تشخیص احساسات از سیگنال‌های گفتار براساس روش‌های فیلتر

نرجس یزدانیان[۱] - حمید محمودیان[۲]

(۱) کارشناس ارشد - دانشکده مهندسی برق، واحد نجف‌آباد، دانشگاه آزاد اسلامی، نجف‌آباد، ایران

(۲) استادیار - دانشکده مهندسی برق، واحد نجف‌آباد، دانشگاه آزاد اسلامی، نجف‌آباد، ایران

تاریخ دریافت: ۱۳۹۴/۸/۲۶ تاریخ پذیرش: ۱۳۹۵/۵/۲۷

خلاصه: گفتار ابزار اولیه ارتباط بین انسانها است. با افزایش تراکنش میان انسان و ماشین نیاز به محاوره خودکار این دو و حذف کاربر انسانی مورد توجه قرار گرفته است. هدف از انجام این تحقیق، تعیین یک مجموعه از ویژگی‌های تأثیر گذار در تشخیص احساسات مبتنی بر سیگنال صحبت می‌باشد. در این مقاله، سیستمی طراحی گردید که شامل سه بخش اصلی، استخراج ویژگی، انتخاب ویژگی و طبقه‌بند می‌باشد. پس از استخراج ویژگی‌های پرکاربردی چون ضرایب کپسترال فرکانسی مل[۱] (MFCC)، ضرایب پیشگویی خطی[۲] (LPC)، ضرایب پیشگویی خطی ادراکی[۳] (PLP)، فرکانس فرمنت، نرخ عبور از صفر، ضرایب کپسترال، فرکانس گام، میانگین، جیتر، شیمر، انرژی، ضرایب تبدیل فوریه، کمترین مقدار در هر پنجره، بیشترین مقدار در هر پنجره، دامنه هر سیگنال و انحراف از معیار، در مرحله بعد به کمک روش‌های فیلتر چون معیار همبستگی پیرسون، آزمون t[۴]، رلیف و بهره اطلاعاتی به انتخاب و رتبه‌بندی ویژگی‌های تاثیرگذار در تشخیص احساسات پرداخته‌ایم. سپس نتایج بصورت زیرمجموعه‌ای از ویژگی‌ها به عنوان ورودی به یک سیستم طبقه‌بند داده شده است. در این مرحله از طبقه‌بند ماشین بردار پشتیبان چندگانه برای طبقه‌بندی هفت کلاس احساسی استفاده شده است. بر اساس نتایج بدست آمده روش انتخاب ویژگی رلیف به همراه طبقه‌بند ماشین بردار پشتیبان چندگانه دارای بیشترین میزان دقت طبقه‌بندی برای تشخیص احساسات مورد نظر با نرخ تشخیص ٪ ۹۳۹۴/ می‌باشد.

کلمات کلیدی: تشخیص احساسات از گفتار، استخراج ویژگی، انتخاب ویژگی، روش فیلتر، ماشین بردارهای پشتیبان.

Emotion Recognition of Speech Signals Based on Filter Methods

Narjes Yazdanian[(1)] - Hamid Mahmodian[(2)]

(1) MSc - Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran.

yazdanian.narjes@gmail.com

(2) Assistant Professor - Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran.

mahmoodian_hamid@yahoo.com

Speech is the basic mean of communication among human beings. With the increase of transaction between human and machine, necessity of automatic dialogue and removing human factor has been considered. The aim of this study was to determine a set of affective features the speech signal is based on emotions. In this study system was designs that include three mains sections, features extraction, features selection and classification. After extraction of useful features such as, mel frequency cepstral coefficient (MFCC), linear prediction cepstral coefficients (LPC), perceptive linear prediction coefficients (PLP), ferment frequency, zero crossing rate, cepstral coefficients and pitch frequency, Mean, Jitter, Shimmer, Energy, Minimum, Maximum, Amplitude, Standard Deviation, at a later stage with filter methods such as Pearson Correlation Coefficient, T-test, relief and information gain, we came up with a method to rank and select effective features in emotion recognition. Then Result, are given to the classification system as a subset of input. In this classification stage, multi support vector machine are used to classify seven type of emotion. According to the results, that method of relief, together with multi support vector machine, has the most classification accuracy with emotion recognition rate of 93.94%

Index Terms: Speech emotion recognition, features extraction, features selection, filter method, support vector machine.

