



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)  
دانشکده مهندسی کامپیوتر

گزارش درس روش پژوهش و ارائه

بررسی سیستم‌های تشخیص حالت چهره برای شناسایی احساسات پایه‌ای

نگارش  
کیمیا حسینی

استاد راهنما  
دکتر مهدی صدیقی

خرداد ۱۴۰۳

## تقدیر و تشکر

لازم میدانم از زحمات دلسوزانه جناب آقای دکتر مهدی صدیقی تشکر کنم که مطالب بسیاری به من آموختند. از زحماتشان صمیمانه سپاسگزارم و برایشان آرزوی توفیق روزافزون دارم.

کیما حسینی

خرداد ۱۴۰۳

## چکیده

تشخیص احساس‌های چهره<sup>۱</sup> یک حوزه بسیار مهم بینایی کامپیوتر و تعامل انسان و کامپیوتر است که احساس‌های انسان را از طریق حالت‌های چهره شناسایی و تفسیر می‌کند. این گزارش به بررسی سیستم‌های تشخیص احساس چهره موجود با بیان تکنیک‌های مختلف یادگیری ماشین پرداخته است. در این گزارش سه گام اصلی اکتشاف مجموعه داده‌ها، اجرای روش‌های مختلف در پیش‌پردازش مانند ترازبندی<sup>۲</sup> و عادی‌سازی<sup>۳</sup> چهره، استخراج ویژگی مانند الگوهای باینری محلی<sup>۴</sup> و هیستوگرام گرادیان‌های جهت‌دار<sup>۵</sup>، و طبقه‌بندی مانند ماشین‌های بردار پشتیبان<sup>۶</sup> و شبکه‌های عصبی کانولوشنی<sup>۷</sup> و بازگشتی<sup>۸</sup> مطرح شده‌اند.

برای ارزیابی، مجموعه داده‌های کلاسیک این حوزه مانند CK+ و FER2013 مطرح شده‌اند. نهایتاً با توجه به رویکردهای مورد بررسی و ادغام آنها می‌توان پیشرفت‌های چشمگیری در دقت بدست آورد.

## واژه‌های کلیدی:

تشخیص احساس‌های چهره، یادگیری ماشین، استخراج ویژگی، هیستوگرام گرادیان‌های جهت‌دار، ماشین‌های بردار پشتیبان، شبکه‌های عصبی کانولوشنی

---

<sup>۱</sup> Face emotion recognition (FER)

<sup>۲</sup> Alignment

<sup>۳</sup> Normalization

<sup>۴</sup> Local binary pattern (LBP)

<sup>۵</sup> Histogram of oriented gradients (HOG)

<sup>۶</sup> Support vector machines (SVM)

<sup>۷</sup> Convolutional neural network (CNN)

<sup>۸</sup> Convolutional neural network (CNN)

صفحه	فهرست مطالب
أ	چکیده.....
۴	فصل اول مقدمه .....
۸	فصل دوم مروری بر سیستم‌های تشخیص حالت چهره .....
۹	۱-۲- طبقه‌بندی احساس‌های پایه‌ای.....
۱۰	۲-۲- کاربردهای FER در صنایع مختلف.....
۱۲	فصل سوم داده‌ها و پیش‌پردازش .....
۱۳	۱-۳- مجموعه داده‌های رایج در FER.....
۱۶	۲-۳- روش‌های جمع‌آوری داده.....
۱۶	۳-۳- تکنیک‌های پیش‌پردازش تصویر.....
۱۷	۱-۳-۳- تشخیص چهره.....
۱۹	۲-۳-۳- ترازبندی و عادی‌سازی.....
۲۱	فصل چهارم روش‌های استخراج ویژگی.....
۲۲	۱-۴- ویژگی‌های هندسی و شکلی.....
۲۲	۱-۴-۱- نقاط عطف چهره.....
۲۳	۲-۴-۱- ویژگی‌های مبتنی بر تقارن.....
۲۴	۲-۴-۲- ویژگی‌های بافتی.....
۲۴	۱-۴-۲- هیستوگرام گرادیان‌های جهت‌دار (HOG).....
۲۵	۲-۴-۲- الگوهای باینری محلی (LBP).....
۲۷	فصل پنجم الگوریتم‌های طبقه‌بندی و شناسایی احساس‌ها.....
۲۸	۱-۵- روش‌های کلاسیک طبقه‌بندی.....
۲۸	۱-۵-۱- ماشین بردار پشتیبان (SVM).....
۲۹	۲-۵-۱- درخت تصمیم‌گیری.....
۳۰	۲-۵- یادگیری عمیق برای طبقه‌بندی احساس‌ها.....
۳۱	۱-۵-۲- شبکه‌های عصبی کانولوشنی (CNN).....
۳۲	۲-۵-۲- شبکه‌های عصبی بازگشتی (RNN).....
۳۴	فصل ششم جمع‌بندی و نتیجه‌گیری.....
۳۵	منابع و مراجع.....

صفحه

فهرست اشکال

شکل ۳-۱- نه مجموعه داده برای تشخیص احساس‌های چهره.....	۱۵
شکل ۳-۲- ویژگی‌های Haar-like.....	۱۷
شکل ۳-۳- مثالی از یکسان‌سازی هیستوگرام.....	۲۰
شکل ۴-۱- استفاده از ASM برای یافتن نقاط عطف چهره‌ها.....	۲۳
شکل ۴-۲- هیستوگرام گرادیان‌های جهت‌دار.....	۲۴
شکل ۴-۳- تقسیم بخش موردنظر تصویر به بلوک‌ها و سلول‌ها و تشکیل آرایه‌ای از هیستوگرام در هر پنجره.....	۲۵
شکل ۴-۴- فرایند LBP.....	۲۶
شکل ۵-۱- خط بهینه در طبقه بندی SVM.....	۲۹
شکل ۵-۲- معماری شبکه عصبی کانولوشنی در FER.....	۳۱

## فصل اول

### مقدمه

## مقدمه

تشخیص احساس‌های چهره به فرایند شناسایی و تحلیل احساس‌های انسانی بر اساس تغییرهای چهره با استفاده از بینایی ماشین و هوش مصنوعی اطلاق می‌شود. این فناوری از مدل‌های محاسباتی برای تشخیص و تفسیر شش احساس اصلی انسان شامل غم، تعجب، نفرت، شادی، ترس، و خشم بهره می‌برد. FER عمدتاً از تصاویر چهره استفاده می‌کند. زیرا حالت‌های بصری یکی از کانال‌های اصلی اطلاعات در ارتباطات بین فردی هستند [۱].

اهمیت FER در گستره وسیع کاربردهای آن و توانایی قابل توجه در تعامل انسان و کامپیوتر<sup>۱</sup> است. احساس‌ها نقش حیاتی در ارتباط‌ها و فرایندهای تصمیم‌گیری انسانی ایفا می‌کنند. با امکان تشخیص و پاسخ به احساس‌های انسانی، سیستم‌های FER می‌توانند تجربه کاربری را در حوزه‌های مختلف بهبود ببخشند. امروزه روش‌شناسی تشخیص چهره به طور گسترده‌ای در خدمت‌دهی به مشتری‌ها، خدمات آموزشی و کاربردهای مرتبط با امنیت به کار می‌رود [۲]. در واقعیت افزوده<sup>۲</sup> [۳] و واقعیت مجازی<sup>۳</sup> [۴] سیستم‌های تشخیص احساس‌های چهره می‌توانند با تطبیق محیط مجازی با احساس‌های کاربر، تجربه‌های کاربری واقعی‌تر و جذاب‌تری ایجاد کنند. از دیگر کاربردهای سیستم‌های FER می‌توان به افزایش ایمنی و راحتی رانندگان با نظارت بر وضعیت‌های احساسی آنها و ارائه هشدارها یا مداخله‌ها در صورت نیاز اشاره کرد [۵]. در محیط‌های درمانی برای نظارت بر وضعیت احساسی بیماران و ارائه مداخلات به موقع از سیستم‌های تشخیص حالت چهره استفاده می‌شود.

سیستم‌های تشخیص حالت چهره از روش‌های مختلفی استفاده می‌کنند که به طور کلی به دو دسته تکنیک‌های مبتنی بر ویژگی و مبتنی بر مدل تقسیم می‌شوند [۲]. برای روش‌های استخراج ویژگی می‌توان الگوریتم‌های هیستوگرام گرادیان‌های جهت‌دار و الگوهای باینری محلی را نام برد. این روش‌ها از ویژگی‌های هندسی، که روابط مکانی بین نقاط کلیدی چهره را تحلیل می‌کنند، تا ویژگی‌های ظاهری که بر اطلاعات بافت و رنگ از تمام یا بخش‌های خاصی از چهره تمرکز دارند، متفاوت هستند. تکنیک‌های مبتنی بر مدل، شبکه‌های عصبی کانولوشنی، شبکه‌های عصبی بازگشتی، و ماشین بردار پشتیبان را شامل می‌شود.

<sup>1</sup> Human-computer interaction (HCI)

<sup>2</sup> Augmented Reality (AR)

<sup>3</sup> Virtual reality (VR)

یکی از گام‌های مهم برای اجرای الگوریتم‌های FER پیش‌پردازش داده‌هاست. تکنیک‌های پیش‌پردازش برای بهبود کیفیت تصاویر ورودی و اطمینان از استخراج دقیق ویژگی‌ها بسیار مهم هستند. مراحل پیش‌پردازش عموماً شامل تشخیص چهره، تراز کردن و عادی‌سازی است. در گام اول ابتدا در تصویرها چهره‌های موجود در آن تشخیص داده می‌شوند. در گام دوم به شناسایی نقاط کلیدی چهره و تراز کردن آن پرداخته می‌شود. در مرحله آخر عادی‌سازی انجام می‌شود تا چهره در یک جهت ثابت همسو شود. عادی‌سازی کمک می‌کند تا تأثیر عوامل اضافی کاهش یابد و تمرکز تحلیل بر روی حالات چهره باشد [۱].

