



Persian words recognition based on facial electromyogram signals

Sharifi, Pooria¹ / Soltanizadeh, Hadi^{2*} / Maleki, Ali³

¹ - Ph.D. Student, Electrical and Computer Engineering Department, Semnan University, Semnan, Iran

² - Associate Professor, Electrical and Computer Engineering Department, Semnan University, Semnan, Iran

³ - 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

Address	Electrical Group, Electrical and Computer Engineering Department, Semnan University, Semnan, Iran		
Postal Code	19111-35131	Tel	02331532697
E-Mail	H_soltanizadeh@semnan.ac.ir	Fax	02331532697



تشخیص کلمات فارسی بر اساس سیگنال‌های الکترومایوگرام چهره

شریفی، پوریا^۱ / سلطانی زاده، هادی^{۲*} / مالکی، علی^۳

^۱ - دانشجوی دکتری، گروه مهندسی برق، دانشکده برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران

^۲ - دانشیار، گروه مهندسی برق، دانشکده برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران

^۳ - 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

واژه‌های کلیدی

از دست دادن صدا و حنجره یک معضل بزرگ برای افراد با اختلالات گفتاری محسوب می‌شود. این اتفاق عواقب جدی و منفی بر کیفیت زندگی فردی و گروهی این اشخاص به ویژه در محیط‌های کاری ایجاد می‌کند. توسعه یک سیستم هوشمند بر پایه سیگنال‌های الکترومایوگرام با قابلیت تشخیص گفتار (بدون استفاده از صوت) می‌تواند دریچه امیدی برای افرادی که حنجره و صدای خود را بر اثر سرطان از دست داده‌اند باشد. اگرچه در این حوزه پژوهش‌هایی برای زبان‌های مختلف انجام شده است اما برای زبان فارسی پژوهشی صورت نگرفته است. در این مقاله برای اولین بار، بازنشاسی واژگان فارسی با استفاده از الکترومایوگرام عضلات چهره انجام پذیرفت. بدین منظور، سیگنال‌های sEMG از هشت عضله چهره شش داوطلب هنگام بیان دوازده کلمه زبان فارسی جمع‌آوری شد. سپس ویژگی‌های VAR ، MFL ، $DAMV$ ، $LTKE$ ، IQR و $Cardinality$ از هر کانال و هر پنجره از سیگنال استخراج گردید و ۴۳۲ ویژگی حاصل از هر سیگنال با استفاده از روش تحلیل مؤلفه اصلی به ۴۹ ویژگی تقلیل یافت. در نهایت به منظور بازنشاسی دوازده کلمه زبان فارسی، ویژگی‌ها به طبقه‌بندهای SVM ، KNN و RF داده شد. میانگین صحت طبقه‌بندی به ترتیب ۸۳/۱۶٪، ۸۱/۹۱٪ و ۷۸/۹۷٪ به دست آمد. ارزیابی نتایج این مقاله گویای آن است که با استفاده از سیگنال‌های EMG می‌توان کلمات محدود زبان فارسی را با صحت خوبی بازنشاسی نمود.

الکترومایوگرام چهره
کلمات فارسی
رابط گفتاری بی‌صدا
بازنشاسی کلمات

*نویسنده‌ی مسئول

گروه برق، دانشکده برق و کامپیوتر، دانشگاه سمنان، سمنان، ایران

۰۲۳۳۱۵۳۲۶۹۷

تلفن

۳۵۱۳۱-۱۹۱۱۱

نشانی
کد پستی

۰۲۳۳۱۵۳۲۶۹۷

دورنگار

H_soltanizadeh@semnan.ac.ir

پست الکترونیک

۱- مقدمه

گفتار طبیعی‌ترین وسیله ارتباطی بین انسان‌ها است و به طور فزاینده‌ای در دستگاه‌های امروزی برای تعامل استفاده می‌شود. بنابراین قابلیت گفتار برای هر انسانی و برای هر جامعه‌ای از اهمیت فوق‌العاده‌ای برخوردار است. با این حال، گفتار به طور معمول باید واضح، شفاف و قابل شنیدن باشد، خصوصاً اگر قرار است از طریق دستگاه‌هایی مانند تلفن همراه یا یک تشخیص دهنده 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. اول، شرایطی که فرد در یک محیط شلوغ مثل کتابخانه، رستوران، مترو، اتوبوس یا جلسه عمومی می‌خواهد ارتباط گفتاری برقرار کند. دوم، شرایطی که نیاز به برقراری ارتباطی محرمانه مثل وارد کردن رمز صوتی برای یک سیستم است. سوم، شرایطی که افراد دچار معلولیت گفتاری مثل افراد دچار به سرطان حنجره قصد برقراری ارتباط صوتی با افراد دیگر یا دستیارهای صوتی را دارند. بنابراین اهمیت توسعه روش‌هایی جایگزین برای شرایط بروز اختلال گفتاری به طور فزاینده‌ای افزایش می‌یابد.

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

رابط گفتار بی‌صدا^۶ 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 می‌گیرد. از آنجایی 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 نیستند 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 پنج کلمه زبان انگلیسی پرداخته‌اند. سیگنال‌های الکترومایوگرام با استفاده از هفت الکتروود و از سه عضله صورت ثبت شده است. ویژگی‌های تبدیل فوریه زمان کوتاه از سیگنال‌ها استخراج شده است. روش‌های تحلیل مؤلفه‌های اصلی^۷ (PCA) و تعبیه خطی محلی^۸ (LLE) به منظور کاهش ابعاد ویژگی‌ها استفاده و با یکدیگر مقایسه شده است. با استفاده از الگوریتم طبقه‌بند k نزدیک‌ترین همسایه به صحت تشخیص ۷۴/۶٪ برای مجموعه کلمات پنج تایی دست پیدا کرده است.

وانگا و همکارانش در سال ۲۰۲۱ [۱۳]، به بازشناسی نه کلمه چینی با استفاده از عضلات چهره پرداخته‌اند. تمام داده‌های EMG با استفاده از ده الکتروود و از شش ماهیچه صورت با فرکانس نمونه‌برداری ۱۰۰۰ هرتز از چهار داوطلب جوان سالم در حالی که کلمات: man, kuai, you, zuo, hou, qian, er, yi, ting را تلفظ می‌کردند ثبت شده است. به منظور استخراج ویژگی خودکار، تصاویر طیف هر سیگنال به مدل یادگیری عمیق از پیش آموزش دیده ResNet18, MobileNet-V1 و Xception داده شده است. در نهایت ویژگی‌های استخراج شده به طبقه‌بندهای حافظه کوتاه مدت دوطرفه^۹ (Bi-LSTM) و واحد بازگشتی گیت^{۱۰} (GRU) داده شده و صحت ۸۷ درصد حاصل شده است.

