

# کلاس‌بندی لهجه‌های فارسی با استفاده از شبکه‌های عصبی

اعظم ربیعی<sup>۱</sup>، سعید ستایشی<sup>۲</sup>

دانشگاه آزاد اسلامی واحد دولت‌آباد

azamrabiei@yahoo.com

## چکیده

این مقاله به کلاس‌بندی بعضی از لهجه‌های فارسی با کمک یک مدل مهندسی شامل پردازش، استخراج ویژگی‌ها و یک کلاس‌بندی کننده عصبی می‌پردازد. داده‌هایی از پنج لهجه مختلف فارسی جمع‌آوری شده و طی آزمایشات مختلف عملکرد کلاس‌بندی کننده عصبی در مقابل دو کلاس‌بندی کننده آماری دیگر بررسی شد. اثر افزایش تعداد کلاس‌ها در عملکرد کلاس‌بندی‌های مختلف لهجه نیز در این مقاله نشان داده شده است.

واژه‌های کلیدی: کلاس‌بندی لهجه - لهجه فارسی - شبکه عصبی مصنوعی - استخراج ویژگی

## مقدمه

فرهنگی یکسان مردمی است که در یک موقعیت جغرافیایی قرار دارند و ممکن است الگوهای گوناگونی را بصورت مشترک در گفتارشان داشته باشند. این الگوها شامل تلفظ‌های مختلف و فاکتورهای آکوستیکی شامل استفاده از صداهای ثابت و واکدار خاص و چگونگی تغییرشان وقتی با هم در یک کلمه یا گروهی از کلمات ترکیب می‌شوند است. این الگوها همچنین شامل موارد دیگری مثل استرس، سرعت گفتار و موزون‌بودن آن نیز است که همه این موارد در لهجه یک شخص با هم جمع می‌گردد [۲]. اینک نو و همکارش در سال ۲۰۰۷ [۱] اثر سابقه لهجه شنونده در درک و فهم لهجه را نشان داد. نتایج این تحقیق نشان می‌دهد که آشنایی قبلی شنونده با لهجه صدای گفته شده در تشخیص لهجه بسیار موثر است.<sup>۱</sup>

مدل‌های استفاده شده برای دسته‌بندی لهجه‌ها در تحقیقات انجام شده را می‌توان به دو دسته زیر تقسیم کرد:

- مدل‌های زیستی بر اساس سیستم شنوایی انسان

تشخیص گفتار اتوماتیک، یکی از مسائل قدیمی مطرح‌شده در علم کامپیوتر بوده است و پیاده‌سازی یک سیستم کارا برای انجام این کار، زمینه‌ی بسیاری از تحقیقات در پردازش سیگنال و هوش مصنوعی است. سیگنال گفتار علاوه بر متن گفته شده، حاوی اطلاعات و ویژگی‌های مهمی راجع به گوینده از جمله سن و جنسیت گوینده، احساسات و استرس، لهجه و گویش و سلامتی گوینده می‌باشد [۱]. دقت یک سیستم تشخیص گفتار، وقتی لهجه یا گویش گوینده با آنچه سیستم با آن آموزش دیده، متفاوت باشد بسیار کاهش می‌یابد. تشخیص درست یک لهجه گوینده می‌تواند برای کاهش نرخ خطا در سیستم‌های تشخیص گفتار بسیار کمک‌کننده باشد. لهجه، ویژگی‌های گفتاری یک زبان و الگوی تلفظ و تاکید است؛ در صورتی که گویش به گرامر و مجموعه لغات برمی‌گردد. هوانگ و همکارانش در سال ۲۰۰۷ [۵]، اولین حرکت در زمینه تشخیص گویش اتوماتیک را آغاز کردند؛ ولی در این مقاله ما به بررسی گویش گوینده نمی‌پردازیم. مواردی که در لهجه گوینده موثر است برگرفته از سابقه اجتماعی-

<sup>۱</sup> مربی

<sup>۲</sup> دانشیار-دانشگاه صنعتی امیرکبیر

- مدل‌های مهندسی شامل پیش‌پردازش، استخراج ویژگی‌ها و یک کلاس‌بندی کننده

به عنوان یک نمونه از مدل‌های زیستی برای دسته‌بندی لهجه‌ها می‌توان به مرجع [۳] اشاره کرد. در این مرجع، تشخیص لهجه، بر اساس مدل مجرای صوتی در لهجه‌های مختلف و تصویری که از مجرای صوتی در ادای هر آوا در دسترس است، انجام شده است. ما در این تحقیق، مدل‌های زیستی را بررسی نخواهیم کرد و در ادامه به مرور مدل‌های مهندسی در تشخیص لهجه خواهیم پرداخت. جدول ۱ تحقیقات انجام شده در کلاس‌بندی لهجه‌ها، در سال‌های اخیر را به همراه روش استفاده شده در آن‌ها نشان می‌دهد.

