



Computational Intelligence in Electrical Engineering  
Vol. 11, No. 1, 2020  
Research Paper

## Alzheimer Speech Signal Analysis of Persian speaking Alzheimer's patients

Mahdiyeh Rahmani<sup>1</sup>, Maryam Momeni<sup>2</sup>

<sup>1</sup> Faculty of Engineering, University of Arak, Arak, Iran  
matie.rah@gmail.com

<sup>2</sup> Faculty of Engineering, University of Arak, Arak, Iran  
m-momeni@araku.ac.ir

### Abstract:

Alzheimer's is a type of brain dementia that gradually reduces mental abilities of the patient. The lack of memory, decision-making disorder, and mistakes in choosing the correct vocabulary are the early symptoms of Alzheimer's disease. Therefore, extensive studies have been conducted on the diagnosis of Alzheimer's disease using the non-invasive speech signal recognition method. Identifying of Alzheimer's disease is dependent on culture and language, speech content, gender, age, accent, and many other factors. Therefore, Alzheimer's speech signal has been studied in various languages. The purpose of this paper is to recognize Alzheimer's patients from healthy people by the use of their speech signal processing in Persian using the combination of time, frequency, and frequency-temporal features. In this paper, after pre-processing, the speech features extracted using the wavelet packet as a frequency-temporal feature next to Mel frequency Cepstral coefficients, zero crossing rate, spectral roll off, band width, root mean square and spectral centroid frequency. Finally, the extracted features have been classified by the support vector machine which achieves recognition precision of 96% on Persian healthy and Alzheimer's speaker experiments. The acceptable results demonstrate the applicability of the proposed non-invasive and low-cost algorithm for the diagnosis of Persian-speaking Alzheimer's patients.

**Keywords:** Alzheimer's disease, Speech Signal, Discrete Wavelet Transform, Support Vector Machine.



2252-083X/ © 2020 The Authors. Published by University of Isfahan

This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



<http://dx.doi.org/10.22108/isee.2019.109797.1109>

## تحلیل سیگنال گفتار بیماران آلزایمری فارسی زبان

مهدیه رحمانی<sup>۱</sup>، مریم مؤمنی<sup>۲</sup>

۱- کارشناسی مخابرات، دانشکده فنی و مهندسی، دانشگاه اراک، اراک، ایران

matie.rah@gmail.com

۲- استادیار، گروه مهندسی برق، دانشکده فنی و مهندسی، دانشگاه اراک، اراک، ایران

m-momeni@araku.ac.ir

**چکیده:** آلزایمر یک نوع اختلال عملکرد مغزی است که به تدریج توانایی‌های ذهنی بیمار تحلیل می‌رود؛ از جمله علائم اولیه این بیماری فقدان حافظه، اختلال در تصمیم‌گیری و اشتباه در انتخاب واژگان درست است؛ بنابراین، پردازش سیگنال گفتار این بیماری توجه بسیاری از پژوهشگران را در دهه اخیر جلب کرده است. تشخیص بیماری آلزایمر با استفاده از سیگنال گفتار به فرهنگ و زبان و محتوای گفتار، جنسیت، سن، لهجه و بسیاری از عوامل دیگر وابسته است؛ از این رو، سیگنال گفتار بیماران آلزایمری در زبان‌های مختلف بررسی شده است. هدف این مقاله تشخیص بیماران آلزایمری از افراد سالم با استفاده از پردازش سیگنال گفتار آنها در زبان فارسی با ترکیب ویژگی‌های زمانی، فرکانسی و زمانی - فرکانسی است. در این مقاله پس از پیش‌پردازش سیگنال گفتار فارسی با بهره‌گیری از بسته موجک، به‌عنوان ویژگی زمان - فرکانس در کنار ضرایب کپسترال فرکانس مل، نرخ عبور از صفر، افت طیف، پهنای باند، انرژی سیگنال و فرکانس مرکز طیفی، ویژگی‌های سیگنال گفتار بیماران آلزایمری و افراد سالم استخراج شدند و دقت طبقه‌بندی نتایج با ماشین بردار پشتیبان، نتیجه ۹۶٪ را دربرداشت. نتایج پذیرفتنی نشان‌دهنده الگوریتم پیشنهادی غیرتهاجمی و کم‌هزینه در تشخیص بیماران آلزایمری فارسی‌زبان است.

**واژه‌های کلیدی:** بیماری آلزایمر، تبدیل موجک گسسته، سیگنال گفتار، ماشین بردار پشتیبان

### ۱- مقدمه

واقع شود. همچنین، تجزیه و تحلیل سیگنال گفتار در بسیاری از موارد برای ایجاد ارتباط بهتر و آسان تر بین انسان و ماشین صورت گرفته است [۲]. این حقیقت سبب شده است پژوهشگران علاقه زیادی به سیگنال گفتار داشته باشند تا بتوانند ارتباط بین ماشین و انسان را به این طریق به وجود آورند که پردازش خودکار گفتار (ASR) این امکان را برای پژوهشگران فراهم کرده است [۳].

همچنین، تشخیص بیماری آلزایمر با استفاده از سیگنال گفتار به فرهنگ و زبان و محتوای گفتار، جنسیت، سن، لهجه و بسیاری از عوامل دیگر وابسته است [۴]؛ از این رو سیگنال گفتار بیماران آلزایمری در زبان‌های فارسی [۵]، اسپانیایی [۶]، انگلیسی [۷]، فرانسوی [۸]، کاتالان، چینی، باسک، عربی، پرتغالی [۹] و روسی [۱۰] بررسی شده است. به طور کلی بررسی سیگنال گفتار بیماران آلزایمری شامل دو مرحله استخراج ویژگی و طبقه بندی است [۱۱]. به دست

آلزایمر یک نوع اختلال عملکرد مغزی است که به تدریج توانایی‌های ذهنی بیمار تحلیل می‌رود. علائم اولیه این بیماری فقدان حافظه، اختلال در تصمیم‌گیری و اشتباه در انتخاب واژگان درست است. یکی از نخستین بخش‌های تأثیر گرفته از این بیماری، قسمتی از قشر مغز است که توانایی پردازش صحبت را دارد [۱]؛ بنابراین، پردازش سیگنال گفتار این بیماری می‌تواند در تشخیص آن مفید

۱ تاریخ ارسال مقاله: ۱۳۹۶/۱۲/۰۵

تاریخ پذیرش مقاله: ۱۳۹۸/۰۹/۰۴

نام نویسنده مسئول: مریم مؤمنی

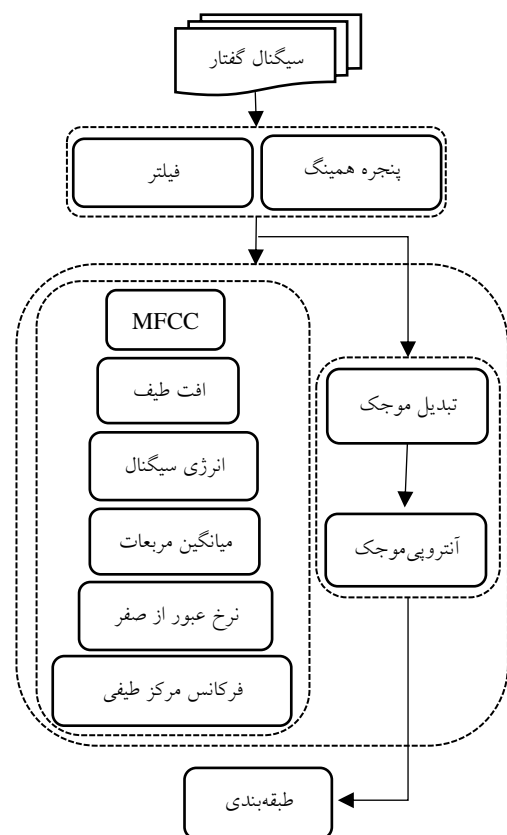
نشانی نویسنده مسئول: ایران - اراک - دانشگاه اراک - دانشکده فنی و مهندسی - گروه مهندسی برق

از صفر ( $ZC^V$ )، میانگین مربعات ( $RMS^A$ )، پهنای باند ( $BW^A$ ) و افت طیف ( $SR^1$ ) می‌توانند مفید واقع شوند [۲۰].

در مقاله حاضر نیز از این ویژگی‌ها استفاده شده است.

