



دانشگاه صنعتی شریف  
دانشکده‌ی مهندسی برق

گزارش نهایی پروژه‌ی کارشناسی  
مهندسی برق گرایش بیوالکتریک

بهبود فرآیند استخراج سیگنال فتوپلتیسموگرافی به صورت راه  
دور با استفاده از تشخیص وزن دار نواحی پوست صورت و بدن

نگارش  
زهرا ملکی

استاد راهنما  
دکتر بابک خلج

استاد درس  
دکتر اقلیدس

بهار ۱۴۰۴

---





## قدردانی

با تشکر از استاد راهنما پروژه دکتر خلیج و همکارانم آقای امیرحسین اکبری و آقای امیرحسین بینش.



# بهبود فرآیند استخراج سیگنال فتوپلته ایسموگرافی به صورت راه دور با استفاده از تشخیص وزن دار نواحی پوست صورت و بدن

## چکیده

این پروژه به بررسی بهبود استخراج پالس ویو در سیگنال های فتوپلته ایسموگرافی از راه دور (rPPG) با استفاده از یک روش توسعه یافته جداسازی پوست می پردازد. هدف اصلی این تحقیق بهبود دقت استخراج سیگنال های rPPG در شرایط مختلف نوری و حرکتی است. برای این منظور، از یک پایپ لاین مبتنی بر الگوریتم دیپلر برای جداسازی دقیق پوست استفاده شده است. داده های مورد استفاده شامل مجموعه های ویدئویی و سیگنال های PPG همزمان با شرایط نوری و حرکتی متغیر جمع آوری شده اند. به منظور استخراج سیگنال های rPPG، از ماسک های پیوسته و الگوریتم پوز استفاده شده است تا به طور مؤثر سیگنال های قلبی از تصاویر استخراج شوند. این روش ها دقت و قابلیت اطمینان سیستم های نظارت بر ضربان قلب غیرتماسی را به طور قابل توجهی بهبود می بخشد.

کلمات کلیدی: فتوپلته ایسموگرافی از راه دور (rPPG) - جداسازی پوست - استخراج پالس - نظارت بر ضربان قلب.





# فهرست مطالب

۳	مقدمه
۳	۱ فوتوپلتیسموگرافی
۴	۲ فوتوپلتیسموگرافی از راه دور
۵	۳ چطور این سیگنال را بدست بیاوریم؟
۸	۴ کاربر های این تکنولوژی
۱۱	روند دستیابی به این سیگنال
۱۱	۱ ساختن دیتاست های مطلوب
۱۲	۲ روش های بدست آوردن سیگنال از تصویر
۱۴	۳ روش های استخراج ضربان قلب
۱۷	متد های پردازش فیلم
۱۷	۱ روش های کلاسیک
۱۸	۲ روش های یادگیری عمیق
۲۱	۳ روش ارائه شده این پژوهش
۲۳	نتایج
۲۳	۱ پیکربندی آزمایش
۲۴	۲ مجموعه داده ها
۲۴	۳ معیارهای ارزیابی
۲۹	جمع بندی
۳۱	مراجع



## فهرست تصاویر

۴	بدست آوردن سیگنال PPG [۱]	۱
۵	بدست آوردن سیگنال rPPG [۲]	۲
۶	سمت راست مدل لندکارکر. چپ مدل چند کلاسه مدیاپایپ برای استخراج نواحی پوست [۳، ۴]	۳
۸	روش کلی استخراج سیگنال	۴
۱۱	درست کردن دیتاست جهت یادگیری مدل	۱
۱۳	مرحله به مرحله استخراج سیگنال ضربان از سیگنال RGB طبق روش امیت [۵]	۲
۱۴	الگوریتم پوز [۲].	۳
۱۷	خروجی روش چرف [۶].	۱
۱۸	خروجی روش لولست [۷].	۲
۱۹	شماتیکی از مدل دیپلب [۸].	۳
۲۰	خروجی روش امیت [۵]	۴
۲۵	نمایی از خروجی های سه مدل روی فریم های دیتاست ما.	۱
۲۶	سیگنال های استخراج شده با استفاده از لندمارکر در بالا. در میانه چند کلاسه. در پایین مدل نهایی ما	۲
۲۷	خروجی مدل روی داده هایی از دیتاست کوکو	۳
۲۸	عملکرد مدل ها در دیتاست ها با متد های استخراج مختلف.	۴



## فهرست جداول



## مقدمه

### ۱ فوتوپلتیسموگرافی

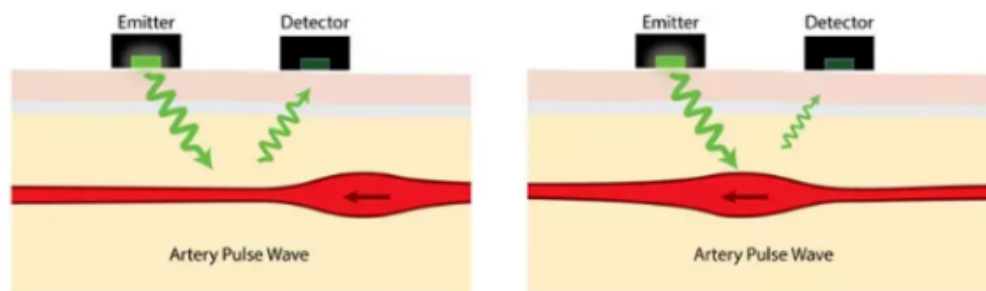
فوتوپلتیسموگرافی (PPG) یک روش نوری برای اندازه‌گیری تغییرات حجم خون در رگ‌های خونی سطحی پوست است که بر اساس تغییرات جذب و بازتاب نور هنگام عبور موج نبضی خون از رگ‌ها کار می‌کند. این روش از یک منبع نوری مانند منبع نور برای انتشار نور و یک حسگر نوری برای دریافت نور بازتاب‌شده استفاده می‌کند. میزان نوری که توسط خون در رگ‌ها جذب می‌شود، به صورت متناوب تغییر می‌کند، زیرا در هر سیکل ضربان قلب، حجم خون درون رگ‌های سطحی پوست افزایش و کاهش می‌یابد. این تغییرات در جذب و بازتاب نور به صورت یک سیگنال نوری ثبت می‌شود که حاوی اطلاعات مهمی در مورد فعالیت‌های همودینامیکی بدن است.

سیگنال PPG معمولاً به دو مؤلفه تقسیم می‌شود: یک مؤلفه متناوب که به تغییرات ضربان قلبی مرتبط است و یک مؤلفه ثابت که نشان‌دهنده جذب کلی نور توسط بافت‌ها و خون غیریالسدار است. مؤلفه متناوب دارای یک الگوی نوسانی است که به ریتم قلب مرتبط است و اطلاعاتی درباره ضربان و پمپاژ خون ارائه می‌دهد. در مقابل، مؤلفه ثابت معمولاً پایدارتر است و به میزان کلی خون درون بافت‌ها و تغییرات غیرمتناوب آن مربوط می‌شود. استخراج صحیح مؤلفه متناوب از میان سایر نویزهای محیطی یکی از چالش‌های پردازش سیگنال PPG است.

در فرایند جمع‌آوری سیگنال PPG، عواملی مانند ضخامت پوست، رنگدانه‌های پوستی، میزان نور محیط و حتی حرکت بدن می‌توانند بر کیفیت سیگنال تأثیر بگذارند. بنابراین، برای به دست آوردن اطلاعات دقیق، معمولاً از الگوریتم‌های پردازش سیگنال مانند فیلترهای دیجیتال، پردازش موجک و تبدیل فوریه برای حذف نویزها و بهبود دقت استخراج سیگنال استفاده می‌شود. برخی از روش‌های پیشرفته‌تر حتی از یادگیری ماشین و شبکه‌های عصبی برای بهبود عملکرد سیستم‌های مبتنی بر PPG بهره می‌برند. یکی از چالش‌های مهم در تحلیل سیگنال PPG حساسیت آن به حرکات بدن است. هنگام حرکت، پوست و رگ‌های خونی ممکن است تغییر موقعیت دهند، که این امر باعث ایجاد تغییرات غیرواقعی در سیگنال می‌شود. برای غلبه بر این مشکل، روش‌های پردازشی خاصی مانند الگوریتم‌های تطبیقی و استفاده از حسگرهای تکمیلی مانند شتاب‌سنج‌ها وژیروسکوپ‌ها پیشنهاد شده‌اند. این حسگرهای تکمیلی می‌توانند حرکات‌های ناخواسته را شناسایی کرده و در پردازش سیگنال PPG اصلاحاتی را اعمال کنند.

از دیگر چالش‌های این روش، وابستگی آن به شرایط محیطی مانند شدت نور محیط است. در محیط‌هایی با نور متغیر، ممکن است میزان نوری که به پوست می‌رسد تغییر کند و این امر بر کیفیت سیگنال اثر بگذارد. برای مقابله با این مشکل، بسیاری از دستگاه‌های مبتنی بر PPG از منابع نوری خاصی مانند نورهای مادون قرمز استفاده می‌کنند که نسبت به تغییرات نور محیطی حساسیت کمتری دارند. همچنین، روش‌های نرم‌افزاری مانند الگوریتم‌های نرمال‌سازی نیز می‌توانند کمک کنند تا اثر تغییرات نور محیطی بر سیگنال کاهش یابد.

یکی از مفاهیم پیشرفته‌تر مرتبط با PPG، فوتوپلتیسموگرافی از راه دور (rPPG) است که بدون نیاز به تماس فیزیکی، از طریق دوربین‌های معمولی اقدام به ثبت و پردازش سیگنال PPG می‌کند. در این روش، تغییرات جزئی در رنگ پوست که ناشی از نوسانات جریان خون است، با استفاده از الگوریتم‌های پردازش تصویر و یادگیری ماشین استخراج می‌شوند. این تکنیک، امکان اندازه‌گیری پارامترهای همودینامیکی را حتی از طریق فیلم‌های ضبط‌شده یا دوربین‌های



شکل ۱: بدست آوردن سیگنال PPG [۱]

نظارتی فراهم می‌کند، اما همچنان با چالش‌هایی مانند تاثیر نور محیط، کیفیت تصویر و حرکات بدن مواجه است. از لحاظ تاریخی، PPG برای نخستین بار در دهه ۱۹۳۰ توسعه یافت، اما در دهه‌های اخیر با پیشرفت فناوری حسگرها و الگوریتم‌های پردازش سیگنال، دقت و قابلیت‌های آن بهبود یافته است. امروزه، بسیاری از پژوهش‌ها روی بهینه‌سازی استخراج اطلاعات از این سیگنال متمرکز هستند و ترکیب آن با سایر فناوری‌های حسگری می‌تواند کاربردهای آن را گسترش دهد [۹، ۱۰].

## ۲ فوتوپلتیسموگرافی از راه دور

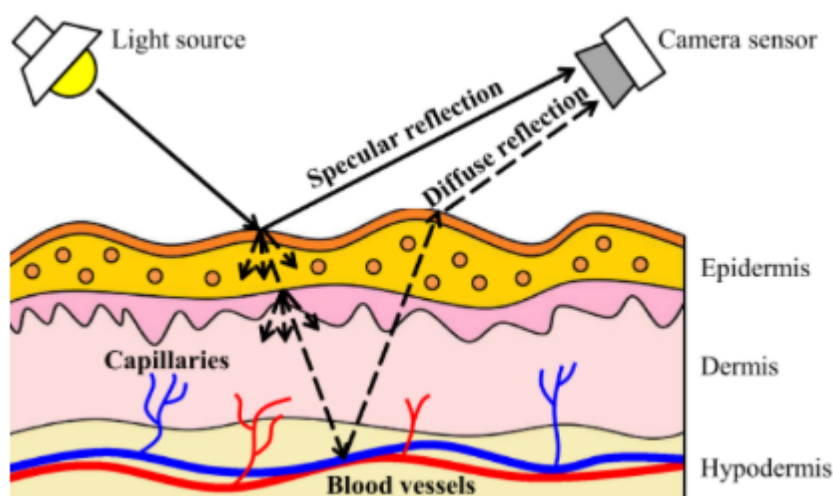
فوتوپلتیسموگرافی از راه دور (rPPG) یک روش نوری غیرتماسی برای اندازه‌گیری تغییرات حجم خون در رگ‌های سطحی پوست است. این روش بر پایه همان اصول فوتوپلتیسموگرافی (PPG) عمل می‌کند، اما به جای استفاده از حسگرهای فیزیکی که مستقیماً روی پوست قرار می‌گیرند، از دوربین‌های دیجیتال و تکنیک‌های پردازش تصویر برای استخراج اطلاعات مربوط به تغییرات جریان خون بهره می‌برد.

