



پردیس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

تشخیص عواطف توسط دستگاه‌های پوشیدنی

نگارنده

علی عبداللہی اصل

استاد راهنما: دکتر باقر باباعلی

پایان‌نامه برای دریافت درجه کارشناسی
در رشته علوم کامپیوتر

مرداد ۱۴۰۳

چکیده

تشخیص عواطف یکی از کاربردهای روزافزون هوش مصنوعی است. این کار می‌تواند در صنعت روانشناسی و خصوصا حوزه ترکیبی محاسبات عاطفی نقش بسیار پررنگی ایفا کند. همچنین هر روز پیشرفت بیشتری را در مورد پوشیدنی‌های هوشمند و افزایش استفاده از آنها را شاهد هستیم.

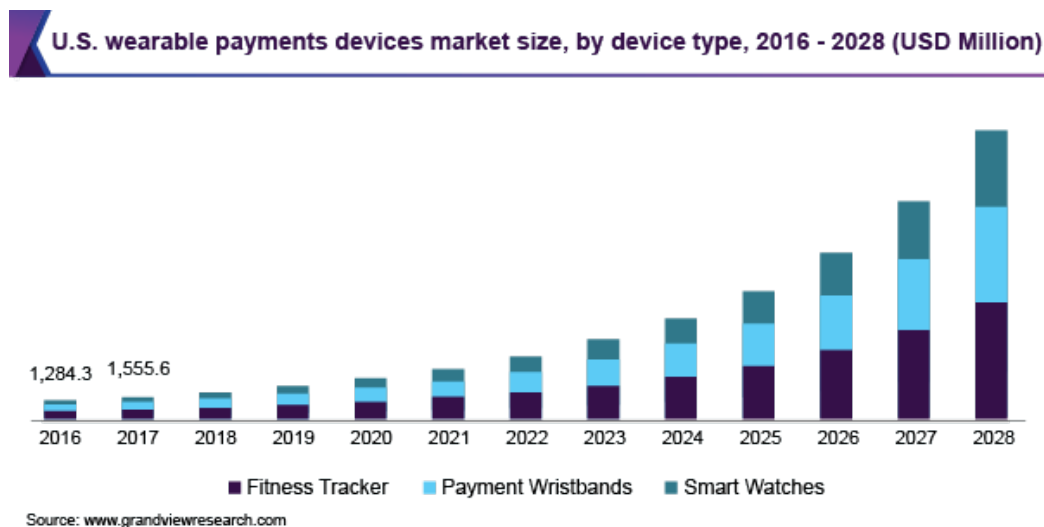
سپاسگزاری

از استاد راهنمای گرانقدر جناب آقای دکتر باباعلی که با راهنمایی های خود مرا در انجام این پروژه یاری دادند
کمال تشکر را دارم.

پیشگفتار

تشخیص احساسات به عنوان یکی از موضوعات مهم در حوزه روانشناسی و علوم عصبی به شمار می‌آید. فهم و شناسایی حالت‌های احساسی افراد می‌تواند به بهبود کیفیت زندگی، مدیریت استرس، و ارتقاء سلامت روان کمک کند. تشخیص صحیح احساسات به پزشکان، روانشناسان، و محققان امکان می‌دهد تا به طور دقیق‌تری به بررسی واکنش‌های احساسی و رفتاری افراد بپردازند و راه‌حل‌های موثرتری برای مشکلات روانی ارائه دهند.

در سال‌های اخیر، استفاده از دستگاه‌های پوشیدنی به طور چشمگیری افزایش یافته است. در شکل ۱ نمودار پیشرفت بازار که به علت رشد مصرف این دستگاه‌های پوشیدنی نمایش داده شده‌اند. همچنین پیشبینی آن تا سال ۲۰۲۸ نمایانگر این است که این دستگاه‌ها نقش به مراتب پررنگ‌تری حتی نسبت به امروز خواهند داشت.



شکل ۱: نمودار بازار دستگاه‌های پوشیدنی از سال ۲۰۱۶ تا پیشبینی آن تا سال ۲۰۲۸ [۱]

این دستگاه‌ها با قابلیت‌های پیشرفته‌ای که دارند، امکان مانیتورینگ مداوم وضعیت جسمی و روانی افراد

را فراهم می‌کنند. از جمله مزایای این دستگاه‌ها می‌توان به راحتی استفاده، قابلیت حمل، و ارائه اطلاعات دقیق و به‌روز اشاره کرد. این ویژگی‌ها باعث شده‌اند تا دستگاه‌های پوشیدنی به ابزارهای ضروری در حوزه سلامت و تحقیقاتی تبدیل شوند.

در این گزارش، از دو دستگاه پوشیدنی استفاده شده است: دستبند Empatica E4 و بند سینه‌ای Respiban. دستبند Empatica E4 با سنسورهای مختلفی که دارد، می‌تواند اطلاعات مربوط به فعالیت‌های الکتریکی پوست، ضربان قلب، و دمای بدن را جمع‌آوری کند. بند سینه‌ای Respiban نیز با اندازه‌گیری دقیق تنفس و فعالیت قلبی، اطلاعات جامعی از وضعیت فیزیولوژیکی فرد ارائه می‌دهد. این دو دستگاه در تشخیص حالات احساسی پایه، استرس، و خوشحالی مورد استفاده قرار گرفته‌اند و داده‌های جمع‌آوری شده از آن‌ها برای تحلیل دقیق‌تر احساسات به کار گرفته می‌شوند.

ترانسفورمرها به عنوان یکی از پیشرفته‌ترین معماری‌های مدل‌های یادگیری عمیق، انقلابی در هوش مصنوعی ایجاد کرده‌اند. این مدل‌ها با استفاده از مکانیسم توجه، توانایی تحلیل، در نظر گرفتن زمینه و پردازش موازی داده‌ها را به طور همزمان دارند که منجر به بهبود قابل توجهی در عملکرد و سرعت مدل‌های یادگیری می‌شود. ترانسفورمرها در کاربردهای مختلفی از جمله پردازش زبان طبیعی، ترجمه ماشینی، و تولید متن‌های خودکار به کار گرفته می‌شوند و توانسته‌اند دقت و کارایی مدل‌های هوش مصنوعی را به سطحی بی‌سابقه برسانند.

در این گزارش، ما از مدل ترانسفورمر BioT که یک مدل قدرتمند برای تحلیل سیگنال‌های بیولوژیکی است، استفاده می‌کنیم. مدل BioT با بهره‌گیری از ساختار ترانسفورمر، قادر است تا اطلاعات پیچیده و متنوعی که از سیگنال‌های بیولوژیکی نظیر ضربان قلب، تنفس، و فعالیت‌های الکتریکی پوست به دست می‌آید را به طور دقیق تحلیل و تفسیر کند. این مدل به ما امکان می‌دهد تا با دقت بالاتری به تشخیص حالات احساسی پایه، استرس، و خوشحالی بپردازیم و نتایج بهتری در این حوزه به دست آوریم.

