

## Autism diagnosis using child facial image and deep neural networks

Kambiz Moradi <sup>1</sup>, Mozafar Bag-Mohammadi <sup>1,\*</sup> and Mojtaba Karami <sup>1</sup>

<sup>1</sup> Faculty of Engineering, Ilam University, Ilam, Iran

E-mails: [k.moradi5@ilam.ac.ir](mailto:k.moradi5@ilam.ac.ir) ; [mozafarbm@yahoo.com](mailto:mozafarbm@yahoo.com) ; [m.karami@ilam.ac.ir](mailto:m.karami@ilam.ac.ir)

\* Corresponding Author

### Abstract

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder that affects physical, social, and language skills. There is no specific medicine to treat ASD. To diagnose this disorder, doctors consider the child's behavior and his developmental history. Autism is often difficult to diagnose because it can have a wide range of symptoms. Children with autism often have subtle facial features that distinguish them from normal children. Early diagnosis and appropriate medical intervention can significantly improve the lives of children with autism and save a lot of money. Deep neural networks (DNN) are a type of machine learning algorithm that can be used to automatically extract features from images as well as classify them. In this research, a practical solution for autism screening is proposed using images of Iranian children's faces through the use of a deep neural network based on transfer learning with 97% detection accuracy. The proposed method can also be used to improve the accuracy of clinical diagnosis.

### Keywords

Autism spectrum disorder, deep learning, artificial intelligence, ASD, CNN.

## تشخیص اوتیسم از روی تصویر صورت کودک با استفاده از شبکه‌های عصبی عمیق

کامبیز مرادی<sup>۱</sup>، مظفر بگ محمدی<sup>۱\*</sup> و مجتبی کرمی<sup>۱</sup>

<sup>۱</sup> گروه مهندسی کامپیوتر، دانشگاه ایلام، ایلام، ایران

ایمیل نویسندگان: [k.moradi5@ilam.ac.ir](mailto:k.moradi5@ilam.ac.ir); [mozafarbm@yahoo.com](mailto:mozafarbm@yahoo.com); [m.karami@ilam.ac.ir](mailto:m.karami@ilam.ac.ir)

### چکیده

اختلال طیف اوتیسم یک اختلال رشد عصبی پیچیده است که مهارت‌های بدنی، اجتماعی و زبانی را تحت تأثیر قرار می‌دهد. هیچ داروی خاصی برای درمان این بیماری وجود ندارد. برای تشخیص این اختلال، پزشکان رفتار و تاریخچه رشد کودک را در نظر می‌گیرند. اوتیسم اغلب به‌سختی قابل تشخیص است، زیرا می‌تواند با طیف گسترده‌ای از علائم ظاهر شود. کودکان مبتلا به اوتیسم اغلب دارای ویژگی‌های ظریفی در صورت هستند که آن‌ها را از کودکان معمولی در حال رشد متمایز می‌نماید. تشخیص زودهنگام و مداخله پزشکی مناسب می‌تواند به‌طور چشمگیری زندگی کودکان مبتلا به اوتیسم را بهبود بخشد و صرفه‌جویی زیادی از لحاظ هزینه نیز داشته باشد. شبکه‌های عصبی عمیق نوعی الگوریتم یادگیری ماشینی هستند که می‌توان از آن‌ها برای استخراج ویژگی‌ها از تصاویر و همچنین دسته‌بندی آن‌ها استفاده شوند. در این تحقیق یک راه‌حل عملی غربالگری اوتیسم با استفاده از تصاویر چهره کودکان ایرانی از طریق استفاده از یادگیری عمیق مبتنی بر یادگیری انتقالی با دقت تشخیص ۹۷٪ پیشنهاد می‌شود. همچنین می‌توان از روش پیشنهادی برای بهبود دقت تشخیص بالینی استفاده نمود.

**کلمات کلیدی:** اختلال طیف اوتیسم، یادگیری عمیق، هوش مصنوعی، CNN، ASD

نام نویسنده مسئول: مظفر بگ محمدی

ایمیل نویسنده مسئول: [mozafarbm@yahoo.com](mailto:mozafarbm@yahoo.com)

تاریخ ارسال مقاله: ۱۴۰۳/۰۷/۲۰

تاریخ(های) اصلاح مقاله: ۱۴۰۳/۰۸/۲۲

تاریخ پذیرش مقاله: ۱۴۰۳/۰۹/۰۳

### ۱- مقدمه

اوتیسم (درخودماندگی) یا اختلال طیف اوتیسم (ASD<sup>۱</sup>) اختلالی عصبی که مهارت‌های بدنی، اجتماعی و زبانی را تحت تأثیر قرار می‌دهد. همان‌طور که از اصطلاح «طیف» برمی‌آید، طیف وسیعی از علائم و سطوح شدت دارد [۱]. این اختلالات در حال حاضر در طبقه‌بندی آماری بین‌المللی بیماری‌ها و مشکلات مربوط به سلامتی به‌عنوان اختلالات روانی و رفتاری، در دسته اختلالات رشدی فراگیر قرار می‌گیرند [۲].

نشانه‌های اوتیسم معمولاً پیش از ۲.۵ سالگی بروز می‌یابد. افراد مبتلا نسبت به محبت بی‌اعتنا یا بیزار به نظر می‌رسند. آن‌ها ممکن است وابستگی بی‌موردی به بعضی از اشیاء پیدا کنند. فردی که از اختلال طیف اوتیسم رنج می‌برد به‌طور کلی قادر به برقراری تعامل اجتماعی و ارتباط با افراد دیگر نیست [۳]. زندگی یک فرد مبتلا معمولاً در تمام طول عمر او تحت تأثیر قرار می‌گیرد. علائم این مشکل ممکن است تا آخر عمر ادامه داشته باشد. درمان کامل افراد مبتلا به این بیماری امکان‌پذیر نیست [۴]. اگرچه اوتیسم در هر سنی قابل تشخیص است، اما علائم آن به‌طور کلی در دو سال اول زندگی ظاهر و در طول زمان ایجاد می‌شود. نرخ مبتلایان به اوتیسم در سراسر جهان متعدد است و با سرعت بسیار بالایی در حال افزایش است. بر اساس مطالعه‌ی که زیدان و همکاران [۵] انجام داده‌اند، تخمین زده می‌شود که در سراسر جهان از هر ۱۰۰ کودک یک نفر مبتلا به اوتیسم است. در ایران نیز آمار با میانگین جهانی برابر است.

فرآیند تشخیص اوتیسم شامل یک ارزیابی جامع از اندازه‌گیری‌های بالینی و مشاهدات متنوع پزشک از رفتار کودک در سن سه سال یا بالاتر از آن است [۶]. به‌عنوان بخشی از تشخیص بالینی از روش‌های تصویربرداری

پرتومغناطیسی مغز، نوار مغز، آزمایش‌های ژنتیک، مطالعات رفتاری و معاینه فیزیکی صورت استفاده می‌شود. فرآیندهای تشخیص استاندارد اوتیسم شامل ارزیابی ضریب هوشی IQ<sup>۲</sup>، بهره هوشی کلامی با استفاده از مقیاس رتبه‌بندی اوتیسم در دوران کودکی (CARS<sup>۳</sup>)، ابزار تشخیص اوتیسم - رتبه‌بندی ADI<sup>۴</sup> (R) و مقیاس هوش اختصاری وکسلر (WAIS<sup>۵</sup>) است. اخیراً برای تسریع روند تشخیص اوتیسم، کارهای تحقیقاتی متعددی انجام شده است.

نتایج تحقیقات نشان می‌دهند که اوتیسم به دلیل نقص در رشد جنینی مغز ایجاد می‌شود و این نقص در شکل‌گیری فیزیکی صورت افراد مبتلا به اوتیسم قابل مشاهده است [۷]. با توجه به اینکه اوتیسم یک اختلال رشد عصبی است و رشد عصبی با رشد مغز مرتبط است. رشد مغز را می‌توان در نشانه‌های چهره افراد اوتیسم در مقایسه با افراد غیر اوتیسم مشاهده کرد. تغییرات مغز در چهره منعکس می‌شود [۸]. افراد مبتلا به اوتیسم در نقاط و مختصات نشانه‌های چهره تفاوت‌هایی با افراد غیرمبتلا دارند [۹]. یافتن تفاوت‌های نشانه‌های چهره در افراد اوتیسم و ناتوانی‌های رشدی مختلف، به‌عنوان نشانگر زیستی در تشخیص اوتیسم، کمک می‌کند. دانشمندان دانشگاه میسوری در مطالعه‌ی [۱۰] با استفاده از ویژگی‌های صورت کودکان غیر مبتلا و مبتلا را موردبررسی قرار دادند و دریافتند که کودکان مبتلا به اوتیسم در مقایسه با کودکان غیر مبتلا دارای ویژگی مشترکی در صورت همچون پهن بودن غیرمعمول صورت فوقانی، چشم‌های پهن و ناحیه میانی کوتاه‌تر شامل گونه‌ها و بینی هستند.

DSM-5 یک راهنمای تشخیصی و آماری برای اختلال‌های روانی است. تمام روشهای تشخیصی برای اختلال‌های روانی باید الزامات DSM-5 را لحاظ کرده باشند. در روش پیشنهادی، ما نظر پزشک در خصوص ابتلا به این اختلال را ملاک قرار داده‌ایم. در این تحقیق ما یک مجموعه داده از تصاویر کودکان مبتلا و غیرمبتلا در ایران را گردآوری کرده‌ایم. سپس، با استفاده از شبکه‌های

<sup>4</sup> Autism Diagnosis Instrument—Rating

<sup>5</sup> Weschler Abbreviated Intelligence Scale

<sup>1</sup> Autism Spectrum Disorder

<sup>2</sup> Intellectual Quotient

<sup>3</sup> Childhood Autism Rating Scale

رشد معمولی (برگرفته از چندین مهدکودک و مدرسه ابتدایی در چین) مربوط به کودکان دو تا ۱۲ ساله و از یک نژاد است. مدل آن‌ها دارای دقت طبقه‌بندی ۹۵٪ و F1-Score برابر ۹۵٪ است. آن‌ها همچنین به این نتیجه رسیدند که برای راه‌حل‌های یادگیری عمیق مبتنی بر تصویر صورت، مجموعه داده‌های مخصوص نژاد باید برای توسعه مدل ساخته شود تا خطاهای طبقه‌بندی به دلیل تفاوت‌های آنترپومتریکی<sup>۶</sup> بین نژادها حذف شود. لذا، اطلاعات نژادی باید به‌عنوان پیش‌نیاز استفاده از یک مدل قابل اجرا در تشخیص یا غربالگری اوتیسم شناخته شود.

بیری و همکاران در [۱۳] یک مدل Mobilenet برای استخراج ویژگی و دسته‌بندی تصاویر برای طبقه‌بندی کودکان به‌عنوان سالم یا بالقوه اوتیسم با دقت ۹۴.۶٪ با استفاده از یادگیری عمیق معرفی کردند. آن‌ها برای تحقیق خود، از مجموعه داده‌ای که در Kaggle[14] قرار دارد و شامل بیش از ۳۰۰۰ تصویر از کودکان اوتیسم و غیر اوتیسم است استفاده کردند. مدل آن‌ها با استفاده از ۳۰۱۴ تصویر آموزش‌دیده و آزمایش‌شده است که به‌طور مساوی بین کودکان مبتلا به اوتیسم و کودکان بدون اوتیسم تقسیم‌شده است. درنهایت به‌دقت بالای ۹۴.۶۴ درصد دست یافتند. به این معنی که مدل در ۹۵٪ مواقع توانسته است کودکان مبتلا به اوتیسم یا عادی را به‌درستی شناسایی کند.