در rPPG، نور محیط یا یک منبع نوری مصنوعی (مانند LED یا نور خورشید) به سطح پوست تابیده می‌شود. بخشی از این نور توسط پوست و بافت‌های زیرین جذب می‌شود، در حالی که بخش دیگر بازتاب می‌یابد. بازتاب نور شامل دو نوع اصلی است: بازتاب آینه‌ای که از سطح پوست بدون تغییر خاصی منعکس می‌شود و بازتاب پراکنده که وارد بافت‌های زیرین پوست شده و پس از تعامل با خون و سایر بافت‌ها، به سمت دوربین بازمی‌گردد. این بازتاب پراکنده حامل اطلاعات مهمی درباره تغییرات حجم خون در عروق سطحی است که در هر ضربه قلب رخ می‌دهد.

سیگنال rPPG از طریق پردازش و تجزیه تحلیل تغییرات رنگ پوست در طول زمان استخراج می‌شود. خون دارای هموگلوبین است که نوری با طول موج‌های خاصی را جذب و بازتاب می‌دهد. در هر سیکل ضربان قلب، مقدار خون درون رگ‌ها تغییر می‌کند و این تغییرات باعث تغییر در شدت نور بازتابی از پوست می‌شود. یک دوربین استاندارد، معمولاً با نرخ فریم بالا، از چهره یا قسمت دیگری از پوست تصویربرداری می‌کند و الگوریتم‌های پردازش تصویر این تغییرات را استخراج می‌کنند تا سیگنال rPPG تولید شود.

یکی از ویژگی‌های اصلی rPPG، توانایی آن در اندازه‌گیری تغییرات همودینامیکی بدن بدون نیاز به تماس فیزیکی است. در این روش، بخش‌هایی از پوست مانند پیشانی، گونه‌ها یا دست‌ها که بیشترین میزان جریان خون را دارند، برای استخراج سیگنال مورد استفاده قرار می‌گیرند. برای بدست آوردن یک سیگنال پایدار و دقیق، معمولاً فیلترهای پردازش تصویر برای حذف نویزهای محیطی مانند تغییرات نور، حرکات سر و بدن، و تغییرات جزئی در وضعیت چهره اعمال می‌شوند.





شکل ۲: بدست آوردن سیگنال rPPG [۲]

پردازش سیگنال rPPG معمولاً شامل مراحل مختلفی است. ابتدا، ویدئوی ضبط شده به فریم‌های جداگانه تقسیم می‌شود. سپس، نواحی حاوی اطلاعات غنی‌تر پوستی شناسایی شده و تغییرات رنگی در این نواحی طی زمان بررسی می‌شوند. برای حذف نویز، الگوریتم‌هایی مانند تحلیل مؤلفه‌های مستقل (ICA) و تحلیل مؤلفه‌های اصلی (PCA) به کار گرفته می‌شوند تا مؤلفه‌های اصلی سیگنال استخراج شوند. در نهایت، با اعمال فیلترهای دیجیتال و تکنیک‌های پردازش سیگنال، سیگنال نهایی rPPG به دست می‌آید که اطلاعات مرتبط با جریان خون را در بر دارد.

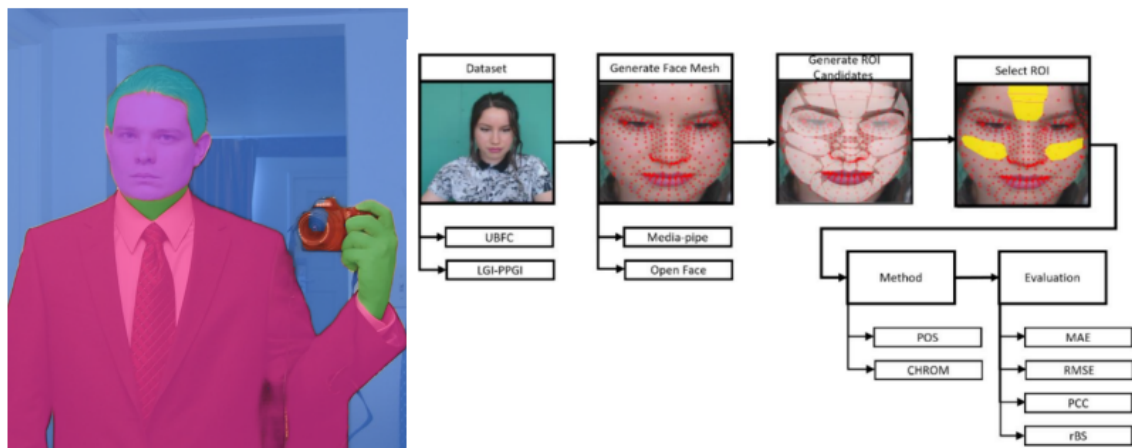
یکی از چالش‌های rPPG، حساسیت آن به شرایط نوری محیط است. تغییرات در شدت نور محیطی ممکن است بر کیفیت سیگنال اثر بگذارد، زیرا نور منبع اصلی اطلاعات در این روش است. همچنین، حرکت‌های بدن و تغییرات در وضعیت پوست (مانند تغییر زاویه صورت یا پلک زدن) ممکن است باعث ایجاد نویز در سیگنال شوند. برای کاهش این مشکلات، معمولاً روش‌های مبتنی بر مدل‌سازی آماری و یادگیری ماشین برای بهبود دقت استخراج سیگنال و کاهش اثر نویز استفاده می‌شوند که این پژوهش هم مبتنی یکی از این روش‌هاست.

تاریخچه rPPG به اوایل دهه ۲۰۰۰ بازمی‌گردد، زمانی که پژوهشگران دریافتند که تغییرات جزئی در شدت نور بازتابی از پوست می‌تواند برای استخراج اطلاعات مربوط به فعالیت قلبی-عروقی مورد استفاده قرار گیرد. با پیشرفت دوربین‌های دیجیتال و الگوریتم‌های پردازش تصویر، دقت و قابلیت‌های این روش بهبود یافت و امروزه rPPG به عنوان یک تکنیک قابل اعتماد برای اندازه‌گیری تغییرات فیزیولوژیکی بدن بدون نیاز به تماس مستقیم با پوست شناخته می‌شود [۲، ۱۱، ۵].

### ۳ چطور این سیگنال را بدست بیاوریم؟

#### ۱.۳ جمع‌آوری داده‌ها

برای استخراج سیگنال rPPG، در مرحله اول باید مجموعه‌ای از ویدئوهای چهره افراد ثبت کنیم. یکی از نکات کلیدی در این مرحله، همزمان‌سازی ضبط ویدئو با ثبت سیگنال PPG واقعی است. برای این کار، علاوه بر تصویربرداری از صورت، باید از سنسورهای فوتوپلتیسموگرافی استفاده کنیم تا سیگنال واقعی PPG همزمان با ویدئو ضبط شود. این



شکل ۳: سمت راست مدل لندکارکر. چپ مدل چند کلاسه مدیاپایپ برای استخراج نواحی پوست [۳، ۴]

همزمانی باعث می‌شود در مراحل بعد، سیگنال rPPG استخراج‌شده با داده‌های مرجع مقایسه و اعتبارسنجی شود. تحقیقات نشان داده است که نرخ فریم دوربین باید حداقل ۳۰ هرتز یا بالاتر باشد. زیرا نرخ‌های فریم پایین نمی‌توانند تغییرات سریع سیگنال PPG را ثبت کنند و باعث از دست رفتن اطلاعات مهم می‌شوند. فرکانس استاندارد ضربان قلب بین ۷۰ تا ۳ هرتز (۴۲ تا ۱۸۰ ضربان در دقیقه) است، بنابراین حداقل باید ۱۰ برابر این مقدار داده در هر ثانیه ثبت شود تا از تئوری نمونه‌برداری نایکوئیست پیروی کند.

در این پژوهش، ما از یک دوربین (Raspberry-Pi) استفاده می‌کنیم، زیرا این دوربین قابلیت تصویربرداری با نرخ فریم مناسب و کیفیت کافی را دارد. همچنین، شرایط نور محیط در این مرحله بسیار مهم است، زیرا تغییرات نور می‌تواند باعث کاهش کیفیت سیگنال نهایی شود.

## ۲.۳ پردازش ویدیو

پس از ضبط ویدئو، اولین گام در پردازش تصویر جداسازی نواحی پوست در زمان واقعی است. دلیل این کار این است که تنها نواحی پوستی صورت شامل اطلاعات مفید برای استخراج rPPG هستند. مطالعات قبلی نشان داده‌اند که سیگنال rPPG در قسمت‌های پیشانی و گونه‌ها قوی‌تر از سایر نواحی است. بنابراین، شناسایی دقیق این نواحی اهمیت زیادی دارد. در روش‌های سنتی، از روش‌های پردازش تصویر مانند آستانه‌گذاری رنگی و مدل‌های رنگی برای جداسازی پوست از پس‌زمینه استفاده شده است. اما این روش‌ها به نور محیط و رنگ پوست حساس هستند.

در این پژوهش، ما از مدل‌های مبتنی بر یادگیری عمیق برای بخش‌بندی پوست استفاده می‌کنیم. میدانیم که این سیگنال در نواحی از صورت مثل پیشانی و گونه‌ها قوی‌تر است. همچنین شرایط نوری و حرکات فرد در کیفیت سیگنال خروجی موثر هستند و هدف از این پژوهش آن است که با استفاده از مدل‌های مبتنی بر یادگیری عمیق مدل جدا کننده پوستی درست کنیم که با دادن وزن‌های مختلف به نقاط مختلف پوست سعی کند تا سیگنال خروجی نسبت به این شرایط مقاوم‌تر باشد. این روش باعث افزایش مقاومت مدل در برابر تغییرات نور و حرکات فرد می‌شود. در روش‌های قبلی، بیشتر تمرکز بر روی انتخاب نواحی خاصی از پوست مانند پیشانی بوده است که در شرایط نوری نامناسب یا حرکت زیاد عملکرد ضعیفی داشته‌اند. مدل پیشنهادی ما می‌تواند با استفاده از یادگیری توزیع پویا در قسمت‌های مختلف پوست، سیگنال خروجی را مقاوم‌تر کند.

### ۳.۳ روش‌های استخراج سیگنال

برای استخراج سیگنال rPPG، روش‌های مختلفی توسعه یافته‌اند که به دو دسته کلی تقسیم می‌شوند: روش‌های مبتنی بر معادلات جبر خطی و روش‌های مبتنی بر یادگیری عمیق. تعدادی از روش‌های مبتنی بر معادلات جبر خطی در زیر لیست شده‌اند:

۱. روش پوز: در این روش، سیگنال rPPG با استفاده از ترکیب خطی کانال‌های رنگی RGB استخراج می‌شود. فرض بر این است که تغییرات ناشی از ضربان قلب در یک صفحه عمود بر رنگ پوست قرار دارند. با استفاده از تبدیل ماتریسی، سیگنال مورد نظر استخراج می‌شود [۲].

۲. روش کروم: این روش با استفاده از کانال رنگ سبز، سیگنال rPPG را استخراج می‌کند [۱۲].

۳. روش ال جی ای: در این روش، سیگنال rPPG با استفاده از الگوریتمی که نسبت به تغییرات محلی مقاوم است، استخراج می‌شود [۵].

۴. روش امیت: این روش با استفاده از تجزیه ماتریسی، سیگنال rPPG را با ایجاد یک ماتریس متعامد با مؤلفه‌های خطی غیر همبسته استخراج می‌کند [۵].

۵. روش پی سی ای: در این روش، با استفاده از تحلیل مؤلفه‌های اصلی، سیگنال rPPG از ترکیب کانال‌های رنگی استخراج می‌شود [۱۳].

روش‌های مبتنی بر یادگیری عمیق (مانند روش دیپ-فیز) که در این روش، از شبکه توجه کانولوشنی برای استخراج سیگنال rPPG استفاده می‌شود. شبکه با استفاده از داده‌های ویدئویی آموزش دیده و قادر به استخراج سیگنال rPPG با دقت بالا است.

