

طراحی و جمع‌آوری دادگان گفتاری به‌عنوان گام نخست بومی‌سازی تشخیص هوشمند اوتیسم در کودکان ایرانی

مریم علیزاده و شیما طیبیان

در زمینه تشخیص اوتیسم مبتنی بر پردازش گفتار افراد صورت گرفته است. افراد مبتلا به اوتیسم از نوعی اختلال رشدی رنج می‌برند که با علائمی همچون ناتوانی در برقراری ارتباطات اجتماعی نمایان می‌شود [۱]. شکل ۱ زنجیره تولید گفتار در انسان را نشان می‌دهد. همان طور که مشاهده می‌کنید، تولید گفتار با شکل‌گیری ایده بیان کلمات، عبارات یا جملات از مغز انسان شروع می‌شود. سپس فرمان‌هایی از مغز به اندام‌های گفتاری مانند لب‌ها، دندان‌ها، تارهای صوتی و زبان جهت قرارگرفتن آنها در موقعیت مشخص و به‌منظور بیان آوای کلمات تشکیل‌دهنده ایده مذکور صادر می‌شود. در پایان، آوای متناظر با ایده مذکور توسط اندام‌های گفتاری ادا می‌شوند. افراد مبتلا به اوتیسم در همان آغاز این فرایند، یعنی مغز دچار اختلال هستند که این اختلال می‌تواند بر صدور فرمان به اندام‌های گفتاری و در نتیجه بیان یا نحوه بیان واحدهایی آوایی و کلمات و جملات شامل آنها و حتی مکالمه‌های دوفره و یا داستان‌گویی تأثیر گذارد. در این بین، واحدهای آوایی که شامل مصوت "ی" هستند، بیشترین مشکل را تولید می‌کنند. این به دلیل عدم برقراری ارتباط کودکان مبتلا به اوتیسم با دنیای اطراف و عدم توانایی بیان واحدهای آوایی تداعی‌کننده فرم خنده در صورت است [۲]. در حالی که افراد سالم توانایی خندیدن همزمان با گفتار را دارند، افراد مبتلا به اوتیسم قادر به انجام این کار نیستند [۱]. از سوی دیگر تشخیص فرکانس گام و گفتار افراد مبتلا به اوتیسم در محیط‌های نویزی نسبت به افراد سالم سخت‌تر است؛ زیرا افراد مبتلا به اوتیسم، نسبت سیگنال به نویز (SNR) پایین‌تری دارند [۳]. از طرف دیگر، کودکان مبتلا به اوتیسم دارای اختلالاتی در پردازش و درک ویژگی‌های پایه‌ای^۲ مثل فرکانس گام گفتار هستند که منجر به اختلالاتی در درک و تولید لحن صدا و احساسات مانند غم و شادی می‌شود [۴]. پس تشخیص ابتلای افراد به اوتیسم می‌تواند مبتنی بر تحلیل و پردازش گفتارشان صورت پذیرد.

فضای خالی پژوهشی در این حوزه در کنار اهمیت بیشتر تشخیص بیماری اوتیسم در قیاس با گذشته، انگیزه‌های لازم جهت پرداختن به این حوزه را ایجاد کرده است. از این رو در بخش ابتدایی این مقاله، تحلیل و بررسی رویکردهای تشخیص بیماری اوتیسم مبتنی بر روش‌های پردازش گفتار، محور اصلی کار قرار گرفته و از سوی دیگر، فضای خالی پژوهشی در این حوزه در داخل ایران بسیار بیشتر نمایان است؛ به طوری که متأسفانه در داخل کشور تحقیقاتی در زمینه تشخیص اوتیسم از طریق گفتار منتشر نشده که یکی از دلایل اصلی آن، کمبود دادگان در این حوزه است. لذا گام نخست بومی‌سازی تشخیص اوتیسم مبتنی بر پردازش گفتار،

چکیده: اختلال طیف اوتیسم، نوعی اختلال رشدی به شمار می‌آید که از طریق علائمی مانند ناتوانی در برقراری ارتباط اجتماعی، خود را نشان می‌دهد. بنابراین بارزترین نشانه افراد مبتلا به اوتیسم، اختلال گفتار است. این مقاله در بخش اول به بررسی و مرور مطالعات انجام‌شده برای تشخیص خودکار اوتیسم بر اساس پردازش گفتار افراد مشکوک به ابتلا می‌پردازد. با توجه به بررسی‌های انجام‌شده، رویکردهای اصلی پردازش گفتار برای تشخیص اوتیسم به دو گروه تقسیم می‌شوند. گروه اول با پردازش پاسخ‌ها یا احساسات افراد مورد آزمایش در پاسخ به سؤالات یا داستان پرسشگر، افراد مبتلا به اوتیسم را تشخیص می‌دهند. گروه دوم، افراد مبتلا به اوتیسم را از طریق میزان نرخ دقت بازشناسی گفتارشان در سیستم‌های تشخیص خودکار گفتار از افراد سالم تفکیک می‌کنند. علی‌رغم پژوهش‌های زیاد انجام‌شده در این حوزه در خارج از ایران، پژوهش‌های اندکی داخل ایران انجام شده‌اند که مهم‌ترین دلیل آن، عدم وجود دادگان غنی متناسب با نیازمندی‌های تشخیص اوتیسم مبتنی بر پردازش گفتار افراد مبتلا است. در بخش دوم پژوهش حاضر به روند طراحی، جمع‌آوری و ارزیابی یک مجموعه دادگان گفتاری مستقل از گوینده برای تشخیص اوتیسم در کودکان ایرانی به‌عنوان گام نخست بومی‌سازی حوزه مذکور پرداخته‌ایم.

کلیدواژه: تشخیص اوتیسم، پردازش گفتار، یادگیری ماشین، دادگان گفتاری، کودکان، زبان فارسی.

۱- مقدمه

پایش خودکار سلامت، یکی از زمینه‌های مطرح در حوزه هوش مصنوعی در سال‌های اخیر است و در این میان، تشخیص ناهنجاری‌های گفتار بیش از پیش مورد توجه قرار گرفته است. با وجود پیشرفت‌های گوناگون در جوامع امروزی، هنوز تشخیص ناهنجاری گفتار توسط پزشکان و گفتاردرمانگران صورت می‌پذیرد. این امر منجر به وابستگی به نیروی انسانی متخصص و تجهیزات مرتبط با این حوزه شده و احتمال رخداد خطا در تشخیص را افزایش می‌دهد. منشأ ناهنجاری‌های گفتار، طیفی وسیع از بیماری‌ها مانند آلزایمر، پارکینسون، ام اس، سکته‌های مغزی و قلبی، تومور مغزی، تومور و کیست حنجره و اوتیسم است. بنا بر بررسی‌های انجام‌شده، علی‌رغم تحقیقات گسترده‌ای که در زمینه تشخیص بیماری‌هایی مانند آلزایمر، پارکینسون، ام اس، تومور و کیست حنجره مبتنی بر روش‌های پردازش گفتار انجام شده است، تحقیقات کمی

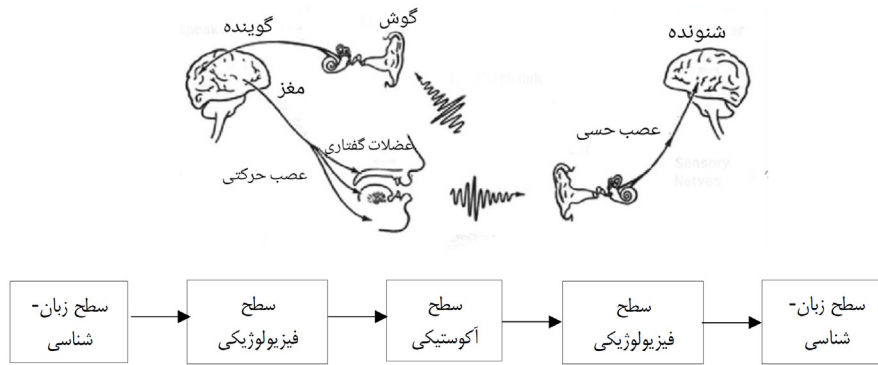
این مقاله در تاریخ ۲۹ بهمن ماه ۱۴۰۱ دریافت و در تاریخ ۱۲ اردیبهشت ماه ۱۴۰۲ بازنگری شد.

مریم علیزاده، پژوهشکده فضای مجازی، دانشگاه شهید بهشتی، تهران، ایران، (email: mar_alizadeh@sbu.ac.ir)

شیما طیبیان (نویسنده مسئول)، پژوهشکده فضای مجازی، دانشگاه شهید بهشتی، تهران، ایران، (email: sh_tabibian@sbu.ac.ir)

1. Signal to Noise Ratio

2. Basic Features



شکل ۱: زنجیره گفتار [۵].

دو نوع آزمایش به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته‌اند [۱۰]. همچنین استفانی شلینسکی و همکارانش از آزمون بازشناسی گفتار در حضور نویز و آزمون بررسی درک فرکانس گام^۲ گفتار توسط افراد مورد آزمایش برای تفکیک افراد سالم از افراد مبتلا به اوتیسم بهره برده‌اند [۳]. سعید صدیق و همکارانش از طریق بررسی ارتباطات اجتماعی مانند رفتارهای تکراری و کلیشه‌ای به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته‌اند [۱۱]. ونبو لو و همکارانش بر اساس واکنش افراد به اسامی (سرعت واکنش افراد، مدت برقراری ارتباط چشمی و جهت سر) به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته‌اند [۱۲]. ولاراتنا و همکارانش از طریق دسته‌بندی احساسات گفتار کودکان به هشت کلاس مختلف غم، انزجار، تعجب، خنثی، شادی، آرامش، ترس و خشم، افراد سالم را از افراد مبتلا به اوتیسم تفکیک کرده‌اند [۱۳].

