

تاریخ دریافت مقاله: ۹۸/۰۱/۱۸

تاریخ پذیرش مقاله: ۹۸/۰۶/۲۳

روش‌های شناسایی احساسات چهره، کاربردها و چالش‌ها

مائده شریف نژاد

کارشناسی ارشد مهندسی کامپیوتر- دانشکده فنی - دانشگاه گیلان - رشت - ایران
پست الکترونیکی: m_sharifnejad@msc.guilan.ac.ir

اسدالله شاه بهرامی*

دانشیار گروه مهندسی کامپیوتر - دانشکده فنی - دانشگاه گیلان - رشت - ایران
پست الکترونیکی: shahbahrami@guilan.ac.ir

علیرضا آکوشیده

استادیار دپارتمان برق و کامپیوتر، آموزشکده شهید چمران - دانشگاه فنی و حرفه‌ای استان گیلان - رشت - ایران
پست الکترونیکی: akushide@tvu.ac.ir

چکیده:

جهت‌دار و الگوی دودویی محلی و روش‌های طبقه‌بندی حالات چهره مانند ماشین بردار پشتیبان و شبکه عصبی پیچشی معرفی می‌شوند. مجموع مطالعات و بررسی‌ها نشان داد که روش‌های مبتنی بر شبکه عصبی پیچشی نسبت به بقیه روش‌ها دارای بالاترین دقت در تشخیص حالات چهره است. ولی تنوع داده‌های تصویر چهره از نظر سن، جنسیت، نژاد، شرایط انسداد، تغییرات روشنائی، دوری و نزدیکی و همچنین تشخیص حالات دیگر از قبیل هیجان، تأسّف و خستگی از چالش‌های پیش رو است که پژوهش در این زمینه ادامه دارد.

واژه‌های کلیدی: شناسایی حالات چهره، شناسایی احساسات چهره، استخراج ویژگی، طبقه‌بندی

۱. مقدمه

در سال‌های اخیر فناوری هوش مصنوعی در بسیاری

در سال‌های اخیر شناسایی احساسات مختلف چهره به دلیل کاربردهای فراوان مانند تشخیص خستگی رانندگان، شناسایی وضعیت روحی بیمار، کشف دروغ و سیستم آموزش خودکار مورد توجه بسیاری در تحقیقات بینایی ماشین قرار گرفته است. به دلیل تنوع افراد و تغییرپذیری حالات چهره، شناسایی این احساسات با دقت بالا هنوز هم به‌عنوان یک مسئله چالش‌برانگیز مطرح است. در این مقاله حالات مختلف چهره مانند خشم، ترس، انزجار، شادی، غم و تعجب معرفی می‌شوند و نه پایگاه داده تصویری که در پژوهش‌های تشخیص حالات چهره مورد استفاده قرار گرفته‌اند مورد مقایسه قرار می‌گیرند. یک سیستم تشخیص حالات چهره دارای دو بخش مهم استخراج ویژگی و طبقه‌بندی حالات است. برخی از روش‌های استخراج ویژگی مانند مدل ظاهری، هیستوگرام گرادینان

* نویسنده مسئول

از کاربردها از قبیل اتومبیل بدون سرنشین، تعامل انسان- کامپیوتر، آموزش و حمل و نقل هوشمند پیشرفت زیادی داشته است. در این بین شناخت احساسات انسان نقش بسیار مهمی در کاربردهای مختلف هوش مصنوعی دارد. در نتیجه مطالعه شناسایی خودکار احساسات انسان و بهبود روش‌های آن به وسیله ماشین یکی از زمینه‌های تحقیقاتی مهم در هوش مصنوعی و یادگیری ماشین است. به‌طور اساسی انسان‌ها احساسات خود را با روش‌های مختلف مانند زبان گفتار، حالات چهره، زبان بدن و کلمات بیان می‌کنند [۱، ۴۱، ۴۲]. مطالعات نشان می‌دهد که برای درک احساسات افراد، ۳۸ درصد اطلاعات به وسیله ریتم گفتار، ۷ درصد از طریق کلمات و ۵۵ درصد به وسیله حالات چهره به دست می‌آید [۴]. به همین دلیل این موضوع محققان را تشویق می‌کند تا در زمینه تجزیه و تحلیل شناسایی حالات مختلف چهره مطالعه کنند. در سال ۱۹۷۱ شش احساس پایه شامل انزجار^۱، ترس^۲، شادی^۳، تعجب^۴، ناراحتی^۵ و خشم^۶ مطرح شد که در جوامع جهانی و قومیت‌های مختلف ثابت است [۴۳]. بر اساس تحقیقات موجود، به‌طور کلی مراحل شناسایی حالات چهره^۷ را می‌توان به دو بخش استخراج ویژگی^۸ و طبقه‌بندی^۹ حالات تقسیم‌بندی کرد. در مرحله اول ویژگی‌های هندسی یا ظاهری صورت از تصویر ورودی چهره استخراج می‌شود. سپس در مرحله دوم با توجه به ویژگی‌های استخراج‌شده، از یک طبقه‌بند استفاده می‌شود تا به تصویر ورودی یک برچسب از شش حالت پایه اختصاص داده شود. در دهه‌های گذشته الگوریتم‌های بسیاری برای استخراج ویژگی و طبقه‌بندی مطرح‌شده است که هر یک کارایی متفاوتی دارند [۲، ۴۷-۴۵]. به‌تازگی نیز برای شناسایی حالات چهره از روش‌های مبتنی بر شبکه عصبی^{۱۰} استفاده‌شده است که

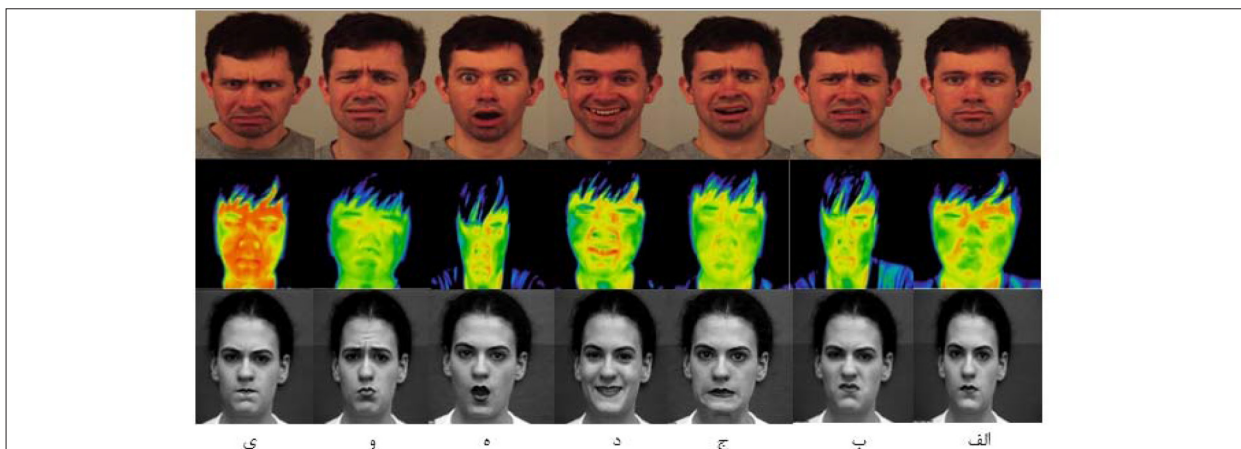
- 1- Disgust
- 2- Fear
- 3- Joy
- 4- Surprise
- 5- Sadness
- 6- Anger
- 7- Facial Expressions Recognition (FER)
- 8- Feature Extraction
- 9- Classification
- 10- Neural Network (NN)

برخی از روش‌های رایج آن شبکه باور عمیق^{۱۱} و شبکه عصبی پیچشی^{۱۲} است. این روش‌ها منجر به پیشرفت بسیاری در زمینه مطالعات پردازش تصویر^{۱۳} شده است [۴۸، ۲۱].

هدف این مقاله بررسی برخی از روش‌های مهم استخراج ویژگی، طبقه‌بندی حالات و همچنین جدیدترین پژوهش‌های انجام‌شده در زمینه شناسایی حالات چهره است. علاوه بر این پایگاه داده‌های استفاده‌شده در این زمینه، از جنبه‌های مختلفی مانند تعداد نمونه، تعداد افراد، جنسیت، سن، نژاد و وضوح تصاویر مورد بحث و بررسی قرار می‌گیرند و در نهایت برخی از چالش‌های موجود مطرح می‌گردد. بدین منظور جهت جمع‌آوری این مطالعات از روش بازبینی نظام‌یافته ادبیات^{۱۴} و رشته جستجوهای مختلفی مانند شناسایی حالات و احساسات چهره، الگوریتم‌های استخراج ویژگی، الگوریتم‌های طبقه‌بندی و پایگاه داده‌های حالات چهره در برخی از پایگاه‌های علمی از قبیل آی‌تریپل‌ای^{۱۵}، اشپرینگر^{۱۶}، الزویور^{۱۷}، ساینس‌دایرکت^{۱۸}، گوگل اسکولار^{۱۹} و مجلات علمی و پژوهشی داخلی استفاده‌شده است. در نهایت ۴۸ منابع مختلف کتاب، مقالات پژوهشی چاپ‌شده در مجلات معتبر ISI و مجلات معتبر داخلی انتخاب‌شده است.