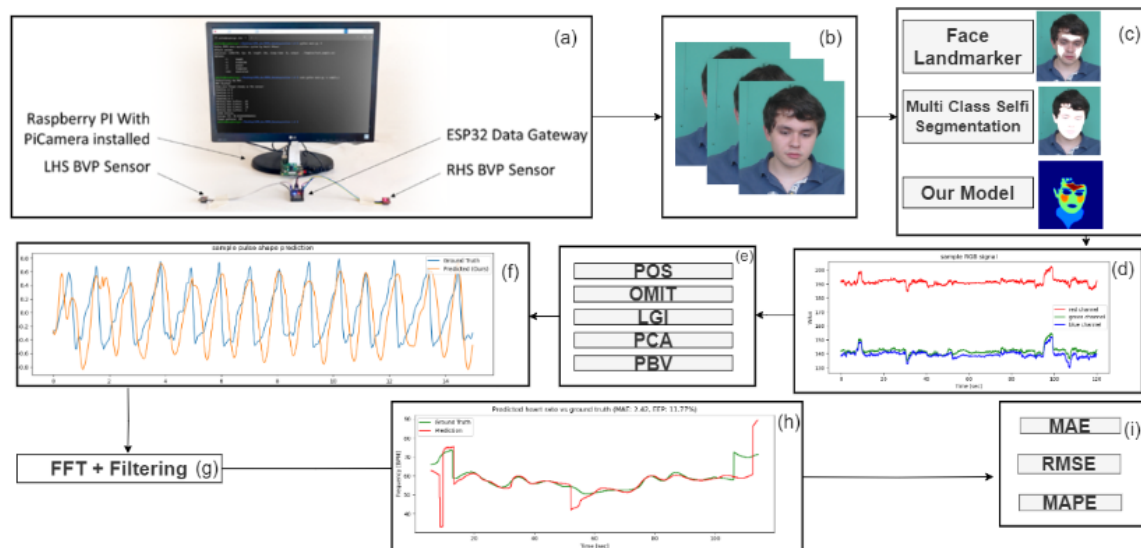
### ۴.۳ استخراج پالس قلب

پس از استخراج سیگنال rPPG خام، نیاز است که سیگنال ضربان قلب از آن استخراج شود. برای این منظور، از فیلترهای دیجیتال و روش‌های پردازش سیگنال مانند فیلترهای باند-گذر برای جداسازی مؤلفه‌های فرکانسی مرتبط با ضربان قلب استفاده می‌شود. سپس با استفاده از روش‌های تحلیل فرکانسی مانند تبدیل فوریه، فرکانس اصلی سیگنال شناسایی و نرخ ضربان قلب محاسبه می‌شود.

### ۵.۳ استخراج نرخ ضربان قلب

در این مرحله، با استفاده از سیگنال پالسی استخراج‌شده، نرخ ضربان قلب محاسبه می‌شود. برای این منظور، از روش‌های تحلیل فرکانسی مانند تبدیل فوریه سریع یا تبدیل ویولت استفاده می‌شود تا فرکانس اصلی نوسانات شناسایی شود. با محاسبه تعداد نوسانات در هر دقیقه، می‌توان نرخ ضربان قلب را به دست آورد. با پیاده‌سازی این مراحل، می‌توان سیگنال rPPG را با دقت بالا استخراج کرده و از آن برای پایش سلامت فرد استفاده کرد.

پایه‌لاین استخراج سیگنال در شکل ۴ نمایش داده شده است. (a) دریافت داده‌ها (b) جمع‌آوری مجموعه داده‌های ویدیویی هم‌زمان با سیگنال‌های PPG (فوتوپلتیسموگرافی) (c) فرایند جداسازی پوست یا انتخاب ناحیه موردنظر (ROI) در این پژوهش، دو مدل پیشرفته موجود را با روش پیشنهادی خود برای جداسازی پوست مقایسه می‌کنیم.



شکل ۴: روش کلی استخراج سیگنال

(d) استخراج سیگنال RGB با میانگین‌گیری از پیکسل‌های ناحیه پوست (e) اعمال روش‌های استخراج سیگنال rPPG (سیگنال PPG بازبازی شده از ویدیو) بر سیگنال RGB (f) مقایسه سیگنال rPPG استخراج شده با سیگنال مرجع PPG (g) تخمین ضربان قلب (h) تحلیل ضربان قلب در طول زمان (i) ارزیابی دقت تخمین ضربان قلب با استفاده از معیارهای آماری

## ۴ کاربرهای این تکنولوژی

سیگنال rPPG به دلیل قابلیت اندازه‌گیری غیر تماسی تغییرات حجم خون در پوست، کاربردهای گسترده‌ای در نظارت و پایش سلامت دارد. یکی از مهم‌ترین موارد استفاده آن، پایش وضعیت نوزادان است. در روش‌های سنتی برای بررسی علائم حیاتی نوزاد، از حسگرهای فیزیکی مانند پالس اکسیمتر یا کمربندهای نظارتی استفاده می‌شود که ممکن است باعث ناراحتی و تحریک پوست حساس آن‌ها شود. اما با استفاده از rPPG، می‌توان بدون نیاز به تماس مستقیم، تغییرات ضربان قلب و نرخ تنفس نوزاد را تحت نظر گرفت. این فناوری به والدین امکان می‌دهد تا به‌طور مداوم سلامت فرزند خود را کنترل کرده و در صورت بروز تغییرات غیرعادی، هشدارهای لازم دریافت کنند. همچنین، این روش می‌تواند در شناسایی زودهنگام سندرم مرگ ناگهانی نوزاد (SIDS) مؤثر باشد.

یکی دیگر از حوزه‌هایی که استخراج سیگنال rPPG اهمیت بالایی دارد، نظارت بر سلامت رانندگان است. تغییرات در ضربان قلب و سطح اکسیژن خون می‌تواند نشان‌دهنده خستگی، خواب‌آلودگی یا استرس راننده باشد. این اطلاعات در سیستم‌های هوشمند خودرو مورد استفاده قرار می‌گیرند تا در مواقع ضروری هشدار داده شود. مطالعات نشان داده‌اند که بسیاری از تصادفات رانندگی به دلیل کاهش سطح هوشیاری راننده رخ می‌دهند و با استفاده از rPPG می‌توان این وضعیت را زودتر تشخیص داد و از وقوع حوادث جلوگیری کرد. علاوه بر این، این روش می‌تواند برای پایش سلامت رانندگان مبتلا به بیماری‌های قلبی و پیش‌بینی حملات قلبی در حین رانندگی استفاده شود. پایش سلامت بیماران، یکی دیگر از کاربردهای مهم rPPG است. در بیمارستان‌ها، اندازه‌گیری علائم حیاتی بیماران معمولاً با استفاده از تجهیزات

متصل به بدن آن‌ها انجام می‌شود که می‌تواند محدودیت‌هایی از نظر راحتی بیمار و هزینه تجهیزات ایجاد کند. با استفاده از دوربین‌های مبتنی بر rPPG، می‌توان بدون نیاز به سیم و سنسورهای فیزیکی، ضربان قلب، سطح اکسیژن خون و نرخ تنفس بیماران را پایش کرد. این روش به‌ویژه برای نظارت بر بیماران سالمند و افرادی که در خانه تحت مراقبت هستند، مفید است. بیمارانی که دچار مشکلات حرکتی هستند یا نیاز به مراقبت طولانی‌مدت دارند، می‌توانند از سیستم‌های نظارتی مبتنی بر rPPG بهره ببرند تا وضعیت سلامت آن‌ها در هر لحظه تحت کنترل باشد.

فناوری rPPG در حوزه ورزش و تناسب‌اندام نیز مورد توجه قرار گرفته است. ورزشکاران و افراد فعال در حوزه تناسب‌اندام نیاز دارند تا ضربان قلب و وضعیت فیزیولوژیکی خود را در طول تمرین بررسی کنند. روش‌های سنتی معمولاً از حسگرهای ضربان قلب روی مچ یا سینه استفاده می‌کنند که ممکن است در برخی از فعالیت‌های ورزشی محدودیت‌هایی ایجاد کند. با استفاده از دوربین‌های rPPG، می‌توان به‌صورت غیر تماسی، میزان ضربان قلب را در طول تمرین بررسی کرد و تغییرات آن را در زمان واقعی تحلیل نمود. این داده‌ها به ورزشکاران کمک می‌کند تا شدت تمرینات خود را بهینه‌سازی کنند و از فشار بیش‌ازحد بر سیستم قلبی-عروقی خود جلوگیری کنند. همچنین، این روش می‌تواند به تحلیل میزان استرس فیزیولوژیکی و بهبود فرآیند ریکاوری پس از تمرین کمک کند. یکی از زمینه‌های نوظهور در استفاده از rPPG، تحلیل سلامت روان و پایش احساسات افراد است. مطالعات نشان داده‌اند که تغییرات در الگوی ضربان قلب می‌تواند اطلاعات ارزشمندی درباره سطح استرس، اضطراب و احساسات فرد ارائه دهد. این روش می‌تواند در محیط‌های کاری برای ارزیابی میزان استرس کارکنان، در کلاس‌های آموزشی برای بررسی تمرکز و در محیط‌های بالینی برای پایش بیماران مبتلا به اضطراب یا افسردگی استفاده شود. همچنین، با ادغام این فناوری در سیستم‌های واقعیت مجازی، می‌توان واکنش‌های فیزیولوژیکی کاربران را اندازه‌گیری کرد و تجربه کاربری را بهبود بخشید.

یکی دیگر از کاربردهای پزشکی این فناوری، کمک به تشخیص و پایش بیماری‌ها است. بیماری‌های قلبی و عروقی یکی از دلایل اصلی مرگ و میر در جهان هستند و تشخیص زودهنگام علائم حیاتی غیرعادی می‌تواند از بسیاری از مشکلات جلوگیری کند. rPPG می‌تواند برای تشخیص تغییرات غیرطبیعی در ریتم قلب و پایش بیماران مبتلا به فشار خون بالا یا آریتمی قلبی استفاده شود. همچنین، این روش در تشخیص بیماری‌های تنفسی مانند آپنه خواب مؤثر است، زیرا تغییرات در نرخ تنفس به‌وضوح در سیگنال rPPG منعکس می‌شود. علاوه بر این، این فناوری در تشخیص بیماری‌های عفونی که بر سیستم قلبی-عروقی تأثیر می‌گذارند، می‌تواند مفید باشد.

فناوری rPPG به‌طور مداوم در حال توسعه است و با پیشرفت الگوریتم‌های یادگیری عمیق، دقت و قابلیت‌های آن بهبود یافته است. امروزه بسیاری از دستگاه‌های هوشمند از جمله تلفن‌های همراه و دوربین‌های امنیتی می‌توانند این فناوری را پیاده‌سازی کنند. استفاده از شبکه‌های عصبی کانولوشنی و مدل‌های پردازش تصویر برای بهینه‌سازی استخراج سیگنال rPPG، امکان افزایش دقت در شرایط نوری مختلف و در هنگام حرکات بدن را فراهم کرده است. با رشد این تکنولوژی، انتظار می‌رود که در آینده‌ای نزدیک، پایش سلامت غیرتماسی به استاندارد در پزشکی و سلامت عمومی تبدیل شود.



# روند دستیابی به این سیگنال

## ۱ ساختن دیتاست های مطلوب

در مرحله اول برای آموزش مدل های یادگیری عمیق به یک دیتاست با تصاویر انسان و پوست نیاز داریم. چندین دیتاست عمومی برای این منظور وجود دارد، اما تعداد تصاویر آن ها بسیار کم است و تنها سناریوهای محدودی را شامل می شوند. علاوه بر این، بسیاری از این دیتاست ها از دسترس خارج شده اند یا برای اهداف خاصی طراحی شده اند که ممکن است برای ما مناسب نباشند. به همین دلیل تصمیم گرفتیم که دیتاست مخصوص به خودمان را بسازیم. برای ساخت این دیتاست از دیتاست معروف به نام های کوکو [۱۴] استفاده کردیم و تصاویر انسان ها را از این دیتاست ها استخراج کردیم.

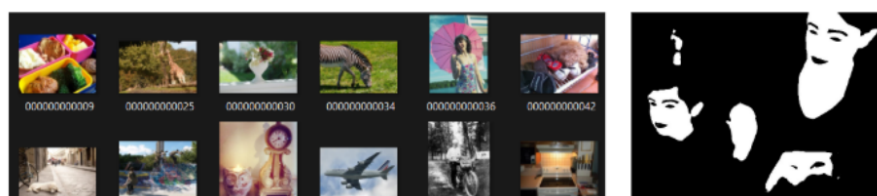
دیتاست کوکو (COCO) یکی از معروف ترین دیتاست ها برای تشخیص اشیاء است که شامل تصاویری با اشیاء مختلف از جمله انسان ها در موقعیت های مختلف می باشد. این دیتاست شامل تصاویری است که در آن ها انسان ها با پس زمینه های مختلف و در موقعیت های طبیعی حضور دارند. این ویژگی ها به ما این امکان را می دهند که انسان ها را در شرایط مختلف از نظر پس زمینه و نور استخراج کنیم.

