

تشخیص عواطف توسط دستگاه‌های پوشیدنی

۲۳ تیر 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.

فهرست مطالب

۳	۱	عواطف
۳	۲	مجموعه داده
۳	۱.۲	دستگاه Respiban
۴	۲.۲	ساعت Empatica E4
۴	۳.۲	پرسشنامه‌ها
۵	۴.۲	طبقه‌بندی
۶	۳	کارهای مشابه
۶	۴	پیش‌پردازش
۶	۱.۴	ساختار مجموعه داده
۶	۲.۴	تمیزسازی داده‌ها
۸	۵	روش‌های آموزش
۸	۱.۵	مدل BioT
۸	۲.۵	ساختار مدل
۹	۱.۲.۵	امیدینگ اولیه
۹	۲.۲.۵	بخش ترانسفورمر
۹	۳.۵	مراحل خوراندن داده به مدل
۱۰	۴.۵	استخراج ویژگی‌ها
۱۰	۶	نتایج
۱۰	۱.۶	۲ کلاس
۱۰	۲.۶	۳ کلاس

۱ عواطف

۲ مجموعه داده

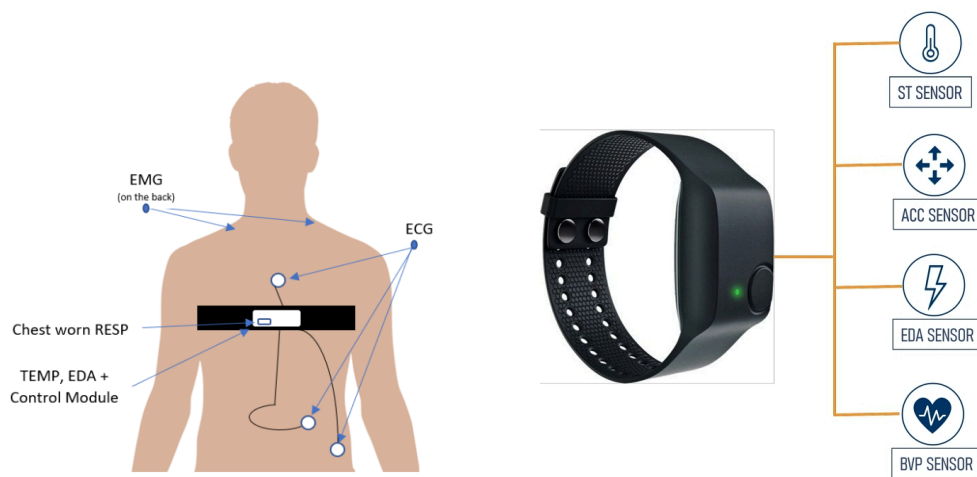
مجموعه داده [1] WESAD یکی از کامل‌ترین مجموعه‌های داده برای تشخیص عواطف است. بیشترین تمرکز و استفاده از این مجموعه برای تشخیص استرس بوده‌است. با این وجود، به جز کلاس استرس و عادی، برای تشخیص کلاس‌های خوشحالی و آرامش نیز می‌توان از این داده‌ها استفاده نمود. علاوه بر آن‌ها، هر شخص پرسشنامه‌هایی نیز پر کرده که این هم می‌تواند باعث خلق مدل‌های جدیدی شود. دو دستگاه اصلی برای فراهم آوردن این داده‌ها مورد استفاده قرار گرفته‌اند: ۱) مچ‌بند Empatica E4 به بسیاری از دانشگاه‌های سراسر دنیا از آن استفاده می‌کنند [2] و ۲) دستگاه Respiban که یکی از پیشرفته‌ترین سنسورهای تحقیقاتی است که بر روی سینه نصب می‌شود [3].

۱.۲ دستگاه Respiban

این دستگاه می‌تواند ۶ عامل را اندازه‌گیری کند. فرکانس ورودی این دستگاه برای همه سنسورهایش ۷۰۰ هرتز می‌باشد. سنسورهای آن به شرح زیر است:

۱. مختصات یابی (Accelerometer)

۲. نوار قلب (Electrocardiogram)



شکل 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.