نتایج مطالعات نشان می‌دهد که احساسات چهره به وسیله انقباض یک یا چند عضله صورت می‌گیرد که به علت بی‌ثباتی احساسی، اجزای صورت مانند پلک‌ها، ابرو، بینی، دهان، چانه و دهان تغییر شکل می‌یابد و از این طریق شش حالت پایه چهره به‌علاوه حالت طبیعی^{۲۰} و غرور^{۲۱} با دقت مناسبی قابل‌شناسایی است. در حال حاضر تلاش برای تشخیص احساسات بیشتر مانند هیجان، حقارت،

- 11- Deep Belief Network (DBN)
- 12- Convolutional Neural Network (CNN)
- 13- Image Processing
- 14- systematic literature review (SLR)
- 15- Institute of Electrical and Electronics Engineers (IEEE)
- 16- Springer
- 17- Elsevier
- 18- ScienceDirect
- 19- Google Scholar
- 20- Natural
- 21- Pride



شکل ۱: مقایسه انواع احساسات چهره در فضاها رنگی مختلف. به ترتیب از ردیف اول تا سوم، طبقه‌بندی احساسات چهره در فضای رنگ RGB (تصاویر پایگاه داده KDEF [۱۸])، طبقه‌بندی احساسات چهره در فضای رنگ خاکستری (تصاویر پایگاه داده CK+ [۳۰]). انواع احساسات: (الف) طبیعی، (ب) انزجار، (ج) ترس، (د) شادی، (ه) تعجب، (و) ناراحتی، (ی) خشم

تعاملات اجتماعی و سیگنال‌های فیزیولوژیکی را نیز منعکس می‌کند.

۲-۱. حالات مختلف چهره

در تحقیقات نشان داده شده است که شش احساس اصلی، یعنی انزجار، ترس، شادی، تعجب، ناراحتی و خشم در قومیت‌ها و فرهنگ‌های انسانی جهانی هستند [۵]. در شکل ۱ این شش احساس پایه به همراه احساس طبیعی در فضاها رنگی مختلف نشان داده شده است. اگرچه انسان‌ها می‌توانند احساسات چهره را به راحتی و بدون تأخیر تشخیص دهند اما این مسئله برای ماشین‌همچنان یک چالش به‌شمار می‌رود [۱]. مطالعات زیادی بر روی احساسات پایه صورت گرفته است؛ اما با این وجود احساسات انسان را می‌توان توسط انواع دیگری نیز تعریف کرد. نمونه‌ای از این احساسات در شکل ۲ و ۳ نشان داده شده است. در شکل ۲ تعدادی حالات چهره از پایگاه داده QLFW^{۲۲} نشان داده شده است و همچنین شکل ۳ نشان‌دهنده احساسات ترکیبی است. احساسات ترکیبی، ترکیبی از دو احساس پایه است. به‌طور مثال می‌توان با احساس خوشحالی تعجب کرد و یا با احساس عصبانیت تعجب کرد. در حال حاضر ۲۱ دسته از احساسات متمایز تعریف شده است [۶]. همچنین احساسات دیگری مانند

تأسف، خستگی و غیره هم چنان ادامه دارد. انتخاب یک روش استخراج ویژگی که در شرایط مختلف مانند رنگ‌های متفاوت پوست، سن، جنسیت و شرایط روشنایی کارآمد باشد، بسیار دشوار و از چالش‌های پیش رو است. همچنین بررسی‌ها نشان داد که روش‌های مبتنی بر شبکه‌های عصبی مانند یادگیری عمیق دارای دقت بالاتری در شناسایی احساسات چهره^{۲۳} در مقایسه با دیگر روش‌ها هستند. در ادامه این مقاله در بخش دوم، حالات مختلف چهره، کاربردها، نرم‌افزارها و پایگاه داده‌های مختلف مطرح می‌گردد. در بخش سوم مراحل شناسایی این حوزه و مهم‌ترین الگوریتم‌های موجود مورد بررسی قرار می‌گیرند. در بخش چهارم برخی کارهای انجام شده در این زمینه عنوان می‌گردد؛ و در انتها، در بخش پنجم به جمع‌بندی و نتیجه‌گیری پرداخته می‌شود.

۲. حالات مختلف چهره و کاربردهای آن

چهره نقش اساسی در شناسایی افراد و نمایش احساسات آن‌ها در سطح جامعه دارد. یکی از روش‌های ارتباط غیرکلامی که توسط آن می‌توان حالات خلقی و ذهنی فرد را درک کرد، حالات چهره است [۳]. حالات چهره نه تنها احساسات را منعکس می‌کند، بلکه سایر فعالیت‌های ذهنی،

23- Quality Labeled Faces in the Wild (QLFW)

22- Facial Emotion Recognition (FER)



شکل ۲: نمایش احساسات مختلف در افراد (تصاویر پایگاه داده QLFW [۲۹])



شکل ۳: نمایش احساسات ترکیبی: (الف) با خوشحالی متعجب شدن، (ب) با خوشحالی منجر شدن، (ج) با ناراحتی ترسیدن، (د) با ناراحتی عصبانی شدن، (ه) با ناراحتی متعجب شدن، (و) با ناراحتی منجر شدن، (ز) با ترس عصبانی شدن، (ح) با ترس متعجب شدن، (ط) با ترس منجر شدن، (ی) با عصبانیت متعجب شدن، (ک) با عصبانیت منجر شدن، (ل) با انزجار متعجب شدن [۶]

داده‌های تصاویر حالات چهره تحت عنوان تصاویر محکزن^{۳۲} و تحت نام‌های،^{۳۳}KDEF،^{۳۴}DISFA،^{۳۵}CE^{۳۶}،^{۳۷}CK+،^{۳۸}BP4D-Spontaneous،^{۳۹}BU-3DEF،^{۴۰}B+،^{۴۱}JAFFE،^{۴۲}MMI تهیه شده‌اند؛ که مقایسه این پایگاه تصاویر حالات چهره از نظر تعداد نمونه، تعداد افراد، جنسیت، سن، نژاد و وضوح تصاویر در جدول ۱ نشان داده شده است. هرکدام از این پایگاه داده‌ها بر اساس کاربرد و زمینه مطالعاتی محقق می‌تواند حائز اهمیت باشد و مورد استفاده واقع شود. شناسایی و تجزیه و تحلیل تصاویر حالات چهره به دلایل مختلف از قبیل تفاوت سنی در افراد، وجود انسداد^{۴۳}، سالخوردگی، تغییرات روشنایی^{۴۴}، دوری و نزدیکی

حقارت^{۴۵}، شرم^{۴۶}، ناامیدی^{۴۷}، تأسف^{۴۸}، هیجان^{۴۹} و دیوانگی^{۵۰} وجود دارند که برای آن‌ها پایگاه داده معتبری وجود ندارد. در طی دهه گذشته، به شناسایی احساسات چهره در بین موضوعات جالب و چالش برانگیز زمینه بینایی کامپیوتر^{۵۱}، تشخیص الگو^{۵۲} و زمینه مربوط به آن توجه زیادی شده است [۷].

۲-۲. پایگاه داده‌های شناسایی حالات چهره

برای پیاده‌سازی روش‌های شناسایی احساسات چهره وجود یک پایگاه داده استاندارد برای اعتبارسنجی نتایج از اهمیت بسزایی برخوردار است. با توجه به اهمیت و نقش تصویر چهره در شناسایی افراد و تشخیص حالات گوناگون در محیط‌ها و شرایط مختلف، مجموعه

32- Benchmark
33- Cohn-Kanade Dataset (CK+)
34- Compound Emotion (CE)
35- Denver Intensity of Spontaneous Facial Action (DISFA)
36- Karolinska Directed Emotional Face (KDEF)
37- Binghamton-Pittsburgh 3D Dynamic Spontaneous (BP4D-Spontaneous)
38- Binghamton University 3D Facial Expression (BU-3DEF)
39- Japanese Female Facial Expressions (JAFFE)
40- Occlusion
41- Illumination

24- Contempt
25- Embarrass
26- Frustrated
27- Sorry
28- Excited
29- Mad
30- Computer Vision
31- Pattern Recognition

تصویر، تغییر زوایای چهره مقابل دوربین، استفاده از لواز م‌آرایشی، انجام عمل‌های جراحی، وجود احساساتی که در پایگاه داده موجود نیستند و غیره امری پیچیده و دشوار است [۳۱]. در ادامه برخی از این چالش‌ها به‌طور مختصر مطرح شده‌اند.

از آنجایی که شکل و بافت صورت از نظر هویت، جنسیت، سن و گروه قومی در جوامع بشری متفاوت است، انتخاب جمعیت به دلیل این تنوع بین فردی در جمع‌آوری پایگاه داده‌ها مهم است. به‌عنوان مثال میانگین باز شدن چشم بین آسیایی‌ها و قفقازی‌ها متفاوت است و همچنین فاصله اجزای چهره در رده‌های سنی مختلف، متفاوت است. بدین ترتیب برای توسعه روشی که مقاوم با تغییرپذیری بین فردی باشد، پایگاه داده باید شامل طیف گسترده‌ای از گروه‌های قومی، توزیع مناسب سن و جنسیت در بین افراد باشد؛ یعنی تغییرپذیری بین فردی تا حد ممکن باشد. همچنین تهیه تصاویر در شرایط آزمایشگاهی نیز شامل پس‌زمینه، تغییرات روشنایی، تغییر در وضعیت سر و انسداد است. پس‌زمینه و شرایط نوری در مورد تغییرات محیطی است. درحالی‌که تغییرات سر و انسداد در مورد تغییر و تنوع درون شخصی است. منبع نور می‌تواند ویژگی اجزای چهره را تحت تأثیر قرار دهد و در مواردی موجب کاهش دقت شناسایی گردد. به همین دلیل به‌منظور کاهش تغییرات روشنایی روش‌های پیش‌پردازش تصویر مانند نرمال‌سازی DCT^{۴۲}، متعادل‌سازی هیستوگرام^{۴۳} و غیره در تصویر ورودی اعمال می‌شود. انسداد در تصویر به معنی وجود مانع در شیء است و می‌تواند شامل عینک، مو بر روی صورت، دست در مقابل صورت، لباس و غیره باشد. همچنین تصویر چهره ممکن است دارای زاویه متفاوت نسبت به دوربین باشد و همین امر سبب می‌شود برخی از ویژگی‌های چهره مانند چشم یا بینی به‌طور جزئی یا کامل دیده نشود. برای غلبه بر این مشکل از ویژگی‌هایی مانند انتقال^{۴۴}، چرخش^{۴۵} و مقیاس‌گذاری^{۴۶} استفاده می‌شوند که

42- DCT Normalization
43- Histogram Equalization
44- Translation
45- Rotation
46- Scaling

از مواردی هستند که در بحث ارزیابی مقاوم بودن روش‌ها مورد استفاده قرار می‌گیرند.

