



## Persian Words Recognition based on Facial Electromyogram Signals

Sharifi, Pooria<sup>1</sup> / Soltanizadeh, Hadi<sup>2\*</sup> / Maleki, Ali<sup>3</sup>

<sup>1</sup> - Ph.D. Student, Electrical and Computer Engineering Department, Semnan University, Semnan, Iran

<sup>2</sup> - Associate Professor, Electrical and Computer Engineering Department, Semnan University, Semnan, Iran

<sup>3</sup> - Associate Professor, Biomedical Engineering Department, Semnan University, Semnan, Iran

### ARTICLE INFO

DOI: 10.22041/ijbme.2023.560736.1796

Received: 22 August 2022

Revised: 26 January 2023

Accepted: 7 February 2023

### KEYWORDS

*Facial Electromyogram*

*Persian Words*

*Silent Speech Interface*

*Word Recognition*

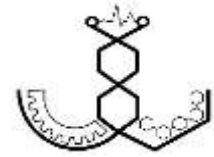
### ABSTRACT

Losing of voice and larynx is a major problem for people with speech disorders. It creates serious and negative consequences on the quality of individual and group life of these people, especially in working environments. The development of an intelligent system based on electromyogram signals with the ability to recognize speech (without using sound) can be a window of hope for people who lost their larynx and voice due to cancer. Although progress and studies in this field are growing in our country and in different languages, but these studies have not been done for the Persian language. In this article, for the first time, recognition of Persian words was done using electromyogram of facial muscles. For this purpose, sEMG signals were collected from eight facial muscles and six volunteers while speaking twelve Persian words. Then, MFL, VAR, DAMV, LTKE, IQR and Cardinality features were extracted from each channel and each window from the signal, and the 432 features from each signal were reduced to 33 features using the PCA principal component analysis method. Finally, in order to recognize twelve Persian words, the features were given to SVM, KNN and RF classifiers. The average classification accuracy was 83.16%, 81.91% and 78.97%, respectively. Our evaluation in this article gives the hope that by using EMG signals it is possible to recognize the limited words of Persian language.

### \*Corresponding Author

<b>Address</b>	Electrical and Computer Engineering Department, Semnan University, Semnan, Iran
<b>Postal Code</b>	35131-19111
<b>E-Mail</b>	<i>h_soltanizadeh@semnan.ac.ir</i>
<b>Tel</b>	+98-23-31532697
<b>Fax</b>	+98-23-31532697





## تشخیص کلمات فارسی بر اساس سیگنال‌های الکترومایوگرام چهره

شریفی، پوریا<sup>۱</sup> / سلطانی‌زاده، هادی<sup>۲\*</sup> / مالکی، علی<sup>۳</sup>

- <sup>۱</sup> - دانشجوی دکتری، گروه مهندسی برق، دانشکده‌ی مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران  
<sup>۲</sup> - دانشیار، گروه مهندسی برق، دانشکده‌ی مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران  
<sup>۳</sup> - دانشیار، گروه مهندسی پزشکی، دانشگاه سمنان، سمنان، ایران

### مشخصات مقاله

شناسه‌ی دیجیتال: 10.22041/ijbme.2023.560736.1796

پذیرش: ۱۸ بهمن ۱۴۰۱

بازنگری: ۶ بهمن ۱۴۰۱

ثبت در سامانه: ۳۱ مرداد ۱۴۰۱

### چکیده

از دست دادن صدا و حنجره یک معضل بزرگ برای افراد با اختلالات گفتاری است. این اتفاق عواقب جدی و منفی بر کیفیت زندگی فردی و گروهی این اشخاص به ویژه در محیط‌های کاری ایجاد می‌کند. توسعه‌ی یک سیستم هوشمند بر پایه‌ی سیگنال‌های الکترومایوگرام با قابلیت تشخیص گفتار (بدون استفاده از صوت) می‌تواند دریچه‌ی امیدبخشی برای افرادی که حنجره و صدای خود را بر اثر سرطان از دست داده‌اند باشد. اگر چه در این حوزه پژوهش‌هایی برای زبان‌های مختلف انجام شده اما برای زبان فارسی پژوهشی صورت نگرفته است. در این مقاله برای اولین بار، بازنشاسی واژگان فارسی با استفاده از الکترومایوگرام عضلات چهره انجام پذیرفته است. بدین منظور سیگنال‌های sEMG از ۸ عضله‌ی چهره‌ی ۶ داوطلب هنگام بیان ۱۲ کلمه‌ی زبان فارسی جمع‌آوری شده است. سپس ویژگی‌های MFL، VAR، DAMV، LTKE، IQR و Cardinality از هر کانال و هر پنجره از سیگنال استخراج گردیده و ۴۳۲ ویژگی حاصل از هر سیگنال با استفاده از روش تحلیل مولفه‌ی اصلی به ۴۹ ویژگی تقلیل یافته است. در نهایت به منظور بازنشاسی ۱۲ کلمه‌ی زبان فارسی، ویژگی‌ها به طبقه‌بندهای SVM، KNN و RF داده شده است. میانگین صحت طبقه‌بندی به ترتیب ۸۳/۱۶٪، ۸۱/۹۱٪ و ۷۸/۹۷٪ به دست آمده است. ارزیابی نتایج این مقاله گویای آن است که با استفاده از سیگنال‌های EMG می‌توان کلمات محدود زبان فارسی را با صحت خوبی بازنشاسی نمود.

### واژه‌های کلیدی

الکترومایوگرام چهره  
کلمات فارسی  
رابط گفتاری بی‌صدا  
بازنشاسی کلمات

### \*نویسنده‌ی مسئول

نشانی گروه مهندسی برق، دانشکده‌ی مهندسی برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران

کد پستی ۳۵۱۳۱-۱۹۱۱۱

پست الکترونیک h\_soltanizadeh@semnan.ac.ir

تلفن +۹۸-۲۳-۳۱۵۳۲۶۹۷

دورنگار +۹۸-۲۳-۳۱۵۳۲۶۹۷



## ۱- مقدمه

گفتار طبیعی‌ترین وسیله‌ی ارتباطی بین انسان‌ها است و از آن به طور فزاینده‌ای در دستگاه‌های امروزی برای تعامل استفاده می‌شود. بنابراین قابلیت گفتار برای هر انسان و برای هر جامعه از اهمیت فوق‌العاده‌ای برخوردار است. با این حال گفتار به طور معمول باید واضح، شفاف و قابل شنیدن باشد، به خصوص اگر قرار است از طریق دستگاه‌هایی مانند تلفن همراه یا یک تشخیص دهنده‌ی گفتار مثل دست‌یارهای صوتی پردازش شود. در شرایطی ممکن است استفاده از گفتار برای برقراری ارتباط دچار اختلال شود که در ادامه به سه مورد از آن اشاره می‌شود. ۱- شرایطی که فرد در یک محیط شلوغ مثل کتابخانه، رستوران، مترو، اتوبوس یا جلسه‌ی عمومی می‌خواهد ارتباط گفتاری برقرار کند. ۲- شرایطی که نیاز به برقراری ارتباطی محرمانه مثل وارد کردن رمز صوتی برای یک سیستم است. ۳- شرایطی که افراد دچار معلولیت گفتاری مثل افراد مبتلا به سرطان حنجره قصد برقراری ارتباط صوتی با افراد دیگر یا دست‌یارهای صوتی را دارند. بنابراین اهمیت توسعه‌ی روش‌هایی جای‌گزین برای شرایط بروز اختلال گفتاری به طور فزاینده‌ای افزایش یافته است.

در سال‌های اخیر یک رویکرد جدید امیدوار کننده به نام رابط‌های جای‌گزین و تقویتی<sup>۱</sup> (AAC) ظهور کرده که وسیله‌ای کمکی برای بازگرداندن ارتباط زبانی با رمزگشایی گفتار از سیگنال‌های بیولوژیک غیرصوتی تولید شده در هنگام تولید گفتار است. یکی از روش‌های شناخته شده‌ی ارتباط گفتاری بی‌صدا، لب‌خوانی است. انواع زیادی از روش‌های سنجش برای گرفتن سیگنال‌های بیولوژیک مرتبط با گفتار وجود دارد که از آن جمله می‌توان به تصویربرداری از مجرای صوتی<sup>۲</sup> [۱، ۲]، ردیابی مغناطیسی حرکات مفصل گفتار<sup>۳</sup> (EMA) [۳، ۴، ۵]، الکترومایوگرافی سطحی<sup>۴</sup> (sEMG) [۶، ۷، ۸] (که فعالیت‌های الکتریکی محرک عضلات صورت را با استفاده از الکترودهای سطحی ثبت می‌کند) و الکتروانسفالوگرافی<sup>۵</sup> (EEG) [۹] (که فعالیت عصبی را در مناطق تشریحی مغز درگیر در تولید گفتار ثبت می‌کند) اشاره کرد.

رابط گفتار بی‌صدا<sup>۱</sup> وسیله‌ای است که امکان برقراری ارتباط گفتاری را بدون استفاده از صدای تولید شده توسط شخص فراهم می‌کند. به این ترتیب که یک نوع لب‌خوانی الکترونیکی

انجام می‌دهد. این کار با استفاده از امواج فراصوت، دوربین‌ها و یا ابزارهای الکترومغناطیسی به منظور ردیابی حرکات زبان و لب‌ها انجام می‌شود. از آن‌جا که رابط‌های گفتاری بی‌صدا ارتباطات گفتاری را بدون اتکا به سیگنال صوتی امکان‌پذیر می‌کنند، اساساً ابزار جدیدی جهت بازیابی قابلیت‌های ارتباطی برای افراد دارای اختلال گفتاری محسوب می‌شوند. هم‌چنین سیگنال‌های بیولوژیک نسبت به سروصدای نویزی محیط حساس نبوده و از سیگنال گفتاری صوتی مستقل هستند. به عنوان مثال این سیگنال‌های بیولوژیک حتی در صورت عدم صداسازی نیز قابل ضبط هستند [۱۰، ۱۱].

