

# تشخیص لهجه‌های گفتار زبان فارسی از روی سیگنال گفتار با استفاده از روش‌های استخراج ویژگی کارآمد و ترکیب طبقه‌بندها

مجتبی شریف نوقابی\*، حسین مروی و دانیال دارابیان

گروه الکترونیک، دانشکده مهندسی برق و رباتیک، دانشگاه صنعتی شاهرود، شاهرود، ایران

## چکیده

تشخیص لهجه از روی شکل موج گفتار یکی از شاخه‌های تاحدودی جدید در علم پردازش گفتار است. تشخیص لهجه می‌تواند تا حد زیادی باعث بهبود سامانه‌های بازساخت گفتار شود. همانند هر سامانه بازشناسی، فرآیند تشخیص لهجه نیز شامل سه مرحله پیش‌پردازش، استخراج ویژگی و طبقه‌بندی است. در این مقاله سه روش کارآمد استخراج ویژگی شامل دامنه مرکزی طیفی (SCM)، مشتق نخست آن ( $\Delta SCM$ ) و تبدیل Zak روی سیگنال گفتار اعمال شده‌اند و کارایی این روش‌ها با روش‌های متدوال استخراج ویژگی مانند ضرایب مل-کپستروم و مشتقات آن مقایسه شده است. علاوه بر این یک ویژگی جدید که با ایجاد تغییراتی در الگوریتم محاسبه ضرایب مل-کپستروم به دست می‌آید برای تشخیص لهجه‌ها در محیط‌های نوفه‌ای معرفی شده است. برای مرحله طبقه‌بندی از پنج طبقه‌بند مختلف، شامل MLP، SVM و RBF، PNN، KNN و ترکیب این طبقه‌بندها با یکدیگر استفاده شده است. نتایج آزمایش‌ها بیان‌گر بهبود نرخ بازشناسی لهجه‌ها با روش‌های پیشنهادی هستند.

واژگان کلیدی: دامنه مرکزی طیفی، ترکیب طبقه‌بندها، لهجه فارسی، ماشین بردار پشتیبان، ضرایب مل-کپستروم بهبود یافته.

## ۱- مقدمه

در زبان‌شناسی به نحوه تلفظ کلمات در یک گروه خاص از افراد یک مکان یا ملیت لهجه می‌گویند. مواردی که در لهجه یک گوینده مؤثر است برگرفته از سابقه اجتماعی-فرهنگی یکسان مردمی است که در یک موقعیت جغرافیایی قرار دارند و ممکن است الگوهای گوناگونی را به صورت مشترک در گفتارشان داشته باشند.

فرآیند بازساخت خودکار گفتار یکی از زمینه‌های دانش هوش مصنوعی است. پژوهش‌ها نشان می‌دهد که تغییرات گفتار که مربوط به گوینده است، شامل تغییر در جنسیت، سن، لهجه، سلامتی و یا عدم سلامتی گوینده از دلایل تنزل کارایی سامانه‌های بازشناسی گفتار هستند. که از میان این ویژگی‌ها دو ویژگی جنسیت و لهجه، بیشترین تأثیر را در این کاهش بازدهی دارند. هنگامی که یک سامانه تشخیص گفتار با یک لهجه خاص آموزش ببیند، آن‌گاه با

لهجه‌ای غیر از لهجه‌ای که با آن آموزش دیده برخورد کند، شاهد کاهش چشم‌گیر بازدهی آن سامانه خواهیم بود. بنابراین اهمیت این مسأله ما را به پیاده‌سازی سامانه‌ای برای تشخیص لهجه‌ها از روی شکل موج گفتار ترغیب می‌کند. به‌طور معمول برای تشخیص جنسیت، از مدل‌های مربوط به خود آن استفاده می‌شود. برای تشخیص لهجه نیز مدل‌های مختلفی ارائه شده است که روند کلی این مدل‌ها شامل سه مرحله استخراج ویژگی، طبقه‌بندی و تشخیص می‌باشد.

اگر نگاهی گذرا به تاریخچه و سابقه تشخیص لهجه‌ها از طریق گفتار داشته باشیم، با دسته‌بندی مشاهدات در دو حوزه لهجه‌های خارجی و فارسی به این نتیجه می‌رسیم که در حوزه لهجه‌های خارجی پژوهش‌های بیشتری انجام شده است که در این میان بر روی تشخیص لهجه‌های انگلیسی

## ۲- سامانه تشخیص لهجه

همان‌طور که در قبل بیان شد هر سامانه بازشناسی لهجه همانند هر سامانه خودکار بازشناسی گفتار شامل سه مرحله پیش‌پردازش، استخراج ویژگی و طبقه‌بندی است. در مرحله پیش‌پردازش به‌طور معمول عملیاتی مانند حذف قسمت سکوت از گفتار انجام می‌شود. این کار با روش‌هایی مانند نرخ عبور از صفر و تعیین آستانه انرژی انجام می‌شود. در نهایت با اعمال یک فیلتر پیش‌تأکید این مرحله به پایان می‌رسد.

## ۳- روش‌های متداول استخراج ویژگی

مهم‌ترین بخش یک سامانه تشخیص لهجه استخراج ویژگی است. در مرحله استخراج ویژگی باید به دو نکته توجه بیشتری داشت. نخست این که باید قسمت‌هایی از سیگنال گفتار انتخاب شود که شامل بیشترین اطلاعات درباره لهجه آن باشد و نکته دوم این است که ویژگی استخراج شده باید قادر باشد اطلاعات مناسبی در مورد لهجه گفتار به ما ارائه کند و ما با استفاده از این ویژگی در مرحله طبقه‌بندی شاهد جداسازی بین لهجه‌های مختلف به بهترین نحو باشیم.

### ۳-۱- ضرایب مل-کپستروم (MFCC)

یکی از ویژگی‌های متداول که از سیگنال گفتار استخراج و در اغلب سیستم‌های تشخیص گفتار و گوینده از آن استفاده می‌شود، ضرایب مل-کپستروم (MFCC)<sup>۱</sup> است. مراحل محاسبه این ضرایب در شکل (۱) نشان داده شده است.

پس از مرحله پیش‌پردازش به دلیل غیرایستاد بودن سیگنال گفتار، آن را قاب‌بندی می‌کنیم تا ایستاد فرض شود. طول قاب‌ها در این مرحله به‌طور تقریبی ۲۵ میلی ثانیه و بدون هم‌پوشانی در نظر گرفته شده است. دلیل عدم قراردادن هم‌پوشانی بین قاب‌ها، نخست این که نتایج آزمایش‌هاست که بیان‌گر نتیجه بهتر در حالت بدون هم‌پوشانی است و دوم این که میانگین‌گیری است که از هر چند فریم متوالی به عمل می‌آید. با توجه به این که ممکن است هر قاب به تنهایی شامل یک آوای کامل نباشد تا در تشخیص لهجه مفید باشد از بردارهای ویژگی مربوط به هر هفت قاب متوالی میانگین گرفته و آن را به عنوان یک بردار ویژگی شامل ضرایب مل-کپستروم و سایر ویژگی‌ها در نظر