همچنین، تبدیل موجک در تحلیل سیگنال گفتار استفاده می‌شود؛ زیرا تجزیه و تحلیل را با رزولوشن چندمنظوره انجام می‌دهد و امروزه در تحلیل سیگنال‌های بیولوژیکی بسیار استفاده می‌شود [۲۱، ۲۲]. بسته موجک، تعمیم‌یافته تبدیل موجک است و به دلیل حجم بالای اطلاعات استخراج‌شده، آنتروپی در گره‌های درخت موجک به‌عنوان بردار ویژگی استفاده می‌شود [۲۳-۲۸]. در مقاله حاضر نیز از بسته موجک برای تکمیل بردار ویژگی استفاده شده است؛ زیرا بسته موجک به‌طور چشمگیری تفاوت بین سیگنال‌ها را نشان می‌دهد [۲۶].

در این مقاله، پس از اخذ سیگنال گفتار و پیش پردازش، ویژگی‌های زمانی، فرکانسی و زمانی - فرکانسی سیگنال گفتار بررسی و طبقه‌بندی شده و در نهایت، تحلیل نتایج صورت گرفته است (شکل ۱).



شکل (۱): بلوک دیاگرام مراحل انجام کار

آوردن بهترین نتیجه و استخراج درست ویژگی در مرحله نخست بر روند انجام مرحله دوم تأثیر می‌گذارد. در مرحله نخست باید اطلاعاتی از سیگنال گفتار استخراج شود که حداکثر تفاوت را بین ویژگی‌های افراد سالم و بیماران آلزایمری داشته باشد.

نشان داده شده است که ویژگی‌های نوایی مثل فرکانس پایه، فرمت‌ها [۱۲، ۱۳]، ویژگی‌های زبان و فرکانس مانند ضرایب کپسترال فرکانس مل ( $MFCC^1$ ) در این حوزه بسیار پرکاربرد است [۱۴]. در مرجع [۹]، ابعاد فرکتال و پارامترهای خطی برای بهبود عملکرد سیستم تشخیص بیماری آلزایمر بررسی می‌شود. در مرجع [۸]، نخستین نشانگرهای صوتی از سیگنال گفتار سه سطح از بیماران آلزایمری استخراج شده و اختلال شناختی بیماران آلزایمری در مراحل اولیه و افراد سالم ارزیابی شده است. در مرجع [۱۵]، ویژگی‌های وابسته به تغییرات زمانی سیگنال گفتار در بیماران آلزایمری ارزیابی شده است. در مرجع [۱۶]، اهمیت حروف بی‌صدا در دستیابی به نتایج مطلوب در تشخیص بیماری نشان داده شده است. در مرجع [۱۷]، شدت و سطح بیماری آلزایمر با استفاده از روان صحبت کردن و تشخیص احساس بیماران بررسی شده که برای شناسایی احساس از ویژگی‌های پروزودیک<sup>۲</sup> و زبان‌شناختی سیگنال گفتار استفاده شده است. در مرجع [۱۰]، ویژگی‌هایی از سیگنال گفتار همچون نقش کلمات در جمله، گرامر صحیح، تلفظ کلمات، روانشناسی لغات و  $MFCC$  بیماران روسی زبان بررسی شده است. در مرجع [۶]، ویژگی‌هایی همچون تعداد فعل، اسم، حروف ربط و روش‌های آماری از سیگنال گفتار بیماران آلزایمری اسپانیایی زبان بررسی شده و طبقه‌بندی با استفاده از ماشین بردار پشتیبان ( $SVM^3$ ) با کرنل تابع پایه شعاعی ( $RBF^4$ ) صورت گرفته است.

مطالعات نشان می‌دهند حدود دو سوم از سیگنال گفتار صدادار است و نقش بسیار مهمی در گفتار دارد که به دلیل طبیعت متناوب می‌تواند شناسایی و استخراج شود [۱۸، ۱۹]. بررسی‌های انجام‌شده نشان می‌دهند طنین حروف صدادار و بی‌صدا در بیماران آلزایمری تفاوت بسیاری دارند؛ بنابراین، ویژگی‌هایی که حروف صدادار در آنها نقش مهمی بر عهده دارند، مانند انرژی، فرکانس مرکزی طیف ( $SC^1$ )، نرخ عبور

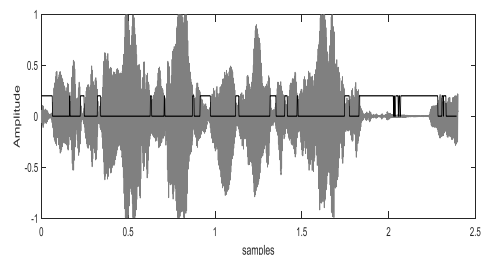
## ۲- روش

### ۲-۱- داده

صدای ضبط شده از بیماران آلزایمری در مرکز نگهداری بیماران آلزایمری شایستگان با میکروفون **HP MMDEVAPI SWD** جمع آوری شده است. اطلاعات به دست آمده شامل ۱۰۰ داده از افراد سالم و بیماران آلزایمری در فاصله سنی ۵۰-۸۵ سال است. داده های پردازش شده در محیطی دوستانه، جمع آوری و از بیماران سؤالاتی مشخص در آرامش پرسیده شده است. مدت زمان مکالمه ضبط شده هریک از مصاحبه شوندهگان حدود ۷۰ دقیقه و در فرمت **.wav** است. این پژوهش، مجوز اخلاقی از مرکز مربوطه دارد. داده های اخذ شده با پردازشگر **i5-5200U CPU 2.20GHz** و با نرم افزار متلب ۲۰۱۸ (**MATLAB**) پردازش شده اند.

### ۲-۲- پیش پردازش

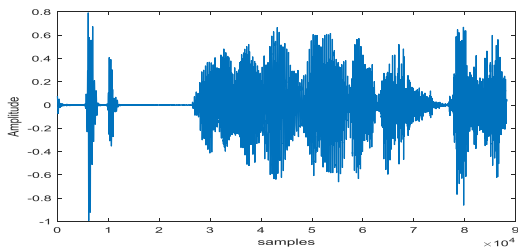
ابتدا نویز موجود در داده های جمع آوری شده، حذف و سیگنال گفتار نرمالیزه می شود. سپس با توجه به ماهیت غیرایستنا سیگنال گفتار، با استفاده از پنجره همینگ، به قسمت های ۵۰-۳۰ میلی ثانیه با همپوشانی ۵۰٪، تقسیم و سکوت در سیگنال گفتار با عبور از صفر و انرژی سیگنال حذف می شود. شکل موج سیگنال گفتار و سکوت (نمودار پالس) در شکل (۲) مشاهده می شود؛ هر کجا الگوریتم تشخیص سکوت بدهد (مکان هایی که مقدار انرژی و نرخ عبور از صفر، از آستانه های مربوطه کمتر است)، نمودار پالس مقدار ۰/۲ می گیرد، در غیر این صورت مقدار آن صفر است. پس از حذف سکوت و فیلتر کردن، سیگنال گفتار با نرخ یکسان نمونه برداری می شود.



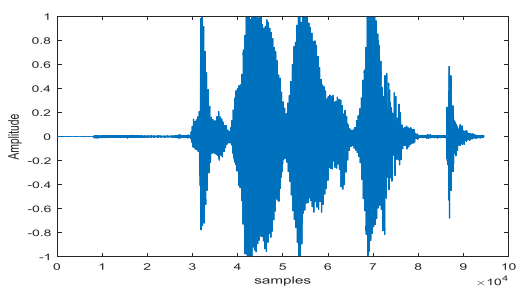
شکل (۲): شکل موج سیگنال گفتار و سکوت (نمودار پالس)

## ۲-۳- استخراج ویژگی

گفتار بیماران آلزایمری آرام تر و با تعداد مکث های زیادی است و برای پیدا کردن کلمه مناسب مدت زمان زیادی را صرف و بیشتر جملات خود را ناتمام رها می کنند. در سیگنال گفتار افراد سالم، حروف صدادار به خوبی مشهودند و تعداد بیشتری حرف صدادار دارد؛ در حالی که در سیگنال گفتار بیماران آلزایمری حروف صدادار کاهش یافته اند و در نتیجه، انرژی سیگنال کمتر می شود؛ زیرا حروف صدادار دارای سطح بالایی از انرژی اند. همچنین، بیان جملات مشابه در بیماران آلزایمری سرعت کمتری نسبت به افراد سالم دارد. بنابراین، سبب افزایش طول سیگنال گفتار برای جملات و کلمات مشابه می شود؛ در نتیجه، این ویژگی ها بر انرژی و پهنای باند سیگنال تأثیر می گذارند. در شکل (۳)، سیگنال گفتار بیمار آلزایمری و فرد سالم نشان داده شده است. انرژی سیگنال در بیمار آلزایمری نسبت به فرد سالم ۰/۷۸۸۰ است.