با وجود پیشرفت‌های قابل توجه، همچنان چالش‌ها و محدودیت‌های زیادی در زمینه FER باقی مانده است. یکی از مسائل عمده، نیاز به مجموعه داده‌های بزرگ و توان محاسباتی زیاد برای آموزش مدل‌های یادگیری عمیق است [۱]. علاوه بر این، انتخاب پارامترهای مناسب، مانند نرخ یادگیری و اندازه‌های هسته<sup>۴</sup> نیاز به تخصص و تجربه زیادی دارد [۱].

انتظار می‌رود در تحقیقات آینده در FER بر غلبه بر محدودیت‌های فعلی و بررسی کاربردهای جدید تمرکز شود. توسعه رویکردهای پویا که CNNها را با شبکه‌های حافظه طولانی کوتاه مدت<sup>۵</sup> ترکیب می‌کنند، می‌تواند شناسایی ویژگی‌ها در دنباله‌های ویدئویی را بهبود بخشد و منجر به عملکرد بهتر در FERهای پویا شود. همچنین سیستم‌های تشخیص حالت چهره می‌توانند دامنه احساسات فراتر از شش احساس اصلی را تشخیص دهند. ترکیب روش‌های چندوجهی که داده‌های چهره، گفتار و فیزیولوژیکی را ترکیب می‌کنند می‌تواند دقت و قابلیت اطمینان سیستم‌های تشخیص احساس‌ها را افزایش دهد.

استفاده از سیستم‌های FER چالش‌های اخلاقی زیادی را به همراه دارد. مسائل حریم خصوصی از اهمیت بالایی برخوردار هستند، زیرا استفاده گسترده از FER می‌تواند منجر به نظارت غیرمجاز و جمع‌آوری داده‌های شخصی بدون رضایت شود. همچنین خطر سوءاستفاده در نظارت و اجرای قانون وجود دارد، جایی که سوگیری الگوریتمی<sup>۶</sup> ممکن است منجر به برخورد ناعادلانه با گروه‌های خاصی شود. برای رسیدگی به این چالش‌های اخلاقی، توسعه سیاست‌های حریم خصوصی قوی و تلاش‌ها برای اطمینان از عدالت و شفافیت در فناوری‌های FER ضروری است.

<sup>۴</sup> kernel

<sup>۵</sup> Long Short-Term Memory (LSTM)

<sup>۶</sup> Algorithmic bias



اکنون به توضیح اجمالی از هر فصل این پژوهش پرداخته می‌شود. پس از فصل مقدمه در فصل دوم، پس از تعریف سیستم‌های FER به طبقه‌بندی احساسات پایه‌ای و کاربردهای FER در صنایع مختلف پرداخته شده است. در فصل بعد، معرفی مجموعه داده‌های رایج در FER، بررسی روش‌های جمع‌آوری داده و تکنیک‌های پیش‌پردازش بررسی شده است. در فصل چهارم، به روش‌های استخراج ویژگی‌ها پرداخته شده است. در فصل بعدی بررسی الگوریتم‌های طبقه‌بندی مانند ماشین بردار پشتیبان، درخت تصمیم‌گیری و یادگیری عمیق مورد بررسی قرار گرفته شده است. در نهایت خلاصه‌ای از یافته‌های پژوهش ارائه گردیده است.

## فصل دوم

### مروری بر سیستم‌های تشخیص حالت چهره

## مروری بر سیستم‌های تشخیص حالت چهره

سیستم‌های تشخیص حالت چهره ابزارهای مهمی در تحلیل احساسات انسانی و تعاملات دیجیتالی هستند. این سیستم‌ها با استفاده از تکنیک‌های بینایی کامپیوتر و هوش مصنوعی حالت‌های مختلف چهره را تشخیص و تحلیل می‌کنند تا وضعیت‌های احساسی افراد مانند شادی، غم، خشم و تعجب را تعیین کنند. در ادامه به طبقه‌بندی احساسات پایه‌ای و کاربردهای FER در صنایع مختلف پرداخته می‌شود تا دیدگاه جامعی از این فناوری ارائه شود.

### ۲-۱- طبقه‌بندی احساس‌های پایه‌ای

احساس‌های پایه‌ای جزء اساسی تجربه انسانی هستند و نقش مهمی در ارتباط‌ها و تعامل‌های بین فردی ایفا می‌کنند. بر اساس نظریه احساسات پایه، شش احساس اصلی وجود دارد که به طور جهانی توسط انسان‌ها شناخته و ابراز می‌شوند: شادی، غم، خشم، ترس، نفرت، و تعجب.

شادی: شادی با ویژگی‌های لبخند زدن، گونه‌های بالاآمده، چهره آرام و گاهی خنده بیان می‌شود. شادی به دلیل حالت چهره مشخص، یکی از آسان‌ترین احساسات برای شناسایی است.

غم: این احساس با اشک، پلک‌های پایین‌آمده و کمی کشیدن گوشه‌های لب‌ها به سمت پایین شناسایی می‌شود.

خشم: خشم از طریق ابروهای گره‌خورده، لب‌های محکم بسته‌شده و فک فشرده و گاهی فریاد یا رفتار تهاجمی بیان می‌شود.

ترس: از طریق چشمان بازشده، دهان باز و حالت بدنی تنش‌زا بیان می‌شود.

تعجب: این احساس با ابروهای بالارفته، چشمان بازشده و دهان باز شناسایی می‌شود.

نفرت و انزجار: از طریق بینی چین‌خورده، لب‌های بالاآمده و گاهی اوقات زبان بیرون‌زده نشان داده می‌شود.

## ۲-۲- کاربردهای FER در صنایع مختلف

فناوری تشخیص احساس‌ها از طریق چهره به دلیل توانایی تحلیل و تفسیر دقیق احساس‌های انسانی، به طور فزاینده‌ای در صنایع مختلف مورد استفاده قرار می‌گیرد. در ادامه چند صنعت که FER در آنها کاربرد گسترده‌ای دارد بررسی شده است.

FER نقش مهمی در تقویت تعامل انسان و کامپیوتر با قادر ساختن سیستم‌ها برای پاسخگویی به حالت‌های احساسی کاربران دارد. این قابلیت باعث می‌شود تا تعامل‌های طبیعی‌تری بین انسان و ماشین وجود داشته باشد. به عنوان مثال، فناوری ردیابی نگاه چشم در حال حاضر به اندازه کافی پیشرفته است که چندین راه حل تجاری مانند Tobii در دسترس است [۶]. این کاربرد به ویژه در خدمت‌دهی به مشتری‌ها استفاده می‌شود؛ جایی که درک و پاسخ به احساس‌های مشتری می‌تواند به طور قابل توجهی تجربه کاربری را بهبود ببخشد.

فناوری تشخیص احساسات چهره کاربردهای قابل توجهی در صنعت سلامت و مراقبت‌های بهداشتی نیز دارد. تجزیه و تحلیل حالت‌های چهره بیماران می‌تواند برای تشخیص و نظارت بر شرایط سلامت روان بسیار مهم باشد. از این فناوری برای مشاهده حالت‌های عاطفی بیماران، به ویژه در افرادی که ممکن است در بیان کلامی خود مشکل داشته باشند، مانند افراد مبتلا به اوتیسم یا اختلالات عصبی خاص استفاده می‌شود [۷]. در سال‌های اخیر، سیستم‌های تشخیص مبتنی بر بینایی کامپیوتری در چندین بیمارستان و کلینیک‌های پوست‌شناسی مورد استفاده قرار گرفته‌اند [۸].

همچنین صنعت خودرو از فناوری FER برای افزایش ایمنی و راحتی رانندگان استقبال کرده است. سیستم‌های پیشرفته دستیار راننده<sup>۱</sup> از تشخیص احساس‌های چهره برای نظارت بر حالت چهره رانندگان و تشخیص علائم خستگی، حواس پرتی، یا استرس استفاده می‌کند. FER با هشدار دادن به رانندگان یا انجام اقدامات اصلاحی، مانند تنظیم محیط خودرو برای جلوگیری از تصادفات، به تجربه رانندگی ایمن‌تر کمک می‌کند. در این فرایند اجزای صورت که از اهمیت بیشتری برخوردارند و به عنوان مؤثرترین آنها برای خواب‌آلودگی در نظر گرفته می‌شوند، در فریم‌های توالی ویدیویی استخراج و ردیابی می‌شوند [۵].

صنایع سرگرمی و بازی نیز از فناوری FER بهره می‌برند. در بازی‌های ویدیویی و واقعیت مجازی، می‌توان از FER برای ایجاد تجربیات موثرتر و تعاملی با تطبیق محیط بازی و چالش‌ها با وضعیت احساسی بازیکن استفاده

<sup>۱</sup> Advanced driving assistance systems (ADAS)

کرد [۴]. به عنوان مثال، اگر تشخیص داده شود که بازیکن خسته است یا اگر بازیکن ناامید به نظر برسد، بازی می‌تواند چالش برانگیزتر شود.