فهرست مطالب

۳	مجموعه داده	۱
۳	۱.۱ دستگاه Respiban	
۴	۲.۱ ساعت Empatica E4	
۴	۳.۱ پرسشنامه‌ها	
۶	۴.۱ طبقه‌بندی	
۷	کارهای مشابه	۲
۷	۱.۲ روش شناسی	
۱۲	پیش‌پردازش	۳
۱۲	۱.۳ ساختار مجموعه داده	
۱۲	۲.۳ تمیزسازی داده‌ها	
۱۴	آموزش	۴
۱۴	۱.۴ مدل BioT	
۱۴	۲.۴ ساختار مدل	
۱۵	۱.۲.۴ امیدینگ اولیه	
۱۶	۲.۲.۴ بخش ترانسفورمر	
۱۶	۳.۴ مراحل خوراندن داده به مدل	
۱۷	۴.۴ استخراج ویژگی‌ها	
۱۷	۵.۴ سنجش	
۱۹	نتایج	۵
۱۹	۱.۵ ۲ کلاس	
۱۹	۲.۵ ۳ کلاس	
۱۹	۱.۲.۵ سیگنال‌های سینه‌ای	
۲۰	۲.۲.۵ سیگنال‌های مچ	
۲۱	۳.۲.۵ جفت سیگنال‌ها	

۴.۲.۵ روند ۲۱

فصل ۱

مجموعه داده

مجموعه داده [2] WESAD یکی از کامل ترین مجموعه های داده برای تشخیص عواطف است. بیشترین تمرکز و استفاده از این مجموعه برای تشخیص استرس بوده است. با این وجود، به جز کلاس استرس و عادی، برای تشخیص کلاس های خوشحالی و آرامش نیز می توان از این داده ها استفاده نمود. علاوه بر آن ها، هر شخص پرسشنامه هایی نیز پر کرده که این هم می تواند باعث خلق مدل های جدیدی شود. دو دستگاه اصلی برای فراهم آوردن این داده ها مورد استفاده قرار گرفته اند: ۱) مچ بند E4 Emptica به بسیاری از دانشگاه های سراسر دنیا از آن استفاده می کنند [۳] و ۲) دستگاه Repiban که یکی از پیشرفته ترین سنسورهای تحقیقاتی است که بر روی سینه نصب می شود [۴].

۱.۱ دستگاه Respiban

این دستگاه می تواند ۶ عامل را اندازه گیری کند. فرکانس ورودی این دستگاه برای همه سنسورهایش ۷۰۰ هرتز می باشد. سنسورهای آن به شرح زیر است:

۱. مختصات یابی (Accelerometer)
۲. نوار قلب (Electrocardiogram)
۳. فعالیت الکتریکی پوست (Electrodermal Activity)
۴. برق ماهیچه نگار (Electromyogram)
۵. تنفس (Respiration)
۶. دما (Temperature)



شکل ۱.۱: شکل راست: ساعت Empatica E4 را نشان می‌دهد. این ساعت به دلیل سنسورهای کامل، زیبایی و راحتی استفاده کاربرد زیادی در تحقیقات دارد. شکل چپ: دستگاه Respiban و نحوه قرارگیری آن روی سینه و محل هر یک از سنسورها را نشان می‌دهد.

۲.۱ ساعت Empatica E4

این ساعت شامل سنسورهای مختصات یابی، فشار خون^۱، دما، فعالیت الکتریکی پوست است. هر یک از این سنسورها با فرکانس متفاوتی اندازه‌گیری شده‌اند. در جدول مقدار فرکانس هر یک از سنسورها آورده شده است.

فرکانس	سنسور
۳۲	ACC
۶۴	BVP
۴	EDA
۴	Temp

۳.۱ پرسشنامه‌ها

علاوه بر دو دستگاه گفته شده، هر یک از سوژه‌های آزمایش، پرسشنامه‌هایی را پر کردند. این پرسشنامه‌ها در جهت دریافت اطلاعات بیشتر در مورد احساسات اشخاص به کار گرفته شدند، اگرچه در هیچ‌یک از مقالات

^۱Blood Volume Pressure

بررسی شده، محققان از این پرسشنامه ها استفاده ای نکردند. در قسمت های پیش رو این پرسشنامه ها را بررسی می کنیم:

PANAS

سوژه می بایست به ۲۶ حس در پرسشنامه، از ۱ تا ۵ امتیاز دهد. این احساسات عبارتند از: فعال، پریشانی، علاقه مند، الهام شده، رنجیده، گناهکار، ترسیده، دشمنی، هیجان زده، مغرور، کج خلق، مشتاق، شرمنده، هوشیار، نگران، مصمم، متوجه، عصبی، وحشت زده، استرسی، خسته، خوشحال، عصبانی، آزردن و ناراحت.

STAI

در این پرسشنامه، سوژه به هر یک از سوال های زیر از ۱ تا ۴ نمره می دهد:

۱. من احساس راحتی می کنم
۲. من احساس نگرانی می کنم
۳. من عصبی هستم
۴. من ریلکس هستم
۵. من احساس دلواپسی می کنم
۶. من احساس رضایت می کنم

SAM

این تست شدت و خوب یا بد بودن احساسات را می سنجد. شخص دو سوال را در مقیاس ۱ تا ۹ پاسخ می دهد: ۱) حس من چقدر خوب است و ۲) شدت این حس چقدر است.

SSSQ

این تست که کوتاه شده تست استاندارد SSSQ است، در زمان های استرس از شرکت کنندگان گرفته شده است. پرسش شونده ها به سوال های زیر از ۱ تا ۵ نمره می دهند:

۱. من متعهد به رسیدن به اهداف عملکردی ام هستم
۲. من می خواهم در این کار موفق شوم
۳. من انگیزه برای انجام این کار را دارم
۴. من خودم را بروز می دهم
۵. من نگران تفکرات دیگران در مورد خودم هستم
۶. من متوجه تاثیری که روی بقیه می گذارم هستم

۴.۱ طبقه‌بندی

این مجموعه داده عواطف انسان‌های مورد بررسی را در ۴ طبقه شناسایی کرده است: (۱) حالت معمولی^۲، (۲) استرس، (۳) خوشحالی^۳ و (۴) آرامش^۴.