ساعد و همکاران [۱۵] یک مدل یادگیری عمیق مبتنی بر یادگیری انتقالی را برای تشخیص اوتیسم با استفاده از ویژگی‌های صورت در کودکان اوتیسمی و عادی بر روی مجموعه داده عمومی [۱۴] پیشنهاد کردند. مدل‌های یادگیری عمیق این مقاله بر اساس سه مدل از پیش آموزش‌دیده VGG19، Nasnetmobile و Xception برای تشخیص اوتیسم با استفاده از تصاویر ویژگی‌های صورت است. آن‌ها از معیارهای ارزیابی استاندارد مانند دقت، ویژگی و حساسیت برای ارزیابی نتایج سه مدل یادگیری عمیق استفاده کردند. مدل Xception بالاترین نتیجه دقت ۹۱٪ را داشت. پس از آن VGG19 و Nasnetmobile به ترتیب ۸۰٪ و ۷۸٪ دقت به دست آوردند.

علم و همکاران در سال ۲۰۲۲ [۱۶] یک مطالعه باهدف یافتن بهترین مدل یادگیری انتقالی برای طبقه‌بندی اوتیسم با در نظر گرفتن پنج مدل رایج مبتنی بر CNN برای تنظیم فرآیندها و بهینه‌سازیها انجام دادند. آن‌ها مدل‌های VGG19، Efficientnetb0، Xception، Mobilenetv2 و Resnet50v2 را بر روی مجموعه داده عمومی موجود در [۱۴] آزمایش کردند. مدل اصلاح‌شده Xception بهترین عملکرد را با دقت ۹۵٪ نشان داد. در تحقیق انجام‌شده توسط احمد و همکاران [۱۷] از CNN، ویژگی‌های صورت و یادگیری انتقالی برای شناسایی کودکان اوتیسمی استفاده شد. محققان سه مدل انتقالی (Mobilenet، Xception و Inceptionv3) را روی مجموعه داده عمومی [۱۴] اعمال کردند و برای Mobilenet به دقت ۹۵٪ رسیدند.

رحمان و همکاران [۱۸] مطالعه‌ای باهدف ایجاد یک مدل مبتنی بر CNN که بتواند اوتیسم را با استفاده از تصاویر صورت کودکان پیش‌بینی کند، انجام دادند. از مدل Mobilenet به‌عنوان یک مدل پایه برای مقایسه عملکرد دو مدل دیگر یعنی Xception و Efficientnets استفاده کردند. در این مطالعه به کاربرد ویژگی‌های صورت به‌عنوان نشانگر زیستی برای تشخیص دقیق کودکان اوتیسم از کودکان غیر اوتیسم پرداخته و از مجموعه داده عمومی [۱۴] استفاده کردند. آن‌ها پنج مدل طبقه‌بندی کننده اوتیسم مبتنی بر CNN ساختند و از یک طبقه‌بندی کننده DNN برای پیش‌بینی استفاده کردند. آن‌ها به این نتیجه رسیدند که مدل Xception عملکرد بهتری نسبت به سایر مدل‌ها دارد و به AUC ۹۶.۶۳٪ حساسیت ۸۸.۴۶٪ و NPV ۸۸٪ دست‌یافتند. آن‌ها نرخ تشخیص

عصبی کانولوشن<sup>۶</sup> (CNN) مبتنی بر یادگیری انتقالی، داده‌ها را با یکی از معماری‌های طراحی‌شده دوباره آموزش دادیم. نتایج شبیه‌سازی حاکی از دقت ۹۷٪ مدل در تشخیص کودکان مبتلا به اوتیسم است. می‌توان از این مدل برای تشخیص زودهنگام اوتیسم و کاهش هزینه‌های مربوطه استفاده نمود. در ادامه‌ی این کار، قصد داریم تا با راه‌اندازی یک وب‌سایت به عموم مردم اجازه دهیم تا با آپلود تصاویر کودک خود از مدل استفاده نمایند. در ادامه‌ی مقاله، ابتدا کارهای مرتبط معرفی خواهند شد. بخش سوم به چالش‌ها اختصاص دارد. در بخش چهارم ادبیات مورد نیاز توضیح داده می‌شود. روش پیشنهادی در بخش پنجم توضیح داده می‌شود. بخش ششم به نتایج شبیه‌سازی و بخش آخر نیز به نتیجه‌گیری اختصاص دارد.

در نهایت، این مطالعه پتانسیل استفاده از شبکه‌های عصبی عمیق و یادگیری انتقالی برای تشخیص زودهنگام اوتیسم در کودکان ایرانی را نشان می‌دهد، که می‌تواند به مداخلات مؤثرتر و نتایج بهتر برای کودکان مبتلا و تحقیقاتی گسترده‌تر در این زمینه منجر شود.

## ۲- پیشینه تحقیق

در زمینه تشخیص اوتیسم با استفاده از هوش مصنوعی تحقیقات متعددی انجام و از روش‌های مختلفی استفاده شده است که در اینجا چند مورد از روش‌هایی که مبتنی بر شبکه‌های عصبی و یادگیری عمیق هستند را مورد بررسی قرار می‌دهیم. علاوه بر این موارد، از هوش مصنوعی در روش‌های تصویربرداری پرتو مغناطیسی و روش‌های مبتنی بر تحلیل رفتار اشخاص نیز برای تشخیص اوتیسم استفاده می‌گردد.

پروینا و ماتا [۱۱] روشی برای تشخیص اوتیسم از روی تصاویر صورت با استفاده از شبکه‌های عصبی مصنوعی باهدف ایجاد یک متدولوژی تشخیص برای اوتیسم ارائه کردند. برای تجزیه و تحلیل نشانه‌های چهره افراد اوتیسمی، مراحل مختلفی مانند تشخیص چهره از تصاویر، تشخیص نشانه چهره، تبدیل نشانه‌ها به‌عنوان ویژگی و استخراج ویژگی را انجام دادند. آن‌ها برای کودکان بین ۱ تا ۱۰ سال، مجموعه داده‌ای خود را از ۵۰ تصویر از کودکان اوتیسم (جمع‌آوری‌شده از یک گروه Facebook به نام Autism Parenting Hub) و ۵۰ تصویر از کودکان معمولی در حال رشد (جمع‌آوری‌شده از منابع آنلاین) جمع‌آوری کرده بودند. الگوریتم‌های تشخیص نشانه‌های چهره مانند CNN، ماشین محدود بولتزمن (RBM) و رمزگذارهای خودکار RBM برای شناسایی نقاط کلیدی در چهره و مقایسه یا طبقه‌بندی چهره‌ها استفاده شدند. بدین ترتیب ۶۸ علامت مشخصه برای پیش‌بینی چهره به دست آوردند. برای استخراج و انتخاب ویژگی‌های موردنیاز و مهم و مرتبط با اوتیسم برای طبقه‌بندی و تشخیص اوتیسم از دو روش استفاده کردند. در روش اول ۸ ویژگی را با محاسبه میانگین ارزش نقطه عطف نقاط یک منطقه خاص و در روش دیگر ۱۵ ویژگی را با یافتن فاصله اقلیدسی بین نقاط شاخص انتخاب کردند. درنهایت با استفاده از نشانه‌های چهره افراد مورد آزمایش و آموزش یک مدل مناسب توانستند اوتیسم را با دقت ۸۰ درصد تشخیص دهند.

لو و پرکوفسکی [۱۲] یک راه‌حل غربالگری اوتیسم با استفاده از تصاویر چهره از روش یادگیری عمیق مبتنی بر VGG16 (به‌عنوان یک مدل از پیش آموزش‌دیده برای یادگیری انتقالی) ارائه کردند. مجموعه داده مورد استفاده، شامل حدود ۶۰۰ تصویر از صورت کودکان دارای اوتیسم با تشخیص بالینی برگرفته از مرکز توان‌بخشی اوتیسم الیم (که در کودکان مبتلا به اوتیسم تخصص دارد و مقر آن در شاندونگ، چین است) و ۵۶۱ تصویر از کودکان با

<sup>6</sup> Convolutional Neural Networks

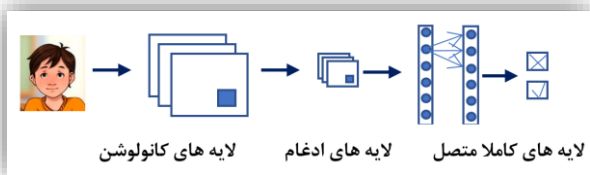
<sup>۷</sup> آنترپومتری یا انسان‌سنجی به اندازه‌گیری بخش‌های مختلف بدن انسان برای شناسایی تفاوت‌ها و دسته‌بندی‌های فیزیکی افراد گفته می‌شود

تمایزی مانند چشم‌های پهن تر و وسط صورت صاف تر دارند [۱۰]. با این حال، این مطالعات بر اندازه‌گیری‌های دستی و ارزیابی‌های ذهنی تکیه داشتند که می‌تواند زمان‌بر و مستعد خطا باشد. فرآیند تشخیص تصویر شامل شناسایی و طبقه‌بندی اشیاء یا الگوهای درون یک تصویر است. در سال‌های اخیر، پیشرفت‌ها در یادگیری ماشین و بینایی کامپیوتر، امکان خودکار سازی فرآیند تجزیه و تحلیل ویژگی‌های صورت را فراهم کرده است. CNN نوعی الگوریتم یادگیری عمیق هستند که با دستیابی به عملکرد پیشرفته در کارهای مختلف می‌توانند الگوهای موجود در داده‌ها از جمله تصاویر را تشخیص دهند. توانایی آن‌ها در یادگیری خودکار ویژگی‌ها از داده‌های خام، آن‌ها را برای کارهای بصری پیچیده که در آن فنون بینایی تصویر سنتی کم می‌آورند، مناسب می‌سازد.

تشخیص بیماری از روی تصاویر یک کار چالش برانگیز است. در سال‌های اخیر، یادگیری عمیق و به ویژه CNN به عنوان یک ابزار قدرتمند برای تجزیه و تحلیل تصاویر ظاهر شده است. CNN ها نوعی از شبکه‌های عصبی هستند که برای کارهای پردازش تصویر مناسب هستند، زیرا می‌توانند یاد بگیرند که ویژگی‌هایی مناسب را از تصاویر استخراج کنند. CNN ها با وجود مزایای فراوانی که دارند، محدودیت‌هایی نیز دارند. یکی از این محدودیت‌ها هزینه محاسباتی بالا است که می‌تواند زمان آموزش و استنباط را برای مجموعه داده‌های بزرگ یا مدل‌های پیچیده بسیار طولانی کند. دوم، CNN ها به داده‌های برجسته دار زیادی نیاز دارند. این محدودیت برای تشخیص بیماری‌هایی مثل اوتیسم که تعداد داده‌های آن‌ها کم است یک چالش جدی است.

#### ۴- شبکه‌های عصبی کانولوشنال یا CNN

CNN ها [۲۶] نوعی شبکه عصبی مصنوعی هستند که نشان داده‌اند در حل وظایف پیچیده مانند طبقه‌بندی تصویر، از جمله طبقه‌بندی تصاویر دودویی، مؤثر هستند. CNN به طور خاص برای یادگیری و استخراج خودکار ویژگی‌های مرتبط از تصاویر طراحی شده و برای کارهایی مانند تمایز بین کلاس‌های مختلف مناسب است. معماری یک CNN از چندین لایه، از جمله لایه‌های کانولوشن، لایه‌های ادغام<sup>۸</sup> و لایه‌های کاملاً متصل تشکیل شده است. شکل ۱ ساختار یک شبکه عصبی کانولوشن را نمایش می‌دهد. این لایه‌ها به صورت متوالی کار می‌کنند تا به تدریج نمایش‌های سلسله مراتبی تصاویر ورودی را یاد بگیرند و ویژگی‌های آن‌ها را استخراج کنند.