۲-۳. برخی از کاربردهای شناسایی حالات چهره

امروزه با پیشرفت فن‌آوری، شناسایی احساسات چهره، در زمینه تعاملات انسان-کامپیوتر به‌سرعت در حال رشد است. این موضوع در زمینه‌های بسیاری مانند پزشکی، راهنمایی و رانندگی، آموزش، نیروی انتظامی، روانشناسی و غیره کاربرد دارد که در ادامه هرکدام از موارد مطرح‌شده به‌طور مختصر مورد بررسی قرار می‌گیرند [۸-۱۰].

پزشکی: این حوزه بر اختراع روبات رفاه اجتماعی متمرکز شده است تا به فردی که از لحاظ فیزیکی غیرفعال است و یا در رختخواب یا صندلی چرخ‌دار است، کمک نماید. برای مثال، اختراع روبات پرستار می‌تواند با ارزیابی احساسات بیمار در بیمارستان کمک نماید. همچنین این موضوع می‌تواند در تشخیص و درمان بیماری‌هایی مانند اوتیسم، شیزوفرنی یا ناراحتی‌های روحی ناشی از حوادث مختلف مؤثر باشد. به‌طور مثال با استفاده از شناسایی حالات چهره می‌توان به بیماران اوتیسم کمک کرد تا بتوانند احساسات انسان‌های دیگر را تشخیص دهند.

راهنمایی و رانندگی: تشخیص خستگی و خواب‌آلودگی فرد و شناسایی وضعیت روحی افراد و نظارت بر خواب‌آلودگی یا وضعیت عاطفی راننده برای ایمنی و کاهش تصادفات جاده‌ای ضروری است.

آموزش: در سیستم‌های مشاوره خودکار، کشف دروغ و سیستم آموزش خودکار کاربرد قابل توجهی دارد. علاوه بر این با تشخیص ناامیدی دانش‌آموزان می‌توان به بهبود تجربیات یادگیری الکترونیکی کمک کند.

نیروی انتظامی: تشخیص حقیقت یا فریب احتمالی در بازجویی‌های پلیس یا مصاحبه شغلی یکی دیگر از کاربردهای این حوزه است.

روانشناسی: تشخیص افسردگی از حالات صورت یکی از کاربردهای مهم در تجزیه و تحلیل روان‌شناسی است.

جدول ۱: برخی از پایگاه داده‌های موجود در شناسایی حالات چهره

منبع	نام پایگاه داده	تعداد نمونه	تعداد افراد	توزیع جنسیت	توزیع سن	توزیع نژاد	وضوح تصویر (پیکسل)	توضیحات
[۳۰]	CK+ ¹	۵۹۳ توالی تصویر	۱۲۳	اکثریت زن	۱۸-۳۰	نژادهای مختلف	۶۴۰×۴۸۰ ۶۴۰×۴۹۰	شامل ۶۸ لندمارک برای هر تصویر - شامل شش حالت پایه به‌علاوه حالت طبیعی و غرور
[۶]	CE ²	۵۰۶۰	۲۳۰	۱۳۰ زن ۱۰۰ مرد	میانگین ۲۳	شامل اکثر نژادها از جمله قفقازی، آسیایی، آفریقایی، اسپانیایی	۳۰۰۰×۴۰۰۰	شامل ۲۲ دسته از احساسات پایه و احساسات ترکیبی - تصاویر رنگی - تصاویر بدون انسداد، عینک و موی صورت
[۳۳]	DISFA ³	۴۸۴۵ فریم ویدئو	۲۷	۱۲ زن ۱۵ مرد	بزرگسال	نژادهای مختلف	۱۰۲۴×۷۶۸	شامل ۶۶ لندمارک برای هر تصویر
[۴۰]	KDEF ⁴	۴۹۰۰ تصویر	۷۰	X	X	X	۵۶۲×۷۶۲	شامل شش حالت پایه به‌علاوه حالت طبیعی از پنج زاویه
[۲۴]	BP4D-Spon-tanous ⁵	X	۴۱	۲۳ زن ۱۸ مرد	۱۸-۲۹	آسیایی، آفریقایی - آمریکایی، اسپانیایی	۱۰۴۰×۱۳۲۹	پایگاه داده سه‌بعدی
[۳۲]	BU-3DEF ⁶	۲۵۰۰ تصویر	۱۰۰	۵۶ زن ۴۴ مرد	۱۸-۷۰	X	۱۰۴۰×۱۳۲۹	پایگاه داده سه‌بعدی شامل شش حالت پایه به‌علاوه حالت طبیعی
[۲۴]	B+ ⁷	۱۶۱۲۸ تصویر	۲۸	X	X	X	۳۲۰×۲۴۳	-
[۲۰]	JAFFE ⁸	۲۱۳ تصویر	۱۰	زن	X	ژاپن	۲۵۶×۲۵۶	شامل شش حالت پایه به‌علاوه حالت طبیعی
[۲۳]	MMI ⁹	۲۹۰۰ ویدئو و ۷۴۰ تصویر	۷۵	X	X	X	۷۲۰×۵۷۶	شامل شش حالت پایه به‌علاوه حالت طبیعی

1-<http://www.consortium.ri.cmu.edu/ckagree/>

2-http://cbcs1.ece.ohio-state.edu/dbform_compound.html

3-<http://mohammadmahoor.com/disfa/>

4-<https://www.emotionlab.se/kdef/>

5-http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html

6-http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html

7-<http://vision.ucsd.edu/content/extended-yale-face-database-b-b>

8-<http://www.kasrl.org/jaffe.html>

9-<https://mmifacedb.eu/>

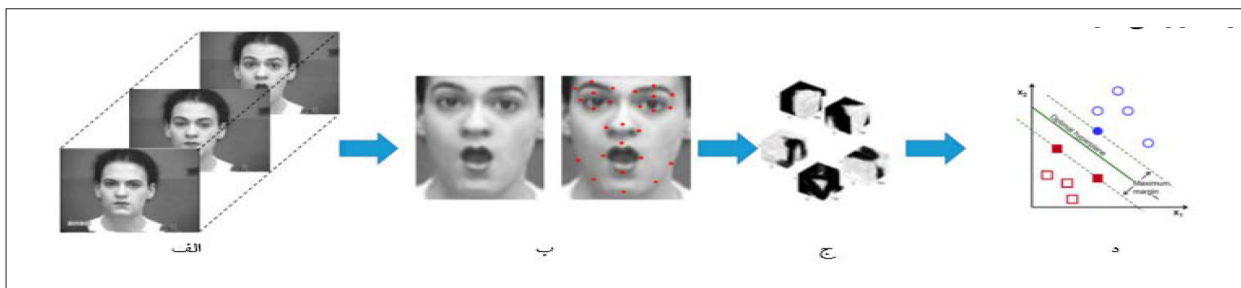
۲-۴. نرم‌افزارهای شناسایی حالات چهره

در زمینه شناسایی حالات چهره نرم‌افزارهای مختلفی همانند رخسار، نرم‌افزار شرکت مایکروسافت و فیس ریدر^{۴۷} وجود دارد؛ که در ادامه هرکدام مورد بررسی قرار می‌گیرد. رخسار^{۴۸}: یکی از نرم‌افزارهای داخلی که برای تشخیص

احساسات عاطفی چهره مورد استفاده قرار می‌گیرد، برنامه رخسار است. این نرم‌افزار با استفاده از علم بینایی ماشین و پردازش تصویر، تصاویر را تحلیل می‌کند و میزان احساسات مختلف چهره‌های موجود در تصاویر را به صورت ریاضی نمایش می‌دهد. یکی از ویژگی‌های

47- FaceReader

48- <https://iranapps.ir/app/ir.rooyan.rokhsar>



شکل ۴: مراحل شناسایی حالات چهره: (الف) تصویر ورودی، (ب) آشکارسازی چهره و اجزای آن، (ج) استخراج ویژگی‌ها از اجزای چهره، (د) طبقه‌بندی حالت چهره با استفاده از الگوهای آموزش داده شده [۲۴]

برای کاربردهای اشاره‌شده مفید باشند. به طور مثال نرم‌افزار کنترل احساس بیماران و یا تشخیص میزان توجه حضار به سخنرانی و غیره که در آینده این فناوری به مرحله‌ای می‌رسد که می‌تواند به روبات‌ها در انجام دقیق‌تر وظایف محوله در خانه یا محل کار کمک کند.

۳. مراحل شناسایی حالات چهره

شناسایی حالات چهره بعد از تشخیص چهره و اجزای آن، از دو مرحله استخراج ویژگی از اجزای چهره و طبقه‌بندی حالات چهره تشکیل شده است [۲۴]. یک روش کلی شناسایی حالات چهره، در شکل ۴ نشان داده شده است. بعد از اخذ تصویر ورودی، چهره شناسایی می‌گردد و ویژگی‌های مهم از چهره آشکار شده، استخراج می‌گردد و در نهایت طبقه‌بندی با استفاده از ویژگی‌های استخراج شده صورت می‌گیرد. برای انجام هر کدام از این مراحل از الگوریتم‌های خاصی استفاده می‌شود که در ادامه مورد بررسی مختصری قرار می‌گیرند.