راتنوسکی و هم‌کارانش در سال ۲۰۲۱ [۱۲] به بازشناسی ۵ کلمه‌ی زبان انگلیسی پرداخته‌اند. در این مقاله سیگنال‌های الکترومایوگرام با استفاده از ۷ الکتروود و از ۳ عضله‌ی صورت ثبت شده و ویژگی‌های تبدیل فوری‌ی زمان کوتاه از سیگنال‌ها استخراج شده است. از روش‌های تحلیل مولفه‌های اصلی<sup>۶</sup> (PCA) و تعبیه‌ی خطی محلی<sup>۷</sup> (LLE) به منظور کاهش ابعاد ویژگی‌ها استفاده شده و این روش‌ها با هم مقایسه شده است. با استفاده از الگوریتم طبقه‌بند k نزدیک‌ترین همسایه، صحت تشخیص ۷۴/۶٪ برای مجموعه‌ی ۵ کلمه حاصل شده است.

وانگا و هم‌کارانش در سال ۲۰۲۱ [۱۳] به بازشناسی ۹ کلمه‌ی چینی با استفاده از عضلات چهره پرداخته‌اند. در این مقاله تمام داده‌های EMG با استفاده از ۱۰ الکتروود و از ۶ ماهیچه‌ی صورت با فرکانس نمونه‌برداری ۱۰۰۰ هرتز از ۴ داوطلب جوان سالم در حالی که کلمات kuai, you, zuo, hou, qian, er, yi, man و ting را تلفظ می‌کردند ثبت شده است. به منظور استخراج ویژگی خودکار، تصاویر طیف هر سیگنال به مدل یادگیری عمیق از پیش آموزش داده شده‌ی MobileNet-V1، ResNet18 و Xception داده شده است. در نهایت ویژگی‌های استخراج شده به طبقه‌بندهای حافظه‌ی کوتاه‌مدت دوطرفه<sup>۸</sup> (Bi-LSTM) و واحد بازگشتی گیت<sup>۹</sup> (GRU) داده شده و صحت ۸۷ درصد حاصل شده است.

چاندراشک در سال ۲۰۲۱ [۱۴] به تشخیص و بازشناسی ۶ حرف انگلیسی A, B, E, I, O و U مبتنی بر سیگنال‌های الکترومایوگرام پرداخته است. در این مقاله سیگنال‌های الکترومایوگرام با استفاده از ۳ الکتروود قرار گرفته در زیر چانه با فرکانس نمونه‌برداری ۲۰۰۰ هرتز از یک داوطلب سالم ثبت

<sup>۶</sup> Silent Speech Interface<sup>۷</sup> Principal Component Analysis<sup>۸</sup> Locally Linear Embedding<sup>۹</sup> Bidirectional Long-Short Term Memory<sup>۱۰</sup> Gated Recurrent Unit<sup>۱</sup> Augmentative and Alternative Communication<sup>۲</sup> Vocal Tract<sup>۳</sup> Electromagnetic Articulography<sup>۴</sup> Surface Electromyography<sup>۵</sup> Electroencephalogram



صفر تا ۱۰ انجام دادند. در این مطالعه در ابتدا سیگنال‌های الکترومایوگرام از ۵ موقعیت ماهیچه‌های صورت و گردن ثبت شده است در حالی که ۱۰ شرکت کننده کلمات را در دو حالت با صدا و بی‌صدا بیان کردند. تعداد ۲۲ ویژگی از سیگنال‌های الکترومایوگرام استخراج شده و ویژگی‌های مناسب جهت طبقه‌بندی انتخاب شده است. نتایج نشان می‌دهد که طبقه‌بند خطی حداقل مربعات فیشر<sup>۲</sup> (FLDA) و طبقه‌بند نرمال بیز<sup>۳</sup> (LBN) بهترین نتیجه را داشته‌اند. میانگین صحت به ترتیب برای حالت صدادار و بی‌صدا ۹۰/۰۱٪ و ۷۹/۱۸٪ بوده است.

جونگ و هم‌کارانش در سال ۲۰۱۹ [۱۸] یک سیستم تشخیص گفتار مبتنی بر سیگنال‌های الکترومایوگرام را با استفاده از ۵ کانال به منظور طبقه‌بندی ۹ هجای تایلندی توسعه داده‌اند. در این مقاله ۶ ویژگی از ۳ دسته ویژگی، یعنی میانگین قدر مطلق<sup>۴</sup> (MAV) و طول شکل موج<sup>۵</sup> (WL) از ویژگی‌های مبتنی بر دامنه، نرخ عبور از صفر<sup>۶</sup> (ZC) و میانگین فرکانس<sup>۷</sup> (MNF) از ویژگی‌های مبتنی بر فرکانس، و کشیدگی<sup>۸</sup> و چولگی<sup>۹</sup> از ویژگی‌های مبتنی بر آمار استخراج شده است. سپس از یک ماشین یادگیری افراطی رگرسیون طیفی<sup>۱۰</sup> (SRELM) برای کاهش ابعاد ویژگی‌ها از ۳۰ به ۸ استفاده شده است. در نهایت ویژگی‌های کاهش یافته با استفاده از طبقه‌بند شبکه‌ی عصبی پیش‌خور با اعتبارسنجی ۵-fold طبقه‌بندی شده است. میانگین صحت طبقه‌بندی به دست آمده در داوطلبان سالم و دیزارتیک به ترتیب ۹۴/۵٪ و ۸۹/۴٪ بوده است.

مینگزینگ و هم‌کارانش در سال ۲۰۲۱ [۱۹] به بررسی مشارکت عضلات مختلف به منظور تشخیص ۱۰ کلمه‌ی انگلیسی و چینی با استفاده از سیگنال‌های EMG پرداخته‌اند. در این پژوهش از ۱۲۰ الکتروود با فاصله‌ی نزدیک روی عضلات چهره و گردن از ۱۸ داوطلب بومی چینی استفاده شده است. الگوریتم مورد استفاده جهت طبقه‌بندی تحلیل تفکیک کننده‌ی خطی بوده و از الگوریتم انتخاب رو به جلو متوالی برای تجزیه و تحلیل موقعیت بهینه‌ی حسگرها استفاده شده است. نتایج نشان می‌دهد که توپوگرافی انرژی الکترومایوگرام می‌تواند به تجسم فعالیت‌های ماهیچه‌ای پویا در طول فرایند صحبت کردن کمک کند و الگوهای انقباض ماهیچه‌ای به طور قابل توجهی برای کارهای مختلف صحبت کردن مشاهده شود. استفاده از تعداد الکتروود برابر در دو ناحیه‌ی گردن و صورت

شده است. به منظور استخراج ویژگی، تصاویر مربوط به طیف سیگنال‌ها به یک شبکه‌ی عمیق گوگل‌نت از پیش آموزش دیده، داده شده است. در نهایت سیگنال‌ها با استفاده از ۸ طبقه‌بند متفاوت مورد ارزیابی قرار گرفته است. تعداد ۸۰٪ داده‌ها برای آموزش و ۲۰٪ برای آزمون در نظر گرفته شده است. نتایج نهایی نشان می‌دهد که بهترین نرخ تشخیص مربوط به طبقه‌بند ماشین بردار پشتیبان با صحت ۸۰/۱۰٪ است.

یانگ و هم‌کارانش در سال ۲۰۲۰ [۱۵] به بازشناسی ۱۰ کلمه‌ی چینی مبتنی بر توابع درون‌یابی پرداختند. در این مقاله داده‌ها با استفاده از ۵ کانال و الکتروود، از ماهیچه‌ی صورت ۲ داوطلب آقا در حالی که کلمات shi, ren, men, le, de, bu را تلفظ می‌کردند ثبت شده است. سیگنال‌ها با دو مدل مورد ارزیابی قرار گرفته است. در مدل اول سیگنال‌ها پنجره‌گذاری شده و از هر پنجره ویژگی‌ها استخراج شده و در مدل دوم ویژگی‌ها از کل سیگنال و بدون پنجره‌گذاری استخراج شده است. ویژگی‌ها از دو حوزه‌ی زمان و زمان-فرکانس استخراج شده است. پس از کاهش ابعاد بردار ویژگی‌ها، سیگنال‌ها با استفاده از سه الگوریتم k نزدیک‌ترین همسایه، جنگل تصادفی و مدل مخفی مارکوف طبقه‌بندی شده که بهترین نتیجه برای طبقه‌بند مارکوف با صحت ۸۵٪ بوده است. هانگ و هم‌کارانش در سال ۲۰۲۰ [۱۶] با استفاده از شبکه‌های عصبی حافظه‌دار دوطرفه به بازشناسی ۴ کلمه‌ی چینی به منظور کنترل ربات پرداختند. در این مقاله سیگنال‌ها با استفاده از ۱۰ الکتروود با فرکانس نمونه‌برداری ۱۰۰۰ هرتز از ۷ داوطلب در حالی که کلمات man و kuai, yi, er را تلفظ می‌کردند ثبت شده است. استخراج ویژگی با استفاده از تبدیل فوریه‌ی زمان کوتاه و شبکه‌ی CNN انجام گرفته است. سپس ویژگی‌های استخراج شده به دو بخش آموزش (۷۰٪) و آزمایش (۳۰٪) تقسیم شده و جهت تشخیص کلمه‌ی مورد نظر به طبقه‌بند BLSTM داده شده است. به منظور ارزیابی روش پیشنهادی، نتایج با الگوریتم‌های LSTM، CNN و جنگل تصادفی<sup>۱</sup> مقایسه شده است. نتایج نشان می‌دهد که بهترین نرخ بازشناسی مربوط به طبقه‌بند BLSTM با صحت ۹۷/۱۱٪ است.