(الف)



(ب)

شکل (۳): سیگنال گفتار بیمار آلزایمری (الف)

و فرد سالم (ب)

در این مطالعه، ویژگی های استخراج شده به سه دسته تقسیم می شوند:

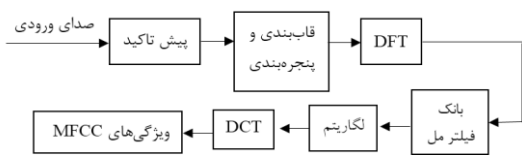
می‌شود به  $m$  باند فرعی تقسیم شود. با فرض کمترین فرکانس  $f_m$  و بیشترین فرکانس  $f_M$  در  $m$  امین باند فرعی، SC در  $m$  امین باند فرعی محاسبه می‌شود.

SR، ۹۵٪ از توزیع طیفی توان را مشخص می‌کند. این اندازه گیری قسمت های صدا دار و بی صدای گفتار را از هم متمایز می‌کند. کلمات بی صدا محدود و وسیعی از طیف فرکانسی را شامل می‌شوند؛ اما بیشتر انرژی مربوط به کلمات بی صدا در باندهای پایینی قرار دارند [۲۹].

محاسبه ضرایب MFCC، روش رایج و معروف در زمینه استخراج ویژگی سیگنال گفتار است. مطالعات نشان می‌دهند سیستم شنوایی انسان، سیستم خطی نیست؛ به این صورت که برای مؤلفه‌های زیر ۱۰۰۰ هرتز، به صورت خطی و برای مؤلفه‌های بالای ۱۰۰۰ هرتز به صورت لگاریتمی عمل می‌کند [۳۰]. این امر سبب آشکار شدن ویژگی‌ها و مشخصات آکوستیکی سیگنال گفتار می‌شود. براساس رابطه (۴)، فرکانس  $f$  در مقیاس مل حاصل می‌شود:

$$mel(f) = (2595 \log_{10} (1 + \frac{f}{700})) \quad (4)$$

بلوک دیاگرام محاسبه MFCC در شکل (۴) نمایش داده شده است. برای محاسبه ضرایب، ابتدا سیگنال گفتار با اعمال پنجره همینگ با طول ۲۵۶ نمونه پنجره گذاری می‌شود که در ۱۲۸ نمونه همپوشانی دارند. استفاده از پنجره همینگ باعث می‌شود کمترین اطلاعات فرکانسی از دست برود.



شکل (۴): بلوک دیاگرام محاسبه MFCC

بعد از مرحله پنجره گذاری تبدیل فوری سریع روی هر فریم، اعمال و سپس بانک فیلتری مثلثی در مقیاس مل اعمال می‌شود و در مرحله نهایی با اعمال تبدیل کسینوس با رابطه (۵)، ضرایب از مقیاس مل به مقیاس زمانی تبدیل می‌شوند:

$$G_n = \sum_{j=1}^p m_k \cos(n(k - \frac{1}{2})\frac{\pi}{j}) \quad (5)$$

• حوزه زمان: ZC و RMS؛

• حوزه فرکانس: BW، SR، SC و MFCC؛

• حوزه زمان - فرکانس: درخت موجک.

استخراج ویژگی ZC یکی از رایج ترین روش ها برای تحلیل سیگنال گفتار است. نرخ عبور از صفر، نرخ تغییرات علامت در طول یک سیگنال است؛ یعنی نرخی که سیگنال از مثبت به صفر تا منفی یا از منفی به صفر تا مثبت تغییر می‌کند. به کمک رابطه زیر برای هر فریم نرخ عبور از صفر به دست می‌آید:

$$ZCR_i = \frac{1}{N} \sum_{n=1}^{N-1} \frac{|sgn[s_i(n)] - sgn[s_i(n-1)]|}{2} \quad (1)$$

در رابطه بالا  $s_i(n)$  نمونه‌های سیگنال مربوط به فریم  $i$ ام و  $N$  طول فریم است که برابر با ۴۸۰ نمونه (۳۰ میلی ثانیه) در نظر گرفته شده‌اند.

به منظور محاسبه مقدار RMS سیگنال های متناوب، روش های مختلفی وجود دارند که در اینجا از انتگرال یکپارچه سازی داده های پیوسته و از رابطه زیر به دست آمده است [۲۷]:

$$f_{RMS} = \sqrt{\frac{1}{T} \int_{t_1}^{t_1+T} f^2(t) dt} \quad (2)$$

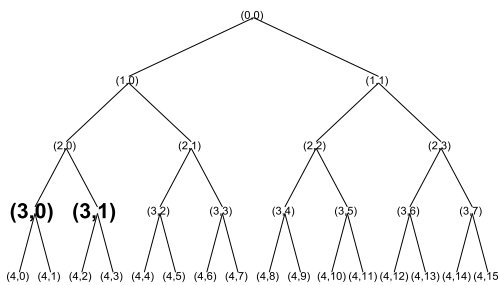
که در آن  $f_{RMS}$  مقدار RMS سیگنال  $f(t)$  در بازه بین  $t_1$  و  $t_1 + T$  دوره تناوب سیگنال  $f(t)$  است.

SC تخمینی از مرکز ثقل طیف در هر باند فرعی است؛ در اصل به عنوان یک ویژگی برای سیستم تشخیص گفتار ارائه شده است. طیف مرکزی فرکانس، میانگین وزن های فرکانسی باندهای فرعی را محاسبه می‌کند که در آن وزن ها نرمالیزه شده اند و بیان کننده انرژی هریک از مؤلفه های فرکانسی اند [۲۸]:

$$SC_m = \frac{\sum_{f=l_m}^{u_m} f |W_m[f] S[f]|}{\sum_{f=l_m}^{u_m} |W_m[f] S[f]|} \quad (3)$$

که  $S[f]$  طیف سیگنال گفتار است و می‌تواند با استفاده از فیلتر گابور<sup>۱۱</sup> که در حوزه فرکانس با  $W_m[f]$  نمایش داده

تقریب ادامه می یابد؛ درحالی که در بسته موجک همین روند روی هر دو گروه تقریب و جزییات تکرار می شود. بنابراین، تجزیه و تحلیل سیگنال به کمک بسته موجک نسبت به تبدیل موجک اطلاعات بیشتری خواهد داشت [۳۳]. افزایش عمق پیشروی بسته موجک سبب می شود تحلیل سیگنال به صورت جزئی تر و کامل تر بررسی شود؛ اما حجم و زمان محاسبات افزایش می یابد. شکل (۵)، درخت بسته موجک با عمق پیشروی ۴ شامل ۳۱ گره را نشان می دهد.



شکل (۵): درخت بسته موجک با عمق پیشروی ۴

حجم بالای ضرایب و چند بعدی بودن آنها سبب می شود از آنتروپی شانون استفاده شود. آنتروپی معیاری از اطلاعات موجود در هر گره است. محاسبه اعداد و ویژگی های استخراج شده از گره ها در بسته موجک، با استفاده از آنتروپی شانون، اطلاعات بسیار ارزشمندی را در اختیار ما قرار می دهد و کاهش حجم اطلاعات و محاسبات را سبب می شود [۳۴].