۳-۱. آشکارسازی چهره و اجزای آن

مرحله آشکارسازی چهره باعث می‌شود چهره در صحنه‌های پیچیده با زمینه‌های مختلط پیدا شود. در ابتدا تصویر چهره و اجزای چهره مانند چشم‌ها، بینی، ابروها و دهان از تصویر ورودی شناسایی می‌شود. در واقع این مرحله موقعیت و اندازه چهره را مشخص می‌کند. در جدول ۲ برخی از روش‌های تشخیص چهره و اجزای آن مانند آدابوست^۱، ماشین بردار پشتیبان^۲، شبکه عصبی

منحصربه‌فرد این نرم‌افزار قابلیت تحلیل حداکثر ۶۴ چهره در یک تصویر است. این نرم‌افزار قادر است هشت احساس خشم، انزجار، ترس، خوشحالی، تعجب، ناراحتی، خواری و سردی را تشخیص دهد. تصاویر گرفته شده با دوربین گوشی، تصاویر ذخیره شده از صفحات وب، تصاویر ذخیره شده از نرم‌افزارهای شبکه‌های اجتماعی و گپ‌زنی؛ قابل تحلیل خواهد بود.

نرم‌افزار شرکت مایکروسافت^۹: مایکروسافت نرم‌افزاری ارائه کرده است که قادر به تشخیص احساسات افراد در تصاویر مختلف است. این نرم‌افزار شش حالت پایه چهره به همراه حالت حقارت و طبیعی را با دقت بسیار بالایی تشخیص می‌دهد. استفاده از این نرم‌افزار نیازی به دانلود و نصب ندارد و کاربران می‌توانند تنها با بارگذاری تصاویر خود در وبگاه این شرکت، پس از چند لحظه احساسات مختلف خود را به صورت درصدی مشاهده کنند [۱۱].

نرم‌افزار فیس ریدر^{۱۰}: این نرم‌افزار برای تحلیل چهره است که می‌تواند حالات عاطفی را در چهره شخص شناسایی کند. این سیستم می‌تواند شش حالت اصلی به همراه حالت تمسخر را تشخیص دهد. این برنامه همچنین می‌تواند تعیین کند که حالات عاطفی فرد مثبت است یا منفی. در این نرم‌افزار جنسیت، سن، نژاد، میزان موهای صورت (ریش یا سبیل) و این که فرد عینک به چشم دارد یا خیر، نیز تشخیص داده می‌شود [۱۲].

نرم‌افزارهایی که در این زمینه تولید می‌شوند، می‌توانند

49- <https://azure.microsoft.com/en-us/services/cognitive-services/face/>

50- <https://www.noldus.com/human-behavior-research/products/facereader>

51- AdaBoost

52- Support Vector Machine (SVM)

جدول ۲: برخی از روش‌های متداول تشخیص چهره و اجزای آن [۱۴-۱۶]

بخش بندی چهره	تشخیص چهره	ثبت چهره	موقعیت‌یابی چهره	رون
×	✓	×	✓	آدابوست
×	✓	×	✓	ماشین بردار پشتیبان
×	✓	×	✓	شبکه عصبی پیچشی
✓	×	×	✓	نقشه برجستگی چهره
✓	×	×	✓	جنگل تصادفی
×	×	✓	×	مدل شکل فعال
×	×	✓	×	مدل ظاهر فعال
×	×	✓	×	میدان تصادفی شرطی
×	×	✓	×	رگرسیون خطی

ویژگی مبتنی بر ایستا یا پویا: چنانچه فقط از یک تصویر به عنوان ورودی استفاده شود، روش ایستا و اگر توالی^{۶۶} تصاویر در نظر گرفته شود، به طوری که اولین تصویر با یک حالت خنثی آغاز شود و در نهایت با یکی از شش حالت پایه تمام شود، روش پویا است. در جدول ۳ با توجه به دسته‌بندی‌های مطرح‌شده برخی از روش‌های استخراج ویژگی مانند هیستوگرام هرمی گرادیان جهت‌دار^{۶۷}، کوانتیزه‌ساز فاز محلی^{۶۸}، موقعیت لندمارک^{۶۹}، فاصله لندمارک^{۷۰}، فیلترگابور^{۷۱}، تبدیل کسینوسی دوبعدی^{۷۲}، الگوی دودویی محلی^{۷۳} و غیره نشان داده شده است.

از مهم‌ترین روش‌های استخراج ویژگی می‌توان به روش‌های مبتنی بر مدل شکل و ظاهر فعال، الگوی دودویی محلی و هیستوگرام گرادیان جهت‌دار^{۷۴} اشاره کرد؛ که در ادامه این روش‌ها به طور مختصر بیان شده‌اند.

مدل ظاهر فعال: این روش تعمیم‌یافته مدل شکل فعال و یکی از روش‌های موقعیت‌یابی دقیق اشیاء است.

- 66- Sequence
- 67- Pyramid Histogram of Oriented Gradient (PHOG)
- 68- Local Phase Quantization (LPQ)
- 69- Landmark location
- 70- Landmark distance
- 71- Gabor Filter
- 72- two-Dimension Discrete Cosine Transform (2D-DCT)
- 73- Local Binary Pattern (LBP)
- 74- Histograms of Oriented Gradients (HOG)

پیچشی، نقشه برجستگی چهره^{۷۵}، جنگل تصادفی^{۷۶}، مدل شکل فعال^{۷۷}، مدل ظاهر فعال^{۷۸}، میدان تصادفی شرطی^{۷۹} و رگرسیون خطی^{۸۰} نشان داده شده است.

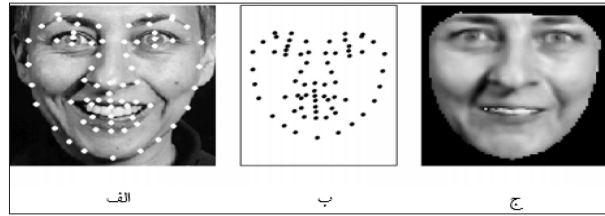
۳-۲. استخراج ویژگی از اجزای چهره

استخراج ویژگی فرآیندی است که در آن با انجام عملیاتی بر روی تصویر، ویژگی‌های بارز و تعیین‌کننده آن مشخص می‌شود. هدف استخراج ویژگی این است که داده‌های خام به شکل قابل‌استفاده‌تری برای پردازش درآیند. روش‌های مبتنی بر استخراج ویژگی یک فضای چندبعدی را به یک فضای با ابعاد کمتر نگاشت می‌دهند؛ و بردار ویژگی را تشکیل می‌دهند. کاهش ابعاد داده‌ها به معنی کم کردن تعداد ویژگی‌های تصویر است. برای روش‌های استخراج ویژگی دسته‌بندی‌های مختلفی مانند مبتنی بر مدل ظاهری^{۸۱} یا مبتنی بر هندسه^{۸۲}، مبتنی بر ویژگی‌های عمومی^{۸۳} یا محلی^{۸۴} و مبتنی بر ایستا^{۸۵} یا پویا^{۸۶} ارائه شده است.

ویژگی مبتنی بر مدل ظاهری یا مبتنی بر شکل هندسه چهره: در روش‌های مبتنی بر مدل ظاهری از فیلترهایی بر روی بافت چهره استفاده می‌شود که این فیلترها می‌توانند بر روی قسمت خاصی از چهره مانند چشم، دهان، بینی و یا بر روی کل چهره اعمال شوند و در روش‌های مبتنی بر هندسه برای استخراج ویژگی از فاصله بین اجزای صورت استفاده می‌شود. به طور مثال فاصله بین چشم‌ها، فاصله ابرو تا چشم، فاصله بینی تا دهان و غیره.

مبتنی بر ویژگی‌های عمومی یا محلی: ویژگی‌های عمومی اطلاعات را از کل ناحیه صورت و ویژگی‌های محلی از قسمت‌های موردعلاقه^{۸۷} چهره استخراج می‌کنند.

- 53- Facial Saliency Map (FSM)
- 54- Random Forests (RF)
- 55- Active Shape Model (ASM)
- 56- Active Appearance Model (AAM)
- 57- Conditional Random Fields (CRF)
- 58- Linear Regression
- 59- Appearance Model
- 60- Geometric
- 61- Global Feature
- 62- Local Feature
- 63- Static
- 64- Dynamic
- 65- Region of Interested (ROI)



شکل ۵: نمونه‌ای از عملکرد روش مدل ظاهر فعال: (الف) تعیین ۶۸ نقطه کلیدی چهره، (ب) استخراج مدل شکل چهره، (ج) استخراج بافت چهره [۳۹]

جدول ۳: برخی روش‌های متداول استخراج ویژگی [۱۴-۱۶]

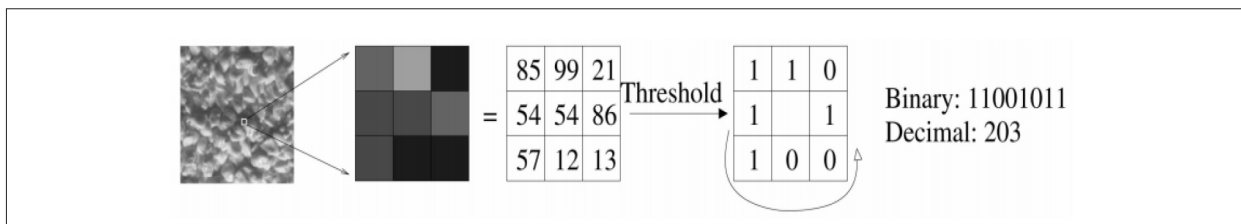
نام روش	مبنی بر ظاهر	مبنی بر هندسه	پویا	ایستا	محلی	عمومی
هیستوگرام هرمی گرادیان جهت‌دار	✓	×	×	✓	×	✓
کوانتیزه ساز فاز محلی	✓	×	×	✓	×	✓
موقعیت لندمارک	×	✓	×	✓	×	✓
فاصله لندمارک	×	✓	×	✓	×	✓
فیلتر گابور	✓	×	×	✓	×	✓
تبدیل کسینوسی دوبعدی	✓	×	×	✓	×	✓
الگوی دودویی محلی	✓	✓	✓	✓	×	✓
LBP-TOP ^{۱۰}	✓	✓	✓	✓	×	✓
LPQ-TOP ^{۱۱}	✓	×	✓	✓	×	✓
میانگین شدت روشنایی ^{۱۲}	✓	×	✓	✓	✓	×
Bow Hist	✓	×	✓	×	✓	×
جریان نوری ^{۱۳}	×	✓	✓	×	×	✓
FFD ^{۱۴}	×	✓	✓	×	×	✓
FAP ^{۱۵}	×	✓	✓	×	✓	×