از میان پژوهش‌های دسته دوم می‌توان به پژوهش دایکی میتسوموتو و همکارانش اشاره نمود که از ویژگی‌های صوتی مانند پرکننده‌ها، خنده و ویژگی‌های عروضی^۳ جهت تشخیص اوتیسم استفاده کرده‌اند. به دلیل اینکه افراد مبتلا به اوتیسم، توانایی استفاده از پرکننده‌ها و یا صحبت همراه با خنده را ندارند، استخراج این ویژگی‌ها به‌عنوان بهترین معیار جهت تشخیص افراد مبتلا به اوتیسم به‌کار رفته است [۱]. دونگ سین زو و همکارانش از پردازش سیگنال گفتار مبتنی بر ۳۹ مدل برای واج‌های انگلیسی مانند a، t و هفت مدل برای مکث، گریه، تردید و نفس کشیدن استفاده کرده‌اند [۱۴]. لینا جی پیلای و همکارانش از ۱۴ واحد آوایی در زبان مالایالم^۴ و اختلاف نرخ بازشناسی گفتار افراد جهت تشخیص افراد مبتلا به اوتیسم استفاده کرده‌اند [۱۵]. جان ژانگ و همکارانش با بررسی سیگنال‌های EEG در دو حالت گفتار و سکوت به تشخیص افراد مبتلا به اوتیسم پرداخته‌اند [۱۶]. سیگنال‌های گفتاری کودکان مبتلا به اوتیسم دارای اختلالات عروضی شامل فرکانس گام غیرعادی^۵، لحن یکنواخت و الگوهای تأکید زبانی غیرمرتبط است که در گفتار آنها نمایان می‌شود. ناتان آ چی و همکارانش با اتکا بر همین امر و ضبط و ثبت رفتارهای معمول و طبیعی کودکان به تفکیک آنها پرداخته‌اند [۱۷]. آیدا خزایی و همکارانش با استفاده از ویژگی‌های گریه به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته‌اند [۱۸]. تانیا تالکار و همکارانش به محاسبه ارتباط بین ویژگی‌های آکوستیکی شامل ویژگی‌های سطح بالا و سطح پایین شامل فرکانس اصلی (F_0)، شدت و ارتباط F_0 و شدت در بیان

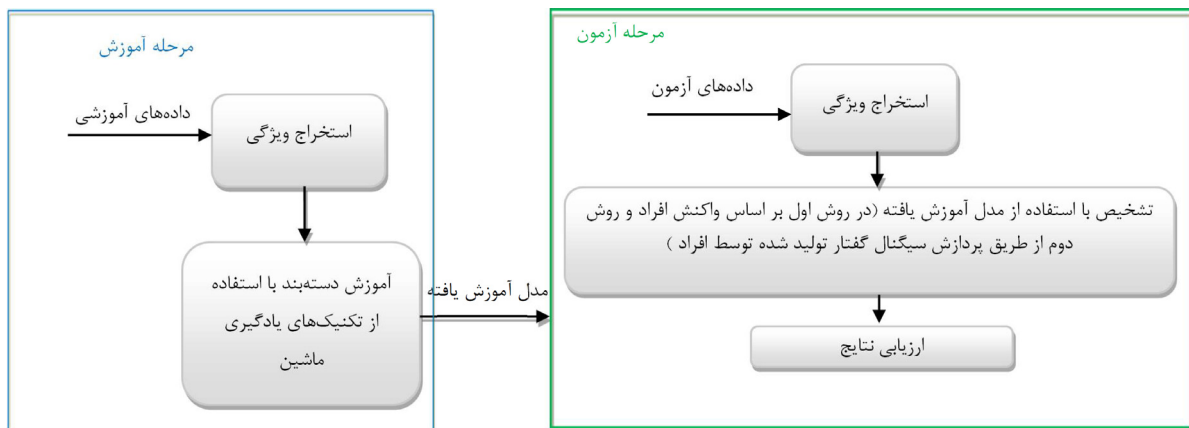
افراد مبتلا را به جمع‌آوری دادگان گفتاری مناسب اختصاص داده که در این مقاله به آن پرداخته‌ایم. در بخش دوم، کارهای پیشین انجام‌شده در این حوزه را مرور خواهیم کرد و چارچوبی را برای تشخیص اوتیسم مبتنی بر روش‌های پردازش گفتار ارائه می‌دهیم. تحقیقات انجام‌شده نشان‌دهنده این موضوع هستند که دادگان معتبری در این زمینه در ایران جمع‌آوری نشده است؛ لذا در بخش سوم به معرفی دادگان جدید تحت نظر درمانگر اوتیسم به‌عنوان نخستین گام بومی‌سازی این حوزه در ایران و ارزیابی دادگان ارائه‌شده با استفاده از مدل مخفی مارکوف^۱ (HMM) پرداخته‌ایم. در پایان، مقاله در بخش چهارم جمع‌بندی شده است.

۲- مرور ادبیات پیشین

رویکردهای موجود در حوزه تشخیص اوتیسم مبتنی بر پردازش گفتار به دو دسته تقسیم می‌شوند. دسته اول از طریق پردازش نوع گفتار و احساسات بداهه افراد در واکنش به داستان، جملات و یا سؤالات فرد پرسشگر به تشخیص و تفکیک افراد مبتلا به اوتیسم از افراد سالم می‌پردازد. دسته دوم از طریق پردازش و تحلیل سیگنال گفتار تولیدشده توسط افراد مورد آزمایش به تشخیص و تفکیک افراد مبتلا به اوتیسم از افراد سالم می‌پردازد. از میان پژوهش‌های دسته اول می‌توان به پژوهش جان بنگ و همکارانش اشاره نمود که از طریق خواندن داستان و نشان دادن عکس و نهایتاً با بررسی گفتار و رفتار بداهه و نوع واکنش احساسی کودک، امکان ابتلا به اوتیسم کودکان را تحلیل کرده‌اند [۶]. استفانی شلینسکی و همکارانش از طریق بیان ۱۳۴ کلمه دوهجایی آلمانی شامل احساسات شادی، غم، ترس، عصبانیت، انزجار و یا حالت خنثی و پردازش واکنش افراد مورد آزمایش به بررسی امکان ابتلای آنها به اوتیسم پرداخته‌اند [۴]. هانا درایمالا و همکارانش در انجام پژوهششان پاسخ کودکان به سؤالات مورد پرسش را که باید دربرگیرنده احساسات غم، ترس، عصبانیت، انزجار یا شادی باشد برای تشخیص ابتلا به اوتیسم استفاده کرده‌اند [۷]. استفانی شلینسکی از طریق نوع پاسخ و واکنش به سؤال‌ها، دقت حاصل از تشخیص هویت، درک احساسات و گفتار، به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته است [۸]. آلیس برد و همکارانش از طریق داستان‌سرایی که توسط ربّاتی انسان‌نما صورت گرفته و بررسی رفتارهای اجتماعی و واکنش‌های احساسی نسبت به وقایع داستان «جادوگر شهر اوز» به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته‌اند [۹]. النا لیاکس و همکارانش با بررسی دقت تشخیص معنای واژگان گفتار ادا شده و دقت تشخیص سن و جنسیت گویندگان در قالب

2. Vocal Pitch Perception
3. Prosodic Features
4. Malayalam
5. Atypical Pitch
6. Fundamental Frequency

1. Hidden Markov Model



شکل ۲: فرایند کلی تشخیص افراد مبتلا به اوتیسم از طریق گفتار.

۲-۲ ویژگی‌های مطرح برای تشخیص اوتیسم مبتنی بر رویکردهای پردازش گفتار

متناسب با دسته‌بندی رویکردهای تشخیص اوتیسم مبتنی بر روش‌های پردازش گفتار برای استخراج ویژگی نیز دو رویکرد لحاظ می‌شود. در رویکرد اول، ویژگی‌های مورد نیاز بر اساس واکنش‌های افراد و در رویکرد دوم، مستقیماً از سیگنال گفتار افراد استخراج می‌شوند. از سوی دیگر، ویژگی‌های استخراج‌شده به سه دسته ویژگی‌های عروزی، آکوستیکی و واکنش و عکس‌العمل‌های اجتماعی تقسیم می‌شوند.

بر اساس مطالعات انجام‌شده در رویکرد اول، ویژگی‌های آکوستیکی عمدتاً شامل ویژگی‌های استخراج‌شده از مجموعه‌های IS-09-emotion [۳۱]، IS ۱۰۰-parading [۳۲] و COMPARE [۳۳] در [۹]، ویژگی‌های استخراج‌شده از مجموعه‌های eGeMAPS^۳ و COMPARE در [۶] و ویژگی‌های MFCC در [۱۱] و [۱۲] هستند. همچنین ویژگی‌های عروزی شامل واریانس فرکانس و امتیاز $F1^4$ مانند [۱۳] و شیمر، جیتر، فرکانس گام، مکث قبل از تغییر گوینده در گفتمان دونفره و پرکننده‌ها در [۳۴] هستند. در برخی از پژوهش‌های رویکرد اول نیز از ترکیب ویژگی‌های آکوستیکی و عروزی مانند فرکانس گام و تن صدا و درک سکوت^۵ استفاده شده است [۸]. در رویکرد دوم، ویژگی‌های آکوستیکی عمدتاً شامل MFCC، نرخ گذار از صفر^۶ (ZCR)، فرمانت‌های اول تا پنجم^۷، انرژی، F_0 ، ویژگی‌های فرکانس غالب^۸ (FD_۱ و FD_۲)، زیر و بمی صدا^۹، ضرایب کپستروم پیش‌بینی خطی^{۱۰} (LPCC) و شدت^{۱۱} در [۱۵]، [۳۵] و [۳۶] هستند. همچنین مرکز طیفی^{۱۲}، فرکانس گام، میانگین انحراف از لگاریتم توان^{۱۳}، انحراف استاندارد^{۱۴}، F_0 ، ضریب چولگی طیف^{۱۵}

واحدهای آوایی مانند "pa-ta-ka"، ویدئو و دست‌خط مشتق‌شده از افراد سالم و مبتلا به اوتیسم پرداخته‌اند [۱۹]. آبیجیت موهانتا و همکارانش با ارائه پنج واحد آوایی شامل /a/، /e/، /i/، /o/ و /u/ و در نظر گرفتن ناحیه تغییر واحدهای آوایی، افراد سالم را از افراد مبتلا به اوتیسم تشخیص داده‌اند [۲۰]. در پژوهش دیگری، تصاویر گیاهان، حیوانات و شماره‌های انگلیسی به کودکان نشان داده شده و کودکان نام آنها را گفته‌اند. قالب کلماتی که ادا می‌شود شامل صامت-مصوت-صامت (CVVC)^۱ یا شامل صامت-مصوت-صامت^۲ (CVC) است [۲۱].