سوژه	ثانیه‌های مفید
۲	۲۸۸۴
۳	۲۹۳۰
۴	۲۹۶۵
۵	۳۰۰۶
۶	۲۹۸۴
۷	۲۹۸۳
۸	۳۰۰۰
۹	۲۹۸۵
۱۰	۳۰۶۸
۱۱	۳۰۱۴
۱۳	۳۰۱۶
۱۴	۳۰۱۶
۱۵	۳۰۲۲
۱۶	۳۰۰۸
۱۷	۳۰۰۲

جدول ۱.۱: ثانیه‌های مفید هر یک از سوژه‌ها

در جدول ۱.۱ ثانیه‌های مفید هر یک از سوژه‌ها آورده شده است. منظور از ثانیه‌های مفید، آنهایی است که کلاس‌های آنها حالت پایه، استرس، خوشحالی و یا آرامش است. در دادگان دو کلاس بی‌نام دیگر وجود دارد که آنها می‌بایست حذف شوند.

کلاس‌گذاری در بیشترین فرکانس ممکن (۷۰۰) صورت گرفته و برای هر یک از سنسورها برای دستیابی به کلاس موردنظر باید آن را به فرکانس آن سنسور تبدیل کنیم.

baseline^۲
Amusement^۳
Meditated^۴

فصل ۲

کارهای مشابه

کارهای زیادی با استفاده از این مجموعه داده برای تشخیص عواطف صورت گرفته است. گرچه این دادگان در ۴ کلاس گردآوری شده، بیشتر کارها ۲ کلاسه یا ۳ کلاسه (حالت عادی، استرس و خوشحالی) هستند. همچنین از پرسشنامه‌های موجود در دادگان بهره‌جندانی برده نشده است. در جدول ۱.۲ کارهای مشابه که از این دادگان استفاده کردند آورده شده است.

۱.۲ روش شناسی

در ادامه به بررسی روش استفاده شده در هر یک از کارهای نام برده شده می‌پردازیم:

- Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection[2]:

در اینجا از استخراج ویژگی‌های پیچیده‌ای استفاده گردیده است. برای ACC میانگین و انحراف معیار برای هر یک از ابعاد شناسایی و با هم جمع شده و علاوه بر آن نقطه اوج هر یک نیز محاسبه گردیده. از سیگنال‌های ECG و BVP میانگین و واریانس آنها و همچنین از نقاط پیک آنها ضربان قلب و از زمان‌های ضربان قلب، تغییرات آن^۱ به دست می‌آید. برای EDA ابتدا یک فیلتر پایین‌گذر^۲ ۵ هرتزی بر روی آن اعمال و میانگین و واریانس محاسبه می‌شود. همچنین دو بخش تونیک و فازیک این سیگنال (به نام‌های Skin Conductance Level و Skin Conductance Response) با توجه به کار [۱۰] استخراج شدند. بر روی سیگنال EMG ابتدا یک فیلتر بالاگذر^۳ اعمال و سپس نقاط پیک شناسایی شدند و چندین ویژگی دیگر با توجه به کار [۱۱] استخراج شدند.

^۱Heart Rate Variability
^۲low pass filter
^۳high pass filter

نام کار	سیگنال‌ها	پنجره	عملکرد f-1
Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection[2]	Extracted features from all the signals	پنجره: ۶۰ ثانیه، قدم: ۱/۴ ثانیه	2class: 91.47, 3class: 72.51
Transformer-based Self-supervised Multimodal Representation Learning for Wearable Emotion Recognition[8]	Wrist BVP, EDA, and Temp	پنجره: ۶۰ ثانیه (۴ هرتز)، قدم: ۱/۴ ثانیه	2class: 93.69, 3class: 82.01
Affective State Recognition with Convolutional Autoencoders[9]	All the signals	پنجره: ۱ ثانیه، قدم: ۱ ثانیه	3class: 82.82
Stress Detection by Machine Learning and Wearable Sensors[6]	Manual features from all the chest signals	پنجره: ۱۰ ثانیه، قدم: ۱۰ ثانیه	2class: 83.34, 3class: 65.73
A Transformer Architecture for Stress Detection from ECG[7]	ECG	پنجره: ۳۰ ثانیه، قدم: ۱ ثانیه	2class: 83.3

جدول ۱.۲: کارهای مشابه بر روی دادگان WESAD

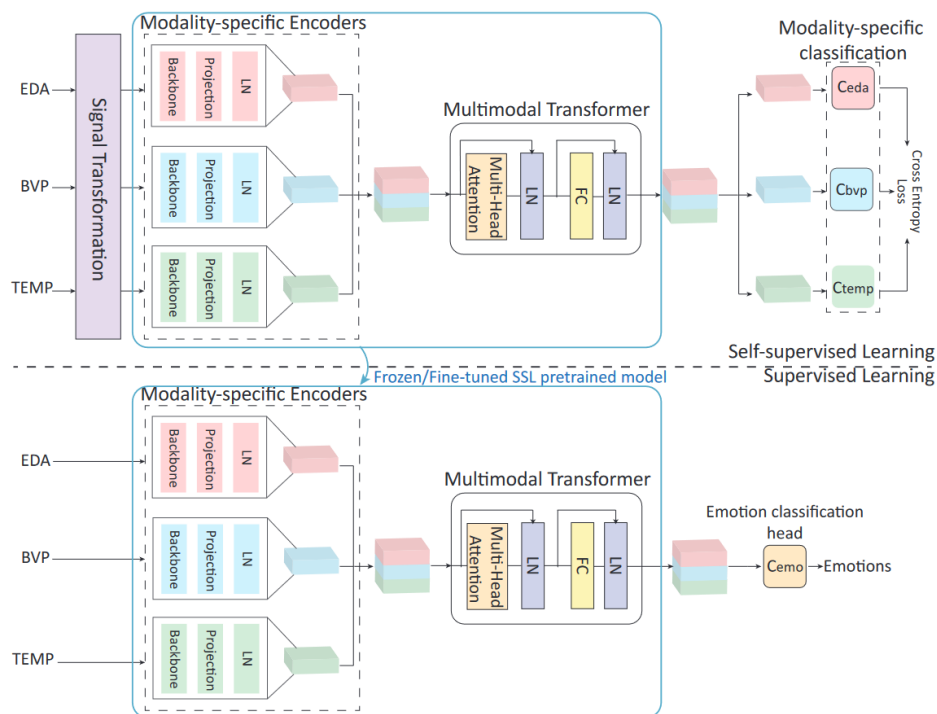
برای Resp ابتدا یک فیلتر میان‌گذر^۴ بر روی آن اعمال شدند. پیک‌ها شناسایی و میانگین و واریانس دم و بازدم‌ها محاسبه شدند. علاوه بر آنها نسبت دم به بازدم، حجم تنفس، نرخ تنفس و مدت زمان تنفس نیز محاسبه گردیدند. برای دما، میانگین و واریانس و بیشینه و کمینه و شیب آن محاسبه گردیده بود. برای انجام کلاسیفیکیشن،^۵ روش دسته‌بندی درخت تصمیم^۵ جنگل تصادفی^۶، k-همسایه-نزدیک^۷، تحلیل تشخیصی خطی^۸ و AdaBoost استفاده شدند.