نویسنده مسئول: نرجس یزدانیان، کارشناس ارشد - دانشکده مهندسی برق، واحد نجف‌آباد، دانشگاه آزاد اسلامی، نجف‌آباد، ایران، yazdanian.narjes@gmail.com

۱- مقدمه

پردازش گفتار به عنوان یکی از زیرشاخه‌های پردازش سیگنال به سرعت در حال گسترش است. علی‌رغم پیشرفت‌های گسترده در حوزهٔ پردازش گفتار، استخراج و درک احساس پنهان در گفتار انسان از یک سو و تولید گفتار احساسی مناسب از سوی دیگر، همچنان یکی از چالش‌های مهم برای ساخت ماشین‌های هوشمند محسوب می‌شود [۱] و [۲].

به‌طور کلی فرآیند سیستم تشخیص احساسات از گفتار شامل سه بخش اصلی استخراج ویژگی، انتخاب ویژگی و طبقه‌بندی می‌باشد. مرحله اول در این سیستم، استخراج ویژگی می‌باشد که بر روی فریم‌های متوالی سیگنال روش‌های مختلف استخراج ویژگی از قبیل ضرایب کپستروم فرکانس مل، فرکانس گام، ضرایب پیشگویی خطی، فرمنت و غیره اعمال می‌شود [۳]- [۵]. پس از استخراج ویژگی‌ها به دلیل اینکه تعدادی از آن‌ها تأثیر کمتری در طبقه‌بندی داشته و حتی موجب تخریب عملکرد سیستم تشخیص و طبقه‌بندی می‌شوند، بنابراین بر روی ماتریس آموزشی فرآیند انتخاب و رتبه‌بندی ویژگی صورت می‌گیرد و در پایان طبقه‌بند ماشین بردار پشتیبان چندکلاسه ارائه شده و مجموعه‌ای از ویژگی‌ها که دارای بیشترین تاثیر در طبقه-بندی می‌باشند به نحوی که علاوه بر داشتن بیشترین میزان درستی تشخیص احساسات، دارای تعداد ویژگی انتخابی کم‌تری نیز هستند، به عنوان مجموعه ویژگی نهایی به کمک روش وزن‌دهی پیشنهادی انتخاب می‌شوند. به منظور اینکه نتایج به دست آمده به واقعیت نزدیک باشند اعتبارسنج ضربدری K-Fold بر کلیه مراحل انتخاب ویژگی و طبقه‌بند اعمال شده است. نخستین تلاش‌ها برای اختراع سیستمی برای بازشناسی خودکار احساسات توسط ماشین، از سال ۱۹۹۵ شروع شده است. تیسچر نیز در این ۱۹۹۵ با استخراج ۳۹ ویژگی طیفی و عروضی با کمک آنالیز رگرسیون چندگانه به بازشناسی ۱۵ حالات احساسی مجزا پرداخت [۶].

در سال ۲۰۰۰ پتروشین با ترکیب ویژگی‌های فرکانس گام، فرکانس فرمنت‌ها و انرژی استخراج شده از نمونه‌های صوتی جمع‌آوری شده از ۳۰ بازیگر غیرحرفه‌ای و با استفاده از طبقه‌بند شبکه عصبی مصنوعی به طبقه‌بندی حالات احساسی خشم، ترس، شادی، ناراحتی و طبیعی با میزان درستی ۷۰٪ رسید [۷].

وروردیس و همکارانش در سال ۲۰۰۸ با استخراج ویژگی‌های فرکانس گام، انرژی و فرکانس‌های فرمنت، از سیگنال‌های صوتی منتخب از پایگاه‌های داده DES[5] و SUSAS و با استفاده از طبقه‌بندی کننده بیز[۶] به نرخ تشخیص احساسات کمتر از ۷۰٪ رسید و در مطالعات خود روشی مناسب برای توصیف مشکلات استفاده از روش‌های انتخاب ویژگی و طبقه‌بند کننده ارائه نمود [۸].