۳. فعالیت الکتریکی پوست (Electrodermal Activity)

۴. برق‌ماهیچه‌نگار (Electromyogram)

۵. تنفس (Respiration)

۶. دما (Temperature)

۲.۲ 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

۳.۲ 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 بررسی شده، محققان از این پرسشنامه‌ها استفاده‌ای نکردند. در قسمت‌های پیش‌رو این پرسشنامه‌ها را بررسی می‌کنیم:

^۱Blood Volume Pressure

PANAS

سوژه می‌بایست به ۲۶ حس در پرسشنامه، از ۱ تا ۵ امتیاز دهد. این احساسات عبارتند از: فعال، پریشانی، علاقه مند، الهام‌شده، رنجیده، گناهکار، ترسیده، دشمنی، هیجان‌زده، مغرور، کج‌خلق، مشتاق، شرمند، هوشیار، نگران، مصمم، متوجه، عصبی، وحشت‌زده، استرسی، خسته، خوشحال، عصبانی، آزرده‌شدن و ناراحت.

STAI

در این پرسشنامه، سوژه به هر یک از سوال‌های زیر از ۱ تا ۴ نمره می‌دهد:

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

SAM

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

SSSQ

این تست که کوتاه‌شده تست استاندارد SSSQ است، در زمان‌های استرس از شرکت‌کنندگان گرفته شده است. پرسش‌شوندگان به سوال‌های زیر از ۱ تا ۵ نمره می‌دهند:

۱. من متعهد به رسیدن به اهداف عملکردی‌ام هستم
۲. من می‌خواهم در این کار موفق شوم
۳. من انگیزه برای انجام این کار را دارم
۴. من خودم را بروز می‌دهم
۵. من نگران تفکرات دیگران در مورد خودم هستم
۶. من متوجه تاثیری که روی بقیه می‌گذارم هستم

۴.۲ 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.

در جدول 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.

baseline^۲
Amusement^۳
Meditated^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: ثانیه‌های مفید هر یک از سوژه‌ها

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.

۳ 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 نشده‌است. در جدول ۲ کارهای مشابه که از این دادگان استفاده کردند آورده شده‌است.

۴ پیش‌پردازش

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

۱.۴ 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 $H_z 700$ در طول ۶۴۹۳ ثانیه هستند. آرایه‌های ACC و BVP به ترتیب ۳۲ و ۶۴ هرتز و دو سیگنال دیگر هر دو ۴ هرتز هستند.

۲.۴ 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 ۳ تغییر می‌دهیم.

```

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

```

شکل ۲: 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 ۳: ساختار دادگان پس از تغییر

نام کار	سیگنال‌ها	پنجره	عملکرد f-1
Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection[1]	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[7]	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[8]	All the signals	پنجره: ۱ ثانیه، قدم: ۱ ثانیه	3class: 82.82
Stress Detection by Machine Learning and Wearable Sensors	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[5]	ECG	پنجره: ۳۰ ثانیه، قدم: ۱ ثانیه	2class: 83.3

جدول ۲: کارهای مشابه بر روی دادگان WESAD

۵ روش‌های آموزش

۱.۵ 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 می‌توان برای انواع سیگنال های حیاتی بهره گرفت. در شکل ۴ ساختار این مدل را مشاهده می‌کنید. در نیم تصویر بالا، ماژول توکنایز کردن سیگنال‌هاست، که با انجام دوباره نمونه‌گیری^۵، نرمال کردن، توکنایز کردن و تخت کردن^۶، آن را تبدیل به جملات می‌کند.