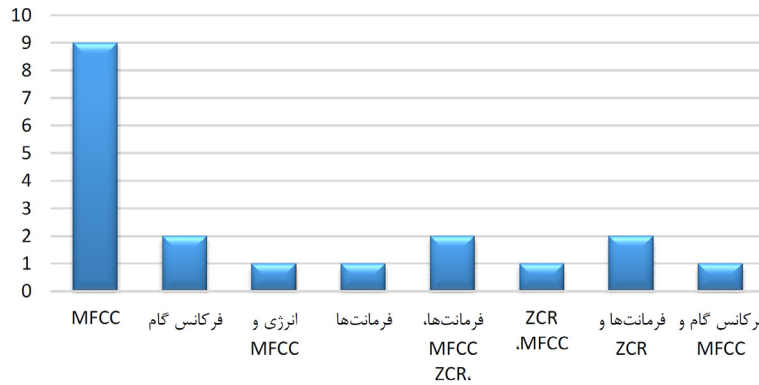
شکل ۲ فرایند کلی تشخیص مبتلا به اوتیسم را از طریق پردازش گفتار افراد مشکوک به ابتلا نشان می‌دهد. همان‌طور که در شکل آمده است، فرایند تشخیص افراد مبتلا به اوتیسم از طریق گفتار دارای دو مرحله آموزش و آزمون است. در مرحله آموزش ابتدا دادگان آموزش، وارد شده و سپس ویژگی‌هایی با قابلیت ایجاد بیشترین تمایز در بین دسته‌های مربوط به داده‌های ورودی استخراج می‌گردند. ویژگی‌های استخراج‌شده به منظور آموزش دسته‌بند استفاده می‌شوند. در مرحله آزمون ابتدا استخراج ویژگی از داده‌ها انجام می‌گیرد و پس از آن، دسته‌بند آموزش‌یافته در مرحله آموزش بر روی ویژگی‌های استخراج‌شده از دادگان آزمون اعمال می‌شود. خروجی حاصل از دسته‌بند به‌طور مستقیم (بر اساس تحلیل واکنش افراد) یا غیرمستقیم (بر اساس دقت بازشناسی گفتار عبارات ادا شده توسط افراد مورد آزمایش) برای تفکیک افراد سالم از افراد مبتلا به اوتیسم استفاده می‌شود. در ادامه به معرفی دادگان، ویژگی‌ها و رویکردهای دسته‌بندی مطرح این حوزه می‌پردازیم.

۱-۲ دادگان مطرح برای تشخیص اوتیسم مبتنی بر رویکردهای پردازش گفتار

دادگان مورد استفاده جهت تشخیص اوتیسم از طریق گفتار به دو دسته تقسیم می‌شوند. دادگان نوع اول، حاوی صداها یا سخنانی هستند که برای کودکان پخش می‌شوند و بر اساس تحلیل نوع واکنش‌های دریافت‌شده به تشخیص اوتیسم می‌پردازند. دادگان نوع دوم حاوی گفتار ضبط‌شده از افراد مبتلا به اوتیسم و سالم هستند که بر اساس دقت بازشناسی گفتار، تشخیص ابتلا یا عدم ابتلا به اوتیسم صورت می‌گیرند. جدول ۱ مجموعه‌ای از رایج‌ترین پایگاه داده‌های انتشار یافته و استفاده‌شده در این حوزه را نشان می‌دهد.

1. Consonant-Vowel-Vowel-Consonant
2. Consonant-Vowel-Consonant

3. Geneva Minimalistic Acoustic Parameter
4. F1 Score
5. Non-Vocal Pitch Perception
6. Zero Crossing Rate
7. The First, Second, Third, Fourth, and Fifth Formants
8. Dominant Frequencies
9. Pitch
10. Linear Prediction Cepstrum Coefficients
11. Intensity
12. Kurtosis of Centroid
13. Mean of Deviation of Log Power
14. Standard Deviation
15. Skewness



شکل ۳: توزیع پرتکرارترین ویژگی‌های استفاده‌شده در مقالات مورد مطالعه.

جدول ۱: محبوب‌ترین دادگان مورد استفاده برای تشخیص اوتیسم بر اساس تکنیک‌های پردازش گفتار.

محتویات دادگان	مرجع استفاده‌کننده از دادگان	نام پایگاه داده	رویکرد تشخیص اوتیسم
ضبط و ثبت گفتار بداهه شامل سه دسته از احساسات (مثبت، خنثی و منفی) از کودکان دارای اختلالات رشد (توسط نشان‌دادن عکس و تعریف داستان مربوط به آن عکس) کلمات معنادار دو یا سه‌هجایی بدون دارابودن احساسات مثبت یا منفی (خنثی) داده‌های صوتی جمع‌آوری‌شده طی سه تا پنج جلسه کوتاه روزانه. کودکان در یک برنامه آموزشی تشخیص احساسات به رهبری انسان یا ربات بر اساس کتاب کار «آموزش کودکان مبتلا به اوتیسم در خواندن ذهن» [#] شرکت کرده‌اند.	[۶]	[۲۲] CPESD*	
۲۵۴۲ فایل گفتاری کوتاه از چهار نوع لحن بیان برای ارزیابی توانایی‌های کودکان در تقلید از انواع لحن بیان	[۲۳]	ETL-WD**	
فایل‌های طولانی (۱۰ تا ۱۵ دقیقه‌ای) کودکانی با رشد ذهنی غیرعادی (مانند کودکان مبتلا به اوتیسم، سندروم دان و ...) در سنین بین ۴ تا ۱۶ سال؛ حاوی گفتار آزمایشگر، والدین و گاهی اوقات کودکان دیگر به زبان روسی	[۹]	[۹] DE-ENIGMA	دسته اول
فایل‌های صوتی و تصویری گردآوری‌شده در آمریکای شمالی همراه با هشت احساس مختلف (شادی، غم، خنثی، تعجب، آرامش، خشم، ترس و انزجار)	[۲۵]	[۲۴] CPSD###	
عکس‌العمل‌های کودکان هنگام صدازدن نام کودکان	[۱۰] و [۲۶]	[۱۰] AD-Child. Ru	
الگوهای رفتاری مرتبط با افراد مبتلا به اوتیسم شامل ارتباطات و ویژگی‌های رفتاری کلیشه‌ای و تکراری، کیفیت واکنش‌های اجتماعی، تماس چشمی غیرعادی، میزان ارتباطات اجتماعی متقابل و طرز ایستادن	[۲۸]	[۲۷] RAVDESS [§]	
ارائه و معرفی پنج حرف صدادار انگلیسی شامل /u/ و /o/ /i/ /e/ /a/	[۱۲]	Response to Name Dataset [۱۲]	
	[۳۰] و [۱]	[۲۹] ADOS module	
	[۲۱] و [۲۰]	Autism Speech Dataset [۲۰]	دسته دوم

* Child Pathological & Emotional Speech Database

** Japanese Phonetically-Balanced Word Speech Database، تولیدشده توسط مؤسسه ملی علوم و فناوری پیشرفته صنعتی و منتشرشده توسط کنسرسیوم (consortium) منابع گفتاری مؤسسه ملی انفورماتیک در ژاپن

Teaching Children with Autism to Mind-Read

Child Pathological Speech Database

\$ The Ryerson Audio-Visual Database of Emotional Speech and Song

«ایم» یا «آه» در زمانی که حرفی برای گفتن نداریم و یا در حال فکر هستیم، اشاره کرد [۱]. شکل ۳ توزیع پرتکرارترین ویژگی‌های استفاده‌شده در مقالات مورد مطالعه را نشان می‌دهد. بر اساس شکل، گرایش به استفاده از ویژگی‌های MFCC و یا ترکیب آن با سایر ویژگی‌ها بسیار زیاد است.

۳-۲ روش‌های دسته‌بندی مطرح برای تشخیص اوتیسم مبتنی بر رویکردهای پردازش گفتار

دسته‌بندهای مورد استفاده در حوزه اوتیسم مبتنی بر رویکردهای پردازش گفتار را می‌توان به دو گروه تقسیم کرد. در گروه اول، بر اساس

و ضریب کشیدگی طیف^۱ به‌عنوان ترکیب ویژگی‌های عروزی و آکوستیکی در [۱] استفاده شده‌اند. بر اساس تحقیقات انجام‌شده، گرایش به استفاده از ویژگی‌های آکوستیکی و یا ترکیبات آنها در هر دو رویکرد بیشتر از ویژگی‌های عروزی است و همچنین بیش از ۵۰ درصد تحقیقات انجام‌شده در هر دو رویکرد از ویژگی‌های آکوستیکی بیشتر از سایر ویژگی‌ها بهره گرفته‌اند. از میان ویژگی‌های منتسب به واکنش‌ها و عکس‌العمل‌های اجتماعی که غالباً جزو ویژگی‌های رویکرد اول محسوب می‌شوند، می‌توان به خندیدن هنگام صحبت، استفاده از صداهایی مثل

1. Kurtosis

جدول ۲: روش‌های مختلف دسته‌بندی مورد استفاده جهت تشخیص اوتیسم مبتنی بر گفتار در رویکرد اول.

مرجع	سال	تکنیک‌های دسته‌بندی	نرخ تشخیص گفتار/احساسات (%)		دقت تشخیص (%)	
			افراد مبتلا به اوتیسم	افراد سالم	افراد مبتلا به اوتیسم	افراد سالم
[۶]	۲۰۱۷	GAN و مقایسه با سایر دسته‌بندها مانند SVM خطی، SVM همراه با کرنل تابع پایه شعاعی* (RBF) و MLP	گزارش نشده	گزارش نشده	۱۰	۸۹
[۹]	۲۰۱۷	شبکه عصبی پیچشی** (CNN) (برای دسته‌بندی تصاویر) و SVM (برای دسته‌بندی گفتار)	گزارش نشده	گزارش نشده	۷۳٫۷	گزارش نشده
[۱۲]	۲۰۱۷	درخت تصمیم	گزارش نشده	گزارش نشده	۹۳#	گزارش نشده
[۷]	۲۰۱۸	جنگل تصادفی، SVM (برای دسته‌بندی گفتار) و CNN (برای دسته‌بندی تصاویر)	گزارش نشده	گزارش نشده	۷۸	گزارش نشده
[۱۱]	۲۰۱۹	حافظه کوتاه‌مدت ماندگار### (LSTM)	۸۴٫۷	گزارش نشده	گزارش نشده	گزارش نشده
[۴]	۲۰۱۹	رویکردهای دسته‌بندی ارائه‌شده در سیستم ANOVA	۶۵ (فرکانس گام)، ۶۲ (احساسات گفتار) و ۸۱ (درک بدون فرکانس گام)	۳۶ (فرکانس گام)، ۸۲ (احساسات گفتار) و ۸۲ (درک بدون فرکانس گام)	گزارش نشده	گزارش نشده
[۲۸]	۲۰۲۰	استفاده از دسته‌بند SVM و مقایسه با دسته‌بندهای درخت تصمیم، پرسپترون و رگرسیون لجستیک	گزارش نشده	گزارش نشده	۷۷	گزارش نشده
[۱۳]	۲۰۲۱	CNN	۹۰	گزارش نشده	گزارش نشده	گزارش نشده

* Radial Basis Function

** Convolutional Neural Networks

این مقاله از تشخیص خودکار نام در هنگام صدازدن مبتنی بر تشخیص گفتار استفاده می‌کند و به این ترتیب، پایگاه داده "Response to Name Dataset" را جهت تشخیص اختلال طیف اوتیسم ارائه داده است. همچنین از طریق تشخیص صورت و تخمین موقعیت سر با در نظر گرفتن سرعت واکنش، مدت برقراری ارتباط چشمی و جهت سر استفاده کرده است.