شکل ۱ ساختار یک شبکه عصبی کانولوشن

لایه‌های کانولوشن بلوک‌های اصلی سازنده CNN ها هستند. آن‌ها از فیلترها یا هسته‌های متعددی تشکیل شده‌اند. کار هر فیلتر تشخیص الگوها یا ویژگی‌های خاص مانند لبه‌ها، گوشه‌ها یا بافت‌ها در تصویر ورودی است. با اعمال چندین فیلتر در هر لایه کانولوشن، CNN ها می‌توانند طیف وسیعی از ویژگی‌ها را در مقیاس‌های فضایی مختلف بیاموزند.

لایه‌های ادغام معمولاً بعد از لایه‌های کانولوشن درج می‌شوند تا ابعاد فضایی نقشه‌های ویژگی را کاهش دهند و در عین حال مهم‌ترین اطلاعات خود

هر مدل را با سطح اطمینان ۹۵٪ مقایسه کردند که Efficientb0 بالاترین امتیازها را دریافت کرد (۵۹.۳۳٪ برای کلاس اوتیسم و ۵۸.۶۷٪ برای کلاس غیر اوتیسم).

#### ۳- چالش‌ها

تشخیص اوتیسم به روش‌های کنونی دارای چالش‌های متعددی است که در ادامه به آن‌ها اشاره می‌گردد:

روش‌های تصویربرداری پرتو مغناطیسی مغز، نوار مغز و آزمایش‌های ژنتیک به دلیل تنوع و عدم تعریف روشن از ویژگی‌های آن برای تشخیص اوتیسم دارای محدودیت‌های مثل در دسترس نبودن، هزینه‌ی بالا، زمان طولانی و مشکل در تفسیر هستند. کودکان خردسال یا آن‌هایی که اضطراب قابل توجهی دارند نمی‌توانند در حین انجام روش تصویربرداری پرتو مغناطیسی ثابت بمانند. این موضوع منجر به کاهش کیفیت تصویر و زمان اسکن طولانی‌تر می‌شود. حتی ممکن است برای گرفتن تصاویر به داروی آرام‌بخش و یا بیهوشی نیاز داشته باشند، که خود دارای عوارض جانبی و خطرات بالقوه‌ای برای فرد مبتلا است.

از طرف دیگر، تشخیص اوتیسم در کودکان از طریق پرسش‌نامه یا چالش‌های متعددی از جمله اعتبار محدود، عدم شفافیت گزینه‌ها، پایایی و روایی پایین پرسش‌نامه، تنوع پرسش‌نامه‌ها و سوگیری فرهنگی و زبانی بالقوه همراه است.

در حقیقت درمان افراد مبتلا به اوتیسم جزو پرهزینه‌ترین و سخت‌ترین درمان‌ها هستند [۱۹، ۲۰]. در یک مطالعه [۲۱] در مورد مداخلات رفتاری فشرده اولیه در هلند نشان داد که اگر مداخله زودهنگام قبل از ۳۰ ماهگی شروع شود، می‌تواند تا بیش از یک میلیون یورو برای هر فرد صرفه‌جویی در هزینه‌های درمان در طول عمر بیمار داشته باشد.

تأثیر این بیماری بر بنیاد خانواده‌ها و کیفیت زندگی آن‌ها بسیار زیاد و گاه ویرانگر است [۲۲، ۲۳].

پیش‌آگهی برای افراد مبتلا به اوتیسم از لحاظ تاریخی ضعیف بوده است [۲۴، ۲۵]. در اغلب موارد، افراد به‌اشتباه این بیماری را جزء مشکلات فرهنگی و رفتاری طبقه‌بندی می‌کنند.

در روش پیشنهادی مشکلاتی مثل در دسترس نبودن، هزینه‌ی بالا، زمان طولانی تشخیص، پایایی و روایی پایین پرسش‌نامه، سوگیری فرهنگی و زبانی و تفسیر نامناسب وجود ندارند. به علاوه، تشخیص احتمال ابتلا به آسانی و به سرعت انجام می‌گردد. یافته‌ها نشان می‌دهند اگر تشخیص اوتیسم در سنین پایین صورت گیرد، می‌تواند سازوکار درمان را بهبود بخشد و منجر به کاهش قابل توجه هزینه‌ها به خصوص در جوامع توسعه‌نیافته گردد [۲۱].

#### ۳-۱- راهکار

برای تشخیص زودهنگام کودکان مبتلا به اوتیسم، وجود یک راه‌حل غربالگری عینی، ارزان با دسترسی آسان و قابل درک در هر خانواده ضروری است. تشخیص اوتیسم به زمان و هزینه قابل توجهی نیاز دارد. تشخیص زودهنگام می‌تواند از وخامت بیشتر وضعیت بیمار جلوگیری و به کاهش هزینه‌های طولانی‌مدت ناشی از تأخیر در تشخیص کمک کند. بنابراین به یک ابزار آزمون غربالگری کارآمد، دقیق و آسان نیاز داریم که بتواند ویژگی‌های اوتیسم در یک فرد را پیش‌بینی و انجام ارزیابی جامع اوتیسم را پیشنهاد کند. ایده استفاده از ویژگی‌های صورت برای تشخیص اوتیسم جدید نیست. مطالعات قبلی نشان داده‌اند که افراد مبتلا به اوتیسم ویژگی‌های صورت

را در هر کانال ورودی اعمال می‌کند، درحالی‌که کانولوشن نقطه‌ای یک کانولوشن  $1 \times 1$  را در همه کانال‌ها انجام می‌دهد. با جدا کردن فیلتر فضایی از اختلاط کانال، Mobilenet به تعادل خوبی بین کارایی و دقت دست می‌یابد. معماری‌های Mobilenet توسط دو فرآیند تعریف می‌شوند: ضرب‌کننده عرض (A) و ضرب‌کننده وضوح (P). ضرب عرض تعداد کانال‌ها را در هر لایه کنترل می‌کند و آن را از مقدار پایه معماری اصلی کاهش می‌دهد. این پارامتر امکان معامله بین اندازه و دقت مدل را فراهم می‌کند. ضرب رزولوشن وضوح تصویر ورودی را کاهش می‌دهد و نیازهای محاسباتی را بیشتر کاهش می‌دهد.

معماری اصلی Mobilenet در مقایسه با سایر مدل‌های محبوب مانند VGG یا Resnet، عملکردی پیشرفته را در وظایف طبقه‌بندی Imagenet با پارامترهای کمتر و هزینه محاسباتی کمتر به دست آورد. همچنین عملکرد عالی در سایر وظایف بینایی مانند تشخیص اشیا و تقسیم‌بندی معنایی را نشان داد. Mobilenet از زمان معرفی خود به دلیل کارایی آن در دستگاه‌های تلفن همراه به‌طور گسترده در برنامه‌های مختلف مورد استفاده قرار گرفته است. Mobilenet به یک انتخاب استاندارد برای کارهای بینایی کامپیوتری در زمان واقعی<sup>۹</sup> در تلفن‌های هوشمند تبدیل شده است و برنامه‌هایی مانند تشخیص چهره، واقعیت افزوده و طبقه‌بندی تصویر روی دستگاه را بدون تکیه بر پردازش مبتنی بر ابر فعال می‌کند.

Inceptionv3 [۲۹] یک معماری شبکه عصبی عمیق است که محبوبیت قابل توجهی در زمینه بینایی کامپیوتر به دست آورده است. این معماری توسط محققان گوگل توسعه یافته است و به دلیل عملکرد استثنایی خود در وظایف طبقه‌بندی تصاویر شناخته شده است. Inceptionv3 بر اساس موفقیت نسخه قبلی خود، Inceptionv1، با معرفی چندین بهبود و بهینه‌سازی استوار است. معماری Inceptionv3 بر اساس مفهوم "Inception Modules" است. این ماژول‌ها برای ثبت اطلاعات در مقیاس‌ها و وضوح‌های مختلف در یک تصویر طراحی شده‌اند. آن‌ها از چندین لایه کانولوشن موازی با اندازه‌های مختلف فیلتر تشکیل شده‌اند که سپس به هم متصل می‌شوند تا یک نقشه ویژگی غنی را تشکیل دهند. این موضوع به شبکه اجازه می‌دهد تا به‌طور مؤثر اطلاعات محلی و جهانی را از یک تصویر ورودی بگیرد.

یکی از نوآوری‌های کلیدی در Inceptionv3 استفاده از لایه‌های "Bottleneck" است. این لایه‌های گلوگاه، ابعاد نقشه‌های ویژگی را قبل از اعمال پیچیدگی‌های گران‌قیمت کاهش می‌دهند و در نتیجه پیچیدگی محاسباتی را بدون از دست رفتن دقت کاهش می‌دهند. این تکنیک به آموزش شبکه‌های عمیق‌تر با منابع محاسباتی محدود کمک می‌کند. یکی دیگر از جنبه‌های مهم Inceptionv3، ادغام طبقه‌بندی کننده‌های کمکی است. این طبقه‌بندی در مراحل میانی شبکه قرار می‌گیرند و به مبارزه با مشکل ناپدید شدن گرادیان در طول آموزش کمک می‌کنند. آن‌ها با ارائه سیگنال‌های نظارتی اضافی، شبکه را تشویق می‌کنند تا ویژگی‌های متمایز بیشتری را بیاموزد.

Inceptionv3 همچنین تکنیکی به نام "Factorized 7x7 Convolutions" را معرفی می‌کند. به جای استفاده از یک لایه کانولوشن  $7 \times 7$ ، آن را به دو کانولوشن  $1 \times 7$  و  $7 \times 1$  متوالی تجزیه می‌کند. این فاکتورسازی هزینه محاسباتی را کاهش می‌دهد و درعین حال قدرت بینایی را حفظ می‌کند. این معماری همچنین شامل بهینه‌سازی‌های دیگری مانند نرمال‌سازی دسته‌ای است که به همگرایی سریع‌تر و تکنیک‌های منظم‌سازی تهاجمی مانند حذف تصادفی (Dropout) برای جلوگیری از بیش‌برازش در طول آموزش کمک می‌کند.

با توجه به نتایج موجود در مقاله تحقیقاتی در Inceptionv3 با عنوان "بازاندیشی در معماری اولیه برای بینایی کامپیوتر" [۲۹]، این معماری به

را حفظ کنند. حداکثر ادغام رایج‌ترین نوع عملیات ادغام است که حداکثر مقدار را در یک منطقه کوچک (مثلاً  $3 \times 3$ ) انتخاب می‌کند و بقیه را دور می‌ریزد. این عملیات نمونه‌برداری، به کاهش پیچیدگی محاسباتی کمک می‌کند و ویژگی‌های آموخته‌شده را تا تغییرات فضایی کوچک قوی‌تر می‌کند. لایه‌های کاملاً متصل معمولاً در انتهای شبکه اضافه می‌شوند تا بر اساس ویژگی‌های استخراج‌شده، استدلال و تصمیم‌گیری در سطح بالا انجام شود. این لایه‌ها هر نورون را از یک لایه به هر نورون در لایه بعدی متصل می‌کنند و امکان یادگیری روابط غیرخطی پیچیده بین ویژگی‌ها را فراهم می‌کنند.