در محیط‌های آموزشی، FER می‌تواند برای بهبود تعامل و یادگیری دانش‌آموزان استفاده شود. با شناخت پاسخ‌های عاطفی دانش‌آموزان در طول درس، معلم‌ها می‌توانند روش‌های تدریس خود را متناسب با نیازهای فردی تنظیم کنند. به طور مثال FER می‌تواند با شناسایی احساس‌هایی مثل گیجی یا بی‌حوصلگی دانش‌آموزان در لحظه، به حفظ توجه و انگیزه آنها در کلاس کمک کند.

در نتیجه، سیستم‌های تشخیص چهره نقش کلیدی در تجزیه و تحلیل احساس‌های انسان و بهبود تجربه کاربری در تعاملات دیجیتالی دارند. تکنیک‌های تشخیص حالت چهره از بینایی کامپیوتر و هوش مصنوعی برای تشخیص و شناسایی حالت‌های چهره و احساس‌های پایه‌ای استفاده می‌کنند. کاربردهای FER در صنایع مختلف بسیار گسترده است، از بهبود تعامل انسان و کامپیوتر و خدمات مشتری گرفته تا مراقبت‌های بهداشتی، ایمنی خودرو، سرگرمی و آموزش. از آنجایی که فناوری FER همچنان در مسیر توسعه قرار دارد، ادغام آن در بخش‌های مختلف صنایع با افزایش درک و واکنش مناسب‌تر ماشین به احساس‌های انسانی همراه خواهد بود.

## فصل سوم

### داده‌ها و پیش‌پردازش

## داده‌ها و پیش‌پردازش

در این فصل، مجموعه داده‌ها به همراه روش‌های پیش‌پردازی که معمولاً در سیستم‌های تشخیص حالت چهره استفاده می‌شوند، بررسی می‌شوند. ابتدا به مجموعه داده‌های عمومی و همچنین سایر مجموعه‌های داده که بر دامنه‌های خاص تمرکز می‌کنند، پرداخته می‌شود. سپس نحوه جمع‌آوری داده‌ها و چالش‌های بدست آوردن داده‌های باکیفیت مورد بررسی قرار می‌گیرد. بخش آخر پیش‌پردازش است که جزء مهمی از سیستم‌های FER است. این مرحله شامل تشخیص چهره، تراز چهره، و عادی‌سازی چهره است.

### ۳-۱- مجموعه داده‌های رایج در FER

تشخیص احساس‌های چهره در درجه اول به مجموعه داده‌های قوی و متنوع برای آموزش و ارزیابی مدل‌ها متکی است. چنین مجموعه‌های داده طیف گسترده‌ای از حالت‌های چهره را در شرایط مختلف ارائه می‌دهند. از این رو دقت و قابلیت اطمینان سیستم‌های FER را بهبود می‌بخشند. برخی از این مجموعه داده‌ها عمومی هستند و در بسیاری از مطالعه‌های مربوط به FER کاربرد دارند. برخی دیگر تخصصی هستند و هدفشان استفاده در شرایط خاصی است.

#### مجموعه داده‌های عمومی

مجموعه داده CK+ (Cohn-Kanade) شامل مجموعه‌ای از حالت‌های چهره استخراج‌شده در محیط‌های کنترل‌شده است. ۵۹۳ سکانس از ۱۲۰ سوژه را در خود جای داده است. این مجموعه داده شامل تصاویر از همه نژادها و جنسیت‌هاست. تصاویر با کیفیت بالا و با نشانه‌گذاری مناسب نقاط چهره باعث می‌شود که این مجموعه داده برای آموزش و اعتبارسنجی در FER بسیار مفید باشد [۹].

یکی دیگر از مجموعه داده‌های پرکاربرد FER2013 است که با استفاده از Google Image Search API ایجاد شده است. این یک پایگاه داده شامل ۳۵۸۸۷ تصویر صورت سیاه و سفید است که هر کدام با یکی از هفت دسته احساسی: شادی، غم، خشم، ترس، تعجب، نفرت؛ انزجار، و خنثی برچسب زده شده‌اند. این مجموعه داده شامل یک مجموعه آموزشی، یک مجموعه تست عمومی و یک مجموعه تست خصوصی است که معیار کاملی را برای ارزیابی مدل‌های FER فراهم می‌کند. از آنجایی که FER2013 بسیار متنوع و بزرگ است، برای آموزش مدل‌هایی از یادگیری عمیق که به داده‌های زیادی نیاز دارند، مناسب است [۹].

## مجموعه داده‌های تخصصی FER

جدا از مجموعه داده‌های عمومی، چندین مورد دیگر برای مقابله با چالش‌های خاص در تشخیص حالت چهره همچون انسداد و تغییر در روشنایی یا برای تشخیص حالت‌های ظریف و غیرپایه‌ای پیشنهاد شده‌اند.

مجموعه داده Bosphorus از این نظر منحصر به فرد است که شامل تصاویر دو بعدی و سه بعدی مربوط به فرایندهای مربوط به تشخیص احساس‌ها، تشخیص حرکت چهره و بازسازی چهره سه بعدی است. این پایگاه داده شامل اطلاعاتی از ۱۰۵ موضوع و ۴۶۶۶ تصویر است که وضعیت‌های مختلف را پوشش می‌دهد. هر فرد در این مجموعه داده حداقل ۳۵ احساس چهره ثبت شده دارد و تعداد قابل توجهی از آنها بازیگران حرفه‌ای هستند. در این مجموعه داده، حالت‌های مختلف سر و انسدادهای مختلف صورت همچون عینک، دست، مو، سبیل و ریش نیز وجود دارد [۹].

مجموعه داده SMIC<sup>1</sup> بر روی حالت‌های ریز، حرکات بسیار سریع و غیرارادی صورت متمرکز است که احساسات واقعی را آشکار می‌کند. این مجموعه داده شامل ۱۶۴ سکانس ویدیویی از ۱۶ سوژه است [۹].

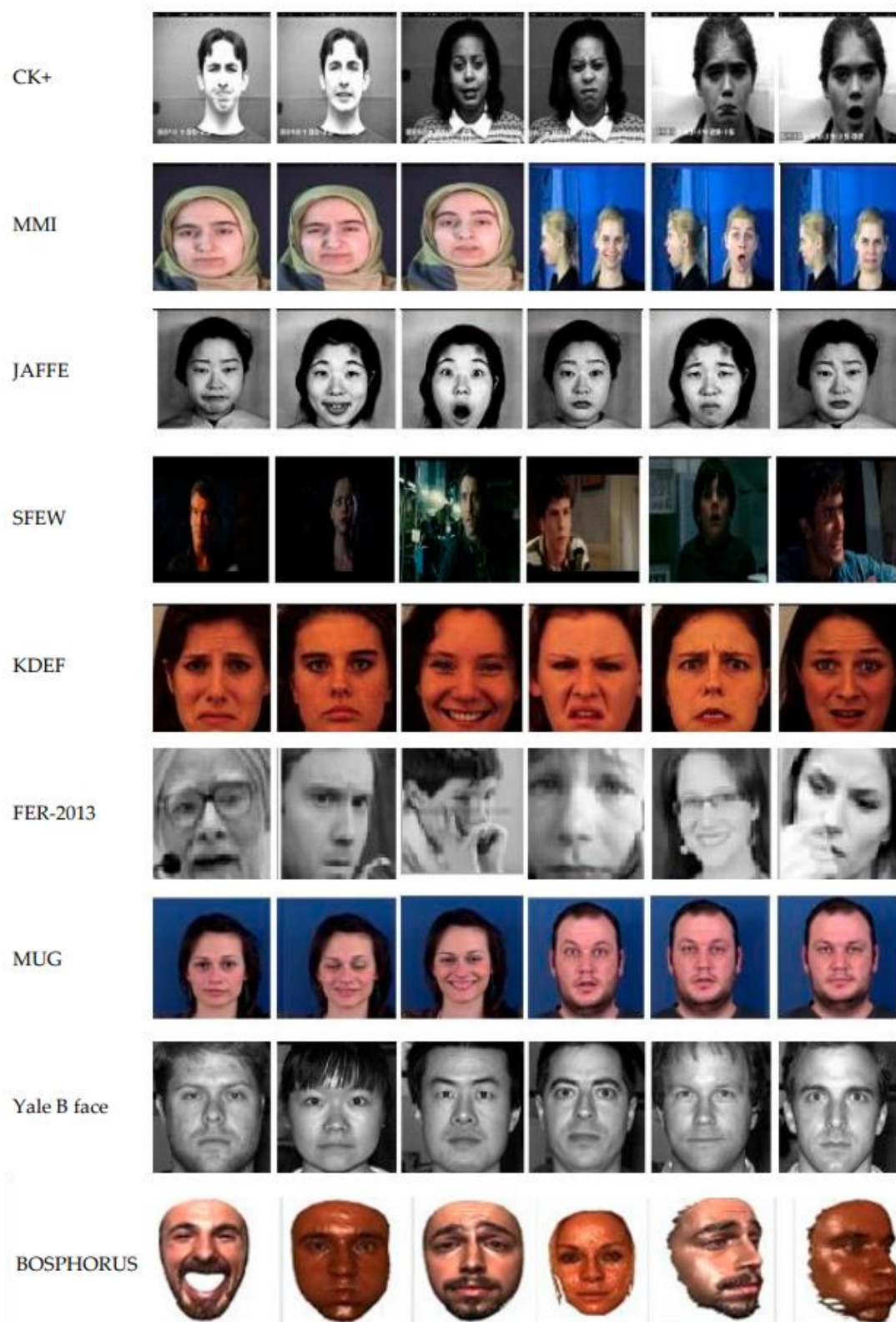
مجموعه داده‌های BBC از آزمایش لبخند جعلی که در وبسایت BBC انجام شد، بدست آمده است. این پایگاه داده شامل ۲۰ ویدیو، ۱۰ مورد واقعی و ۱۰ مورد لبخند ساختگی، است که با وضوح  $286 \times 314$  پیکسل با سرعت ۲۵ فریم در ثانیه جمع‌آوری شده است [۹].