بوزکورت و همکارانش در سال ۲۰۰۹ تحقیقات خود را با استخراج ویژگی‌های اسپکترال، عروضی و ویژگی‌ها بر اساس مدل مخفی مارکف انجام و موفق به طبقه‌بندی پنج حالات احساسی با نرخ تشخیص ۶۳٪ شدند [۹]. یانگ و همکارانش با استخراج ویژگی‌هایی نظیر انرژی، فرمنت و نرخ عبور از صفر و طبقه‌بند مدل مخلوط گوسی، نرخ تشخیص حالت احساسی شادی را ۵۲/۷٪ گزارش کردند. این محققان آزمایشات خود را برروی سیگنال‌های صوتی انتخاب شده از پایگاه داده برلین انجام دادند [۱۰]. در سال ۲۰۱۰ یانگ و همکارانش با استخراج ویژگی‌هایی نظیر انرژی، فرمنت و نرخ عبور از صفر و طبقه‌بند مدل مخلوط گوسی، نرخ تشخیص حالت احساسی شادی را ۵۲/۷٪ گزارش کردند. این محققان آزمایشات خود را برروی سیگنال‌های صوتی انتخاب شده از پایگاه داده برلین انجام دادند [۱۱].

در سال ۲۰۱۱ انریکو البورنوز با استخراج ویژگی‌های طیفی و MLS[7] با استفاده از طبقه‌بند کننده مدل مخلوط گوسی و مدل مخفی مارکوف به تشخیص احساسات شادی، عصبانیت، ترس، خواب‌آلود، ناراحتی و طبیعی از سیگنال‌های گفتار استخراج شده از پایگاه داده برلین پرداختند و به این نتیجه رسیدند که ترکیب ویژگی‌های طیفی و عروضی مانند انرژی، فرکانس گام و ضرایب فرکانس مل سپستروم، باعث بهبود نرخ تشخیص احساسات از گفتار می‌گردد و میزان درستی تشخیص احساسات را ۶۶/۸۳٪ اعلام کردند[۱۲]. در مرجع [۱۳]، در سال ۲۰۱۲ با استخراج ویژگی‌های ضرایب فرکانس مل سپستروم، فرکانس گام، انرژی و نرخ عبور از صفر از سیگنال‌های صوتی پایگاه‌های داده برلین، DES و سبریان با کمک طبقه بندهای DAG - SVM، UDT و **3**DEC[۸] موفق به طبقه‌بندی حالات احساسی خشم، شادی ناراحتی، تعجب و طبیعی با نرخ تشخیص ۶۳/۲٪ شدند و به این نتیجه رسیدند که برای بازشناسی احساسات طبقه‌بند **3**DEC نسبت به سایر طبقه‌بندهای استفاده شده در تحقیقاتش از عملکرد بهتری برخوردار می‌باشد.

مرجع [۱۴] در سال ۲۰۱۲، مقایسه بین ویژگی‌های ضرایب فرکانس مل سپستروم و SBC با استفاده از طبقه‌بندی کننده مدل مخلوط گوسی برای طبقه‌بندی حالات احساسی شادی، ترس، عصبانیت، تنفر، ناراحتی و طبیعی صورت گرفته است. نتایج نشان می‌دهد که استفاده از ویژگی‌های SBC نرخ تشخیص احساسات را در مقایسه با ضرایب فرکانس مل سپستروم از ۵۱٪ به ۷۰٪افزایش می‌دهد. همچنین نشان داده شده است که ویژگی‌های SBC[9] در مقایسه با ضرایب فرکانس مل سپستروم، حساسیت کمتری نسبت به نویز دارند. در سال ۲۰۱۳ مایانک باهارگاوا و همکارانشان با استخراج ویژگی‌های انرژی، فرکانس گام و ضرایب فرکانس مل سپستروم از سیگنال‌های صوتی منتخب از پایگاه داده برلین و با استفاده از طبقه‌بند کننده ماشین بردار پشتیبان و شبکه عصبی مصنوعی به ترتیب به دقت اندازه‌گیری ۲۷/۸۰٪ و ۶/۸۰٪ رسیدند [۱۵].

در سال ۲۰۱۵ رضا اسدی و همکارش برای تشخیص شش احساس مختلف و به بررسی تأثیر انواع مختلف ویژگی‌ها نظیر، MFCC و TEO[10] با استفاده از ماشین بردار پشتیبان و k- نزدیکترین همسایگی پرداختند. ویژگی‌های TEO تاثیر بیشتری در دقت تشخیص داشتند. در این آزمایش بهترین نتیجه مربوط به ماشین بردار پشتیبان با کرنل گوسی، ٪ ۹۳/۹۴ گزارش شده است [۱۶].