آموزش CNN شامل دو مرحله اصلی است: انتشار به جلو و پس انتشار. در انتشار روبه‌جلو، یک تصویر ورودی به‌صورت متوالی از هر لایه عبور داده می‌شود و هر لایه با استفاده از وزن‌ها و سوگیری‌های آموخته‌شده، ورودی خود را به یک نمایش سطح بالاتر تبدیل می‌کند. خروجی نهایی از آخرین لایه کاملاً متصل به دست می‌آید که نشان‌دهنده احتمالات کلاس پیش‌بینی شده است. در طول عملیات مرحله‌ی پس انتشار، از تفاوت بین برجسب‌های کلاس پیش‌بینی شده و واقعی برای محاسبه تلفات یا خطا استفاده می‌شود. سپس این خطا به عقب در شبکه منتشر می‌شود و وزن‌ها و سوگیری‌های هر لایه را با استفاده از الگوریتم‌های بهینه‌سازی مانند نزول گرادیان تنظیم می‌کند. این روند تکراری تا زمانی ادامه می‌یابد که شبکه به مجموعه‌ای از وزن‌های بهینه همگرا شود که تلفات کلی را به حداقل می‌رساند.

شبکه‌های CNN ابزار قدرتمندی برای وظایف طبقه‌بندی تصاویر دودویی هستند. آن‌ها می‌توانند به‌طور خودکار ویژگی‌های مرتبط را از تصاویر از طریق چندین لایه فیلتر کانولوشنال و عملیات ادغام بیاموزند و استخراج کنند. با آموزش بر روی مجموعه داده‌های بزرگ با انتشار پس‌زمینه، این شبکه‌ها می‌توانند به نتیجه‌ی دلخواه برسند. یکی از کارهای اساسی در CNN‌های عمیق، معماری Alexnet است که توسط کریژفسکی و همکاران [۲۷] پیشنهاد شده است. در سال ۲۰۱۲، Alexnet با برنده شدن در چالش تشخیص تصویری در مقیاس بزرگ (ILSVRC) با یک حاشیه بزرگ به پیشرفت قابل توجهی در طبقه‌بندی تصاویر دست یافت. از آن زمان، پیشرفت‌های زیادی در CNN‌های عمیق برای طبقه‌بندی تصاویر دودویی انجام شده است. محققان معماری‌های مختلفی مانند Vggnet، Googlenet (Inception)، Resnet و Densenet را پیشنهاد کرده‌اند که عملکرد پیشرفته‌ای را در مجموعه داده‌های معیار مانند Imagenet به دست آورده‌اند. در ادامه برخی از معماری‌های رایج CNN معرفی می‌شوند.

Mobilenet [۲۸] یک معماری شبکه عصبی عمیق است که به‌طور خاص برای برنامه‌های موبایل و بینایی کامپیوتری طراحی شده است. این معماری توسط محققان گوگل در سال ۲۰۱۷ معرفی شد و هدف آن ارائه مدل‌های کارآمد و سبک‌وزن است که می‌توانند بر روی دستگاه‌های دارای محدودیت منابع مانند گوشی‌های هوشمند اجرا شوند. این معماری به دلیل توانایی آن در دستیابی به دقت بالا با حفظ نیازهای محاسباتی پایین، محبوبیت قابل توجهی به دست آورده است.

ایده کلیدی پشت Mobilenet استفاده از کانولوشن‌های قابل تفکیک عمیق است که تعداد پارامترها و محاسبات را در مقایسه با لایه‌های کانولوشن سنتی به میزان قابل توجهی کاهش می‌دهد. با توجه به مقاله [۲۸]، این تکنیک هم‌ردپای حافظه و هم‌هزینه محاسباتی شبکه را کاهش می‌دهد و درعین حال دقت را حفظ می‌کند.

کانولوشن‌های قابل تفکیک عمیق از دو عملیات مجزا تشکیل شده‌اند: یک کانولوشن عمقی و به دنبال آن یک کانولوشن نقطه‌ای. کانولوشن عمقی فیلتر

<sup>9</sup> Real time

آن را برای کارهای مختلف آسان می‌کند. این انعطاف‌پذیری در تحقیقات یادگیری عمیق بسیار مهم است، جایی که معماری‌ها اغلب نیاز به سفارشی‌سازی یا گسترش برای مطابقت با نیازهای خاص دارند. پس Xception یک معماری برای شبکه‌های عصبی عمیق است که پیچیدگی‌های قابل تفکیک عمیق را به‌عنوان راهی برای بهبود کارایی و عملکرد معرفی می‌کند. ساختار مدولار، قابلیت‌های یادگیری بازنمایی بهبود یافته و عملکرد پیشرفته آن را به ابزاری ارزشمند در زمینه بینایی کامپیوتر تبدیل کرده است.

#### ۴-۱- بهبود کارایی آموزش در CNN

بهبود کارایی آموزش در CNN به معنای افزایش دقت، سرعت و پایداری مدل در فرآیند یادگیری است. این بهبود می‌تواند از طریق روش‌ها و تکنیک‌های مختلفی حاصل شود که به مدل کمک می‌کنند تا بهتر و سریع‌تر به نتایج مطلوب برسد. برخی از این تکنیک‌ها عبارتند از:

- **تقویت داده:** با ایجاد تغییرات کوچک در تصاویر موجود، می‌توان مجموعه داده‌های آموزشی را بزرگتر و متنوع‌تر کرد که به جلوگیری از بیش‌برازش کمک می‌کند [۳۱].
  - **نرمال‌سازی دسته‌ای:** این تکنیک باعث می‌شود که شبکه عصبی سریع‌تر و پایدارتر آموزش ببیند. نرمال‌سازی دسته‌ای با تنظیم میانگین و واریانس خروجی لایه‌ها، به بهبود همگرایی کمک می‌کند [۳۱].
  - **استفاده از لایه‌های Dropout:** این لایه‌ها به صورت تصادفی برخی از نورون‌ها را در طول آموزش غیرفعال می‌کنند تا از وابستگی بیش از حد شبکه به نورون‌های خاص جلوگیری شود و به این ترتیب، شبکه به طور کلی قوی‌تر و مقاوم‌تر می‌شود [۳۱].
  - **تنظیمات بهینه‌سازی:** استفاده از الگوریتم‌های بهینه‌سازی پیشرفته مانند Adam، Rmsprop و تنظیم مناسب نرخ یادگیری می‌تواند به بهبود کارایی آموزش کمک کند [۳۲].
  - **توقف زودهنگام:** یک تکنیک منظم‌سازی<sup>۱۰</sup> است که برای جلوگیری از بیش‌برازش در CNN استفاده می‌شود و به این صورت عمل می‌کند که در طول فرآیند آموزش، عملکرد مدل روی مجموعه داده‌های اعتبارسنجی را نظارت می‌کند و زمانی که عملکرد مدل روی این مجموعه داده‌ها بهبود نمی‌یابد یا شروع به بدتر شدن می‌کند، آموزش را متوقف می‌کند [۳۳].
  - **یادگیری انتقالی<sup>۱۱</sup>:** استفاده از مدل‌های پیش‌آموزش دیده<sup>۱۲</sup> و تنظیم آن‌ها برای وظایف خاص می‌تواند زمان آموزش را کاهش داده و دقت مدل را افزایش دهد [۳۴].
- در ادامه در مورد یادگیری انتقالی و توقف زودهنگام بیشتر بحث خواهد خواهد شد.

#### ۴-۲- یادگیری انتقالی

یادگیری انتقالی [۳۵] تکنیکی در یادگیری عمیق است که به مدل‌ها اجازه می‌دهد از دانش کسب‌شده در یک کار برای بهبود عملکرد در یک کار دیگر استفاده کنند. در هنگام استفاده از CNN برای طبقه‌بندی تصاویر دودویی، این روش می‌تواند زمانی که داده‌های برجسب‌گذاری شده کمی در دسترس است، بسیار مؤثر باشد [۳۴]. به‌جای آموزش مدل از ابتدا، از یک مدل از پیش آموزش‌دیده (که قبلاً روی مجموعه داده بزرگی آموزش دیده است) به‌عنوان

عملکرد پیشرفته‌ای در مجموعه داده‌های معیار مختلف مانند Imagenet دست یافت و از نظر دقت و کارایی محاسباتی از مدل‌های قبلی پیشی گرفت. این مقاله همچنین بینش‌های دقیقی در مورد انتخاب‌های طراحی و نتایج تجربی ارائه می‌دهد که از اثربخشی Inceptionv3 پشتیبانی می‌کنند. Inceptionv3 به‌طور گسترده در کاربردهای مختلف از جمله طبقه‌بندی تصویر، تشخیص اشیا و تقسیم‌بندی تصویر مورد استفاده قرار گرفته است. تطبیق‌پذیری و عملکرد بالا آن را به انتخابی محبوب برای محققان و متخصصان در زمینه بینایی کامپیوتر تبدیل کرده است. به‌طور کلی، Inceptionv3 یک معماری شبکه عصبی عمیق است که از ماژول‌های اولیه، لایه‌های Bottleneck، طبقه‌بندی‌کننده‌ی کمکی و سایر بهینه‌سازی‌ها برای دستیابی به عملکرد پیشرفته در وظایف طبقه‌بندی تصویر استفاده می‌کند. انتخاب‌های طراحی آن توسط تحقیقات و آزمایش‌های دقیق پشتیبانی می‌شود و آن را به یک انتخاب قابل اعتماد برای برنامه‌های بینایی کامپیوتر تبدیل می‌کند.

Xception (مخفف «Extreme Inception») یک معماری برای شبکه‌های عصبی عمیق است که توسط فرانسوا شولت، خالق کتابخانه یادگیری عمیق Keras پیشنهاد شده است. نتایج این کار در یک مقاله تحقیقاتی با عنوان "Xception: یادگیری عمیق با پیچیدگی‌های قابل جداسازی عمیق" در سال ۲۰۱۷ منتشر شده است [۳۰]. هدف معماری Xception بهبود کارایی و عملکرد CNN با معرفی یک رویکرد جدید به نام پیش‌های قابل جداسازی عمیق است [۳۰]. در CNN های سنتی، هر لایه کانولوشن هم پیچیدگی فضایی و هم پیچیدگی کانال را انجام می‌دهد. این بدان معنی است که فیلترهای مورد استفاده در عملیات کانولوشن برای همه کانال‌های ورودی اعمال می‌شود. با این حال، Xception این رویکرد سنتی را با پیچیدگی‌های قابل تفکیک عمیق جایگزین می‌کند، که این دو عملیات را به لایه‌های جداگانه تقسیم می‌کند.

اولین گام در پیچیدگی‌های قابل تفکیک عمیق، پیچیدگی عمقی است، که در آن هر کانال ورودی با مجموعه‌ای از فیلترها به‌طور مستقل در هم پیچیده می‌شود. این عملیات با استفاده از پارامترهای کمتر در مقایسه با پیچیدگی‌های سنتی، اطلاعات مکانی را ضبط می‌کند و در عین حال پیچیدگی محاسباتی را کاهش می‌دهد. مرحله دوم پیچیدگی نقطه‌ای است که در آن پیچیدگی‌های ۱×۱ برای ترکیب کانال‌های خروجی از مرحله قبل در نقشه ویژگی نهایی اعمال می‌شود. Xception با جدا کردن کانولوشن‌های فضایی و کانالی، تعداد پارامترها و محاسبات مورد نیاز را کاهش می‌دهد که منجر به بهبود کارایی و زمان‌های آموزشی سریع‌تر می‌شود. این طراحی معماری همچنین امکان یادگیری بازنمایی بهتر را فراهم می‌کند، زیرا استخراج ویژگی‌های دقیق‌تر را امکان‌پذیر می‌سازد.