درواقع مدل ظاهر فعال یک توصیفگر متداول چهره است که از تجزیه و تحلیل مؤلفه‌های اصلی^{۷۵} برای کاهش ابعاد و درعین حال حفظ عناصر مهم و بافت تصاویر چهره استفاده می‌کند. وقتی که این روش بر روی تصاویر اعمال می‌شود، می‌تواند به‌طور توأم تغییرات شکل^{۷۶} و بافت^{۷۷} چهره را با توجه به نقاط و اجزای کلیدی چهره مانند نوک بینی، چشم‌ها، دهان و ابرو استخراج کند. مدل‌های متفاوتی از نظر تعداد این نقاط وجود دارد. یک مدل، پنج نقطه‌ای است که فقط چشم‌ها و پائین بینی را شناسایی می‌کند. مدل دیگر

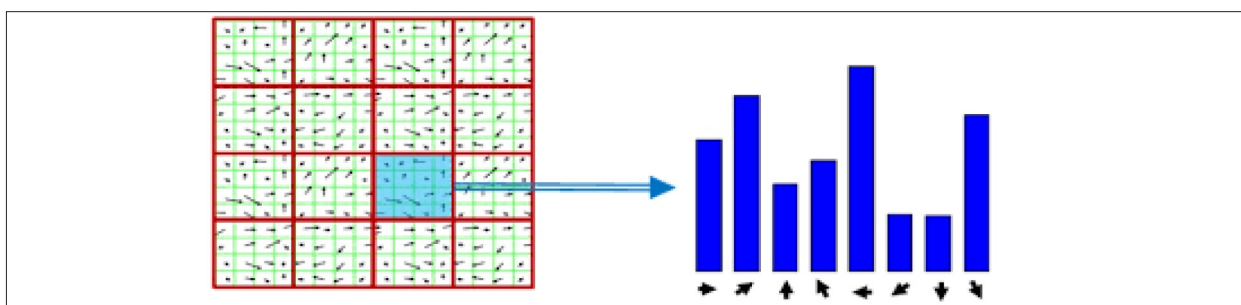
75- Principal Component Analysis (PCA)
76- Shape
77- Texture

شامل ۶۸ نقطه کلیدی است که شامل تشخیص چشم‌ها، بینی، دهان و مدل چهره می‌شود [۳۸، ۳۹]. در شکل ۵ (الف) تصویر چهره با ۶۸ نقطه کلیدی نشان داده شده است. مدل ظاهر فعال در سه مرحله اصلی تولید می‌شود: ۱- ایجاد یک مدل شکل آماری برای مدل‌سازی تغییرات شکل یک شیء با استفاده از مجموعه‌ای از تصاویر آموزشی ۲- ایجاد یک مدل بافت برای مدل‌سازی تغییرات بافت که توسط شدت پیکسل‌ها نمایش داده می‌شود. ۳- ایجاد مدل ظاهر نهایی توسط ترکیب کردن مدل‌های شکل و بافت. در شکل ۵، شکل و بافت تصویر پس از اعمال الگوریتم مدل ظاهر فعال نشان داده شده است. در این شکل هر مدل تصویر با استفاده یک بردار $S=[x_1, y_1, \dots, x_n, y_n]$ به طول $2n$ تعریف می‌شود که n تعداد نقاط کلیدی تصویر است.

الگوی دودویی محلی: این الگوریتم یک توصیفگر محلی و یک قدرتمند برای توصیف بافت است. این توصیفگر از یک پنجره مثلاً با ابعاد 3×3 تشکیل می‌شود که ابعاد این پنجره می‌تواند گسترش یابد. برای انجام این کار یک دایره همسایگی (P, R) را تعریف می‌کنیم؛ که P تعداد همسایه‌ها و R شعاع همسایگی است. مرکز این پنجره بر روی هر پیکسل قرار داده شده و برای هر پیکسل در این پنجره، مقدار آن را با هر یک از هشت همسایه P_n ($n=0,1,\dots,7$) مقایسه می‌کنیم. چنانچه مقدار پیکسل مرکزی از همسایه‌اش بیشتر بود، مقدار یک و در غیر این صورت مقدار صفر را جایگزین می‌کنیم؛ و پیکسل‌ها را به‌صورت یک دایره در جهت عقربه‌های ساعت و یا خلاف جهت، دنبال می‌کنیم. برچسب LBP برای هر پیکسل مرکزی (x,y) تصویر $f(x,y)$ طبق رابطه ۱ محاسبه می‌شود. نتیجه به شکل یک رشته دودویی هشت بیتی درآمده و سپس به مقدار دهی معادل تبدیل می‌شود. در نهایت هیستوگرامی از مقادیر به‌عنوان بردار ویژگی‌ها تولید خواهد شد. در شکل ۶ نمایی از عملکرد توصیفگر الگوی دودویی محلی نشان داده شده است. این روش با استفاده از عملگرهای



شکل ۶: مثالی از نحوه عملکرد توصیفگر الگوی دودویی محلی [۳۴]



شکل ۷: چگونگی ایجاد بردار ویژگی هیستوگرام با استفاده روش گرادیان جهت‌دار [۳۷]

داده شده است.

$$g = \sqrt{g_x^2 + g_y^2} \quad (۳)$$

$$\theta = \arctan \frac{g_y}{g_x} \quad (۴)$$

غیرخطی، بردار ویژگی با ابعاد کم را استخراج می‌کند [۳۴].

$$LBP_{P,R}(x, y) = \sum_{p=0}^{P-1} s((f(x_p) - f(x, y))) 2^p \quad (۱)$$

که در این معادله $s(z)$ تابع آستانه است و به صورت

زیر مقداردهی می‌شود.

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \quad (۲)$$

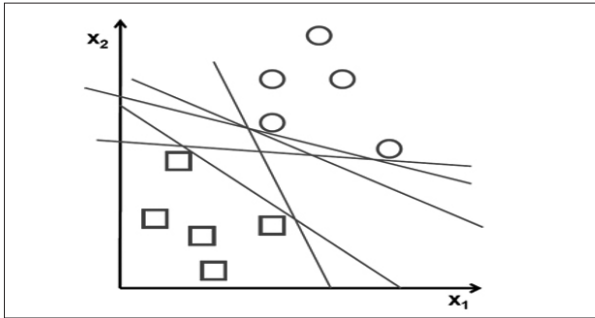
۳-۳. طبقه‌بندی حالات چهره

بعد از استخراج بردار ویژگی باید حالت چهره تشخیص داده شود که این حالت می‌تواند یکی از شش حالت اصلی و یا حالت‌هایی مثل درد، خواب‌آلودگی و دل‌بستگی عاطفی یا شاخص‌های اختلال روان‌پزشکی باشد. برخی از معمول‌ترین روش‌هایی که برای طبقه‌بندی وجود دارند، آدابوست، ماشین بردار پشتیبان، شبکه عصبی پیچشی و شبکه خودمزمگذار^{۸۰} می‌باشند؛ که در ادامه هر کدام به‌طور مختصر توضیح داده شده‌اند. همچنین در جدول ۴ برخی از روش‌های طبقه‌بندی مانند طبقه‌بندی شبکه بیزین^{۸۱} شبکه عصبی، K نزدیک‌ترین همسایه^{۸۲}، شبکه عصبی بازگشتی^{۸۳} ماشین بردار پشتیبان، جنگل تصادفی، شبکه عصبی پیچشی، رگرسیون هسته^{۸۴}، حافظه طولانی کوتاه‌مدت^{۸۵} و فیلتر ذرات^{۸۶} مورد بررسی قرار گرفته است.

هیستوگرام گرادیان جهت‌دار: در این روش تصویر به سلول‌های مربعی کوچک مثلاً 8×8 تقسیم می‌شود. در هر سلول اندازه^{۷۸} و زاویه گرادیان^{۷۹} با توجه به رابطه‌های ۳ و ۴ تعیین می‌گردد؛ و به ازای هر پیکسل گرادیان افقی و عمودی محاسبه می‌شود که می‌تواند به‌طور مؤثری ویژگی شکل یک تصویر را ضبط کند. متناسب با مقادیر اندازه و زاویه در هر سلول بردار هیستوگرام محاسبه می‌شود. سپس با استفاده از هیستوگرام حاصل یک بردار ویژگی تشکیل می‌شود. در نهایت با کنار هم قرار دادن بردارهای ویژگی سلول‌ها یک بردار ویژگی بزرگ‌تر ساخته می‌شود که بردار ویژگی نهایی حاصل از روش هیستوگرام گرادیان جهت‌دار است [۳۷]. در شکل ۷ چگونگی ایجاد بردار ویژگی هیستوگرام با استفاده از روش گرادیان جهت‌دار نمایش

78- Magnitude Gradient
79- Direction Gradient

80- Auto-encoder
81- Bayesian Network Classifiers (BNC)
82- K-Nearest Neighbors (KNN)
83- Recurrent Neural Network (RNN)
84- Kernel Regression
85- Long Short-term Memory (LSTM)
86- Particle Filter



شکل ۸: جداکننده داده‌ها در الگوریتم SVM [۲۲]