سریسووان و هم‌کارانش در سال ۲۰۲۰ [۱۷] یک مقایسه از عمل کرد ۸ طبقه‌بند برای تشخیص گفتار مبتنی بر سیگنال‌های الکترومایوگرام برای بازشناسی ۱۱ کلمه‌ی تایلندی شامل اعداد

<sup>۱</sup> Zero Crossing

<sup>۲</sup> Mean Frequency

<sup>۳</sup> Kurtosis

<sup>۴</sup> Skewness

<sup>۱۰</sup> Spectral Regression Extreme Learning Machine

<sup>۱</sup> Random Forest

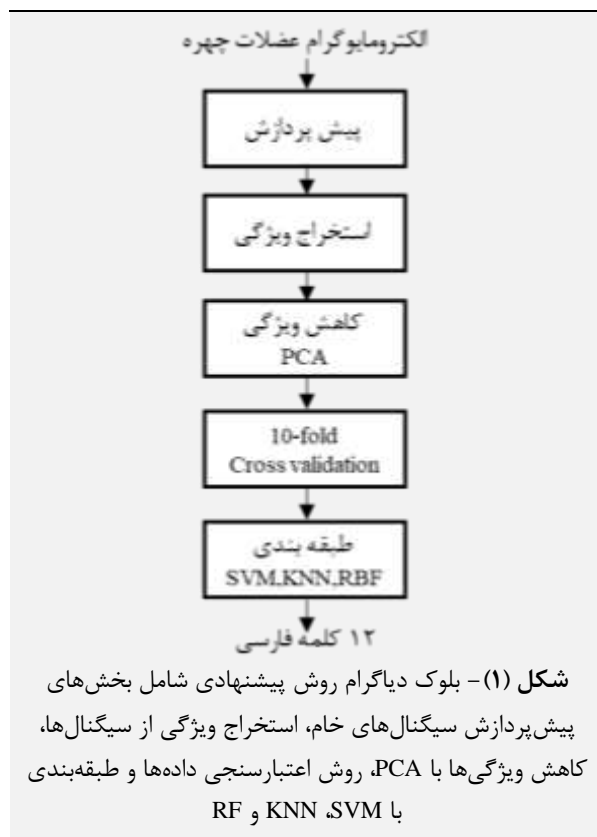
<sup>۲</sup> Fishers Linear Discriminant Analysis

<sup>۳</sup> Linear Bayes Normal

<sup>۴</sup> Mean Absolute Value

<sup>۵</sup> Waveform Length

و سپس با استفاده از الگوریتم تحلیل مولفه‌ی اصلی، ویژگی‌ها تقلیل یافته است. در نهایت با روش اعتبارسنجی ۱۰-fold و ماشین بردار پشتیبان<sup>۱</sup>، نزدیک‌ترین همسایه‌ی  $k^2$  و جنگل تصادفی، ۱۲ کلمه‌ی فارسی طبقه‌بندی شده است.



## ۲-۱- ثبت سیگنال الکترومایوگرام سطحی از عضلات صورت

با توجه به این که هیچ پایگاه داده‌ی سیگنال EMG حاصل از عضلات صورت شخص در حین گفتار کلمات فارسی وجود ندارد در این مقاله ثبت سیگنال و ایجاد پایگاه داده از عضلات صورت داوطلب‌ها در حین گفتار انجام شده است. سیگنال‌های ثبت و جمع‌آوری شده در این مقاله از ۳۰ داوطلب شامل ۱۵ مرد در بازه‌ی سنی ۲۰-۴۵ سال و ۱۵ خانم با بازه‌ی سنی ۲۰-۳۴ سال به دست آمده است. شرکت کنندگان هیچ‌گونه سابقه‌ی بیماری یا اختلال اسکلتی عضلانی مزمن نداشته و قبل از ثبت سیگنال رضایت خود را به صورت کتبی برای شرکت در این تحقیق از طریق پرسش‌نامه‌ی مورد تایید کمیته‌ی اخلاق اعلام نموده‌اند. این مطالعه توسط کمیته‌ی اخلاق دانشگاه علوم پزشکی سمنان با شناسه‌ی IR.SEMUMS.REC.1401.103 بررسی و تایید شده است. داوطلب‌ها روی صندلی نشسته و موقعیت خود را ثابت کرده تا از اثرات نامطلوب حرکت‌های ناخواسته بر

نشان می‌دهد که نتایج مربوط به گردن خیلی بهتر از صورت است. علاوه بر این تنها با انتخاب ۱۵ الکتروود بهینه که عمدتاً به جای صورت روی گردن توزیع شده است نرخ طبقه‌بندی می‌تواند به بالاتر از ۹۰٪ برسد.

سیونگ چا و هم‌کارانش در سال ۲۰۲۲ [۲۰] یک سیستم تشخیص گفتار بی‌صدا به منظور بازشناسی ۶ کلمه‌ی کره‌ای بر اساس EMG ثبت شده در اطراف چشم را برای ۲۰ داوطلب پیاده‌سازی کرده‌اند. در این مقاله ۶ الکتروود EMG روی یک فریم پلاستیکی با شکلی مشابه با پد به یک هدست VR تجاری متصل شده است. ویژگی ریشه‌ی میانگین مربعات (RMS) با استفاده از یک پنجره‌ی زمانی لغزشی ۰/۱ ثانیه بدون هم‌پوشانی از سیگنال‌ها استخراج شده است. با استفاده از مدل یادگیری عمیق بهینه شده (me-bLSTM) ۶ کلمه‌ی کره‌ای بی‌صدا با معادل انگلیسی Menu, Escape, Previous, Next, Home و Select با صحت بالای ۹۲/۵۳٪ طبقه‌بندی شده است. روش مقاله‌ی حاضر تشخیص گفتار بی‌صدا توسط EMG است، جایی که پتانسیل‌های الکتریکی عضلات چهره‌ی کاربر توسط الکتروودهای سطحی ثبت می‌شود. این امر امکان پردازش گفتار را حتی در صورت عدم تولید سیگنال صوتی فراهم می‌کند که می‌تواند در مکان‌های شلوغ با صداها‌ی نویزی موثر باشد [۲۰]. توسعه‌ی یک سیستم ارتباط گفتاری از طریق سیگنال‌های الکترومایوگرام مبتنی بر زبان فارسی می‌تواند در زمینه‌های زیادی مفید باشد. در این مقاله برای اولین بار به بازشناسی گفتار زبان فارسی مبتنی بر الکترومایوگرام چهره پرداخته شده است. هم‌چنین در کارهای انجام شده در این حوزه، این مطالعه دارای بیش‌ترین تعداد داوطلب با توزیع جنسیتی برابر است. در ادامه در بخش ۲ توضیح در مورد نحوه‌ی ثبت سیگنال، ایجاد پایگاه داده، استخراج ویژگی، کاهش ویژگی‌ها و طبقه‌بندی ارائه شده است. در بخش ۳ نتایج حاصل از طبقه‌بندی کلمات زبان فارسی و مقایسه با سایر پژوهش‌ها برای زبان‌های دیگر بیان شده است. در نهایت یافته‌ها و بحث و نتیجه‌گیری در بخش‌های ۴ و ۵ مطرح شده است.

## ۲- مواد و روش‌ها

شمای کلی از روش پیشنهادی در شکل (۱) نشان داده شده است. در ابتدا سیگنال‌های عضلات چهره‌ی هر داوطلب ثبت شده و سپس پیش‌پردازش‌های لازم جهت حذف نویز و آرتیفکت‌های حرکتی روی تمام سیگنال‌ها اعمال شده است. در مرحله‌ی بعد از هر سیگنال ویژگی‌های مناسب استخراج شده

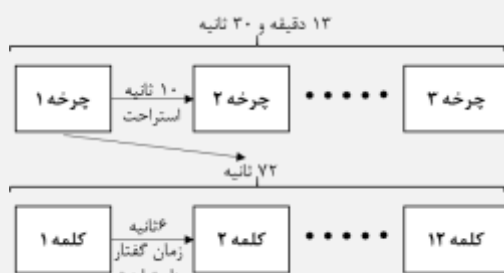
<sup>۲</sup> K Nearest Neighbor

<sup>۱</sup> Support Vector Machine

برای یک داوطلب در شکل (۳) نشان داده شده است. دوازده کلمه‌ی بیان شده در این آزمون شامل ماه‌های سال به زبان و تقویم فارسی (فروردین، اردیبهشت، خرداد، تیر، مرداد، شهریور، مهر، آبان، آذر، دی، بهمن و اسفند) است. در جدول (۱) کلمات به همراه تلفظ آن‌ها ارائه شده است.



شکل (۲) - (الف) مکان قرارگیری ۸ الکتروود، (ب) دستگاه EMG مورد استفاده جهت ثبت سیگنال



شکل (۳) - الگواری ثبت سیگنال

جدول (۱) - معرفی و تلفظ ۱۲ کلمه‌ی فارسی

شماره‌ی کلاس	کلمه‌ی فارسی	تلفظ
۱	فروردین	/farvardin/
۲	اردیبهشت	'ordibehest/
۳	خرداد	/xordād/
۴	تیر	/tir/
۵	مرداد	/mordād/
۶	شهریور	/šahrivar/
۷	مهر	/mehr/
۸	آبان	'ābān/
۹	آذر	'āzar/
۱۰	دی	/dey/
۱۱	بهمن	/bahman/
۱۲	اسفند	'esfand/

سیگنال‌های الکترومایوگرام (آرتیفکت حرکتی) جلوگیری شود. برای ثبت سیگنال الکترومایوگرام سطحی چهره، از دستگاه ثبت سیگنال الکترومایوگرام Biometric DataLog MWX8 با ۸ کانال و الکتروودهای فعال استفاده شده است. پیش از اتصال الکتروودها، به منظور کاهش امپدانس پوست-الکتروود، محل الکتروودگذاری با الکل تمیز و سپس الکتروودها با آرایش دوقطبی قرار داده شده است. الکتروود مرجع<sup>۱</sup> نیز در ناحیه‌ی استخوانی پشت گوش نصب شده است. الکتروود اول روی عضله‌ی ریسوریوس<sup>۲</sup>، الکتروود دوم روی عضله‌ی دپرسور انگولی اوریس<sup>۳</sup>، الکتروود سوم روی عضله‌ی پلاتیسم<sup>۴</sup>، الکتروود چهارم روی عضله‌ی منتالیس<sup>۵</sup>، الکتروود پنجم روی عضله‌ی بوکسیناتور<sup>۶</sup>، الکتروود ششم روی عضله‌ی لویتورلی سوپریوریس<sup>۷</sup>، الکتروود هفتم روی عضله‌ی زیگوماتیکوس مینور و مازور<sup>۸</sup> و الکتروود هشتم روی عضله‌ی جوشی<sup>۹</sup> قرار داده شده است. موقعیت قرارگیری الکتروودها و نمای ظاهری دستگاه ثبت سیگنال در شکل (۲) نشان داده شده است.