تمرکز بیشتری وجود داشته است. به‌عنوان نمونه در سال ۱۹۹۶/ارسلان و هانسن با ارائه یک سامانه طبقه‌بندی لهجه برای زبان انگلیسی یکی از برجسته‌ترین و نخستین پژوهش‌های انجام‌شده در این زمینه را انجام دادند. (ارسلان، ۱۹۹۶). فاریا در سال ۲۰۰۵ تلاش دیگری برای طبقه‌بندی لهجه‌ها با استفاده از روش‌های SVM و GMM انجام داد. در این پژوهش علاوه بر ویژگی‌های صوتی گفتار از ویژگی‌های کلامی آن نیز استفاده شده است. (فاریا، ۲۰۰۵) پدرسن و دیدریچ در سال ۲۰۰۷ طبقه‌بندی لهجه را با کمک SVM و یادگیری درخت تصمیم برای دو لهجه عربی و هندی از زبان انگلیسی انجام داد. (پدرسن، ۲۰۰۷)، علاوه بر این پژوهش‌های دیگری نیز در حوزه تشخیص لهجه صورت گرفته است. (چن، ۲۰۰۱)، (کامف، ۱۹۹۶) و (زنگ، ۲۰۰۵). در حوزه لهجه‌های فارسی، نتایج مشاهدات ما نشان می‌دهد که تاکنون تعداد سه پژوهش در این زمینه انجام شده است، در نخستین پژوهش پنج لهجه اصفهانی، تهرانی، آذری، کردی و مازندرانی انتخاب شده و با استخراج ویژگی‌های ضرایب مل-کپستروم و مشتقات اول و دوم آن و همچنین انرژی هر قاب در مرحله طبقه‌بندی از شبکه عصبی احتمالاتی و ماشین بردار پشتیبان استفاده شده است. (قلی‌پور، ۱۳۹۱) در دومین پژوهش صورت‌گرفته سه لهجه تهرانی، اصفهانی و کرمانشاهی برای انجام آزمایش‌های مختلف انتخاب شده‌اند. همانند بیشتر مقالات، این مقاله نیز در مرحله استخراج ویژگی از ویژگی‌های متداولی همچون ضرایب مل-کپستروم، فرکانس‌های فرمنت و انرژی استفاده کرده است. در مرحله طبقه‌بندی سه طبقه‌بند مختلف شامل KNN، SVM و MLP به کار گرفته شده است (ربیعی، ۲۰۱۰). در آخرین پژوهش بررسی شده پنج لهجه اصفهانی، تهرانی، آذری، شمالی و جنوبی برای انجام آزمایش‌ها برگزیده شده است. در این مقاله ویژگی‌های ضرایب مل-کپستروم و مشتق نخست و دوم آن، انرژی هر قاب، سه فرکانس فرمنت نخست و SDC از سیگنال گفتار دارای لهجه استخراج شده است. هدف اصلی این مقاله ترکیب دو طبقه‌بند GMM و PRLM است که با انجام این کار نرخ تشخیص لهجه‌های مختلف نسبت به حالتی که این طبقه‌بندها به‌طور تکی استفاده می‌شوند، به اندازه مطلوبی افزایش می‌یابد. (جلالوند، ۲۰۱۲)

<sup>۱</sup> Mel Frequency Cepstral Coefficient

### ۳-۳- فرکانس‌های فرمنت

یکی دیگر از ویژگی‌های حوزه فرکانس که از سیگنال گفتار استخراج می‌شود، فرکانس‌های فرمنت هست. فرمنت‌ها پارامترهای مربوط به لوله صوتی هستند که برای هر آوایی مقدار فرکانس فرمنت متفاوتی در مقایسه با آوای دیگر وجود دارد که مشخصه همان آوا به شمار می‌رود. به‌طور معمول از سه یا دو فرکانس فرمنت نخست ( $2F$  یا  $3F$ ) به‌عنوان ویژگی در سامانه‌های تشخیص گفتار و لهجه استفاده می‌شوند. برای محاسبه این ویژگی روش‌های مختلفی وجود دارد که در اینجا با استفاده از رابطه بازگشتی ضرایب پیش‌بینی خطی ( $LPC^1$ ) فرکانس‌های فرمنت محاسبه شده‌اند.

### ۳-۴- پیش‌گویی خطی مبتنی بر درک انسان ( $PLP^2$ )

یکی از معایب آنالیز LPC آن است که فرآیند و نحوه عملکرد دستگاه شنوایی انسان را در محاسبه ویژگی‌ها منظور نمی‌کند. به‌عبارت‌دیگر LPC در تمامی فرکانس‌ها سیگنال گفتار را به یک‌صورت تخمین می‌زند که این مطابق با دستگاه شنوایی انسان نیست. برای حل این مشکل و به منظور هماهنگ کردن روش پیش‌گویی خطی با دستگاه شنوایی انسان روش پیش‌گویی خطی مبتنی بر درک انسان پیشنهاد شد. (گریگر، ۲۰۱۱) برای محاسبه این ویژگی به جای مقیاس مل از یک مقیاس جدید به نام بارک استفاده شده است. برخی از آزمایش‌های انجام‌شده بیان‌گر این مسئله هست که روش MFCC نسبت به PLP دارای نتایج بهتری در تشخیص گفتار است؛ اما روش PLP نسبت به تغییرات تعداد ضرایب و تعداد فیلترهای مورد استفاده در محاسبه ضرایب دارای نتایج پایدارتری نسبت به MFCC است. علاوه‌براین ویژگی PLP در برابر نوفه مقاومت بیشتری از خود نشان می‌دهد و مصون‌تر است. (فان، ۲۰۰۰) به‌طور کلی این آنالیز به دلیل کم‌بودن پیچیدگی محاسباتی آن یک ویژگی کاراست و سیگنال گفتار را با بعدی کمتر از خود سیگنال نمایش می‌دهد. و این ویژگی‌ها باعث شده است که از این آنالیز در تشخیص گفتار مستقل از گوینده بیشتر استفاده شود.

می‌گیریم. برای به‌دست آوردن تعداد قاب‌های متوالی سعی شده است، بهترین تعداد قاب با بهترین طول که همان ۲۵ میلی‌ثانیه است، انتخاب شود.



(شکل-۱) : مراحل محاسبه MFCC

### ۳-۲- مشتق نخست و دوم ضرایب مل-کپستروم

با اعمال سایر مراحل شکل (۱) ضرایب مل-کپستروم حاصل می‌شود. علاوه بر ضرایب مل-کپستروم از مشتق نخست ( $\Delta MFCC$ ) و مشتق دوم آن ( $\Delta\Delta MFCC$ ) نیز به‌عنوان ویژگی استفاده می‌شود که از روابط (۱) و (۲) به‌دست می‌آیند (بهره‌وا، ۲۰۱۲).

$$\Delta MFCC = D[n] = C[n+m] - C[n-m] \quad (1)$$

$$\Delta\Delta MFCC = DD[n] = D[n+m] - D[n-m] \quad (2)$$

که در این روابط  $C$  بردار ضرایب مل-کپستروم و  $n$  شماره قاب است.  $m$  در عمل به‌طور تقریبی ۲ یا ۳ در نظر می‌گیرند.

<sup>1</sup> Linear Prediction Cepstral

<sup>2</sup> Perceptual Linear Prediction

ویژگی‌هایی که بیان شد متداول‌ترین ویژگی‌هایی هستند که در یک سامانه تشخیص گفتار و لهجه به کار می‌روند.

## ۴- روش‌های پیشنهادی استخراج ویژگی

در این قسمت از مقاله به بیان توضیحات و جزئیات بیشتری از روش‌های پیشنهادی برای مرحله استخراج ویژگی می‌پردازیم.

### ۴-۱- دامنه مرکزی طیفی و مشتق نخست آن (SCM و $\Delta SCM$ )

ویژگی دامنه مرکزی طیفی نخستین بار در سال ۲۰۱۰ به عنوان یک ویژگی در تشخیص گوینده استفاده شده است. این ویژگی و مشتق نخست آن با اعمال تغییراتی در مراحل محاسبه ضرایب مل-کستروم به دست می‌آیند.

برای محاسبه  $SCM^1$  با توجه به شکل (۱) به جای مرحله محاسبه انرژی از رابطه (۳) دامنه‌های مرکزی طیفی را محاسبه کرده و سپس لگاریتم و تبدیل کسینوسی ضرایب حاصل را به دست می‌آوریم (کارن، ۲۰۱۰).

$$M_k = \frac{\sum_{f=l_k}^{u_k} f |S[f] \omega_k[f]|}{\sum_{f=l_k}^{u_k} f} \quad (3)$$

در این رابطه  $l_k$  فرکانس لبه پایین و  $u_k$  فرکانس لبه بالای هر فیلتربانک، مثلثی شکل است.  $S[f]$  تبدیل یافته هر قاب از سیگنال گفتار از حوزه زمان به حوزه فرکانس و  $\omega_k[f]$  نشان دهنده هر کدام از فیلتربانک‌های مثلثی شکل هستند.

همانند ضرایب مل-کستروم برای ویژگی اخیر نیز مشتقاتی وجود دارد که می‌توان آن‌ها را به عنوان یک ویژگی در کنار سایر ویژگی‌ها استفاده کرد. در این مقاله از مشتق نخست ویژگی  $SCM$  استفاده شده است که با توجه به رابطه (۴) به دست می‌آید.

$$\Delta SCM = D[n] = S[n+m] - S[n-m] \quad (4)$$

که در این رابطه  $S$  بردار ضرایب دامنه مرکزی طیفی و  $n$  شماره قاب است. همانند رابطه محاسبه  $\Delta MFCC$ ،  $m$  را در عمل به طور تقریبی ۲ یا ۳ در نظر می‌گیرند.