SFEW<sup>2</sup> مجموعه داده‌ای است که برای ارائه شرایط واقعی‌تر و چالش برانگیزتر برای FER طراحی شده است. تصاویر موجود در مجموعه داده‌ها از فیلم‌ها استخراج می‌شوند، بنابراین احساس‌های هفت‌گانه چهره تحت شرایط مختلف نور، تغییرات حالت سر، و انسداد نشان داده می‌شوند. SFEW شامل برچسب‌هایی برای هفت احساس اصلی است و می‌تواند برای توسعه و آزمایش سیستم‌های FER که برای کاربردهای واقعی در نظر گرفته شده‌اند بسیار مفید باشد [۹].

<sup>1</sup> Spontaneous micro-expression corpus

<sup>2</sup> Static facial expressions in the wild





شکل ۳-۱: نه مجموعه داده برای تشخیص احساس‌های چهره [۹].

### ۳-۲- روش‌های جمع‌آوری داده

جمع‌آوری داده در همه راه‌های ممکن برای ساختن مدل‌هایی برای سیستم‌های کارآمد در تشخیص احساس‌های چهره، بسیار مهم است تا بتواند به طور قوی عمل کرده و نتایج را تعمیم دهد. در ادامه چند تکنیک برای جمع‌آوری داده‌ها در این زمینه بررسی خواهند شد.

روش استخراج داده از وب<sup>۳</sup> از ابزارهای خودکار برای جمع‌آوری تصاویر چهره در دسترس در اینترنت استفاده می‌کند. به عنوان مثال، مجموعه داده FER2013 شامل مجموعه‌ای غنی از تصاویر چهره برچسب‌گذاری شده با استفاده از Google Image Search API است [۹].

روش دیگر استفاده از یک محیط کنترل‌شده برای عکاسی حالت و احساس چهره است. به این صورت که افراد برای نشان دادن احساسات مورد نظر حاضر می‌شوند. در این روش از دوربین‌های با وضوح بالا و نورپردازی‌های مناسب استفاده می‌شود که کیفیت تصاویر و برچسب‌گذاری دقیق آنها را تضمین می‌کند. مجموعه داده CK+ از این روش به وجود آمده است [۹].

از طریق جمع‌سپاری، تصاویر حالت‌های مختلف صورت از شرکت‌کنندگانی از سرتاسر جهان با پلتفرم‌های آنلاین مانند Amazon Mechanical Turk جمع‌آوری می‌شود. شرکت‌کنندگان تصاویر یا ویدیوهایی از چهره‌شان آپلود می‌کنند، سپس توسط خودشان یا گروه دیگری از افراد برچسب‌های تصاویر مشخص می‌شود تا مجموعه داده غنی‌تر و متنوع‌تر شود [۱۰].

### ۳-۳- تکنیک‌های پیش‌پردازش تصویر

تکنیک‌های پیش‌پردازش تصویر در تشخیص حالت چهره ضروری هستند زیرا کیفیت تصاویر را افزایش می‌دهند و اطمینان می‌دهند که ویژگی‌های لازم برای تشخیص دقیق احساس‌ها برجسته می‌شوند. در مرحله پیش‌پردازش به طور کلی سه گام برداشته می‌شود: تشخیص چهره در تصویر، ترازبندی چهره، و عادی‌سازی.

<sup>3</sup> Web scraping

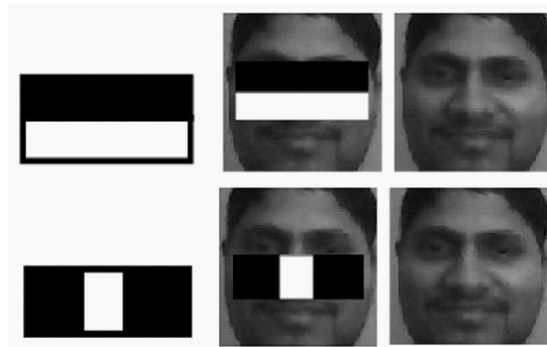
### ۳-۱-۳- تشخیص چهره

تشخیص چهره یکی از اجزای مهم FER است که شامل تشخیص و شناسایی چهره‌ها در یک تصویر می‌شود. در ادامه الگوریتم‌های مختلفی که برای این منظور استفاده می‌شوند بررسی می‌شود.

#### الگوریتم Viola-Jones

به دلیل کارایی بالای الگوریتم Viola-Jones معمولاً از آن برای تشخیص چهره در زمان واقعی<sup>۴</sup> استفاده می‌شود. در این الگوریتم به طور کلی چهار گام طی می‌شود. گام اول بدست آوردن ویژگی‌های Haar-like است. ویژگی‌های Haar-like قالب‌هایی هستند که الگوهایی مانند لبه‌ها، خطوط و یا مستطیل‌ها را در تصویر شناسایی می‌کنند. این ویژگی‌ها مجموع شدت<sup>۵</sup> پیکسل‌ها را در مناطق مستطیلی مجاور مقایسه می‌کنند [۱۱]. مثالی از این مورد در تصویر ۳-۲ دیده می‌شود.

$$f = \sum(\text{white area}) - \sum(\text{black area})$$



شکل ۳-۲: ویژگی‌های Haar-like. در تصویر اول ناحیه چشم‌ها تیره‌تر از گونه‌ها و در تصویر بعد ناحیه چشم‌ها تیره‌تر از بینی است [۱۱].

گام دوم انتگرال تصویر است. انتگرال تصویر یک ساختار داده‌ای است که به طور موثر مجموع مقادیر پیکسل را در یک قالب مستطیلی محاسبه می‌کند. بنابراین، ویژگی‌های Haar-like را می‌توان با استفاده از این نمایش تصویر یکپارچه، بسیار کارآمد محاسبه کرد [۱۱].

گام سوم AdaBoost است. در این الگوریتم، AdaBoost بهترین ویژگی‌ها را انتخاب می‌کند و با ترکیب چندین طبقه‌بندی ضعیف و تشکیل یک طبقه‌بندی قوی با اقدامات زیر، طبقه‌بندی‌کننده‌ها را آموزش می‌دهد.

<sup>۴</sup> Real-time

<sup>۵</sup> Intensity

۱- تخصیص وزن اولیه به تمام نمونه‌های تمرینی.

۲- روی هر ویژگی یک طبقه‌بندی ضعیف آموزش می‌دهد و به خطای آن نگاه می‌کند.

۳- سپس طبقه‌بندی‌کننده با کمترین خطا را انتخاب کرده و وزن نمونه را به‌روز می‌کند.

۴- به تعداد مشخصی گام ۲ و ۳ تکرار خواهد شد.

گام آخر طبقه‌بندی‌کننده‌های آبشاری<sup>۶</sup> است که در ابتدا نواحی غیر چهره را رد می‌کند و تشخیص را سرعت می‌بخشد. این آبشار از چند مرحله تشکیل شده است. هر مرحله از یک طبقه‌بندی قوی تشکیل شده است که با استفاده از Adaboost آموزش دیده است. نهایتاً ناحیه‌ای از تصویر که تمام مراحل را پشت سر بگذارد به عنوان صورت یا چهره طبقه‌بندی می‌شود.

### تجزیه و تحلیل اجزای اصلی<sup>۷</sup>

یکی از روش‌های آماری قوی که در کاهش ابعاد استفاده می‌شود، تجزیه و تحلیل مؤلفه‌های اصلی است که با PCA شناخته می‌شود. در این روش فضای اولیه با ابعاد بالا به فضایی با ابعاد پایین‌تر تبدیل می‌شود و در عین حال تنوع را تا حد امکان حفظ می‌کند.

هدف اصلی PCA کاهش ابعاد مجموعه داده است و با طی پنج گام تنها ویژگی‌هایی را حفظ می‌کند که عوامل مهم واریانس هستند.

۱- استانداردسازی: میانگین داده‌ها صفر و انحراف معیار یک می‌شود.

$$z_i = \frac{x_i - \mu}{\sigma} \quad (۰-۳)$$

$X_i$  مقدار اولیه،  $\mu$  میانگین، و  $\sigma$  انحراف معیار است.

۲- محاسبه ماتریس کوواریانس: ماتریس کوواریانس محاسبه می‌شود، که چگونگی تغییر متغیرهای مجموعه داده از میانگین نسبت به یکدیگر مشخص شود.

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (۱-۳)$$

$n$  تعداد نقاط داده،  $X_i$  نقاط داده و  $\mu$  میانگین است.

<sup>۶</sup> Cascading classifiers

<sup>۷</sup> Principal component analysis (PCA)

۳- مقادیر ویژه و بردار ویژه: این مرحله شامل محاسبه مقادیر ویژه و بردارهای ویژه ماتریس کوواریانس است. به عبارت دیگر، بردارهای ویژه (مولفه‌های اصلی) جهت‌های فضای ویژگی جدید را تعیین می‌کنند، در حالی که مقادیر ویژه بزرگی آنها را تعیین می‌کنند.

$$Cv = \lambda v \quad (2-3)$$

$C$  ماتریس کوواریانس،  $v$  بردار ویژه و  $\lambda$  مقدار ویژه است.