سیگنال‌های الکترومایوگرام تحت تاثیر نویز برق شهر و آرتیفکت حرکتی قرار می‌گیرند. عمده‌ی انرژی طیف فرکانسی سیگنال EMG خام در محدوده‌ی ۱۰ تا ۲۵۰ هرتز است [۲۱]. هم‌چنین محدوده‌ی فرکانسی مربوط به آرتیفکت حرکتی از ۰ تا ۱۰ هرتز و فرکانس نویز برق ۵۰ هرتز است [۲۱]. با استفاده از فیلتر ناچ ۵۰ هرتز نویز مربوط به برق شهر روی تمام سیگنال‌ها تضعیف شده است. هم‌چنین برای حذف آرتیفکت حرکتی، یک فیلتر بالاگذر باترورت مرتبه‌ی ۶ با فرکانس قطع ۱۰ هرتز روی تمام کانال‌ها اعمال شده است [۲۱، ۲۲].

برای ثبت سیگنال از داوطلب‌ها، یک برنامه‌ی اندروید طراحی شده است تا داوطلب‌ها در زمان‌های مقرر هر کلمه‌ای که روی برنامه به نمایش در می‌آید را بیان کنند. کلمات با فاصله‌ی ۶ ثانیه از هم بیان می‌شوند. پس از بیان هر مجموعه‌ی ۱۲ کلمه‌ای، بدون آن‌که ثبت متوقف شود به داوطلب ۱۰ ثانیه استراحت داده شده و مجدد ۱۲ کلمه با فواصل ۶ ثانیه بیان می‌شود. این روند برای هر داوطلب ۱۰ بار تکرار شده است. در نهایت از هر داوطلب ۱۲۰ کلمه شامل ۱۲ کلاس و ۱۰ تکرار به دست آمده است (۱۲۰=۱۲×۱۰). تعداد کل داده‌ها برای داوطلب‌ها برابر با ۳۶۰۰ کلمه شامل ۱۲ کلاس، ۱۰ تکرار و ۳۰ داوطلب است (۳۶۰۰=۱۲×۱۰×۳۰). الگواری ثبت سیگنال

<sup>۶</sup> Buccinator

<sup>۷</sup> Levator Labii Superioris

<sup>۸</sup> Zygomaticus Minor and Major

<sup>۹</sup> Masseter

<sup>۱</sup> Reference Electrode

<sup>۲</sup> Risorius

<sup>۳</sup> Depressor Anguli Oris

<sup>۴</sup> Platysma

<sup>۵</sup> Mentalis

## ۲-۲- استخراج ویژگی

میانگین مربعات سیگنال EMG در حالت استراحت (بدون انقباض) است که با رابطه‌ی (۱) تعریف می‌شود [۲۸].

$$\varepsilon = R \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i)^2} \quad (1)$$

در این رابطه  $r_i$  نمونه‌های سیگنال در حالت استراحت و  $N$  تعداد نمونه‌ها است. در مقاله‌ی [۲۸] مقدار  $R$  از ۰ تا ۰۶ پیشنهاد شده است. در مقاله‌ی حاضر مقدار آستانه برابر ۰/۰۱ میلی‌ولت و مقدار  $R$  برابر با ۰/۶ تنظیم شده است. هم‌چنین در ویژگی ششم  $Q_1$  معادل با چارک ۲۵ و  $Q_3$  معادل با چارک ۷۵ است. جهت محاسبه‌ی ویژگی ششم هر پنجره به طول ۱۲۸ نمونه در ابتدا از کوچک به بزرگ مرتب شده و سپس چارک ۷۵-ام از چارک ۲۵-ام کسر شده است.

برای استخراج ویژگی، ابتدا هر سیگنال پنجره‌گذاری شده و ویژگی‌ها برای هر پنجره‌ی لغزنده به طول ۱۲۸ نمونه و با هم‌پوشانی ۲۵٪ محاسبه شده است [۲۳، ۲۴]. هر کانال از هر سیگنال به ۹ پنجره تقسیم شده است. در این مقاله از ویژگی‌های حداکثر طول فراکتال<sup>۱</sup> (MFL)، واریانس الکترومایوگرام<sup>۲</sup>، تفاضل میانگین مقدار مطلق<sup>۳</sup> (DAMV)، لگاریتم عمل‌گر انرژی Teager-Kaiser (LTKEO)، کاردینالیته<sup>۴</sup> (CARD) و محدوده‌ی بین چارکی<sup>۵</sup> (IQR) استفاده شده که این ویژگی‌ها به همراه روابط آن‌ها در جدول (۲) ارائه شده است. در این روابط  $x$  بیان‌گر هر نمونه در یک پنجره از سیگنال،  $i$  بیان‌گر شماره‌ی آن نمونه و  $N$  بیان‌گر طول پنجره یا تعداد کل نمونه‌ها در آن پنجره از سیگنال است. مقدار آستانه در ویژگی پنجم برابر با حاصل ضرب  $R$  در مجذور

جدول (۲) - روابط و اسامی ویژگی‌های استخراج شده

شماره	نام ویژگی	رابطه‌ی ریاضی	مرجع
۱	حداکثر طول فراکتال	$MFL = \log \left( \sqrt{\sum_{i=1}^{N-1} (x_{i+1} - x_i)^2} \right)$	[۲۴]
۲	واریانس EMG	$VAR = \frac{1}{N-1} \sum_{i=1}^N x_i^2$	[۲۵]
۳	میانگین قدر مطلق تفاضل	$DAMV = \frac{1}{N-1} \sum_{i=1}^{N-1}  x_{i+1} - x_i $	[۲۶]
۴	لگاریتم عمل‌گر انرژی Kaiser-Teager	$\log(\Psi) = \log \left( \sum_{i=2}^{N-1} x_i^2 - x_{i-1}x_{i+1} \right)$	[۲۷]
۵	کاردینالیته	Step 1: $y_i = \text{sort}(x_i)$ , $i = 1: N$ , Step 2: $CARD = \sum_{i=1}^{N-1}  y_i - y_{i+1}  > \varepsilon$	[۲۸]
۶	محدوده‌ی بین چارکی	$IQR = Q_3 - Q_1$	[۲۹]

ویژگی  $i$ -ام و  $f_{i_{\max}}$  بیش‌ترین مقدار ویژگی  $i$ -ام می‌باشد [۲۷]. در نهایت ماتریس مجموعه‌ی داده‌های ویژگی برای آزمایش و آموزش به ترتیب برابر با  $۳۶۰ \times ۴۳۲$  و  $۳۲۴۰ \times ۴۳۲$  شده است.

$$f_{i_{\text{norm}}} = \frac{f_i - f_{i_{\min}}}{f_{i_{\max}} - f_{i_{\min}}} \quad (2)$$

## ۲-۳- کاهش ویژگی‌ها با استفاده از تحلیل مولفه‌های اصلی

تحلیل مولفه‌ی اصلی به وسیله‌ی دنباله‌هایی از ترکیبات خطی متعامد از متغیرهای اولیه با نام مولفه‌های اصلی که واریانس

با توجه به این که از هر سیگنال الکترومایوگرام ۶ ویژگی استخراج شده، طول هر بردار ویژگی برابر با ۴۳۲ بوده که برابر با حاصل ضرب ۶ ویژگی، ۹ پنجره و ۸ کانال است. ویژگی‌ها در مجموعه‌ی داده ممکن است شامل مقادیری باشند که در محدوده‌ی متفاوتی قرار گیرند. این ویژگی‌ها با مقادیر بزرگ ممکن است نسبت به ویژگی‌های با مقادیر کم اثر بسیار زیادتری در تابع هزینه داشته باشند. این مشکل با نرمالیزه نمودن ویژگی‌ها طوری که مقادیر آن‌ها در محدوده‌های مشابه قرار گیرند برطرف می‌شود. در این مقاله ویژگی‌های استخراج شده به روش Min-Max و با رابطه‌ی (۲) نرمالیزه شده است. در این رابطه  $f_i$  ویژگی که قرار است نرمالیزه شود،  $f_{i_{\min}}$  کم‌ترین مقدار

<sup>۱</sup> Cardinality

<sup>۵</sup> Interquartile Range

<sup>۱</sup> Maximum Fractal Length

<sup>۲</sup> Variance Of EMG

<sup>۳</sup> Difference Absolute Mean Value

تابع کرنل خطی در رابطه‌ی (۳)، تابع کرنل چندجمله‌ای در رابطه‌ی (۴) و تابع کرنل پایه‌ی شعاعی در رابطه‌ی (۵) نشان داده شده است [۳۰].

$$K(X_i, X_j) = 1 + X_i^T X_j \quad (۳)$$

$$K(X_i, X_j) = (1 + X_i^T X_j)^p \quad (۴)$$

$$K(X_i, X_j) = \exp\left(-\frac{\|X_i - X_j\|^2}{2\sigma^2}\right) \quad (۵)$$

در این روابط  $X$  بردار داده‌ها و  $p$  درجه‌ی چندجمله‌ای است. کرنل خطی ساده‌ترین نوع کرنل است و برای داده‌های جدایی‌پذیر خطی، عمل‌کرد خوبی نشان می‌دهد. عمل‌کرد کرنل چندجمله‌ای به درجه‌ی  $p$  چندجمله‌ای وابسته است زیرا میزان جدایی‌پذیری داده‌ها به  $p$  بستگی دارد [۳۰]. در این مقاله عمل‌کرد چندجمله‌ای برای درجه‌ی ۴ ارزیابی شده است. به منظور طبقه‌بندی با استفاده از KNN معیار فاصله، اقلیدسی و تعداد همسایه برابر با ۱۳ تنظیم شده است.