### ۴-۲- اعمال تبدیل Zak به سیگنال گفتار

ویژگی دیگری که در این مقاله از آن به عنوان یک ویژگی جدید استفاده شده است، تبدیل Zak است. این تبدیل قادر است ضرایبی از سیگنال گفتار، هم از حوزه زمان و هم از حوزه فرکانس استخراج کند. رابطه (۵) نحوه محاسبه این تبدیل را نشان می‌دهد (اوسلندر، ۱۹۹۱).

$$Z_{sig}(t, v) = \sum_{n=-\infty}^{n=+\infty} S(t+n) e^{-j2\pi n v} \quad (5)$$

در این رابطه منظور از  $S$  همان سیگنال گفتار است،  $v$  دوره تناوب را نشان می‌دهد و  $t$  نشان دهنده قرار داشتن سیگنال در حوزه زمان است. با اعمال این تبدیل به هر قاب سیگنال گفتار یک ماتریس در خروجی خواهیم داشت که شامل  $N$  سطر و  $M$  ستون خواهد بود.  $N$  تعداد ضرایب حاصل از تبدیل در حوزه زمان و  $M$  تعداد ضرایب حاصل در حوزه فرکانس است.

با پایان یافتن مرحله استخراج ویژگی و تشکیل بردار ویژگی‌های مختلف که از ترکیب ویژگی‌های استخراج شده به دست می‌آیند، وارد مرحله طبقه‌بندی می‌شویم.

### ۵- طبقه‌بندی

پس از تشکیل بردار ویژگی‌ها به منظور دسته‌بندی و تشخیص آن‌ها باید از یک طبقه‌بند مناسب استفاده کرد. در این مقاله ابتدا از پنج طبقه‌بند مختلف شامل شبکه عصبی پرسپترون چندلایه (MLP)، شبکه عصبی احتمالاتی (PNN)، طبقه‌بند  $k$  نزدیک‌ترین همسایه (KNN)، ماشین بردار پشتیبان (SVM)، و شبکه توابع بنیادی شعاعی (RBF) استفاده شده است. در مرحله بعد بعضی از طبقه‌بندهای به کار گرفته شده به دو روش مختلف و به صورت دوتایی و سه‌تایی با یکدیگر ترکیب شده‌اند. هدف این است که با مقایسه نتایج مربوط به هر کدام از این طبقه‌بندها و ترکیب آن‌ها، یک حالت را که دارای عملکرد بهتر و بهینه‌تری است، انتخاب کنیم. ابتدا توضیح مختصری درباره هر کدام از طبقه‌بندها ارائه می‌شود و سپس به بررسی نتایج خواهیم پرداخت.

### ۵-۱- شبکه $MLP^2$

این شبکه شامل سه لایه به نام‌های ورودی، مخفی و خروجی است که تعداد سلول‌های هر لایه به روش سعی و

<sup>2</sup> Multi-Layer Perceptron

<sup>1</sup> Spectral Centroid Magnitude

این نمونه‌ها  $k_n$  نزدیک‌ترین همسایه‌های  $x$  هستند. در حالت کلی  $k$  را به صورت  $k_n$  در نظر می‌گیریم که  $k_n$  تابعی تعریف شده از  $n$  است. اگر چگالی نقاط آموزش اطراف  $x$  زیاد باشد، سلول کوچک می‌شود و بنابراین نتیجه به دست آمده نتیجه بهتری است و در صورتی که چگالی نقاط آموزش اطراف  $x$  کم باشد، سلول بزرگ می‌شود.

#### ۵-۴- طبقه‌بند SVM<sup>۳</sup>

SVM در واقع یک طبقه‌بندی‌کننده دودویی است که دو طبقه را با استفاده از یک مرز خطی از هم جدا می‌کند. در این روش با استفاده از تمامی باندها و یک الگوریتم بهینه‌سازی، نمونه‌هایی که مرز طبقه‌ها را تشکیل می‌دهند، به دست می‌آورند. این نمونه‌ها را بردارهای پشتیبان گویند. تعدادی از نقاط آموزشی را که کمترین فاصله تا مرز تصمیم‌گیری دارند، به عنوان زیرمجموعه‌ای برای تعریف مرزهای تصمیم‌گیری و به عنوان بردار پشتیبان می‌توان در نظر گرفت (واپنایک، ۱۹۹۱). البته قبل از تقسیم خطی باید دقت کرد که برای این که ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند، داده‌ها را با تابعی مشخص به فضایی با ابعاد بالاتر منتقل می‌کنیم. برای محاسبه مرز تصمیم‌گیری دو طبقه به طور کامل جدا از هم از روش حاشیه بهینه استفاده می‌شود.

#### ۵-۵- شبکه RBF<sup>۴</sup>

RBF روشی برای تقریب توابع است و یادگیری با آن ارتباط نزدیکی با شبکه‌های عصبی مصنوعی دارد. در این روش فرضیه یادگرفته شده به صورت رابطه (۶) است.

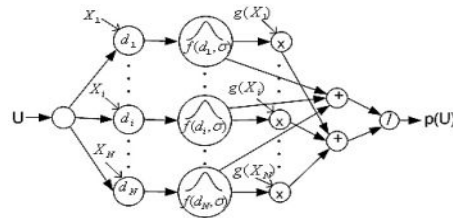
$$\hat{f}(x) = w_0 + \sum_{u=1}^k w_u K_u(d(x_u, x)) \quad (6)$$

$w_0$  نشان‌دهنده بردار وزن نخستین و  $w_u$  بردار وزن‌های به‌روزشده است.  $x$  خروجی مطلوب،  $x_u$  ورودی تابع و  $d$  فاصله بین ورودی و خروجی مطلوب است. و در نهایت  $\hat{f}(x)$  خروجی شبکه RBF را به‌ازای هر ورودی نشان می‌دهد. در این روش از تعداد  $k$  تابع کرنل برای تقریب تابع استفاده می‌شود. تابع کرنل به‌طور معمول به صورت تابع گاوسی رابطه (۷) با واریانس  $\sigma_u^2$  انتخاب می‌شود.

خطا مشخص می‌شود. این شبکه بر مبنای الگوریتم پس‌انتشار خطا آموزش می‌بیند. بدین ترتیب که خروجی‌های واقعی با خروجی‌های دلخواه مقایسه و وزن‌ها با الگوریتم پس‌انتشار به صورت تحت نظارت تنظیم می‌شوند تا الگوی مناسب به وجود آید. روند یادگیری هنگامی متوقف می‌شود که مجموع کل خطا از مقدار آستانه تعیین شده کمتر شود و یا تعداد کل دوره آموزش به پایان برسد (هیگین، ۲۰۰).

#### ۵-۲- شبکه PNN<sup>۱</sup>

این طبقه‌بند از نوع شبکه‌های عصبی RBF است و بر پایه تابع چگالی احتمال نمایی و قانون تصمیم‌گیری Bayes عمل می‌کند. PNN یک شبکه پیش‌رونده سه‌لایه بوده و از یادگیری بانظارت استفاده می‌کند. لایه نخست ورودی‌ها را دریافت، لایه میانی بردار احتمال را بر پایه توزیع هر طبقه تعیین و لایه آخر مقدار بیشینه احتمال را انتخاب کرده و طبقه مربوطه را تعیین می‌کند. این شبکه به صورت موازی، محاسبات را انجام می‌دهد، آموزش ساده‌ای دارد و به دلیل سرعت بالا در کاربردهای بی‌درنگ مورد استفاده قرار می‌گیرد (قلیپور علی، ۱۳۹۱). شکل (۲) تصویری از این شبکه را نشان می‌دهد.



شکل (۲): نحوه عملکرد شبکه PNN (قلیپور علی، ۱۳۹۱)

#### ۵-۳- شبکه KNN<sup>۲</sup>

KNN یک الگوریتم آموزش با سرپرستی است. در حالت کلی از این الگوریتم به دو منظور تخمین تابع چگالی توزیع داده‌های آموزش و طبقه‌بندی داده‌های آزمون بر اساس الگوهای آموزش استفاده می‌شود. برای تخمین  $p(x)$  از روی  $n$  نمونه آموزش توسط الگوریتم KNN می‌توانیم یک سلول به مرکزیت  $x$  ایجاد کرده و اجازه دهیم شعاع این سلول تا حدی گسترش پیدا کند که  $k_n$  نمونه آموزش را در برگیرد.

<sup>1</sup> Probabilistic Neural Network

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

<sup>3</sup> Support Vector Machine

<sup>4</sup> Radial Basis Function



$$K_u(d(x_u, x)) = \frac{1}{e^{2\sigma_u^2 d(x_u, x)^2}} \quad (7)$$

نشان داده شده است در صورتی که تعداد کافی تابع کرنل گاوسی انتخاب شود، با استفاده از این شبکه می‌توان هر تابعی را با خطای تاحدودی کم تقریب زد. از مزایای این شبکه می‌توان به آموزش آسان‌تر آن نسبت به شبکه‌های عصبی معمولی، که از روش پس‌انتشار استفاده می‌کنند، نام برد.