چاندراشک در سال ۲۰۲۱ [۱۴]، به تشخیص و بازشناسی شش حرف انگلیسی A, B, E, I, O, U مبتنی بر سیگنال‌های الکترومایوگرام پرداخته است. سیگنال‌های الکترومایوگرام با استفاده از سه الکتروود قرار گرفته در زیر چانه با فرکانس نمونه‌برداری ۲۰۰۰ هرتز از یک داوطلب سالم ثبت شده است. به منظور استخراج ویژگی، تصاویر مربوط به طیف سیگنال‌ها به

^۶ Silent speech interface^۷ Principal component analysis^۸ Locally Linear Embedding^۹ Bidirectional long-short term memory^{۱۰} Gated recurrent unit^۱ Augmentative and alternative communication^۲ vocal tract^۳ Electromagnetic articulography^۴ Surface electromyography^۵ Electroencephalogram



سیگنال‌های الکترومایوگرام به منظور بازشناسی ۱۱ کلمه تایلندی شامل اعداد صفر تا ده انجام دادند. در ابتدا، سیگنال‌های الکترومایوگرام از پنج موقعیت ماهیچه‌های صورت و گردن ثبت شده‌اند در حالی که ده شرکت کننده کلمات را در دو حالت با صدا و بی‌صدا بیان کردند. ۲۲ ویژگی از سیگنال‌های الکترومایوگرام استخراج شده است و ویژگی‌های مناسب جهت طبقه‌بندی انتخاب شده است. نتایج نشان می‌دهند که طبقه‌بند خطی حداقل مربعات فیشر^{۱۲} (FLDA) و طبقه‌بند نرمال بیز^{۱۳} (LBN) بهترین نتیجه را به دست آوردند. میانگین صحت به ترتیب برای حالت صدادار و بی‌صدا برابر با ۰/۹۰/۰۱ و ۰/۷۹/۱۸ بوده است.

جونگ و همکارانش در سال ۲۰۱۹ [۱۸]، یک سیستم تشخیص گفتار مبتنی بر سیگنال‌های الکترومایوگرام را با استفاده از پنج کانال به منظور طبقه‌بندی نه هجای تایلندی توسعه دادند. شش ویژگی از سه دسته ویژگی، یعنی میانگین قدر مطلق^{۱۴} (MAV) و طول شکل موج^{۱۵} (WL) از ویژگی‌های مبتنی بر دامنه، نرخ عبور از صفر^{۱۶} (ZC) و میانگین فرکانس^{۱۷} (MNF) از ویژگی‌های مبتنی بر فرکانس، و کشیدگی^{۱۸} و چولگی^{۱۹} از ویژگی‌های مبتنی بر آمار استخراج شدند. پس از آن، یک ماشین یادگیری افراطی رگرسیون طیفی^{۲۰} (SRELM) برای کاهش ابعاد ویژگی‌ها از ۳۰ به ۸ مورد استفاده قرار گرفت. در نهایت، ویژگی‌های کاهش یافته با استفاده از طبقه‌بند شبکه عصبی پیش‌خور با اعتبارسنجی ۵-fold طبقه‌بندی شدند. میانگین صحت طبقه‌بندی به دست آمده در داوطلبان سالم و دیزارتیک به ترتیب ۰/۹۴/۵ و ۰/۸۹/۴ به دست آمده است.

مینگزینگ و همکارانش در سال ۲۰۲۱ [۱۹]، به بررسی مشارکت عضلات مختلف به منظور تشخیص ده کلمه انگلیسی و چینی با استفاده از سیگنال‌های EMG پرداختند. از ۱۲۰ الکتروود با فاصله نزدیک بر روی عضلات چهره و گردن از هجده داوطلب بومی چینی استفاده شده است. الگوریتم مورد استفاده جهت طبقه‌بندی تحلیل تفکیک کننده خطی است، و الگوریتم انتخاب رو به جلو متوالی برای تجزیه و تحلیل موقعیت بهینه حسگرها استفاده شده است. نتایج نشان می‌دهد که توپوگرافی انرژی الکترومایوگرام می‌تواند به تجسم فعالیت‌های ماهیچه‌ای پویا در طول فرآیند صحبت کردن کمک کند، و الگوهای انقباض ماهیچه‌ای به طور قابل توجهی برای کارهای مختلف

یک شبکه عمیق گوگل‌نت از پیش آموزش دیده، داده شده است. در نهایت سیگنال‌ها با استفاده از هشت طبقه‌بند متفاوت مورد ارزیابی قرار گرفته است. ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد برای آزمون در نظر گرفته شده است. نتایج نهایی نشان می‌دهد که بهترین نرخ تشخیص مربوط به طبقه‌بند ماشین بردار پشتیبان با صحت ۰/۸۰/۱۰ است.

یانگ و همکارانش در سال ۲۰۲۰ [۱۵]، به بازشناسی ده کلمه چینی مبتنی بر توابع درون‌یابی پرداخته‌اند. داده‌ها با استفاده از پنج کانال و هفت الکتروود، از ماهیچه صورت دو داوطلب آقا در حالی که کلمات: you, yi, wu, shi, ren, men, le, de, bu, zai را تلفظ می‌کردند ثبت شده است. سیگنال‌ها بر اساس دو مدل مورد ارزیابی قرار گرفته است. در مدل اول سیگنال‌ها پنجره‌گذاری شد و از هر پنجره ویژگی‌ها استخراج شد و در مدل دوم ویژگی‌ها از کل سیگنال و بدون پنجره‌گذاری استخراج شده است. ویژگی‌ها از دو حوزه زمان و زمان فرکانس استخراج شده است. در نهایت پس از کاهش ابعاد بردار ویژگی‌ها، سیگنال‌ها با استفاده از سه الگوریتم k نزدیک‌ترین همسایه و جنگل تصادفی و مدل مخفی مارکوف طبقه‌بندی شده است. بهترین نتیجه مربوط به طبقه‌بند مارکوف با صحت ۰/۸۵ بوده است.

هانگ و همکارانش در سال ۲۰۲۰ [۱۶]، با استفاده از شبکه‌های عصبی حافظه‌دار دوطرفه به بازشناسی چهار کلمه چینی به منظور کنترل ربات پرداختند. سیگنال‌ها با استفاده از ده الکتروود با فرکانس نمونه‌برداری ۱۰۰۰ هرتز از هفت داوطلب در حالی که کلمات: man, kuai, yi, er را تلفظ می‌کردند ثبت شده است. استخراج ویژگی با استفاده از تبدیل فوریه زمان کوتاه و شبکه CNN انجام گرفته است. سپس ویژگی‌های استخراج شده به دو بخش آموزش (۷۰ درصد) و آزمایش (۳۰ درصد) تقسیم شده است و جهت تشخیص کلمه مورد نظر به طبقه‌بند BLSTM داده شده است. به منظور ارزیابی روش پیشنهادی نتایج با الگوریتم‌های LSTM، CNN، جنگل تصادفی^{۱۱} مقایسه شده است. نتایج نشان می‌دهد که بهترین نرخ بازشناسی مربوط به طبقه‌بند BLSTM با صحت ۰/۹۷/۱۱ است.