## ۲-۴- طبقه بندی

عملکرد سیستم های ASR مستقیماً به ابزار طبقه بندی استفاده شده بستگی دارد. به طور کلی دو نوع ابزار طبقه بندی کننده وجود دارد؛ مولد<sup>۱۲</sup> و تشخیصی<sup>۱۳</sup> [۳۵]. مدل مولد سعی می کند پراکندگی داده ها را در یک کلاس خاص تخمین بزند، داده های رقابتی را نادیده می گیرد و فقط اطلاعات مرتبط را در نظر می گیرد؛ مدل مخفی مارکف از این نوع است. مدل تشخیصی سعی می کند تابع طبقه بندی را پیش بینی کند و خطاهای طبقه بندی را تا حد ممکن کاهش دهد؛ SVM [۳۶] و شبکه های عصبی [۳۷] از این نوع اند. اساس کار SVM افزایش فاصله بین نمونه ها و مشخص کردن مرز طبقه بندی است. این فاصله، حاشیه

که در آن  $C_n$  ضریب نام ضرایب MFCC،  $m_j$  خروجی  $j$ امین کانال بانک فیلتر،  $p$  تعداد کانال های بانک فیلتر است و  $n = 1, 2, \dots, p$  است. در این مقاله تعداد کانال های فیلتر، ۴۰ و تعداد ضرایب انتخابی ۱۹ در نظر گرفته شده است.

تبدیل موجک اطلاعات دقیقی از سیگنال گفتار را دربردارد و این اجازه را می دهد تا در مقیاس های متفاوت سیگنال بررسی شود [۳۱]. تابع موجک برای ایجاد تعادلی بین حوزه زمان (بازه محدود) و حوزه فرکانس (پهنای باند محدود) به کار می رود. با تحلیل و تغییر اندازه موجک مادر، اجزای فرکانس پایین و بالا را به ترتیب در مقیاس های بزرگ و کوچک می توان مشاهده کرد [۳۲]. تبدیل موجک از رابطه (۶) به دست می آید:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \omega\left(\frac{t-b}{a}\right) \quad (6)$$

که در آن،  $\omega(t)$  تبدیل ویولت اصلی است و  $a$  و  $b$  به ترتیب ضرایب مقیاس و جابجایی اند. معمولاً ضرایب  $a$  و  $b$  در تبدیل موجک گسسته به شکل زیر تعیین می شوند:

$$\begin{aligned} a &= a_0^j \\ b &= kb_0 a_0^j \end{aligned} \quad (7)$$

که در آن  $j$  و  $k$  عدد صحیح اند؛ بنابراین، تبدیل موجک گسسته از رابطه زیر به دست می آید:

$$\psi_{j,k} = a_0^{-j/2} \psi(a^{-j}t - kb_0) \quad (8)$$

گرچه پارامترهای این تبدیل گسسته اند، این تبدیل در حوزه زمان پیوسته است و ضرایب تجزیه موجک به کمک رابطه زیر محاسبه می شود:

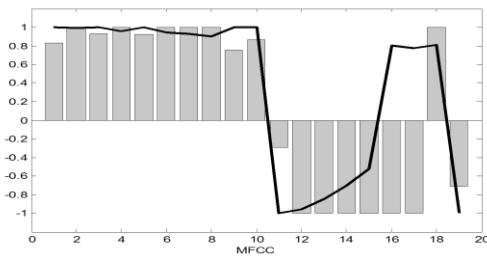
$$\begin{aligned} d_{j,k} &= \langle f_\omega(t), \psi_{j,k}(t) \rangle \\ &= \frac{1}{a_0^{j/2}} \int f_\omega(t) \psi(a_0^{-j}t - kb_0) dt \end{aligned} \quad (9)$$

که

$$f_\omega(t) = \sum_j \sum_k \langle f_\omega(t), \psi_{j,k}(t) \rangle \psi_{j,k}(t) \quad (10)$$

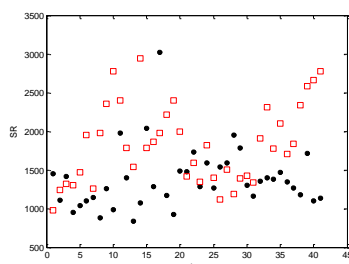
تبدیل موجک کمک شایان توجهی در تحلیل و استخراج ویژگی سیگنال می کند. سیگنال در تبدیل موجک به دو گروه تقریب و جزییات دسته بندی می شود و این روند روی نتایج

است. ضرایب ابتدایی (ضریب ۱-۱۱) نشان‌دهنده چگونگی توزیع انرژی در طیف سیگنال گفتار است و دیگر ضرایب (۱۲-۱۹) جزئیات طیفی سیگنال را بررسی می‌کنند؛ به همین دلیل استفاده از این ضرایب در تحلیل سیگنال گفتار مطلوب است. در شکل (۶) نیز این ضرایب تفاوت بیشتری در دو گروه دارند.

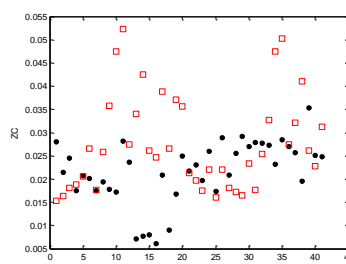


شکل (۶): نسبت ضرایب MFCC در دو گروه افراد سالم (خط ممتد) و بیماران آلزایمری (مستطیل)

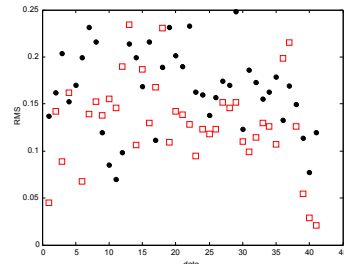
ضرایب حاصل در بردار ویژگی از MFCC، در دو گروه به صورت چشم‌گیری تغییر کرده‌اند که تفاوت‌های فرکانسی سیگنال‌های گفتار ارزیابی شده در بیماران آلزایمری و افراد سالم فارسی‌زبان، به این امر بسیار کمک کرده‌اند. این تفاوت‌ها که از ویژگی‌های حروف و تأثیرات بیماری آلزایمری بر سیگنال گفتار حاصل می‌شوند، ویژگی‌های مناسب برای طبقه‌بندی در نظر گرفته شده‌اند.



(ج)

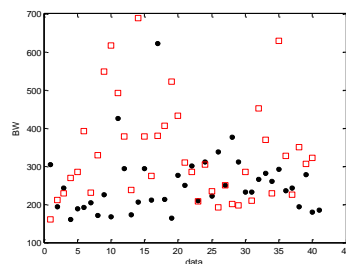


(ب)

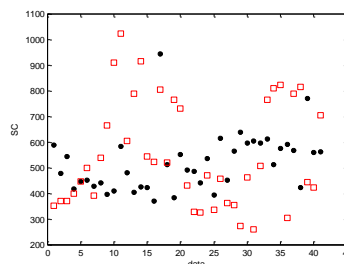


(الف)

شکل (۷): پراکندگی ویژگی‌های RMS (الف)، ZC (ب)، SR (ج)، SC (د) و BW (ه) مربوط به بیماران آلزایمری (مربع) و افراد سالم (نقطه)



(د)



(ه)

شناخته می‌شود و با افزایش آن قادر به تعمیم الگوهای ناشناخته است. راه‌حل افزایش حاشیه، به SVM این اجازه را می‌دهد تا بیشترین طبقه‌بندی‌های غیرخطی را در حضور نویز، یکی از مشکلات ASR، انجام دهد. همچنین، SVM‌ها مشکلات همگرایی و پایداری معمول که بیشتر شبکه‌های عصبی دارا هستند را ندارند. مفهوم اساسی نهفته در SVM، کاهش خطای ساختاری است [۳۱]. یک دستگاه یادگیری به گونه‌ای انتخاب می‌شود که علاوه بر به حداقل رساندن خطای آزمون، تعمیم‌پذیری مناسبی در طبقه‌بندی و تخمینی از نسبت بردارهای طبقه‌بندی‌شده بر کل بردارهای آموزش داشته باشد [۴].

### ۳- نتایج

کارایی روش پیشنهادی در این مقاله با استفاده از داده‌های جمع‌آوری شده از بیماران آلزایمری فارسی‌زبان بررسی شده است. شکل (۶)، اختلاف میانگین ویژگی‌های استخراج‌شده نوزده ضریب MFCC در افراد سالم و آلزایمری را نشان می‌دهد که نرخ تغییرات در طیف سیگنال گفتار را مشخص می‌کند. مقادیر مثبت ضرایب MFCC نشان‌دهنده وجود انرژی بیشتر است و متعاقباً ضرایب منفی انرژی بسیار کمی را در سیگنال گفتار نشان می‌دهند [۳۸]. مطابق شکل (۶)، تعداد ضرایب مثبت در افراد سالم بیشتر است که نشان‌دهنده انرژی بیشتر در سیگنال گفتار این افراد

عمق پیشروی‌های متفاوت درخت موجک مادر db4 در تشخیص بیماری آلزایمر در جدول (۲) نشان داده شده‌اند. بسته موجک با عمق پیشروی ۴ به سیگنال گفتار اعمال و اطلاعات موجود در گره‌های مشخص شده در شکل (۵)، در شکل (۹) نشان داده شده است.