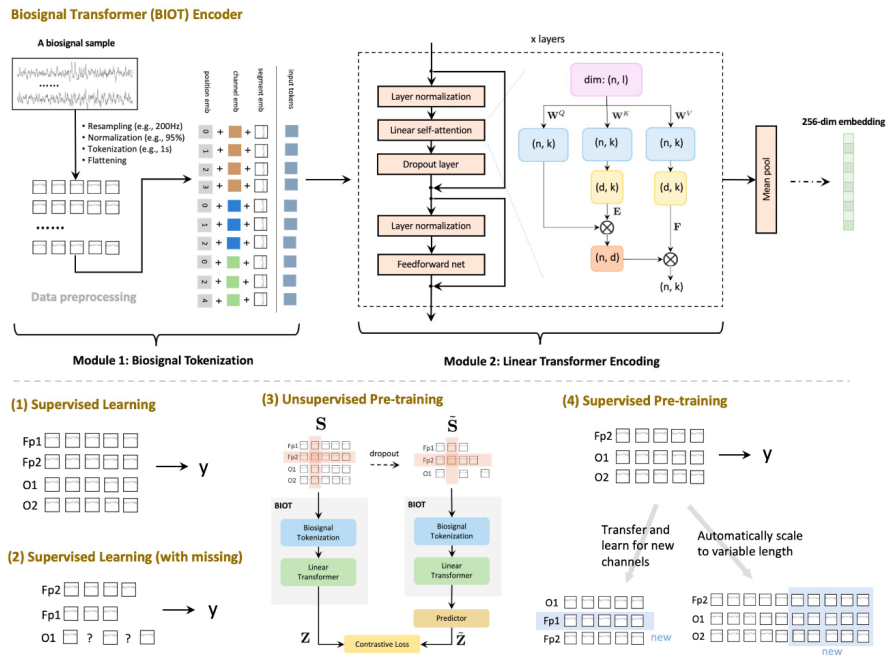
سپس تعامل بین این جملات با استفاده از ماژول ترنسفورمر خطی (نیم‌تصویر پایین) یاد گرفته می‌شوند. این مدل به صورت با نظارت می‌تواند داده‌های مختلف کامل و ناقص را برای پیش‌آموزش و fine-tuning بپذیرد.

۲.۵ ساختار مدل

مدل از دو بخش انکودر و سر کلاسیفیکیشن^۷ تشکیل شده است.

سر کلاسیفیکیشن از یک تابع فعالساز ELU^۸ و یک لایه خطی تشکیل شده است که با تبدیل $xA^T + b$ انکودینگ ساخته شده را به یک وکتور با اندازه تعداد کلاس‌ها می‌برد که بیانگر احتمال این است که آن ورودی در هر کدام از آن کلاس‌ها باشد.

resampling^۵
flattening^۶
Classification Head^۷
Exponential Linear Unit^۸



شکل ۴: معماری شبکه BioT

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 پیش‌رو به تفصیل این دو بخش توضیح داده شده‌اند.

۱.۲.۵ امبدینگ اولیه

۲.۲.۵ بخش ترانسفورمر

بخش انکودر نیز ابتدا یک امبدینگ از ورودی ساخته و سپس آن را به یک ترانسفورمر با توجه خطی^۹ پاس می‌دهد. سپس خروجی ترانسفورمر که یک تنسور به ابعاد $(batch_size, combined_eqien, emb_size)$ 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.

۳.۵ 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 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)

^۹ Linear Attention Transformer

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 تأثیری ندارد). در نهایت ابعاد هر یک از سیگنال‌ها به شکل (۲۰۰, ۴) برابر ثانیه‌ها) در می‌آید.

برای تغییر فرکانس از تابع resample از پکیج scipy بهره بردیم. این تابع از متود فوریه برای تبدیل فرکانس یک سیگنال استفاده می‌کند.

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

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.

۴.۵ 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 [۴] از پنجره‌هایی به طول ۱ ثانیه، ۱۰ ثانیه [۵]، و حتی ۳۰ ثانیه [۶] استفاده کردند. در مورد آخر، یکی از علل طول زیاد پنجره به دلیل استفاده از مدل transformer و بهره‌گیری از زمینه^{۱۰} است.

۶ نتایج

۱.۶ 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 ۳ 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 [۷] 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.

۲.۶ ۳ کلاسه

با بهره‌گیری از این مدل قدرتمند به نتایج بسیار بهتری در کلاس‌بندی با ۳ کلاس خواهیم رسید. ۳ کلاس عبارتند از استرس، خوشحالی و عادی.