• Transformer-based Self-supervised Multimodal Representation Learning for Wearable Emotion Recognition[8]

ابتدا چندین تبدیل روی هر پنجره اتفاق می‌افتد:

band pass filter^۴
Decision Tree^۵
Random Forest^۶
K-Nearest-Neighbor^۷
Linear Discriminant Analysis^۸

۱. جایگشت: قسمت‌های مختلف پنجره جدا شده و سپس با جایگشتی دیگر به هم چسبانده می‌شوند.
۲. پیچ و تاب زمانی: پنجره سیگنال به n قسمت تقسیم شده و نیمی از آنها منبسط و نیمی دیگر منقبض می‌شوند.
۳. برش: با تقسیم پنجره به n قسمت، یکی را حذف کرده و دوباره نمونه‌گیری می‌کنیم.



شکل ۱.۲: معماری مدل مقاله [۸]. مدل ترانسفورمر اصلی از یک توجه چندکله‌ای (Multi-Head Attention) و سپس نرمال‌ساز لایه (Layer Normalization) و سپس یک لایه کامل متصل (Fully Connected) و دوباره یک نرمال‌ساز لایه دیگر.

معماری مدل در شکل ۱.۲ آورده شده است. مرحله آموزش شامل دو مرحله بوده. اول پیش‌آموزش آن روی داده‌گان PRESAGE [۱۲] انجام می‌شود. در مرحله دوم از انکودر مدل پیش‌آموزش داده شده برای کلاسیفیکیشن استفاده می‌شود.

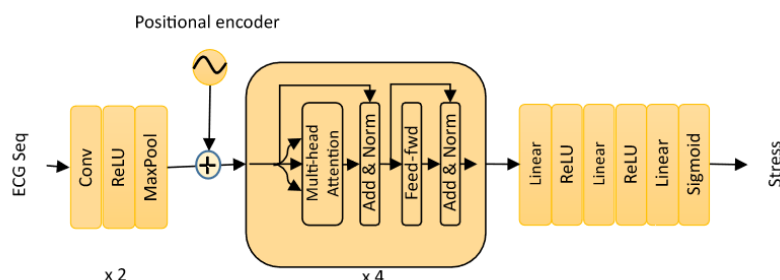
- [9] Affective State Recognition with Convolutional Autoencoders: ۳ اتوانکودر مورد استفاده قرار می‌گیرند. یکی برای سیگنال‌های سینه‌ای که ۸۰ ویژگی هدف آن است. برای سیگنال BVP مچ هم یک اتوانکودر که ۴۰ ویژگی را استخراج می‌کند. در انتها برای

سیگنال‌های مچی EDA و دما نیز یکی دیگر که تنها ۴ ویژگی را تحویل دهد.

در نهایت این ۱۲۴ ویژگی با هم تجميع شده و به دسته‌بندی‌کننده SVM خورانده می‌شود تا دسته بندی را انجام دهد.

- [6] Stress Detection by Machine Learning and Wearable Sensors:
در این کار بیشینه، کمینه، میانگین و انحراف معیار پنجره‌های ۱۰ ثانیه‌ای بدون اشتراک به دست آورده شده است و سپس با بهره‌گیری از ۵ روش دسته‌بندی جنگل تصادفی، k -همسایه-نزدیک، تحلیل تشخیصی خطی، AdaBoost و ماشین بردار پشتیبان^۹، آن‌ها را دسته‌بندی کرده‌اند.
- [7] A Transformer Architecture for Stress Detection from ECG:
در این مدل تنها از سیگنال‌های ECG استفاده شد. در شکل ۲.۲ ساختار این مدل نمایش داده شده‌است.

مدل از ۳ بخش تشکیل شده: بخش کانولوشن، بخش ترانسفورمر و بخش کاملاً متصل.



شکل ۲.۲: ساختار مدل [۷]

در بخش کانولوشن که هدف آن استخراج ویژگی‌هاست، دو لایه مشابه دارد که هر یک از یک کانولوشن و در ادامه آن تابع فعالساز ReLU و یک لایه حداکثر فشرده‌ساز^{۱۰} قرار دارد.

برای پاس دادن خروجی بخش کانولوشن ابتدا از یک انکودینگ مکانی^{۱۱} استفاده می‌شود. بعد از آن بخش ترانسفورمر قرار دارد که از ۴ لایه با ساختار زیر تشکیل می‌شود:

Support Vector Machine^۹
Max-Pooling^{۱۰}
Positional Encoding^{۱۱}

۱. لایه چندسر خودتوجه^{۱۲}

۲. لایه dropout و نرمال‌ساز

۳. لایه کاملاً متصل پیشخور^{۱۳}

۴. دوباره لایه dropout و نرمال‌ساز

پس از بخش ترانسفورمر، ۳ لایه خطی خواهیم داشت که دو تایی اول تابع فعال‌ساز ReLU و آخری را تابع سیگموید همراهی می‌کند که مسئولیت دسته‌بندی بر عهده آن است.

^{۱۲} Multi-Head Self-Attention
^{۱۳} feed-forward

فصل ۳

پیش پردازش

در بخش‌های پیش‌رو در مورد کارهای مورد نیاز برای آماده‌سازی داده‌ها برای آموزش مدل‌ها بحث می‌کنیم

۱.۳ ساختار مجموعه داده

در دادگان WESAD داده‌های تجمیع‌شده و همگام‌شده را برای هر سوژه در یک فایل pkl فراهم آورده‌اند. این فایل یک دیکشنری به صورت زیر است.
آرایه label و تمام آرایه‌های سنسورهای chest، به طول ۴،۵۴۵،۱۰۰ هستند، که همه $Hz700$ در طول ۶۴۹۳ ثانیه هستند. آرایه‌های ACC و BVP به ترتیب ۳۲ و ۶۴ هرتز و دو سیگنال دیگر هر دو ۴ هرتز هستند.