## ۵-۶- ترکیب طبقه‌بندها<sup>۱</sup>

هدف نهایی طراحی یک سامانه شناسایی الگو رسیدن به بهترین عملکرد طبقه‌بندی برای مسأله موجود است. (هیکنین، ۱۹۹۹) از آن جایی که هیچ طبقه‌بندی به‌طور کامل قادر به حل تمام مسائل نیست، ترکیب طبقه‌بندها به‌عنوان روشی برای کاهش خطای طبقه‌بندی و افزایش کارایی سامانه‌ها پیشنهاد شده است.

ایده اصلی این طرح، ایجاد طبقه‌بندهای گوناگون با ناحیه‌های خطای متفاوت است تا سامانه ترکیبی با بهره‌گیری از نقاط قوت تک‌تک طبقه‌بندها و رفع خطای ایجادشده یک طبقه‌بند، با اطلاعات به‌دست‌آمده از طریق طبقه‌بندهای دیگر، خطای کل سامانه را کاهش دهد (دابتنیچ و ۱۹۹۵) و (فروند، ۱۹۹۷).

برای ترکیب طبقه‌بندها، روش‌ها و طرح‌های مختلفی پیشنهاد شده است که از آن جمله می‌توان روش‌های سری، موازی و ترکیبی را نام برد. در این مقاله از روش موازی استفاده شده است. در این روش همه طبقه‌بندها به‌طور هم‌زمان با هم ترکیب می‌شوند. یکی از ساده‌ترین فرآیندهایی که در مرحله ترکیب استفاده می‌شود، روش میانگین‌گیری است؛ بدین صورت که نتیجه حاصل از تمام طبقه‌بندهای موردنظر برای ترکیب کردن با هم جمع و سپس بر تعداد طبقه‌بندهای ترکیب‌شده تقسیم می‌شوند.

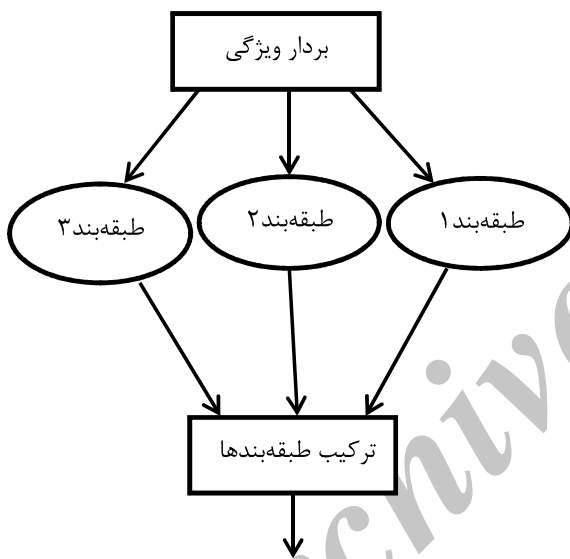
روش دیگر این است که برای هر کدام از طبقه‌بندها یک وزن خاص تعیین و سپس این وزن را در نتیجه نهایی هر کدام از طبقه‌بندها ضرب کنیم. وزن‌ها باید به‌گونه‌ای باشند که مجموع وزن‌های به‌کاررفته برابر یک شود. بهترین وزن ممکن برای هر طبقه‌بند را از روش سعی و خطا به‌دست می‌آوریم. پس از ضرب کردن وزن در نظر گرفته‌شده برای هر طبقه‌بند در طبقه‌بند موردنظر،

<sup>۱</sup> Combination classifier

نتیجه حاصل را با هم جمع می‌کنیم و سپس نرخ بازشناسی را با ترکیب جدید به‌دست می‌آوریم.

در این پژوهش از هر دو روش مذکور استفاده شده است. روش میانگین‌گیری برای زمانی که سه طبقه‌بند را با هم ترکیب کرده‌ایم و روش تعیین وزن هنگامی که دو طبقه‌بند با یکدیگر ترکیب ترکیب شده است، استفاده می‌شود.

نکته دیگری که در مورد ترکیب طبقه‌بندها باید در نظر گرفت، این است که بردار ویژگی که هر کدام از طبقه‌بندها با آن آموزش دیده‌اند هم می‌تواند متفاوت باشد و هم به‌طور کامل مشابه هم باشند که در اینجا ویژگی استخراج‌شده برای آموزش هر کدام از طبقه‌بندها با هم یکسان است. شکل (۳) روش استفاده‌شده را نشان می‌دهد.

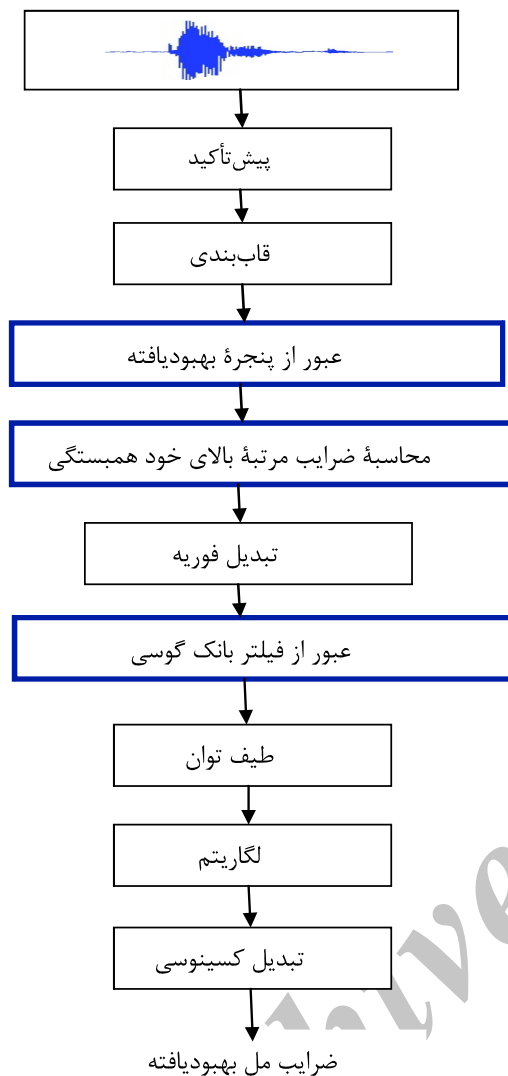


(شکل - ۳) : روش استفاده‌شده برای ترکیب طبقه‌بندها

روش ترکیب طبقه‌بندها، طبق مشاهدات ما، تاکنون یکبار در مورد تشخیص لهجه‌های فارسی به کار گرفته شده است که در آن دو طبقه‌بند GMM و PRLM با یکدیگر به روش موازی ترکیب و باعث بهبود نتایج تا حد مطلوبی (حدود ۲۰ درصد) شده‌اند. (جلالوند، ۲۰۱۲)

## ۶- تشخیص لهجه‌ها در محیط نوفه‌ای

یکی از عواملی که ممکن است در هر سامانه تشخیص گفتاری باعث کاهش کارایی و همچنین نرخ بازشناسی آن سامانه شود، وجود نوفه در محیط است. حال اگر دو عامل



(شکل - ۴): روش پیشنهادی برای محیط نوفه‌ای ( ضرایب مل - کپستروم بهبودیافته)

ویژگی این نوفه با فرض ناهمبسته بودن نسبت به سیگنال اصلی این است که تابع خودهمبستگی مربوط به آن تا حدود زیادی نسبت به زمان بدون تغییر و نزدیک به صفر است. بنابراین تابع خودهمبستگی به صورت رابطه (۱۰) بیان خواهد شد.

$$R_{XX}(m, k) = R_{SS}(m, k) + R_{dd}(m) \quad (10)$$

در این رابطه  $k$  شماره هر قاب است. رابطه (۱۰) نشان می‌دهد که تابع خودهمبستگی نوفه ناهمبسته به سیگنال، مستقل از فریم است و با توجه به ناچیز بودن آن

لهجه و نوفه در یک گفتار کنار هم قرار گیرند، شرایط دشوارتری برای تشخیص آن ایجاد خواهد شد. بنابراین نیاز است تا روشی برای حذف نوفه از یک سیگنال گفتار لهجه‌دار قبل از تشخیص نوع لهجه آن پیشنهاد شود. برای حذف نوفه از یک سیگنال گفتار، تاکنون روش‌های متنوعی پیشنهاد شده است که اغلب آن‌ها با ایجاد تغییراتی در مراحل محاسبه ضرایب مل - کپستروم به دست آمده‌اند. این تغییرات را به سه دسته مختلف تقسیم‌بندی می‌توان کرد:

۱- مدل‌های بهبودیافته شامل تغییر در بلوک‌های پایه این الگوریتم

۲- مدل‌های بهبودیافته شامل یک بلوک تکمیل‌کننده که به الگوریتم پایه اضافه شده است.