با وجود فعالیت‌های گسترده در زمینه پردازش احساسات از سیگنال گفتار و کاربردهای وسیع آن، لازم است ترکیبی از روش‌های گذشته و

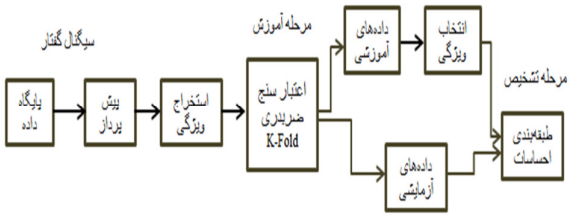
روش‌های جدید صورت گیرد تا دقت سیستم تشخیص احساس بهبود یابد. بنابر مطالعات گذشته در سیستم‌های بازشناسایی احساسات، از ویژگی‌های استخراج شده به صورت دسته‌ای استفاده شده است اما در این مطالعه سعی بر آن است که با استفاده از روش‌های آماری و انتخاب ویژگی‌های منفرد از گروه‌های مختلف ویژگی‌ها و ترکیب این ویژگی‌های تأثیرگذار در تشخیص احساسات به افزایش راندمان جهت طبقه‌بندی احساسات از سیگنال گفتار رسید.

در ادامه‌ی مقاله در بخش۲ سیستم تشخیص احساسات شرح داده خواهد شد. در بخش سوم روش پیشنهادی را بیان نموده و در ادامه در بخش ۴ به تحلیل نتایج خواهیم پرداخت.

۲- سیستم تشخیص احساس

سیستم تشخیص احساس از روی گفتار معمولاً از دیدگاه تشخیص الگو در سه بخش استخراج ویژگی، انتخاب ویژگی و طبقه‌بندی مورد بررسی قرار می‌گیرد.

همان طور که در شکل (۱) نشان داده شده است، در اولین مرحله پس از پیش‌پردازش سیگنال که شامل فریم‌بندی و پنجره‌گذاری می‌باشد، برای تشخیص احساسات گوینده نیاز به آن است که ویژگی‌های قابل مقایسه‌ای از سیگنال گفتار مد نظر استخراج گردد. در مرحله دوم ویژگی‌های موثر در تشخیص احساس توسط الگوریتم انتخاب ویژگی شناسایی و رتبه‌بندی شده و در آخرین مرحله نیز طبقه‌بند با توجه به ویژگی‌های انتخاب شده احساس موجود در گفتار ادا شده را مشخص می‌کند.



شکل (۱): فرایند سیستم تشخیص احساس

Fig. (1): The process system emotion recognition

۲-۱- استخراج ویژگی

برای تشخیص احساسات گوینده نیاز به آن است که ویژگی‌های قابل مقایسه‌ای از سیگنال گفتار استخراج گردد و بر اساس آن ویژگی‌ها، عملیات تشخیص و طبقه‌بندی صورت پذیرد. از جمله ویژگی‌های پرکاربرد در زمینه پردازش احساسات از روی سیگنال گفتار، ضرایب کپسترال فرکانس مل (MFCC)، ضرایب کپسترال پیشگویی خطی (LPC)، ضرایب پیشگویی خطی ادراکی (PLP)، فرکانس فرمنت[۱۱]، نرخ عبور از صفر[۱۲]، ضرایب کپسترال[۱۳]، فرکانس گام[۱۴]، جیتر[۱۵] و شیمر[۱۶] می‌باشند که از فریم‌های متوالی سیگنال گفتار استخراج می‌شوند [۱۷-۲۳].

۲-۲- انتخاب ویژگی

پس از فرایند استخراج ویژگی و تشکیل ماتریس بردار ویژگی، در مرحله دوم انتخاب ویژگی‌های تاثیرگذار به منظور بهبود میزان درستی تشخیص احساسات گوینده صورت می‌گیرد. روش‌های انتخاب ویژگی به دو دسته رتبه‌بندی و بسته‌بندی تقسیم می‌شوند [۲۴]. روش‌های رتبه‌بندی یا فیلتر، مستقل از الگوریتم‌های یادگیری بوده و بر اساس یک معیار رتبه‌بندی خاص عمل می‌کند، به همین علت این روش‌ها از نظر محاسباتی کاراتر بوده و برای پردازش داده‌های با ابعاد بالا مناسب هستند. در مقابل به مجموعه روش‌هایی که از تابع ارزیابی مبتنی بر نرخ خطای طبقه‌بندی کننده استفاده می‌کنند روش‌های بسته‌بندی می‌گویند، این روش‌ها از الگوریتم‌های یادگیری خاصی جهت ارزیابی کارآیی زیرمجموعه ویژگی‌های مختلف استفاده می‌کنند. درنتیجه عموماً هزینه محاسباتی بالاتری نسبت به روش‌های فیلتر دارند [۲۵] و [۲۶].