در مقاله [۳۰] نشان داده شده است که معماری Xception به عملکرد پیشرفته‌ای در کارهای مختلف بینایی کامپیوتری مانند طبقه‌بندی تصویر و تشخیص اشیا دست می‌یابد. در واقع، در طول ارزیابی خود از مدل‌های پیشرفته قبلی در مجموعه داده Imagenet عملکرد بهتری داشت. نویسندگان همچنین نشان داده‌اند که Xception می‌تواند به‌عنوان یک استخراج‌کننده ویژگی برای انتقال وظایف یادگیری استفاده شود و تطبیق‌پذیری آن فراتر از طبقه‌بندی تصاویر را نشان می‌دهد.

یکی از مزایای Xception توانایی آن در یادگیری ویژگی‌های انتزاعی بیشتر به دلیل معماری عمیق‌تر آن در مقایسه با مدل‌های قبلی مانند Inception است. پیچیدگی‌های قابل تفکیک عمیق امکان استفاده کارآمدتر از ظرفیت مدل را فراهم می‌کند و شبکه را قادر می‌سازد تا الگوها و روابط پیچیده را در داده‌ها ثبت کند. علاوه بر این، Xception یک ساختار ماژولار دارد که سازگاری و اصلاح

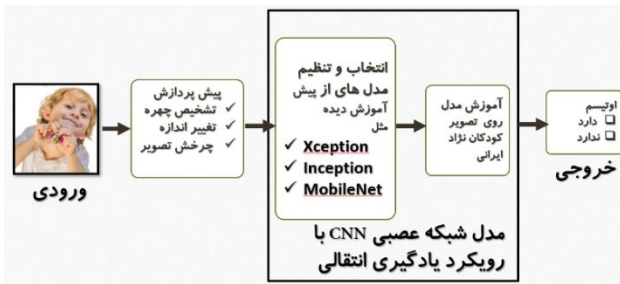
<sup>12</sup> Pre-trained Models

<sup>10</sup> Regularization

<sup>11</sup> Transfer Learning

## ۵- روش پیشنهادی

در این مطالعه، یادگیری عمیق مبتنی بر یادگیری انتقالی را با استفاده از سه مدل (Xception، Mobilenet و Inception) برای تشخیص اوتیسم از طریق از ویژگی‌های صورت کودکان اوتیسم و عادی، پیشنهاد می‌کنیم. این مدل‌ها از ویژگی‌های قابل توجهی که از تصاویر چهره استخراج می‌کنند برای تعیین اینکه آیا کودک مبتلا به اوتیسم است یا نیست بهره می‌برند. یکی از مزایای استفاده از الگوریتم‌های یادگیری عمیق، توانایی استخراج جزئیات بسیار کوچک از یک تصویر است که شخص نمی‌تواند با چشم غیرمسلح متوجه آن شود. **شکل ۲ چارچوب کلی مطالعه** شامل جمع‌آوری داده‌ها، پیش‌پردازش و بارگذاری داده‌ها، آماده‌سازی و آموزش مدل را نشان می‌دهد.



شکل ۲ چارچوب کلی مطالعه

برای استفاده از CNN و اعمال یادگیری انتقال در طبقه‌بندی تصاویر دودویی، این مراحل را دنبال می‌کنیم:

۱. **انتخاب یک مدل CNN از پیش آموزش دیده:** چندین معماری محبوب CNN وجود دارد که روی مجموعه داده‌های بزرگ از قبل آموزش داده شده‌اند، مانند VGG16، Resnet50 و Inceptionv3. انتخاب مدل از پیش آموزش دیده به عواملی مانند منابع محاسباتی و الزامات خاص کار بستگی دارد.
۲. **حذف آخرین لایه کاملاً متصل:** آخرین لایه(های) مدل از پیش آموزش دیده، مسئول طبقه‌بندی تصاویر به دسته‌های خاص از مجموعه داده اصلی آن است. از آنجایی که ممکن است کار طبقه‌بندی تصاویر دودویی دارای کلاس‌ها یا برچسب‌های متفاوتی نسبت به موارد موجود در مجموعه داده اصلی باشد، این لایه(ها) را حذف کرده و با لایه‌های جدیدی جایگزین می‌کنیم که متناسب با وظیفه ما باشد.
۳. **منجمد کردن لایه‌های قبلی:** لایه‌های اولیه مدل از پیش آموزش دیده ویژگی‌های سطح پایینی را که عموماً برای بسیاری از کارها قابل اجرا هستند، نشان می‌دهند. برای جلوگیری از اصلاح این ویژگی‌های آموخته‌شده در طول تمرین، وزن این لایه‌ها را منجمد می‌کنیم. این به ما اجازه می‌دهد تا روی تنظیم دقیق لایه‌های بعدی که مخصوص کار ما هستند تمرکز کنیم.
۴. **افزودن لایه‌های کاملاً متصل جدید:** به بالای مدل از پیش آموزش دیده، لایه‌های کاملاً متصل جدیدی اضافه می‌کنیم که به‌طور خاص برای طبقه‌بندی تصاویر دودویی آموزش داده می‌شوند. این لایه‌های جدید یاد می‌گیرند که ویژگی‌های استخراج‌شده از مدل از پیش آموزش دیده را به کلاس‌های خروجی مورد نظر نگاشت کنند.
۵. **آموزش و تنظیم دقیق:** ما کل شبکه را آموزش می‌دهیم، هم لایه‌های جدید اضافه‌شده و هم لایه‌های بعدی مدل از پیش آموزش دیده که منجمد نشده‌اند. وزن‌های این لایه‌ها با استفاده از مجموعه داده برچسب‌گذاری‌شده مخصوص کار طبقه‌بندی تصاویر دودویی به‌روزرسانی می‌شوند. با دنبال کردن این مراحل، یادگیری انتقالی به ما امکان می‌دهد هم از دانش عمومی که توسط مدل از پیش آموزش دیده گرفته شده است و هم از دانش خاصی که از مجموعه داده‌های برچسب‌گذاری‌شده‌مان آموخته‌ایم، بهره

نقطه شروع استفاده می‌شود. این مدل از پیش آموزش دیده ویژگی‌های مفید را از تصاویر استخراج کرده و الگوهای بصری کلی مانند لبه‌ها، بافت‌ها و اشکال را درک کرده است.

ایده اصلی این است که ویژگی‌های سطح پایین آموخته‌شده توسط مدل از پیش آموزش دیده برای کارهای جدید نیز قابل استفاده هستند. این ویژگی‌ها به ما اجازه می‌دهند که مدل‌های دقیق‌تری با داده‌های آموزشی کمتر بسازیم. یادگیری انتقالی با استفاده از وزن‌های مدل اول به‌عنوان نقطه شروع و تنظیم دقیق پارامترها با داده‌های کار جدید، می‌تواند در زمان و منابع صرفه‌جویی کند و عملکرد مدل را بهبود بخشد. این رویکرد به‌ویژه وقتی داده‌های کمی برای کار جدید در دسترس است، بسیار مفید است.

## ۳-۴- توقف زود هنگام

توقف زود هنگام در CNN تکنیکی است که برای بهبود کارایی آموزش استفاده می‌شود. با توقف زود هنگام فرآیند آموزش در صورت رعایت شرایط خاص متوقف می‌شود. این می‌تواند به کاهش منابع محاسباتی و زمان مورد نیاز برای آموزش کمک کند و درعین حال عملکرد خوبی به دست آورد [۳۶]. یکی از کاربردهای توقف زود هنگام در CNN در زمینه انتخاب مدل است. به جای آموزش یک شبکه برای تعداد ثابتی از دوره‌ها و سپس انتخاب بهترین مدل با عملکرد بر اساس دقت اعتبارسنجی، توقف زود هنگام امکان نظارت بر دقت اعتبارسنجی در طول آموزش و توقف فرآیند را به محض شروع به کاهش، می‌دهد. این به جلوگیری از بیش‌برازش کمک می‌کند و تضمین می‌کند که مدل بیش‌ازحد لازم آموزش داده نمی‌شود. این روش به ما اجازه می‌دهد که مدل را در زمانی که دقت آن در مجموعه اعتبارسنجی بهبود نمی‌یابد متوقف کنیم.

به‌طور کلی، توقف زود هنگام در CNN می‌تواند به سرعت بخشیدن به فرآیند آموزش، کاهش هزینه‌های محاسباتی و بهبود کارایی کلی بدون به خطر انداختن عملکرد کمک کند. در روش توقف زود هنگام اصطلاح «مانیتور» به معیاری اطلاق می‌شود که مدل برای ارزیابی عملکرد خود در طول آموزش استفاده می‌کند. زمانی که از تلفات اعتبارسنجی به‌عنوان مانیتور استفاده می‌شود، به این معنی که مدل در حین آموزش، تلفات اعتبارسنجی را ردیابی می‌کند. تلفات اعتبارسنجی معیاری است که نشان می‌دهد مدل در مجموعه‌های از داده‌های جداگانه که آموزش ندیده‌اند، چقدر خوب عمل می‌کند. با نظارت بر تلفات اعتبارسنجی، مدل می‌تواند از تطبیق بیش از حد داده‌های آموزشی جلوگیری کند و اطمینان حاصل کند که به خوبی به داده‌های جدید و دیده نشده تعمیم می‌یابد.

پارامتر «Patience» در روش توقف زود هنگام به تعداد دوره‌هایی اشاره دارد که مدل قبل از توقف آموزش در صورت عدم بهبود در معیار نظارت شده منتظر می‌ماند. به عبارت دیگر، اگر تلفات اعتبارسنجی برای تعداد معینی از دوره‌ها (مثلاً ۱۰ دوره) کاهش پیدا نکند، فرآیند آموزش قبل از اینکه مدل کل دوره‌ها را آموزش ببیند، متوقف می‌شود. این می‌تواند برای جلوگیری از بیش‌برازش و صرفه‌جویی در زمان با پرهیز از دوره‌های غیرضروری آموزش در زمانی که هیچ پیشرفت دیگری انجام نمی‌شود، مفید باشد. به عنوان مثال، اگر در یک کار از تلفات اعتبارسنجی به‌عنوان مانیتور و تنظیم Patience بر روی مقدار ۱۰ استفاده کنید یعنی به CNN خود دستور می‌دهید که عملکرد خود را بر اساس تلفات اعتبارسنجی زیر نظر داشته باشد و در صورت عدم مشاهده بهبود در تلفات اعتبار برای ۱۰ دوره متوالی، آموزش را متوقف کند. این کمک می‌کند تا اطمینان حاصل شود که مدل ما به خوبی به داده‌های جدید تعمیم می‌یابد و از بیش‌برازش جلوگیری می‌کند و درعین حال در زمان آموزش نیز صرفه‌جویی می‌کند.