برای استفاده از این دیتاست، انسان ها را از تصاویر آن ها استخراج کردیم و سپس با استفاده از مدل های مدیاپایپ که در فصل بعد به تفصیل توضیح داده می شود، پردازش های لازم را انجام دادیم. در ابتدا از مدل جداسازی سلفی چند کلاسه (Selfie (Multi-Class Segmentation) استفاده کردیم تا پوست بدن و صورت انسان ها را جدا کنیم. این مدل به ما این امکان را می دهد که بخش های مختلف بدن و صورت را با دقت جدا کنیم. سپس از مدل لندمارکر استفاده کردیم تا نواحی مختلف صورت مانند چشم ها و دهان را استخراج کنیم. با استفاده از این ابزار، توانستیم انوتیشن نواحی خاص مانند پیشانی و گونه ها را انجام دهیم و به هر ناحیه وزن های متفاوتی اختصاص دادیم. این کار به ما کمک کرد تا در فرآیند یادگیری، نواحی مانند پیشانی و گونه ها که معمولاً سیگنال های قوی تری دارند، تأثیر بیشتری در مدل داشته باشند. [۴]

پس از تولید این دیتاست، آن را برای آموزش مدل های یادگیری عمیق استفاده کردیم تا بتوانیم سیگنال های مورد نظر را با دقت بالاتری استخراج کنیم.

دیتاست بعدی که برای آموزش و ارزیابی مدل استفاده کردیم، دیتاست ویدیویی همزمان با سیگنال استخراج شده

COCO:



شکل ۱: درست کردن دیتاست جهت یادگیری مدل

از حسگر است. برای تولید این دیتاست از دوربین (Raspberry Pi) استفاده کردیم که همزمان با حسگرهای ثبت کننده سیگنال ضربان قلب، اطلاعات را ثبت می کند. این دیتاست شامل ویدئوهایی است که در سناریوهای مختلف ثبت شده اند، از جمله:

- فرد در حالت ساکن
- فرد در حال صحبت کردن
- فرد در حال حرکت دادن صورت
- فرد در حال ورزش کردن

این ویدئوها به ما این امکان را می دهند که فرآیند استخراج سیگنال های فیزیولوژیکی از تغییرات صورت و پوست را به طور دقیق بررسی و آزمایش کنیم. تمامی این ویدئوها همزمان با سیگنال های PPG واقعی از حسگر ثبت شده اند که می توانند به عنوان مرجع برای ارزیابی عملکرد مدل استفاده شوند. علاوه بر این، دیتاست های آماده دیگری نیز با تنوع محدود و تعداد داده های کم وجود دارند که مدل ما بر روی آن ها نیز تست شده است. این دیتاست ها شامل ویدئوهایی با شرایط متفاوت و محدود هستند که به ما کمک می کنند تا مدل را در شرایط مختلف ارزیابی کنیم و عملکرد آن را در دنیای واقعی بسنجیم. تمامی این دیتاست ها با توجه به نیازهای خاص پروژه جمع آوری و پردازش شده اند تا بهترین عملکرد را در استخراج سیگنال های rPPG از ویدئوها به دست آوریم.

## ۲ روش های بدست آوردن سیگنال از تصویر

استخراج سیگنال rPPG از ویدئو های ثبت شده نیازمند روش های پردازش سیگنال و بینایی کامپیوتری است که بتوانند اطلاعات ضربان قلب را از تغییرات ظریف رنگ پوست استخراج کنند.

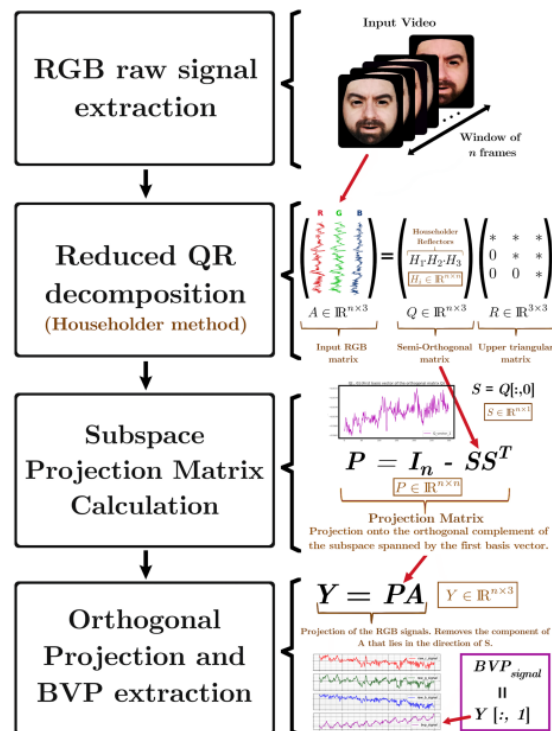
### ۱.۲ روش امیت در استخراج سیگنال

این روش جدید که از تجزیه ماتریسی QR برای تبدیل سیگنال های خام RGB به سیگنال های rPPG استفاده می کند. این روش به طور خاص مقاوم در برابر نویزهای ناشی از فشرده سازی ویدئو است. با استفاده از این تبدیل، سیگنال BVP به صورت بهینه از تغییرات رنگ پوست استخراج می شود [۵].

### ۲.۲ روش کروم

این روش مبتنی بر مدل بازتاب دوکروماتیک است که سیگنال rPPG را از بازتاب پخش شده (Diffuse-Reflection) پوست استخراج می کند، در حالی که اثرات بازتاب آینه ای (Specular-Reflection) که حاوی اطلاعات ضربانی نیست، حذف می شود. در این روش، با استفاده از بردار رنگ پوست استاندارد، تعادل رنگی انجام شده و تغییرات کرومینانسی استخراج (Chrominance) می شود [۱۲].





شکل ۲: مرحله به مرحله استخراج سیگنال ضربان از سیگنال RGB طبق روش امیت [۵]

## ۳.۲ روش پوز

این روش از یک صفحه‌ی متعامد با بردار رنگ پوست در فضای RGB استفاده می‌کند تا سیگنال ضربان را استخراج کند. این روش با فرض اینکه تغییرات ناشی از PPG در جهتی قرار دارند که متعامد بر رنگ پوست است، یک فیلتر خطی اعمال می‌کند [۲].

## ۴.۲ روش فیزفورمر

این روش از مدل‌های یادگیری عمیق برای استخراج سیگنال rPPG استفاده می‌کند. برخلاف روش‌های کلاسیک که مبتنی بر پردازش سیگنال و تبدیل‌های ریاضی هستند، این روش با استفاده از شبکه‌های عصبی کانولوشنی (CNN) و مدل‌های مبتنی بر یادگیری خودنظارتی، مستقیماً سیگنال ضربان را از ویدئو استخراج می‌کند.

این روش مبتنی بر ترنسفورمر است و از معماری‌های توجه درونی (Self-Attention) برای تحلیل سیگنال‌های زمانی استفاده می‌کند. این مدل نسبت به نویزهای حرکتی و تغییرات نوری مقاوم‌تر بوده و می‌تواند ارتباطات بلند مدت را در سری‌های زمانی rPPG استخراج کند.

---

**Algorithm 1:** Plane-Orthogonal-to-Skin.

---

**Input:** A video sequence containing  $N$  frames

```
1: Initialize:  $\mathbf{H} = \text{zeros}(1, N)$ ,  $l = 32$  (20 frames/s camera)
2: for  $n = 1, 2, \dots, N$  do
3:    $\mathbf{C}(n) = [R(n), G(n), B(n)]^\top \leftarrow$  spatial averaging
4:   if  $m = n - l + 1 > 0$  then
5:      $\mathbf{C}_n^i = \frac{\mathbf{C}_{m \rightarrow n}^i}{\mu(\mathbf{C}_{m \rightarrow n}^i)} \leftarrow$  temporal normalization
6:      $\mathbf{S} = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{pmatrix} \cdot \mathbf{C}_n \leftarrow$  projection
7:      $\mathbf{h} = \mathbf{S}_1 + \frac{\sigma(\mathbf{S}_1)}{\sigma(\mathbf{S}_2)} \cdot \mathbf{S}_2 \leftarrow$  tuning
8:      $\mathbf{H}_{m \rightarrow n} = \mathbf{H}_{m \rightarrow n} + (\mathbf{h} - \mu(\mathbf{h})) \leftarrow$  overlap-adding
9:   end if
10: end for
Output: The pulse signal  $\mathbf{H}$ 
```

---

شکل ۳: الگوریتم پوز [۲].

## ۳ روش های استخراج ضربان قلب

روش های استخراج ضربان قلب از سیگنال های PPG یا rPPG از جمله مهم ترین ابزارها در تجزیه و تحلیل وضعیت فیزیولوژیکی افراد هستند. یکی از روش های رایج برای استخراج ضربان قلب، شناسایی قله ها در سیگنال است. در این روش، ابتدا سیگنال PPG پردازش می شود تا نویزهای مربوط به حرکت و تغییرات نور محیط کاهش یابد. سپس، با استفاده از الگوریتم های شناسایی قله، نقاط اوج سیگنال که نشان دهنده ضربان قلب هستند، شناسایی می شوند. این قله ها معمولاً با استفاده از مشتق اول و دوم سیگنال یا الگوریتم هایی برای شناسایی نقاط بیشینه شناسایی می شوند. با شمارش تعداد این قله ها در واحد زمان، می توان نرخ ضربان قلب (BPM) را محاسبه کرد. این روش به دلیل سادگی و سرعت، بسیار محبوب است، اما در شرایطی که سیگنال تحت تأثیر حرکت های شدید یا تغییرات نوری قرار گیرد، دقت آن کاهش می یابد.

روش دیگر برای استخراج ضربان قلب، استفاده از تبدیل فوریه سریع (FFT) است. در این روش، سیگنال PPG به دامنه فرکانس تبدیل می شود و فرکانس غالب که معمولاً در محدوده ۷۰.۰ تا ۳ هرتز قرار دارد، به عنوان نشان دهنده ضربان قلب شناسایی می شود. با این روش می توان سیگنال های پیچیده تر را تحلیل کرد و نویزهایی که در سیگنال وجود دارند را کاهش داد. پس از اعمال FFT، طیف فرکانسی به دست آمده نشان دهنده تغییرات ضربانی است که می تواند برای محاسبه ضربان قلب مورد استفاده قرار گیرد. این روش از دقت بالاتری نسبت به روش شناسایی قله ها برخوردار است و می تواند سیگنال هایی که تحت تأثیر نویز و حرکت قرار دارند را بهتر تجزیه و تحلیل کند. البته نیاز به محاسبات بیشتری دارد و برای سیگنال های پیچیده تر ممکن است به تنظیمات خاص نیاز داشته باشد. روش های مبتنی بر یادگیری عمیق در حال حاضر یکی از پیشرفته ترین تکنیک ها برای استخراج ضربان قلب هستند. این روش ها از مدل های پیچیده تری مانند شبکه های عصبی حافظه بلندمدت (LSTM) برای تحلیل سیگنال های rPPG استفاده

می‌کنند. LSTM برای تحلیل داده‌های زمانی و نوسانات ضربانی مناسب است. این مدل‌ها به طور خودکار ویژگی‌های پیچیده را از داده‌های ورودی استخراج می‌کنند و می‌توانند سیگنال‌های rPPG را از تغییرات بسیار ظریف رنگ پوست شبیه‌سازی کنند. یکی از مزایای این روش‌ها، دقت بالای آن‌ها در شرایط محیطی پیچیده مانند حرکت سریع یا تغییرات نوری است. همچنین، این روش‌ها نیازی به پردازش‌های دستی ویژگی‌ها ندارند و قادر به یادگیری از داده‌های بزرگ و پیچیده هستند. با این حال، نیاز به داده‌های آموزشی زیاد و منابع محاسباتی بالا از معایب این روش‌ها محسوب می‌شود.

در نهایت، انتخاب روش مناسب برای استخراج ضربان قلب به شرایط خاص سیگنال و منابع موجود بستگی دارد. برای سیگنال‌های ساده‌تر با نویز کم، روش‌های شناسایی قله‌ها و FFT مؤثر هستند، اما در شرایط پیچیده‌تر، روش‌های یادگیری عمیق می‌توانند بهترین عملکرد را ارائه دهند.