۲.۳ تمیزسازی دادگان

همانطور که در بالا گفته شد، برخی کلاس‌های داده بلااستفاده هستند. در قدم اول این‌ها حذف می‌شوند و تنها ثانیه‌های مفید باقی می‌مانند. سپس برای زیباسازی ساختار ذخیره داده، آن را به شکل ۲.۳ تغییر می‌دهیم.

```

`Sx.pkl`
├── subject
├── label
├── signal
│   ├── chest
│   │   ├── ACC
│   │   ├── ECG
│   │   ├── EMG
│   │   ├── EDA
│   │   ├── Temp
│   │   └── Resp
│   └── wrist
│       ├── ACC
│       ├── BVP
│       ├── EDA
│       └── TEMP

```

شکل ۱.۳: ساختار اولیه دادگان WESAD

```

`Sx_n0.pkl`
├── label
├── chest_ACC
├── chest_ECG
├── chest_EMG
├── chest_EDA
├── chest_Temp
├── chest_Resp
├── wrist_ACC
├── wrist_BVP
├── wrist_EDA
└── wrist_TEMP

```

شکل ۲.۳: ساختار دادگان پس از تغییر

فصل ۴

آموزش

۱.۴ مدل BioT

مدل BioT برای کار با داده های EEG طراحی و ساخته شده است. اما همانطور که از نام آن پیداست (Bio Transformer) از آن می توان برای انواع سیگنال های حیاتی بهره گرفت. در شکل ۱.۴ ساختار این مدل را مشاهده می کنید. در نیم تصویر بالا، ماژول توکنایز کردن سیگنال هاست، که با انجام دوباره نمونه گیری^۱، نرمال کردن، توکنایز کردن و تخت کردن^۲، آن را تبدیل به جملات می کند. سپس تعامل بین این جملات با استفاده از ماژول ترنسفورمر خطی (نیم تصویر پایین) یاد گرفته می شوند. این مدل به صورت با نظارت می تواند داده های مختلف کامل و ناقص را برای پیش آموزش و fine-tuning بپذیرد.

۲.۴ ساختار مدل

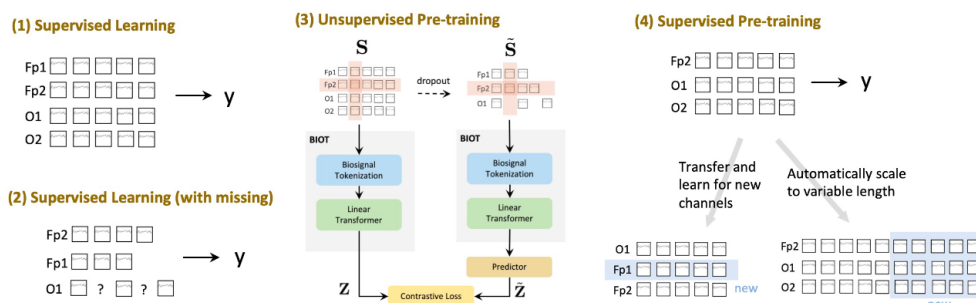
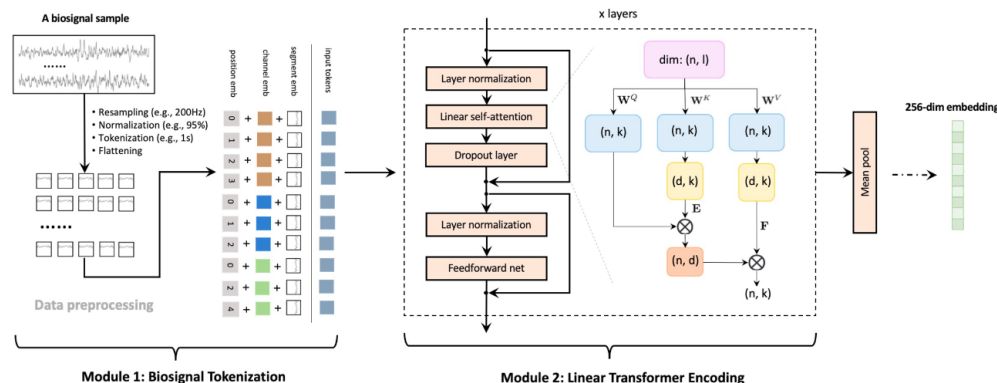
مدل از دو بخش انکودر و سر کلاسیفیکیشن^۳ تشکیل شده است.

سر کلاسیفیکیشن از یک تابع فعالساز ELU^۴ و یک لایه خطی تشکیل شده است که با تبدیل $xA^T + b$ انکودینگ ساخته شده را به یک وکتور با اندازه تعداد کلاس ها می برد که بیانگر احتمال این است که آن ورودی در هر کدام از آن کلاس ها باشد.

بخش انکودر نیز ابتدا یک امبدینگ اولیه ساخته و سپس آن را به ترانسفورمر داده تا برای پنجره مورد نظر یک امبدینگ نهایی بسازد. در قسمت پیش رو به تفصیل این دو بخش توضیح داده شده اند.

resampling^۱
flattening^۲
Classification Head^۳
Exponential Linear Unit^۴

Biosignal Transformer (BIOT) Encoder



شکل ۱.۴: معماری شبکه BioT

۱.۲.۴ امبدینگ اولیه

برای محاسبه یک امبدینگ اولیه، برای هر یک از کانال‌ها، فرایند زیر محاسبه می‌گردد:

۱. بر روی سیگنال آن کانال، STFT^۵ اعمال می‌گردد.

۲. سپس بر روی آن Patch Frequency embedding صورت می‌گیرد. برای این امر ابتدا یک جایگشت بر روی ابعاد سیگنال اعمال شده و سپس با یک تبدیل خطی^۶ آن را به ابعاد ۲۵۶ می‌بریم.

۳. در نهایت انکودینگ مکانی^۷ بر آن افزوده می‌شود.

پس از انجام این فرایندها بر روی سیگنال‌های هر کانال آن‌ها را با هم میانگین می‌گیریم و امبدینگ اولیه آماده می‌شود.

Short-Time Fourier Transform^۵
Linear^۶
Positional Encoding^۷

۲.۲.۴ بخش ترانسفورمر

بخش انکودر نیز ابتدا یک امبدینگ از ورودی ساخته و سپس آن را به یک ترانسفورمر با توجه خطی^۸ پاس می‌دهد. سپس خروجی ترانسفورمر که یک تنسور به ابعاد $(batch_size, combined_eqn, emb_size)$ است را با میانگین‌گیری در بعد اول تبدیل به تنسوری به ابعاد $(batch_size, emb_size)$ می‌کنیم، تا برای هر پنجره از هر سیگنال‌ها به یک امبدینگ برسیم که برگرفته از اطلاعات کل پنجره در طول زمان است.

۳.۴ مراحل خوراندن داده به مدل

برای آماده‌سازی داده خام و خوراندن آن به مدل BioT می‌بایست چندین کار انجام داد.

پس از تمیزسازی داده‌ها که در بالا گفته شد، برای هر سوژه یک فایل pickle ساخته می‌شود که مجموع ۱۵ فایل می‌شود. هر یک از این فایل‌ها یک دیکشنری به فرمت شکل ۲.۳ است، که هر یک از آنها یک تنسور است. طول تنسور تک بعدی label ۴ برابر ثانیه‌های آن سوژه و ابعاد باقی تنسورهای سیگنال‌ها برابر (فرکانس سیگنال، تعداد کانال‌ها، ۴ برابر ثانیه‌ها) است.