کودکان	تعداد داده آموزشی	تعداد داده اعتبارسنجی	تعداد داده آزمون	کل داده‌ها
غیر اوتیسم	۳۶۰	۴۰	۱۰۰	۵۰۰
اوتیسم	۳۶۰	۴۰	۱۰۰	۵۰۰
مجموع	۷۲۰	۸۰	۲۰۰	۱۰۰۰

### ۵-۳- پیش‌پردازش داده‌ها

هدف از پیش‌پردازش داده‌ها، تمیز کردن، برش دادن تصاویر، حذف هرگونه نویز و افزایش کیفیت تصویر است. این کار به دلیل جمع‌آوری داده‌ها از منابع مختلف و تنوع بالای تصاویر، اهمیت بسیاری داشت. با توجه به اینکه تصاویر از طریق خانواده‌ها و مراکز جمع‌آوری شدند، قبل از آنکه برای آموزش مدل‌های یادگیری عمیق استفاده شوند، باید پیش‌پردازش می‌شدند. مجموعه‌ی تصاویر چهره را از تصویر اصلی با استفاده از نرم‌افزار فتوشاپ برش و آن‌ها را به نسبت (۱:۱) و اندازه ۲۲۴×۲۲۴ پیکسل تغییر دادیم. تصاویر بی کیفیت از مجموعه حذف شدند. با توجه به اینکه برای انتخاب داده‌های از روش **K-Fold** استفاده و مقدار **k** را برابر ۵ قرار داده شده است، تصاویر به سه دسته آموزش، اعتبارسنجی و آزمون به ترتیب با نسبت ۸۰ درصد آموزش و اعتبارسنجی و ۲۰ درصد آزمون تقسیم شدند. به طوری که از ۵۰۰ تصویر ۳۶۰ تصویر برای آموزش، ۴۰ تصویر برای اعتبارسنجی و ۱۰۰ تصویر برای آزمون تخصیص داده شدند.

برای مقیاس‌گذاری، از روش نرمال‌سازی استفاده کردیم. به این صورت که مقادیر پیکسل تمام تصاویر از بازه [۰،۲۵۵] به بازه [۰،۱] تغییر مقیاس دادیم. این روش مقیاس‌گذاری، به قابلیت مقایسه و بهبود عملکرد مدل در آموزش و آزمون کمک کرد. در نهایت، پیش‌پردازش داده‌ها در بهبود دقت کار طبقه‌بندی توسط مدل‌های یادگیری عمیق، تأثیر مثبتی داشت.

### ۵-۴- انتخاب مدل

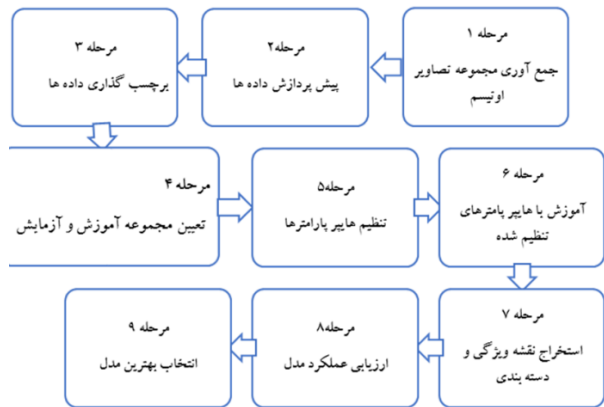
در اینجا از مدل‌های CNN استفاده می‌کنیم، زیرا CNN می‌تواند یک تصویر را دریافت کند، به اشیاء مختلف درون تصویر اهمیت بدهد و سپس اشیاء درون تصویر را از یکدیگر متمایز کند. علاوه بر این، در CNN‌ها پیش‌پردازش در مقایسه با روش‌های دیگر حداقل کار و زمان لازم دارد [۳۷]. برای آموزش از سه مدل Inceptionv3، Xception و Mobilenet با تأکید بر روی مدل Mobilenet با توجه به سبک بودن حجم و موفقیت در آزمایش‌های دیگر، استفاده کردیم. برای Mobilenet، از پایه‌ترین نسخه Mobilenet با آلفای ۱ و ضریب عمق ۱ استفاده کردیم. برای انجام پیش‌بینی‌های دودویی، دولایه کاملاً متصل به انتهای مدل اضافه کرده ایم. اولین لایه یک لایه مترکم با ۵۱۲ نورون است (نرمال‌سازی  $L2 = 0.015$ ، فعال‌ساز ReLU) که سپس به لایه پیش‌بینی متصل می‌شود که فقط دو خروجی دارد (فعال‌ساز Softmax). حذف تصادفی ۰.۵ را به لایه اول اعمال کردیم تا از بیش‌برازش جلوگیری کند. خروجی نهایی یک طبقه‌بندی دودویی از «اوتیسم» یا «غیر اوتیسم» است.

برای اجرای برنامه از زبان برنامه‌نویسی پایتون و پلتفرم کولاب استفاده شد. کولاب<sup>۱۴</sup> یک پلتفرم گوگل است که کاربر را قادر می‌سازد تا کد را بر روی یک GPU اختصاصی اجرا کند و این فرصت را فراهم می‌کند تا مجموعه داده کاربر را به اشتراک بگذارد و نتایج را منتشر کند. سخت‌افزار مورد استفاده GPU T4 8 Gb و RAM 12 Gb است. جهت آموزش از مدل‌های یادگیری انتقال عمیق

ببریم. این رویکرد اغلب منجر به عملکرد بهتر در مقایسه با آموزش یک CNN از ابتدا می‌شود، به‌خصوص زمانی که داده‌های برچسب‌گذاری شده محدودی در دسترس داریم.

### ۵-۱- روش کار

در این کار، با استفاده از مدل‌های یادگیری عمیق، تشخیص اوتیسم در کودکان ایرانی با دقت قابل قبولی انجام گرفته است و می‌تواند به‌عنوان یک شیوه جدیدی در تشخیص اوتیسم از روی ویژگی‌های صورت استفاده شود و در تشخیص اوتیسم در کودکان مفید باشد. در شکل ۳ فلوجارت مراحل انجام کار آورده شده است.



شکل ۳ فلوجارت مراحل انجام کار

### ۵-۲- مجموعه داده

از مجموعه داده‌ای شامل تصاویر صورت کودکان اوتیسم و کودکان عادی نژاد ایرانی استفاده شده است. این مجموعه داده، از طریق شبکه اجتماعی تلگرام با اعلام آمادگی والدین داری فرزند اوتیسم جهت شرکت در اینکار، مرکز آموزش استثنایی بشارت و توانبخشی حس پنهان کرمانشاه برای کودکان اوتیسم<sup>۱۳</sup> و کودکان مدارس عادی برای کودکان غیر اوتیسم جمع‌آوری شده است. دیتاست مجموعاً شامل ۱۰۰۰ تصویر چهره است که نیمی از آن‌ها مربوط به کودکان اوتیسم و نیمی دیگر از کودکان غیر اوتیسم می‌باشند.

روش اعتبارسنجی متقابل K-Fold یکی از تکنیک‌های مهم در یادگیری ماشین و یادگیری عمیق است که برای ارزیابی مدل‌ها و انتخاب داده‌ها استفاده می‌شود. در این روش، داده‌ها به **K** قسمت تقسیم می‌شوند. هر بار یکی از این بخش‌ها به‌عنوان داده آزمون و باقی بخش‌ها به‌عنوان داده آموزش استفاده می‌شوند. این فرآیند **K** بار تکرار می‌شود و هر بار یک قسمت متفاوت به‌عنوان داده آزمون در نظر گرفته می‌شود. این روش کمک می‌کند تا اطمینان حاصل شود که مدل روی تمامی داده‌ها خوب عمل می‌کند و وابسته به انتخاب خاصی از داده‌ها نیست. در پایان، نتایج هر **K** بار آزمون تجمیع شده و میانگین نتایج به‌عنوان عملکرد نهایی مدل گزارش می‌شود. در این کار برای انتخاب داده‌ها از روش **K-Fold** استفاده شده است. جدول ۱ توزیع داده توزیع داده‌ها را نشان می‌دهد.

### جدول ۱ توزیع داده‌ها

<sup>14</sup> Google Colab

<sup>۱۳</sup> با تعهد عدم انتشار تصاویر بصورت اخلاقی و تعهد کتبی صورت گرفته است

		مقادیر واقعی	
		مثبت	منفی
مثبت	TP	FP	مقیاس
	FN	TN	
منفی			

سایر معیارهای مهم ارزیابی که با استفاده از ماتریس سردرگمی محاسبه می‌شوند عبارت‌اند از دقت<sup>۱۷</sup> که با استفاده از فرمول (۲)، بازخوانی<sup>۱۸</sup> با روش فرمول (۳) و رتبه F1 با استفاده از فرمول (۴) محاسبه می‌شوند. دقت نسبت موارد مثبت پیش‌بینی شده درست را از بین همه موارد مثبت پیش‌بینی شده اندازه‌گیری می‌کند. درحالی‌که بازخوانی نسبت نمونه‌های مثبت درست پیش‌بینی شده را به همه نمونه‌های مثبت واقعی اندازه‌گیری می‌کند. امتیاز F1 ترکیبی از دقت و بازخوانی است که معیاری متعادل بین آن‌ها فراهم می‌کند.

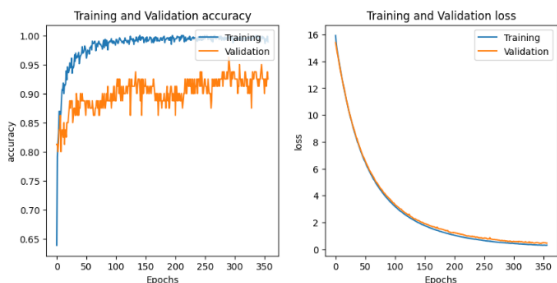
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{فرمول (۱)}$$

$$Precision = \frac{TP}{TP + FP} \quad \text{فرمول (۲)}$$

$$Recall = \frac{TP}{TP + FN} \quad \text{فرمول (۳)}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{فرمول (۴)}$$

با توجه به اینکه مدل Mobilenet باهدف ایجاد یک مدل سبک، سریع و با دقت قابل قبول، مورد استفاده در دستگاه‌های با منابع محدود مانند موبایل‌ها و بردهای مینی کامپیوتری و کاربردهای بلادرنگ طراحی شده است، نسبت به سایر مدل‌های مورد آزمایش عملکرد بهتری داشت. نمودار دقت و تلفات داده‌های انتخاب شده برای آموزش و اعتبارسنجی در شکل ۴ نشان داده شده است.



شکل ۴ نمودار دقت و تلفات داده‌های آموزشی و اعتبارسنجی

کار را با تنظیم فرآیندها (تنظیم لایه‌ها، بهینه‌ساز) بر روی مدل Mobilenet پنج بار و هر بار بر روی مجموعه داده‌های پنج‌گانه که قبلاً توضیح داده شد اجرا و به نتایج جدول ۱ ماتریس سردرگمی برای داده‌های اصلی رسیدیم. با استفاده از تقویت داده تعداد تصاویر مجموعه آموزش را برای هر کلاس به ۲۰۰۰ تصویر افزایش دادیم و به نتایج جدول ۲ ماتریس سردرگمی برای داده‌های تقویت شده رسیدیم که تفاوت آنچنانی نداشت.

در کتابخانه Keras API و برای تجزیه و تحلیل عملکرد مدل‌ها از چندین کتابخانه پردازش داده مانند Tensorflow, Matplotlib, Sklearn, Pandas و Numpy استفاده کردیم.

مجموعه پارامترهای بهینه آموزش مدل را که پس از انجام آزمایش‌های متعدد به دست آوردیم به صورت زیر است: اندازه دسته ۶۴ و نرخ یادگیری  $2 \times 10^{-5}$  است. از بهینه‌ساز تنظیم شده آدام برای تنظیم وزن استفاده کردیم. این بهینه‌ساز وزن‌های شبکه را بر اساس نزول گرادیان به گونه‌ای تنظیم می‌کند که عملکرد تابع تلفات را به حداقل می‌رساند و درنهایت توانایی آن را برای دسته‌بندی دقیق تصاویر دودویی بهبود می‌بخشد. درنهایت، آنتروپی متقابل دودویی را به عنوان تابع تلفات در نظر گرفتیم. برای ۵۰۰ دوره، مدل را با استفاده از توقف زودهنگام با تنظیمات Patience برابر ۱۰ و تلفات اعتبارسنجی آموزش داده و مدل را اعتبارسنجی کردیم.