#### جدول ۱- تحقیقات انجام شده در کلاس‌بندی لهجه

مرجع	ویژگی‌های استفاده شده	راه حل
[۴]	MFCC، انرژی و مشتقات اولشان	HMM
[۲]	MFCC، انرژی و مشتقات اول و دومشان	SVM و یادگیری درخت تصمیم
[۶]	MFCC، انرژی و سه فرکانس اول فرمنت	یک معیار فاصله
[۷]	MFCC، انرژی و سه فرکانس اول فرمنت	یک کلاس‌بندی کننده فازی-گوسی
[۸]	استفاده از ویژگی‌های کلامی علاوه بر ویژگی‌های صوتی	GMM و SVM
[۹]	فرکانس‌های فرمنت دوم و سوم	GMM
[۱۰]	انرژی، فرکانس پایه و سه فرکانس اول فرمنت، پهنای باندهای سه فرکانس اول فرمنت و مشتقات اول و دومشان	HMM‌های کلاس-آوایی
[۱۱]	فاصله بین توقف شل و توقف کامل، طول کلمه، زیر و بمی صدا و فرکانس‌های فرمنت دوم و سوم	SVM و HMM

ارسلان و هانسن در سال ۱۹۹۶ [۴] با ارائه یک سیستم کلاس‌بندی لهجه برای زبان انگلیسی یکی از برجسته‌ترین و اولین تحقیقات انجام شده در این زمینه را انجام دادند. این تحقیق نشان داد که هر چه طول و تعداد کلمات گفته شده بیشتر باشد، دقت سیستم کلاس‌بندی گفتار بیشتر می‌شود. این مرجع همچنین نشان داد بازدهی در سیستم تشخیص گفتار با قابلیت دسته‌بندی لهجه‌ها بطور چشمگیری بیشتر از سیستم تشخیص گفتار بدون شناخت

لهجه است. پدیرسن و دیدریچ در سال ۲۰۰۷ [۲] کلاس‌بندی لهجه را با کمک SVM<sup>۱</sup> و یادگیری درخت تصمیم برای دو لهجه عربی و هندی از زبان انگلیسی انجام داد. در این مطالعه، کلاس‌بندی بر اساس متوسط ویژگی‌ها روی چند فریم کنار هم انجام شد. به منظور از بین بردن تفاوت‌های گفتارها و برای یکسان‌سازی، از تفاضل متوسط کپسترال و نرمال‌سازی انرژی در این تحقیق استفاده شده است. اولاً و کاری در سال ۲۰۰۷ [۶]، کلاس‌بندی دیگری از گفتار بر اساس یک معیار فاصله انجام دادند. ایده این روش از آنجا گرفته شده است که زمانی که یک گوینده غیر بومی شروع به یادگیری زبان دومی می‌کند، تلفظ آوای زبان مادری خود را به زبان جدید القا می‌کند. این جایگذاری منجر به ابهاماتی بین مرزهای آواها می‌گردد و شباهت بین آوای مختلف را افزایش می‌دهد. این روش از این اطلاعات برای انتقال نقاط داده به فضایی که فاصله اقلیدسی بین شباهت‌ها حداقل و بین تفاوت‌ها حداکثر است استفاده می‌کند. اولاً و کاری همچنین در تحقیقات دیگری [۷]، همین کار را با یک کلاس‌بندی کننده گوسی فازی (FGMM)<sup>۲</sup> انجام دادند. فاریا در سال ۲۰۰۵ [۸]، تلاش دیگری برای کلاس‌بندی لهجه‌ها را روی پایگاه داده Fisher<sup>۳</sup> با متدهای GMM و SVM انجام داد. در این تحقیق علاوه بر ویژگی‌های صوتی گفتار، از ویژگی‌های کلامی<sup>۴</sup> گفتار نیز استفاده شده است. ایده استفاده از ویژگی‌های کلامی بر اساس این فرضیه است که گویندگان غیر بومی دنباله کلماتی را تولید می‌کنند که اساساً متفاوت از زبان تولید شده توسط یک گوینده بومی است. این تحقیق نشان داد که افراد غیر بومی، مجموعه لغات محدودتری را استفاده می‌کنند و گفتارشان قابل پیش‌بینی‌تر است. در تحقیق ساده دیگری از دشپنده و همکارانش [۹]، ابتدا بخش‌های سکوت از گفتار حذف شده‌اند و پس از استخراج ویژگی‌ها، کلاس‌بندی دو گروه زبان انگلیسی امریکایی و هندی با کمک GMM انجام شده است. فانگ و کت نیز در سال ۱۹۹۹ [۱۰]، یک روش ترکیبی، هم بر اساس ویژگی و هم بر اساس مدل را ارائه