Long Short-Term Memory

شبکه‌های عصبی عمیق به تشخیص افراد مبتلا به اوتیسم پرداخته است. در این بخش، دادگان و تحقیقات موجود در حوزه تشخیص اوتیسم از طریق گفتار را مورد بررسی قرار داده‌ایم. متأسفانه داخل ایران تحقیقاتی در زمینه تشخیص اوتیسم از طریق گفتار منتشر نشده که از دلایل اصلی آن، کمبود دادگان در این حوزه است. لذا گام نخست بومی‌سازی تشخیص خودکار اوتیسم را به جمع‌آوری دادگان گفتاری مناسب اختصاص دادیم. بخش بعدی، جزئیات مربوط به طراحی و جمع‌آوری دادگان و نتایج تحلیل و ارزیابی این دادگان را توضیح می‌دهد.

۳- جمع‌آوری دادگان گفتاری مستقل از گوینده به زبان فارسی برای تشخیص کودکان مبتلا به اوتیسم (PERSIANSICHASD)

برای جمع‌آوری دادگان گفتاری مناسب علاوه بر مطالعات و تحقیقات انجام‌شده که نتایج آنها در بخش‌های قبلی منعکس شد، به مشاوره با یک متخصص درمانگر در حوزه اوتیسم پرداختیم. توجه به بازه سنی کودکان در این پژوهش از آنجا نشأت می‌گیرد که هرچه تشخیص ابتدای به اوتیسم در سنین پایین‌تری انجام شود، احتمال درمان و کیفیت تعاملات اجتماعی افراد افزایش می‌یابد. لازم به ذکر است که دادگان گفتاری PersianSICHASD^۲ به‌طور مختصر در یک مقاله کنفرانسی بین‌المللی به زبان انگلیسی [۲] معرفی شده است. اما در مقاله حاضر این دادگان با جزئیات بیشتری، معرفی و با رویکردی متفاوت از رویکرد مقاله کنفرانسی، ارزیابی و تحلیل شده است. همچنین مرور کامل و تحلیل و دسته‌بندی پژوهش‌های انجام‌شده در حوزه اوتیسم برای اولین بار در کار حاضر ارائه شده‌اند. در ادامه روند جمع‌آوری و ارزیابی دادگان آمده است.

تحلیل واکنش فرد به یک گفتار، موزیک یا سیگنال خاص به تفکیک افراد سالم از افراد مبتلا به اوتیسم می‌پردازند. در گروه دوم بر اساس استخراج ویژگی گفتار بیان‌شده توسط خود فرد و تحلیل گفتارشان با استفاده از انواع دسته‌بندها به تفکیک افراد سالم از افراد مشکوک به ابتلا به اوتیسم پرداخته می‌شود. معمولاً در این گروه، اختلاف بارز نرخ بازشناسی خودکار گفتار افراد سالم و افراد مبتلا به اوتیسم برای تفکیک این دو دسته از هم استفاده می‌شود.

جدول‌های ۲ و ۳، روش‌های مختلف دسته‌بندی در دو رویکرد مذکور را ارائه کرده‌اند. در بعضی از تحقیقات مورد بررسی در جدول ۲ (به‌ویژه در دسته اول)، روش‌های دسته‌بندی به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته‌اند. در این مقالات، نتایج ارزیابی برحسب دقت تشخیص افراد سالم و دقت تشخیص افراد مبتلا به اوتیسم گزارش شده‌اند. با این حال در سایر مطالعات انجام‌شده (به‌ویژه در دسته دوم)، نتایج به‌دست‌آمده از روش‌های دسته‌بندی باید تجزیه و تحلیل شوند تا افراد سالم از افراد مبتلا به اوتیسم تفکیک گردند. نتایج ذکرشده معمولاً نرخ بازشناسی صحیح^۱ گفتار افراد سالم یا مبتلا به اوتیسم است. تفاوت بارز بین نرخ بازشناسی صحیح برای جداسازی افراد سالم از افراد مبتلا به اوتیسم استفاده می‌شود.

شکل ۴ توزیع استفاده از انواع دسته‌بندها و ترکیباتشان را در مقالات مطالعه‌شده نشان می‌دهد. مطابق با جدول ۲ و نمودار توزیع دسته‌بندها در شکل ۴، در اکثر تحقیقات بررسی‌شده از دسته‌بند SVM یا ترکیبی از این دسته‌بند با سایر دسته‌بندها برای تشخیص افراد مبتلا به اوتیسم از افراد سالم استفاده شده است. محققان از روش‌های دیگری نیز مانند پردازش تصویر به تشخیص افراد مبتلا به اوتیسم پرداخته‌اند. مثلاً [۳۷] با به‌کارگیری تصاویر MRI از مغز فرد مشکوک به ابتلا و با استفاده از

جدول ۳: روش‌های مختلف دسته‌بندی مورد استفاده جهت تشخیص اوتیسم مبتنی بر گفتار در رویکرد دوم.

مرجع	سال	تکنیک‌های دسته‌بندی	نرخ تشخیص گفتار/احساسات (%)		دقت تشخیص (%)	
			افراد مبتلا به اوتیسم	افراد سالم	افراد مبتلا به اوتیسم	افراد سالم
[۱۴]	۲۰۰۹	روش بدون نظارت خوشه‌بندی k -means*	گزارش نشده	گزارش نشده	۸۷٫۴	گزارش نشده
[۳۸]	۲۰۱۳	SVM	گزارش نشده	گزارش نشده	۹۶٫۱۷	گزارش نشده
[۳۴]	۲۰۱۶	SVM و رگرسیون بردار پشتیبان	گزارش نشده	گزارش نشده	۶۹٫۴۲	۹۳٫۵۸
[۱۵]	۲۰۱۷	DNN همراه با خود رمزگذار	۵۶	۹۸	گزارش نشده	گزارش نشده
[۳۰]	۲۰۲۱	SVM، جنگل تصادفی، k نزدیک‌ترین همسایه و دسته‌بند بیز ساده (NB)**	گزارش نشده	گزارش نشده	۸۵	گزارش نشده
[۱۶]	۲۰۱۹	استفاده از ابزار عدم تطابق منفی (MMN)# و دامنه پاسخ عدم تطابق (MMR)##	۷۰٫۵	۶۹٫۳۸	گزارش نشده	گزارش نشده
[۳۹]	۲۰۱۹	شبکه‌های عصبی	۷۲ (تشخیص الگو)	گزارش نشده	۷۸ (عبارات گفتاری)	گزارش نشده
[۱]	۲۰۱۹	SVM	گزارش نشده	گزارش نشده	۷۹	۶۴

این مقاله چندین وظیفه را تعیین می‌کند:
 (۱) دقت $AUC = 1.00$ برای فرمانت‌ها، F_0 و ...
 (۲) دقت $AUC: 0.84$ حاصل از بیان واحدهای آوایی خاص مانند 'pa-ta-ka'.
 (۳) بیان صامت‌های مشخصی که بهترین نتیجه را در فرمانت‌ها (۸۸) $AUC: 0.84$ و F_0 و شدت دارند.
 (۴) گفتار آزاد دارای بهترین نتیجه در F_0 ($AUC: 1.00$) است.
 (۵) تمام وظایف گفتاری دارای بهترین نتیجه در F_0 ($AUC: 1.00$) است.
 میانگین همبستگی 0.72 با نمرات صحیح ADOS@
 دقت میانگین دخترها و پسرها: 0.81 @
 جنگل تصادفی: 0.70
 CNN: 0.79

* Unsupervised K-Means Clustering

** Naive Bayes

Mismatch Negativity

Mismatch Response

\$ There is no clear difference between people with autism and healthy people

\$\$ Gaussian Mixture Model

@ True ADOS Scores

@@ دقت 0.857 درصد برای 24 پسر مبتلا به اوتیسم و دقت 0.714 درصد برای 7 دختر مبتلا به اوتیسم

۳-۱ طراحی و جمع‌آوری دادگان

بر اساس مشاوره‌ها و نظرات متخصص درمانگر در حوزه اوتیسم، از ویژگی‌های افراد مبتلا به اوتیسم، کم‌توانی در تولید بعضی از واحدهای آوایی می‌باشد. با توجه به نقاط ضعف افراد مبتلا به اوتیسم، پیشنهادهایی از طرف درمانگر جهت تشخیص ابتلا به اوتیسم مطرح گردید که شامل موارد زیر است:

- ۱) بیان کلمه‌ای توسط سیستم و درخواست بیان کلماتی نزدیک به آن توسط فرد مشکوک به ابتلا
 - ۲) بیان صفت ساده‌ای به کودک و درخواست از او برای ساخت موصوف و صفت
 - ۳) بیان یک واحد آوایی به کودک و درخواست جهت وصل کردن واحد آوایی دوم به آن
 - ۴) بیان کلمه‌ای به کودک و درخواست از او برای تکرار واحد
- ۵) بیان کلمه‌ای به کودک و درخواست ساخت جمله با آن کلمه
 - ۶) بیان یک واحد آوایی (به‌عنوان مثال "ت") به کودک و درخواست بیان دو نوع غذا، اسباب‌بازی، گل و ... که با آن واحد آوایی شروع می‌شود.
 - ۷) بیان کلمه‌ای به کودک و درخواست بیان متضاد آن
 - ۸) بیان کلمه‌ای بی‌معنا به کودک و درخواست بیان کلمه معنادار
 - ۹) بیان یک نوع غذا به کودک و درخواست بیان یک نوع غذای دیگر
 - ۱۰) بیان یک عدد به کودک و درخواست گفتن دو عدد قبل یا بعد از آن
 - ۱۱) درخواست از کودک جهت بیان کلمه‌ای با لب‌هایش یا انتهای حلقش یا زبانش و یا بالعکس (مثلاً سؤال شود که واحد آوایی "ب" با کدام اندام گفتاری بیان می‌شود).