جدول (۱): نرخ تشخیص به‌ازای انواع موجک db

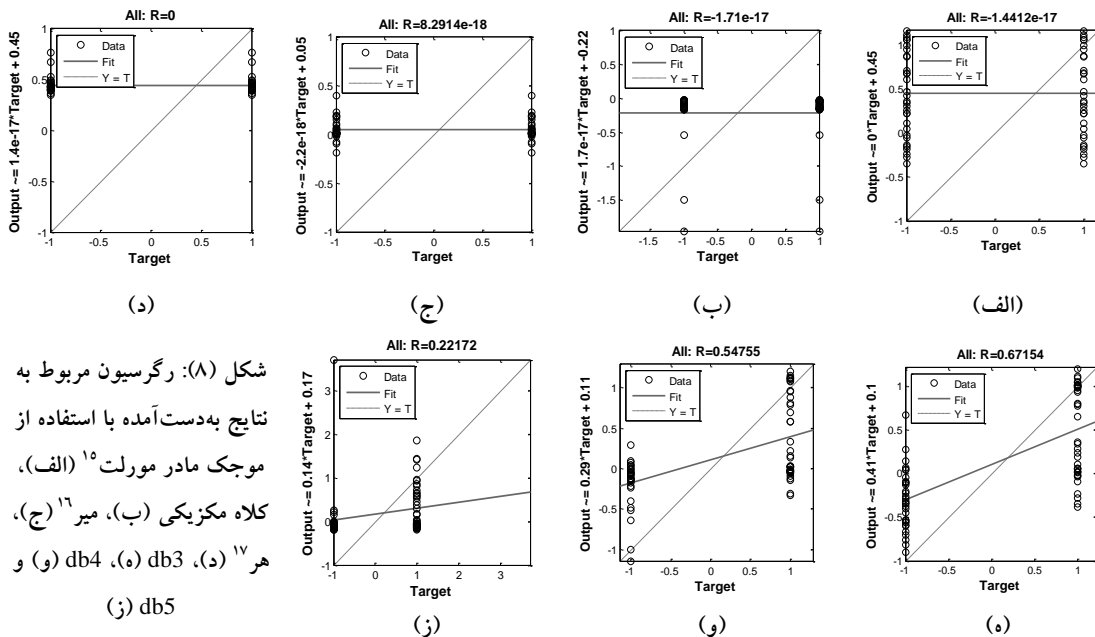
نوع موجک	نوع آنتروپی	درصد تشخیص
Db2	شانون	۸۳/۷۸
Db3	شانون	۸۶/۴۸
Db4	شانون	۹۱/۸۹
Db5	شانون	۸۹/۱۸

جدول (۲): نرخ تشخیص به‌ازای عمق پیشروی‌های متفاوت موجک مادر db4

عمق پیشروی	ضرایب آنتروپی	درصد تشخیص
۲	۷	۸۰
۳	۱۵	۸۶/۸
۴	۳۱	۸۹
۵	۶۳	۸۲/۵

شکل (۷) نشان‌دهنده پراکندگی ویژگی‌های RMS، SC، SR، ZC و BW است و میزان تفاوت ویژگی‌های استخراج شده در هر دو گروه را نشان می‌دهد؛ برای مثال، در شکل (۷-الف) که تفاوت ویژگی‌های استخراج شده حاصل از RMS را نشان می‌دهد، سیگنال گفتار از دو گوینده آقا (فرد سالم و آلزایمری) با گفتن جمله «من یک کشاورز هستم» به دست آمده است. پس از محاسبه ویژگی از سیگنال گفتار مشاهده می‌شود ویژگی استخراج شده بین افراد سالم (نقطه در شکل ۷-الف) و آلزایمری (مربع در شکل ۷-الف) متفاوت است که از این ویژگی در کنار دیگر ویژگی‌های به‌دست آمده برای طبقه‌بندی استفاده شده است. همان‌طور که در شکل (۸) مشاهده می‌شود، رگرسیون موجک‌های مادر دایبچیز (db) نسبت به سایر موجک‌ها بالاتر است. همچنین، آنتروپی شانون موجک مادر db4 بهترین عملکرد را بین سایر موجک‌های db داشته است (جدول ۱)؛ بنابراین، در این مطالعه از این موجک استفاده شده است.

با توجه به تأثیر عمق پیشروی در بسته موجک بر نتایج به‌دست آمده، عمق پیشروی‌های متفاوتی، بررسی و سپس ویژگی‌های استخراج شده با SVM ارزیابی و طبقه‌بندی شد.



شکل (۸): رگرسیون مربوط به نتایج به‌دست آمده با استفاده از موجک مادر مورلت<sup>۱۵</sup> (الف)، کلاه مکزیکی (ب)، میر<sup>۱۶</sup> (ج)، هر<sup>۱۷</sup> (د)، db3 (ه)، db4 (و) و db5 (ز)



$$\text{دقت} = \frac{TP}{TP + FP} \quad (14)$$

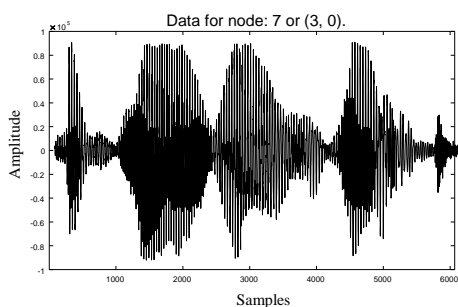
که در آن، TP مثبت درست، TN منفی درست، FP مثبت نادرست و FN منفی نادرست است. نتایج صحت طبقه بندی برای بردارهای ویژگی متفاوتی در جدول (۳) نشان داده شده‌اند و نتایج به دست آمده براساس تغییر نوع کرنل بررسی شده‌اند. همان‌طور که در جدول (۳) نشان داده شده است، بهترین نتایج از طبقه بندی با کرنل چندجمله‌ای درجه ۲ حاصل شده‌اند. چهار معیار صحت، حساسیت، تشخیص و دقت به همراه بازه اطمینان (CI)  $^{(A)}$  ۹۵٪ برای بردار ویژگی که بهترین عملکرد را در طبقه بندی SVM با کرنل چندجمله‌ای درجه ۲ داشته است، به ترتیب برابر  $(CI = /۸۹/۲ - /۹۹/۰)$ ،  $۹۳/۰$ ،  $(CI = /۸۵/۶ - /۹۷/۲)$  و  $(CI = /۹۰/۵ - /۹۹/۵)$  و  $۴۰/۰$  (CI =  $/۳۰/۵ - /۵۰/۳$ ) شده‌اند.

گره فرد (۳،۰) و زوج (۳،۱) به ترتیب خروجی‌های فیلتر بالاگذر و پایین‌گذر را نشان می‌دهند. شکل‌های (۹-ب) و (۹-د)، جزئیات سیگنال و شکل‌های (۹-الف) و (۹-ج) کلیات سیگنال گفتار را نشان می‌دهند. جزئیات سیگنال دو گروه نشان‌دهنده تفاوت فرکانسی زیاد دو گروه و در نتیجه، تفاوت حروف صادادار آنهاست. پس از پیش‌پردازش و انتخاب موجک مادر مناسب، نسبت داده‌های آموزش و تست ۶۰ به ۴۰ درصد انتخاب شد و با استفاده از طبقه بند SVM با توابع کرنل مختلف داده‌ها طبقه بندی شده‌اند. به کمک روابط (۱۱-۱۴) چهار معیار صحت، حساسیت، تشخیص و دقت عملکرد طبقه بندی ارزیابی شدند:

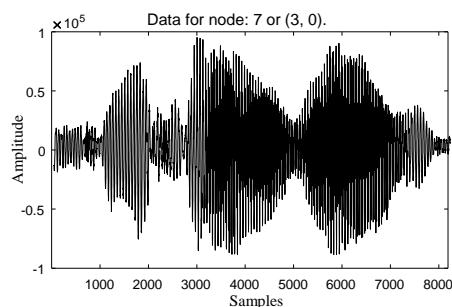
$$\text{صحت} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (11)$$

$$\text{حساسیت} = \frac{TP}{TP + FN} \quad (12)$$

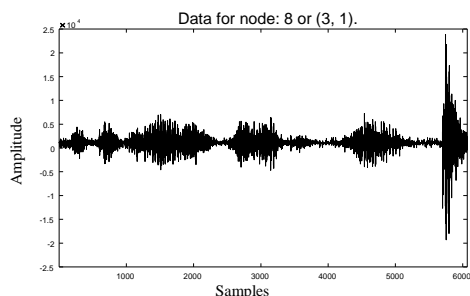
$$\text{تشخیص} = \frac{TN}{TN + FP} \quad (13)$$



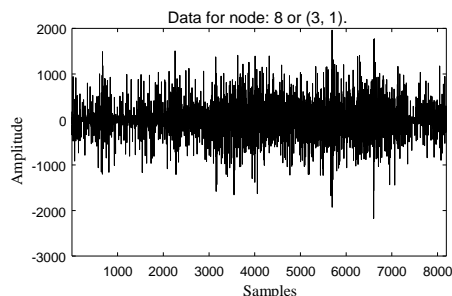
(ج)



(الف)



(د)



(ب)

شکل (۹): اطلاعات موجود در گره‌های مشخص شده در درخت موجک شکل (۵)، شکل‌های (الف) و (ب) به ترتیب نشان‌دهنده سیگنال گره‌های (۳،۰) و (۳،۱) مربوط به بیمار آلزایمری و شکل‌های (ج) و (د) نظیر گره‌های گفته شده در فرد سالم

جدول (۳): نتایج صحت طبقه‌بندی با SVM

نوع کرنل	ZC+SR+SC +BW+RMS	MFCC	Wavelet Entropy	MFCC+ Wavelet Entropy	MFCC+ ZC+SR+SR +BW+RMS	MFCC+ ZC+SR+SR+BW+RMS Wavelet Entropy
چندجمله‌ای درجه ۲	٪۸۰	٪۸۶	٪۸۷	٪۹۱	٪۸۴	٪۹۳
چندجمله‌ای درجه ۳	٪۸۲	٪۸۷	٪۸۹	٪۸۹	٪۸۰	٪۹۰
چندجمله‌ای درجه ۴	٪۹۱	٪۸۴	٪۸۴	٪۸۴	٪۷۹	٪۸۲
چندجمله‌ای درجه ۵	٪۹۰	٪۸۳	٪۹۱	٪۸۳	٪۷۵	٪۸۳
چندجمله‌ای درجه ۶	٪۸۹	٪۷۱	٪۸۹	٪۶۸	٪۵۲	٪۷۳
شبکه عصبی توابع ۰/۰۰۱=شعاعی	٪۶۹	٪۵۴	٪۷۹	٪۷۳	٪۶۸	٪۷۳
توابع ۰/۰۱=شعاعی	٪۷۵	٪۷۵	٪۷۵	٪۷۵	٪۵۰	٪۷۵
توابع ۰/۱=شعاعی	٪۷۳	٪۷۵	٪۷۵	٪۷۵	٪۵۰	٪۷۵

## ۴- بحث و نتیجه‌گیری

گفتار یکی از مهم‌ترین راه‌های ارتباط انسانی است و به‌دلیل ماهیت آن، برای تشخیص زوال عقل استفاده می‌شود [۳۹]. در این مقاله، روش پیشنهادی با استفاده از داده‌های جمع‌آوری شده از بیماران آلزایمری فارسی‌زبان مطالعه شده است. در اینجا هدف بررسی ویژگی‌های مستقل از گوینده است و تنوع جملات در داده‌های تست و آموزش رعایت شده است. به این صورت که جملات یکسان در اخذ داده افراد سالم و آلزایمری استفاده شده‌اند؛ درحالی‌که تنوع جملات در هر دو گروه به‌منظور داشتن عوامل تأثیرگذار در پردازش سیگنال گفتار (لهجه، جنسیت، سن و سطح بیماری) و تشکیل پایگاه داده عمومی‌تر برای کسب نتایج قابل اطمینان رعایت شده است.

این مقاله از ASR، روشی سودمند برای پردازش سیگنال گفتار بیماران آلزایمری، بهره گرفته است. انتخاب روش دقیق و درست برای استخراج بردار ویژگی از سیگنال گفتار تأثیر چشمگیری بر نتایج به دست آمده خواهد داشت؛

بنابراین در این مقاله سعی بر این است که از پردازش گفتار در علم پزشکی و تشخیص بیماری آلزایمر در مراحل اولیه استفاده شود. ویژگی‌هایی انتخاب و ارزیابی شده‌اند که بتوانند کمک شایان توجهی به این موضوع داشته باشند. ویژگی‌های مطرح شده در این مقاله، ویژگی‌های نوایی و فرکانسی ZC, SR, SC, BW و RMS و همچنین، ویژگی‌های حاصل از بررسی فرکانس مل و استخراج ضرایب کپسترال آن و ضرایب حاصل از آنروپی موجک‌اند که امروزه در پردازش سیگنال گفتار بسیار استفاده می‌شوند. پس از استخراج ویژگی، طبقه‌بندی با SVM انجام گرفته است و کرنل‌های مختلف این روش ارزیابی شده‌اند. مطابق جدول (۳)، بردار ویژگی از آنروپی بسته موجک، MFCC, ZC, SR, SC, BW و RMS تشکیل شد و با کرنل‌های مختلفی ارزیابی شدند. کرنل توابع شعاعی در هر ۶ بردار ویژگی نتایج خوبی نسبت به سایر کرنل‌ها نداشته است. بردار تشکیل شده از ویژگی‌های ZC, SR, SC, BW

بررسی شده است و در [۹] زبان‌های مختلف بررسی شده‌اند. در مراجع [۴۳، ۴۱، ۵۰]، ویژگی‌های آکوستیک و در [۴۴، ۴۵] ویژگی‌های فرکانسی مانند MFCC استخراج و ارزیابی شده‌اند. ویژگی آکوستیک، گفتاری و زبانی به صورت دستی استخراج می‌شوند که بسیار زمان‌برند و دقت بالایی ندارد. مرجع [۵] با وجود عملکرد خوب، پیچیدگی محاسباتی دارد. در مقاله حاضر ویژگی‌های متداول در پردازش سیگنال گفتار، همچون بسته موجک در کنار ضرایب کپسترال فرکانس مل، نرخ عبور از صفر، افت طیف، پهنای باند، انرژی سیگنال و فرکانس مرکز طیفی سیگنال گفتار بیماران آلزایمری و افراد سالم فارسی زبان به طور خودکار استخراج شده‌اند و در این طبقه بندی، دقت به دست آمده با ماشین بردار پشتیبان ۹۶٪ بوده است.

و RMS بهترین نتیجه را در طبقه بندی با صحت ۹۱٪ و با استفاده از کرنل چندجمله‌ای درجه ۴ داشته است. نتیجه ۹۱٪ بار دیگر در کرنل چندجمله‌ای درجه ۲ برای بردار ویژگی MFCC به همراه آنتروپی بسته موجک ظاهر شده است؛ اما اضافه کردن ویژگی‌های BW، SC، SR، ZC و RMS به بردار ویژگی مذکور نتیجه ۹۳٪ را داشته است که بهترین نتیجه حاصل از این طبقه بندی بوده است. جدول (۴)، خلاصه برخی مطالعات انجام شده در زمینه تحلیل سیگنال گفتار بیماران آلزایمری را نشان می‌دهد. همان‌طور که در جدول (۴) مشاهده می‌شود، در مراجع [۵۰، ۵۸]، ویژگی‌های زبانی و آماری سیگنال گفتار، پردازش و در [۴۲، ۴۱] ویژگی‌های گفتاری و زبانی تحلیل شده‌اند. در [۷] نیز گفتار از لحاظ احساس و عواطف

جدول (۴): خلاصه برخی مطالعات انجام شده در زمینه تحلیل سیگنال گفتار بیماران آلزایمری