که می‌توان آن‌ها را با γ برچسب‌گذاری کرد. در این حالت نمونه‌ها به شکل زوج‌های مرتب (x_i, y_i) بیان می‌شود که $i=1, \dots, n$ و y_i نیز دارای مقادیر $\{-1, 1\}$ است. همان‌طور که در شکل ۸ مشاهده می‌شود این داده‌ها را می‌توان توسط چندین جداکننده تفکیک و طبقه‌بندی کرد. تعدادی از این نمونه‌های آموزشی که کمترین فاصله تا مرز تصمیم‌گیری دارند، به‌عنوان بردار پشتیبان در نظر گرفته می‌شوند. در فضای دوبعدی، بردارهای پشتیبان، یک خط، در فضای سه‌بعدی یک صفحه و در فضای n بعدی یک ابرصفحه را شکل خواهند داد. در واقع هدف اصلی این روش ایجاد یک ابرصفحه برای جداسازی طبقه‌ها (طبقه‌بندی داده‌ها) است به‌گونه‌ای که این ابرصفحه بتواند نسبت به نمونه‌های موجود در طبقه‌ها بیشترین فاصله را داشته باشد. محاسبه مرز تصمیم‌گیری دو رده کاملاً جدا از هم از روش حاشیه بهینه استفاده می‌شود. در این روش مرز خطی بین دو رده به‌گونه‌ای انتخاب می‌شود که تمام نمونه‌های رده با برچسب $+1$ در یک‌طرف مرز و تمام نمونه‌های رده -1 در طرف دیگر مرز واقع شوند. یک مرز تصمیم‌گیری خطی را در حالت کلی می‌توان به‌صورت رابطه ۶ نوشت. در این معادله x یک نقطه روی مرز تصمیم‌گیری و w یک بردار n بعدی عمود بر مرز تصمیم‌گیری است [۲۲].

$$wx=b=0 \quad (6)$$

شبکه عصبی پیچشی: این روش توسط سه لایه پردازش اصلی، لایه پیچش^{۸۹}، لایه ادغام^{۹۰} و لایه

جدول ۴: دسته‌بندی برخی از روش‌های متداول طبقه‌بندی [۱۴-۱۶]

روش	categorical	Continuous	ایستا	پویا
طبقه‌بندی شبکه بیزین	✓	✗	✓	✗
شبکه عصبی	✓	✗	✓	✗
K نزدیک‌ترین همسایه	✓	✗	✓	✗
شبکه عصبی بازگشتی	✗	✓	✓	✗
ماشین بردار پشتیبان	✓	✗	✓	✓
جنگل تصادفی	✓	✗	✓	✓
شبکه عصبی پیچشی	✓	✗	✓	✗
رگرسیون هسته	✗	✓	✓	✗
حافظه طولانی کوتاه‌مدت	✓	✗	✓	✗
فیلتر ذرات	✗	✓	✗	✓

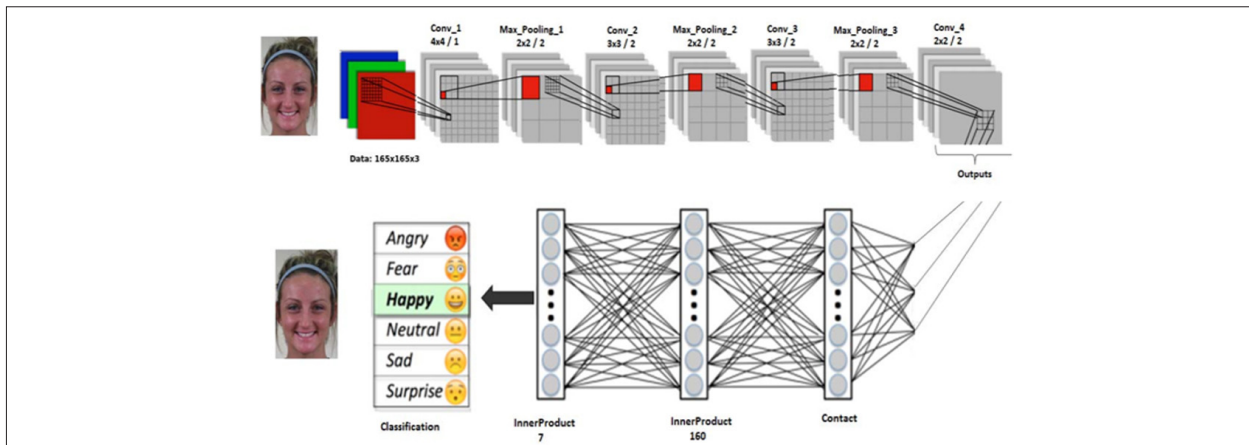
آدابوست: این روش یک الگوریتم تهاجمی^{۸۷} است که در هر مرحله یک طبقه‌بندی ضعیف را انتخاب می‌کند. در هر مرحله t به هرکدام از نمونه‌های آموزشی یک مقدار وزن $(d^t = (d_1^t, \dots, d_n^t))$ تعلق می‌گیرد؛ و بر اساس مجموعه وزن‌های (d^t) ، یادگیرنده ضعیف h_t ایجاد می‌شود. در هر تکرار t مجموعه وزن نمونه‌ها با توجه به رابطه ۵ به‌روزرسانی می‌شوند. بدین‌صورت که مقدار وزن برای نمونه‌هایی که به‌درستی طبقه‌بندی شوند، کم و برای نمونه‌هایی که اشتباه طبقه‌بندی شوند، زیاد می‌شود. مقدار خطای آموزش نیز در هر مرحله برابر با مجموعه داده‌های اشتباه طبقه‌بندی می‌شود. بنابراین با این روش در هر مرحله روی نمونه‌هایی که در مراحل قبل اشتباه طبقه‌بندی شده‌اند، تمرکز می‌شود [۳۶].

$$D_n^{(t+1)} = D_n^t \frac{\exp\{-\alpha_t y_n h_t(x_n)\}}{Z_t} \quad (5)$$

ماشین بردار پشتیبان: این روش یکی از روش‌های یادگیری بانظارت^{۸۸} است که از آن برای طبقه‌بندی استفاده می‌کنند. الگوریتم SVM دو رده را با استفاده از یک مرز خطی از هم جدا می‌کند. در این حالت فرض می‌شود که مجموعه‌ای از نمونه‌های آموزشی تفکیک‌پذیر وجود دارند

89- Convolutional Layer
90- Pooling Layer

87- Aggressive
88- Supervised Learning



شکل ۹: یک معماری نمونه از شبکه عصبی پیچشی [۱۹]

لایه ادغام، لایه تمام‌اتصال وجود دارد. این لایه نقشه ویژگی‌های دوبعدی را به منظور دسته‌بندی تصاویر به بردار ویژگی یک‌بعدی تبدیل می‌کند. در هر شبکه عصبی پیچشی، آموزش با استفاده از دو مرحله پیشرو^{۹۷} و پس انتشار^{۹۸} انجام می‌گیرد. در مرحله اول تصویر ورودی به شبکه داده می‌شود. در واقع این همان عمل ضرب نقطه‌ای بین ورودی و پارامترهای هر نورون و نهایتاً اعمال عملیات پیچشی در هر لایه است. برای این کار خروجی شبکه را با استفاده از یک تابع خطا^{۹۹} با پاسخ صحیح مقایسه کرده و میزان خطا محاسبه می‌شود. در مرحله بعدی بر اساس میزان خطای محاسبه‌شده، مرحله پس انتشار آغاز می‌شود. در این مرحله شیب هر پارامتر با توجه به قاعده زنجیره‌ای^{۱۰۰} محاسبه و تمامی پارامترها با توجه به تأثیری که بر خطای ایجادشده در شبکه دارند، تغییر پیدا می‌کنند. پس از به‌روزرسانی پارامترها مرحله پیشرو شروع می‌شود و بعد از تکرار تعداد مناسبی از این مراحل آموزش شبکه به پایان می‌رسد.

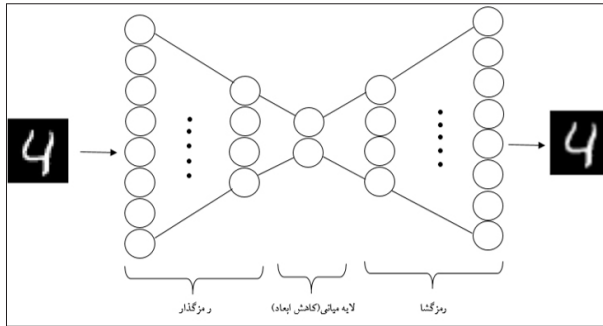
شبکه خودرمزگذار: خود رمزگذار یک الگوریتم یادگیری بدون نظارت^{۱۰۱} است که برای تخمین بردار ویژگی، روش انتشار رو به عقب^{۱۰۲} را اعمال می‌کند. شکل ۱۰ یک خودرمزگذار با یک لایه میانی را نشان می‌دهد. در این شکل خودرمزگذار از سه لایه ورودی،

تمام‌اتصال^{۹۱} ساخته شده است که هرکدام از این لایه‌ها دارای وظایفی هستند. در شکل ۹ یک معماری کلی شبکه عصبی پیچشی برای دسته‌بندی تصاویر نشان داده شده است که در آن تصویر شخص به‌عنوان ورودی به شبکه داده‌شده و در لایه خروجی از میان شش حالت خشم، ترس، شادی، طبیعی، ناراحتی و تعجب، تصویر ورودی در دسته‌بندی حالت شادی قرار گرفته است. در حالت کلی شبکه عصبی پیچشی یک شبکه عصبی سلسله‌مراتبی است که لایه‌های پیش و ادغام آن به‌صورت یک‌درمیان از هم قرار دارند و بعد از آن‌ها لایه تمام‌اتصال وجود دارد. هسته اصلی این نوع شبکه، لایه پیش است و شامل تعدادی نورون است که در سه بعد عرض، ارتفاع و عمق قرار گرفته‌اند و خروجی آن را می‌توان به‌صورت یک مجموعه سه‌بعدی از نورون‌ها و به‌صورت یک صفحه که نقشه ویژگی^{۹۲} نام دارد، نشان داد. این شبکه برای پیش^{۹۳} تصویر ورودی و نقشه ویژگی‌های میانی از هسته‌های^{۹۴} مختلفی استفاده می‌کند [۲۸].