۴- مرتب سازی بردارهای ویژه: بردارهای ویژه طبق مقادیر ویژه به صورت نزولی مرتب می‌شوند. بردارهای ویژه با بالاترین مقادیر ویژه مهم‌ترین اطلاعات را دارند.

۵- تصویر کردن روی فضای ویژگی جدید:  $k$  تا از اول لیست بردارهای ویژه برداشته می‌شود تا ماتریس جدید  $W$  تشکیل شود و داده اولیه روی این فضای  $k$  بعدی تصویر می‌شود.

$$Y = W^T X \quad (3-3)$$

### ۳-۳-۲- ترازبندی و عادی سازی

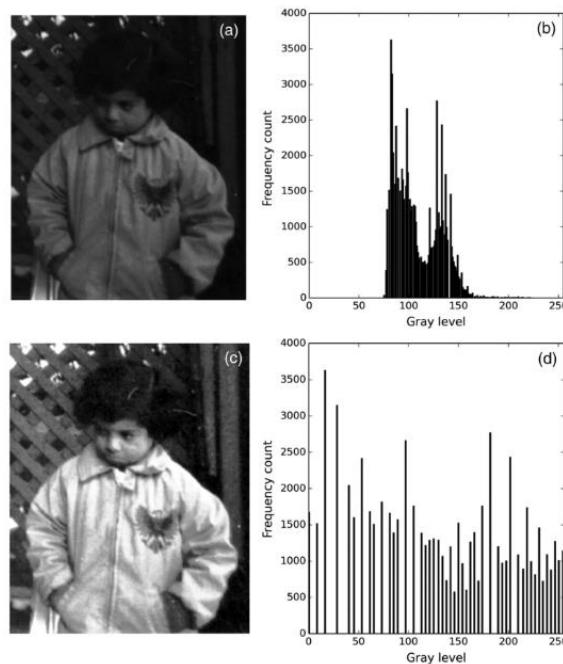
دو فرآیند تراز چهره و عادی سازی، مراحل بسیار مهم پیش‌پردازش داده‌ها در سیستم های تشخیص احساسات چهره هستند.

در تراز کردن، تصاویر صورت شناسایی شده در گام قبلی به گونه‌ای تنظیم می‌شود که چشم‌ها، بینی، و دهان همه در موقعیت‌های استاندارد قرار گیرند، بنابراین می‌تواند تغییراتی که به علت وضعیت سر و یا زاویه دوربین وجود دارد را کاهش دهد. تبدیل‌های هندسی با روش‌های چرخش، مقیاس بندی و ترجمه تصاویر با توجه به نقاط کلیدی شناسایی شده در چهره، موقعیت‌یابی ثابت ویژگی‌ها را تضمین می‌کنند.

رویکرد ما به شناسایی چهره فرض می‌کند که به تصاویر آموزشی به خوبی تراز شده از هر فرد، که تحت نورپردازی متغیر گرفته شده‌اند، دسترسی داریم [۱۲]. در نتیجه در عادی سازی تلاش می‌شود تا شرایط نوری، تضاد<sup>۸</sup> و اندازه تصاویر یکسانی برای تصاویر ایجاد شود تا هرگونه عوامل خارجی که ممکن است باعث کاهش دقت تشخیص شوند را کاهش دهد.

<sup>۸</sup> contrast

یکی از روش‌های عادی‌سازی، یکسان‌سازی هیستوگرام است که با توزیع مجدد مقادیر شدت، بهبود تضاد در تصاویر را قادر می‌سازد. این روش می‌تواند توزیع روشنایی را در یک تصویر بهبود بخشد و بنابراین ویژگی‌های چهره و جزئیات ظریف یک تصویر را مشخص کند. از این روش در تصاویری که در شرایط نوری مختلف گرفته شده‌اند استفاده ویژه‌ای می‌شود [۱۳]. همانطور که در تصویر ۳-۳ دیده می‌شود (a) تصویر ورودی و (b) هیستوگرام سیاه و سفید آن، که نشان می‌دهد توزیع آن در محدوده کوچکی متمرکز است. در حالی که یکسان‌سازی هیستوگرام تضاد کلی تصویر (c) را با توزیع یکنواخت‌تر مقادیر پیکسل (d) بهبود می‌بخشد [۱۳].



شکل ۳-۳: مثالی از یکسان‌سازی هیستوگرام [۱۳].

یکی از بخش‌های مهم سیستم‌های تشخیص حالت چهره مجموعه داده‌ها و روش‌های پیش‌پردازش دقیق است. استفاده از مجموعه داده‌های متنوع با کیفیت بالا، تعمیم دقیق‌تر نتایج برای داده‌های جدید را فراهم می‌کند. تشخیص چهره، تراز چهره، و عادی‌سازی چهره، مراحل پیش‌پردازش برای افزایش کیفیت تصویر و برجسته کردن ویژگی‌های مهم برای تشخیص دقیق‌تر هستند.

## فصل چهارم

### روش‌های استخراج ویژگی

## روش‌های استخراج ویژگی

استخراج ویژگی در تشخیص احساسات صورت پس از پیش‌پردازش صورت می‌گیرد. سیستم‌های FER قادر خواهند بود حالت‌های عاطفی پایه‌ای را از طریق استخراج ویژگی از تصاویر صورت تجزیه و تحلیل و تفسیر کنند. در این فصل، روش‌های مختلف استخراج ویژگی، مانند رویکردهای هندسی و شکل‌محور، رویکردهای مبتنی بر بافت و تکنیک‌های پیشرفته با استفاده از یادگیری عمیق مورد بررسی قرار می‌گیرند. بنابراین، این روش‌ها باید به خوبی درک شوند تا بتوان یک سیستم FER قوی و دقیق را که توانایی مناسب در سناریوهای مختلف در دنیای واقعی داشته باشد، توسعه داد.

### ۴-۱- ویژگی‌های هندسی و شکلی

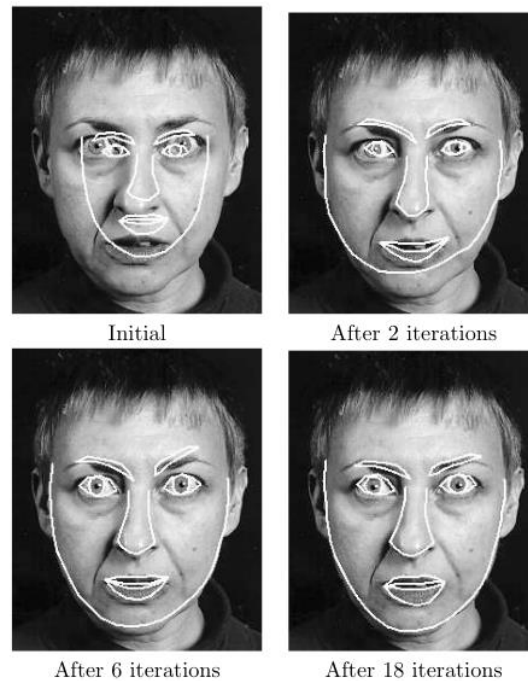
از جمله مهم‌ترین روش‌های استخراج ویژگی‌ها در FER، ویژگی‌های هندسی و شکلی است، زیرا شرح مفصلی از روابط فضایی و ویژگی‌های ساختاری اجزای صورت را تشکیل می‌دهند. این اطلاعات معمولاً از شناسایی نقاط کلیدی چهره و تجزیه و تحلیل تقارن آن به دست می‌آیند.

#### ۴-۱-۱- نقاط عطف چهره

نقاط کلیدی صورت، مکان‌هایی روی چهره هستند که روی تشکیل احساس‌های پایه‌ای در چهره موثر هستند مانند گوشه‌های چشم، نوک بینی و گوشه‌های دهان. برای تشخیص نقاط عطف چهره، الگوریتم‌هایی برای مکان‌یابی نقاط کلیدی روی صورت ارائه شده‌اند. مدل‌های شکلی فعال<sup>۱</sup> یکی از معروف‌ترین رویکردها برای تشخیص شکل در تصاویر، به‌ویژه تشخیص نقاط کلیدی چهره است. این رویکرد شامل یک مرحله آموزشی برای ساخت یک مدل است. مرحله بعدی آن برآزش است که ابتدا مدل در جایی مناسب در تصویر جدید قرار می‌گیرد سپس به صورت تکرارشونده تصویر با مدل ساخته‌شده همگرا می‌شود تا مدل کاملاً با تصویر منطبق شود و نقاط عطف چهره در تصاویر جدید شناسایی شوند [۱۴]. مثالی از این روش در شکل ۴-۱ نمایش داده شده است.

<sup>۱</sup> Active Shape Models





شکل ۴-۱: استفاده از *ASM* برای یافتن نقاط عطف چهره [۱۴].

#### ۴-۱-۲- ویژگی‌های مبتنی بر تقارن

ویژگی‌های مبتنی بر تقارن از تقارن طبیعی در چهره انسان برای ایجاد تشخیص دقیق‌تر احساسات استفاده می‌کنند. چهره انسان به طور معمول متقارن است و هرگونه انحراف در آن به نوعی سرنخی از حالت‌های عاطفی بدست می‌دهد. با تأکید بر ویژگی‌های متقارن خود صورت، سیستم‌های FER می‌توانند کوچک‌ترین تغییرات را در حالت چهره که به احساس‌های خاص اشاره می‌کنند، ثبت کنند.