## ۲-۵- معیارهای کمی ارزیابی

عمل‌کرد طبقه‌بند با استفاده از معیارهای کمی صحت<sup>۴</sup> و حساسیت<sup>۵</sup> مورد ارزیابی قرار گرفته که این معیارها به ترتیب در روابط (۶) و (۷) معرفی شده است [۳۱].

$$\text{Accuracy} = \frac{\text{تعداد نمونه‌های درست پیش‌بینی شده}}{\text{تعداد کل نمونه‌ها}} \quad (۶)$$

$$\text{Sensitivity}_i = \frac{TP_i}{(TP_i + FN_i)} \quad (۷)$$

پارامتر صحت، اساسی‌ترین معیار اندازه‌گیری کیفیت یک طبقه‌بند بوده و برابر با تعداد نمونه‌های درست پیش‌بینی شده تقسیم بر تعداد کل نمونه‌ها است. معیار حساسیت که نرخ پاسخ‌های مثبت درست<sup>۶</sup> نیز نامیده می‌شود به معنی نسبتی از موارد درست در هر کلاس بوده که طبقه‌بند آن‌ها را به درستی از همان کلاس تشخیص داده است. در واقع به ازای هر کلاس یک مقدار حساسیت وجود دارد که از رابطه‌ی (۷) به دست می‌آید. در این رابطه  $TP_i$  مثبت واقعی (تعداد نمونه‌هایی که عضو کلاس  $i$  بوده و عضو کلاس  $i$  تشخیص داده شده است)،  $FN_i$  منفی کاذب<sup>۷</sup> (نمونه‌ی عضو کلاس  $i$  که به اشتباه در کلاس

ماکسیم دارند به دست می‌آید [۲۹]. تحلیل مولفه‌ی اصلی روشی برای استخراج مولفه‌های اصلی از مجموعه‌ی بزرگ ویژگی‌های استخراج شده است. در واقع یک مجموعه با بعد پایین از مولفه‌ها از یک مجموعه‌ی دارای بعد بالا از ویژگی‌ها استخراج می‌شود. اولین مولفه، یک ترکیب خطی از ویژگی‌ها است که جهت بیش‌ترین تغییرات در داده‌ها را تعیین می‌کند. مولفه‌های بعدی نیز از مفهومی مشابه آن‌چه بیان شده، پیروی می‌کنند. به عبارت دیگر آن‌ها مقدار واریانس باقی‌مانده را بدون آن که با مولفه‌های پیشین دارای همبستگی شوند، در خود حفظ می‌کنند. [۲۹].

پیش از مرحله‌ی طبقه‌بندی، ویژگی‌ها با استفاده از PCA کاهش یافته است. نتایج بر اساس تعداد مولفه‌های اصلی از ۱ مولفه تا ۶۰ مولفه مورد ارزیابی و بررسی قرار گرفته که در بخش ۳ در شکل (۵) گزارش شده است.

## ۲-۴- طبقه‌بندی

ویژگی‌های استخراج شده به طبقه‌بند داده شده تا عملیات آموزش و آزمایش انجام گرفته و بتوان با اعمال داده‌های آزمایش به طبقه‌بند پیش‌بینی کرد که هر سیگنال مربوط به کدام کلاس است. بدین منظور از طبقه‌بندهای SVM، KNN و RF استفاده شده است. به منظور اعتبارسنجی از روش 10-fold استفاده شده که در این اعتبارسنجی، داده‌ها به  $K$  زیرمجموعه افزاز شده که از این  $K$  زیرمجموعه هر بار یکی برای اعتبارسنجی و  $K-1$  زیرمجموعه‌ی دیگر برای آموزش به کار رفته است. این روال  $K$  بار تکرار شده و تمام داده‌ها دقیقاً یک بار برای اعتبارسنجی به کار رفته است. تعداد داده‌های آموزش و آزمایش برای هر fold به ترتیب برابر با ۳۲۴۰ و ۳۶۰ است.

ماشین بردار پشتیبان یک طبقه‌بند دوتایی است اما این امکان وجود دارد تا با روش‌های متفاوتی از این کلاس‌بند به منظور حل مسائل چندکلاسه نیز استفاده شود. در این مقاله از روش یک کلاس در برابر همه استفاده شده است [۳۰]. در استفاده از الگوریتم SVM، تعیین کرنل مناسب که وظیفه‌ی نگاشت مناسب داده‌ها در فضای ویژگی‌ها را بر عهده دارد حائز اهمیت است. بدین منظور ابتدا الگوریتم با استفاده از سه کرنل پایه‌ی شعاعی<sup>۱</sup>، چندجمله‌ای<sup>۲</sup> و خطی<sup>۳</sup> مورد آزمایش قرار گرفته است. از آن‌جا که کرنل چندجمله‌ای دارای نتایج بهتری بوده تمام نتایج بر اساس این کرنل ارائه شده است.

<sup>۴</sup> Sensitivity

<sup>۵</sup> True Positive Rate

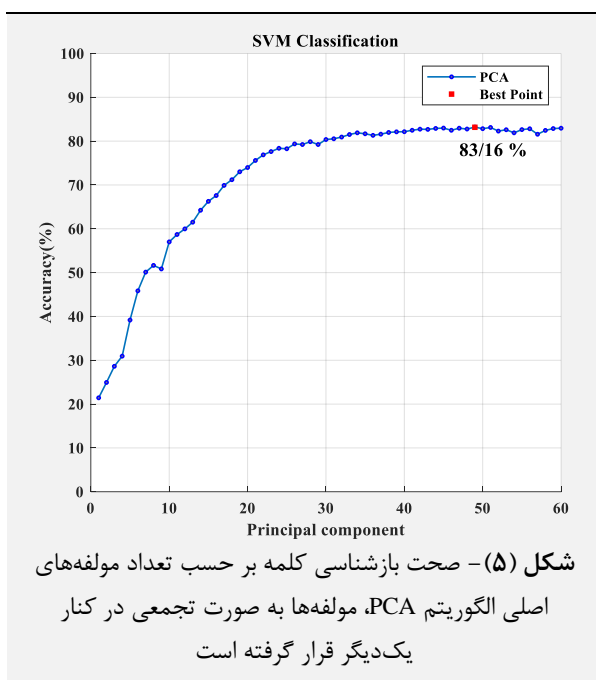
<sup>۷</sup> False Negative

<sup>۱</sup> Radial Basis Function

<sup>۲</sup> Polynomial

<sup>۳</sup> Linear

<sup>۴</sup> Accuracy

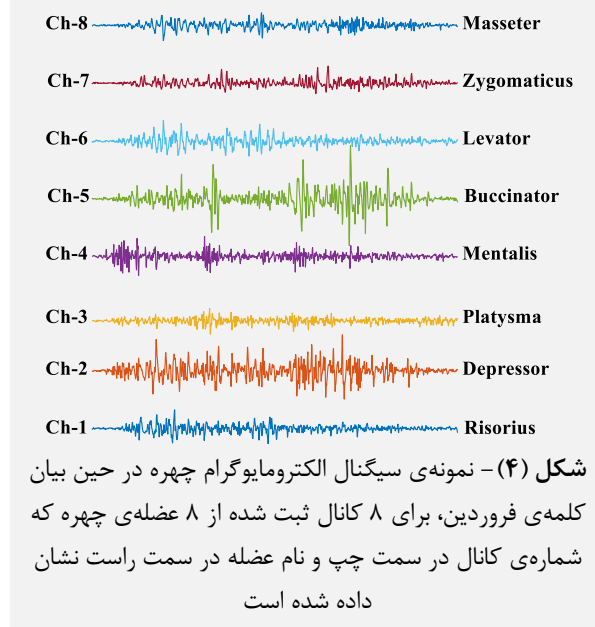


مشاهده می‌شود که پس از مولفه‌ی ۴۰ که بیان‌گر قرار گرفتن ۴۰ مولفه‌ی اصلی اول در کنار یکدیگر است، صحت به طور معناداری افزایش پیدا نمی‌کند. به این دلیل که مولفه‌های ۴۰ به بعد دارای اطلاعات کم یا کم‌اثر هستند، با این وجود جهت اطمینان تا مولفه‌ی ۶۰ مورد بررسی قرار گرفته است. میزان صحت بازشناسی کلمه برای ۱۲ کلمه‌ی فارسی با رسیدن به ۴۹ مولفه به بالاترین مقدار ۸۳/۱۶٪ رسیده است. از این رو تعداد مولفه‌های اصلی مورد استفاده در ادامه برابر با ۴۹ است. در شکل (۶) ماتریس اغتشاش حاصل از طبقه‌بندی ۱۲ کلاس با SVM بدون کاهش ابعاد ارائه شده که در آن محور افقی نشان دهنده‌ی کلاس‌های هدف و محور عمودی نشان دهنده‌ی کلاس‌های شناسایی شده بوده و هم‌چنین قطر اصلی ماتریس اغتشاش نشان دهنده‌ی تعداد داده‌های درست تشخیص داده شده می‌باشد. مشاهده می‌شود که کم‌ترین تعداد داده‌ی درست تشخیص داده شده مربوط به شناسایی کلاس ۱۰ یعنی کلمه‌ی دی با ۱۹۲ مورد بوده که ۶۹ مورد از داده‌های آزمایش به اشتباه کلاس ۴ (تیر) تشخیص داده شده است. از طرفی کلاس ۴ (تیر) بعد از کلمه‌ی دی دارای کم‌ترین تعداد داده‌ی درست تشخیص داده شده با ۲۱۵ مورد بوده که بیش‌ترین میزان خطای آن ۴۸ مورد می‌باشد که به اشتباه کلاس ۱۰ (دی) شناسایی شده است. کلاس‌های ۲ (اردیبهشت)، ۱ (فروردین)، ۸ (آبان) و ۶ (شهریور) به ترتیب دارای بالاترین نرخ بازشناسی هستند. در شکل (۷) ماتریس اغتشاش حاصل از کاهش ابعاد و طبقه‌بندی با SVM گزارش شده است. مشاهده می‌شود که با استفاده از الگوریتم PCA و کاهش ابعاد از ۴۳۲ ویژگی به ۴۹