# متد های پردازش فیلم

## ۱ روش های کلاسیک

### ۱.۱ روش چرف

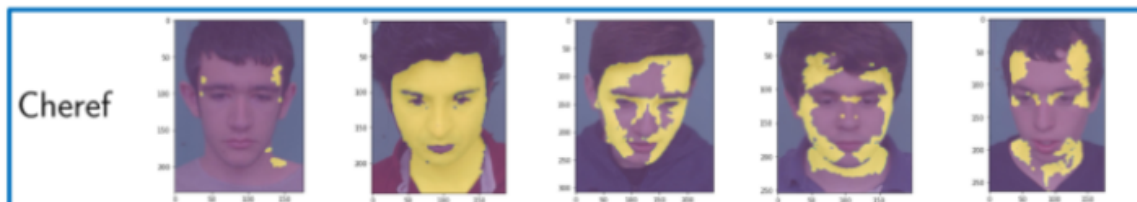
روش چرف (Cherref) برای استخراج پوست به ترکیب آستانه گذاری در چندین فضای رنگی می پردازد. این روش ویژگی های خاص رنگ پوست انسان را در رنگ های مختلف بررسی کرده و با استفاده از آستانه گذاری مناسب در چندین فضای رنگی، نواحی پوست را شناسایی می کند. این روش یکی از ساده ترین روش ها برای شناسایی پوست است، چرا که بر اساس تغییرات رنگی پوست در تصاویر عمل می کند [۶].

### ۲.۱ روش لولست

روش لولست (Levelset) برای شناسایی پوست به ترکیب آستانه گذاری رنگی با تقسیم بندی سطحی (Level set segmentation) و ردیابی می پردازد. در این روش، ابتدا آستانه گذاری رنگی برای شناسایی نواحی پوست اعمال می شود، سپس الگوریتم set Level برای بهبود دقت و افزایش توانایی در ردیابی حرکت یا تغییرات در تصویر استفاده می شود. این روش می تواند با دقت بیشتری نواحی پوست را شبیه سازی کند و در برابر تغییرات جزئی در موقعیت صورت یا بدن مقاوم تر است. از ویژگی های این روش، توانایی آن در ردیابی نواحی پوست به صورت دینامیک در ویدئوهای متحرک است. [۷]

### ۳.۱ مشکلات روش های چرف، لولست و الگوریتم های آستانه گذاری رنگی

هر یک از روش های استخراج پوست که در تصاویر مطرح شده اند، با چالش ها و مشکلات خاص خود روبه رو هستند. یکی از مشکلات مشترک در روش چرف، حساسیت به تغییرات نوری محیطی است. این روش که بر اساس آستانه گذاری رنگی در فضا های رنگی مختلف عمل می کند، در شرایط نوری متفاوت مانند نور ضعیف یا تغییرات ناگهانی نور محیط



شکل ۱: خروجی روش چرف [۶].



Example skin masks for UBFC dataset



Example skin masks for BP4D dataset

شکل ۲: خروجی روش لولست [۷].

نمی‌تواند به درستی پوست را شناسایی کند. همچنین، چون این روش ساده است، در شرایط پیچیده که رنگ پوست با پس‌زمینه مشابه باشد، دقت شناسایی کاهش می‌یابد و احتمال اشتباه در استخراج پوست بیشتر می‌شود.

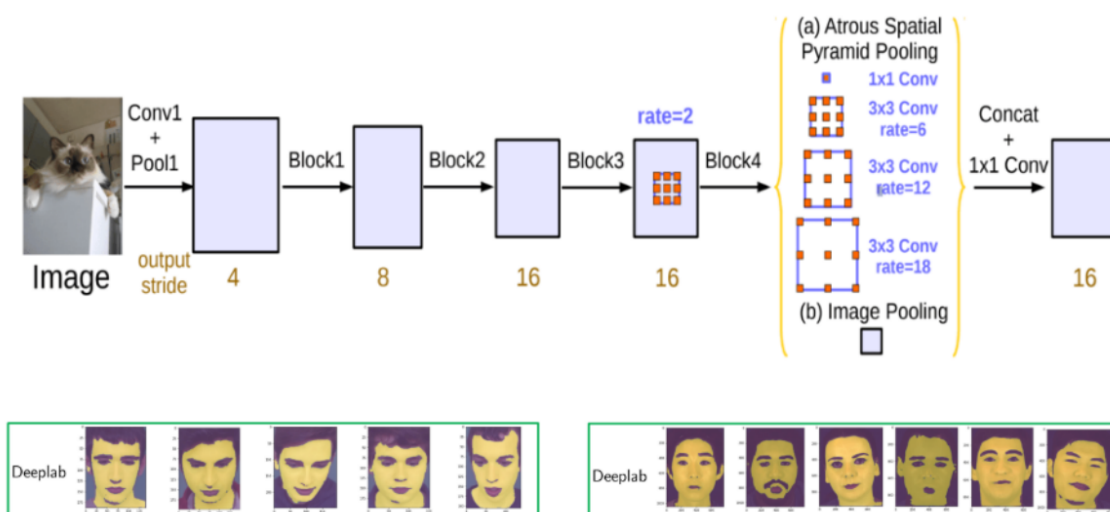
در روش لولست که از ترکیب آستانه‌گذاری رنگی و تقسیم‌بندی سطحی استفاده می‌کند، مشکلات عمده شامل پیچیدگی پردازشی بالا است. این روش نیاز به محاسبات بیشتری دارد و زمان‌بر است، که باعث می‌شود در سیستم‌های با منابع محاسباتی محدود عملکرد مناسبی نداشته باشد. علاوه بر این، این روش حساس به حرکات شدید صورت یا بدن است. در شرایطی که فرد حرکت شدیدی داشته باشد یا تغییرات ناگهانی در موقعیت صورت رخ دهد، ممکن است مدل قادر به شناسایی دقیق نواحی پوست نباشد. همچنین، تنظیم دقیق پارامترهای الگوریتم مانند انتخاب آستانه‌ها و مقادیر مربوط به تقسیم‌بندی سطحی برای کارایی بهتر، به زمان و تلاش زیادی نیاز دارد.

روش الگوریتم‌های آستانه‌گذاری رنگی که برای استخراج پوست در تصاویر و ویدئوها استفاده می‌شود، با چالش‌های حساسیت به تغییرات نور و پس‌زمینه‌های پیچیده روبه‌رو است. تغییرات در شدت نور و یا وجود پس‌زمینه‌هایی با رنگ مشابه پوست می‌تواند به‌طور قابل توجهی دقت شناسایی پوست را کاهش دهد. علاوه بر این، این روش‌ها معمولاً از آستانه‌گذاری‌های ثابت استفاده می‌کنند که برای همه شرایط نوری و محیطی مناسب نیستند. به‌ویژه در موقعیت‌هایی که رنگ پوست فرد مشابه رنگ‌های موجود در پس‌زمینه یا اجسام اطراف است، این روش‌ها می‌توانند نواحی پوست را به اشتباه شناسایی کنند. در نهایت، مشکلات دیگری مانند انتخاب دقیق آستانه‌ها برای هر فضای رنگی (HSV، RGB، YCbCr) وجود دارد که نیازمند تنظیمات خاص و آزمایش‌های دقیق است تا بهترین نتایج حاصل شود.

## ۲ روش‌های یادگیری عمیق

### ۱.۲ روش دیپلب

دیپلب یکی از مدل‌های قدرتمند بخش‌بندی تصویر در یادگیری عمیق است که برای شناسایی دقیق نواحی مختلف یک تصویر استفاده می‌شود. این مدل به‌طور ویژه برای بخش‌بندی اشیاء و ویژگی‌های خاص تصاویر طراحی شده و در زمینه استخراج ویژگی‌های صورت و پوست انسان در تصاویر و ویدئوها کاربرد دارد.



شکل ۳: شماتیکی از مدل دیپلب [۸].

مدل دیپلب از شبکه‌های کانولوشنی (CNN) برای تشخیص و استخراج ویژگی‌ها از تصویر استفاده می‌کند. یکی از ویژگی‌های برجسته این مدل، استفاده از Convolution Atrous (که به آن Convolution Dilated نیز گفته می‌شود) است. این نوع کانولوشن به مدل این امکان را می‌دهد که ویژگی‌ها را از مناطق بزرگتر تصویر استخراج کند بدون اینکه وضوح تصویر کاهش یابد. این فرآیند به مدل این اجازه را می‌دهد که بتواند ویژگی‌های متعدد و مقیاس‌پذیر را از تصویر جمع‌آوری کند. برای تقویت دقت و قابلیت استخراج ویژگی‌های چند مقیاسی، مدل از Pyramid Spatial Pooling استفاده می‌کند. این روش به‌طور خاص برای تصاویر با مقیاس‌های مختلف و با ویژگی‌های متفاوت از هر مقیاس طراحی شده است.

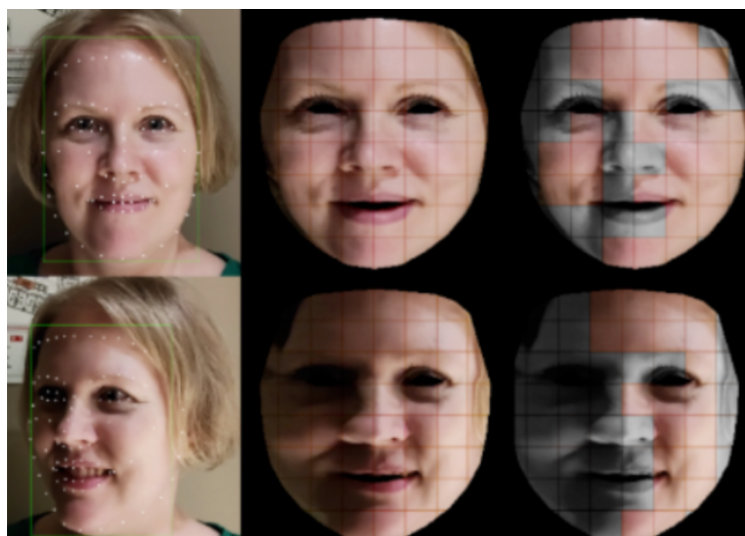
این مدل ابتدا تصویر ورودی را به لایه‌های مختلف کانولوشن و سپس به لایه‌های Pyramid Spatial Atrous منتقل می‌کند که در این لایه‌ها، ویژگی‌های تصویری با نرخ‌های مختلف برای استخراج اطلاعات چند مقیاسه پردازش می‌شود. سپس ویژگی‌های مختلف به یک لایه کاملاً متصل شده متصل می‌شوند و نتیجه نهایی در قالب بخش‌بندی به مدل داده می‌شود [۸].

این تحقیق نشان داده که با داشتن دیتاست مناسب این مدل خروجی خوبی می‌دهد و مشکل نبود دیتاست مناسب با عکس‌های کافی برای آموزش این شبکه است [۱۵].

## ۲.۲ روش مدیا پایپ

مدیاپایپ (MediaPipe) ابزاری است که توسط گوگل برای پردازش داده‌های ویدئویی و تصویری در زمان واقعی طراحی شده است. این ابزار به‌ویژه در شناسایی ویژگی‌های صورت و ردیابی آن‌ها در ویدئوهای زنده کاربرد دارد. مدل لندمارکر یکی از زیرمجموعه‌های مدیاپایپ است که برای شناسایی نقاط کلیدی صورت به کار می‌رود.

مدل لندمارکر در مدیاپایپ با استفاده از الگوریتم‌های ردیابی ویژگی‌های صورت، اقدام به شناسایی نقاط کلیدی در صورت می‌کند. این نقاط معمولاً شامل قسمت‌هایی مانند چشم‌ها، بینی، دهان، پیشانی، و خط فک است. این مدل می‌تواند این نقاط را در ویدئوهای زنده یا تصاویر ثابت شناسایی کند. مدل لندمارکر از شبکه‌های عصبی پیشرفته



شکل ۴: خروجی روش امیت [۵]

و الگوریتم‌های ردیابی برای شناسایی و ردیابی دقیق این نقاط استفاده می‌کند. مدیپایپ قادر است در زمان واقعی ویژگی‌های صورت را شناسایی کرده و تغییرات آن‌ها را در شرایط مختلف محیطی و نوری ردیابی کند. این مدل از خروجی‌های دقیق برای شناسایی ویژگی‌های صورت در محیط‌های متحرک و متنوع استفاده می‌کند [۳].

مدل جداسازی سلفی چندکلاسه (Selfie (Multi-Class Segmentation) یکی دیگر از مدل‌های مدیپایپ است که برای تقسیم‌بندی و شناسایی نواحی مختلف صورت و بدن در تصاویر و ویدئوها کاربرد دارد. این مدل به‌طور خاص برای تفکیک پس‌زمینه و بخش‌های مختلف صورت طراحی شده است. در این مدل، مدیپایپ از آستانه‌گذاری‌های چند کلاسه استفاده می‌کند. در این مدل تصویر به شش بخش پس‌زمینه، مو، پوست صورت، پوست بدن، لباس و اشیاء تقسیم می‌شود. این مدل علاوه بر اینکه می‌تواند. همچنین، این مدل از شبکه‌های عصبی پیشرفته برای بخش‌بندی دقیق و ردیابی ویژگی‌های صورت در هر فریم از ویدئو استفاده می‌کند. ما از این مدل می‌توانیم استفاده کنیم تا پوست صورت و بدن را بدست آوریم اما این مدل قادر به جداسازی چشم‌ها و دهان و وزن دهی متفاوت به مکان‌های مختلف پوست نیست [۴].