جدول ۴: واحدهای آوایی پیشنهاد شده جهت تفکیک افراد مبتلا به اوتیسم از افراد سالم.

شماره	واحدهای آوایی	نمایش معادل در سیستم IPA
۱	ای	e□
۲	آی	a□
۳	یه	əh
۴	ای	i□
۵	اوی	u□i
۶	ای	□□i
۷	اوو	□□vva
۸	ایی	i□jjə
۹	گپگ	gəpəgə
۱۰	کشش صامت "س"	ss
۱۱	کشش صامت "ش"	□□

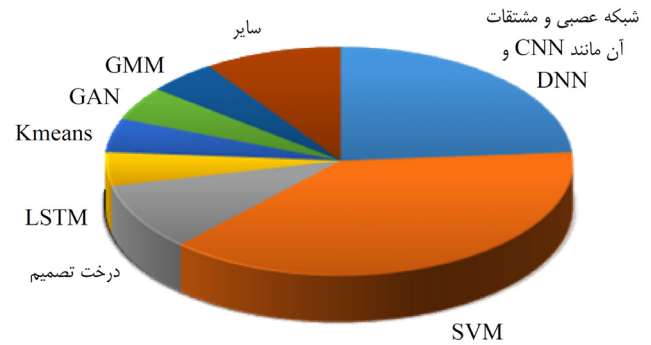
بردار پشتیبان یا روش‌های جدیدی مانند حافظه کوتاه مدت ماندگار یا حتی رویکردهای بازشناسی گفتار پایانه به پایانه استفاده کنیم. هرچه دادگان ضبط شده از کودکان، تناسی بیشتر با سطح توانمندی‌های کودکان مبتلا به اوتیسم در بیان گفتار داشته باشد، اختلاف نرخ دقت بازشناسی گفتار بین کودکان سالم و مبتلا به اوتیسم بالاتر خواهد بود؛ زیرا کودکان سالم، واحدهای آوایی را بدون مشکل خاصی تلفظ می‌کنند؛ لیکن کودکان مبتلا به اوتیسم یا قادر به تلفظ آنها نبوده یا آنها را بسیار ناقص ادا می‌کنند. بنابراین رویکرد بازشناسی گفتار، مشروط به انتخاب واحدهای آوایی مناسب، فارغ از نوع روش مورد استفاده (یادگیری عمیق یا روش‌های سنتی)، نرخ بازشناسی گفتار بسیار پایینی برای گفتار ضبط شده کودکان مبتلا به اوتیسم دارند و اختلاف نرخ دقت بازشناسی گفتار میان کودکان سالم و مبتلا به اوتیسم، عدد قابل توجهی خواهد بود. با توجه به مشاهده ضعف بارز کودکان مبتلا به اوتیسم در بیان واحدهای آوایی انتخاب شده، حجم محدود مجموعه دادگان جمع‌آوری شده در این پژوهش در قیاس با نیازمندی‌های آموزشی رویکردهای یادگیری عمیق به دادگان بزرگ، کمبود تعداد واحدهای آوایی مورد نیاز (یازده واحد) و دقت بازشناسی کافی و قابل قبول رویکردهای مبتنی بر مدل مخفی مارکوف به‌عنوان یکی از رویکردهای همچنان مطرح در حوزه بازشناسی گفتار در سطح کلمه با دادگان محدود، نیازی به صرف هزینه بالا برای آموزش رویکرد بازشناسی گفتار مبتنی بر روش‌های یادگیری عمیق که معمولاً برای بازشناسی گفتار محاوره‌ای مورد استفاده قرار می‌گیرند، نبوده، روش‌های مبتنی بر مدل‌های مخفی مارکوف با توجه به هدف پژوهش حاضر از دقت مورد نیاز برخوردار خواهند بود. ارزیابی نتایج بر اساس دو معیار دقت^۲ و صحت^۳ انجام شده که با استفاده از (۱) و (۲) محاسبه می‌شوند

$$Accuracy = \frac{H - I}{N} \quad (1)$$

$$Correctness = \frac{H}{N} \quad (2)$$

که H تعداد تشخیص‌های درست، N تعداد کل کلمات قابل مقایسه و I تعداد خطاهای درج کلمات است.

برای تنظیم پارامترهای مدل مخفی مارکوف، دادگان گویندگان سالم به دو بخش آموزش و آزمون تقسیم شده‌اند. مجموعه آزمون گویندگان



شکل ۴: توزیع استفاده از انواع دسته‌بندها و ترکیباتشان در مقالات مطالعه شده.

(۱۲) بیان واحدهای آوایی سایشی مانند "س" و "ش" به مدت دو یا سه ثانیه همراه با کشش

(۱۳) بیان یک واحد آوایی مثل تاتایا و درخواست تغییر نسبت‌های

واحدهای آوایی‌شان و تکرار یک تعداد مشخص از آن واحد آوایی

(۱۴) درخواست بیان با لحن افتان یا خیزان کلمه و یا واحد آوایی

(۱۵) بیان حروف یک کلمه معنادار و درخواست از کودک برای بیان درست کلمه. مثلاً می‌گوییم ک ف ش و کودک باید بگوید کفش.

بر اساس مقالات مطالعه شده و جلسات پرسش و پاسخی که با فرد درمانگر گذاشته شد، واحدهای آوایی جدول ۴ جهت تفکیک افراد مبتلا به اوتیسم از افراد سالم، محور جمع‌آوری دادگان قرار گرفتند. پیشنهادهای دیگر به دلیل ایجاد سوگیری در نتایج حاصل از آنها و سختی در برقراری ارتباط با افراد مبتلا به اوتیسم استفاده نشدند.

برای ویرایش دادگان PersianSICHASD از نرم‌افزار کول ادیت^۱ استفاده گردید و هر فایل ضبط شده در محیط نرم‌افزار، باز و نمایش داده می‌شود. چنانچه صداهایی از قبیل کلیک، سرفه، باز و بسته شدن در و ... در بخش‌های سکوت فایل وجود داشته باشد، با استفاده از امکانات نرم‌افزار حذف می‌شود و سایر نویزهای محیطی در فایل با همان شدت باقی می‌مانند. البته با توجه به آنکه ضبط صدای گویندگان در محیط تمیز و عاری از نویز انجام شده است، میزان این نویز قابل توجه نیست و بنابراین می‌توان ادعا کرد که تمام فایل‌ها بدون هیچ گونه نویز قابل توجهی جمع‌آوری شده‌اند. برچسب‌گذاری دادگان در سطح واحد آوایی، ادا شده و به صورت دستی انجام گردیده و همچنین برای قسمت‌های سکوت از برچسب sil استفاده شده است. بنابراین دادگان آموزش و آزمون گویندگان سالم حاوی ۳۳۰ فایل با پسوند wav (۳۰ فایل به‌ازای هر واحد آوایی) و ۳۳۰ فایل با پسوند lab (برچسب متناظر با هر فایل واحد آوایی) می‌باشد. همچنین دادگان آزمون حاوی فایل‌های گویندگان مبتلا به اوتیسم حاوی ۸۸ فایل با پسوند wav (۸ فایل به‌ازای هر واحد آوایی) و ۸۸ فایل با پسوند lab (برچسب متناظر با هر فایل واحد آوایی) است.

۳-۲ ارزیابی دادگان PersianSICHASD مبتنی بر HMM

برای ارزیابی دادگان PersianSICHASD از یک سیستم بازشناسی واحدهای آوایی مبتنی بر مدل مخفی مارکوف استفاده شده است. هدف اصلی از ارزیابی دادگان جمع‌آوری شده، تحلیل و بررسی تناسب واحدهای آوایی انتخاب شده در دادگان با سطح توانمندی کودکان مبتلا به اوتیسم است. برای انجام این تحلیل، اختلاف نرخ بازشناسی گفتار در کودکان سالم و کودکان مبتلا به اوتیسم مورد توجه است. برای بازشناسی گفتار می‌توانیم از روش‌های سنتی مانند مدل‌های مخفی مارکوف و ماشین

2. Accuracy
3. Correctness

1. Cool Edit

جدول ۵: بهترین خروجی حاصل از مقدار p متغیر بین ۱۰- تا ۱۲۰- به‌ازای وضعیت‌های ۴، ۶، ۸، ۱۰ و ۱۲ و تعداد توابع گوسی ثابت ۱۶ و تعداد ویژگی‌های MFCC ۳۹.

دقت	صحت	p	وضعیت	تعداد ضرایب ویژگی	تعداد مخلوط گوسی
۹۳/۶	۹۳/۹	-۱۲۰	۴	۳۹	۱۶
۹۲/۹	۹۳/۶	-۱۵۰	۶	۳۹	۱۶
۹۱/۲	۹۲/۲	-۱۲۰	۸	۳۹	۱۶
۸۹/۲	۸۹/۹	-۱۶۰	۱۰	۳۹	۱۶
۹۰/۵	۹۱/۲	-۱۳۰	۱۲	۳۹	۱۶

جدول ۶: بهترین خروجی حاصل از تعداد گوسی متغیر بین اعداد ۴، ۸، ۱۶ و ۳۲ به‌ازای تعداد وضعیت ثابت چهار، تعداد توابع گوسی ثابت ۱۶ و تعداد ضرایب ویژگی ثابت ۳۹.

دقت	صحت	تعداد مخلوط گوسی	p	وضعیت	تعداد ضرایب ویژگی
۹۳/۶۰	۹۴/۶۱	۴	-۱۲۰	۴	۳۹
۹۴/۲۸	۹۵/۶۲	۸	-۱۲۰	۴	۳۹
۹۳/۶۰	۹۳/۹۴	۱۶	-۱۲۰	۴	۳۹
۹۳/۲۷	۹۳/۹۴	۳۲	-۱۲۰	۴	۳۹

جدول ۷: بهترین خروجی حاصل از تغییر تعداد ضرایب ویژگی بین اعداد ۱۳، ۲۶، ۳۹ و ۵۲ به‌ازای تعداد وضعیت ثابت چهار، تعداد توابع گوسی ثابت ۸ و جریمه ثابت ۱۲۰-.