سریسووان و همکارانش در سال ۲۰۲۰ [۱۷]، یک مقایسه از عملکرد هشت طبقه‌بند برای تشخیص گفتار مبتنی بر

^{۱۶} Zero crossing

^{۱۷} Mean frequency

^{۱۸} Kurtosis

^{۱۹} Skewness

^{۲۰} Spectral regression extreme learning machine

^{۱۱} Random forest

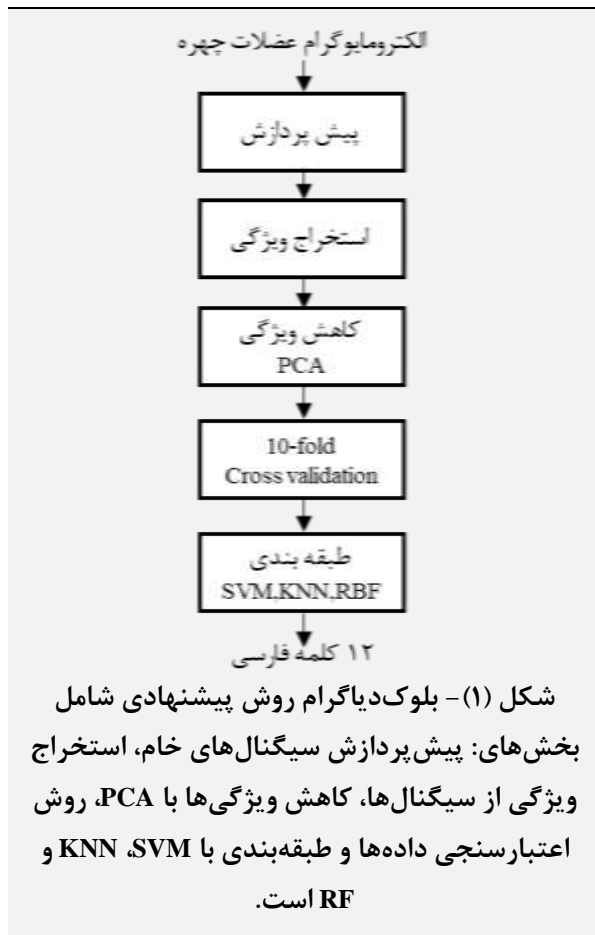
^{۱۲} Fishers linear discriminant analysis

^{۱۳} Linear Bayes normal

^{۱۴} Mean absolute value

^{۱۵} Waveform length

سپس پیش پردازش‌های لازم جهت حذف نویز و آرتیفکت‌های حرکتی بر روی تمامی سیگنال‌ها اعمال شد. در مرحله بعد از هر سیگنال ویژگی‌های مناسب استخراج شد و بعد با استفاده از الگوریتم تحلیل مؤلفه اصلی ویژگی‌ها تقلیل یافت. در نهایت با روش اعتبارسنجی 10-fold و ماشین بردار پشتیبان^{۲۱}، نزدیک‌ترین همسایه k^{۲۲} و جنگل تصادفی^{۲۳} دوازده کلمه فارسی طبقه‌بندی شد.



۲-۱- ثبت سیگنال الکترومایوگرام سطحی از

عضلات صورت

با توجه به اینکه هیچ پایگاه داده سیگنال EMG حاصل از عضلات صورت شخص در حین گفتار کلمات فارسی وجود ندارد اقدام به ثبت سیگنال و ایجاد پایگاه داده از عضلات صورت داوطلب‌ها در حین گفتار نمودیم. سیگنال‌های ثبت و جمع‌آوری شده در این مقاله از سی داوطلب که شامل پانزده مرد در بازه سنی ۲۰-۴۵ سال و پانزده خانم با بازه سنی ۲۰-۳۴ می‌باشد. مشارکت‌کنندگان هیچ‌گونه سابقه بیماری یا اختلال اسکلتی عضلانی مزمن نداشته و قبل از ثبت سیگنال رضایت خود را به صورت کتبی برای شرکت در این تحقیق از

صحت کردن مشاهده شود. استفاده از تعداد الکتروود برابر در دو ناحیه گردن و صورت 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 شده‌اند، نرخ طبقه‌بندی می‌تواند به بالاتر از ۹۰ درصد برسد. سیونگ چا و همکارانش در سال ۲۰۲۲ [۲۰]، یک سیستم تشخیص گفتار بی‌صدا به منظور بازشناسی شش کلمه کره‌ای بر اساس EMG ثبت شده در اطراف چشم برای بیست داوطلب پیاده‌سازی کردند. شش الکتروود EMG بر روی یک فریم پلاستیکی با شکلی مشابه با پد به یک هدست VR تجاری متصل شده است. ویژگی ریشه میانگین مربعات (RMS)، با استفاده از یک پنجره زمانی لغزشی ۰/۱ ثانیه بدون همپوشانی از سیگنال‌ها استخراج شده است. با استفاده از مدل یادگیری عمیق بهینه شده (me-bLSTM)، شش کلمه کره‌ای بی‌صدا با معادل انگلیسی Home, Menu, Escape, Previous, Next, Select با صحت بالای ۹۲/۵۳ درصد طبقه‌بندی شده است. روش این مقاله تشخیص گفتار بی‌صدا توسط sEMG است، جایی که پتانسیل‌های الکتریکی عضلات چهره کاربر توسط الکتروودهای سطحی ثبت می‌شود. این امر امکان پردازش گفتار را حتی در صورت عدم تولید سیگنال صوتی امکان‌پذیر می‌کند که می‌تواند روش ویژه‌ای در مکان‌های شلوغ با صداهای نویزی باشد [۲۰].

توسعه یک سیستم ارتباط گفتاری از طریق سیگنال‌های الکترومایوگرام مبتنی بر زبان فارسی می‌تواند در زمینه‌های متعددی مفید واقع شود. این مقاله برای اولین بار به بازشناسی گفتار زبان فارسی مبتنی بر الکترومایوگرام چهره می‌پردازد. همچنین در کارهای انجام شده در این حوزه، این مطالعه دارای بیشترین تعداد داوطلب با توزیع جنسیتی برابر است. ساختار این مقاله به شرح زیر است: توضیح در مورد نحوه ثبت سیگنال، ایجاد پایگاه داده، استخراج ویژگی، کاهش ویژگی‌ها و طبقه‌بندی در بخش دوم ارائه شده است. سپس نتایج حاصل از طبقه‌بندی کلمات زبان فارسی و مقایسه با پژوهش‌های دیگران برای زبان‌های دیگر در بخش سوم بیان شده است. در نهایت یافته‌ها و بحث و نتیجه‌گیری در بخش چهارم و پنجم مطرح شده است.