پس از هر لایه پیش یک لایه ادغام قرار می‌گیرد و هدف استفاده از لایه ادغام کاهش اندازه نقشه ویژگی و پارامترهای شبکه است. برای پیاده‌سازی لایه ادغام روش‌های مختلفی وجود دارد که معمولاً از تابع ماکزیمم^{۹۵} و تابع میانگین^{۹۶} استفاده می‌شود. در نهایت بعد از آخرین

97- Feed forward
98- Back Propagation
99- Loss Function
100- Chain Rule
101- Unsupervised
102- Back Propagation

91- Fully Connected Layer
92- Feature Map
93- Convolve
94- Kernel
95- Max Pooling
96- Average Pooling



شکل ۱۱: معماری خودرمزگذار عمیق

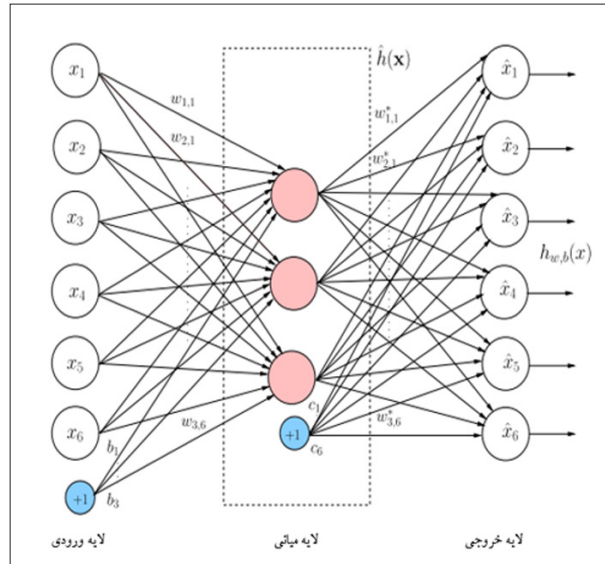
استفاده و دقت شناسایی حاصل از نتایج بیان شده است. به منظور ارزیابی شناسایی حالات چهره، معیارهای آن به چهار روش دقت^{۱۰۷}، بازخوانی^{۱۰۸}، صحت^{۱۰۹} و اندازه^{۱۱۰} F طبقه‌بندی می‌شوند. در این بین مهم‌ترین و عمومی‌ترین معیار برای تعیین کارایی یک الگوریتم طبقه‌بندی، دقت است که نشان می‌دهد طبقه بند طراحی شده چند درصد از کل مجموعه رکوردهای آزمایشی را به درستی طبقه‌بندی کرده است. دقت طبقه‌بندی با استفاده از رابطه ۷ به دست می‌آید که نسبت تشخیص‌های درست (TP+TN) به کل داده‌ها (N) را بیان می‌کند. در این رابطه TN تعداد درست منفی در پایگاه داده و TP تعداد درست مثبت است. همان‌طور که در جدول ۵ نشان داده شده است، در [۲۵] که از روش شبکه عصبی پیچشی در مراحل استخراج ویژگی و طبقه‌بندی استفاده کرده است، دارای بالاترین دقت شناسایی است. اگرچه استفاده از شبکه عصبی باعث افزایش دقت شده است اما از طرفی هزینه محاسباتی بالایی را به همراه دارد. همچنین انتخاب مناسب نرخ یادگیری، اندازه هسته فیلترهای پیچشی و انتخاب تعداد لایه‌های مناسب از چالش‌های این روش است.

$$Accuracy = \frac{TP+TN}{N} \quad (7)$$

۵. نتیجه‌گیری

تجزیه و تحلیل حالات مختلف چهره مانند خشم، ترس، انزجار، شادی، غم، تعجب، غرور، هیجان، تأسف، درد،

107- Accuracy
108- Recall
109- Precision
110- F-Measure



شکل ۱۰: معماری خودرمزگذار با یک لایه میانی [۴]

میانی و خروجی تشکیل شده است. شبکه خود رمزگذار عمیق^{۱۰۳} همان‌طور که در شکل ۱۱ مشاهده می‌شود از چندین لایه و دو بخش رمزگذار^{۱۰۴} و رمزگشا^{۱۰۵} تشکیل می‌شود. این شبکه در بخش رمزگذار به عنوان مثال ارقام دست‌نویس را به عنوان داده‌ی ورودی بدون برچسب^{۱۰۶} دریافت کرده و آن را به فضای ویژگی‌ها نگاشت می‌دهد و در بخش رمزگشا فضای ویژگی را به فضای ورودی نگاشت می‌دهد و سعی در بازنمایی مجدد ورودی‌ها در خروجی دارد به صورتی که کمترین اختلاف ممکن را با مقدار ورودی داشته باشد. در واقع بخش اصلی خودرمزگذار لایه میانی آن است که به عنوان ویژگی استخراج شده برای دسته‌بندی استفاده می‌شود؛ و در همین لایه کاهش ابعاد داده‌ی ورودی انجام می‌شود [۴, ۳۵].

۴. مقایسه برخی از پژوهش‌های انجام‌شده در زمینه شناسایی حالات چهره

جدول ۵ برخی از تحقیقات اخیر با هدف بررسی کارایی آن‌ها در استخراج ویژگی و طبقه‌بندی شناسایی حالات چهره را نشان می‌دهد. همچنین پایگاه داده مورد

103- Deep Auto-encoder
104- Encoder
105- Decoder
106- Unlabeled

جدول ۵: برخی از کارهای انجام شده در شناسایی حالات چهره

منبع	الگوریتم استخراج ویژگی	الگوریتم طبقه‌بندی ویژگی	پایگاه داده ارزیابی شده	دقت شناسایی
[۲۰]	ترکیب SIFT و SboF ^{۱۶}	ماشین بردار پشتیبان	CK+ JAFFE	۹۸/۵ درصد ۹۸/۳ درصد
[۲۶]	تجزیه و تحلیل مؤلفه‌های اصلی	فاصله اقلیدسی	Indian Face	۹۱/۲۶ درصد
[۲۵]	شبکه عصبی پیچشی	شبکه عصبی پیچشی	CK+	۹۹/۳۱ درصد
[۲۷]	ترکیب الگوریتم مدل ظاهر فعال، هیستوگرام گرادیان جهت‌دار و تجزیه و تحلیل مؤلفه‌های اصلی	شبکه عصبی DSAE ^{۱۷}	CK+	۹۵/۷۹ درصد
[۴]	ویژگی‌های هندسی، الگوی دودویی محلی و خود رمزگذار	شبکه عصبی خودسازمان‌ده ^{۱۸}	MMI +CK	۹۷/۵۵ درصد ۹۸/۹۵ درصد
[۲۳]	الگوی دودویی محلی و شبکه عصبی پیچشی	شبکه عصبی پیچشی	CK+ JAFFE	۹۷ درصد ۹۲/۲ درصد
[۲۱]	ترکیب الگوریتم SIFT و GWO ^{۱۹}	شبکه عصبی GWO	CK+ JAFFE	۹۱/۲۲ درصد ۸۹/۷۹ درصد

نهایتاً طبقه‌بندی حالات مختلف چهره داشتند. ولی نتایج با استفاده از یک و یا حداکثر دو پایگاه داده تصویری به دست آمده است که داده‌های موجود دارای بسیاری از حالت‌های دنیای واقعی مانند انسداد، تغییرات روشنایی، دوری و نزدیکی و تغییر زوایای چهره نسبت به دوربین و غیره نبودند و همچنین نتایج برای حالات دیگر مانند هیجان، تأسف، خستگی و غیره محاسبه نشده است. لذا استفاده از چندین پایگاه داده تصویری متنوع و تشخیص حالات مختلف چهره از پژوهش‌های است که باید در این زمینه ادامه یابد.

منابع

- [۱] ح. صادقی، ف. آیت‌اللهی و ا. اسدالله‌راعی، «شناسایی حالت چهره با استفاده از نرمالیزاسیون هندسی و تبدیل موجک مختلط Dual-Tree»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۵، شماره ۳، پاییز ۹۴.
- [۲] ص. ابراندوست، س. مشگینی و س. نصیرزاده، «بازشناسی چهره با استفاده از آنالیز تفکیک خطی بر پایه موجک‌های هار و گابور و ماشین بردار پشتیبان»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۷، شماره ۴، زمستان ۹۶.
- [۳] س. شفیعی‌پور، ه. سیدعربی و ع. آقاگل‌زاده، «شناسایی چهره بر اساس ظاهر در رشته‌های ویدئویی»، مجله مهندسی برق دانشگاه تبریز، جلد ۴۵، شماره ۲، تابستان ۹۴.
- [4] A. Majumder, L. Behera, and V. K. Subramanian, "Automatic facial expression recognition system using deep network-based data fusion," IEEE Transactions on Cybernetics, vol. 48, no. 1, pp. 103–114, 2018.

خواب‌آلودگی و خستگی کاربردهای مختلفی در زمینه‌های آموزش، پزشکی، راهنمایی و رانندگی، روانشناسی و غیره دارد. انسان‌ها از توانایی شگفت‌انگیزی برای رمزگشایی این حالات چهره در فرهنگ‌های مختلف و شرایط متنوع در زمان بسیار کوتاه برخوردارند؛ اما این موضوع برای ماشین که به‌طور خودکار بتواند این حالت‌های متفاوت را تشخیص دهد با چالش‌های مختلفی روبرو است. یکی از اصلی‌ترین چالش‌ها مربوط به الگوریتم‌های استخراج ویژگی و طبقه‌بندی مناسب است. در این مقاله برخی از مهم‌ترین الگوریتم‌های استخراج ویژگی مانند مدل ظاهری، هیستوگرام گرادیان جهت‌دار و الگوی دودویی محلی و طبقه‌بندی حالات مختلف چهره مانند ماشین بردار پشتیبان و شبکه عصبی پیچشی که بیشتر برای شش حالت پایه خشم، ترس، انزجار، شادی، غم و تعجب استفاده شده‌اند معرفی گردیدند. به‌علاوه نه پایگاه داده تصویری که در این زمینه، تشخیص حالات مختلف چهره مورد استفاده قرار گرفته‌اند از نظر تعداد افراد، جنسیت، سن، نژاد و وضوح تصاویر مورد مقایسه قرار گرفتند. گرچه مطالعات انجام شده نشان داد که روش‌های مبتنی بر شبکه عصبی پیچشی بالاترین دقت را در استخراج ویژگی و

pp. 1004–1012, 2018.