مراجع	سال پژوهش	ضرایب ویژگی استفاده شده	زبان	روش تشخیص گفتار	دقت (بر حسب درصد)
[۹]	۲۰۱۵	فرکانس مرکز طیفی، انرژی، کوتاه مدت، حروف صدادار و بی صدا، بعد فرکانس	فرانسوی، انگلیسی، کاتالان، اسپانیایی، چینی، باسک، عربی، پرتغالی	k نزدیک ترین همسایه شبکه عصبی پروسپترون چند لایه	۹۰
[۶]	۲۰۱۶	فعل، اسم، حروف اضافه و ربط، ویژگی های آماری	اسپانیایی	SVM	۸۸
[۵]	۲۰۱۶	ویژگی های آکوستیک و استنتاج فازی و بهینه سازی	فارسی	میانگین مربع خطا و نمودارهای کانطور	۹۹/۹۶
[۴]	۲۰۱۵	نرخ کلمات پرسشی، تعداد فعل، اسم و حروف اضافه، فرکانس فعل، فرکانس اسم	انگلیسی	SVM، بیز ساده، درخت تصمیم گیری، k نزدیک ترین همسایه	۹۳/۷۹
[۷]	۲۰۱۳	ویژگی های احساس و عاطفی	انگلیسی	شبکه عصبی پروسپترون چند لایه درخت تصمیم گیری	۹۳/۰۲
[۸]	۲۰۱۵	میانگین، مدت زمان، انحراف استاندارد ویژگی های صدا، سکوت، متناوب بودن و نامتناوب بودن سیگنال گفتار	فرانسوی	k نزدیک ترین همسایه بیز ساده	۹۱/۴۷
[۴۲]	۲۰۱۶	ویژگی های گفتاری و زبانی	آلمانی	SVM	۸۷/۵۹
				F-score	۸۰

مراجع سال پژوهش	ضرایب ویژگی استفاده شده	زبان	روش تشخیص گفتار	دقت (بر حسب درصد)
[۴۰] ۲۰۱۷	استخراج خطاهای گفتاری، ویژگی‌های آماری	آلمانی	k نزدیک‌ترین همسایه تحلیل خطی SVM	۶۲/۲۳
[۴۳] ۲۰۱۸	ویژگی‌های آکوستیک	انگلیسی	مدل مخفی مارکوف مدل ترکیبی گوسی	۶۰/۶
[۴۵] ۲۰۱۸	MFCC و جیتر و شیمر محلی ویژگی‌های آکوستیک (نرخ گفتار،	انگلیسی	شبکه عصبی پیچشی	۷۳/۶
[۴۱] ۲۰۱۸	سکوت و بردار (i) و ویژگی‌های زبانی	آلمانی	-	۸۱/۹
[۴۴] ۲۰۱۹	MFCC	مجارستانی	شبکه عصبی	۷۳/۳
مقاله حاضر	ترکیب ویژگی‌های زمان، فرکانس و زمانی - فرکانسی	فارسی	ماشین بردار پشتیبان	۹۶

- مشاهده می‌شود نتیجه حاصل از این پژوهش در زبان فارسی با کارهای مشابه در زبان‌های دیگر مقایسه پذیر است؛ همچنین، استخراج ویژگی‌ها به صورت خودکار و مستقل از گوینده‌اند.
- گسترش این کار می‌تواند امکان ابتلا به این بیماری را در افراد پیش بینی کند و این مسئله کمک شایان توجهی در خصوص بهبود زندگی افراد خواهد داشت. برای دستیابی به اطلاعات دقیق تر می‌توان افرادی را بررسی کرد که در سطح پایین تری از این بیماری قرار دارند و بیماری آنها به مرحله حاد پیشرفت نکرده است. صحت عملکرد این ساختار را با علم پزشکی نیز می‌توان بررسی کرد. در عصر حاضر، پیش‌بینی‌های صورت گرفته در علم پزشکی بر اساس عوامل ژنتیکی، تحصیلات، حرفه افراد، منطقه آب‌وهوایی و سن افراد است که در صورت بررسی این موضوع به شکلی بسیار کامل تر این پیش‌بینی شکل دقیق تر و مفیدتری به خود خواهد گرفت.
- مراجع**
- [1] M .Nobakht, S. N. Hoseini, P. Mortazavvi, I. Sohrabi, B. Esmailzade, N. Rahbar Rooshandel, S. Omidzahir, "Neuropathological Changes in Brain Cortex and Hippocampus in a Rat Model of Alzheimer's Disease", Iranian Biomedical Journal, Vol. 15, PP. 51-58, 2011.
- [2] A. Khodabakhsh, F. Yesil, E. Guner, C. Demiroglu, "Evaluation of Linguistic and Prosodic Features for Detection of Alzheimer's Disease in Turkish Conversational Speech", EURASIP
- Journal on Audio, Speech, and Music Processing, Vol. 2015, No. 1, Mar. 2015.
- [3] M. El Ayadi, M. S. Kamel, F. Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases", Pattern Recognition, Vol. 44, No. 3, PP. 572-587, Mar. 2011.
- [4] S. J. Chaudhari, R. M. Kagalkar, "Automatic Speaker Age Estimation and Gender Dependent Emotion Recognition", International Journal of Computer Applications, Vol. 117, No. 17, PP. 5-10, May 2015.
- [5] M. Nasrolahzadeh, Z. Mohammadpoori, J. Haddadnia, "Analysis of Mean Square Error Surface and its Corresponding Contour Plots of Spontaneous Speech Signals in Alzheimer's Disease with Adaptive Wiener Filter", Computers in Human Behavior, Vol.61, PP. 364-371, Aug. 2016.
- [6] L. Hernandez-Dominguez, E. Garcia-Cano, S. Ratte, G. Sierra-Martinez, "Detection of Alzheimer's Disease Based on Automatic Analysis of Common Objects Descriptions", Proceedings of the 7th Workshop on Cognitive Aspects of Computational Language Learning, pp.10-15,11 Aug 2016.
- [7] K. Lopez-de-Ipiña, J. B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C. M. Travieso, M. Ecay-Torres, P. Martinez-Lage H. Eguiraun, "On Automatic Diagnosis of Alzheimer's Disease based on Spontaneous Speech Analysis and Emotional Temperature", Cognitive Computation, Vol. 7, PP. 44-55, 2015.
- [8] A. König, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, V. Manera, F. Verhey, P. Aalten, P. H. Robert, R. David, "Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease", Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, Vol. 1, No. 1, PP. 112-124, Mar. 2015.
- [9] K. López-de-Ipiña, J. Solé-Casals, H. Eguiraun, J.