<sup>۱</sup> Support Vector Machine

<sup>۲</sup> Fuzzy Gaussian Mixture Model

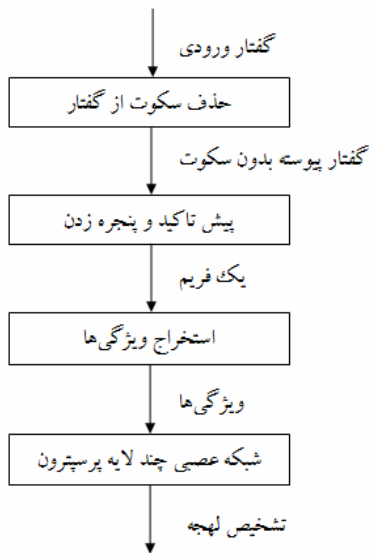
<sup>۳</sup> <http://www ldc.upenn.edu/projects/ears>

<sup>۴</sup> Lexical

تحقیق ارائه خواهد شد و نهایتاً به مباحث مربوط به پیاده‌سازی و نتایج آزمایشات خواهیم پرداخت.

### مدل استفاده شده در این تحقیق

تشخیص و کلاس‌بندی صداهای لهجه‌دار در این تحقیق، با کمک شبکه‌های عصبی مصنوعی انجام شد. بلوک دیاگرام مدل استفاده شده در این مقاله در شکل ۱ نشان داده شده است.



شکل ۵- مدل استفاده شده در این مقاله

#### • حذف سکوت از گفتار

گفتار استفاده شده برای این کلاس‌بندی، یک گفتار پیوسته است. این گفتارهای پیوسته از حذف فریم‌های سکوت از گفتار ورودی بدست می‌آید. منظور از فریم سکوت فریمی است که انرژی آن از ۰/۱۵ انرژی کل گفتار کمتر باشد.

#### • پیش‌تاکید و پنجره زدن

منظور از مرحله پیش‌تاکید، استفاده از فیلتر  $1 - \alpha \cdot Z^{-1}$  است که در آن معمولاً  $\alpha = 0.97$  در نظر گرفته می‌شود. با کمک این فیلتر، فرکانس‌های فرمنت و دیگر ویژگی‌ها مشخص‌تر بدست می‌آیند. فریم‌های ۲۵ میلی‌ثانیه‌ای گفتار پیوسته در این آزمایش با شیفت ۱۰ میلی‌ثانیه‌ای از یکدیگر تفکیک شده‌اند؛ به عبارت دیگر هر فریم با فریم

کردند. برای کلاس‌بندی سریع لهجه‌ها با داده‌های کم، به‌جای استفاده از HMM بر اساس آوا برای تشخیص، این مقاله، آموزش HMM های کلاس-آوایی را پیشنهاد می‌کند. مجموعه آواها در این تحقیق به ۶ کلاس (توقفی، سایشی- انسدادی، سایشی، تودماغی، نیمه‌صدادار و صدادار) تقسیم شده است.