^{۱۰}context

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 ۳: نتایج BioT با پنجره ۶۰ ثانیه و قدم‌های ربع ثانیه‌ای. سیگنال‌های BVP، EDA و دمای مج استفاده شدند.

در جدول ۴ نتایج برای ۳ کلاس با تنظیمات مشابه ۳ آورده شده است. امتیاز f-1 به طور فوق‌العاده‌ای به بیش از ۹۰٪ رسید که بسیار بهتر از امتیازهای ۳ کلاسه دیگر مقالات است!

در شکل 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 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^{۱۱} می‌باشد.

^{۱۱} Over fit

Accuracy	امتیاز F-1	سؤزه
۹۱.۰	۹۲.۰	۲
۷۳.۰	۷۸.۰	۳
۹۶.۰	۹۶.۰	۴
۹۸.۰	۹۸.۰	۵
۸۰.۰	۸۴.۰	۶
۸۳.۰	۸۴.۰	۷
۸۳.۰	۸۶.۰	۸
۷۳.۰	۸۰.۰	۹
۸۷.۰	۸۷.۰	۱۰
۷۵.۰	۷۳.۰	۱۱
۹۶.۰	۹۶.۰	۱۳
۵۹.۰	۵۷.۰	۱۴
۹۶.۰	۹۶.۰	۱۵
۰.۱	۰۰.۱	۱۶
۷۲.۰	۷۰.۰	۱۷
۹۰.۰	۹۱.۰	مجموع

جدول ۴: همانند ۳، با تفاوت اینکه در اینجا ۳ کلاس را مقایسه کردیم.

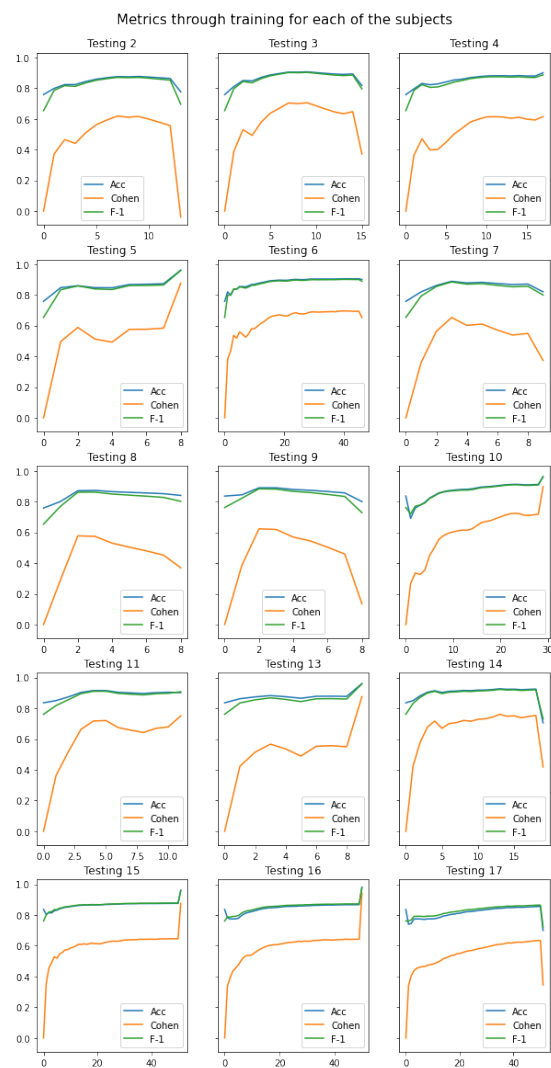
مراجع

- Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., and Van Laerhoven, [۱]
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 [۲]
Detection Using Weighted Score-Level Fusion of Multiple Sensor. In Pro-
ceedings 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., [۳]
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 learn- [۴]
ing 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 [۵]
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. [۶]
(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 [v]
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 [a]
Convolutional Autoencoders. In 2022 44th Annual International Conference
of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 4664-
4667). IEEE.



شکل ۵: روند پیشرفت دقت و معیارهای دیگر در طول فرایند یادگیری. نقطه آخر این نمودارها بیانگر دقت بر روی تست است.