علت ۴ برابر شدن ثانیه‌ها استفاده از ربع ثانیه به عنوان واحد زمانی است. همچنین تعداد کانال‌ها برای همه به جز ACC برابر ۱ است.

تغییر فرکانس فرکانس مناسب برای مدل BioT برابر ۲۰۰ هرتز است. بنابراین برای هر یک از سیگنال‌ها فرکانس سیگنال را به ۲۰۰ تبدیل می‌کنیم. همچنین در کانال‌های مختلف میانگین می‌گیریم (در حقیقت این کار به جز بر ACC بر سیگنال دیگری تاثیری ندارد). در نهایت ابعاد هر یک از سیگنال‌ها به شکل (۲۰۰، ۴ برابر ثانیه‌ها) در می‌آید.

برای تغییر فرکانس دو تابع مختلف استفاده کردیم:

۱. یکی از بهترین روش‌ها برای این کار بهره‌گیری از PCA است. این روش را تنها برای سیگنال‌های سینه می‌شود استفاده کرد، چراکه PCA تنها می‌تواند ابعاد را کم کند و سیگنال‌های می‌چند در بیشترین حالت ۶۴ هرتز هستند.

۲. روش دیگر استفاده از تابع resample از پکیج scipy بود. این تابع از متود فوریه برای تبدیل فرکانس یک سیگنال استفاده می‌کند. از این روش می‌توان برای همه سیگنال‌ها استفاده کرد.

نرمال‌سازی با توجه به خود مدل BioT، نرمال‌سازی به این صورت انجام می‌شود که سیگنال‌ها بر چندک ۹۵- صدم تقسیم می‌شود.

^۸Linear Attention Transformer

تقسیم دادگان همانند تمامی مقالات و کارهای انجام شده بر روی این دادگان، ما هم از روش LOSO استفاده می‌کنیم. بدین ترتیب به ازای هر یک از سوژه‌ها، آن را به عنوان تست و از ۱۴ تای باقی‌مانده، ۱۰ تا را به عنوان داده آموزشی و ۴ تای باقی‌مانده را به عنوان داده validation استفاده می‌کنیم.

۴.۴ استخراج ویژگی‌ها

برای تبدیل داده به فرم مناسب برای آموزش مدل‌های ماشین لرنینگ، یکی از روش‌های پرکاربرد و محبوب، استخراج ویژگی از پنجره‌های سری زمانی است. در کارهای مختلف از پنجره‌های با طول‌های متفاوت استفاده می‌کنند. برای مثال در [۵] از پنجره‌هایی به طول ۱ ثانیه، ۱۰ ثانیه [۶]، و حتی ۳۰ ثانیه [۷] استفاده کردند. در مورد آخر، یکی از علل طول زیاد پنجره به دلیل استفاده از مدل transformer و بهره‌گیری از زمینه^۹ است.

۵.۴ سنجش

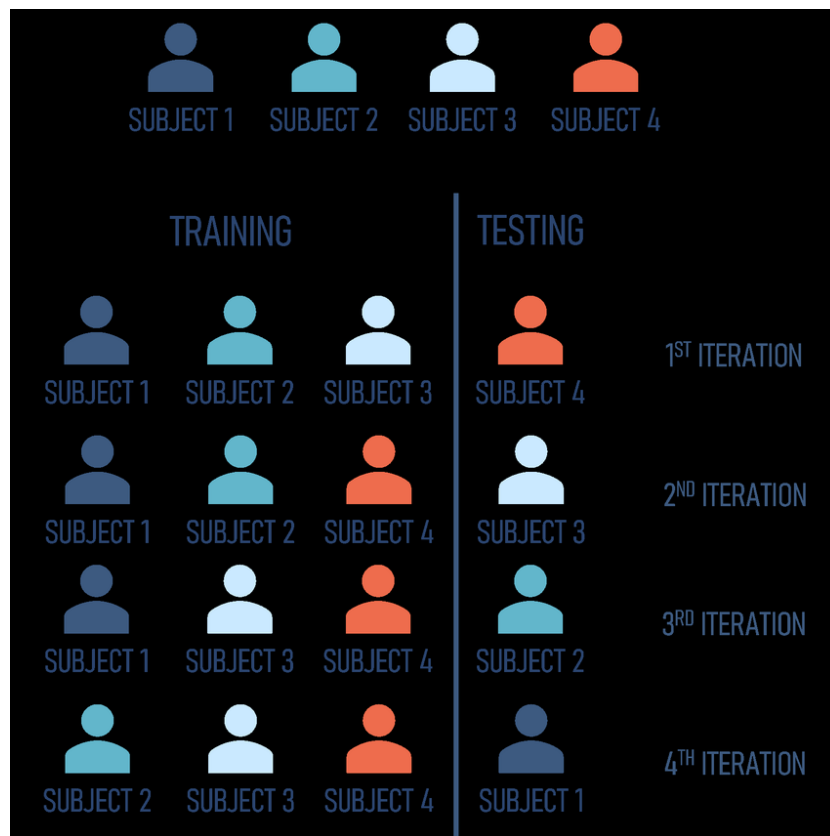
برای سنجش مدل همانند تمامی کارهای بررسی شده، از روش LOSO^{۱۰} و معیار F-1 استفاده می‌کنیم.

LOSOS در روش LOSO هر مرتبه یک سوژه را به عنوان تست و سوژه‌های دیگر برای آموزش و صحت‌سنجی استفاده می‌کنیم. در انتها نمره‌های به دست آمده را میانگین گرفته و نمره نهایی را تعیین می‌کنیم. در شکل ۲.۴ این روش به صورت گرافیکی نمایان شده‌است.

F-1 یکی از معیارهای پرکاربرد در سنجش مدل‌های هوش مصنوعی امتیاز F-1 است. چراکه در آن هم دقت و هم صحت در نظر گرفته می‌شوند و این می‌تواند مدل‌ها را بهتر از معیار دقت به تنهایی ارزیابی کند. این نمره از میانگین هارمونیک دقت و فراخوانی به دست می‌آید که در رابطه زیر نمایش داده شده‌است:

$$F-1 = \frac{2}{\frac{1/Precision}{1/Recall}} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

context^۹
Leave-One-Subject-Out^{۱۰}



شکل ۲.۴: شکل توضیح دهنده روش LOSO

فصل ۵

نتایج

۱.۵ ۲ کلاسه

برای مقایسه آن با هر یک از کارها، این مدل را با استفاده از سیگنال‌ها و پنجره‌های هر یک از کارها مقایسه می‌کنیم.