۲- مواد و روش‌ها

شمای کلی از روش پیشنهادی در شکل (۱) نشان داده شده است. در ابتدا سیگنال‌های عضلات چهره هر داوطلب ثبت شد.

^{۲۳} Random forest

^{۲۱} Support vector machine

^{۲۲} K nearest neighbor



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

به منظور ثبت سیگنال از داوطلب‌ها، یک برنامه اندروید طراحی شد تا داوطلب‌ها در زمان‌های مقرر شده هر کلمه‌ای که بر روی برنامه به نمایش در می‌آید را بیان کنند. کلمات با فاصله شش ثانیه از هم بیان می‌شوند. پس از بیان هر مجموعه‌ی ۱۲ کلمه‌ای، بدون آنکه ثبت متوقف شود به داوطلب ۱۰ ثانیه استراحت داده شد و مجدد مطابق قبل، دوازده کلمه با فواصل شش ثانیه بیان می‌شوند. این روند برای هر داوطلب ده بار تکرار شد.

در نهایت از هر داوطلب ۱۲۰ کلمه شامل دوازده کلاس و ده تکرار خواهیم داشت (۱۲۰ = ۱۰ × ۱۲). تعداد کل داده‌ها برای داوطلبین برابر با ۳۶۰۰ کلمه شامل دوازده کلاس، ده تکرار و سی داوطلب خواهد بود (۳۶۰۰ = ۱۰ × ۱۲ × ۳۰). الگواره ثبت سیگنال برای یک داوطلب در شکل (۳) نمایش داده شده است. دوازده کلمه شامل ماه‌های سال به زبان و تقویم فارسی به شرح زیر است: فروردین، اردیبهشت، خرداد، تیر، مرداد، شهریور، مهر، آبان، آذر، دی، بهمن و اسفند. در جدول (۱) کلمات به همراه تلفظ آن نمایش داده شده است.

جدول (۱) - معرفی و تلفظ دوازده کلمه فارسی

شماره کلاس	کلمه فارسی	تلفظ
۱	فروردین	/farvardin/
۲	اردیبهشت	/'ordibehešt/
۳	خرداد	/xordād/
۴	تیر	/tir/
۵	مرداد	/mordād/

طریق پرسش‌نامه مورد تایید توسط کمیته اخلاق اعلام نموده‌اند. این مطالعه توسط کمیته اخلاق دانشگاه علوم پزشکی سمنان با شناسه IR.SEMUMS.REC.1401.103 بررسی و تایید شده است. داوطلبین بر روی صندلی نشسته و موقعیت خود را ثابت کرده تا از اثرات نامطلوب حرکت‌های ناخواسته^{۲۴} بر سیگنال‌های الکترومایوگرام (آرتیفکت حرکتی) جلوگیری شود. برای ثبت سیگنال الکترومایوگرام سطحی چهره، از دستگاه ثبت سیگنال الکترومایوگرام DataLog MWX8 Biometric، با هشت کانال و الکتروودهای فعال استفاده شد. قبل از اتصال الکتروودها، به منظور کاهش امپدانس پوست-الکتروود، محل الکتروودگذاری با الکل تمیز و سپس الکتروودها با آرایش دوقطبی قرار گرفتند. الکتروود مرجع^{۲۵} نیز در ناحیه استخوانی پشت گوش نصب گردید. الکتروود اول بر روی عضله ریسوریوس^{۲۶}، الکتروود دوم روی عضله دپرسور آنگولی اوریس^{۲۷}، الکتروود سوم روی عضله پلاتیسم^{۲۸}، الکتروود چهارم روی عضله منتالیس^{۲۹}، الکتروود پنجم روی عضله بوکسیناتور^{۳۰}، الکتروود ششم روی عضله لویتورلیسی سوپریور^{۳۱}، الکتروود هفتم روی عضله زیگوماتیکوس مینور و ماژور^{۳۲} و الکتروود هشتم روی عضله جوشی^{۳۳} قرار گرفته است. موقعیت قرارگیری الکتروودها و نمای ظاهری دستگاه ثبت سیگنال در شکل (۲) نشان داده شده است.

سیگنال‌های الکترومایوگرام تحت تأثیر نویز برق شهر و آرتیفکت حرکتی قرار می‌گیرند. عمده انرژی طیف فرکانسی سیگنال EMG خام در محدوده ۱۰ تا ۲۵۰ هرتز است. [۲۱]. همچنین محدوده 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 بر روی تمام سیگنال‌ها تضعیف شد. همچنین برای حذف آرتیفکت حرکتی، یک فیلتر بالاگذر باترورث مرتبه شش با فرکانس قطع ۱۰ هرتز بر روی تمامی کانال‌ها اعمال گردید [۲۱، ۲۲].

^{۲۹} Mentalis

^{۳۰} Buccinator

^{۳۱} Levator labii superioris

^{۳۲} Zygomaticus minor and major

^{۳۳} Masseter

^{۲۴} Artifact

^{۲۵} Reference Electrode

^{۲۶} Risorius

^{۲۷} Depressor anguli oris

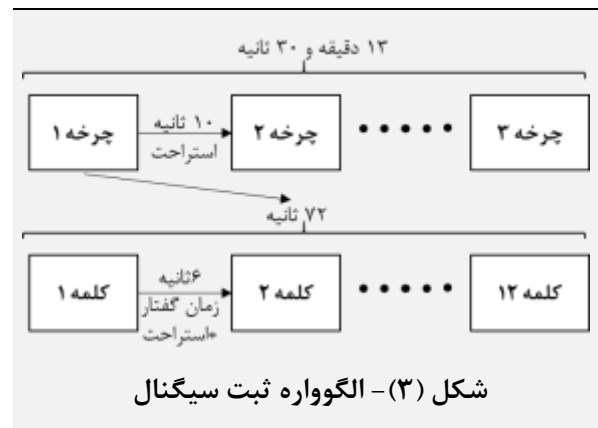
^{۲۸} Platysma

حداکثر طول فراکتال^{۲۴} (MFL)، واریانس الکترومایوگرام^{۲۵}، تفاضل میانگین مقدار مطلق^{۲۶} (DAMV)، لگاریتم عملگر انرژی Teager-Kaiser (LTKEO)، کاردینالیتی^{۲۷} (CARD) و محدوده بین چارکی^{۲۸} (IQR) استفاده شد که در جدول (۲) به همراه روابط نمایش داده شده است. در روابط مربوط به ویژگی‌ها پارامتر x بیانگر هر نمونه در یک پنجره از سیگنال و i بیانگر شماره آن نمونه است، پارامتر N بیانگر طول پنجره یا تعداد کل نمونه‌ها در آن پنجره از سیگنال می‌باشد. مقدار آستانه در ویژگی پنجم برابر است با حاصل ضرب R در مجذور میانگین مربعات سیگنال EMG در حالت استراحت (بدون انقباض) به دست می‌آید که با رابطه (۱) تعریف می‌شود [۲۸].