## ۳.۲ روش امیت

روش امیت برای شناسایی مناطق مهم صورت در تصاویر استفاده می‌شود. این مدل با استفاده از آستانه‌گذاری‌های نوری و ویژگی‌های شدت نور، نواحی مناسب برای شبیه‌سازی پوست را شناسایی می‌کند. در این روش، از آستانه‌گذاری نوری و تحلیل ویژگی‌های شدت نور (L) و زاویه دید استفاده می‌شود. این روش به‌طور ویژه برای تصاویر صورت طراحی شده است و می‌تواند نواحی مهم صورت را از نواحی غیر مفید یا پس‌زمینه جدا کند. این مدل از آستانه‌گذاری‌های کنتراست و نور استفاده می‌کند تا بهترین مناطق برای تحلیل ویژگی‌های صورت انتخاب شوند.

این روش نسبت به شرایط نوری مقاومت بیشتری دارد ولی از نواحی کمی سیگنال استخراج می‌شود. همچنین تنها از پوست صورت استفاده می‌شود و پوست دست و گردن که سیگنال در آن بسیار قوی است نادیده گرفته می‌شود [۵].



### ۳ روش ارائه شده این پژوهش

متوجه شدیم که عملکرد سیستم‌های استخراج سیگنال rPPG در محیط‌های با نور ضعیف به طور قابل توجهی کاهش می‌یابد. در چنین شرایطی، توانایی شناسایی تغییرات در رنگ پوست که مرتبط با ضربان قلب است، محدود می‌شود. همچنین، تأثیر تغییرات نور محیط یکی دیگر از مشکلات اصلی است. نور محیطی می‌تواند سیگنال‌های مربوط به ضربان قلب را تحت تأثیر قرار داده و آن‌ها را محو یا مخفی کند، که این امر باعث کاهش دقت در تشخیص ضربان قلب می‌شود. این مشکلات به‌ویژه در محیط‌هایی که نور ثابت یا قابل پیش‌بینی ندارند، مثل فضای داخلی یا مکان‌های بیرونی با نور طبیعی، مشهودتر هستند.

برای آموزش مدل‌های استخراج سیگنال rPPG نیاز به دیتاست‌های تصویری با کیفیت است. این دیتاست‌ها باید به‌طور کافی شامل تصاویر مختلف از چهره و پوست باشند تا مدل‌ها بتوانند ویژگی‌های مختلف صورت را شناسایی کنند. علاوه بر دیتاست‌های تصویری، دیتاست‌های ویدیویی نیز برای ارزیابی و آزمایش کامل سیستم‌های rPPG ضروری هستند. ویدئوها این امکان را می‌دهند که مدل در شرایط واقعی و در زمان واقعی عملکرد خود را نشان دهد. این داده‌ها باید شامل ویدئوهایی از افراد در شرایط مختلف و با رفتارهای متنوع باشند تا مدل بتواند در تحلیل تغییرات ضربان قلب در محیط‌های پویا و متغیر به‌طور مؤثر عمل کند.

با این حال، دیتاست‌های موجود محدودیت‌هایی دارند که می‌توانند عملکرد مدل‌های یادگیری عمیق را تحت تأثیر قرار دهند. در دیتاست‌های تصویری، یکی از مشکلات عمده عدم وجود ماسک‌های چندکلاسه است که برای شناسایی بخش‌های مختلف صورت مانند پیشانی، گونه‌ها، دهان و چشم‌ها ضروری هستند. علاوه بر این، بسیاری از این دیتاست‌ها داده‌های برچسب‌گذاری شده کافی برای آموزش مدل‌ها ندارند که باعث می‌شود مدل نتواند ویژگی‌ها را با دقت بالا یاد بگیرد. در دیتاست‌های ویدیویی نیز مشکلات دیگری وجود دارد. از جمله اینکه این دیتاست‌ها اغلب تنها شرایط نوری خاصی را پوشش می‌دهند و در سناریوهای با نورپردازی سخت یا حرکات سریع صورت، دقت مدل به شدت کاهش می‌یابد. علاوه بر این، یکی از مشکلات اصلی در این دیتاست‌ها، عدم همزمانی بین سیگنال‌های PPG و ویدئو است، که باعث می‌شود نتایج استخراج شده از مدل‌ها دقت کافی را نداشته باشد. یکی از چالش‌های اصلی در پردازش ویدیویی سیگنال‌های rPPG، گسترش بخش‌بندی تصویر به تحلیل ویدیوهای زمان واقعی است. در سیستم‌های ویدیویی، ماسک‌ها باید قادر باشند در وضعیت‌های مختلف و با تغییرات سریع در زاویه دید یا نور به‌طور مداوم به روز شوند. ما علاقه مند هستیم ماسک‌ها به‌صورت نقشه‌های حرارتی پیوسته نمایش داده شوند تا بتوانند به‌طور دقیق و پیوسته به تغییرات پاسخ دهند.

رویکرد ما در این پروژه بر پایه استفاده از یک مدل مبتنی بر دیپلبر برای بخش‌بندی دقیق پوست و استخراج سیگنال‌های rPPG در زمان واقعی استوار است که با استفاده از دیتاست تصویری تولید شده توسط خودمان (که به دست آوردن ماسک آنان از دو مدل مدیاپایپ همزمان کمک گرفته میشود) آموزش داده میشود. برای آموزش و ارزیابی مدل، دیتاست‌های ویدیویی-PPG همزمان جمع‌آوری شده‌اند که در آن‌ها شرایط مختلف نورپردازی و حرکت و ضربان قلب بالا در نظر گرفته شده است. این دیتاست‌ها به ما کمک کردند تا مدل‌های خود را تحت شرایط واقعی آزمایش کنیم و عملکرد آن‌ها را در محیط‌های متغیر بررسی کنیم.

در مرحله استخراج سیگنال، از ماسک‌های پیوسته با وزن دهی متفاوت به بخش‌های مختلف استفاده شده است تا بتوان سیگنال‌های rPPG را به‌طور مستمر و دقیق و مقاوم به شرایط در طول فیلم‌ها استخراج کرد. به‌علاوه، برای بهبود دقت و پایداری استخراج سیگنال‌های ضربان قلب، از الگوریتم‌های استخراج سیگنال مثل پوز بهره برده‌ایم که قادر است به‌طور مؤثر تغییرات رنگ پوست را شبیه‌سازی کرده و سیگنال rPPG را استخراج کند.



## نتایج

ما پایپ لایینی که در شکل ۴ نشان داده شده است را روی سه مدل جداسازی پوست مورد بررسی قرار می‌دهیم: شامل مدل پیشنهادی ما برای جداسازی پوست کل بدن، مدل چند کلاسه مدیپایپ و مدل لندمارکر این مقایسه بر روی مجموعه داده خودمان، UBFC<sup>۱</sup>، UBFC<sup>۲</sup>، UBFC-PHYS (همان تعداد نمونه موجود در مجموعه داده ما) انجام شده است. تمامی آزمایش‌ها بر روی ۳۰ ثانیه ابتدایی ویدیوها صورت گرفته‌اند. از آنجا که ویدیوهای موجود در مجموعه داده ما ۳۰ ثانیه طول دارند، این مدت‌زمان به صورت یکنواخت در تمام ارزیابی‌ها حفظ شده تا شرایط مقایسه عادلانه باشد.

### ۱ پیکربندی آزمایش

ما یک مجموعه داده جدید rPPG ارائه می‌دهیم. بیشتر مجموعه داده‌های موجود تحت شرایط نوری و محیطی ایده‌آل و با حداقل حرکت سوژه ثبت شده‌اند، که لزوماً بازتاب‌دهنده کاربردهای واقعی نیستند. بر اساس این ملاحظات، ما طراحی و پیاده‌سازی یک دستگاه نمونه‌برداری اختصاصی را ضروری دانستیم؛ دستگاهی که اطمینان حاصل کند داده‌های تصویری و سیگنال‌های ضربان قلب به صورت دقیق و هم‌زمان ثبت می‌شوند. نمایی کلی از پیکربندی این سیستم در بخش (a) از شکل ۴ آمده است.

ما از برد توسعه‌ای (Raspberry-Pi-4B) استفاده کرده‌ایم که دارای پردازنده ۶۴-بیتی با فرکانس ۱/۵ گیگاهرتز و حافظه رم به میزان ۸ گیگابایت است. این سیستم با سیستم عامل (Raspberry-Pi-OS) اجرا شده و از زبان برنامه‌نویسی پایتون برای توسعه سریع و یکپارچه‌سازی آسان بهره می‌برد. برای تصویربرداری، ما از ماژول دوربین (V2-Raspberry-Pi) استفاده کرده‌ایم که تصاویر را با وضوح ۷۲۰×۱۲۸۰ پیکسل و نرخ ۳۰ فریم بر ثانیه (fps) ثبت می‌کند.

برای دریافت داده‌های ضربان قلب، ما از سنسور (MAX۳۰۱۰۲) با کیفیت آزمایشگاهی بهره برده‌ایم که از طریق پروتکل (I2C) ارتباط برقرار می‌کند. با توجه به نبود کتابخانه‌ی پایتون قدرتمند برای این سنسور در رزبری‌پی، ما یک برد توسعه‌ای (ESP۳۲) را به عنوان واسطه به کار گرفته‌ایم. این برد که با زبان سی و توسط آردوینو برنامه‌نویسی شده، نقش یک پل ارتباطی را ایفا می‌کند که داده‌های سنسور را از طریق ارتباط سریال به رزبری‌پی منتقل می‌نماید. علاوه بر این، برای افزایش دقت اندازه‌گیری و کاهش خطاهای احتمالی در ثبت سیگنال ضربان قلب، یک سنسور (MAX۳۰۱۰۲) اضافی نیز در سیستم گنجانده شده تا هم‌زمان داده‌های ضربان قلب از هر دو دست دریافت شوند. هر سنسور (MAX) از چهار پین در (ESP۳۲) استفاده می‌کند، و این برد قادر است دو ارتباط (I2C) را به صورت هم‌زمان و بدون تداخل پشتیبانی کند.

فرآیند نمونه‌برداری به صورت دستی هماهنگ شده و با نرخ فریم ثابت اجرا می‌شود تا اختلاف‌های زمانی احتمالی بین دوربین و سنسورها برطرف شود. دوربین و سنسورها دقیقاً با نرخ ۳۰ فریم بر ثانیه همگام شده‌اند. با توجه به حجم بالای داده‌ها، سیستم ما از معماری چندرسمانی (multi-threaded) برای مدیریت هم‌زمان

داده‌های تصویری و سنسور استفاده می‌کند. پس از ثبت، فریم‌های دوربین به یک ریشه اختصاصی منتقل می‌شوند که مسئول فشرده‌سازی و ذخیره‌سازی ویدیوهاست. این پردازش موازی سبب می‌شود ریشه اصلی آزاد باقی بماند تا بتواند بدون وقفه فریم‌های بعدی را ثبت کند. نمونه‌های دریافتی از سنسور ضربان قلب که نسبتاً حجم کمی دارند، در طول زمان ضبط در حافظه رم ذخیره شده و پس از پایان ضبط، در فایل‌هایی متناظر با ویدیوها روی دیسک ذخیره می‌گردند. نرخ فریم قابل‌دستیابی در این سیستم حداکثر ۳۰ fps است که به واسطه محدودیت‌های ذاتی ماژول دوربین تعیین می‌شود. این طراحی، یک جریان داده‌ی پایدار و با نرخ بالا را تضمین می‌کند که برای تحلیل دقیق rPPG حیاتی است.

## ۲ مجموعه داده‌ها

ما روش پیشنهادی خود را با استفاده از مجموعه داده معرفی شده مان و همچنین دو مجموعه داده عمومی دیگر مورد ارزیابی قرار می‌دهیم: UBFC-rPPG [۱۶] و UBFC-PHYS [۱۷]. این مجموعه داده‌ها شامل ویدیوهای ضبط شده همراه با داده‌های فیزیولوژیکی مرجع برای ارزیابی هستند.

مجموعه داده ما با نام SYNC-rPPG معرفی می‌شود. داده‌ها از ۲۰ فرد ۱۸ تا ۲۵ ساله جمع‌آوری شده است و طول هر ویدیو ۳۰ ثانیه است. تمامی افراد شرکت‌کننده رضایت خود را برای در دسترس عموم قرار گرفتن داده‌هایشان اعلام کرده‌اند. هر فرد در چهار سناریوی متفاوت شرکت کرده است:

در سناریوی اول، فرد در حالت آرام قرار دارد، بدون حرکت سر و با حداقل حالات چهره.  
در سناریوی دوم، از فرد خواسته می‌شود یک متن احساسی را با صدای بلند بخواند یا در مورد یک خاطره مهم صحبت کند.

