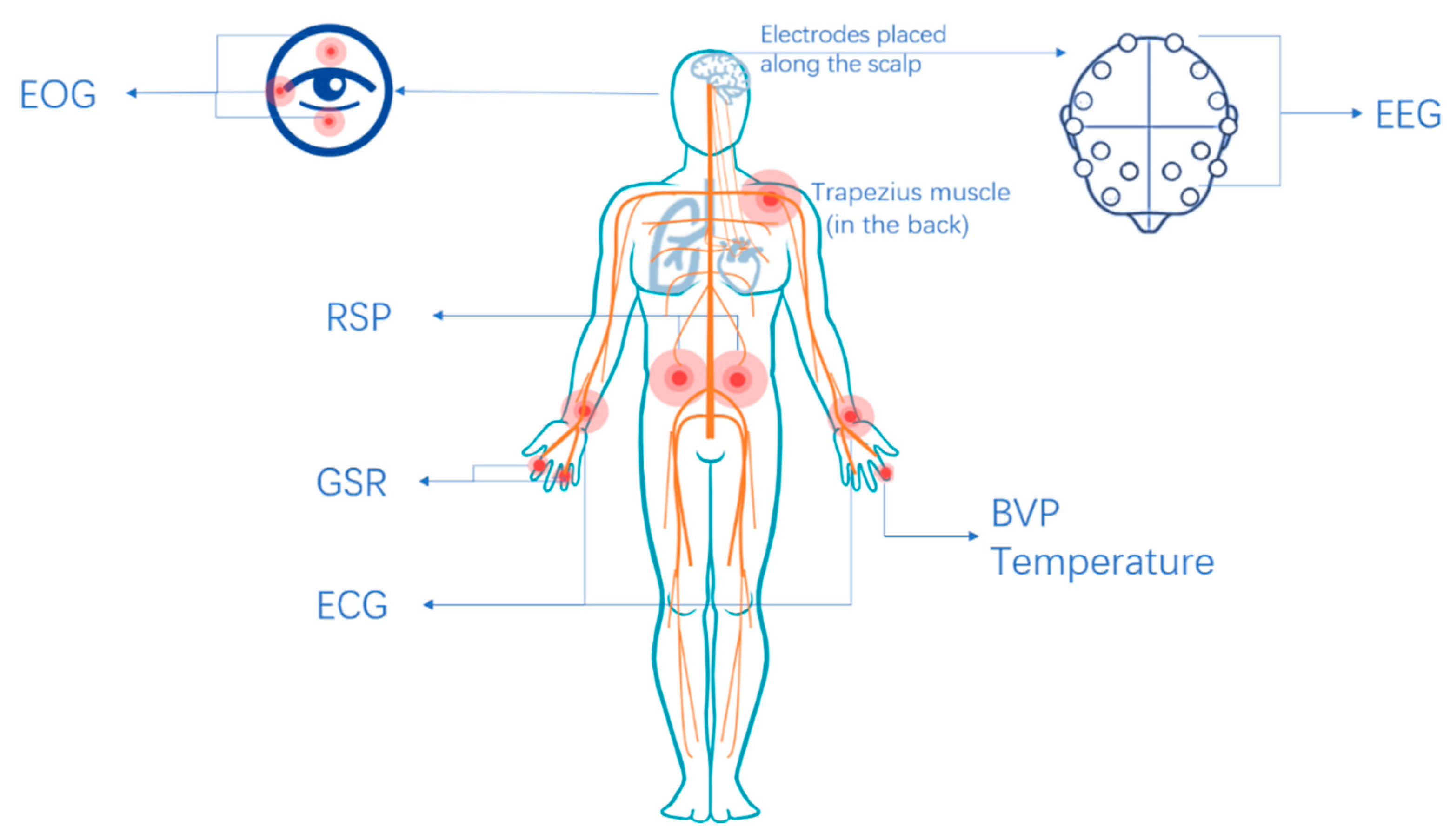


*شکل 5. توزیع احساسات در محورهای ظرفیت (valence) و برانگیختگی (arousal)*

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)