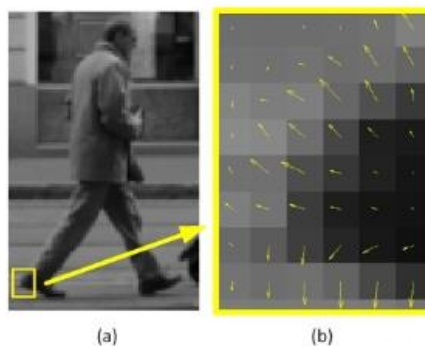
این انحراف‌ها و عدم تقارن با مقایسه نیمه چپ و راست صورت شناسایی می‌شوند. یک رویکرد رایج این است که با استفاده از نقاط عطف یافت‌شده در چهره، خط عمودی وسط چهره در نظر گرفته می‌شود. سپس با این خط عمودی و نقاط متناظر در دو طرف صورت اندازه‌گیری تقارن انجام می‌شود. به طور مثال وقتی کسی لبخند می‌زند، گوشه‌های دهانش به طور متقارن به سمت بالا حرکت می‌کند و این وجود تقارن را می‌توان نشانه‌ای برای تشخیص لبخند دانست و عدم تقارن را می‌توان نشانه‌هایی برای وجود تحقیر یا پوزخند در نظر گرفت.

## ۴-۲- ویژگی‌های بافتی

ویژگی‌های بافتی جزئیات و الگوهای ظریفی را در تصاویر چهره نشان می‌دهند که به احساس‌های مختلف مربوط می‌شود. توصیفگرهای ویژگی به معنای نمایش تصویری است که اطلاعات مفید را استخراج می‌کند و اطلاعات غیر ضروری را از تصویر نادیده می‌گیرد. هیستوگرام گرادیان‌های جهت‌دار و الگوهای باینری محلی دو توصیفگر ویژگی پرکاربرد در استخراج ویژگی‌های بافتی هستند.

### ۴-۲-۱- هیستوگرام گرادیان‌های جهت‌دار (HOG)

این تکنیک برای تبدیل یک تصویر به یک بردار ویژگی با استفاده از اطلاعات گرادیان‌هاست (تصویر ۴-۲) که سپس برای الگوریتم‌های طبقه‌بندی تصویر مانند SVM اعمال می‌شود. چون در شناسایی اشیاء در تصاویر، لبه‌ها و گوشه‌ها مهم‌ترین اطلاعات را می‌دهند اطلاعات هیستوگرام گرادیان‌ها مورد استفاده قرار می‌گیرند [۱۵].



شکل ۴-۲: هیستوگرام گرادیان‌های جهت‌دار. (a) تصویر اولیه (b) تصویر گرادیان‌ها [۱۵]

برای استفاده از HOG، در مرحله پیش‌پردازش معمولاً از نسبت ثابت ۱:۲ مثلاً ۱۰۰×۲۰۰ یا ۵۰۰×۱۰۰۰ پیکسل استفاده می‌شود. سپس گرادیان افقی ( $G_x$ ) و عمودی ( $G_y$ ) در هر پیکسل محاسبه می‌شود و بعد اندازه و جهت گرادیان‌ها بدست آورده می‌شوند [۱۵].

$$M(x, y) = G_x(x, y)^2 + G_y(x, y)^2 \quad (۰-۵)$$

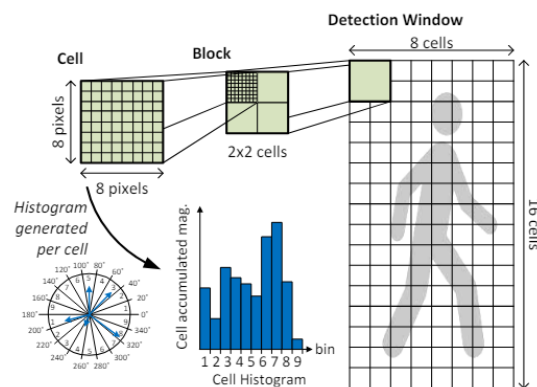
$$\theta(x, y) = \tan^{-1} \left( \frac{G_y(x, y)}{G_x(x, y)} \right) \quad (۱-۵)$$

حالا تصویر به سلول‌ها تقسیم می‌شوند. مثلاً سلول‌های ۸×۸ برای یک تصویر ۶۴×۱۲۸. برای هر سلول، هیستوگرام جهت گرادیان‌ها محاسبه می‌شود. به طور رایج ۹ سطل برای جهت‌های صفر تا ۱۶۰ درجه با فواصل

۲۰ در نظر گرفته می‌شود. هر پیکسل به نسبت اندازه گرادیان خود و جهت آن در هیستوگرام مشارکت می‌کند. بنابراین، اگر گرادیان یک پیکسل بین دو سطل باشد، اندازه آن به تناسب بین این دو بین تقسیم می‌شود [۱۵].

در مرحله آخر می‌توان برای بهبود خروجی از عادی‌سازی بلوک‌ها استفاده کنیم. برای مقاوم‌سازی توصیفگر در برابر تغییرات نوری، بلوک‌هایی به ابعاد مثلاً  $8 \times 8$  یا  $16 \times 16$  پیکسل عادی‌سازی می‌شوند. این کار با محاسبه طول بردار مقادیر هیستوگرام و تقسیم تمامی عناصر بر این طول انجام می‌شود. به این ترتیب، عادی‌سازی به گونه‌ای صورت می‌گیرد که توصیفگر، بدون جهت‌گیری و مقاوم در برابر تغییرات نوری باشد [۱۵].

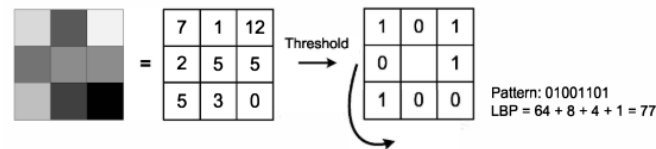
تصویر ۳-۴ روند کلی‌ای از توصیفگر ویژگی هیستوگرام گرادیان‌های جهت‌دار نشان می‌دهد. در اینجا سلول‌ها  $8 \times 8$  پیکسل هستند. هیستوگرام‌های گرادیان در ۹ سطل برای کل محدوده صفر تا  $360^\circ$  درجه برای هر سلول بدست می‌آیند. در اینجا هر بلوک شامل چهار سلول یعنی  $16 \times 16$  پیکسل است. هیستوگرام‌های سلولی در یک بلوک به هم می‌پیوندند و یک بردار ویژگی برای آن بلوک ایجاد می‌کنند [۱۵].



شکل ۳-۴: تقسیم بخش موردنظر تصویر به بلوک‌ها و سلول‌ها و تشکیل آرایه‌ای از هیستوگرام در هر پنجره [۱۵]

#### ۴-۲-۲- الگوهای باینری محلی (LBP)

LBP با مقایسه شدت یک پیکسل مرکزی در یک محله کوچک با پیکسل‌های همسایه آن کار می‌کند. بسته به اینکه شدت آن بیشتر یا کمتر از مقدار آستانه باشد، هر پیکسل همسایه را با یک مقدار باینری رمزگذاری می‌کند. همه این مقادیر باینری در یک عدد باینری کنار هم قرار می‌گیرند و در نهایت بافت آن ناحیه را نشان می‌دهند [۱۶]. شکل ۴-۴ نمونه‌ای از الگوریتم LBP را نشان می‌دهد که برای درک بهتر این الگوریتم توضیح داده خواهد شد.



شکل ۴-۴: فرایند  $LBP$  [۱۶]

برای مقدار آستانه پیکسل وسط در ناحیه یا محله در نظر گرفته می‌شود که در اینجا پنج است. سپس، تمام همسایه‌های آن پیکسل بررسی می‌شوند که مقدار شدت آنها از حد آستانه بیشتر است یا کمتر. اگر بیشتر است برای آن یک قرار داده می‌شود در غیر اینصورت صفر. سپس این مقادیر صفر و یک با شروع از پیکسل سمت چپ و وسط و در خلاف جهت عقربه‌های ساعت کنار یکدیگر قرار می‌گیرند تا برای پیکسل وسطی الگوی باینری را بسازند و نهایتاً به مبنای ده برده می‌شود. این مراحل برای هر پیکسل در تصویر تکرار می‌شود تا هر پیکسل یک مقدار پیدا کند [۱۶].

اکنون از این مقادیر  $LBP$  در ساخت هیستوگرام استفاده می‌شود. با ساختن یک هیستوگرام از الگوهای  $LBP$  می‌توان فراوانی الگوهای بافتی مختلف در تصویر را ثبت کرد. سپس از این هیستوگرام به عنوان یک بردار ویژگی استفاده می‌شود.

استخراج ویژگی در سیستم‌های تشخیص حالت چهره، یک مرحله بسیار ضروری است. روش‌های مبتنی بر هندسه و شکل و روش‌های مبتنی بر بافت تکنیک‌هایی هستند که در  $FER$  برای استخراج ویژگی‌ها از حالت‌های چهره موجود هستند. بنابراین، این روش‌ها باید برای اطمینان از ساخت سیستم‌های قوی و دقیق در  $FER$  مورد تحقیق قرار گیرند.

## فصل پنجم

### الگوریتم‌های طبقه‌بندی و شناسایی احساس‌ها

## الگوریتم‌های طبقه‌بندی و شناسایی احساس‌ها

پس از طی گام‌های قبلی نوبت به طبقه‌بندی چهره‌ها و احساس‌ها می‌رسد. در ادامه چند الگوریتم برای این هدف بررسی شده‌است.