۳- بهبود در پیاده‌سازی سخت‌افزاری این الگوریتم توسط کاستن محاسبات ضرب و جمع در الگوریتم پایه

در این مقاله با ترکیب چند روش از روش‌های استفاده‌شده به نتیجه مطلوبی در کاهش نوفه از یک سیگنال گفتار لهجه‌دار رسیده‌ایم. شکل (۴) نحوه به دست آمدن ویژگی جدید را که شامل سه تغییر نسبت به الگوریتم نشان داده شده در شکل است، نشان می‌دهد.

تغییرات ایجادشده به شرح زیر است.

الف) در بیشتر الگوریتم‌های بررسی‌شده برای مرحله عبور از پنجره از یک پنجره همینگ ساده استفاده شده است؛ اما پنجره‌ای که ما در اینجا استفاده می‌کنیم، با رابطه (۸) نشان داده می‌شود. (سahیدولا، ۲۰۱۲)

$W(n)$  یک پنجره همینگ ساده با طول مناسب است.

در پنجره پیشنهادشده دو پارامتر پراکندگی طیفی و عرض بخش اصلی پنجره افزایش می‌یابد که یک تغییر مطلوب است؛ اما هم‌گرایی بخش‌های جانبی کاهش پیدا می‌کند که این یک تغییر نامطلوب است؛ اما در مجموع تغییرات ایجادشده حاکی از بهبود نتایج است و بنابراین این تغییر را اعمال می‌کنیم. (سahیدولا، ۲۰۱۲)

ب) نوفه اضافه‌شده به سیگنال را می‌توان با رابطه (۹) بیان کرد.

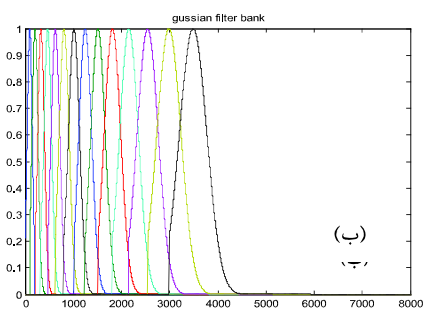
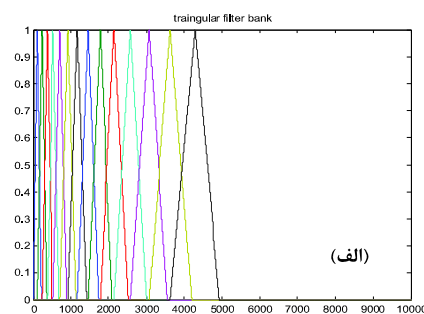
$$w_{new}(n) = nw(n) \quad (8)$$

$$X(n) = S(n) + d(n) \quad (9)$$

که در آن  $s(n)$  سیگنال ورودی و  $d(n)$  نوفه اضافی شونده به سیگنال است.

می‌توان با این روش اثر نوفه را کاهش داد. تخریب سیگنال توسط نوفه در ضرایب مرتبه پایین تابع خودهمبستگی، بیشتر از ضرایب مرتبه بالا است، بنابراین با حذف آن‌ها، اثر نوفه کاسته خواهد شد.

(ج) آخرین تغییر ایجاد شده در محاسبه ضرایب مل-کپستروم، تغییر نوع فیلتربانک از مثلثی به گوسی است. در فیلتربانک مثلثی اطلاعات بخش‌هایی از قاب که در نقاط ابتدایی و انتهایی و خارج از زیربخش‌ها، قرار می‌گیرند از دست می‌روند؛ زیرا مثلث‌ها در خارج از زیرباندها وزنی ندارند؛ اما اگر به جای این فیلتربانک از یک فیلتربانک گوسی که متقارن نیز می‌باشد، استفاده کنیم، به دلیل وجود وزن در خارج از زیرباندهای آن، مانع از دست رفتن اطلاعات در این بخش‌ها می‌شود. مزیت دیگر آن نسبت به فیلتربانک مثلثی، شروع و اتمام آن با شیب کمتر و ملایم‌تر است. این فیلتربانک با ایجاد همبستگی بیشتر بین زیرباندها اطلاعات از دست رفته در مرزها را کاهش داده و در بهبود الگوریتم و به دنبال آن بالا بردن نرخ بازشناسی سامانه خودکار پردازش گفتار مؤثر خواهد بود. شکل (۵) این دو نوع فیلتربانک را نشان می‌دهد.



(شکل - ۶): الف: فیلتربانک مثلثی، ب: فیلتربانک گوسی

جملات لهجه‌دار مربوط به این مقاله از پایگاه داده فارسی‌دات پژوهشگاه توسعه فناوری‌های پیشرفته خواجه نصیرالدین طوسی که معتبرترین پایگاه دادگان گفتار زبان فارسی است، استخراج شده‌اند. این پایگاه داده شامل ۶۰۸۰ جمله است که توسط ۳۰۴ نفر گوینده ایرانی با ده لهجه مختلف بیان شده است. از هر گوینده ۲۰ جمله ضبط گردیده است. در این پایگاه داده گویندگان از لحاظ سن، جنس، سطح تحصیلات و لهجه‌هایشان متمایز شده‌اند که شرایط خوبی را برای انجام پژوهش‌های مختلف فراهم کرده است. در این پژوهش برای شناسایی لهجه‌ها پنج لهجه مختلف انتخاب شده است. در انتخاب نوع لهجه‌ها سعی شده است، لهجه بیشتر مردم مناطق ایران را از لحاظ جغرافیایی شامل شود. این لهجه‌ها شامل تهرانی، ترکی، اصفهانی، شمالی و جنوبی هستند.

## ۷-۲- آزمایش نخست: مقایسه لهجه‌ها به صورت دوتایی

در آزمایش نخست با انتخاب ویژگی MFCC از میان ویژگی‌های متداول و ویژگی SCM از بین ویژگی‌های پیشنهادی و با به کار بردن تمامی طبقه‌بندهای معرفی شده در این مقاله، نرخ بازشناسی لهجه‌ها به صورت دوتایی به دست آمده‌اند. هدف از این آزمایش مقایسه عملکرد ویژگی پیشنهادی با ویژگی‌های متداول در حالی که دو طبقه داریم، است.

جدول (۱) نتایج را برای ویژگی MFCC نشان می‌دهد.

(جدول - ۱): نرخ تشخیص لهجه‌ها به صورت دوتایی با

طبقه‌بندهای مختلف و ویژگی متداول MFCC

میانگین	RBF	SVM	KNN	PNN	MLP	طبقه‌بند
69.23	72.21	<b>82.62</b>	30.46	80.4	80.05	اصفهانی و تهرانی
76.95	74.95	83.03	62.1	<b>83.3</b>	81.4	تهرانی و ترکی
69.13	70.59	80.55	32.04	81.65	<b>80.86</b>	ترکی و اصفهانی
74.07	76.57	81.65	46.26	82.34	<b>83.55</b>	شمالی و جنوبی
70.59	71.24	<b>82.75</b>	35.91	81.79	81.26	تهرانی و جنوبی
70.15	69.79	79.72	40.68	<b>80.55</b>	80.05	تهرانی و شمالی
70.28	74.31	80.13	35.03	80.96	<b>80.99</b>	ترکی و جنوبی
69.81	72.21	<b>83.58</b>	34.9	78.62	79.78	ترکی و شمالی
75.15	75.12	<b>82.89</b>	55.27	81.37	81.13	اصفهانی و شمالی
74.18	71.08	80.82	55.03	<b>82.34</b>	81.67	اصفهانی و جنوبی



### ۷-۳- آزمایش دوم: مقایسه همه لجه‌ها در حضور ویژگی‌های مختلف

در آزمایش دوم، عملکرد بردار ویژگی‌های مختلف متداول و پیشنهادی در حضور همه لجه‌های منتخب این مقاله سنجیده شده است.

ابتدا با تشکیل بردار ویژگی‌هایی از MFCC و مشتقات آن و همچنین فرکانس‌های فرمنت، نرخ بازشناسی را در حضور طبقه‌بندهای مختلف به دست می‌آوریم. جدول (۳) نتایج حاصل از این آزمایش را نشان می‌دهد.