در این مقاله از روش‌های فیلتر چون، معیارهای همبستگی پیرسون، آزمون t، رلیف و بهره اطلاعاتی برای انتخاب و رتبه‌بندی ویژگی‌ها استفاده شده است. سپس نتایج به صورت مجموعه‌ای از ویژگی‌های رتبه‌بندی شده به عنوان ورودی به یک سیستم طبقه‌بند داده می‌شود.

۲-۲-۱- ضریب همبستگی پیرسون[۱۷]

یکی از روش‌های انتخاب ویژگی ضریب همبستگی پیرسون می‌باشد که توسط کارل پیرسون معرفی شده‌است [۲۷]. این ضریب، میزان همبستگی که با $\mathrm{Corr}(X,Y)$ نمایش داده می‌شود، بین دو متغیر فاصله‌ای یا نسبی را محاسبه کرده و مقدار آن بین ۱+ و ۱- می‌باشد.

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\delta_x + \delta_y} \qquad (۱)$$

در (۱)، X بیانگر بردار نمونه‌ها و Y بیانگر برچسب کلاس‌های مربوطه می‌باشد. همچنین δ_X و δ_Y به ترتیب انحراف معیار نمونه‌ها و انحراف معیار برچسب کلاس های مربوطه را نمایش می‌دهند.

۲-۲-۲ آزمون t[۱۸]

یکی دیگر از روش‌های انتخاب ویژگی، آزمون t می‌باشد که ویژگی‌ها را بر اساس معیار میانگین و واریانس رتبه‌بندی می‌نماید. در واقع در این روش، ویژگی‌هایی برتر هستند که موجب افزایش فاصله بین میانگین دو کلاس و کاهش پراکندگی درون کلاسی گردند.

$$t = \frac{|\mu_1 - \mu_2|}{\delta_1 + \delta_2} \qquad (۲)$$

در معادله (۲) مقادیر، μ_1، μ_2، δ_x و δ_y به ترتیب بیانگر میانگین و انحراف معیار نمونه‌های کلاس یک و دو می‌باشند.

۲-۲-۳ رلیف[۱۹]

روش رلیف، از یک راه حل آماری برای انتخاب ویژگی استفاده می‌کند. همچنین یک روش مبتنی بر وزن است که از الگوریتم‌های مبتنی بر

نمونه الهام گرفته است. روش کار به این صورت است که از میان مجموعه نمونه‌های آموزشی، یک زیرمجموعه نمونه انتخاب می‌شود. الگوریتم به صورت تصادفی یک نمونه از این زیرمجموعه را انتخاب می‌کند، سپس برای هر یک از ویژگی‌های این نمونه، نزدیکترین برخورد و نزدیکترین شکست را بر اساس معیار اقلیدسی پیدا می‌کند. نزدیکترین برخورد، نمونه‌ای است که کمترین فاصله اقلیدسی را در میان سایر نمونه‌های هم‌کلاس با نمونه انتخاب شده دارد. نزدیکترین شکست نیز نمونه‌ای است که کمترین فاصله اقلیدسی را در میان نمونه‌های غیر هم‌کلاس با نمونه انتخاب شده دارد. الگوریتم پس از تعیین نزدیکترین برخورد و نزدیکترین شکست، وزن‌های ویژگی‌ها را به روزرسانی می‌کند. این به‌روزرسانی به این صورت است که مربع اختلاف بین مقدار ویژگی مورد نظر در نمونه انتخاب شده و نمونه نزدیکترین برخورد از وزن ویژگی کم می‌شود و در عوض مربع اختلاف بین مقدار ویژگی در نمونه انتخاب شده و نزدیکترین شکست به وزن ویژگی اضافه می‌شود. هر چه مقدار این وزن بزرگتر باشد، ویژگی مورد نظر، بهتر می‌تواند نمونه‌های یک کلاس را از دیگران جدا کند [۲۸].