در مورد لهجه‌های غیر انگلیسی تحقیقات بسیار محدودی صورت گرفته است [۱۲] و [۱۳]. در مورد زبان فارسی نیز هنوز سیستم تشخیص اتوماتیک گفتاری با بازدهی بالا وجود ندارد؛ در ایران در حال حاضر، تعدادی آزمایشگاه پردازش گفتار در دانشگاه‌های شریف [۱۴]، امیرکبیر [۱۵]، علم و صنعت [۱۶] و یزد [۱۷] وجود دارد. نتایج تحقیقات انجام شده در این آزمایشگاه‌ها مقالات متعددی در مجلات و همایش‌های داخلی و خارجی متعدد بوده است. در بین این تحقیقات ایرانی، تحقیقی مستقیماً بر روی لهجه‌های فارسی انجام نشده است. تنها نتایج بعضی از این تحقیقات بر روی تشخیص زبان [۱۸] ممکن است تا حدی به این تحقیق نزدیک باشد.

از مرور اجمالی که بر تحقیقات انجام شده در این زمینه تاکنون داشتیم نکات زیر قابل توجه است:

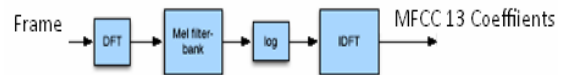
- ۱- از آنجا که ضرایب MFCC در بسیاری از مسائل تشخیص گوینده استفاده شده و نتایج خوبی را نشان داده، در تشخیص لهجه نیز از آن استفاده شده است.
- ۲- سه فرکانس فرمنت اول که مربوط به حرکت زبان است، بیشترین تغییرات آواها در هر زبانی را نشان می‌دهند و در لهجه نقش مهمی ایفا می‌کند.
- ۳- بسیاری از تحقیقات علاوه بر خود ویژگی‌ها، از تغییرات ویژگی‌ها (مشتقات اول و دوم) نیز استفاده می‌کنند.
- ۴- بهترین کلاس‌بندی کننده‌های قابل انعطاف و موثر استفاده شده برای کلاس‌بندی لهجه تاکنون HMM و SVM هستند.

با توجه به خاصیت تطبیق‌پذیری متدهای وفقی در این تحقیق، از شبکه‌های عصبی برای کلاس‌بندی ویژگی‌ها، استفاده شده است. در ادامه، مدل استفاده شده در این

بعدی و قبلی ۱۵ میلی‌ثانیه همپوشانی دارد. فریم‌های گفتار در پنجره همینگ نیز ضرب شده‌اند.

### • استخراج ویژگی‌ها

ویژگی‌های در نظر گرفته شده، شامل انرژی، F2، F3 و ۱۳ ضریب فرکانسی MFCC هستند. فرکانس‌های فرمنت با کمک ضرایب LPC فریم بدست آمده‌اند و ضرایب MFCC نیز طبق بلوک‌دیگرام شکل ۲ محاسبه شده‌اند.



### شکل ۲- مراحل استخراج ویژگی MFCC از گفتار

از آنجا که فریم‌بندی بر اساس زمان صورت گرفته است، برای فشردن داده‌ها، ویژگی‌های استخراج شده در هر چند فریم، میانگین‌گیری می‌شوند، تا ویژگی‌ها بر اساس آوا باشند.

### • شبکه عصبی چندلایه پرسپترون

از یک شبکه عصبی دو لایه پیشرو پرسپترون برای کلاس‌بندی بردارهای ویژگی بدست آمده، می‌توان استفاده کرد. تعداد نرون‌های لایه ورودی برابر اندازه بردار ویژگی و برابر ۱۶ تاست. تعداد نرون‌های لایه مخفی با آزمون و خطا تعیین می‌شود. تعداد نرون‌های لایه آخر نیز برابر تعداد لهجه‌ها یا تعداد کلاس‌هاست. خروجی مطلوب برای نرون  $i$  ام، اگر الگوی ورودی مربوط به کلاس  $i$  ام باشد، برابر ۱ و در غیر اینصورت برابر ۰-۱ است. تابع انتقال لایه مخفی را  $\text{tansig}$  و لایه آخر را خطی در نظر گرفتیم؛ ضمناً از الگوریتم یادگیری گرادینت توام<sup>۱</sup> که یکی از بهترین متدهای یادگیری بر اساس گرادینت نزولی مرتبه دو است، برای آموزش شبکه استفاده شد.