(جدول-۳): نرخ بازشناسی لجه‌ها با ویژگی‌های متداول

ویژگی	طبقه‌بند	MLP	PNN	KNN	SVM	RBF
MFCC		55.63	72.33	22.07	<b>75.3</b>	25.73
$MFCC + \Delta MFCC$		65.95	71.3	21.11	<b>73.73</b>	36.04
$MFCC + \Delta MFCC + \Delta\Delta MFCC$		68.45	70.67	19.53	<b>74.63</b>	68.70
$MFCC + \Delta MFCC + \Delta\Delta MFCC + 2F$		65.86	72.37	21.22	<b>74.09</b>	63.46

در مرحله بعد بردارهایی از ویژگی‌های مختلف پیشنهادی را تشکیل می‌دهیم و آن‌ها را به طبقه‌بندهای مختلف اعمال می‌کنیم تا نرخ بازشناسی به ازای پنج لجه مختلف به دست‌آید. جدول (۴) نتایج حاصل را نشان می‌دهد.

(جدول-۴): نرخ بازشناسی لجه‌ها با ویژگی‌های پیشنهادی

ویژگی	طبقه‌بند	MLP	PNN	KNN	SVM	RBF
SCM		53	71.83	26.93	<b>75.16</b>	36.75
$SCM + \Delta SCM$		58.13	71.70	26.33	<b>74.54</b>	33.31
Zak Transform		48.4	71.65	24.56	<b>74.11</b>	52.61

در جدول (۵) میانگین نرخ بازشناسی به‌ازای هر کدام از ویژگی‌ها با یکدیگر مقایسه شده‌اند.

(جدول-۵): مقایسه میانگین نرخ بازشناسی لجه‌ها با

ویژگی‌های مختلف

ویژگی	میانگین نرخ تشخیص
MFCC	50.21
$MFCC + \Delta MFCC$	53.62
$MFCC + \Delta MFCC + \Delta\Delta MFCC$	<b>60.39</b>
$MFCC + \Delta MFCC + \Delta\Delta MFCC + 2F$	59.4
SCM	52.73
$SCM + \Delta SCM$	52.8
Zak Transform	<b>54.26</b>

همان‌طور که نتایج جدول (۱) نشان می‌دهد، سه طبقه‌بند SVM، MLP و PNN به‌طور میانگین دارای عملکرد بهتری نسبت به سایر طبقه‌بندها هستند. در جدول (۲) نتایج به‌ازای ویژگی پیشنهادی SCM نشان داده شده است.

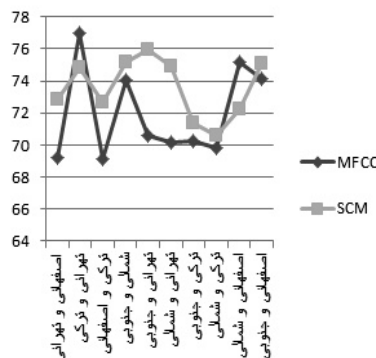
(جدول-۲): نرخ تشخیص لجه‌ها به صورت دوتایی با

طبقه‌بندهای مختلف و ویژگی پیشنهادی SCM

طبقه‌بند	MLP	PNN	KNN	SVM	RBF	میانگین
اصفهانی و تهرانی	78.22	83.11	44.16	<b>83.40</b>	75.44	72.86
تهرانی و ترکی	79.25	80.88	62.38	<b>81.77</b>	69.79	74.81
ترکی و اصفهانی	76	78.96	45.17	<b>79.11</b>	74.31	72.71
شمالی و جنوبی	80.74	82.96	58.98	<b>82.96</b>	70.27	75.18
تهرانی و جنوبی	77.33	84	57.69	<b>84.14</b>	76.73	75.97
تهرانی و شمالی	78.07	82.07	57.17	<b>82.51</b>	74.63	74.89
ترکی و جنوبی	79.55	81.92	46.26	<b>82.51</b>	66.55	71.35
ترکی و شمالی	78.96	80	41.29	<b>80.44</b>	73.18	70.57
اصفهانی و شمالی	74.07	80.59	61.69	<b>80.44</b>	64.62	72.28
اصفهانی و جنوبی	77.77	81.33	63.67	<b>81.62</b>	71.08	75.09

همان‌طور که نتایج دو جدول اخیر نشان می‌دهد، ویژگی پیشنهادی SCM دارای میانگین عملکرد بهتری نسبت به ویژگی MFCC در تشخیص لجه‌ها به‌صورت دوتایی است. همچنین طبقه‌بند KNN نسبت به سایر طبقه‌بندها بازدهی کمتر دارد. شکل (۶) مقایسه میانگین عملکرد این دو ویژگی را نشان می‌دهد.

لازم به ذکر است که به‌دلیل افزایش حجم مقاله از به‌دست‌آوردن نرخ بازشناسی لجه‌ها به‌صورت دوتایی و در حضور دو ویژگی پیشنهادی دیگر، یعنی  $\Delta SCM$  و تبدیل Zak صرف‌نظر می‌کنیم.



(شکل-۶): مقایسه میانگین عملکرد دو ویژگی SCM و MFCC در تشخیص لجه‌ها به‌صورت دوتایی

طبقه‌بندها استفاده می‌کنیم. به‌طور معمول وزنی که به طبقه‌بندی با نرخ بازشناسی کمتر تعلق می‌گیرد از وزنی که به طبقه‌بندی با نرخ بازشناسی بیشتر تعلق می‌گیرد، بیشتر است. به‌عنوان مثال اگر شبکه MLP دارای نرخ بازشناسی ۴۲/۷ درصد و شبکه SVM دارای نرخ بازشناسی ۶۳/۲۴ درصد باشد، وزنی که به MLP تعلق می‌گیرد، برابر ۰/۶ و وزنی که به PNN تعلق می‌گیرد، برابر ۰/۴ است. جدول (۶) نتایج حاصل را نشان می‌دهد.

(جدول - ۶): نرخ بازشناسی لهجه‌ها با ترکیب طبقه‌بندها

به‌صورت دوتایی		MLP	PNN	SVM	MLP + PNN	PNN + SVM	SVM + MLP
ویژگی	طبقه بند						
	$MFCC + \Delta MFCC + \Delta \Delta MFCC$	64.62	71.52	71.23	84.6	<b>87.71</b>	86.59
	Zak Transform	42.7	70.03	63.24	76.57	<b>85.58</b>	83.93

همان‌طور که نتایج جدول نشان می‌دهد، با انجام عمل ترکیب طبقه‌بندها شاهد بهبود نرخ بازشناسی تا حد مطلوبی خواهیم بود. لازم به ذکر است که دلیل تفاوت نرخ بازشناسی به‌ازای هر کدام از طبقه‌بندها در جدول اخیر با مقادیر جدول (۳) و جدول (۴) این است که با هر بار راه‌اندازی طبقه‌بندها، وزن‌های نخستین به‌صورت تصادفی تعیین می‌شود و این باعث تغییر نرخ بازشناسی می‌شود. لازم به ذکر است که بردار وزن نخستین برای هر بار آموزش شبکه، توسط نرم‌افزار به‌صورت خودکار تعیین که باعث پایایی نتایج می‌شود. در جدول (۷) میانگین عملکرد روش ترکیب به‌صورت دوتایی و میزان بهبود نرخ بازشناسی نشان داده شده است.

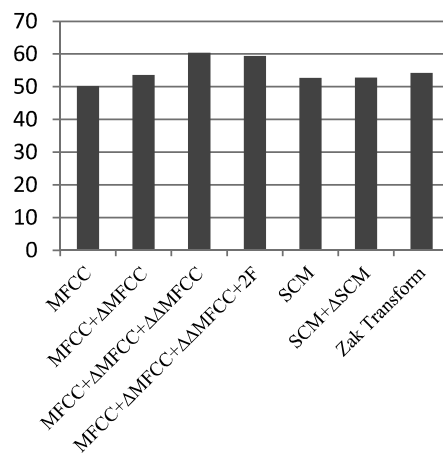
(جدول - ۷): مقایسه میانگین عملکرد طبقه‌بندها

و ترکیب آنها		میانگین بدون ترکیب	میانگین ترکیب به‌صورت دوتایی	میزان بهبود
ویژگی	طبقه‌بند			
	$MFCC + \Delta MFCC + \Delta \Delta MFCC$	69.12	86.3	<b>17.18%</b>
	Zak Transform	58.65	82.02	<b>23.37</b>

در مرحله بعد به‌ازای همان دو ویژگی انتخابی، هر سه طبقه‌بند را با یکدیگر ترکیب می‌کنیم. در این مرحله ترکیب کردن را با روش میانگین‌گیری انجام می‌دهیم، به این

همان‌طور که مقادیر این جدول نشان می‌دهد، در میان ویژگی‌های متداول استفاده‌شده در این مقاله، بردار ویژگی «  $MFCC + \Delta MFCC + \Delta \Delta MFCC$  » دارای بالاترین بازدهی است که دلیل این امر می‌تواند وجود مشتقات ضرایب مل-کپستروم باشد. در میان ویژگی‌های پیشنهادشده تبدیل Zak بیشترین بازدهی را دارد که این امر نیز می‌تواند ناشی از حضور هم‌زمان ویژگی‌های حوزه زمان و فرکانسی در این تبدیل باشد. نمودار میله‌ای شکل (۷) مقایسه این بازدهی را نشان می‌دهد.