در سناریوی سوم، فرد حرکات سریع چرخشی سر را انجام می‌دهد.  
در سناریوی چهارم، سه دقیقه پس از انجام تمرین ورزشی، ضبط انجام می‌شود (مشابه سناریوی اول).  
در مجموع، این مجموعه داده شامل ۸۰ ویدیو است که هر کدام ۳۰ ثانیه طول دارند.  
پایگاه داده UBFC-rPPG [۱۶] با استفاده از وب‌کم (Logitech-C۹۲۰-HD-Pro) با نرخ ۳۰ فریم بر ثانیه و وضوح تصویر ۴۸۰×۶۴۰ و با فرمت RGB بدون فشرده‌سازی ۸-بیتی ضبط شده است. برای دریافت سیگنال PPG از یک دستگاه (CMS۵۰E-pulse-oximeter) استفاده شده است. این پایگاه داده شامل دو بخش است: بخش اول، که شامل ۸ ویدیو است (یکی پس از ورزش و باقی در حالت پایدار) بخش دوم، که شامل ۴۲ ویدیو است، هر یک به مدت حدود یک دقیقه و با حداقل حرکت.

مجموعه داده UBFC-PHYS [۱۷] شامل داده‌هایی از ۵۶ شرکت‌کننده است که هر کدام در سه وظیفه شرکت کرده‌اند: استراحت، صحبت، و حل مسائل ریاضی. شرکت‌کنندگان فیلم‌برداری شده‌اند و همزمان یک میچ‌بند پوشیده‌اند که سیگنال‌های BVP (حجم خون) و EDA (فعالیت الکترو درمال) را ثبت می‌کند.

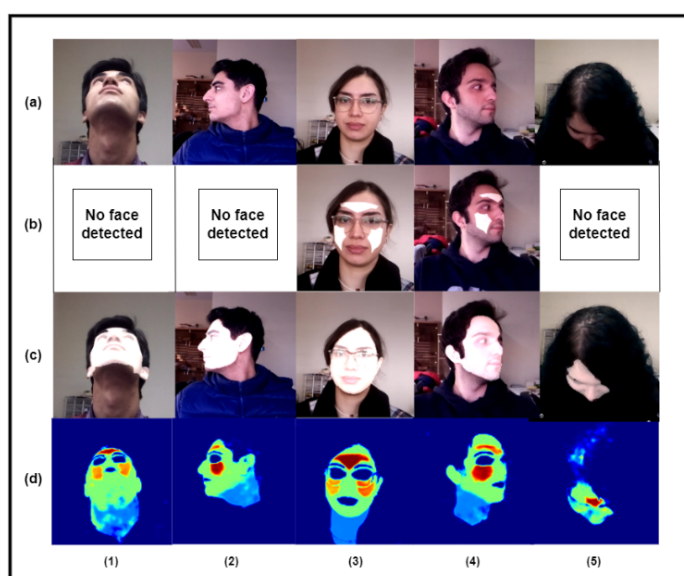
## ۳ معیارهای ارزیابی

برای ارزیابی نرخ ضربان قلب استخراج شده، از سه معیار میانگین خطای مطلق، (MAE) ریشه میانگین مربعات خطا (RMSE)، و میانگین درصد خطای مطلق (MAPE) طبق تعریف در [۱۸] استفاده می‌کنیم.

با توجه به اینکه سیگنال PPG در نوک انگشت به صورت طبیعی دارای تأخیر آناتومیک نسبت به سیگنال‌های rPPG استخراج شده از نواحی صورت و گردن است [۱۹]، ضریب همبستگی پیرسون (PCC) که به طور مستقیم

بین سیگنال‌های استخراج‌شده و مرجع محاسبه می‌شود، ممکن است معیار معناداری نباشد. برای حل این مشکل، ما PCC را در تمام جابه‌جایی‌های زمانی ممکن (در بازه یک ثانیه) محاسبه می‌کنیم و MPCC را به عنوان بیشینه مقدار به دست آمده تعریف می‌کنیم.

از آنجا که داده‌ها را خود ما جمع‌آوری کرده‌ایم، از چالش‌های کار با دستگاه‌های مربوطه آگاه هستیم. همان‌طور که در [۵] نشان داده شده است، سیگنال‌های PPG از حسگرهای تماسی نوک انگشت در مجموعه داده‌های عمومی ممکن است در اثر حرکت یا قطع تماس انگشت دچار وارونگی شده و باعث خطا در تخمین ضربان قلب شوند. در مجموعه داده ما از دو سنسور استفاده شده است. برای افزایش قابلیت اطمینان، میانگین سیگنال این دو سنسور گرفته می‌شود. با این حال، اگر قطع اتصال قابل توجهی در یکی از سنسورها رخ دهد، داده‌های آن سنسور برای آن ویدیو به طور کامل کنار گذاشته می‌شود تا اندازه‌گیری‌های غیرقابل اعتماد بر نتایج تأثیر نگذارند.



شکل ۱: نمایی از خروجی‌های سه مدل روی فریم‌های دیتاست ما.

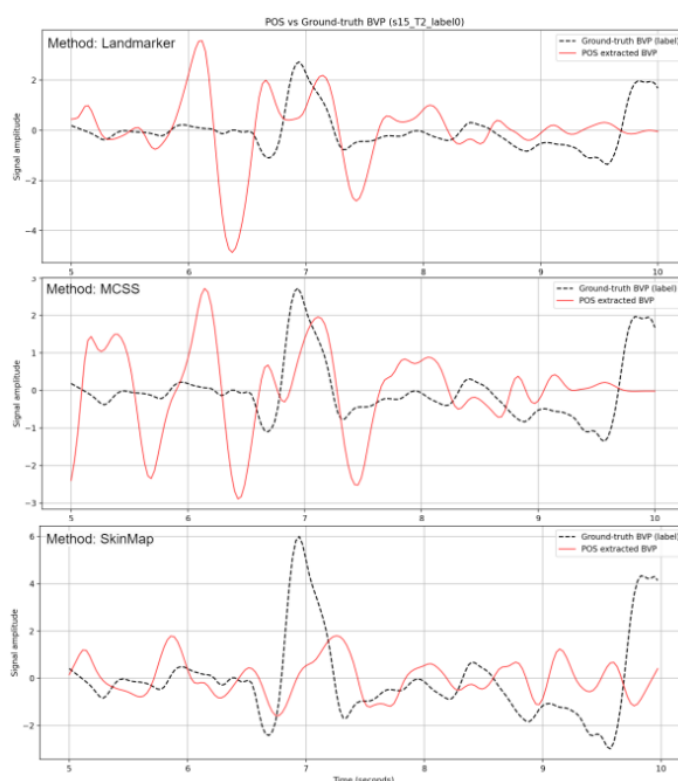
### ۱.۳ نتایج تجربی

برای بررسی عملکرد زمان اجرای مدل، نرخ فریم بر ثانیه (FPS) و تأخیر (latency) را بر روی مجموعه داده‌های منتخب محاسبه کردیم که در؟؟ نمایش داده شده‌اند. این ارزیابی بر روی کارت گرافیک (NVIDIA-RTX-۳۰۶۰) انجام شد و مقادیر ثبت‌شده برای FPS و تأخیر نشان می‌دهند که مدل توانایی عملکرد در زمان واقعی (real-time) را دارد.

در شکل ۱ ما به بررسی مدل‌ها روی فریم‌های خروجی پرداختیم. در این شکل، (a) نمونه‌ای از فریم‌های استخراج‌شده از وظیفه‌ی چرخش سر در مجموعه داده‌ی ما (b) نتایج جداسازی پوست با استفاده از روش لندمارکر، که نواحی سفید نشان‌دهنده نواحی تشخیص داده شده هستند. در برخی فریم‌ها، الگوریتم تشخیص چهره موفق به شناسایی چهره نشده است. (c) نتایج جداسازی با استفاده از روش مدل چند کلاسه مدیاپایپ، که نواحی سفید نواحی پوست تشخیص داده شده را نشان می‌دهند. (d) تصویر نقشه حرارتی از خروجی مدل جداسازی پیشنهادی ما است.

یکی از راه‌های ارزیابی مدل‌ها، اندازه‌گیری میانگین تعداد فریم‌هایی است که در آن‌ها مدل موفق به تنظیم درست ماسک نمی‌شود. این مشکل بیشتر در مدل‌های مبتنی بر ناحیه موردنظر (ROI) در حین حرکت بروز می‌کند، همان‌طور که در ۱ مشاهده می‌شود. در مجموعه داده ما، لندمارکر به‌طور میانگین در سه چهارم فریم در وظیفه‌ی صحبت کردن و ۱۱۸ فریم در وظیفه‌ی چرخش سر در هر ویدیو شکست می‌خورد، درحالی‌که مدل ما (اسکین-مپ) و چند کلاسه بدون خطا عمل می‌کنند.

برای مقایسه‌ی عادلانه، لندمارکر را بر روی مجموعه داده‌های کامل UBFC نیز اعمال کردیم. با این حال، در سناریوهای چالش برانگیز مانند مجموعه داده ما و UBFC-PHYS این مدل نتوانست هیچ سیگنالی بسازد. ما این موارد را کنار گذاشتیم و عملکرد مدل را در بهترین شرایط ارزیابی کردیم، اما با این وجود، در شرایط دشوار عملکرد آن پایین‌تر از سایر مدل‌ها باقی ماند.



شکل ۲: سیگنال‌های استخراج‌شده با استفاده از لندمارکر در بالا. در میانه چند کلاسه. در پایین مدل نهایی ما

همان‌طور که در ۲ نشان داده شده است، نمونه‌هایی از سیگنال‌های استخراج‌شده از یکی از ویدیوهای مجموعه داده ما ارائه شده‌اند. در وظیفه‌ی چرخش سر، آشکار است که مدل ما بهتر توانسته شکل و قله‌های واقعی سیگنال را بازسازی کند و در نتیجه مقادیر (MAE) بهتری ارائه دهد. درحالی‌که در موارد متعددی، مدل لندمارکر نتوانسته چهره را شناسایی کند (مطابق ۱).

در ۴، ما مقایسه‌ی کاملی بین اسکین-مپ و سایر مدل‌های پیشرفته با استفاده از پایپین استخراج دیتاست‌ها به صورت بدون نظارت ارائه داده‌ایم. اسکین-مپ در مجموعه داده‌های ما و UBFC-PHYS که سناریوهای واقعی‌تری دارند، عملکرد بهتری از خود نشان داده است. در مقابل، لندمارکر و Multi-Region در مجموعه داده‌ی UBFC



شکل ۳: خروجی مدل روی داده هایی از دیتاست کوکو

که شرایط نسبتاً ایستا دارند، عملکرد بالاتری دارند. باین حال، اسکین-مپ در این شرایط نیز رقابتی باقی می ماند و در سناریوهای چالش برانگیز دقت خود را حفظ می کند، درحالی که لندمارکر و مدل چند کلاسه چنین توانایی ای ندارند. برای مشاهده عملکرد مدل در سناریوهای دشوار، جدول؟؟ ارائه شده که نتایج براساس مجموعه داده دیتاست ما را نشان می دهد. در سناریوهای صحبت کردن و چرخش سر که چهره به طور کامل قابل مشاهده نیست یا مستقیماً به دوربین نگاه نمی کند، مدل ما در هر دو معیار RMSE و MPCC عملکرد بهتری از سایر مدل ها دارد. مقدار بالاتر MPCC نشان دهنده شباهت بیشتری بین سیگنال ضربان استخراج شده و سیگنال واقعی است. بالاترین مقدار MPCC در وظیفه چرخش سر ثبت شده که نشان دهنده مقاومت بالای مدل ما در برابر حرکات سراسر است. این ارزیابی نشان می دهد که در شرایط ایستا، انتخاب ساده ناحیه موردنظر (ROI) کفایت می کند، اما برای کاربردهای واقعی، این روش ناکافی است. در کاربردهای عملی، مدل ها باید از تمام منابع اطلاعاتی موجود استفاده کنند تا قابل اعتماد باشند. همچنین، برخی روش ها مانند (Multi-Region) چندین مؤلفه نظیر تشخیص چهره، تنظیم چهره، و شناسایی نقاط کلیدی چهره را پیش از انتخاب ROI به کار می برند [5]، درحالی که مدل ما به هیچ گونه تشخیص چهره یا پیش پردازش گسترده ای نیاز ندارد. علاوه بر این، در کاربردهایی که شناسایی دقیق ضربان قلب اهمیت دارد، وجود یک مدل مقاوم می تواند نویر ناگهانی محیطی ناشی از حرکت بدن را فیلتر کرده و کیفیت سیگنال را بهبود بخشد.