دقت	صحت	تعداد ضرایب ویژگی	تعداد مخلوط گوسی	p	وضعیت
۹۱/۵۸	۹۱/۹۲	۱۳	۸	-۱۲۰	۴
۹۲/۵۹	۹۳/۶۰	۲۶	۸	-۱۲۰	۴
۹۴/۲۸	۹۵/۶۲	۳۹	۸	-۱۲۰	۴
۹۳/۶۱	۹۵/۲۸	۵۲	۸	-۱۲۰	۴

مدل در جدول ۷ آمده است. با توجه به نتایج جدول، بهترین تعداد ضرایب ویژگی به‌ازای تعداد وضعیت چهار و تعداد توابع مخلوط گوسی هشت، تعداد ضرایب ویژگی برابر با ۳۹ معادل با ۱۲ ضریب اصلی کپستروم فرکانس مل و یک ضریب انرژی به همراه مشتقات اول یا دومشان می‌باشد. این تنظیم دارای دقت ۹۴/۲۸ درصد و صحت ۹۵/۶۲ درصد بر روی دادگان آزمون کودکان سالم است. در ادامه به ارزیابی دادگان برای هر دو گروه کودکان سالم و مبتلا به اوتیسم با استفاده از تکنیک ارزیابی متقابل چهاربخشی^۴ می‌پردازیم. لازم به ذکر است که معمولاً در ارزیابی رویکردهای مبتنی بر مدل مخفی مارکوف که روش مشخصی برای مقداردهی اولیه پارامترها دارند، از تکنیک ارزیابی متقابل k بخشی استفاده نمی‌شود. دلیل استفاده از این تکنیک در این مقاله برای ارزیابی مدل مخفی مارکوف، محدودبودن حجم دادگان و نگرانی نویسندگان مقاله از سوگیری پارامترهای مدل مخفی مارکوف به دادگان گویندگان انتخاب‌شده در مجموعه آموزش بوده است. با انتخاب تکنیک ارزیابی متقابل چهاربخشی، در هر بار آموزش مدل‌های مخفی مارکوف از دادگان گفتاری گویندگان متفاوتی در مجموعه آموزش استفاده شده و اطمینان بیشتری از صحت نتایج حاصل می‌گردد.

در این بخش، ۱۲ مدل مخفی مارکوف را با استفاده از تنظیمات حاصل از بهترین نتایج جدول ۷ جهت تشخیص افراد مبتلا به اوتیسم آموزش می‌دهیم. بخش مربوط به کودکان سالم در دادگان به چهار قسمت مساوی تقسیم شده و هر بار، سه قسمت از چهار قسمت برای آموزش ۱۲ مدل مخفی مارکوف و یک قسمت از چهار قسمت به همراه بخش مربوط به کودکان مبتلا به اوتیسم در دادگان برای ارزیابی استفاده شده است. نهایتاً برای تحلیل، نتایج مربوط به این چهار ارزیابی میانگین‌گیری شده‌اند. جدول ۸ نتایج حاصل از شیوه ارزیابی متقابل چهاربخشی را برای دادگان PersianSICHASD با استفاده از بهترین تنظیمات حاصل از جدول ۷ (تعداد ضرایب ویژگی ۳۹، وضعیت ۴، تعداد گوسی ۸ و جریمه ۱۲۰-) نشان می‌دهد.

نتایج نشان‌دهنده این موضوع است که افراد مبتلا به اوتیسم در بیان این مجموعه از واحدهای آوایی دارای اختلال هستند. در بین این واحدهای آوایی، نرخ بازشناسی گفتار واحدهای آوایی "ای" و "ی" دارای نرخ بازشناسی گفتار صفر درصد هستند که نشان از عدم توانایی کودکان مبتلا به اوتیسم در بیان این واحدهای آوایی است. دقت بازشناسی گفتار در [۱۵] که به‌عنوان پایه پژوهش ما به‌شمار می‌آید، دارای میانگین نرخ بازشناسی ۵۶ درصد برای افراد مبتلا به اوتیسم است؛ در حالی که HMM دارای میانگین دقت نرخ بازشناسی گفتار ۱۵/۶۸ درصد می‌باشد. افراد مبتلا به اوتیسم در ادای مجموعه واحدهای آوایی پیشنهادشده دارای اختلالاتی هستند؛ لذا انتظار می‌رود که نرخ بازشناسی گفتار برای افراد

سالم حاوی ۹۹ فایل ۹ گوینده اول و مجموعه آموزش حاوی ۲۴۲ فایل ۲۱ گوینده آخر می‌باشد و بدیهی است که گویندگان دو مجموعه آزمون و آموزش کاملاً مستقل از هم هستند. تعداد مدل‌های مخفی مارکوف شامل ۱۲ مدل به‌ازای ۱۱ واحد آوایی و سکوت می‌باشد. به‌منظور انتخاب تعداد بهینه برای وضعیت^۱، ۱۰۰ مجموعه مدل مخفی مارکوف (هر یک حاوی ۱۱ مدل برای واحدهای آوایی و یک مدل سکوت) با انتخاب وضعیت‌های ۴، ۶، ۸، ۱۰ و ۱۲ آموزش دادیم. در هر وضعیت مقدار جریمه (p) را به‌ازای اندازه مقیاس^۳ (s) برابر یک بین ۱۰- تا ۲۰۰- تغییر دادیم. در تمام این وضعیت‌ها تعداد ضرایب ویژگی ثابت و برابر با ۳۹ و همچنین تعداد توابع گوسی نیز ثابت و برابر با ۱۶ هستند. بهترین خروجی حاصل از مقدار p متغیر بین ۱۰- تا ۱۲۰- به‌ازای هر کدام از وضعیت‌ها در جدول ۵ قرار داده شده است. در بین وضعیت‌ها، دقت و صحت حاصل از تعداد وضعیت چهار دارای بیشترین مقدار است.

در مرحله بعد به‌منظور انتخاب تعداد بهینه برای توابع مخلوط گوسی در هر وضعیت، چهار مجموعه مدل مخفی مارکوف (هر یک حاوی ۱۱ مدل برای واحدهای آوایی و یک مدل سکوت) با انتخاب اعداد ۴، ۸، ۱۶ و ۳۲ برای تعداد مخلوط‌های گوسی با تعداد ویژگی‌های MFCC ۳۹ و تعداد وضعیت ثابت چهار آموزش یافته‌اند. نتایج حاصل از این ارزیابی در جدول ۶ آمده است. همچنین جهت انتخاب تعداد ویژگی‌های MFCC، چهار مجموعه مدل مخفی مارکوف (هر یک حاوی ۱۱ مدل برای واحدهای آوایی و یک مدل سکوت) شامل چهار دسته ویژگی مختلف (۱۲ ضریب مل کپستروم و یک ضریب انرژی و در مجموع ۱۳ ضریب) یا ۱۲ ضریب مل کپستروم و یک ضریب انرژی به همراه مشتقات اولشان (در مجموع ۲۶ ضریب) یا ۱۲ ضریب مل کپستروم و یک ضریب انرژی به همراه مشتقات اول یا دوم و سومشان (در مجموع ۵۲ ضریب) استخراج شده‌اند. نتایج حاصل از ارزیابی این چهار

1. State
2. Penalty
3. Scale

جدول ۸: نتایج حاصل از ارزیابی دادگان شامل بیان واحدهای آوایی با تکنیک ارزیابی متقابل چهاربخشی توسط افراد سالم و افراد مبتلا به اوتیسم با استفاده از بهترین تنظیمات حاصل از جدول ۷.

واحدهای آوایی پیشنهادشده	بخش اول (%)		بخش دوم (%)		بخش سوم (%)		بخش چهارم (%)	
	افراد سالم	افراد مبتلا به اوتیسم	افراد سالم	افراد مبتلا به اوتیسم	افراد سالم	افراد مبتلا به اوتیسم	افراد سالم	افراد مبتلا به اوتیسم
EY	۸۵٫۲	۰	۶۸٫۷۱	۰	۷۱٫۷۲	۰	۸۰	۰
AY	۷۷٫۸۱	۳۳	۷۵٫۷۱	۲۵	۷۴	۳۰	۷۹٫۷۱	۱۶٫۶۲
AXHH	۷۶٫۹۵	۰	۷۱٫۸۵	۰	۷۳	۰	۷۹	۰
IY	۷۵٫۶۳	۰	۷۴٫۷۱	۸٫۳۷	۷۶٫۷۱	۰	۷۵	۰
UWY	۸۱٫۹۷	۲۶	۷۹٫۲۸	۸٫۷۵	۸۴٫۱۴	۹٫۲	۸۰	۵٫۷۵
OWY	۸۹٫۸۶	۸	۸۷٫۴۲	۱۷٫۷۵	۸۳	۱۱٫۲۸	۷۹٫۵۷	۱۰
OWVAE	۸۰٫۰۶	۳۵٫۸	۸۱٫۷۱	۲۷	۷۶٫۴۲	۳۴٫۵۷	۷۶٫۴۲	۳۹٫۲۸
IYYAE	۸۰٫۱۷	۷	۷۹٫۱۴	۸٫۱۲	۸۰٫۷۱	۱۶٫۵	۷۸٫۴۲	۹٫۲۸
GAEPAEGAE	۸۴٫۴۷	۴۳٫۳	۸۴٫۲۵	۴۲٫۷۵	۷۶٫۴۲	۶۵	۷۸	۴۴٫۵
S	۵۳٫۷۷	۷٫۱	۶۱٫۲۸	۱۷	۵۹٫۴۲	۰	۶۰٫۸۵	۶٫۸۷
SH	۵۳٫۷۸	۱۵٫۶	۵۴	۲۴	۵۸٫۸۵	۱۸٫۱۴	۷۱٫۵۹	۱۷٫۵۷
میانگین	میانگین افراد سالم در بین ۴ بخش	۷۴٫۸۵	میانگین افراد مبتلا به اوتیسم در بین ۴ بخش	۱۵٫۶۸				

افراد سالم از افراد مبتلا به اوتیسم شامل جیتر، تن صدا و شیمیر می‌باشد. در مرحله بعد، از ویژگی‌های استخراج‌شده جهت آموزش دسته‌بند استفاده می‌شود. برای آموزش دسته‌بند از دسته‌بندهایی استفاده می‌شود که قابلیت ایجاد بیشترین تمایز مابین ویژگی‌های استخراج‌شده از دادگان آموزشی را داشته باشند. محبوب‌ترین روش‌های دسته‌بندی برای تفکیک افراد سالم و مبتلا به اوتیسم شامل ماشین بردار پشتیبان، شبکه عصبی پیچشی، شبکه عصبی عمیق، بیز ساده، رگرسیون لجستیک بیزی، جنگل تصادفی، کا نزدیک‌ترین همسایه و شبکه مولد تخصصی هستند. تحقیقات انجام‌شده در این زمینه نشان می‌دهند که ترکیب روش دسته‌بند مبتنی بر ماشین بردار پشتیبان با سایر روش‌های دسته‌بندی دارای بالاترین دقت است (۹۷/۱ درصد دقت برای تشخیص افراد مبتلا به اوتیسم و ۹۳/۵۸ درصد دقت برای تشخیص افراد سالم). خروجی حاصل از دسته‌بند به‌طور مستقیم یا غیرمستقیم جهت تفکیک افراد سالم از افراد مبتلا به اوتیسم مورد استفاده قرار می‌گیرد. در روش مستقیم بر اساس تحلیل واکنش حسی توسط افراد مورد بررسی نسبت به جملات، کلمات و ... به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته می‌شود. روش غیرمستقیم بر اساس دقت بازشناسی گفتار اداشده توسط افراد مورد بررسی به تفکیک افراد سالم از افراد مبتلا به اوتیسم می‌پردازد.