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

در رابطه (۱) r_i نمونه‌های سیگنال در حالت استراحت و N تعداد نمونه‌ها می‌باشد. در [۲۸] مقدار R از ۰ تا ۶ پیشنهاد شده است. در کار ما مقدار آستانه برابر 0.1 میلی‌ولت و مقدار R برابر با 0.6 تنظیم شده است. همچنین در ویژگی ششم Q_1 معادل با چارک ۲۵ و Q_3 معادل با چارک ۷۵ می‌باشد. جهت محاسبه ویژگی ششم هر پنجره به طول ۱۲۸ نمونه در ابتدا از کوچک به بزرگ مرتب می‌شود و سپس چارک ۱۷۵ام از چارک ۱۲۵ام کسر می‌شود.

۶	شهریور	/sahrivar/
۷	مهر	/mehr/
۸	آبان	/'ābān/
۹	آذر	/'āzar/
۱۰	دی	/dey/
۱۱	بهمن	/bahman/
۱۲	اسفند	/'esfand/



۲-۲ استخراج ویژگی

برای استخراج ویژگی، ابتدا هر سیگنال پنجره‌گذاری می‌شود و ویژگی‌ها برای هر پنجره لغزنده به طول ۱۲۸ نمونه و با همپوشانی ۲۵ درصد محاسبه می‌شود [۲۳، ۲۴]. هر کانال از هر سیگنال به نه پنجره تقسیم شد. در این مقاله از ویژگی‌های

جدول (۲) - روابط و اسامی ویژگی‌های استخراج شده

شماره	نام ویژگی‌ها	عبارت 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	کاردینالیتی	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	محدوده بین چارکی	$IQR = Q_3 - Q_1$	[۲۹]

^{۲۷} Cardinality

^{۲۸} Interquartile Range

^{۲۴} Maximum fractal length

^{۲۵} Variance of EMG

^{۲۶} Difference Absolute Mean Value

ویژگی‌های استخراج شده به طبقه‌بند داده می‌شود تا عملیات آموزش و آزمایش انجام گیرد و بتوان با اعمال داده‌های آزمایش به طبقه‌بند پیش‌بینی کرد که هر سیگنال مربوط به کدام کلاس می‌باشد. بدین منظور از طبقه‌بندهای SVM، KNN و RF استفاده شد. به منظور اعتبارسنجی از روش 10-fold استفاده شد که در این نوع اعتبارسنجی، داده‌ها به K زیرمجموعه افراز می‌شوند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و K-1 تای دیگر برای آموزش به کار می‌روند. این روال K بار تکرار می‌شود و همه داده‌ها دقیقاً یک‌بار برای اعتبارسنجی به کار می‌روند. تعداد داده‌های آموزش و آزمایش برای هر fold به ترتیب برابر با ۳۲۴۰ و ۳۶۰ می‌باشد.

ماشین بردار پشتیبان یک طبقه‌بند دوتایی است. اما این امکان است تا با روش‌های متفاوتی از این کلاس‌بند به منظور حل مسائل چند کلاسه هم استفاده کرد. در این مقاله از روش یک کلاس در برابر همه استفاده شد [۳۰]. در استفاده از الگوریتم SVM تعیین کرنل مناسب که وظیفه نگاشت مناسب داده‌ها در فضای ویژگی‌ها را بر عهده دارد حائز اهمیت است. بدین منظور ابتدا الگوریتم را با استفاده از سه کرنل پایه شعاعی^{۳۹}، چندجمله‌ای^{۴۰} و خطی^{۴۱} مورد آزمایش قرار دادیم. روابط ریاضی هر یک از کرنل‌ها در ادامه شرح داده شده است. از آنجا که کرنل چند جمله‌ای دارای نتایج بهتری بود کلیه نتایج بر اساس این کرنل ارائه شده است.

تابع کرنل خطی در رابطه (۳)، تابع کرنل چند جمله‌ای در رابطه (۴) و تابع کرنل پایه شعاعی در رابطه (۵) نشان داده شده است [۳۰].

$$K(X_i, X_j) = 1 + X_i^T X_j \quad (3)$$

$$K(X_i, X_j) = (1 + X_i^T X_j)^p \quad (4)$$

$$K(X_i, X_j) = \exp\left(-\frac{\|X_i - X_j\|^2}{2\sigma^2}\right) \quad (5)$$

که X ، بردار داده‌ها است و p درجه‌ی چند جمله‌ای است. کرنل خطی ساده‌ترین نوع کرنل می‌باشد و برای داده‌های جدایی‌پذیر خطی، عملکرد خوبی نشان می‌دهد. عملکرد کرنل چند جمله‌ای به درجه‌ی p چند جمله‌ای وابسته است؛ زیرا اینکه داده‌ها چقدر جدایی‌پذیر هستند، به p بستگی دارد [۳۰]. در این مقاله، عملکرد چند جمله‌ای برای درجه ۴ ارزیابی شده است. به منظور طبقه‌بندی با استفاده از KNN معیار فاصله، اقلیدسی و تعداد همسایه برابر با ۱۳ تنظیم شد.

با توجه به اینکه از هر سیگنال الکترومایوگرام شش ویژگی استخراج شد، طول هر بردار ویژگی برابر با ۴۳۲ خواهد شد که برابر است با حاصل ضرب شش ویژگی و نه پنجره و هشت کانال. ویژگی‌ها در مجموعه داده ممکن است شامل مقادیری باشد که در محدوده متفاوتی قرار دارند. این ویژگی‌ها با مقادیر بزرگ ممکن است اثر بسیار زیادتری در تابع هزینه نسبت به ویژگی‌های با مقادیر کم داشته باشند. این مشکل با نرمالیزه نمودن ویژگی‌ها طوری که مقادیر آن‌ها در محدوده‌های مشابه قرار گیرند برطرف خواهد شد. لذا در این مقاله ویژگی‌های استخراج شده به روش Min-Max و با رابطه (۲) نرمالیزه شدند. پارامتر f_i برابر با ویژگی که قرار است نرمالیزه شود، پارامتر $f_{i\min}$ برابر با کمترین مقدار ویژگی نام و پارامتر $f_{i\max}$ برابر با بیشترین مقدار ویژگی نام می‌باشد [۲۷].