### پیاده‌سازی و نتایج تجربی

جامعه آماری این تحقیق، نمونه فایل‌های صوتی گفتارهای لهجه‌داری هستند که شامل بیشترین آواها یا اصطلاحات و

کلمات فارسی باشند. برای این منظور مجموعه اصطلاحات و عبارات زیر در نظر گرفته شده است. دقت کنید که فایل‌های صوتی درگیر با گویش‌ها نمی‌شود و صرفاً عبارات زیر در لهجه‌های مختلف ادا می‌گردد:

«کجا داری می‌ری؟ - کتابمو بُرد - گنج قارون - ژاکت ژاله - چند سالتِه؟ - سوپر آقا مجید - ظاهر زیبا - خط میخی»

داده‌های جمع‌آوری شده در هر لهجه، از سه زن و سه مرد میانسال (نه کودک و نه پیر) جمع‌آوری شده است. ابزار استفاده شده برای تجزیه و تحلیل داده‌ها و پیاده‌سازی، نرم‌افزار Matlab است. از آنجا که بطور متوسط، هر ۷ فریم مجاور، ویژگی‌های تقریباً یکسانی دارند و مربوط به یک آوا هستند، داده‌های بدست آمده از مرحله استخراج ویژگی‌ها در هر ۷ فریم میانگین‌گیری شده‌اند. این عدد توسط آزمون و خطا بدست آمده است. برای یکسان‌سازی در آزمایشات مختلف آموزش شبکه‌های عصبی، مقادیر این ویژگی‌ها، بین ۰-۱ تا ۱ نرمالایز شده‌اند و ضمناً شبکه عصبی آموزش را با یک مجموعه وزن‌های اولیه تصادفی یکسان شروع می‌کند.

در اولین آزمایش، سه لهجه‌ی تهرانی، اصفهانی و کرمانشاهی در نظر گرفته شد. از ۳۰٪ ویژگی‌های استخراج شده، برای آموزش استفاده شد. ۷۰٪ از کل داده‌ها (شامل داده‌های قبلاً آموزش داده شده و جدید) را بصورت تصادفی انتخاب و برای تست در نظر گرفتیم. جدول ۲ نتایج بازدهی این آموزش را نشان می‌دهد. توجه کنید که چون داده‌ها در مرحله تست بصورت تصادفی انتخاب می‌شوند، این آزمایش برای هر لهجه ۱۰ بار تکرار شده و ارقام جدول، میانگین نتایج این ۱۰ بار آزمایش هستند.

### جدول ۲- نتایج بازدهی آموزش شبکه برای سه لهجه

لهجه	لهجه	لهجه	
اصفهانی	کرمانشاهی	تهرانی	
۹۷/۸۵٪	۹۸/۵۱٪	۹۷/۵۳٪	بازدهی روی دادگان آموزشی
۹۶/۵۴٪	۹۷/۷۵٪	۹۵/۵۸٪	بازدهی روی دادگان تست

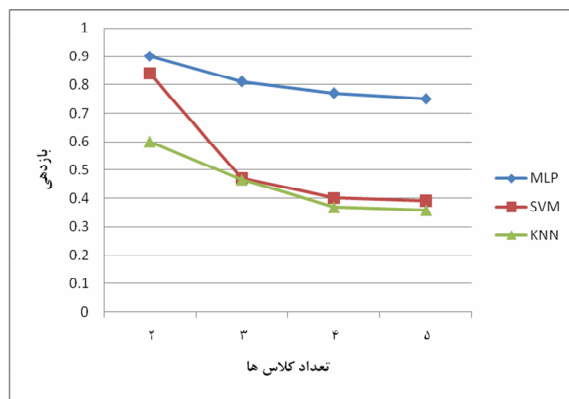
<sup>۱</sup> Conjugate Gradient

جدول نشان می‌دهد که شبکه‌های عصبی قدرت تفکیک بعضی از لهجه‌ها را تا حتی ۹۶٪ دارد.

#### جدول ۴- قدرت تشخیص لهجه‌ها از یکدیگر

مرحله آموزش	مرحله تست	
۹۸/۳۳٪	۸۷/۳۹٪	تشخیص لهجه تهرانی از اصفهانی
۹۸/۶۵٪	۸۸/۵۴٪	تشخیص لهجه تهرانی از کرمانشاهی
۹۸/۰۵٪	۹۶/۲۷٪	تشخیص لهجه تهرانی از کرمانی
۹۸/۳۲٪	۹۵/۲۹٪	تشخیص لهجه تهرانی از چهارمحالی
۹۹/۲٪	۸۴/۰۷٪	تشخیص لهجه اصفهانی از کرمانشاهی
۹۹/۱۴٪	۸۹/۳۸٪	تشخیص لهجه اصفهانی از کرمانی
۹۸/۶۷٪	۹۴/۵۲٪	تشخیص لهجه اصفهانی از چهارمحالی
۹۸/۸۹٪	۸۴/۸۵٪	تشخیص لهجه کرمانشاهی از کرمانی
۹۸/۸۵٪	۸۹/۱۷٪	تشخیص لهجه کرمانشاهی از چهارمحالی
۹۸/۴۵٪	۸۶/۹۳٪	تشخیص لهجه کرمانی از چهارمحالی