مطالعات انجام‌شده حاکی از آن است که تحقیقی در زمینه تشخیص اوتیسم از طریق گفتار برای زبان فارسی و یا کودکان ایرانی منتشر نشده که این کاستی می‌تواند به دلیل عدم وجود دادگان معتبر در این زمینه باشد. فضای خالی در این حوزه در کنار اهمیت بیشتر تشخیص بیماری اوتیسم در قیاس با گذشته، انگیزه‌های لازم جهت پرداختن به این موضوع را ایجاد کرده است. از این رو در ادامه، مرور و تحلیل پژوهش‌های انجام‌شده در حوزه تشخیص اوتیسم بر مبنای رویکردهای پردازش گفتار، بومی‌سازی این حوزه از طریق پرداختن به گام نخست آن یعنی معرفی دادگان محور اصلی کار قرار گرفت. بنابراین با مشورت و همفکری متخصص درمانگر در حوزه اوتیسم، مجموعه دادگانی طراحی و گردآوری شد. این دادگان شامل واحدهای آوایی است که افراد مبتلا به اوتیسم در بیان آنها دارای مشکل بوده یا قادر به تلفظ صحیح این واحدهای آوایی نیستند؛ در حالی که افراد سالم قادر به بیان درست و کامل آن واحدهای آوایی می‌باشند. مثلاً برای اولین بار از ترکیب واحد آوایی "پ" و واحد

مبتلا به اوتیسم کم و نزدیک به صفر باشد که نتایج تأییدکننده این واقعیت است. اختلاف میانگین دقت بازشناسی واحدهای آوایی میان افراد سالم و افراد مبتلا به اوتیسم حدود ۶۰ درصد (۵۹/۱۸) می‌باشد؛ لذا می‌توان با استفاده از همین اختلاف نرخ بازشناسی واحدهای آوایی بین افراد سالم و افراد مبتلا به اوتیسم، تفکیک دو گروه از یکدیگر را انجام داد. انجام این کار و ارائه سیستم تشخیص اوتیسم کودکان بر اساس پردازش گفتارشان، گام دوم از بومی‌سازی تشخیص اوتیسم کودکان ایرانی است که در پژوهش آتی به آن خواهیم پرداخت.

۴- جمع‌بندی

اوتیسم نوعی اختلال رشدی- مغزی به شمار می‌آید. افراد مبتلا به اوتیسم دارای اختلالاتی در لحن و نوع بیان واحدهای آوایی، کلمات و جملات و همچنین در مکالمات دو یا چندنفره و تک‌گویی‌هایی مانند داستان‌سرایی هستند؛ بنابراین تشخیص اوتیسم به‌خوبی از طریق رویکردهای پردازش گفتار انجام می‌شود. دو نوع رویکرد جهت تشخیص و تفکیک افراد سالم از افراد مبتلا به اوتیسم مبتنی بر رویکردهای پردازش گفتار استفاده می‌شود. در رویکرد اول بر اساس تحلیل نوع واکنش حسی افراد اعم از حس غم یا شادی و لحن صدا نسبت به داستان، جملات یا مکالمات، به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته می‌شود. در رویکرد دوم بر اساس دقت بازشناسی گفتار اداشده اعم از واحدهای آوایی، کلمات و یا جملات توسط افراد مشکوک به اوتیسم، به تفکیک افراد سالم از افراد مبتلا به اوتیسم پرداخته می‌شود. فرایند تشخیص افراد مبتلا به اوتیسم از طریق گفتار شامل دو مرحله آموزش و آزمون است. در مرحله آموزش، ابتدا دادگان آموزشی، وارد و سپس جهت استخراج ویژگی، ویژگی‌هایی با قابلیت ایجاد بیشترین تمایز میان دسته‌بندهای هدف از دادگان آموزشی استخراج می‌شود. برای استخراج ویژگی‌های گفتاری جهت تفکیک افراد مبتلا به اوتیسم از افراد سالم، ویژگی‌های آکوستیکی، ویژگی‌های عروضی یا ترکیبی از ویژگی‌های آکوستیکی و عروضی از گفتار استخراج می‌شود. ویژگی‌های آکوستیکی مورد استفاده جهت تفکیک افراد سالم از افراد مبتلا به اوتیسم شامل ضرایب مل کپستروم، نرخ گذار از صفر، فرمانت‌های اول تا سوم و انرژی است. همچنین ویژگی‌های عروضی مورد استفاده جهت تفکیک

- Learning and Knowledge Discovery in Databases*, pp. 193-208, 10-14 Sept. 2018.
- [8] S. Schelinski, Mechanisms of Voice Processing: Evidence from Autism Spectrum Disorder, Ph.D. Thesis, Humboldt University in Berlin, 2018.
- [9] A. Baird, et al., "Automatic classification of autistic child vocalisations: a novel database and results," in *Proc. InterSpeech'17*, pp. 849-853, Stockholm, Sweden, 20-24 Aug. 2017.
- [10] E. Lyakso, et al., "AD-Child. Ru: speech corpus for Russian children with atypical development," in *Proc. Int. Conf. on Speech and Computer, SPECOM'19*, pp. 299-308, Istanbul, Turkey, 20-25 Aug. 2019.
- [11] S. Sadiq, et al., "Deep learning based multimedia data mining for autism spectrum disorder (ASD) diagnosis," in *Proc. Int. Conf. on Data Mining Workshops, ICDMW'19*, pp. 847-854, Beijing, China, 8-11 Nov. 2019.
- [12] W. Liu, T. Zhou, C. Zhang, X. Zou, and M. Li, "Response to name: a dataset and a multimodal machine learning framework towards autism study," in *Proc. 7th Int. Conf. on Affective Computing and Intelligent Interaction, ACII'17*, pp. 178-183, San Antonio, TX, USA, 23-26 Oct. 2017.
- [13] K. Welarathna, V. Kulasekara, K. Pulasinghe, and V. Piyawardana, "Automated sinhala speech emotions analysis tool for autism children," in *Proc. 10th Int. Conf. on Information and Automation for Sustainability, ICIAS'21*, pp. 500-505, Negambo, Sri Lanka, 11-13 Aug. 2021.
- [14] D. Xu, et al., "Automatic childhood autism detection by vocalization decomposition with phone-like units," in *Proc. of the 2nd Workshop on Child, Computer and Interaction, WOCOCI '09*, Article ID: 5, 7 pp., Cambridge, MA, USA, 5-5 Nov. 2009.
- [15] L. G. Pillai and E. Sherly, "A deep learning based evaluation of articulation disorder and learning assistive system for autistic children," *International J. on Natural Language Computing*, vol. 6, no. 5, pp. 19-36, Oct. 2017.
- [16] J. Zhang, Y. Meng, C. Wu, Y. T. Xiang, and Z. Yuan, "Non-speech and speech pitch perception among cantonese-speaking children with autism spectrum disorder: an ERP study," *Neuroscience Letters*, vol. 703, pp. 205-212, Jun. 2019.
- [17] N. A. Chi, et al., "Classifying autism from crowdsourced semistructured speech recordings: machine learning model comparison study," *JMIR Pediatrics and Parenting*, vol. 5, Article ID: e35406, Apr. 2022.
- [18] A. Khozaei, H. Moradi, R. Hosseini, H. Pouretmad, and B. Eskandari, "Early screening of autism spectrum disorder using cry features," *PLoS One*, vol. 15, Article ID: e0241690, Dec. 2020.
- [19] T. Talkar, J. R. Williamson, D. J. Hannon, H. M. Rao, S. Yuditskaya, K. T. Claypool, et al., "Assessment of speech and fine motor coordination in children with autism spectrum disorder," *IEEE Access*, vol. 8, pp. 127535-127545, 2020.
- [20] A. Mohanta and V. K. Mittal, "Acoustic features for characterizing speech of children affected with ASD," in *Proc. IEEE 16th India Council Int. Conf., INDICON'19*, 4 pp., Rajkot, India, 13-15 Dec. 2019.
- [21] A. Mohanta, P. Mukherjee, and V. K. Mirtal, "Acoustic features characterization of autism speech for automated detection and classification," in *Proc. National Conf. on Communications, NCC'20*, 6 pp., Kharagpur, India, 21-23 Feb. 2020.
- [22] F. Ringeval, et al., "Automatic analysis of typical and atypical encoding of spontaneous emotion in the voice of children," in *Proc. 17th Annual Conf. of the Int. Speech Communication Association, ISCA'16*, pp. 1210-1214, San Francisco, CA, USA, 8-12 Sept. 2016.
- [23] I. F. Lin, et al., "Vocal identity recognition in autism spectrum disorder," *PLoS One*, vol. 10, Article ID: e0129451, Jun. 2015.
- [24] F. Ringeval, et al., "Automatic intonation recognition for the prosodic assessment of language-impaired children," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1328-1342, Oct. 2010.
- [25] M. Asgari, A. Bayestehtashk, and I. Shafran, "Robust and accurate features for detecting and diagnosing autism spectrum disorder," in *Proc. Annual Conf. of the Int. Speech Communication Association*, - pp. 191-194, 25-29 Aug. 2013.
- [26] E. Lyakso, et al., "Speech features of 13-15 year-old children with autism spectrum disorders," in *Proc. Int. Conf. on Speech and Computer, SPECOM'20*, pp. 291-303, St. Petersburg, Russia, 7-9 Oct. 2020.
- [27] S. R. Livingstone and F. A. Russo, "The ryerson audio-visual database of emotional speech and song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in north american english," *PLoS One*, vol. 13, Article ID: e0196391, May 2018.