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

در نهایت ماتریس مجموعه داده‌های ویژگی برای آزمایش و آموزش به ترتیب برابر با ۳۶۰×۴۳۲ و ۳۲۴۰×۴۳۲ شد.

۲-۳- کاهش ویژگی‌ها با استفاده از تحلیل مؤلفه‌های اصلی

تحلیل مؤلفه اصلی به وسیله‌ی دنباله‌هایی از ترکیبات خطی متعامد از متغیرهای اولیه با نام مؤلفه‌های اصلی که واریانس ماکزیمم دارند به دست می‌آید [۲۹]. تحلیل مؤلفه اصلی، روشی برای استخراج مؤلفه‌های اصلی از مجموعه بزرگ ویژگی‌های استخراج شده است. در واقع یک مجموعه با بُعد پایین از مؤلفه‌ها از یک مجموعه دارای بُعد بالا از ویژگی‌ها استخراج می‌شود. اولین مؤلفه، یک ترکیب خطی از ویژگی‌ها است که جهت بیشترین تغییرات در داده‌ها را تعیین می‌کند. مؤلفه‌های بعدی نیز از مفهومی مشابه آنچه بیان شد، پیروی می‌کنند. به عبارت دیگر، آن‌ها مقدار واریانس باقیمانده را بدون آنکه با مؤلفه‌های پیشین دارای همبستگی شوند، در خود حفظ می‌کنند. [۲۹].

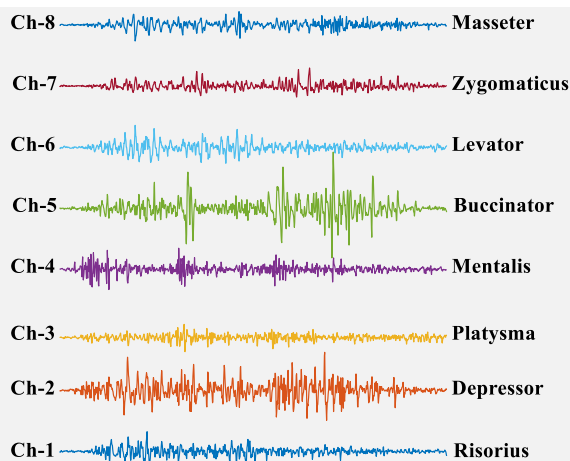
قبل از مرحله طبقه‌بندی، ویژگی‌ها با استفاده از PCA کاهش یافت. نتایج بر اساس تعداد مؤلفه‌های اصلی از یک مؤلفه تا شصت مؤلفه مورد ارزیابی و بررسی قرار گرفت که در بخش ۳- نتایج در شکل (۵) گزارش گردید.

۲-۴- طبقه‌بندی

⁴¹ Linear

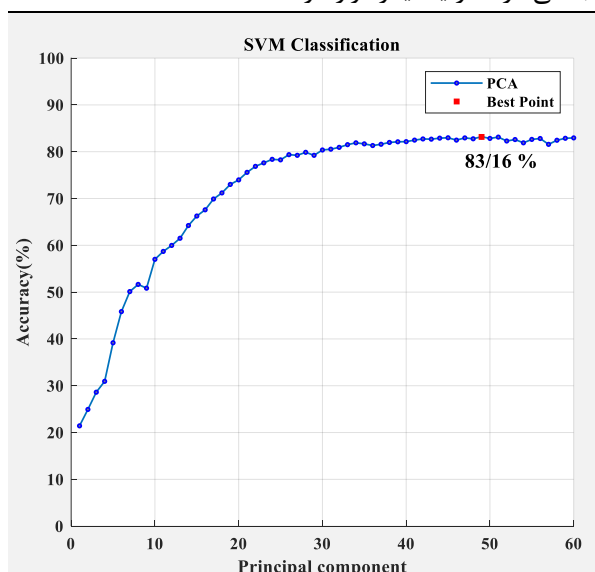
³⁹ Radial basis function

⁴⁰ Polynomial



شکل (۴) - نمونه سیگنال الکترومایوگرام چهره در حین بیان کلمه فروردین، برای هشت کانال ثبت شده از هشت عضله چهره که شماره کانال در سمت چپ و نام عضله در سمت راست نشان داده شده است.

برداری شامل ۴۳۲ ویژگی با استفاده از الگوریتم تحلیل مؤلفه‌های اصلی به ۴۹ مؤلفه اصلی تقلیل یافت تا صحت بازشناسی کلمه بر حسب تعداد مؤلفه‌های اصلی ارزیابی شود. مؤلفه‌های آغازین دارای اطلاعات بیشتر و موثرتری هستند ولی به تنهایی کارآمدی لازم را ندارند. به طور مثال در شکل (۵) مؤلفه اصلی اول به تنهایی دارای صحت ۲۱ درصد در بازشناسی ۱۲ کلمه است. به دلیل ناهمبسته بودن مؤلفه‌های اصلی با یکدیگر، وقتی به طور همزمان و تجمعی از مؤلفه‌ها استفاده می‌شود صحت بازشناسی افزایش می‌یابد. به همین منظور به ترتیب، مؤلفه‌های اصلی از پایین‌ترین به بالاترین به صورت تجمعی در کنار یکدیگر قرار گرفتند.



۲-۵- معیارهای کمی ارزیابی

عملکرد طبقه‌بند با استفاده از معیارهای کمی صحت^{۴۲} و حساسیت^{۴۳} مورد ارزیابی قرار گرفت. این معیارها به ترتیب در روابط (۶) و (۷) معرفی شده است [۳۱]:

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

$$Sensitivity_i = \frac{TP_i}{(TP_i + FN_i)} \quad (۷)$$

پارامتر صحت، اساسی‌ترین معیار اندازه‌گیری کیفیت یک طبقه‌بند است و برابر با تعداد نمونه‌های درست پیش‌بینی شده تقسیم بر تعداد کل نمونه‌ها است. معیار حساسیت که آن را «نرخ پاسخ‌های مثبت درست»^{۴۴} نیز می‌گویند به معنی نسبی از موارد درست در هر کلاس است 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 رابطه (۷) به دست می‌آید. TP_i نشانگر مثبت واقعی است یعنی تعداد نمونه‌هایی که عضو کلاس i هستند و عضو کلاس i تشخیص داده شده‌اند، FN_i نشانگر منفی کاذب^{۴۵} است یعنی نمونه عضو کلاس i باشد اما به اشتباه کلاس دیگری تشخیص داده شود. در رابطه (۷) i نشانگر شماره کلاس‌ها است. دوازده مقدار حساسیت برای هر کلاس به دست آمد. مقادیر حساسیت مربوط به هر کلاس در سطر اول ماتریس اغتشاش در شکل (۶) و (۷) گزارش شد. میانگین مقدار حساسیت برای هر دوازده کلاس در جدول (۳) گزارش گردید. [۳۱، ۳۲].