#### ۶- نتایج شبیه‌سازی

ما در این بخش ابتدا به معرفی معیارهای ارزیابی و سپس به نتایج حاصل از آزمایش مدل‌های ارائه شده در این پژوهش می‌پردازیم. در بخش انتهایی نیز میزان دقت و خطای روش در این کار را با روش‌های دیگر مورد بررسی و مقایسه قرار خواهیم داد. بعد از مرحله آموزش، مدل‌ها جهت ارزیابی مورد آزمایش قرار می‌گیرند تا در شرایطی یکسان، وضعیت آن‌ها در مقایسه با مدل‌های دیگر سنجیده شود. در یادگیری عمیق برای آنکه میزان کارایی یک مدل مورد بررسی قرار گیرد، ابتدا خروجی مدل بر اساس معیارهای استاندارد همان حوزه مورد ارزیابی قرار داده می‌شود. سپس، نتایج حاصل از ارزیابی مدل با نتایج سایر پژوهش‌های مشابه مقایسه می‌شود.

#### ۶-۱- معیارهای ارزیابی شبکه‌های عصبی عمیق

یکی از معیارهای کلیدی ارزیابی شبکه‌های عصبی عمیق در طبقه‌بندی دودویی تصاویر، صحت<sup>۱۵</sup> مدل است. صحت اندازه‌گیری می‌کند که چگونه مدل به درستی برجسب‌های کلاس تصاویر را پیش‌بینی می‌کند [۳۸]. باین حال، صحت ممکن است به تنهایی برای ارزیابی عملکرد یک شبکه عصبی عمیق کافی نباشد. در فرمول (۱) نحوه محاسبه آن توضیح داده شده است. ماتریس سردرگمی<sup>۱۶</sup> [۳۸] یک ابزار تحلیلی در یادگیری ماشین و آمار است که برای ارزیابی عملکرد الگوریتم‌های دسته‌بندی استفاده می‌شود. این ماتریس یک جدول با ترکیبی از مقادیر پیش‌بینی شده و واقعی است. ماتریس سردرگمی به عنوان جدولی تعریف می‌شود که اغلب برای توصیف عملکرد یک مدل طبقه‌بندی بر روی مجموعه‌ای از داده‌های آزمایشی که مقادیر واقعی آن‌ها مشخص است، استفاده می‌شود. درایه‌های ماتریس سردرگمی عبارت‌اند از:

۱. مثبت واقعی (TP): مثبت را پیش‌بینی و درست است.
۲. منفی واقعی (TN): منفی را پیش‌بینی و درست است.
۳. مثبت کاذب (FP): مثبت را پیش‌بینی و نادرست است. (خطای نوع ۱)
۴. منفی کاذب (FN): منفی را پیش‌بینی و نادرست است. (خطای نوع ۲)

جدول ۲ ماتریس سردرگمی را نمایش می‌دهد.

#### جدول ۲ ماتریس سردرگمی

<sup>18</sup> Recall

<sup>15</sup> Accuracy

<sup>16</sup> Confusion Matrix

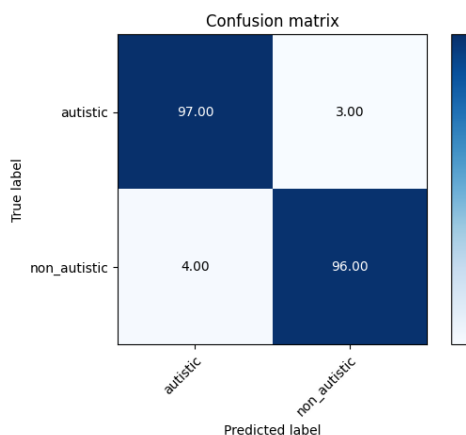
<sup>17</sup> Precision

داده	Accuracy:	Precision:	Recall:	F1 Score:
فولد ۱	۰.۸۹	۰.۹۱	۰.۸	۰.۸۹
فولد ۲	۰.۹۵۵	۰.۹۷	۰.۹	۰.۹۵
فولد ۳	۰.۹۲۵	۰.۹۰	۰.۹	۰.۹۳
فولد ۴	۰.۹۶۵	۰.۹۷	۰.۹	۰.۹۶
فولد ۵	۰.۹۴۵	۰.۹۴	۰.۹	۰.۹۵
می انگین	۰.۹۴	۰.۹۳۴	۰.۹	۰.۹۴

جدول ۴ نتایج به دست آمده برای داده‌های تقویت شده

داده	Accuracy:	Precision:	Recall:	F1 Score:
فولد ۱	۰.۹۰	۰.۹۰	۰.۹۱	۰.۹۱
فولد ۲	۰.۹۴	۰.۹۴	۰.۹۴	۰.۹۴
فولد ۳	۰.۹۲	۰.۹۵	۰.۸۹	۰.۹۲
فولد ۴	۰.۹۴	۰.۹۳	۰.۹۵	۰.۹۴
فولد ۵	۰.۹۳	۰.۹۰	۰.۹۵	۰.۹۳
میان گین	۰.۹۲	۰.۹۲	۰.۹۲	۰.۹۳

جزئیات ماتریس سردرگمی در شکل ۵ نشان داده شده است. برای موارد مثبت واقعی (TP)، مدل می‌تواند به درستی ۹۷ کودک را به‌عنوان مبتلا به اوتیسم پیش‌بینی کند. برای موارد منفی واقعی (TN)، مدل به درستی می‌تواند ۹۶ کودک را به‌عنوان غیراوتیسم پیش‌بینی کند. برای موارد منفی کاذب (FN)، مدل به اشتباه ۳ کودک را به‌عنوان غیراوتیسم پیش‌بینی کرد، اما آن‌ها اوتیسم داشتند. برای موارد مثبت کاذب (FP)، مدل به اشتباه ۴ کودک را به‌عنوان اوتیسم پیش‌بینی کرد، اگرچه آن‌ها اوتیسم نداشتند.



شکل ۵ ماتریس سردرگمی مربوط به بهترین حالت

جدول ۱ ماتریس سردرگمی برای داده‌های اصلی

داده	تعداد دوره آموزش	زمان اجرا	ثابت واقعی	منفی واقعی	منفی کاذب	ثابت کاذب
فولد ۱	۳۴	۰.۰۸	۹	۹	۱	۸
فولد ۲	۵۹	۳۰.۸	۹	۷	۶	۹
فولد ۳	۱۳	۷۵۶	۰	۰	۵	۹
فولد ۴	۷۵	۲۵۰	۹	۷	۴	۹
فولد ۵	۷۲	۴۶۴	۹	۴	۵	۹
می انگین	۵۱	۳۵۲	۹	۴	۶	۹

جدول ۲ ماتریس سردرگمی برای داده‌های تقویت شده

داده	تعداد دوره آموزش	زمان اجرا	ثابت واقعی	منفی واقعی	منفی کاذب	ثابت کاذب
فولد ۱	۳۴	۳۰.۸	۹	۸	۱	۹
فولد ۲	۰.۷	۳۸۳	۹	۷	۶	۹
فولد ۳	۰.۳	۵۱۹	۹	۷	۵	۸
فولد ۴	۶۱	۷۴۲	۹	۵	۷	۹
فولد ۵	۳۰	۰.۶۰	۹	۸	۵	۹
می انگین	۴۷	۴۶۷.۲	۹	۸	۷	۹

همان‌طور که در جداول ۳ و ۴ مشاهده می‌شود، تعداد دوره‌های آموزش برای داده‌های تقویت شده نسبت به داده‌های تقویت نشده کمتر است. در حالی که میانگین نتایج داده‌های تقویت نشده بهتر است. نتایج معیارهای ارزیابی برای داده‌های اصلی در جدول ۳ نتایج به دست آمده برای داده‌های اصلی و برای داده‌های تقویت شده در جدول ۴ نتایج به دست آمده برای داده‌های تقویت شده آورده شده است.

جدول ۳ نتایج به دست آمده برای داده‌های اصلی

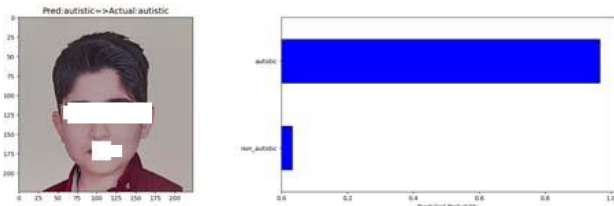
## ۶-۲- بحث

این تحقیق یک مدل برای تشخیص اوتیسم از روی تصویر صورت کودکان با استفاده از CNN مبتنی بر یادگیری انتقالی پیشنهاد می‌کند. معماری‌های CNN این توانایی را دارند که ویژگی‌های درون تصویر را استخراج کنند و با استفاده از این ویژگی‌های مشترک آن‌ها را به دو گروه اوتیسم و غیر اوتیسم طبقه‌بندی کنند. محققان مدل‌های مختلف را برای تصاویر عمومی [۱۴] و اختصاصی [۱۲] اعمال کردند. ما در این کار توانستیم برای اولین بار یک مدل بر پایه Mobilenet برای تشخیص اوتیسم در کودکان ایرانی ارائه کنیم و به نتایج میانگین مندرج در **جدول ۵ میانگین نتایج بدست آمده** برسیم. در بهترین حالت نیز برای مدل اصلاح شده Mobilenet نتایج Accuracy=97%، Precision=97%، recall=96%، AUC\_ROC=97%، و F1 score=96% به دست آمده است.

جدول ۵ میانگین نتایج بدست آمده

Mean AUC	Accuracy	Precision	Recall	F1_Score
۹۸٪	۹۳٪	۹۲٪	۹۳٪	۹۳٪

در آینده با استفاده از این مدل یک برنامه کاربردی جهت نصب بر روی گوشی‌های هوشمند تهیه می‌شود. این برنامه می‌تواند توسط مراکز سنجش کودکان برای ورود به مدرسه در آموزش و پرورش، مراکز سنجش بهزیستی، والدین و مربیان مورد استفاده قرار گیرد. تا با استفاده از یک تصویر ساده بتوانند تشخیص اولیه اوتیسم را همانند شکل ۶ انجام دهند و از نتایج بدست آمده جهت مراجعه به پزشکان متخصص در مغز و اعصاب یا روان پزشکی و تشخیص نهایی استفاده کنند. به عنوان کار آینده، می‌توان اوتیسم را با استفاده از جمعیت ویژگی‌های استخراج شده از تصاویر EEG و FMRI تشخیص داد و از آن در کنار برگه‌های ارزیابی و مدل تشخیص تصاویر چهره در پژوهش حاضر استفاده کرد تا متخصصان بتوانند با سهولت از آن استفاده کنند.



شکل ۶ نمونه خروجی برنامه

## ۸- حمایت

این کار بر اساس تصاویر ارسالی بعضی از والدین کودکان مبتلا و همچنین اداره کل بهزیستی و اداره کل آموزش و پرورش استان کرمانشاه با تعهد عدم انتشار تصاویر انجام گرفت. جا دارد در اینجا از همکاری این مجموعه‌ها تقدیر و تشکر گردد. این کار تحقیقی در آزمایشگاه هوش مصنوعی دانشگاه ایلام انجام گرفته است.

## ۹- محدودیت

در زمینه شرکت کردن در پروژه‌های تحقیقاتی به اندازه کافی برای خانواده و بخصوص مسئولین فرهنگ‌سازی نشده است. ما در اینکار، با ترس مردم و خانواده‌ها از مشارکت در این پروژه تحقیقی و نگرانی آنها بابت در اختیار قرار دادن تصاویر فرزندانشان مواجه شدیم.