آوایی "گ" (واحد آوایی "گِگ")، جهت تشخیص و تفکیک افراد سالم از افراد مبتلا به اوتیسم استفاده شده است. نتایج حاصل از درخواست بیان این واحد آوایی، نشانگر عدم توانایی کودکان مبتلا به اوتیسم در تلفظ این واحد آوایی و همچنین، تمام واحدهای آوایی شامل صامت‌هایی که افراد مبتلا به اوتیسم در تلفظ آنها مشکل دارند یا قادر به تلفظ آن واحدهای آوایی نیستند، است. مثلاً افراد مبتلا به اوتیسم در بیان صامت "ش" یا "س" به صورت کشیده به مدت ۵ ثانیه و یا تلفظ واحدهای آوایی که تداعی‌کننده خنده در صورت فرد هستند (مانند واحدهای آوایی ی) مشکل دارند. همچنین بیان ترکیباتی مانند "ایی"، "آی"، "اِه"، "ای"، "اُی" و "ای کشیده" برای آنها سخت است؛ لذا این افراد به‌سختی با دنیای پیرامون خود ارتباط برقرار می‌کنند. نتایج ارزیابی متقابل چهاربخشی روی این دادگان، مبتنی بر مدل مخفی مارکوف بیانگر آن است که میانگین نرخ بازشناسی گفتار واحدهای آوایی پیشنهادشده برای افراد مبتلا به اوتیسم ۱۵/۶۸ درصد است؛ در حالی که میانگین نرخ بازشناسی گفتار واحدهای آوایی پیشنهادی برای افراد سالم حدود ۷۵ درصد می‌باشد. اختلاف بین افراد سالم و افراد مبتلا به اوتیسم دارای مقدار قابل توجهی (۶۰ درصد) است؛ لذا از این اختلاف بارز نرخ بازشناسی واحدهای آوایی می‌توان جهت تشخیص و تفکیک افراد سالم از افراد مبتلا به اوتیسم استفاده کرد. انجام این کار و ارائه سیستم تشخیص اوتیسم کودکان بر اساس پردازش گفتارشان، گام دوم از بومی‌سازی تشخیص اوتیسم کودکان ایرانی است که در کارهای آتی به آن خواهیم پرداخت. همچنین باید این مسئله را در نظر گرفت که افراد مبتلا به اوتیسم دارای طیفی از ناهنجاری‌ها با شدت‌های مختلف در گفتار (ناتوانی در تلفظ بعضی از واحدهای آوایی، کلمات و ساختار جملات)، رفتار (حرکات ناهنجار دست، پا و چشم) و تعاملات اجتماعی (پاسخ‌های نادرست به جملات) و یا عدم توانایی استفاده از جملات و کلمات در جایگاه مناسب، عدم توانایی خندیدن هم‌زمان با گفتار و استفاده از کلمات پرکننده مانند "ایم" یا "اه" (زمانی که در حین گفتمان دونفره فکر می‌کنند و یا حرفی برای گفتن ندارند) هستند؛ لذا پیشنهاد می‌شود که از تلفیق صدا، حالات بدنی، نوع نگاه، حرکات بدن و تصاویر MRI از مغز فرد جهت بررسی دقیق‌تر این ناهنجاری استفاده شود.

مراجع

- [1] D. Mitsumoto, et al., "Autism spectrum disorder discrimination based on voice activities related to fillers and laughter," in *Proc. 53rd Annual Conf. on Information Sciences and Systems, CISS'19*, 6 pp., Baltimore, MD, USA, 20-22 Mar. 2019.
- [2] M. Alizadeh and S. Tabibian, "A Persian speaker-independent dataset to diagnose autism infected children based on speech processing techniques," in *Proc. 7th Int. Conf. on Signal Processing and Intelligent Systems, ICSPIS'21*, 5 pp., Tehran, Iran, 29-30 Dec. 2021.
- [3] S. Schelinski and K. V. Kriegstein, "Speech-in-noise recognition and the relation to vocal pitch perception in adults with autism spectrum disorder and typical development," *J. of Autism and Developmental Disorders*, vol. 50, no. 1, pp. 356-363, Jan. 2020.
- [4] S. Schelinski and K. V. Kriegstein, "The relation between vocal pitch and vocal emotion recognition abilities in people with autism spectrum disorder and typical development," *J. of Autism and Developmental Disorders*, vol. 49, pp. 68-82, 2019.
- [5] P. P. Denes, *The Speech Chain*, WH Freeman Company, 1993.
- [6] J. Deng, et al., "Speech-based diagnosis of autism spectrum condition by generative adversarial network representations," in *Proc. of the Int. Conf. on Digital Health*, pp. 53-57, Londres, UK, 2-5 Jul. 2017.
- [7] H. Drimalla, et al., "Detecting autism by analyzing a simulated social interaction," in *Proc. Joint European Conf. on Machine*

- [37] Z. Sherkatghanad, et al., "Automated detection of autism spectrum disorder using a convolutional neural network," *Frontiers in Neuroscience*, vol. 13, Article ID: 1325, Jan. 2020.
- [38] S. H. R. E. Motlagh, H. Moradi, and H. Pouretamad, "Using general sound descriptors for early autism detection," in *Proc. 9th Asian Control Conf., ASCC'13*, 5 pp., Istanbul, Turkey, 23-26 Jun. 2013.
- [39] A. Wijesinghe, P. Samarasinghe, S. Seneviratne, P. Yogarajah, and K. Pulasinghe, "Machine learning based automated speech dialog analysis of autistic children," in *Proc. 11th Int. Conf. on Knowledge and Systems Engineering, KSE'19*, 5 pp., Da Nang, Vietnam, 24-26 Oct. 2019.
- [40] M. Eni, et al., "Estimating autism severity in young children from speech signals using a deep neural network," *IEEE Access*, vol. 8, pp. 139489-139500, 2020.
- [28] R. Matin and D. Valles, "A speech emotion recognition solution-based on support vector machine for children with autism spectrum disorder to help identify human emotions," in *Proc. Intermountain Engineering, Technology and Computing, IETC'20*, 6 pp., Orem, UT, USA, 2-3 Oct. 2020.
- [29] C. Küpper, et al., "Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning," *Scientific Reports*, vol. 10, Article ID: 4805, 11 pp., 2020.
- [30] Y. K. Kim, et al., "Analyzing short term dynamic speech features for understanding behavioral traits of children with autism spectrum disorder," in *Proc. Interspeech'21*, pp. 2916-2920, Brno, Czech Republic, 30 Aug.-3 Sept. 2021.
- [31] B. Schuller, S. Steidl, and A. Batliner, "The Interspeech 2009 emotion challenge," in *Proc. Interspeech'09*, pp. 312-315, Brighton, UK, 6-10 Sept. 2009.
- [32] B. Schuller, et al., "The INTERSPEECH 2010 paralinguistic challenge," in *Proc. Interspeech'10*, pp. 2794-2797, Makuhari, Japan, 26-30 Sept. 2010.
- [33] B. Schuller, et al., "The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism," in *Proc. Interspeech'13*, pp. 148-152, Lyon, France, 25-29 Aug. 2013.
- [34] A. Pahwa, G. Aggarwal, and A. Sharma, "A machine learning approach for identification & diagnosing features of neurodevelopmental disorders using speech and spoken sentences," in *Proc. Int. Conf. on Computing, Communication and Automation, ICCCA'16*, pp. 377-382, Greater Noida, India, 29-30 Apr. 2016.
- [35] S. A. Majeed, H. Husain, S. A. Samad, and T. F. Idbeaa, "Mel frequency cepstral coefficients (MFCC) feature extraction enhancement in the application of speech recognition: a comparison study," *J. of Theoretical & Applied Information Technology*, vol. 79, no. 1, pp. 38-56, Sept. 2015.
- [36] A. Mohanta, P. Mukherjee, and V. K. Mirtal, "Acoustic features characterization of autism speech for automated detection and classification," in *Proc. National Conf. on Communications, NCC'20*, 6 pp., haragpur, India, 21-23 Feb. 2020.

مریم علیزاده تحصیلات خود را در مقطع کارشناسی مهندسی کامپیوتر- نرم افزار در سال ۱۳۹۴ به پایان رسانده است. ایشان مقطع کارشناسی ارشد خود را در رشته مهندسی فناوری اطلاعات- گرایش چندرسانه‌ای در سال ۱۳۹۷ در دانشگاه شهید بهشتی گذرانده است. در حال حاضر دانشجوی دکتری رشته مهندسی کامپیوتر- گرایش هوش مصنوعی و رباتیک در دانشگاه شهید بهشتی می‌باشد. زمینه تحقیقاتی مورد علاقه ایشان، پردازش سیگنال و گفتار است.

شیما طیبیان تحصیلات خود را در مقطع کارشناسی رشته مهندسی کامپیوتر- گرایش نرم‌افزار از دانشگاه صنعتی اصفهان در سال ۱۳۸۳ به پایان رسانده است. ایشان مقطع کارشناسی ارشد و دکتری خود را در رشته مهندسی کامپیوتر- گرایش هوش مصنوعی و رباتیک در سال‌های ۱۳۸۶ تا ۱۳۹۲ در دانشگاه علم و صنعت ایران گذرانده است. نام‌برده قبل از پیوستنش به دانشگاه شهید بهشتی در سمت استادیار و عضو هیات علمی پژوهشکده فضای مجازی در سال‌های ۱۳۹۳ الی ۱۳۹۶ استادیار پژوهشگاه هوافضا بوده است. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: بازشناسی گفتار، واژه-یابی گفتار، بهسازی گفتار، تشخیص فرامین صوتی، طراحی واسط‌های کاربری مبتنی بر گفتار، تشخیص احساس از گفتار، روش‌های یادگیری ماشین، پایش سلامت و روش‌های بهینه‌سازی.