دیگری تشخیص داده شده است) و i شماره‌ی کلاس است. تعداد ۱۲ مقدار حساسیت برای هر کلاس به دست آمده است. مقادیر حساسیت مربوط به هر کلاس در سطر اول ماتریس اغتشاش در شکل‌های (۶) و (۷) گزارش شده است. میانگین مقدار حساسیت برای هر ۱۲ کلاس در جدول (۳) ارائه شده است [۳۱، ۳۲].

### ۳- نتایج

در شکل (۴) یک نمونه از سیگنال متعلق به کلمه‌ی فروردین نشان داده شده است. در این شکل محور افقی نشان دهنده‌ی زمان است (۰/۹ ثانیه). تغییرات دامنه‌ی سیگنال بین ۲- تا ۲ میلی‌ولت است. هر یک از سیگنال‌ها که در یک سطر نمایش داده شده نشان دهنده‌ی یک کانال بوده که توسط یک الکتروود از روی یک نقطه از چهره ثبت شده است.



برداری شامل ۴۳۲ ویژگی با استفاده از الگوریتم تحلیل مولفه‌های اصلی به ۴۹ مولفه‌ی اصلی تقلیل یافته است تا صحت بازشناسی کلمه بر حسب تعداد مولفه‌های اصلی ارزیابی شود. مولفه‌های آغازین دارای اطلاعات بیش‌تر و موثرتری هستند اما به تنهایی کارآمدی لازم را ندارند. به طور مثال در شکل (۵) مولفه‌ی اصلی اول به تنهایی دارای صحت ۲۱٪ در بازشناسی ۱۲ کلمه است. به دلیل ناهمبسته بودن مولفه‌های اصلی با یکدیگر، زمانی که به طور هم‌زمان و تجمعی از مولفه‌ها استفاده شود صحت بازشناسی افزایش می‌یابد. به همین منظور به ترتیب، مولفه‌های اصلی از پایین‌ترین به بالاترین به صورت تجمعی در کنار یکدیگر قرار گرفته است.



نامربوط، صحت طبقه‌بندی را بهبود بخشد. ساده‌سازی فضای ویژگی و کاهش ابعاد فضا موجب می‌شود تا طبقه‌بندی ساده‌تر و با پیچیدگی کم‌تری انجام شود و صحت افزایش یابد. به بیان دیگر کاهش پیچیدگی فضای ویژگی‌ها می‌تواند موجب بهبود صحت طبقه‌بندی و کاهش زمان پردازش شود [۳۶].

مولفه، صحت به اندازه‌ی ۰/۹۱٪ افزایش یافته است. تحلیل مولفه‌های اصلی با کاهش دادن ابعاد داده‌ها موجب سرعت بخشیدن به محاسبات می‌شود. هم‌چنین زمانی که ابعاد بالایی با متغیرهای همبسته وجود داشته باشد، PCA می‌تواند با حذف ویژگی‌های افزونه (ویژگی‌های وابسته) و داده‌های نویزی و

SVM without PCA Confusion Matrix

1	273 7.6%	11 0.3%	3 0.1%	3 0.1%	2 0.1%	6 0.2%	3 0.1%	0 0.0%	2 0.1%	0 0.0%	2 0.1%	5 0.1%	38.1% 11.9%
2	7 0.2%	279 7.8%	2 0.1%	1 0.0%	1 0.0%	3 0.1%	0 0.0%	5 0.1%	6 0.2%	1 0.0%	3 0.1%	6 0.2%	38.9% 11.1%
3	1 0.0%	2 0.1%	232 6.4%	7 0.2%	30 0.8%	2 0.1%	0 0.0%	3 0.1%	4 0.1%	6 0.2%	3 0.1%	0 0.0%	30.0% 20.0%
4	2 0.1%	0 0.0%	5 0.1%	215 6.0%	7 0.2%	1 0.0%	16 0.4%	3 0.1%	4 0.1%	69 1.9%	2 0.1%	10 0.3%	34.4% 35.6%
5	3 0.1%	1 0.0%	36 1.0%	8 0.2%	237 6.6%	9 0.2%	6 0.2%	1 0.0%	4 0.1%	3 0.1%	1 0.0%	4 0.1%	75.7% 24.3%
6	2 0.1%	3 0.1%	3 0.1%	1 0.0%	4 0.1%	269 7.5%	4 0.1%	1 0.0%	5 0.1%	2 0.1%	9 0.2%	6 0.2%	37.1% 12.9%
7	0 0.0%	0 0.0%	1 0.0%	4 0.1%	1 0.0%	1 0.0%	230 6.4%	2 0.1%	1 0.0%	11 0.3%	5 0.1%	2 0.1%	39.1% 10.9%
8	1 0.0%	3 0.1%	1 0.0%	3 0.1%	4 0.1%	1 0.0%	7 0.2%	271 7.5%	14 0.4%	4 0.1%	5 0.1%	0 0.0%	36.3% 13.7%
9	2 0.1%	0 0.0%	9 0.2%	2 0.1%	5 0.1%	0 0.0%	2 0.1%	5 0.1%	250 6.9%	5 0.1%	1 0.0%	1 0.0%	38.7% 11.3%
10	0 0.0%	0 0.0%	4 0.1%	48 1.3%	1 0.0%	2 0.1%	22 0.6%	0 0.0%	1 0.0%	192 5.3%	6 0.2%	3 0.1%	38.8% 31.2%
11	4 0.1%	0 0.0%	0 0.0%	1 0.0%	2 0.1%	2 0.1%	5 0.1%	6 0.2%	6 0.2%	1 0.0%	255 7.1%	5 0.1%	38.9% 11.1%
12	5 0.1%	1 0.0%	4 0.1%	7 0.2%	6 0.2%	4 0.1%	5 0.1%	3 0.1%	3 0.1%	6 0.2%	8 0.2%	258 7.2%	33.2% 16.8%
	91.0%	93.0%	77.3%	71.7%	79.0%	89.7%	76.7%	90.3%	83.3%	64.0%	85.0%	86.0%	82.3%
	9.0%	7.0%	22.7%	28.3%	21.0%	10.3%	23.3%	9.7%	16.7%	36.0%	15.0%	14.0%	17.8%
	1	2	3	4	5	6	7	8	9	10	11	12	

شکل (۶) - ماتریس اغتشاش برای شناسایی کلمات (کلاس‌ها) توسط SVM

۹ و ۱۰ روش پیشنهادی عمل کرد پایین‌تری داشته که دلیل آن استفاده از تعداد الکتروود بیش‌تر است. استفاده از الکتروودهای کم‌تر به دلیل مزایای نصب آسان‌تر، زمان آماده‌سازی کم‌تر، مصرف انرژی کم‌تر و سرعت پردازش بالاتر مطلوب است. هم‌چنین فراهم کردن دستگاه ثبت سیگنال EMG با تعداد الکتروود بیش‌تر به معنی افزایش هزینه‌ی دستگاه بوده که در عمل و برای کاربردهای واقعی، نامطلوب است.

جدول (۳) - نتایج طبقه‌بندی بر اساس روش‌ها

روش	صحت	حساسیت
SVM	۸۲/۲۵	۰/۸۲۲۵
SVM+PCA	۸۳/۱۶	۰/۸۳۲۵
KNN+PCA	۸۱/۹۱	۰/۸۱
RF+PCA	۷۸/۹۷	۰/۷۷

در نهایت داده‌ها با استفاده از KNN و RF نیز مورد ارزیابی قرار گرفته است. نتایج طبقه‌بندی ۱۲ کلاس با استفاده از سه الگوریتم SVM، KNN و RF بر اساس معیارهای صحت و حساسیت در جدول (۳) گزارش شده است.