در آخرین آزمایش، اثر افزایش تعداد کلاس‌ها در عملکرد کلاس‌بندی‌های مختلف لهجه توسط MLP، SVM و KNN بررسی شد. شکل ۳ نمودار میانگین بازدهی کلاس‌بندی‌های مختلف بازای افزایش تعداد کلاس‌ها را نشان می‌دهد.



شکل ۳- مقایسه میانگین بازدهی کلاس‌بندی‌های مختلف لهجه بازای افزایش تعداد کلاس‌ها

با افزایش تعداد کلاس‌ها، خطای کلاس‌بندی افزایش می‌یابد؛ این افزایش در کلاس‌بندی‌کننده MLP، کمتر و در دو کلاس‌بندی‌کننده دیگر بیشتر است. علت کاهش سریع بازدهی SVM از دو کلاس به سه کلاس این است که اساساً SVM، روشی برای دسته‌بندی دو کلاس است و برای

در دومین آزمایش، به جای کلاس‌بندی‌کننده وفقی شبکه‌های عصبی از کلاس‌بندی‌کننده‌های آماری SVM<sup>۱</sup> و KNN<sup>۲</sup> (نزدیکترین همسایه) استفاده کردیم. آزمایشات مختلف برای هر سه کلاس‌بندی‌کننده روی یک مجموعه دادگان تست ثابت انجام شد. جدول ۳ نتایج بازدهی کلاس‌بندی این آزمایشات را نشان می‌دهد. توجه کنید که این بازدهی‌ها تا ۳۳٪، شانس هستند (سه کلاس).

#### جدول ۳- بازدهی کلاس‌بندی‌کننده‌های مختلف

میانگین	کرمانشاهی	اصفهانی	تهرانی	
۸۱/۱۵٪	۸۱/۹۹٪	۸۶/۹۵٪	۷۴/۵٪	MLP
۴۷/۰۳٪	۳۹/۷۱٪	۴۹/۲۴٪	۵۲/۱۵٪	SVM
۴۷/۲۲٪	۵۲/۸۷٪	۵۲/۹۳٪	۳۵/۸۶٪	KNN1
۴۶/۰۲٪	۵۲/۴۴٪	۵۳/۲۸٪	۳۲/۳۳٪	KNN2

در روش SVM، آزمایشات زیادی با توابع مختلف جداکننده کلاس‌ها (خطی، چند جمله‌ای، RBF<sup>۳</sup> و ...) انجام شد و نهایتاً بهترین بازدهی‌ها در جدول ۳ قرار گرفتند. در مورد روش KNN نیز آزمایشات مختلف با معیارهای فاصله متفاوت (مربع اقلیدسی، قدر مطلق اندازه فاصله‌ها، همبستگی و ...) انجام شد و دو تا از بهترین نتایج در جدول ۳ نشان داده شده‌اند. این جدول نشان می‌دهد که بهترین نتایج مربوط به شبکه‌های عصبی است.

در آزمایش بعدی، قدرت تشخیص لهجه‌های مختلف از یکدیگر توسط شبکه‌های عصبی بررسی شد. در این آزمایش، تمام حالت‌های مختلف تشخیص پنج لهجه اصفهانی، تهرانی، کرمانشاهی، کرمانی و چهارمحالی از یکدیگر توسط یک کلاس‌بندی‌کننده دوتایی شبکه عصبی تست شد. داده‌های این جدول نیز میانگین ۱۰ بار انجام آزمایش روی داده‌های تست مختلف است. جدول ۴ نتایج بازدهی این آزمایشات را نشان می‌دهد. از آنجایی که بازدهی مرحله آموزش شبکه‌های عصبی در تشخیص لهجه‌های جدول ۴ تقریباً یکسان هستند، بازدهی‌های مرحله تست با تقریب خوبی قابل مقایسه هستند. این