۳- نتایج

در شکل (۴) یک نمونه از سیگنال متعلق به کلمه فروردین نمایش داده شد. محور افقی نشان دهنده زمان است (۰/۹ ثانیه). تغییرات دامنه سیگنال بین منفی ۲ میلی‌ولت تا ۲ میلی‌ولت می‌باشد. هر یک از سیگنال‌ها که در یک سطر نمایش داده شده نشان دهنده یک کانال است که توسط یک الکتروود از روی یک نقطه از چهره، ثبت شده است.

^{۴۴} True Positive Rate

^{۴۵} False Negative

^{۴۲} Accuracy

^{۴۳} Sensitivity



شکل (۵) - صحت بازشناسی کلمه بر حسب تعداد مؤلفه‌های اصلی الگوریتم PCA، مؤلفه‌ها به صورت تجمعی در کنار یکدیگر قرار گرفته‌اند.

ملاحظه شد که بعد از مؤلفه ۴۰ که بیانگر قرار گرفتن ۴۰ مؤلفه اصلی اول در کنار یکدیگر است، صحت به طور معناداری افزایش پیدا نمی‌کند. به این دلیل که مؤلفه‌های ۴۰ به بعد دارای اطلاعات کم یا کم اثر هستند با وجود این، جهت اطمینان تا مؤلفه ۶۰ مورد بررسی قرار گرفت. میزان صحت بازشناسی کلمه برای دوازده کلمه فارسی با رسیدن به ۴۹ مؤلفه به بالاترین مقدار ۸۳/۱۶ درصد رسید. لذا تعداد مؤلفه‌های اصلی مورد استفاده در ادامه ۴۹ است.

در شکل (۶) ماتریس اغتشاش حاصل از طبقه‌بندی دوازده کلاس با SVM بدون کاهش ابعاد گزارش شده است که در آن محور افقی نشان دهنده کلاس‌های هدف و محور عمودی نشان دهنده کلاس‌های شناسایی شده می‌باشد و همچنین قطر اصلی ماتریس اغتشاش نشان دهنده تعداد داده‌های درست تشخیص داده شده می‌باشد. مشاهده می‌شود که کمترین تعداد داده درست تشخیص داده شده مربوط به شناسایی کلاس ده یعنی کلمه دی با ۱۹۲ مورد می‌باشد که ۶۹ مورد از داده‌های آزمایش

به اشتباه کلاس چهارم (تیر) تشخیص داده شده است. از طرفی کلاس چهارم کلمه (تیر) بعد از کلمه دی دارای کمترین تعداد داده درست تشخیص داده شده با ۲۱۵ مورد می‌باشد که بیشترین میزان خطای آن ۴۸ مورد می‌باشد که به اشتباه کلاس ده یعنی کلمه (دی) شناسایی شده است. از طرفی کلاس‌های دو (اردیبهشت)، یک (فروردین)، هشت (آبان) و شش (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 با استفاده از الگوریتم PCA و کاهش ابعاد از ۴۳۲ ویژگی به ۴۹ مؤلفه صحت به اندازه ۰/۹۱ درصد افزایش یافته است. تحلیل مؤلفه‌های اصلی با کاهش دادن ابعاد داده‌ها موجب سرعت بخشیدن به محاسبات می‌شود. علاوه، زمانی که ابعاد بالایی با متغیرهای همبسته داشته باشیم، PCA می‌تواند با حذف ویژگی‌های افزونه (ویژگی‌های وابسته) و داده‌های نویزی و نامربوط، صحت طبقه‌بندی را بهبود بخشد. ساده‌سازی فضای ویژگی و کاهش ابعاد فضا موجب می‌گردد 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].



SVM without PCA Confusion Matrix

Output Class	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%	33.3%	64.0%	35.0%	36.0%
		9.0%	7.0%	22.7%	28.3%	21.0%	10.3%	23.3%	9.7%	16.7%	36.0%	15.0%	14.0%	
		1	2	3	4	5	6	7	8	9	10	11	12	
		Target Class												

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



SVM+PCA Confusion Matrix

Output Class \ Target Class	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	Accuracy
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% 12.7%
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% 13.6%
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% 22.1%
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% 32.7%
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% 20.1%
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% 8.2%
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% 14.1%
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% 12.5%
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% 10.3%
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% 25.9%
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% 13.2%
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% 15.5%
	89.3% 10.7%	93.0% 7.0%	77.3% 22.7%	74.0% 26.0%	76.7% 23.3%	90.0% 10.0%	77.3% 22.7%	88.3% 11.7%	86.7% 13.3%	69.7% 30.3%	90.0% 10.0%	85.7% 14.3%	83.2% 16.8%

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

دهنده تعداد کلاس و محور عمودی نشان دهنده میزان صحت و هر نقطه رنگی نشان دهنده مقدار صحت گزارش شده در یک مرجع می‌باشد. به منظور صحت‌سنجی مناسب، صحت بازشناسی کلمه برای روش پیشنهادی این مقاله به ازای تعداد کلمات از یک تا دوازده محاسبه گردید تا امکان مقایسه با پژوهش‌های پیشین با تعداد کلمات مختلف فراهم گردد. مشاهده می‌شود که روش پیشنهادی در مقایسه با سایر روش‌ها دارای عملکرد بهتری می‌باشد. فقط در تعداد کلاس ۹ و ۱۰ روش پیشنهادی عملکرد پایین‌تری داشته است که به علت استفاده از تعداد الکتروود بیشتر می‌باشد. 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

در نهایت داده‌ها با استفاده از KNN و RF هم مورد ارزیابی قرار گرفت. نتایج طبقه‌بندی دوازده کلاس با استفاده از سه الگوریتم SVM، KNN و RF بر اساس معیارهای صحت و حساسیت در جدول (۳) گزارش شده 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 نتایج حاصل از این مقاله را با مقالات دیگر با زبان‌های مختلف و بر اساس تعداد کلاس‌های متفاوت مقایسه کرده‌ایم که در شکل (۸) نمایش داده شده است، محور افقی نشان