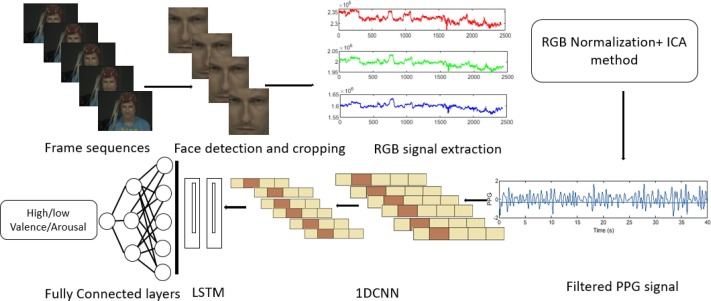
شایان ذکر است که اخیراً روش غیر تماسی برای تخمین و پایش ضربان قلب (HR) توجه تعداد زیادی از محققان را به خود جلب کرده است. در این حالت، شخص مجبور نیست از سنسور روی بدن خود استفاده کند. این تکنیک ساده است و مزایای زیادی دارد. به عنوان مثال می­توان آن را به طور خاص برای افرادی که در طول همه­گیری COVID-19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



در سال 2008، ورکرویس و همکاران [18] 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 اعمال شد. توجه داشته باشید که سیگنال­های PPG ثبت شده با استفاده از تکنیک­های یادگیری عمیق سنتی (یادگیری عمیق) طبقه بندی شدند. شکل زیر یک نمای کلی از روش های مورد استفاده در مطالعه را نشان می دهد.



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 [31] 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 [32] 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 [33] 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 [34] 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 [35] 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 [36] یک مطالعه مقایسه­ای بین دو ناحیه صورت، یعنی فقط گونه­ها و گونه­ها با بینی انجام دادند.

علاوه بر این، غنادیان و همکاران [37] به منظور بهبود روش ICA [31] یک تکنیک یادگیری ماشین را برای شناسایی بهترین جزء اطلاعاتی سیگنال PPG پس از اعمال تکنیک ICA اتخاذ کردند. همچنین، کوریهارا و همکاران [38] 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 [39] 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[42]. 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 [43] 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 [44]، 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 [45]، آنها از طبقه‌بندی‌کننده‌های SVM و LDA برای طبقه‌بندی سطح استرس با استفاده از سیگنال‌های PPG بدون تماس استفاده کردند. اخیراً بنزث و همکاران [46] 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 [39]، 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 [47] 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 [48] 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 [49]، 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 [19]. 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].

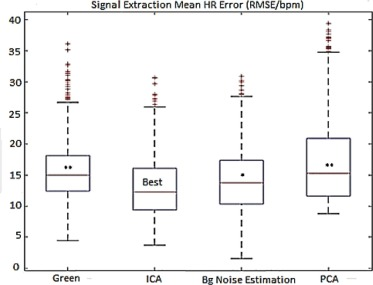
به طور گسترده‌ای پذیرفته شده است که MAHNOB-HCI یک پایگاه داده چندوجهی (شامل داده‌های صوتی، تصویری و فیزیولوژیکی) است که شامل 30 شرکت‌کننده از جنسیت‌ها و قومیت‌های مختلف است. در این حالت، هر یک از شرکت‌کنندگان برای برانگیختن احساسات خود، تعدادی ویدیوی انتخابی را با دقت تماشا می‌کنند و در این حین، از آنها 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 [40] 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.

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



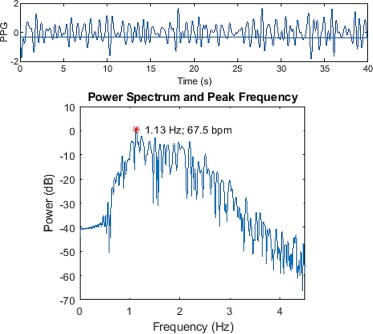
لازم به ذکر است که چندین مرحله برای ایجاد پایگاه داده جدید ما که شامل سیگنال‌های PPG بدون تماس است دنبال شد. ابتدا، ده شرکت‌کننده از پایگاه داده MAHNOB-HCI بارگذاری شد. هر سوژه شامل 20 ویدیوی RGB از چهره جلویی است. سپس، تکنیک ICA بر روی توالی‌های ویدیویی انتخابی با طول 40 ثانیه اعمال شد. پس از آن، نرخ 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[50]. 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[50]. شکل زیر سیگنال PPG و تخمین HR به‌دست‌آمده از سیگنال چگالی طیفی توان را نشان می‌دهد.



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.



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 [51]. 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 [50] 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 زیر نمای کلی از مراحل دنبال شده در مطالعه پیشنهادی را ارائه می­دهد.

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 با توزیع در مقیاس‌های مختلف از 1 تا 9 ارائه کردند.

در مقیاس عاطفی، 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 [52].

سیگنال PPG فیلتر شده نشان داده شده در شکل 2 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.