<sup>1</sup> Support Vector Machine

<sup>2</sup> K-Nearest Neighbor

<sup>3</sup> Radial Basis Function

- 6- Ullah S. and Karray F., "Speaker Accent Classification Using Distance Metric Learning Approach", *IEEE International Symposium on Signal Processing and Information Technology*, 2007.
- 7- Ullah S. and Karray F., "Speaker Accent Classification System Using A Fuzzy Gaussian Classifier", *Conference on Information and Emerging Technologies(ICIET 2007)*, July 2007.
- 8- Faria A., "Accent Classification for Speech Recognition", *Machine Learning for Multimodal Interaction*, pp.285-293, 2005.
- 9- Deshpande Sh., Chikkerur Sh. and Govindaraju V., "Accent Classification in Speech", *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pp.139-143, 2005.
- 10-Fung P. and Kat L. W., "Fast Accent Identification and Accented Speech Recognition", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, pp.221 – 224, Mar 1999.
- 11-Tang H and Ghorbani A., "Accent Classification using Support Vector Machine and Hidden Markov Model", *Advances in Artificial Intelligence Processings*, Vol.2671, pp.629-631, 2003.
- 12-Zheng Y., Sproat R., Guy L., Shafranz I., Zhouz H., Suz Y., Jurafsky D., Starr R. and Yoon S. , "Accent Detection and Speech Recognition for Shanghai-Accented Mandarin", <http://www.stanford.edu/~jurafsky/>.
- 13-Kumpf K. and King R.W., "Automatic Accent Classification of Foreign Accented Australian English Speech", *Fourth International Conference on Spoken Language Proceedings (ICSLP)* Vol. 3, pp.1740-1743, Oct 1996.
- 14-<http://ce.sharif.edu/~spl/>
- 15-<http://ele.aut.ac.ir/~sprl/>
- 16-<http://aspl.iust.ac.ir/>
- 17-<http://pweb.yazduni.ac.ir/pages/Faculties/engineering/elec/SPRL/>
- 18-Ziaei A., Ahadi S.M. and Yeganeh H., "Enhanced Spectral Features for Spoken Language Identification," in *Proc. International Conference on Signal Processing (ICSP08)*, Beijing, 2008.

استفاده‌ی چندکلاسی از آن یکی از این دو تکنیک استفاده می‌شود: یک کلاس در مقابل کل بقیه کلاس‌ها قرار داده شود و این کار برای تمام کلاس‌ها تکرار شود یا کلاس‌ها دو به دو در مقابل هم قرار گیرند.

## نتیجه‌گیری

در این مقاله کلاس‌بندی چند لهجه فارسی با کمک یک مدل مهندسی شامل پیش‌پردازش، استخراج ویژگی‌ها و کلاس‌بندی‌کننده انجام شد. مطالعات انجام شده در این تحقیق، نشان داد که بعضی از ویژگی‌های گفتار مثل انرژی، فرکانس‌های F2، F3 و ضرایب فرکانسی مل نقش مهم‌تری در تشخیص لهجه گوینده دارند. نتایج آزمایشات این مقاله نشان می‌دهد که کلاس‌بندی‌کننده وقتی شبکه عصبی نسبت به دو کلاس‌بندی‌کننده آماری SVM و KNN عملکرد بهتری در تشخیص لهجه دارد. همچنین خطای کلاس‌بندی در شبکه‌های عصبی با افزایش تعداد کلاس‌ها کاهش کمتری نسبت به دو کلاس‌بندی‌کننده دیگر دارد.

## مراجع

- 1- Ikeno A. and Hansen H. L. J., "The Effect of Listener Accent Background on Accent Perception and Comprehension", *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2007, Article ID 76030, 8 pages, 2007.
- 2- Pedersen C. and Diederich J., "Accent Classification Using Support Vector Machines", *6th IEEE/ACIS International Conference on Computer and Information Science*, ICIS 2007.
- 3- Angkitittrakul P., and Hansen H.L. J., "Advances in Phone-Based Modeling for Automatic Accent Classification", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, March 2006.
- 4- Arslan M. L. and Hansen H.L. J., "Language Accent Classification in American English", *Speech Communication*, Revised January 29, 1996.
- 5- Huang R., Hansen H.L. J., and Angkitittrakul P., "Dialect/Accent Classification Using Unrestricted Audio", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, February 2007.