با وصف اینکه اوتیسم در رفتار یک شخص مشخص می‌شود و چون یک اختلال رشدی عصبی است، می‌تواند با سایر اختلالات رشدی عصبی در زندگی واقعی اشتباه گرفته شود. بنابراین تشخیص آن کار ساده‌ای نیست. برای اینکه بتوان اوتیسم را با استفاده از تصاویر چهره تشخیص داد نیاز به جمع‌آوری تصاویر کودکان اوتیسمی و غیر اوتیسمی هست که این تصاویر را باید نشانه‌گذاری کرد. اگر بخواهید با استفاده از الگوریتم‌های یادگیری عمیق برای دسته‌بندی این دو گروه اقدام کنید نیاز به مجموعه‌ای بزرگ از تصاویر دارید که این تصاویر نیز باید دارای ویژگی‌های خاصی از جمله کیفیت، اندازه و... باشند که در غیر این صورت نتایج مناسبی به دست نخواهد آمد. در این تحقیق، مدل‌های مختلف پیش آموزش را با تغییرات مناسب در تنظیم پارامترها، تغییر و افزودن لایه‌ها، نرخ یادگیری و تعیین بهینه‌ساز مناسب با تصاویر گردآوری شده از چهره کودکان نژاد ایرانی آموزش داده شده و به نتایجی به شرح جدول ۳ دست یافتیم.

همان‌طور که در مرور ادبیات گفتیم شش مقاله مرتبط پیدا شد که در چهار مورد [۱۳، ۱۵، ۱۷، ۳۹] از مخزن عمومی تصاویر [۱۴] استفاده شده بود. در یک مورد مارک پرکوفسکی و همکاران [۱۲] از تصاویر چهره کودکان چینی و در یک مورد دیگر پروینا و ماتا در مطالعه‌ی [۱۱] از تعداد خیلی کمی از تصاویر چهره استفاده کردند. آنها ویژگی‌ها را به صورت دستی استخراج کرده بودند. باتوجه به اینکه طبق مطالعه مارک پرکوفسکی و همکاران [۱۲] نژاد در تشخیص اوتیسم با استفاده از تصاویر چهره تأثیر می‌گذارد، بنابراین در این مطالعه منحصراً از تصاویر چهره کودکان ایرانی استفاده شده است. در کار پروینا و ماتا [۱۱] از شبکه‌های عصبی مصنوعی و از ۶۸ علامت مشخصه صورت استفاده شده است که به دقت ۸۰٪ رسیدند. در حالی که شبکه‌های عصبی عمیق می‌توانند مشخصه‌های بیشتری را استخراج کنند.

در چهار مطالعه‌ای که بر روی داده‌های عمومی [۱۴] انجام شد فواز و همکاران [۱۵] با xception دقت ۹۱٪ را به دست آوردند. علیم و همکاران [۱۶] با استفاده از مدل اصلاح شده xception به دقت ۹۵٪ رسیدند. زیاد و همکاران [۱۷] با استفاده از mobilenet به دقت ۹۵٪ روی داده‌های اعتبار سنجی رسیدند. ولی برای داده‌های آزمون نتایج نشان داده نشد. رحمان و همکاران [۱۸] با xception به AUC=96.63% رسیدند. اما آنها نتایج مربوط به دقت و صحت را ارائه نکردند. نهایتاً پرکوفسکی و همکاران [۱۲] توانستند یک مدل را با دقت ۹۵٪ برای تشخیص اوتیسم در کودکان چینی آموزش دهند.

نکته‌ای که در این مطالعات قابل‌ارائه بود نحوه انتخاب داده‌ها بود که هیچ‌گونه توضیحی داده نشده است. در حالی که در این مطالعه انتخاب داده‌ها براساس روش K-Fold بوده است که روشی مورد قبول و شناخته شده در انتخاب داده‌ها هست. در بهترین حالت برای مدل اصلاح شده Mobilenet نتایج Accuracy=97%، Precision=97%، recall=96%، AUC\_ROC=97%، F1 score=96% و Mean AUC=99% به دست آمده است. نتایج امیدوار کننده‌ای هستند و می‌توان با استفاده از آنها برنامه‌ای جهت تلفن همراه هوشمند جهت تشخیص اوتیسم با استفاده از ویژگی‌های صورت توسعه داد. کیفیت تصاویر سن کودکان شرکت کننده در کیفیت مدل تأثیر بسزایی دارد. در اینجا اکثر کودکان شرکت کننده در محدوده سنی ۳±۷ بودند. چنانچه گروه سنی محدودتر و تصاویر با کیفیت‌تر شوند و از پس‌زمینه یکسانی استفاده شود نتایج بهتری به دست خواهد آمد.

## ۷- نتیجه‌گیری

- [21] A. Estes, J. Munson, S. J. Rogers, J. Greenon, J. Winter, and G. Dawson, "Long-term outcomes of early intervention in 6-year-old children with autism spectrum disorder," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 54, no. 7, pp. 580-587, 2015.
- [22] R. Bouma and R. Schweitzer, "The impact of chronic childhood illness on family stress: A comparison between autism and cystic fibrosis," *Journal of clinical psychology*, vol. 46, no. 6, pp. 722-730, 1990.
- [23] M. E. Dunn, T. Burbine, C. A. Bowers, and S. Tantleff-Dunn, "Moderators of stress in parents of children with autism," *Community mental health journal*, vol. 37, pp. 39-52, 2001.
- [24] V. Lotter, "Factors related to outcome in autistic children," *Journal of autism and childhood schizophrenia*, vol. 4, no. 3, pp. 263-277, 1974.
- [25] K. Aldridge et al., "Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes," *Molecular autism*, vol. 2, pp. 1-12, 2011.
- [26] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 201<sup>st</sup> international conference on engineering and technology (ICET), 2017, pp. 1-6: Ieee.
- [27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [28] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.
- [30] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251-1258.
- [31] J. Gu et al., "Recent advances in convolutional neural networks," *Pattern recognition*, vol. 77, pp. 354-377, 2018.
- [32] R. Zaheer and H. Shaziya, "A study of the optimization algorithms in deep learning," in 2019 third international conference on inventive systems and control (ICISC), 2019, pp. 536-539: IEEE.
- [33] L. Prechelt, "Early stopping-but when?," in *Neural Networks: Tricks of the trade*: Springer, 2002, pp. 55-69.
- [34] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, pp. 1-74, 2021.
- [35] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, 2016/05/28 2016.
- [36] L. Prechelt, "Early Stopping - But When?," in *Neural Networks: Tricks of the Trade*, G. B. Orr and K.-R. Müller, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 55-69.
- [37] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, p. 53, 2021/03/31 2021.
- [38] P. Cichosz, "Assessing the quality of classification models: Performance measures and evaluation procedures," *Central European Journal of Engineering*, vol. 1, no. 2, pp. 132-158, 2011/06/01 2011.
- [39] M. S. Alam, M. M. Rashid, R. Roy, A. R. Faizabadi, K. D. Gupta, and M. M. Ahsan, "Empirical Study of Autism Spectrum Disorder Diagnosis Using Facial Images by Improved Transfer Learning Approach," *Bioengineering (Basel, Switzerland)*, vol. 9, no. 11, p. 710, Accessed on: 2022/11/. doi: 10.3390/bioengineering9110710 Available: <http://europepmc.org/abstract/MED/36421111>  
<https://doi.org/10.3390/bioengineering9110710>  
<https://europepmc.org/articles/PMC9687350>  
<https://europepmc.org/articles/PMC9687350?pdf=render>
- [1] C. S. Paula, S. H. Ribeiro, E. Fombonne, and M. T. Mercadante, "Brief report: prevalence of pervasive developmental disorder in Brazil: a pilot study," *Journal of autism and developmental disorders*, vol. 41, pp. 1738-1742, 2011.
- [2] L. C. Nunes et al., "A hybrid model to guide the consultation of children with autism spectrum disorder," in *Research & Innovation Forum 2019: Technology, Innovation, Education, and their Social Impact 1*, 2019, pp. 419-431: Springer.
- [3] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Informatics for Health and Social Care*, vol. 44, no. 3, pp. 278-297, 2019.
- [4] N. Bagatell, "From cure to community: Transforming notions of autism," *Ethos*, vol. 38, no. 1, pp. 33-5, 2010.
- [5] J. Zeidan et al., "Global prevalence of autism: A systematic review update," *Autism Research*, vol. 15, no. 5, pp. 778-790, 2022.
- [6] D. Bone, M. S. Goodwin, M. P. Black, C.-C. Lee, K. Audhkhasi, and S. Narayanan, "Applying machine learning to facilitate autism diagnostics: pitfalls and promises," *Journal of autism and developmental disorders*, vol. 45, pp. 1121-1136, 2015.
- [7] G. Tripi et al., "Cranio-facial characteristics in children with autism spectrum disorders (ASD)," *Journal of Clinical Medicine*, vol. 8, no. 5, p. 641, 2019.
- [8] W. DeMyer, W. Zeman, and C. G. Palmer, "The face predicts the brain: diagnostic significance of median facial anomalies for holoprosencephaly (arhinencephaly)," *Pediatrics*, vol. 34, no. 2, pp. 256-263, 1964.
- [9] K. Aldridge et al., "Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes," (in eng), *Mol Autism*, vol. 2, no. 1, p. 15, Oct 14 2011.
- [10] M. Beary, A. Hadsell, R. Messersmith, and M.-P. Hosseini, "Diagnosis of autism in children using facial analysis and deep learning," *arXiv preprint arXiv:2008.02890*, 2020.
- [11] T. Lakshmi Praveena and N. Muthu Lakshmi, "A methodology for detecting ASD from facial images efficiently using artificial neural networks," in *Advances in Computational and Bio-Engineering: Proceeding of the International Conference on Computational and Bio Engineering*, 2019, Volume 1, 2020, pp. 365-373: Springer.
- [12] A. Lu and M. Perkowski, "Deep learning approach for screening autism spectrum disorder in children with facial images and analysis of ethnoracial factors in model development and application," *Brain Sciences*, vol. 11, no. 11, p. 1446, 2021.
- [13] M.-P. Hosseini, M. Beary, A. Hadsell, R. Messersmith, and H. Soltanian-Zadeh, "Deep learning for autism diagnosis and facial analysis in children," *Frontiers in Computational Neuroscience*, p. 119, 2022.
- [14] G. Piosenka, "Concerns with 'Detect Autism' Dataset ,Kaggle," <https://www.kaggle.com/datasets/cihan063/autism-image-data>, 2021.
- [15] F. W. Alsaade and M. S. Alzahrani, "Classification and detection of autism spectrum disorder based on deep learning algorithms," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [16] M. S. Alam, M. M. Rashid, R. Roy, A. R. Faizabadi, K. D. Gupta, and M. M. Ahsan, "Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach," *Bioengineering*, vol. 9, no. 11, p. 710, 2022.
- [17] Z. A. Ahmed et al., "Facial features detection system to identify children with autism spectrum disorder: deep learning models," *Computational and Mathematical Methods in Medicine*, vol. 2022, 2022.
- [18] K. Mujeeb Rahman and M. M. Subashini, "Identification of autism in children using static facial features and deep neural networks," *Brain Sciences*, vol. 12, no. 1, p. 94, 2022.
- [19] J. W. Jacobson and J. A. Mulick, "System and cost research issues in treatments for people with autistic disorders," *Journal of autism and developmental disorders*, vol. 30, no. 6, p. 585, 2000.
- [20] S. Bryson, "Our most vulnerable citizens: Needs of and service models for people with autism," Toronto: Autism Society Ontario, 1991.