در ابتدا مقایسه را ۲ کلاسه (استرس و غیر استرس) انجام می‌دهیم. در جدول ۱.۵ نتایج با دو روش اول که پنجره‌های یکسانی داشتند مقایسه شدند. میانگین $f1$ -score برابر ۹۱ درصد شد که دقتی نسبتاً خوب محسوب می‌شود. به دلیل محدودیت‌های محاسباتی به جای قدم‌های ربع ثانیه‌ای از قدم‌های ۶ ثانیه‌ای استفاده شد که اینگونه تعداد داده‌های آموزش به نسبت کار [۸] که بهترین کار بوده، بسیار کمتر (حدوداً ۲۵ برابر) بوده ولی با این حال دقت تنها ۲ درصد کمتر بوده که با کاهش اندازه قدم‌ها این درصد نیز بهتر خواهد شد.

۲.۵ ۳ کلاسه

در قدم بعدی این مدل را برای ۳ کلاس تست کردیم. ۳ کلاس عبارتند از استرس، خوشحالی و عادی. در ادامه این کار را با سیگنال‌های مختلف بررسی کردیم:

۱.۲.۵ سیگنال‌های سینه‌ای

اولین آزمایش که نتایج آن در جدول ۲.۵ آمده، را با بهره‌گیری از PCA و سیگنال‌های سینه انجام دادیم. نتایج آن به نسبت کار [۶] که از سیگنال‌های سینه بهره برده بود، حدود ۱۰٪ امتیاز $F-1$ بهتری داشت.

سپس آزمایش مشابهی را این بار با resample به جای PCA انجام دادیم. بدین صورت که به جای اینکه اجزای اصلی پنجره را به دست آوریم، یک نمونه پریودیک به وسیله فوریه از آن بگیریم و از آن استفاده کنیم.

سوژه	امتیاز F-1	Accuracy
۲	0.78	0.70
۳	0.82	0.79
۴	0.90	0.88
۵	0.96	0.96
۶	0.90	0.89
۷	0.82	0.80
۸	0.84	0.80
۹	0.80	0.73
۱۰	0.96	0.96
۱۱	0.90	0.90
۱۳	0.96	0.96
۱۴	0.70	0.73
۱۵	0.96	0.96
۱۶	0.98	0.98
۱۷	0.70	0.72
مجموع	0.91	0.93

جدول ۱.۵: نتایج کلاسیفیکیشن ۲ کلاسه با مدل BioT با پنجره ۶۰ ثانیه و قدم‌های ربع ثانیه‌ای. سیگنال‌های EDA، BVP و دمای مج استفاده شدند.

در جدول ۳.۵ که از resample استفاده شده، نتایج به مراتب بهتری گرفتیم و F-1 score حدود ۹٪ بهتر شد. این پیشرفت قابل توجه به معنای آن است که تابع بهره گرفته از فوریه resample بسیار بهتر PCA عمل می‌کند. در نهایت با سیگنال‌های سینه به تنهایی، در امتیاز F-1 به ۸۷٪ رسیدیم که از مدل تمام سینه [۶] که از مدل‌های کلاسیفیکیشن کلاسیک مانند حدود ۲۰٪ بهتر است

۲.۲.۵ سیگنال‌های مج

در جدول ۴.۵ نتایج برای سیگنال‌های مج که عبارتند از EDA، BVP و دما، آورده شده است.

با استفاده از این سیگنال‌ها به امتیاز F-1 برابر ۸۹.۰ رسیدیم که نتیجه کمی بهتری نسبت به سیگنال‌های سینه داشتیم. به علت resample کردن سیگنال‌ها، مزیت ۷۰۰ هرتزی بودن سیگنال‌های سینه کمرنگ شد. همین امر باعث این شد که اطلاعات بیشتری از سیگنال‌های مج دست نخورده باقی بماند و همین مسبب دقت نسبتاً بهتر در اینجا بوده است.

سوژه	امتیاز F-1	Accuracy
۲	0.75	0.64
۳	0.54	0.57
۴	0.85	0.78
۵	0.23	0.18
۶	0.77	0.67
۷	0.54	0.59
۸	0.85	0.78
۹	0.77	0.67
۱۰	0.77	0.67
۱۱	0.77	0.67
۱۳	0.85	0.78
۱۴	0.77	0.67
۱۵	0.85	0.78
۱۶	0.85	0.81
۱۷	0.77	0.67
مجموع	0.78	0.71

جدول ۲.۵: همانند ۱.۵، با تفاوت اینکه در اینجا ۳ کلاس را مقایسه کردیم. در اینجا تمام سنسورهای سینه و به وسیله PCA بررسی شده‌اند.

۳.۲.۵ جفت سیگنال‌ها

برای بهتر شدن نتایج، اطلاعات بیشتری را به مدل دادیم و بدین ترتیب هم از سیگنال‌های سینه و هم مچ استفاده کردیم تا از مزایای هر دو استفاده کنیم و زمینه غنی‌تری را برای مدل فراهم آوریم تا نتیجه بهتری را شاهد باشیم. در جدول ۵.۵ نتایج حاصله آورده شده‌است.

با استفاده از هر دو گروه سیگنال‌ها به F-1 score فوق‌العاده ۹۳٪ رسیدیم. این امتیاز بسیار بهتر از امتیازی بود که در دیگر کارهاست. در کار [۸] که بهترین کار در کلاسیفیکیشن ۳ کلاسه بوده که آن هم از ترانسفورمرها استفاده کرده‌است، نتیجه آن ۱۰٪ کمتر از این مدل بوده‌است.

۴.۲.۵ روند

در شکل ۱.۵ دقت، امتیاز Cohen و F-1 در طول روند یادگیری برای ۳ کلاس می‌باشد. حد بالای تعداد epoch ها برابر ۵۰ بوده، ولی همانطور که مشاهده می‌کنیم، ۱۲ تا از آنها قبل از این مقدار همگرا شدند و به ۵۰ نرسیدند که ۱۰ تای آنها به زیر epoch ۲۰ برای همگرایی نیاز داشتند.

نقطه آخر هر یک از نمودارها، آن معیار برای داده تست می‌باشد و دلیل جهش ناگهانی در آخر نمودارها

سوژه	امتیاز F-1	Accuracy
۲	0.75	0.64
۳	0.77	0.67
۴	1.0	1.0
۵	1.0	1.0
۶	0.85	0.81
۷	0.77	0.74
۸	0.69	0.74
۹	0.85	0.85
۱۰	0.77	0.67
۱۱	0.77	0.67
۱۳	1.0	1.0
۱۴	0.46	0.46
۱۵	0.85	0.86
۱۶	0.92	0.92
۱۷	0.69	0.63
مجموع	0.87	0.83

جدول ۳.۵: نتایج ۳ کلاسه با تنظیمات مشابه ۲.۵ با تفاوت اینکه به جای استفاده از PCA، از resample استفاده شده است.