الگوریتم رلیف در شکل (۲) بیان شده‌است که در آن، D: مجموعه آموزشی، S: مجموعه ویژگی اصلی، N: تعداد ویژگی‌ها، T: زیرمجموعه ویژگی انتخاب شده، M: تعداد ویژگی‌های انتخاب شده یا تعداد ویژگی‌هایی است که لازم است انتخاب شوند.

۲-۲-۴- بهره اطلاعاتی[۲۰]

یکی دیگر از روش‌های انتخاب ویژگی روش بهره اطلاعاتی می‌باشد. این روش، ویژگی‌ها را بر اساس معیار آنتروپی مرتب می‌نماید. در واقع ویژگی‌هایی برتر هستند که دارای بیشترین بار اطلاعاتی نسبت به سایر ویژگی‌ها می‌باشند [۲۹]. به‌عبارت دیگر بهره اطلاعات برای یک بردار ویژگی نظیر A به صورت زیر تعریف می‌شود:

$$Gain(A) = -\sum_{i=1}^{N} p_i \log_2 p_i \qquad (۳)$$

در معادله (۳)، N تعداد نمونه‌های بردار ویژگی A و p_i برابر است با مقدار نمونه iام تقسیم بر مجموع مقادیر نمونه‌های ویژگی A.

۲-۳- طبقه‌بندی

Relief (D, S, NoSample, Threshold)

(1) T= Ø

(2) Initialize all weights, W_i , to zero.