در نهایت نتایج حاصل از این مقاله با مقالات دیگر با زبان‌های مختلف و بر اساس تعداد کلاس‌های متفاوت مقایسه شده که در شکل (۸) نشان داده شده است. در این شکل محور افقی بیان‌گر تعداد کلاس، محور عمودی بیان‌گر میزان صحت و هر نقطه‌ی رنگی بیان‌گر مقدار صحت گزارش شده در یک مرجع است. برای صحت‌سنجی مناسب، صحت بازنسائی کلمه برای روش پیشنهادی به ازای تعداد کلمات از ۱ تا ۱۲ محاسبه شده است تا امکان مقایسه با پژوهش‌های پیشین با تعداد کلمات مختلف فراهم گردد. مشاهده می‌شود که روش پیشنهادی در مقایسه با سایر روش‌ها دارای عمل کرد بهتری است. تنها در تعداد کلاس

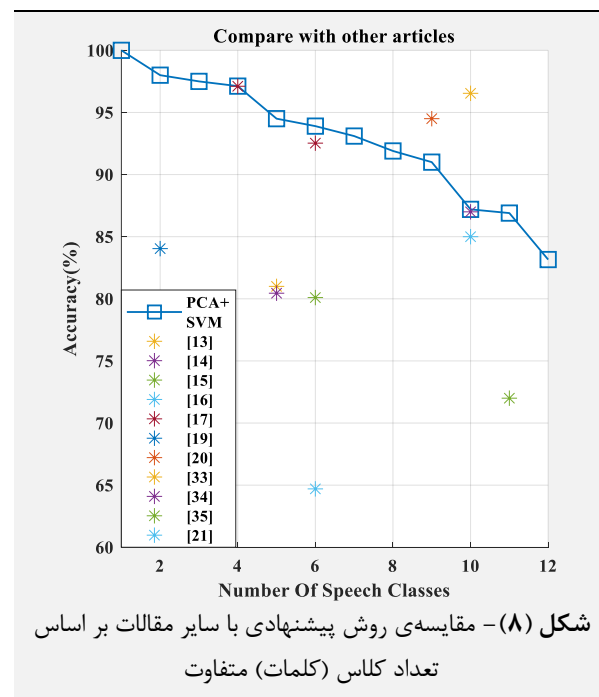
SVM+PCA Confusion Matrix

Output Class	1	268 7.4%	14 0.4%	3 0.1%	1 0.0%	1 0.0%	7 0.2%	0 0.0%	2 0.1%	3 0.1%	1 0.0%	1 0.0%	6 0.2%	87.3%
	2	16 0.4%	279 7.8%	4 0.1%	1 0.0%	2 0.1%	4 0.1%	0 0.0%	8 0.2%	2 0.1%	1 0.0%	1 0.0%	5 0.1%	86.4%
	3	1 0.0%	1 0.0%	232 6.4%	9 0.2%	40 1.1%	2 0.1%	0 0.0%	3 0.1%	0 0.0%	5 0.1%	2 0.1%	3 0.1%	77.9%
	4	2 0.1%	0 0.0%	10 0.3%	222 6.2%	6 0.2%	0 0.0%	21 0.6%	2 0.1%	3 0.1%	54 1.5%	3 0.1%	7 0.2%	67.3%
	5	3 0.1%	0 0.0%	37 1.0%	6 0.2%	230 6.4%	6 0.2%	3 0.1%	1 0.0%	1 0.0%	0 0.0%	0 0.0%	1 0.0%	79.9%
	6	3 0.1%	1 0.0%	3 0.1%	0 0.0%	4 0.1%	270 7.5%	4 0.1%	0 0.0%	2 0.1%	0 0.0%	3 0.1%	4 0.1%	91.8%
	7	0 0.0%	0 0.0%	0 0.0%	5 0.1%	2 0.1%	1 0.0%	232 6.4%	5 0.1%	3 0.1%	14 0.4%	4 0.1%	4 0.1%	85.9%
	8	1 0.0%	1 0.0%	1 0.0%	3 0.1%	3 0.1%	1 0.0%	5 0.1%	265 7.4%	14 0.4%	2 0.1%	5 0.1%	2 0.1%	87.5%
	9	2 0.1%	1 0.0%	4 0.1%	2 0.1%	3 0.1%	3 0.1%	1 0.0%	5 0.1%	260 7.2%	5 0.1%	0 0.0%	4 0.1%	89.7%
	10	0 0.0%	0 0.0%	1 0.0%	43 1.2%	1 0.0%	1 0.0%	21 0.6%	0 0.0%	2 0.1%	209 5.8%	4 0.1%	0 0.0%	74.1%
	11	2 0.1%	1 0.0%	0 0.0%	2 0.1%	6 0.2%	2 0.1%	7 0.2%	5 0.1%	5 0.1%	4 0.1%	270 7.5%	7 0.2%	86.8%
	12	2 0.1%	2 0.1%	5 0.1%	6 0.2%	2 0.1%	3 0.1%	6 0.2%	4 0.1%	5 0.1%	5 0.1%	7 0.2%	257 7.1%	84.5%
		89.3%	93.0%	77.3%	74.0%	76.7%	90.0%	77.3%	88.3%	86.7%	89.7%	90.0%	85.7%	83.2%
		10.7%	7.0%	22.7%	26.0%	23.3%	10.0%	22.7%	11.7%	13.3%	30.3%	10.0%	14.3%	16.8%
		1	2	3	4	5	6	7	8	9	10	11	12	
		Target Class												

شکل (۷) - ماتریس اغتشاش برای شناسایی کلمات (کلاس‌ها) توسط SVM+PCA

صوتی) در حوزه‌ی زمان و با استفاده از سه طبقه‌بند ماشین بردار پشتیبان، نزدیک‌ترین همسایه‌ی K و جنگل تصادفی انجام شده است. از آن‌جا که تا به امروز بازشناسی کلمات زبان فارسی با استفاده از سیگنال الکترومایوگرام انجام نشده و روز به روز تعداد بیماران مبتلا به سرطان حنجره و یا اختلالات گفتاری رو به افزایش است، این تحقیق می‌تواند در راستای کمک به این بیماران مورد استفاده قرار گیرد.

به منظور اجرای روش پیشنهادی به دلیل آن که هیچ پایگاه داده‌ای از سیگنال‌های الکترومایوگرام چهره به زبان فارسی وجود نداشت اقدام به ایجاد پایگاه داده‌ای برای این منظور شده است. پایگاه داده‌ی ایجاد شده شامل سیگنال الکترومایوگرام ۸ عضله‌ی چهره هنگام بیان ۱۲ کلمه‌ی فارسی است. در ابتدا پیش‌پردازش‌های لازم به منظور حذف نویز روی سیگنال‌های ثبت شده اعمال گردیده، سپس سیگنال به پنجره‌هایی تقسیم شده و پس از آن ۶ ویژگی از هر پنجره استخراج شده است. به منظور کم کردن حجم داده‌ها از الگوریتم تحلیل مولفه‌ی اصلی برای کاهش ویژگی‌ها استفاده شده و در نهایت بردارهای ویژگی برای آموزش و آزمایش به طبقه‌بند داده شده است. با مشاهده‌ی نتایج مربوط به ماتریس اغتشاش در شکل (۷) مشخص می‌شود



شکل (۸) - مقایسه‌ی روش پیشنهادی با سایر مقالات بر اساس تعداد کلاس (کلمات) متفاوت

#### ۴- یافته‌ها و بحث

در این مقاله بازشناسی ۱۲ کلمه‌ی رایج از زبان فارسی مبتنی بر سیگنال الکترومایوگرام (بدون استفاده از سیگنال گفتار



که بتواند با حداقل تعداد الکترومد، تعداد کلمات بیش‌تری را با صحت بالاتر طبقه‌بندی کند فراهم گردد. از طرفی با توجه به عدم وجود پایگاه داده‌ی سیگنال‌های الکترومایوگرام عضلات چهره در حین بیان کلمات فارسی، پیشنهاد می‌شود که پایگاه داده‌ای برای تحقیقات بیش‌تر ایجاد شود. در آینده سعی خواهد شد تا تحقیقات روی جملات و عبارات فارسی با دایره‌ی واژگان وسیع‌تر برای تشخیص مبتنی بر سیگنال الکترومایوگرام (بدون استفاده از سیگنال صوتی) انجام شود.

## ۶- مراجع

- [1] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, "Silent speech interfaces," *Speech Commun.*, vol. 52, no. 4, pp. 270–287, 2010, doi: 10.1016/j.specom.2009.08.002.
- [2] T. Schultz, M. Wand, T. Hueber, D. J. Krusienski, C. Herff, and J. S. Brumberg, "Biosignal-Based Spoken Communication: A Survey," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 12, pp. 2257–2271, 2017, doi: 10.1109/TASLP.2017.2752365.
- [3] B. Denby et al., "Recent results in silent speech interfaces," *J. Acoust. Soc. Am.*, vol. 141, no. 5, pp. 3646–3646, 2017, doi: 10.1121/1.4987881.
- [4] M. Wand, J. Koutnik, and J. Schmidhuber, "Lipreading with long short-term memory," in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mar. 2016, pp. 6115–6119. doi: 10.1109/ICASSP.2016.7472852.
- [5] J. S. Chung, A. Senior, O. Vinyals, and A. Zisserman, "Lip reading sentences in the wild," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3444–3450, 2017, doi: 10.1109/CVPR.2017.367.
- [6] L. A. Cheah et al., "Preliminary evaluation of a silent speech interface based on intra-oral magnetic sensing," *BIODEVICES 2016 - 9th Int. Conf. Biomed. Electron. Devices, Proceedings; Part 9th Int. Jt. Conf. Biomed. Eng. Syst. Technol. BIOSTEC 2016*, pp. 108–118, 2016, doi: 10.5220/0005824501080116.
- [7] J. A. Gonzalez et al., "A silent speech system based on permanent magnet articulography and direct synthesis," *Comput. Speech Lang.*, vol. 39, pp. 67–87, 2016, doi: 10.1016/j.csl.2016.02.002.
- [8] F. Bocquet, T. Hueber, L. Girin, C. Savariaux, and B. Yvert, "Real-Time Control of an Articulatory-Based Speech Synthesizer for Brain Computer Interfaces," *PLoS Comput. Biol.*, vol. 12, no. 11, 2016, doi: 10.1371/journal.pcbi.1005119.
- [9] T. Schultz and M. Wand, "Modeling coarticulation in EMG-based continuous speech recognition," *Speech Commun.*, vol. 52, no. 4,

که برخی از کلاس‌ها با یکدیگر تداخل دارند. به عنوان مثال کلاس‌های ۳ (خرداد) و ۵ (مرداد) با یکدیگر تداخل دارند و کلاس‌های ۴ (تیر) و ۱۰ (دی) به یکدیگر نزدیک هستند و سیستم پیشنهادی به اشتباه داده‌های هر یک از این کلمات را با برجسب دیگری شناسایی کرده است. با توجه به این که کلمه‌ی مرداد و خرداد هم از نظر تلفظ و هم از نظر نوع بیان فعال شدن عضلات چهره مشابه هستند لذا حاصل شدن این نتایج قابل توجیه است. از طرفی کلمات تیر و دی نیز نسبت به یکدیگر داده‌های حساسی دارند و سیستم کلمه‌ی تیر را به اشتباه کلمه‌ی دی، و کلمه‌ی دی را به اشتباه کلمه‌ی تیر شناسایی کرده است. بر خلاف دو کلمه‌ی مرداد و خرداد، کلمات تیر و دی از نظر تلفظ و آوا یکسان نبوده و تنها از نظر نوع بیان و فعالیت عضلانی مشابه یکدیگر هستند و به همین جهت نتایج حاصل از این دو کلمه با یکدیگر تداخل داشته است. کلمه‌ی اردیبهشت دارای بالاترین صحت شناسایی بوده که دلیل آن نوع بیان متفاوت این کلمه نسبت به سایر کلمات است و عضلاتی که با بیان این کلمه در چهره‌ی شخص فعال می‌شود با دیگر حالت‌ها تمایز بیش‌تری دارند.