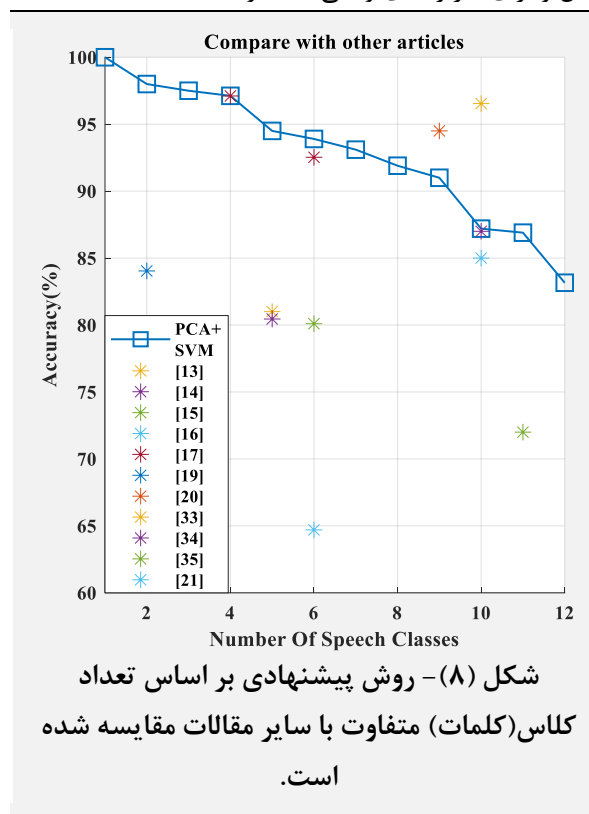
نهایت بردارهای ویژگی برای آموزش و آزمایش به طبقه‌بند اعمال گردید.

با مشاهده نتایج مربوط به ماتریس اغتشاش شکل (۷) مشخص می‌شود که برخی از کلاس‌ها با یکدیگر تداخل دارند به عنوان مثال کلاس‌های ۳ و ۵، که به ترتیب کلمه خرداد و مرداد است با یکدیگر تداخل دارند و کلاس‌های ۴ و ۱۰، که به ترتیب کلمه تیر و دی است به یکدیگر نزدیک هستند و سیستم پیشنهادی به اشتباه داده‌های هر یک از این کلمات را با برچسب دیگری شناسایی می‌کند. با توجه به اینکه کلمه مرداد و خرداد هم از نظر تلفظ و هم از نظر نوع بیان و فعال شدن عضلات چهره مشابه هستند لذا حاصل شدن این نتایج قابل توجیه است. از طرفی کلمات تیر و دی هم نسبت به یکدیگر داده‌های حساسی دارند و سیستم کلمه تیر را به اشتباه کلمه دی، و کلمه دی را به اشتباه کلمه تیر شناسایی می‌کند. برخلاف دو کلمه مرداد و خرداد، کلمات تیر و دی از نظر تلفظ و آوا یکسان نیستند و فقط از نظر نوع بیان و فعالیت عضلانی مشابه یکدیگر هستند و به همین جهت نتایج حاصل از این دو کلمه با یکدیگر تداخل دارند. کلمه اردیبهشت دارای بالاترین صحت شناسایی می‌باشد که علت آن نوع بیان متفاوت آن نسبت به دیگر کلمات می‌باشد و عضلاتی که با بیان این کلمه در چهره 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 روبرو است. در مقایسه با تشخیص گفتار مبتنی بر سیگنال صوتی، قابلیت اطمینان EMG پایین‌تر است و امکان تشخیص تعداد بالای کلمات وجود ندارد. بنابراین، تاکنون، 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 خواهد بود. همچنین دیگر محدودیت 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 خواهد بود. براساس نتایج حاصل از سیستم پیشنهادی می‌توان گفت 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 اینکه تحلیل سیگنال‌های الکترومایوگرام چهره نسبت به سایر عضلات بدن به دلیل نزدیک بودن عضلات به یکدیگر دشوارتر است

الکتروود بیشتر به معنی افزایش هزینه دستگاه خواهد بود که در عمل و برای کاربردهای واقعی، نامطلوب است.



شکل (۸) - روش پیشنهادی بر اساس تعداد کلاس (کلمات) متفاوت با سایر مقالات مقایسه شده است.

۴- یافته‌ها و بحث

در این مقاله، بازشناسی دوازده کلمه رایج از زبان فارسی مبتنی بر سیگنال الکترومایوگرام (بدون استفاده از سیگنال گفتار صوتی) در حوزه زمان و با استفاده از سه طبقه‌بند ماشین بردار پشتیبان، نزدیک‌ترین همسایه K و جنگل تصادفی ارائه شد. از آنجا که تا به امروز بازشناسی کلمات زبان فارسی با استفاده از سیگنال الکترومایوگرام انجام نشده است و روز به روز بیماران مبتلا به سرطان حنجره و یا اختلالات گفتاری رو به افزایش است، لذا می‌تواند این تحقیق در راستای کمک به این بیماران مورد استفاده قرار گیرد.

به منظور اجرای روش پیشنهادی به دلیل آنکه هیچ پایگاه داده‌ای از سیگنال‌های الکترومایوگرام چهره به زبان فارسی وجود نداشت اقدام به ایجاد پایگاه داده‌ای برای این منظور نمودیم. پایگاه داده ایجاد شده شامل سیگنال الکترومایوگرام هشت عضله چهره هنگام بیان دوازده کلمه فارسی می‌باشد. در ابتدا پیش‌پردازش‌های لازم را به منظور حذف نویز بر روی سیگنال‌های ثبت شده اعمال کردیم سپس سیگنال را به پنجره‌هایی تقسیم کرده و سپس شش ویژگی را از هر پنجره استخراج نمودیم. به منظور کم کردن حجم داده‌ها از الگوریتم تحلیل مؤلفه اصلی برای کاهش ویژگی‌ها استفاده شد و در



- 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, 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 پیشنهاد می‌شود تا بر روی محل و تعداد قرارگیری الکترودها و عضلات مطالعات بیشتری انجام گیرد تا بتوان سیستمی را طراحی نمود که بتوان با حداقل تعداد الکترودها، تعداد کلمات بیشتری را با صحت بالاتر طبقه‌بندی کرد. از طرفی با توجه به نبود پایگاه داده سیگنال‌های الکترومایوگرام عضلات چهره در حین بیان کلمات و جملات زبان فارسی، پیشنهاد می‌شود پایگاه داده‌ای به منظور تحقیقات بیشتر ایجاد شود. در آینده سعی می‌کنیم تا تحقیقات خود را بر روی جملات و عبارات فارسی با دایره واژگان وسیع‌تر برای تشخیص مبتنی بر سیگنال الکترومایوگرام و (بدون استفاده از سیگنال صوتی) معطوف نماییم.
- ### 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.*



- 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 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-Octob, pp. 3162–3167, 2020, doi: 10.1109/SMC42975.2020.9282863.
- [17] N. Srisuwan, P. Prukpattaranont, 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 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



Computation,” *Procedia Comput. Sci.*,
vol. 165, pp. 104–111, 2019, doi:
10.1016/j.procs.2020.01.079.

- 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.
- [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.
- [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