### ۵-۱- روش‌های کلاسیک طبقه‌بندی

رویکردهای کلاسیک برای طبقه‌بندی روش‌هایی هستند که کم و بیش از ابزارهای آمار و یادگیری ماشین استفاده می‌کنند مانند ماشین بردار پشتیبان و درخت تصمیم‌گیری.

#### ۵-۱-۱- ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان یکی از شناخته‌شده‌ترین و قوی‌ترین الگوریتم‌ها در یادگیری ماشین با نظارت است. همچنین برای پردازش داده‌ها با ابعاد بالا و روابط پیچیده بسیار کارآمد است و در این پژوهش برای طبقه‌بندی احساس‌ها از روی تصاویر استفاده می‌شود.

هدف اصلی SVM جستجوی یک ابرصفحه<sup>۱</sup> برای جدا کردن نقاط داده در فضای ویژگی بدست آمده است. این امر با در نظر گرفتن به حداکثر رساندن حاشیه<sup>۲</sup> انتخاب می‌شود. حاشیه فاصله بین ابرصفحه از نزدیک‌ترین نقطه داده از هر کلاس است. به حداکثر رساندن حاشیه منجر به عملکرد بهتر طبقه‌بندی می‌شود.

#### داده‌های قابل جداسازی خطی:

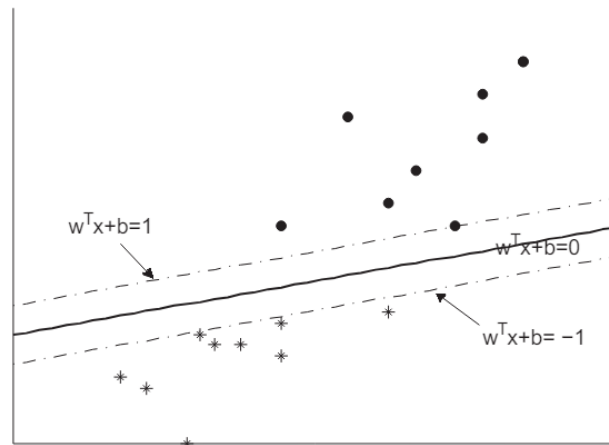
با داشتن دو ویژگی، ابرصفحه یک خط است. همانطور که دیده می‌شود در شکل ۵-۱ چندین خط جداکننده وجود دارد، اما ابرصفحه بهینه آنی است که بیشترین حاشیه را ایجاد می‌کند. مثلاً در شکل ۵-۱ خط وسط پررنگ‌تر انتخاب می‌شود. اگر چنین ابرصفحه‌ای وجود داشته باشد که فاصله‌اش از هر طرف تا نزدیک‌ترین داده، بیشترین باشد به آن حاشیه سخت<sup>۳</sup> گفته می‌شود [۱۷].

---

<sup>۱</sup> Hyperplane

<sup>۲</sup> Margin

<sup>۳</sup> Hard margin



شکل ۵-۱: خط بهینه در طبقه‌بندی SVM [۱۷]

در شرایطی که داده کاملاً قابل تفکیک نیست، SVM جریمه‌ای را برای طبقه‌بندی‌های نادرست در نظر می‌گیرد و هدف علاوه بر به حداکثر رساندن حاشیه، به حداقل رساندن جریمه نیز هست که به این مورد حاشیه نرم<sup>۴</sup> گفته می‌شود [۱۷].

#### داده‌های قابل جداسازی غیرخطی:

برای این داده‌ها، SVM از یک تابع هسته استفاده می‌کند که داده‌ها را به یک فضای ابعادی بالاتر نگاشت می‌کند، جایی که بتوان داده‌ها را به صورت خطی جدا کرد [۱۷].

### ۵-۱-۲- درخت تصمیم‌گیری

درخت‌های تصمیم‌گیری به عنوان یکی از الگوریتم‌های یادگیری ماشین هم برای طبقه‌بندی و هم برای وظایف رگرسیون در نظر گرفته می‌شوند. آنها با تقسیم بازگشتی یک مجموعه داده به زیر مجموعه‌ها مطابق با مقادیر ویژگی‌های ورودی کار می‌کنند تا در نهایت یک مدل درخت‌مانند از این تصمیم‌ها ساخته شود. یکی از نقاط قوت آنها این است که هم برای داده‌های عددی و هم غیرعددی خوب عمل می‌کنند. علاوه بر این، آنها بسیار قابل تفسیر هستند [۱۸].

<sup>۴</sup> Soft margin

در این درخت، گره ریشه کل مجموعه داده را توصیف می‌کند و همچنین به تقسیم اول اشاره دارد. گره‌های داخلی به یک تست ویژگی یا یک شرط تبدیل می‌شوند. برگ‌ها به برچسب‌های کلاس یا به عبارت دیگر پیش‌بینی‌های نهایی تبدیل می‌شوند. ساخت یک درخت با انتخاب بهترین ویژگی انجام می‌شود که داده‌ها را به خالص‌ترین زیر مجموعه‌های ممکن تقسیم می‌کند. یکی از معیارهای اندازه‌گیری خلوص، ناخالصی GINI است [۱۸].

$$GINI(T) = 1 - \sum_{p=0}^{p-1} P(i|t)^2 \quad (۵-۰)$$

T مجموعه داده، p تعداد کلاس‌ها و  $p(i|t)$  فراوانی نسبی کلاس p در مجموعه داده T است [۱۸].

پس از تقسیم مجموعه داده D به دو زیر مجموعه D1 و D2 با اندازه‌های N1 و N2، شاخص جینی به صورت زیر تعریف می‌شود [۱۸]:

$$GINI_{SPLIT(D)} = \left(\frac{N1}{N}\right) * GINI(D1) + \left(\frac{N2}{N}\right) * GINI(D2) \quad (۵-۱)$$

$$GINI_{GAIN(A)} = GINI(D) - GINI_{SPLIT(D)} \quad (۵-۲)$$

با استفاده از رابطه (۵-۱) تصمیم گرفته می‌شود که کدام صفت از بین مجموعه ویژگی‌ها برای تقسیم بهتر است و با استفاده از رابطه (۵-۲) کاهش ناخالصی اعمال می‌شود [۱۸].

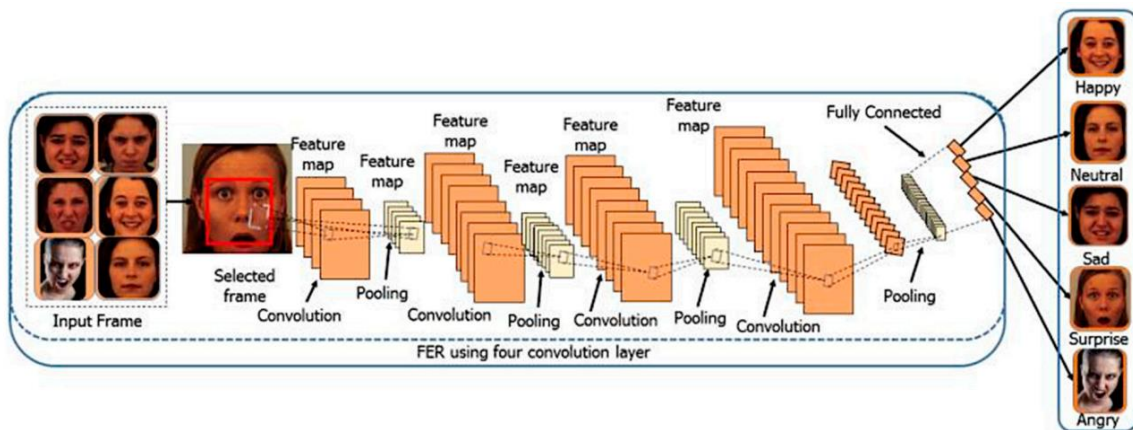
## ۵-۲- یادگیری عمیق برای طبقه‌بندی احساس‌ها

یادگیری عمیق بیشتر حوزه‌های بینایی کامپیوتر را تشکیل می‌دهد. این رویکرد توانایی شبکه‌های عصبی مصنوعی را برای یادگیری خودکار و استخراج ویژگی‌ها از خود داده‌های خام نشان می‌دهد و از این رو به طور قابل توجهی دقت و استحکام بیشتری نسبت به روش‌های سنتی ارائه کرده است. یادگیری عمیق به دلیل توانایی ذاتی آن در مدیریت داده‌های حجیم و مدل‌سازی روابط پیچیده، انتخاب ایده‌آلی برای طبقه‌بندی احساس‌ها است.



## ۵-۲-۱- شبکه‌های عصبی کانولوشنی (CNN)

شبکه‌های عصبی کانولوشنی از مدل‌های یادگیری عمیق هستند که در پردازش تصویر کاملاً موفق بوده‌اند. طبق تصویر ۵-۲ استفاده از لایه‌های متعدد، مانند لایه‌های کانولوشن، لایه‌های ادغام<sup>۵</sup> و لایه‌های کاملاً متصل، یادگیری خودکار سلسله‌مراتبی را از طریق پس‌انتشار<sup>۶</sup> امکان‌پذیر می‌سازد.



شکل ۵-۲: معماری شبکه عصبی کانولوشنی در FER [۹]

لایه‌های کانولوشنی، لایه‌های ورودی را در هم پیچش می‌دهند و به لایه بعدی منتقل می‌کنند. اهمیت چنین عملیاتی در تشخیص الگوهای فضایی مانند لبه‌ها و بافت‌ها یا حتی ویژگی‌های پیچیده‌تر در تصاویر است. در این لایه، هسته‌ها تصویر ورودی را بررسی و پیمایش می‌کنند تا نقشه‌های ویژگی را تولید کنند [۹].