## ۵- نتیجه‌گیری

به عنوان یک فناوری در حال توسعه، استفاده از سیگنال‌های الکترومایوگرام به منظور تشخیص گفتار در حال حاضر با محدودیت‌هایی روبه‌رو است. در مقایسه با تشخیص گفتار مبتنی بر سیگنال صوتی، قابلیت اطمینان EMG پایین‌تر است و امکان تشخیص تعداد بالای کلمات وجود ندارد. بنابراین تا کنون مناسب‌ترین سناریوی کاربردی برای آن، ارسال دستورات فهرست شده به سیستم‌های کنترل از راه دور است. از طرفی تشخیص کلمات به صورت پیوسته در جملات چالش برانگیز است. همچنین محدودیت دیگر، تشخیص واج در بین کلمات است. در واقع به دلیل عدم دسترسی مستقیم به عضلات مرتبط با بیان هر کلمه، بازشناسی گفتار با استفاده از سیگنال‌های EMG پیچیده‌تر است. بر اساس نتایج حاصل از سیستم پیشنهادی می‌توان گفت که در شناسایی کلمات با استفاده از سیگنال‌های الکترومایوگرام چهره عواملی هم‌چون نوع بیان، شدت بیان یا حتی نوع جنسیت می‌تواند تأثیرگذار باشد که باید مورد ارزیابی بیش‌تر قرار گیرد. با توجه به این که تحلیل سیگنال‌های الکترومایوگرام چهره نسبت به سایر عضلات بدن به دلیل نزدیک بودن عضلات به یکدیگر دشوارتر است پیشنهاد می‌شود تا روی محل و تعداد قرارگیری الکترودها و عضلات مطالعات بیش‌تری انجام شود تا امکان طراحی سیستمی



- in a virtual reality environment,” *Virtual Real.*, 2022, doi: 10.1007/s10055-021-00616-0.
- [21] A. Phinyomark and E. Scheme, “EMG pattern recognition in the era of big data and deep learning,” *Big Data Cogn. Comput.*, vol. 2, no. 3, pp. 1–27, 2018, doi: 10.3390/bdcc2030021.
- [22] A. Andrews, E. Morin, and L. McLean, “Optimal electrode configurations for finger movement classification using EMG,” *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009*, pp. 2987–2990, 2009, doi: 10.1109/IEMBS.2009.5332520.
- [23] T. R. Farrell and R. F. Weir, “The optimal controller delay for myoelectric prostheses,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 111–118, 2007, doi: 10.1109/TNSRE.2007.891391.
- [24] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, “Fractal analysis features for weak and single-channel upper-limb EMG signals,” *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11156–11163, 2012, doi: 10.1016/j.eswa.2012.03.039.
- [25] D. Tkach, H. Huang, and T. A. Kuiken, “Study of stability of time-domain features for electromyographic pattern recognition,” *J. Neuroeng. Rehabil.*, vol. 7, no. 1, 2010, doi: 10.1186/1743-0003-7-21.
- [26] K. S. Kim, H. H. Choi, C. S. Moon, and C. W. Mun, “Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions,” *Curr. Appl. Phys.*, vol. 11, no. 3, pp. 740–745, 2011, doi: 10.1016/j.cap.2010.11.051.
- [27] R. N. Khushaba, A. H. Al-Timemy, A. Al-Ani, and A. Al-Jumaily, “A Framework of Temporal-Spatial Descriptors-Based Feature Extraction for Improved Myoelectric Pattern Recognition,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1821–1831, 2017, doi: 10.1109/TNSRE.2017.2687520.
- [28] A. Waris and E. N. Kamavuako, “Effect of threshold values on the combination of EMG time domain features: Surface versus intramuscular EMG,” *Biomed. Signal Process. Control*, vol. 45, pp. 267–273, 2018, doi: 10.1016/j.bspc.2018.05.036.
- [29] A. R. Verma and B. Gupta, “Detecting Neuromuscular Disorders Using EMG Signals Based on TQWT Features,” *Augment. Hum. Res.*, vol. 5, no. 1, 2020, doi: 10.1007/s41133-019-0020-7.
- [30] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, “Surface EMG hand gesture recognition system based on PCA and GRNN,” *Neural Comput. Appl.*, vol. 32, no. 10, pp. 6343–6351, 2020, doi: 10.1007/s00521-019-04142-8.
- [31] M. Vihinen, “How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis”, *BMC Genomics*, vol. 13 Suppl 4, no. Suppl 4, 2012, doi: 10.1186/1471-2164-13-S4-S2.
- pp. 341–353, 2010, doi: 10.1016/j.specom.2009.12.002.
- [10] M. Wand and T. Schultz, “Analysis of phone confusion in EMG-based speech recognition,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 757–760, 2011, doi: 10.1109/ICASSP.2011.5946514.
- [11] M. Wand, M. Janke, and A. T. Schultz, “Tackling speaking mode varieties in EMG-based speech recognition,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2515–2526, 2014, doi: 10.1109/TBME.2014.2319000.
- [12] A. Ratnovsky, S. Malayev, S. Ratnovsky, S. Naftali, and N. Rabin, “EMG-based speech recognition using dimensionality reduction methods,” *J. Ambient Intell. Humaniz. Comput.*, 2021, doi: 10.1007/s12652-021-03315-5.
- [13] Y. Wang, M. Zhang, R. Wu, H. Wang, Z. Luo, and G. Li, “Speech neuromuscular decoding based on spectrogram images using conformal predictors with Bi-LSTM,” *Neurocomputing*, vol. 451, pp. 25–34, 2021, doi: 10.1016/j.neucom.2021.03.025.
- [14] V. Chandrashekhar, “Classification of EMG Signals Using Machine Learning for the Construction of a Silent Speech Interface,” *Young Res.*, vol. 5, no. 1, pp. 266–283, 2021, [Online]. Available: <http://www.theyoungresearcher.com/papers/chandrashekhar.pdf>
- [15] M. Yang and M. Zhang, “An Alternative sEMG based Isolated Word Subvocal Speech Recognition System based on Interpolation Functions,” *Proc. 2020 Int. Conf. Big Data Artif. Intell. Softw. Eng. ICBASE 2020*, pp. 306–309, 2020, doi: 10.1109/ICBASE51474.2020.00071.
- [16] H. Ye et al., “Attention Bidirectional LSTM Networks Based Mime Speech Recognition Using sEMG Data,” *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, vol. 2020-October, pp. 3162–3167, 2020, doi: 10.1109/SMC42975.2020.9282863.
- [17] N. Srisuwan, P. Prukattaranont, and C. Limsakul, “Comparison of Classifiers for EMG based Speech Recognition,” *J. Phys. Conf. Ser.*, vol. 1438, no. 1, 2020, doi: 10.1088/1742-6596/1438/1/012032.
- [18] N. Sae Jong and P. Phukpattaranont, “A speech recognition system based on electromyography for the rehabilitation of dysarthric patients: A Thai syllable study,” *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 234–245, 2019, doi: 10.1016/j.bbe.2018.11.010.
- [19] M. Zhu et al., “Automatic Speech Recognition in Different Languages Using High-Density Surface Electromyography Sensors,” *IEEE Sens. J.*, vol. 21, no. 13, pp. 14155–14167, Jul. 2021, doi: 10.1109/JSEN.2020.3037061.
- [20] H. S. Cha, W. Du Chang, and C. H. Im, “Deep-learning-based real-time silent speech recognition using facial electromyogram recorded around eyes for hands-free interfacing

- [34] S. Ma et al., "Silent Speech Recognition Based on Surface Electromyography," in 2019 Chinese Automation Congress (CAC), Nov. 2019, pp. 4497–4501.
- [35] M. W. Soon, M. I. H. Anuar, M. H. Z. Abidin, A. S. Azaman, and N. M. Noor, "Speech recognition using facial sEMG," Proc. 2017 IEEE Int. Conf. Signal Image Process. Appl. ICSIPA 2017, pp. 1–5, 2017.
- [36] S. Velliangiri, S. Alagumuthukrishnan, and S. I. Thankumar Joseph, "A Review of Dimensionality Reduction Techniques for Efficient Computation," Procedia Comput. Sci., vol. 165, pp. 104–111, 2019, doi: 10.1016/j.procs.2020.01.079.
- [32] G. Purushothaman and R. Vikas, "Identification of a feature selection based pattern recognition scheme for finger movement recognition from multichannel EMG signals," Australas. Phys. Eng. Sci. Med., vol. 41, no. 2, pp. 549–559, 2018, doi: 10.1007/s13246-018-0646-7.
- [33] J. Zhuang et al., "Comparison of Contributions between Facial and Neck Muscles for Speech Recognition Using High-Density surface Electromyography," 2019 IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl. CIVEMSA 2019 - Proc., 2019.