Dataset	Scenario	Metric	Models							
			Spatial Average [33]	Landmarker	MCSS	Full-body	DaapPhys [12]	EfficientPhys [34]	PhysFormer [66]	SkinMap (ours)
SYNC-rPPG	Rest	MAE ↓	11.34±2.48	10.02±2.81	11.51±3.16	10.55±2.15	5.89±2.97	<b>1.85±0.61</b>	11.87±2.48*	6.86±1.62
		RMSE ↓	15.86±9.63	16.08±11.37	18.24±11.91	14.26±8.60	14.54±12.31	3.31±2.27	16.24±9.54	9.97±5.54
		MAPE ↓	15.41±3.59	13.92±4.51	16.41±4.93	14.72±3.44	6.98±3.33	2.39±0.78	16.49±3.89	9.09±2.15
		PCC ↑	-0.04±0.24	0.11±0.23	0.20±0.23	0.039±0.236	0.233±0.229	0.933±0.085	-0.042±0.235	0.497±0.205
		SNR (dB) ↑	-5.49±0.45	-4.82±0.45	-4.75±0.43	-4.74±0.52	-1.66±0.74	-1.44±0.64	-5.21±0.43	-4.20±0.51
	Talking	MAE ↓	13.45±2.39	13.54±2.59	12.83±2.40	<b>12.30±2.04</b>	29.53±4.44*	22.85±4.28	12.39±2.55	12.66±2.17
		RMSE ↓	17.16±9.50	17.81±9.92	16.74±9.31	15.31±8.63	35.59±16.91	29.81±15.64	16.85±9.28	15.95±9.33
		MAPE ↓	15.44±2.72	14.86±2.53	14.35±2.52	13.75±2.11	31.41±4.38	24.65±4.26	13.96±2.97	14.42±2.31
		PCC ↑	0.243±0.229	0.31±0.22	0.32±0.22	0.439±0.212	-0.281±0.226	-0.262±0.227	0.128±0.234	0.242±0.229
		SNR (dB) ↑	-6.57±0.51	-6.09±0.56	-6.34±0.64	-6.21±0.59	-8.35±0.77	-7.24±0.70	-5.95±0.34	-6.15±0.67
	Head Rotation	MAE ↓	14.85±2.10	24.17±3.51	13.80±1.92	13.45±2.47	27.25±2.36*	21.45±3.18	15.21±2.71	<b>11.95±2.13</b>
		RMSE ↓	17.58±8.32	28.82±14.04	16.25±7.91	17.41±9.88	29.22±11.35	25.74±12.80	19.44±10.63	15.29±8.17
		MAPE ↓	19.18±2.93	31.30±4.39	17.74±2.57	17.65±3.40	34.05±2.51	27.80±4.44	20.50±4.24	14.99±2.59
		PCC ↑	-0.028±0.236	0.50±0.20	0.03±0.24	0.170±0.232	-0.072±0.235	-0.335±0.222	0.107±0.234	0.343±0.221
		SNR (dB) ↑	-6.25±0.38	unstable	-6.90±0.47	-5.62±0.44	-9.35±0.60	-7.77±0.48	-6.07±0.41	-5.82±0.49
	After Exercise	MAE ↓	36.47±4.86	29.53±5.33	32.70±5.27	33.05±4.89	45.18±9.34*	37.88±7.69	<b>28.56±5.67</b>	32.96±4.64
		RMSE ↓	42.46±17.89	37.96±18.98	40.31±19.11	39.62±18.05	61.53±31.25	51.16±25.13	38.21±20.06	38.94±17.70
		MAPE ↓	29.05±3.38	22.95±3.80	25.77±3.54	25.82±3.29	34.82±6.73	28.44±5.49	21.98±3.95	26.11±3.07
		PCC ↑	0.241±0.229	0.00±0.24	-0.47±0.21	0.033±0.236	-0.317±0.224	-0.450±0.210	-0.038±0.236	0.312±0.224
		SNR (dB) ↑	-10.77±0.95	-9.64±1.03	-10.18±1.05	-10.02±0.83	-8.65±1.08	-7.34±0.92	-8.92±1.03	-9.49±0.88
UBFC-Phys	Rest	MAE ↓	4.91±1.23	5.13±1.55	5.28±1.52	4.65±1.10	5.57±1.43	<b>3.75±0.98</b>	6.25±1.46*	5.18±1.36
		RMSE ↓	10.13±6.50	12.29±8.53	12.19±8.52	9.19±5.72	11.20±6.57	7.63±5.17	11.87±7.12	10.86±6.95
		MAPE ↓	6.88±1.89	6.83±2.43	7.03±2.41	5.98±1.64	7.47±2.02	5.34±1.52	8.97±2.27	7.27±2.07
		PCC ↑	0.751±0.093	0.577±0.116	0.597±0.113	0.770±0.090	0.718±0.105	0.834±0.083	0.678±0.108	0.717±0.102
		SNR (dB) ↑	0.69±0.71	2.82±0.90	3.06±0.87	2.04±0.91	0.322±0.771	0.71±0.75	-0.75±0.84	0.37±0.80
	Talking	MAE ↓	12.75±1.80	25.00±2.72*	24.85±2.80	16.09±2.01	19.45±2.37	16.91±2.11	18.19±1.95	<b>12.04±1.73</b>
		RMSE ↓	18.20±9.03	31.77±14.53	32.03±13.29	21.64±10.36	25.10±10.69	22.46±10.25	21.98±9.48	17.35±8.90
		MAPE ↓	18.38±3.07	35.87±4.80	35.38±4.44	22.24±2.90	24.31±2.94	23.51±3.42	25.33±3.21	16.82±2.88
		PCC ↑	0.143±0.140	-0.262±0.136	-0.073±0.141	0.193±0.139	-0.062±0.152	-0.126±0.145	0.214±0.158	0.124±0.140
		SNR (dB) ↑	-5.14±0.41	-7.42±0.54	-6.18±0.57	-6.30±0.54	-6.14±0.43	-5.53±0.43	-6.24±0.38	-5.19±0.40
	Arithmetic	MAE ↓	10.31±1.62	22.13±2.51*	20.51±2.27	19.89±2.08	13.18±1.87	12.19±1.89	16.44±2.12	<b>10.12±1.61</b>
		RMSE ↓	15.68±7.74	28.72±12.60	26.34±11.57	24.99±10.63	18.68±9.31	17.99±8.39	21.94±10.02	15.46±7.91
		MAPE ↓	15.06±2.70	35.88±4.84	33.45±4.56	31.37±4.21	16.86±2.26	17.56±3.00	23.29±3.24	14.72±2.70
		PCC ↑	0.325±0.132	-0.166±0.138	0.152±0.138	-0.044±0.140	0.436±0.130	0.248±0.141	-0.024±0.149	0.394±0.129
		SNR (dB) ↑	-4.57±0.36	-6.57±0.60	-6.76±0.59	-6.13±0.54	-4.83±0.44	-4.00±0.47	-5.38±0.34	-4.17±0.36

شکل ۴: عملکرد مدل ها در دیتاست ها با متد های استخراج مختلف.



## جمع‌بندی

این پروژه با هدف استخراج سیگنال ضربان قلب از طریق تصاویر و ویدئوهای صورت (rPPG) و بدون نیاز به تماس فیزیکی، طراحی و پیاده‌سازی شده است. در ابتدا، با استفاده از مدل دیپ‌لَب، یک مدل بخش‌بندی دقیق پوست صورت ایجاد شد تا ویژگی‌های مختلف صورت به‌طور مؤثر از تصاویر استخراج شوند. سپس، برای جمع‌آوری داده‌های آموزشی، دیتاست‌های ویدیویی همزمان با سیگنال‌های rPPG از افراد در شرایط مختلف نورپردازی و حرکت تهیه شد. در مرحله آموزش، مدل روی دیتاست‌های تصویری آموزش داده شد و پارامترهای آن براساس نتایج اولیه تنظیم گردید. پس از آن، مدل بهینه‌شده به ویدئوهای واقعی اعمال شد و برای عملکرد در زمان واقعی بهینه شد. این مرحله به‌ویژه برای اجرای سیستم در شرایط پویا و متغیر اهمیت داشت.

در نهایت، با پیاده‌سازی مدل آموزش‌دیده‌شده روی دیتاست‌های ویدیویی جمع‌آوری‌شده، سیگنال‌های rPPG استخراج و با سیگنال‌های حاصل از سنسورهای PPG مقایسه شدند. این مقایسه نشان داد که روش پیشنهادی دقت و قابلیت اطمینان بالاتری نسبت به روش‌های سنتی ارائه می‌دهد. نتیجه‌گیری این پروژه، تأکید بر این است که استفاده از روش‌های بدون تماس مانند rPPG می‌تواند به‌طور مؤثر برای نظارت بر ضربان قلب در شرایط مختلف محیطی و بدون نیاز به تجهیزات اضافی و تماس فیزیکی، به کار گرفته شود. این روش‌ها به‌ویژه در شرایط پیچیده نوری و حرکتی می‌توانند عملکرد مطلوبی داشته باشند و در نهایت، در نظارت‌های سلامتی و پزشکی کاربرد گسترده‌ای پیدا کنند.



- [1] X. Liu, X. Zhang, G. Narayanswamy, Y. Zhang, Y. Wang, S. Patel, and D. McDuff, "Deep physiological sensing toolbox," *arXiv preprint arXiv:221000716*., .2022
- [2] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol.64, no.7, pp.1491–1479, .2017
- [3] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile gpus," *ArXiv*, vol.abs/190706724., .2019
- [4] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for building perception pipelines," .2019
- [5] C. . Casado and M. B. López, "Face2ppg: An unsupervised pipeline for blood volume pulse extraction from faces," *IEEE Journal of Biomedical and Health Informatics*, vol.27, no.11, pp.5541–5530, .2023
- [6] D. Dahmani, M. Cheref, and S. Larabi, "Zero-sum game theory model for segmenting skin regions," *Image and Vision Computing*, vol.99, p.103925, .2020
- [7] A. Trumpp, S. Rasche, D. Wedekind, M. Schmidt, T. Waldow, F. Gaetjen, K. Plötze, H. Malberg, K. Matschke, and S. Zaunseder, "Skin detection and tracking for camera-based photoplethysmography using a bayesian classifier and level set segmentation," in *Bildverarbeitung für die Medizin*, .2017
- [8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *European Conference on Computer Vision*, .2018

- [9] W. Wang, A. C. den Brinker, and G. de Haan, “Single-element remote-ppg,” *IEEE Transactions on Biomedical Engineering*, vol.66, no.7, pp.2043–2032, .2019
- [10] A. Gudi, M. Bittner, and J. van Gemert, “Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation,” *Applied Sciences*, vol.10, no.23, .2020
- [11] C. S. Pilz, S. Zaunseder, J. Krajewski, and V. Blazek, “Local group invariance for heart rate estimation from face videos in the wild,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.13358–1335, .2018
- [12] G. de Haan and V. Jeanne, “Robust pulse rate from chrominance-based rppg,” *IEEE Transactions on Biomedical Engineering*, vol.60, no.10, pp.2886–2878, .2013
- [13] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak, “Measuring pulse rate with a webcam — a non-contact method for evaluating cardiac activity,” in *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp.410–405, .2011
- [14] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Doll’ar, and C. L. Zitnick, “Microsoft COCO: common objects in context,” *CoRR*, vol.abs/14050312., .2014
- [15] M. Scherpf, H. Ernst, L. Misera, H. Malberg, and M. Schmidt, “Skin segmentation for imaging photoplethysmography using a specialized deep learning approach,” in *2021 Computing in Cardiology (CinC)*, vol.48, pp.4–1, .2021
- [16] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, “Unsupervised skin tissue segmentation for remote photoplethysmography,” *Pattern Recognition Letters*, vol.92, pp.42–35, .2017
- [17] R. Meziati, Y. Benezeth, P. De Oliveira, J. Chappé, and F. Yang, “Ubfc-phys,” .2021
- [18] K. Lee, J. Oh, H. You, and E. C. Lee, “Improving remote photoplethysmography performance through deep-learning-based real-time skin segmentation network,” *Electronics*, vol.12, p.3729, 09 .2023

- [19] J. Allen and A. Murray, “Effects of filtering on multisite photoplethysmography pulse waveform characteristics,” in *Computers in Cardiology, 2004*, pp.488–485, .2004