در این مرحله از آزمایش‌ها به این نکته دست یافتیم که در حضور پنج لهجه، برخی از ویژگی‌های متداول استفاده‌شده کارایی بهتری نسبت به ویژگی‌های پیشنهادی دارند و در میان طبقه‌بندها، ماشین بردار پشتیبان، کارایی بالاتری دارد.



(شکل - ۸): نمودار میله‌ای مقایسه بازدهی ویژگی‌های مختلف در تشخیص لهجه‌ها

#### ۷-۴- آزمایش سوم: ترکیب طبقه‌بندها

در این بخش از مقاله قصد داریم با استفاده از روش ترکیب طبقه‌بندها بازشناسی لهجه‌ها را انجام دهیم. از میان ویژگی‌های متداول و پیشنهادی، ویژگی را انتخاب می‌کنیم که دارای بالاترین بازدهی باشد. این ویژگی‌ها شامل بردار ویژگی «  $MFCC + \Delta MFCC + \Delta \Delta MFCC$  » و ضرایب تبدیل Zak است.

به‌منظور انجام عمل ترکیب کردن، طبقه‌بندهای MLP، PNN و SVM که بالاترین بازدهی را داشته‌اند، ابتدا طبقه‌بندها را به‌صورت دوتایی با یکدیگر ترکیب

می‌کنیم. برای انجام این کار از روش وزن‌دهی به خروجی

## ۷-۵- آزمایش چهارم : تشخیص لهجه‌ها در محیط نوفه‌ای

ویژگی پیشنهادی در این مقاله را که با ایجاد سه تغییر در الگوریتم پایه محاسبه ضرایب مل-کپستروم به دست می‌آید با نام AGMFCC نام‌گذاری می‌کنیم. برای انجام آزمایش از نسبت سیگنال به نوفه‌های مختلف استفاده شده است. در مرحله نخست چهار ویژگی مختلف در کنار طبقه‌بند PNN به کار گرفته شده است. جدول (۹): نرخ بازشناسی لهجه‌ها در محیط نوفه‌ای با طبقه‌بند PNN و ویژگی‌های مختلف نتایج حاصل را نشان می‌دهد.

(جدول- ۹): نرخ بازشناسی لهجه‌ها در محیط نوفه‌ای با

طبقه‌بند PNN و ویژگی‌های مختلف

SNR	ویژگی	MFCC	SCM	Zak T	AGMFCC
Without noise		71.41	69.13	71.54	<b>72.14</b>
25dB		71.05	71.38	70.69	<b>71.65</b>
10dB		70.6	<b>72.08</b>	70.72	70.33
5dB		70.76	71.16	69.12	<b>72.30</b>
0dB		<b>72.23</b>	69.06	67.83	70.49
-5dB		71.25	63.17	67.05	<b>72.41</b>
	میانگین	71.21	69.17	69.49	<b>71.55</b>

برای این که نشان داده شود ویژگی AGMFCC مستقل از طبقه‌بند در محیط نوفه‌ای بهترین عملکرد را دارد، همان آزمایش بالا با طبقه‌بند SVM انجام شده است. (جدول ۱۰): نرخ بازشناسی لهجه‌ها در محیط نوفه‌ای با طبقه‌بند SVM و ویژگی‌های مختلف نتایج را نشان می‌دهد.

(جدول- ۱۰): نرخ بازشناسی لهجه‌ها در محیط نوفه‌ای با

طبقه‌بند SVM و ویژگی‌های مختلف

SNR	ویژگی	MFCC	SCM	Zak T	AGMFCC
Without noise		74.58	73.95	74.29	<b>74.63</b>
25dB		74.13	<b>75.75</b>	74.84	75.21
10dB		73.73	<b>74.18</b>	72.14	73.69
5dB		71.54	<b>75.16</b>	74.23	73.95
0dB		66.97	71.94	71.10	<b>75.92</b>
-5dB		58.52	69.97	68.58	<b>72.12</b>
	میانگین	69.95	73.49	72.53	<b>74.25</b>

همان‌طور که نتایج جدول (۹) و جدول (۱۰) نشان می‌دهد، ویژگی پیشنهادی AGMFCC که برای محیط‌های نوفه‌ای استفاده می‌شود، بدون توجه به نوع طبقه‌بند بهترین

صورت که نتیجه تمام طبقه‌بندهای موردنظر را با هم جمع کرده و بر تعداد آن‌ها تقسیم می‌کنیم و سپس با تعیین یک آستانه خروجی نهایی این ترکیب را به دست می‌آوریم. جدول (۸) نتایج حاصل را نشان می‌دهد.

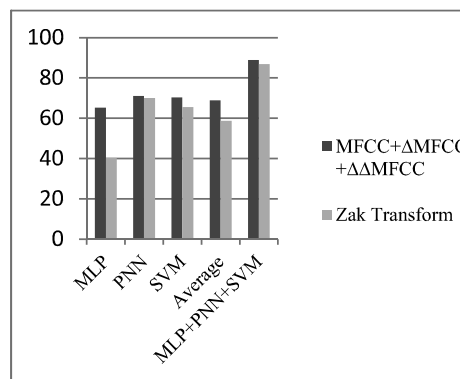
(جدول- ۸): نرخ بازشناسی لهجه‌ها با ترکیب طبقه‌بندها به

صورت سه تایی

ویژگی	طبقه‌بند	MLP	PNN	SVM	MLP+SVM + PNN
	$MFCC + \Delta MFCC + \Delta \Delta MFCC$	65.22	71.05	70.33	<b>88.85</b>
	Zak Transform	40.27	70.03	62.52	<b>86.93</b>

همان‌طور که جدول بالا نشان می‌دهد با انجام این روش ترکیبی نرخ بازشناسی افزایش پیدا می‌کند. برای ویژگی متداول استفاده شده نسبت به میانگین خروجی طبقه‌بندها ۱۹/۹۸ درصد بهبود و برای ویژگی پیشنهادی ۲۸/۱۷ درصد بهبود را شاهد هستیم.

در این جدول نیز بنا به دلیلی که در قبل بیان شد، نرخ بازشناسی حاصل از هر کدام از طبقه‌بندها با مقادیر جدول‌های قبلی مقادیری تفاوت دارد. نمودار شکل (۸) مقایسه نتایج را نشان می‌دهد.



(شکل- ۹): نمودار مقایسه عملکرد طبقه‌بندها و ترکیب آنها به صورت سه تایی

نتایج بخش ۷-۴ بیان‌گر بهبود نتایج با به‌کاربردن روش ترکیب طبقه‌بندهاست و این روش را می‌توان به‌عنوان شیوه‌ای کارآمد در تشخیص گفتار و لهجه استفاده کرد.

## ۹- مراجع

قلی‌پور علی، صدیقی محمدحسین و شمسی موسی، "طبقه‌بندی برخی از لهجه‌های زبان فارسی با استفاده از شبکه عصبی احتمالاتی" بیستمین کنفرانس مهندسی برق ایران، دانشگاه تهران، ۱۳۹۱.

A. Rabiee and S. Setayeshi, "Persian Accents Identification Using an Adaptive Neural Network", 2th .Int. Conf. on Education Tech-nology and Computer Science, pp. 7-11, 2010.

A.Faria "Accent Classification for Speech Recognition", Machine Learning for Multimodal Interaction, pp.285-293, Springer Berlin Heidelberg, 2005.

C.Pedersen and J.Diederich, "Accent Classification Using Support Vector Machines", 6th ICIS, 2007.

F.Phan, M.T.Evangelia.sideman "Speaker Identification using Nerula Network and Wavelets" IEEE Engineering in medicin and Biology Magazine, vol.191, pp.92-101,2000.

H.Behrava, Master's thesis, "Dialect and Accent Recognition". School of Computing, Eastern Finland university, 2012.

J.M.Karen Kua, T.Thiruvanan, M.Nosratighods, E.Ambikairajah and J.Epps. "Investigation of Spectral Centroid Magnitude and Frequency for Speaker Recognition", In Odyssey-2010, pp 007.