لایه‌های ادغام، لایه‌هایی هستند که فقط اطلاعات مهم را نگه می‌دارند و ابعاد هر نقشه ویژگی را کاهش می‌دهند. انواع متداول ادغام شامل در نظر گرفتن حداکثر<sup>۷</sup> و میانگین<sup>۸</sup> است. این مرحله بار محاسباتی را کاهش می‌دهد و بیش‌برازش را کنترل می‌کند.

از لایه‌های کاملاً متصل معمولاً در انتهای شبکه استفاده می‌شود و مسئول طبقه‌بندی تصویر به کلاس تشخیص داده‌شده هستند. در لایه‌های کاملاً متصل، هر نورون به تمام نورون‌های لایه قبلی متصل است.

<sup>5</sup> Pooling

<sup>6</sup> Backpropagation

<sup>7</sup> Max-pooling

<sup>8</sup> Average-pooling

## ۵-۲-۲- شبکه‌های عصبی بازگشتی (RNN)

RNNها اطلاعات متوالی را به صورت یک بخش در هر زمان پردازش می‌کنند. در این روش یک بردار حالت وجود دارد که اطلاعات بخش‌های قبلی در دنباله را نگه می‌دارد.

### حافظه طولانی کوتاه مدت

LSTM یک نوع RNN خاص است که می‌تواند وابستگی‌های طولانی کوتاه‌مدت را یاد بگیرد. شبکه‌های LSTM برای حل مسئله محو شدن گرادیان<sup>۹</sup> معرفی شدند که از نظر معماری از RNNهای سنتی بسیار پیچیده‌تر هستند. این معماری از چند دروازه<sup>۱۰</sup> تشکیل شده است [۱۹].

دروازه فراموشی<sup>۱۱</sup> تصمیم می‌گیرد که وضعیت سلول چه مقدار اطلاعات را فراموش کند. دروازه ورودی کنترل می‌کند که چه اطلاعاتی از ورودی برای به‌روز رسانی وضعیت سلول استفاده می‌شود. دروازه خروجی کنترل می‌کند که چه مقدار از حالت سلول بر اساس ورودی فعلی به عنوان یک خروجی نشان داده می‌شود [۱۹].

دروازه‌های ورودی، فراموشی، و خروجی را می‌توان آموزش داد تا یاد بگیرد به ترتیب چه اطلاعاتی در حافظه ذخیره شود، چه مدت ذخیره شود و چه زمانی آن را بخواند [۱۹]. این دروازه‌ها به شبکه‌های LSTM اجازه می‌دهند تا وضعیت سلولی را حفظ کرده و به‌روزرسانی کنند.

---

<sup>۹</sup> Vanishing gradient

<sup>۱۰</sup> Gate

<sup>۱۱</sup> Forget gate

## فصل ششم

### جمع‌بندی و نتیجه‌گیری

## جمع‌بندی و نتیجه‌گیری

امروزه FER بسیار تکامل یافته است و تکنیک‌ها و روش‌های مختلفی برای کمک به تشخیص قوی و دقیق احساسات به آن وارد شده‌اند. کاربردهای سیستم‌های FER در حوزه‌های مختلف، از تعامل انسان و کامپیوتر و امنیت و نظارت گرفته تا بازاریابی و مراقبت‌های بهداشتی اهمیت دقت در طبقه‌بندی احساس‌ها را نشان می‌دهد. در این پژوهش روش‌ها و گام‌های رایج ایجاد یک سیستم تشخیص احساس‌های پایه‌ای مورد بررسی قرار گرفتند. در گام اول برای ایجاد یک FER، مجموعه داده‌ها و تکنیک‌های پیش‌پردازش مورد بررسی قرار گرفتند. تشخیص چهره، تراز چهره، و عادی‌سازی به عنوان تکنیک‌های پیش‌پردازش مورد مطرح شدند.

در گام بعدی برخی از تکنیک‌های استخراج ویژگی ارائه شدند. در ابتدا به ویژگی‌های هندسی و شکلی پرداخته شد. شناسایی نقاط عطف چهره و تجزیه و تحلیل تقارن، دو موضوع مورد بحث در این بخش بودند. سپس در بخش ویژگی‌های بافتی به الگوریتم‌هایی مانند هیستوگرام گرادیان‌های جهت‌دار و الگوهای باینری محلی پرداخته شد.

در انتها الگوریتم‌هایی برای طبقه‌بندی و شناسایی احساسات، شامل روش‌های کلاسیک و رویکردهای یادگیری عمیق مورد بحث قرار گرفتند. با طبقه‌بندی‌کننده‌های کلاسیک مانند ماشین‌های بردار پشتیبان و درخت‌های تصمیم‌گیری و در یادگیری عمیق با شبکه‌های عصبی کانولوشنی و شبکه‌های عصبی بازگشتی، به طور خاص LSTM، آشنایی ایجاد گشت.

FER همچنان با چالش‌های متعددی روبرو است که فرصت‌هایی را برای تحقیق و توسعه آینده ایجاد می‌کند. یکی از چالش‌های مهم، تنوع در حالت‌های چهره در فرهنگ‌ها، سنین و جنسیت‌های مختلف در مجموعه داده‌ها است. علاوه بر این، موقعیت‌های دنیای واقعی اغلب شامل محیط‌های متنوع و پویا می‌شوند که در آن نور، انسداد و ... می‌توانند تشخیص احساسات را پیچیده‌تر کنند. نگرانی‌های حفظ حریم خصوصی نیز با گسترش فناوری‌های FER به وجود می‌آیند که توسعه دستورالعمل‌های اخلاقی و تکنیک‌های حفظ حریم خصوصی را ضروری می‌کند. علی‌رغم این چالش‌ها، فرصت‌های آینده فراوان است، به‌ویژه با ادغام مدل‌های پیشرفته یادگیری عمیق و در دسترس بودن مجموعه داده‌های بزرگتر و متنوع‌تر.

## منابع و مراجع

1. B. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," *Sensors*, vol. 18, no. 2, p. 401, Jan. 2018, doi: <https://doi.org/10.3390/s18020401>.
2. A. Saxena, A. Khanna, and D. Gupta, "Emotion Recognition and Detection Methods: A Comprehensive Survey," *Journal of Artificial Intelligence and Systems*, vol. 2, no. 1, pp. 53–79, 2020, doi: <https://doi.org/10.33969/ais.2020.21005>.
3. C.-H. Chen, I.-Jui. Lee, and L.-Y. Lin, "Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders," *Research in Developmental Disabilities*, vol. 36, pp. 396–403, Jan. 2015, doi: <https://doi.org/10.1016/j.ridd.2014.10.015>.
4. S. Hickson, N. Dufour, A. Sud, V. Kwatra, and I. Essa, "Eyemotion: Classifying facial expressions in VR using eye-tracking cameras," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 1626–1635.
5. M. A. Assari and M. Rahmati, "Driver drowsiness detection using face expression recognition," *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Kuala Lumpur, Malaysia, 2011, pp. 337–341.
6. S. Zhai, C. Morimoto, and S. Ihde, "Manual and gaze input cascaded (MAGIC) pointing," *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*, 1999.
7. A. T. Wieckowski, L. T. Flynn, J. A. Richey, D. Gracanin, and S. W. White, "Measuring change in facial emotion recognition in individuals with autism spectrum disorder: A systematic review," *Autism*, vol. 24, no. 7, pp. 1607–1628, Jun. 2020.
8. I. Maglogiannis and C. N. Doukas, "Overview of Advanced Computer Vision Systems for Skin Lesions Characterization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 721–733, Sep. 2009.
9. A. R. Khan, "Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges," *Information*, vol. 13, no. 6, p. 268, May 2022.
10. R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, "EMOTIC: Emotions in Context Dataset," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jul. 2017.
11. K. Aashish and A. Vijayalakshmi, "Comparison of Viola-Jones And Kanade-Lucas-Tomasi Face Detection Algorithms," *Oriental journal of computer science and technology*, vol. 10, no. 1, pp. 151–159, Mar. 2017.
12. J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse Representation for Computer Vision and Pattern Recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
13. [9]A. Toet and T. Wu, "Efficient contrast enhancement through log-power histogram modification," *Journal of Electronic Imaging*, vol. 23, no. 6, p. 063017, Dec. 2014.

14. T. F. Cootes, "An Introduction to Active Shape Models," *Image Processing and Analysis*, Jan. 2000.
15. C. Kelly, F. Siddiqui, Burak Bardak, and R. Woods, "Histogram of oriented gradients front end processing: An FPGA based processor approach," *Zenodo (CERN European Organization for Nuclear Research)*, Oct. 2014.
16. J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," *IEEE Xplore*, Oct. 01, 2011.
17. W. Huibing, J. Xiong, Z. Yao, M. Lin, and J. Ren. "Research survey on support vector machine," *10th EAI International Conference on Mobile Multimedia Communications*, pp. 95-103, 2017.
18. F. Z. Salmam, A. Madani, and M. Kissi, "Facial Expression Recognition Using Decision Trees," *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, Mar. 2016.
19. F. Gers, N. N. Schraudolph, J. Schmidhuber, "Learning Precise Timing with LSTM Recurrent Networks," *Journal of Machine Learning Research*, vol. 3, pp. 115-143, Jan. 2002.