- [21] L. Falek, A. Amrouche, L. Fergani, H. Teffahi, A. Djeradi, "Formantic Analysis of Speech Signal by Wavelet Transform", Proceedings of the World Congress on Engineering. London, Vol. 2, PP.1572-1576, 2011.
- [22] N. Erfanian Saeedi, F. Almasganj, "Wavelet Adaptation for Automatic Voice Disorders Sorting", Computers in Biology and Medicine, Vol. 43, No. 6, PP. 699-704, Jul. 2013.
- [23] R. Behroozmand, F. Almasganj, "Optimal Selection of Wavelet-Packet-Based Features using Genetic Algorithm in Pathological Assessment of Patients' Speech Signal with Unilateral Vocal Fold Paralysis", Computers in Biology and Medicine, Vol. 37, No. 4, PP. 474-485, Apr. 2007.
- [24] H. Khadivi Heris, B. S. Aghazadeh, M. Nikkhah-Bahrami, "Optimal Feature Selection for the Assessment of Vocal Fold Disorders", Computers in Biology and Medicine, Vol. 39, No. 10, PP. 860-868, Oct. 2009.
- [25] C. D. P. Crovato, A. Schuck, "The Use of Wavelet Packet Transform and Artificial Neural Networks in Analysis and Classification of Dysphonic Voices", IEEE Transactions on Biomedical Engineering, Vol. 54, No. 10, PP.1898-1900, Oct. 2007.
- [26] N. Erfanian Saeedi, F. Almasganj, F. Torabinejad, "Support Vector Wavelet Adaptation for Pathological Voice Assessment", Computers in Biology and Medicine, Vol. 41, No. 9, PP. 822-828, Sep. 2011.
- [27] S. Poomjan, T. Taengtang, K. Srinuanjan, S. Kamoldilok, C. Ruttanapun, P. Buranasiri, "Proof of using Fourier Coefficients for Root Mean Square Calculations on Periodic Signals", Advanced Studies in Theoretical Physics, Vol. 8, pp. 21-25, 2014.
- [28] P. N. Le, E. Ambikairajah, J. Epps, V. Sethu, E. H. C. Choi, "Investigation of Spectral Centroid Features for Cognitive Load Classification", Speech Communication, Vol. 53, No. 4, PP. 540-551, Apr. 2011.
- [29] J. Vesanto, Esa Alhoniemi. "Clustering of the Self-Organizing Map", Reprinted from IEEE Transactions on Neural Networks, Vol. 11, No. 3, PP. 586-600, 2000.
- [30] S. Akbari, M. H. Ghassemian, "Heart Sound Classification using MFCC and SVM", Iranian Association of Information and Communication Technology, May 20015.
- [31] Y. Li, L. Zhang, B. Li, Y. Xu, S. Wu, X. Wei, X. Liu, R. Lin, Q. Wang, "The Simulation Study of Three Typical Time Frequency Analysis Methods", BIO Web of Conferences, Vol. 8, PP. 02007, 2017.
- [32] L. Salhi, M. Talbi, A. Cherif, "Voice Disorders Identification Using Hybrid Approach: Wavelet Analysis and Multilayer Neural Networks", World Academy of science, Engineering and Technology, Vol. 45, PP.330-339, 2008.
- [33] S. Rekik, D. Guerchi, H. Hamam, S. Selouani, "Audio Steganography Coding Using the Discrete Wavelet Transforms", International Journal of Computer Science and Security (IJCSS) ,Vol. 6, B. Alonso, C. M. Travieso, A. Ezeiza, N. Barroso, M. Ecay-Torres, P. Martinez-Lage, B. Beitiaf, "Feature Selection for Spontaneous Speech Analysis to Aid in Alzheimer's Disease Diagnosis: A Fractal Dimension Approach", Computer Speech and Language, Vol. 30, No.1 ,PP. 43-60, 2015.
- [10] K. C. Fraser, J. A. Meltzer, F. Rudzicz, "Linguistic Features Identify Alzheimer's Disease in Narrative Speech", Journal of Alzheimer's Disease, Vol. 49, No. 2, PP. 407-422, Oct. 2016.
- [11] M. Kalamani, S. Valarmathy, C. Poonkuzhali, R. Karthiprakash, "Comparison Of Cepstral And Mel Frequency Cepstral Coefficients For Various Clean And Noisy Speech Signals", ECE Department, Bannari Amman Institute of Technology, Vol. 2, No. 1, PP.2897-2904, 2014.
- [12] D. Gharavian, M. Sheikhan, F. Ashoftehdel, "Using Neutralized Formant Frequencies to Improve Emotional Speech Recognition", IEICE Electronics Express, Vol. 8, No. 14, PP. 1155-1160, 2011.
- [13] A. Potamianos, P. Maragos, "Speech Formant Frequency and Bandwidth Tracking using Multiband Energy Demodulation", The Journal of the Acoustical Society of America, Vol. 99, No. 6, PP. 3795-3806, Jun. 1996.
- [14] B. J. Mohan, R. Babu, "Speech Recognition using MFCC and DTW", 2014 International Conference on Advances in Electrical Engineering (ICAEE), PP.1-4, Jan. 2014.
- [15] S. Ahmed, A. M. F. Haigh, C. A. de Jager, P. Garrard, "Connected Speech as a Marker of Disease Progression in Autopsy-Proven Alzheimer's Disease", Brain, Vol. 136, No. 12, PP.3727-3737, Oct. 2013.
- [16] J. J. G. Meilán, F. Martínez-Sánchez, J. Carro, J. A. Sánchez, E. Pérez, "Acoustic Markers Associated with Impairment in Language Processing in Alzheimer's Disease", The Spanish journal of psychology, Vol. 15, No. 02, PP.487-94, Jul. 2012.
- [17] K. López-de-Ipiña, J. B. Alonso, N. Barroso, M. Faundez-Zanuy, M. Ecay, J. Solé-Casals, C. M. Travieso, A. Estanga, A. Ezeiza, "New Approaches for Alzheimer's Disease Diagnosis Based on Automatic Spontaneous Speech Analysis and Emotional Temperature", Ambient Assisted Living and Home Care, pp. 407-414, 2012.
- [18] M. Gordon, "Linguistic Aspects of Voice Quality with Special Reference to Athabaskan", In Siri Tuttle & Gary Holton (eds.) Proceedings of the 2001 Athabaskan Languages Conference. Fairbanks: Alaska Native Language Center, PP. 163-178, 2001.
- [19] R. G. Bachu, S. Kopparthi, B. Adapa, B. D. Barkana, "Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal", Department of Electrical Engineering School of Engineering, University of Bridgeport, Bridgeport, USA, PP.279-282, March. 2008.
- [20] T. F. Quatieri, "Discrete-Time Speech Signal Processing: Principles and Practice", 1<sup>st</sup> Ed. Kindle Edition, 2001.

- Automatic Transcriptions in Dementia Detection from Speech", Interspeech 2017, PP. 3117-3121, Aug. 2017.
- [41] J. Weiner, T. Schultz, "Selecting Features for Automatic Screening for Dementia Based on Speech", Lecture Notes in Computer Science, PP. 747-756, 2018.
- [42] J. Weiner, C. Herff, T. Schultz, "Speech-Based Detection of Alzheimer's Disease in Conversational German", Interspeech 2016, PP. 1938-1942, Sep. 2016.
- [43] B. Mirheidari, D. Blackburn, T. Walker, A. Venneri, M. Reuber, H. Christensen, "Detecting Signs of Dementia Using Word Vector Representations", Interspeech 2018, PP. 1893-1897, Sep. 2018.
- [45] T. Warnita, N. Inoue, K. Shinoda, "Detecting Alzheimer's Disease using Gated Convolutional Neural Network from Audio Data", Interspeech 2018, PP. 1706-1710, Sep. 2018.
- [44] G. Gosztolya, V. Vincze, L. Tóth, M. Pákási, J. Kálmán, I. Hoffmann, "Identifying Mild Cognitive Impairment and Mild Alzheimer's Disease Based on Spontaneous Speech using ASR And Linguistic Features", Computer Speech & Language, Vol. 53, pp. 181-197, Jan. 2019.
- No. 1, PP. 79-93, 2012.
- [34] S. Vajapeyam, "Understanding Shannon's Entropy metric for Information", Computer Science, Information Theory, Vol. 1, No. 2014, March. 2014.
- [35] N. Almaadeed, A. Aggoun, A. Amira, "Speaker Identification using Multimodal Neural Networks and Wavelet Analysis", IET Biometrics, Vol. 4, No. 1, PP. 18-28, Mar. 2015.
- [36] T. Matsui, S. Furui, "A Text-Independent Speaker Recognition Method Robust Against Utterance Variations", [Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing, 1991.
- [37] A. Shukla, "Speaker Identification using Wavelet Analysis and Modular Neural Network", Journal of Acoustic Society of India, Vol. 36, No. 1, PP.14-19, 2009.
- [38] X. Pan, H. Zhao, Y. Zhou, "The Application of Fractional Mel Cepstral Coefficient in Deceptive Speech Detection", Peer J., Vol. 3, No.18, PP. e1194, Aug. 2015.
- [39] K. Forbes-McKay, M. F. Shanks, A. Venneri, "Profiling Spontaneous Speech Decline in Alzheimer's Disease: A Longitudinal Study", Acta Neuropsychiatrica, Vol. 25, No. 06, PP.320-327, May 2013.
- [40] J. Weiner, M. Engelbart, T. Schultz, "Manual and

---

<sup>1</sup> Automatic Speech Recognition

<sup>2</sup> Mel Frequency Cepstral Coefficients

<sup>3</sup> Prosodic

<sup>4</sup> Support Vector Machine

<sup>5</sup> Radial Basis Function Kernel

<sup>6</sup> Spectral Centroid

<sup>7</sup> Zero Crossing

<sup>8</sup> Root Mean Square

<sup>9</sup> Band Width

<sup>10</sup> Spectral Roll Off

<sup>11</sup> Gabor

<sup>12</sup> Generative

<sup>13</sup> Discriminative

<sup>14</sup> Daubechies

<sup>15</sup> Morlet

<sup>16</sup> Meyer

<sup>17</sup> Haar

<sup>18</sup> Confidence Interval