K.Kumpf and R.W.King "Automatic accent classification of foreign accented australian English speech" Fourth international conference on spoken language, pp 1740-1743 vol.3, Philadelphia, 1996.

L. Auslander, I. Gertner, R. Tolimieri, "The Discrete Zak Transform Application to Time-Frequency Analysis and Synthesis of Nonstationary Signals" IEEE Trans. on Signal Proc., Vol. 39, No. 4, pp. 825-835, April 1991.

M. Sahidullah and G. Saha, "A novel windowing technique for efficient computation of mfcc for speaker recognition" Arxiv, pp 1206-2437 v1., 2012.

M.L.Arslan and H.L.Hansen, "Language Accent Classification in American English", Speech Communication 18(4): 353-367, 1996.

O.Grigeor, C.Grigeor and V.Velican, "Impaired Speech Evaluation using Mel-Cepstrum Analysis" International Journal Of Circuit, Systems And Signal Processing, 2014.

S.Haykin, "Neural Networks", Macmillan College Publishing Company, 1999.

S.Jalalvand, A.Akbari, and B.Nasersharif, "A classifier combination approach for Farsi accents recognition" 20th Iranian Conference on Electric Engineering, pp.716-720, Tehran, Iran, 2012.

عملکرد را دارد. همان‌طور که نتایج دو جدول اخیر نشان می‌دهد، در شرایط نوبه‌ای ویژگی AGMFCC نسبت به سایر ویژگی‌ها میانگین عملکرد بهتری دارد.

## ۸- نتیجه‌گیری کلی

در این مقاله هدف ما تشخیص لهجه‌های مختلف زبان فارسی از روی شکل موج گفتار بود. برای این منظور سه مرحله پیش‌پردازش، استخراج ویژگی و طبقه‌بندی را اجرا کردیم. در مرحله استخراج ویژگی از تعدادی ویژگی متداول و تعدادی ویژگی جدید شامل SCM،  $\Delta$ SCM و تبدیل Zak استفاده شد. در نهایت در مرحله طبقه‌بندی، پنج طبقه‌بند مختلف شامل MLP، PNN، SVM، KNN و RBF و ترکیبی از آن‌هایی که دارای بالاترین بازدهی بودند، به صورت دوتایی و سه تایی به کار گرفته شد. البته لازم به ذکر است که علاوه بر طبقه‌بندهای مورد استفاده در این مقاله، طبقه‌بندهای دیگری نیز وجود دارند که از آن‌ها می‌توان برای تشخیص گفتار، گوینده و لهجه استفاده کرد که از میان آن‌ها می‌توان به طبقه‌بند PLDA / I-vector اشاره کرد. روش تحلیل تفکیک خطی احتمالاتی (PLDA) و I-vector اغلب برای تشخیص گوینده مستقل از متن استفاده می‌شود و با ایجاد تغییراتی می‌توان آن را برای تشخیص گوینده وابسته به متن نیز استفاده کرد. (استافلایس)

نتایج آزمایش‌هایی که در سه مرحله مختلف انجام شد نشان داد که در بین ویژگی‌های متداول، بردار ویژگی شامل ضرایب مل-کپستروم و مشتق نخست و دوم آن و در بین ویژگی‌های پیشنهادی ویژگی تبدیل Zak دارای بیشترین بازدهی هستند. در بین طبقه‌بندهای مختلف نیز طبقه‌بند SVM دارای میانگین عملکرد بهتری نسبت به سایر طبقه‌بندها است. علاوه بر این انجام عمل ترکیب طبقه‌بندها نیز باعث شد که نرخ بازشناسی پنج لهجه مختلف، بین ۱۷ تا ۲۸ درصد بهبود پیدا کند.

علاوه بر این یک ویژگی جدید که با ایجاد سه تغییر در الگوریتم اصلی محاسبه ضرایب مل-کپستروم به دست می‌آید، و آن را AGMFCC نامیدیم، برای تشخیص لهجه‌ها در محیط نوبه‌ای استفاده شد؛ که نتایج آزمایش‌ها نشان دهنده بهبود عملکرد سامانه تشخیص لهجه‌ها در محیط نوبه‌ای با استفاده از این ویژگی است.



**حسین مروی** مدرک کارشناسی

خود را در رشته مهندسی الکترونیک از دانشگاه فردوسی اخذ کرد. ایشان مقطع کارشناسی ارشد را در دانشگاه شیراز و در رشته مهندسی برق گرایش مخابرات به

پایان رسانیدند. مدرک دکترای تخصصی خود را در کشور انگلستان، دانشگاه Surrey، مرکز CVSSP و در زمینه پردازش گفتار اخذ کردند. نام‌برده هم‌اکنون به‌عنوان هیئت‌علمی با رتبه دانشیاری در دانشگاه صنعتی شاهرود، دانشکده مهندسی برق و رباتیک مشغول تدریس است. زمینه‌های مورد علاقه ایشان، پردازش سیگنال‌ها، پردازش صوت، تشخیص گفتار، تشخیص گوینده و بهسازی گفتار است.

نشانی رایانامه ایشان عبارت است از:

**h.marvi@shahroodut.ac.ir**



**دانیال دارابی**ان در سال ۱۳۸۶ در

رشته دبیر فنی الکترونیک در دانشگاه حکیم سبزواری پذیرفته و با گذراندن واحدهای درسی لازم در سال ۱۳۹۰ در مقطع کارشناسی فارغ‌التحصیل شد. در مهر همان

سال در رشته مهندسی الکترونیک گرایش سیستم در مقطع کارشناسی ارشد دانشگاه صنعتی شاهرود پذیرفته شد؛ و سپس در شهریور ۱۳۹۲ با دفاع از پایان‌نامه خود در زمینه بازساخت مقاوم گفتار با استفاده از ضرایب مل-کپستروم بهبودیافته، تحصیلات در مقطع کارشناسی ارشد را به پایان برد. هم‌اکنون به‌عنوان هنرآموز در آموزش و پرورش و دانشگاه‌های غیرانتفاعی مشغول به تدریس است.

نشانی رایانامه ایشان عبارت است از:

**Danial.darabian1@gmail.com**

Simon Haykin, Neural Networks, Macmillan College Publishing Company, 1999.

T. G. Dietterich and G. Bakiri(1995), "Solving multiclass learning problems via error correcting output codes", J. of Artificial Intelligence research 2, pp263-286.

T.Chen, C.Huang, E.Chang and J.Wang, "Auto-matic Accent Identification Using Gaussian Mixture Models" Conference on Auto-matic Speech Recognition and Understanding, pp.343-346,Italy, 2001.

T.Stafylakis, P. Kenny, P. Ouellet, J. Perez, M. Kockmann, and P. Dumouchel. "I-Vector-/PLDA Variants for Text-Dependent Speaker Recognition." Preprint submitted to Computer, Speech and Language, 2013.

V. Vapnik and A. Chervonenkis, "The necessary and sufficient conditions for consistency in the empirical risk minimization method," Pattern Recognition and Image Analysis, vol. 1, no. 3, pp. 283-305, 1991.

Y. Freund and R. Schapire.(1997) "A decision-theoretic generalization of on-line learning and an application to boosting". Journal of Computer and System Sciences, 55, 1, pp. 119-139.

Y.Zheng, R.Sproat, L.Gu, I.Shafran, H.Zohu, Y.Su, D.Jurafsky, R.Star, S-Y.Yoon, "Accent detection and speech recognition for Shanghai-Accented Mandarin", Interspeech, pp. 217-220, 2005.

**مجتبی شریف نوقابی** در سال

۱۳۸۶ در رشته دبیر فنی الکترونیک در دانشگاه حکیم سبزواری پذیرفته شد و با گذراندن واحدهای درسی لازم در سال ۱۳۹۰ با رتبه چهارم در بین



دانشجویان ورودی خود در مقطع کارشناسی فارغ‌التحصیل شد. در مهر همان سال در رشته مهندسی الکترونیک گرایش سیستم در مقطع کارشناسی ارشد دانشگاه صنعتی شاهرود پذیرفته شد؛ و سپس در بهمن ۱۳۹۲ با دفاع از پایان‌نامه خود در زمینه تشخیص لهجه‌های گفتار زبان فارسی و با کسب رتبه دوم در بین دانشجویان هم‌ورودی خود، تحصیلات در مقطع کارشناسی ارشد را به پایان برد. هم‌اکنون به‌عنوان هنرآموز در آموزش و پرورش و دانشکده‌های فنی مشغول به تدریس است.

نشانی رایانامه ایشان عبارت است از:

**mojtabasharif@chmail.ir**