[21] N. P. Nirmala Sreedharan, B. Ganesan, R. Raveendran, P. Sarala, B. Dennis, and R. Boothalingam R., “Grey Wolf optimisation-based feature selection and classification for facial emotion recognition,” *IET Biometrics*, vol. 7, no. 5, pp. 490–499, 2018.

[22] E. Osuna, R. Freund, and F. Girosit, “Training support vector machines: an application to face detection,” *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2002.

[23] B. Yang, J. Cao, R. Ni, and Y. Zhang, “Facial expression recognition using weighted mixture deep neural network based on double-channel facial images,” vol. 6, pp. 4630–4640, 2018.

[24] B. Ko, “A brief review of facial emotion recognition based on visual information,” *Sensors*, vol. 18, no. 2, p. 401, 2018.

[25] M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, “Facial expression recognition with identity and emotion joint learning,” *IEEE Transactions on Affective Computing*, pp. 1–1, 2018.

[26] A. R. Surve, V. R. Ghorpade, and A. S. Pathe, “Continuous facial emotion recognition system using PCA for ambient living,” *Advances in Intelligent Systems and Computing*, pp. 319–332, 2018.

[27] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie, “Facial expression recognition via learning deep sparse autoencoders,” *Neurocomputing*, vol. 273, pp. 643–649, 2018.

[28] H. A. Rowley, S. Baluja, and T. Kanade, “Neural network-based face detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998.

[29] L. J. Karam and T. Zhu, “Quality labeled faces in the wild (QLFW): a database for studying face recognition in real-world environments,” *Human Vision and Electronic Imaging XX*, 2015.

[30] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010.

[31] J. Kumari, R. Rajesh, and K. M. Pooja, “Facial Expression Recognition: A Survey,” *Procedia Computer Science*, vol. 58, pp. 486–491, 2015.

[32] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and M. J. Rosato, “A 3D Facial expression database for facial behavior research,” *7th International Conference on Automatic Face and Gesture Recognition*, 2006.

[33] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, “DISFA: A Spontaneous Facial Action Intensity Database,” *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151–160, 2013.

[34] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[35] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, “A sparse auto-encoder-based deep neural network approach

[5] Y.-D. Zhang, Z.-J. Yang, H.-M. Lu, X.-X. Zhou, P. Phillips, Q.-M. Liu, and S.-H. Wang, “Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation,” *IEEE Access*, vol. 4, pp. 8375–8385, 2016.

[6] S. Du, Y. Tao, and A. M. Martinez, “Compound facial expressions of emotion,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 15, pp. E1454–E1462, 2014.

[7] G. Wen, H. Chen, D. Cai, and X. He, “Improving face recognition with domain adaptation,” *Neurocomputing*, vol. 287, pp. 45–51, 2018.

[8] H. Ali, M. Hariharan, S. Yaacob, and A. H. Adom, “Facial emotion recognition using empirical mode decomposition,” *Expert Systems with Applications*, vol. 42, no. 3, pp. 1261–1277, 2015.

[9] S. Ebrahimi Kahou, V. Michalski, K. Konda, R. Memisevic, and C. Pal, “Recurrent Neural Networks for Emotion Recognition in Video,” *Proceedings of the ACM on International Conference on Multimodal Interaction*, 2015.

[10] K. Kulkarni, C. Corneanu, I. Ofodile, S. Escalera, X. Baro, S. Hyniewska, J. Allik, and G. Anbarjafari, “Automatic recognition of facial displays of unfeared emotions,” *IEEE Transactions on Affective Computing*, pp. 1–1, 2018.

[11] Microsoft azure, <https://azure.microsoft.com/en-us/services/cognitive-services/face>. [Accessed 31 3 2019].

[12] Noldus, <https://www.noldus.com/human-behavior-research/products/facereader>. [Accessed 31 3 2019].

[13] F.D. I. Torre, W. S. Chu, X. Xiong, F. Vicente, X. Ding, and J. Cohn, “IntraFace,” *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, 2015.

[14] D. Mehta, MFH. Siddiqui, and A. Y. Javaid, “Facial emotion recognition: A survey and real-world user experiences in mixed reality,” *Sensors*, vol. 18, no. 2, p. 416, 2018.

[15] C. A. Corneanu, M. O. Simon, J. F. Cohn, and S. E. Guerrero, “Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1548–1568, 2016.

[16] J. Kumari, R. Rajesh, and K. M. Pooja, “Facial expression recognition: a survey,” *Procedia Computer Science*, vol. 58, pp. 486–491, 2015.

[17] H. Nguyen, K. Kotani, F. Chen, and B. Le, “A Thermal facial emotion database and Its analysis,” *Lecture Notes in Computer Science*, pp. 397–408, 2014.

[18] D. Lundqvist, A. Flykt, and A. Öhman, “The Karolinska Directed Emotional Faces – KDEF,” CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998.

[19] A. Fathallah, L. Abdi, and A. Douik, “Facial expression recognition via deep learning,” *IEEE/ACS 14th International Conference on Computer Systems and Applications*, 2017.

[20] T. Kalsum, S. M. Anwar, M. Majid, B. Khan, and S. M. Ali, “Emotion recognition from facial expressions using hybrid feature descriptors,” *IET Image Processing*, vol. 12, no. 6,

Z. Cui, "Cross-Domain Facial Expression Recognition Based on Transductive Deep Transfer Learning," IEEE Access, vol. 7, pp. 108906–108915, 2019.

[43] B. Fasel and J. Luetttin, "Automatic facial expression analysis: a survey," Pattern Recognition, vol. 36, no. 1, pp. 259–275, Jan. 2003.

[44] M. Li, H. Xu, X. Huang, Z. Song, X. Liu, and X. Li, "Facial Expression Recognition with Identity and Emotion Joint Learning," IEEE Transactions on Affective Computing, pp. 1–1, 2018.

[45] K. Kulkarni, C. Corneanu, I. Ofodile, S. Escalera, X. Baro, S. Hyniewska, J. Allik, and G. Anbarjafari, "Automatic Recognition of Facial Displays of Unfelt Emotions," IEEE Transactions on Affective Computing, pp. 1–1, 2018.

[46] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial Expression Analysis," Handbook of Face Recognition, pp. 247–275, 2005.

[47] T. Wilhelm, H.-J. Böhme, and H.-M. Gross, "Classification of Face Images for Gender, Age, Facial Expression, and Identity," Lecture Notes in Computer Science, pp. 569–574, 2005.

[48] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep Learning Approaches for Facial Emotion Recognition: A Case Study on FER-2013," Smart Innovation, Systems and Technologies, pp. 1–16, Oct. 2017.

for induction motor faults classification," Measurement, vol. 89, pp. 171–178, 2016.

[36] Y. Wu and X. Ai, "Face detection in color images using AdaBoost algorithm based on skin color information," First International Workshop on Knowledge Discovery and Data Mining, 2008.

[37] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

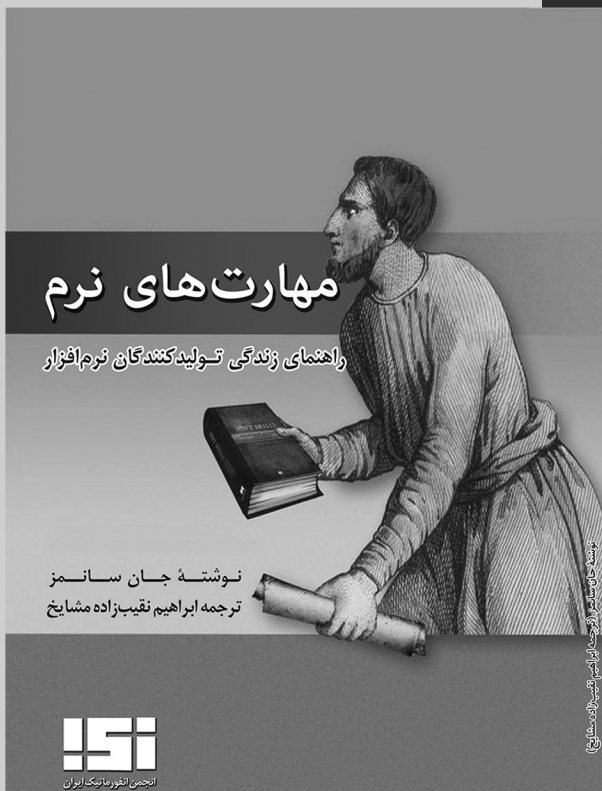
[38] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38–59, 1995.

[39] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Face recognition using active appearance models," Computer Vision — ECCV'98, pp. 581–595, 1998.

[40] D. Lundqvist, A. Flykt, and A. Ohman, "The Karolinska Directed Emotional Faces (KDEF)," CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998.

[41] Q. Li and Y. A. Kim, "Analysis of Facial Emotion Recognition Technology and Its Effectiveness in Human Interaction," Advances in Interdisciplinary Practice in Industrial Design, pp. 217–226, Jun. 2018.

[42] K. Yan, W. Zheng, T. Zhang, Y. Zong, C. Tang, C. Lu, and



جدیدترین کتاب از انتشارات انجمن انفورماتیک ایران

منتشر شد!

مهارت های نرم

برای تهیه کتاب با دفتر انجمن انفورماتیک ایران

تماس بگیرید ۶۶۴۱۲۸۶۱

چاپ اول