همین نکته است. از بین این ۱۵ تا تنها در ۶ سوژه برای داده تست افت بدی را در معیارها شاهد بودیم که بیانگر انطباق بیش از حد^۱ می باشد.

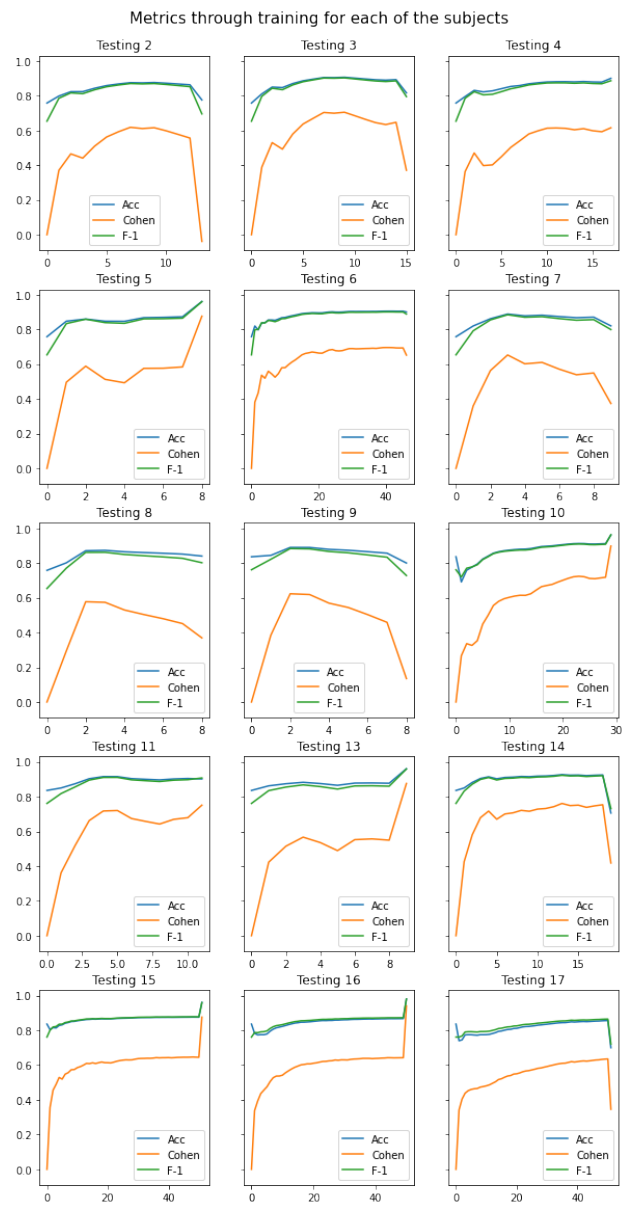
^۱ Over fit

سوژه	امتیاز F-1	Accuracy
۲	0.92	0.92
۳	0.77	0.75
۴	0.85	0.78
۵	0.85	0.78
۶	0.85	0.81
۷	0.85	0.85
۸	1.0	1.0
۹	0.85	0.81
۱۰	0.77	0.67
۱۱	0.77	0.67
۱۳	1.0	1.0
۱۴	0.23	0.087
۱۵	1.0	1.0
۱۶	1.0	1.0
۱۷	0.69	0.63
مجموع	0.89	0.84

جدول ۴.۵: نتایج کلاسیفیکیشن ۳ کلاسه با استفاده از سیگنال‌های مچ. برای تغییر فرکانس از resample بهره گرفته شد. باقی تنظیمات مانند قبل است.

سوژه	امتیاز F-1	Accuracy
۲	0.75	0.73
۳	0.69	0.63
۴	0.92	0.91
۵	0.92	0.91
۶	0.85	0.86
۷	0.85	0.85
۸	0.85	0.86
۹	0.92	0.93
۱۰	0.92	0.92
۱۱	0.77	0.75
۱۳	1.0	1.0
۱۴	1.0	1.0
۱۵	1.0	1.0
۱۶	1.0	1.0
۱۷	0.54	0.55
مجموع	0.93	0.92

جدول ۵.۵: نتایج کلاسیفیکیشن ۳ کلاسه با استفاده از سیگنال‌های سینه و مچ. برای تغییر فرکانس از resample بهره گرفته شد. باقی تنظیمات مانند قبل است.



شکل ۱.۵: روند پیشرفت دقت و معیارهای دیگر در طول فرایند یادگیری. نقطه آخر این نمودارها بیانگر دقت بر روی تست است.

کتاب نامه

grandviewresearch.com [۱]

Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., and Van Laerhoven, K. (2018, October). Introducing wesad, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM international conference on multimodal interaction (pp. 400-408)

Fauzi, M. A., Yang, B., and Yeng, P. (2022, November). Improving Stress Detection Using Weighted Score-Level Fusion of Multiple Sensor. In Proceedings of the 7th International Conference on Sustainable Information Engineering and Technology (pp. 65-71).

Iqbal, T., Redon-Lurbe, P., Simpkin, A. J., Elahi, A., Ganly, S., Wijns, W., and Shahzad, A. (2021). A sensitivity analysis of biophysiological responses of stress for wearable sensors in connected health. IEEE Access, 9, 93567- 93579

Bobade, P., and Vani, M. (2020, July). Stress detection with machine learning and deep learning using multimodal physiological data. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 51-57). IEEE.

Garg, P., Santhosh, J., Dengel, A., and Ishimaru, S. (2021, April). Stress detection by machine learning and wearable sensors. In 26th International Conference on Intelligent User Interfaces-Companion (pp. 43-45).

Behinaein, B., Bhatti, A., Rodenburg, D., Hungler, P., and Etemad, A. (2021, September). A transformer architecture for stress detection from ecg. In Proceedings of the 2021 ACM International Symposium on Wearable Computers (pp. 132-134).

Wu, Y., Daoudi, M., & Amad, A. (2023). Transformer-based self-supervised multimodal representation learning for wearable emotion recognition. *IEEE Transactions on Affective Computing*, 15(1), 157-172.

Rovinska, S., & Khan, N. (2022, July). Affective State Recognition with Convolutional Autoencoders. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 4664-4667). IEEE.

Choi, J., Ahmed, B., & Gutierrez-Osuna, R. (2011). Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE transactions on information technology in biomedicine*, 16(2), 279-286.

Wijsman, J., Grundlehner, B., Penders, J., & Hermens, H. (2013). Trapezius muscle EMG as predictor of mental stress. *ACM transactions on embedded computing systems (TECS)*, 12(4), 1-20.

PRESAGE dataset