(3) For i=1 to NoSample/*Arbitrarily chosen*/

Randomly choose an instance x in D

Finds its nearHit and nearMiss

For j=1 to N

$W_j = W_j - diff(x_j, nearHit_j)^2 + diff(x_j, nearMiss_j)^2$

(4) For j=1 to N

If $W_j \geq$ Threshold

Append feature f_i to T

(5) Return T

شکل (۲): الگوریتم رلیف

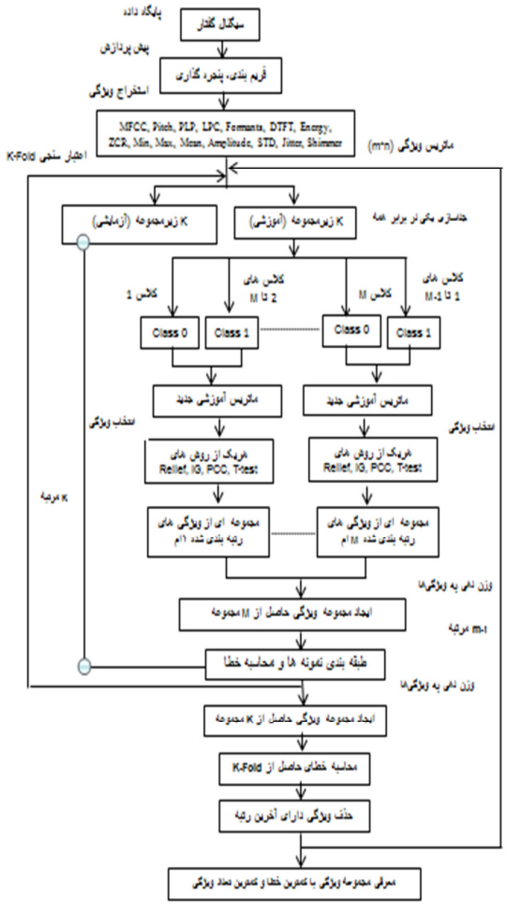
Fig. (2): Relief algorithm

پس از رتبه‌بندی ویژگی‌ها در این مرحله نوبت به تشخیص احساسات گوینده توسط طبقه‌بند می‌رسد. در این مرحله از طبقه‌بند ماشین بردار پشتیبان چندکلاسه (MSVM) برای طبقه‌بندی استفاده شده است. ماشین بردار پشتیبان یکی از روش‌های توانمند می باشد که در حالت پایه تنها برای جداکردن دو کلاس از یکدیگر طراحی شده است، ولی از آن برای طبقه‌بندی چندکلاسه نیز استفاده می‌کنند. ایده اولیه این روش برای اولین بار در سال ۱۹۷۹ توسط وپنیک مطرح شده است. اساس کار SVM بر مبنای حداقل‌سازی خطای ساختاری بنا شده است [۳۰]. در صورتیکه پایگاه داده مد نظر برای تشخیص احساسات دارای دو کلاس متفاوت باشد می‌توان از طبقه‌بند مانند ماشین بردار پشتیبان دو کلاسه (باینری) استفاده نمود، که با توجه به مطالعات پیشین دارای دقت عملکرد بالایی می‌باشد. اما در صورتیکه پایگاه داده مورد نظر دارای احساسات بیش از دو کلاس باشد باید از طبقه‌بندهای چندکلاسه استفاده کرد. بدین منظور طبقه‌بند جدیدی بر مبنای ماشین بردار پشتیبان ارائه شده است. تحت عنوان ماشین بردار پشتیبان چندکلاسه از نوع یکی در مقابل همه[۲۱].

۳- روش پیشنهادی

در این مطالعه به منظور تشخیص احساسات از سیگنال گفتار سیستمی شبیه سازی شده است که مطابق شکل(۳) شامل چهار مرحله اصلی پیش پردازش، استخراج ویژگی، انتخاب ویژگی و طبقه‌بند می‌باشد.

پس از دریافت سیگنال صوتی از پایگاه داده مورد نظر و پیش از استخراج ویژگی‌ها عملیات فریم‌بندی و پنجره‌گذاری سیگنال تحت عنوان پیش‌پردازش اجرا می‌گردد، که در این مطالعه پایگاه داده مد نظر به زبان آلمانی (تحت عنوان پایگاه داده برلین) با ۱۰ گوینده و دارای هفت احساس (کلاس) شامل: نفرت با ۲ نفر گوینده (۱ مرد و ۱ زن)، خواب آلود با ۳ گوینده (۲مرد و ۱ زن)، عصبانیت با ۲ گوینده (۱ مرد و ۱ زن)، ترس با ۲ گوینده (۲ مرد)، شادی با ۲ گوینده (۱ مرد و ۱ زن)، غمگین با ۲ گوینده (۱ مرد و ۱ زن) و طبیعی با ۳ گوینده (۲مرد و ۱ زن) می باشد. در این مقاله طول هر فریم ۵۰ میلی ثانیه (۴۰۰ نمونه) با ۵۰ درصد همپوشانی می‌باشد و پنجره‌گذاری همینگ بر روی هر فریم اعمال می‌شود و سپس مراحل اصلی سیستم تشخیص احساسات شامل استخراج ویژگی، انتخاب ویژگی و طبقه‌بندی بر روی این نمونه‌های صوتی اعمال می‌شود.



شکل (۳): سیستم فرآیند تشخیص احساسات از سیگنال گفتار
Fig. (3): The system of emotion recognition of speech signals process

- به منظور تشکیل ماتریس ویژگی، بر روی هر پنجره، روش‌های استخراج ویژگی به کار گرفته می‌شود. در مطالعات گذشته یکی از پرکاربردترین روش‌ها برای تشکیل بردار ویژگی، ضرایب مل کپستروم، فرکانس گام، انرژی و ضرایب پیشگویی خطی می‌باشد. بنابراین با توجه به مطالعات انجام شده، اولین روش در استخراج ویژگی روش ضرایب مل کپستروم می باشد که در این مطالعه با روش (A1) عنوان می‌شود. روش دیگر ترکیبی از ضرایب مل کپستروم و فرکانس گام می‌باشد که با روش (A2) عنوان می‌گردد. سومین رویکرد، ترکیبی از ضرایب مل کپستروم و فرکانس گام و انرژی است که با روش (A3) بیان می‌شود و آخرین رویکرد ترکیب سه روش فوق‌الذکر (A1 تا A3) به همراه ضرایب پیشگویی خطی می‌باشد که با عنوان روش A4 ارائه می‌گردد. روش پیشنهادی در این مقاله همان‌طور که در شکل (۴) نشان داده شده است علاوه بر روش A4، ترکیبی از ویژگی‌های فرکانس‌های فرمنت، ضرایب تبدیل فوریه، ضرایب پیشگویی خطی ادراکی، نرخ عبور از صفر، میانگین، کمترین مقدار در هر پنجره (Min)، بیشترین مقدار در هر پنجره (Max)، دامنه هر سیگنال، انحراف از معیار، جیتر و شیمر را تحت عنوان (PR-B) می‌باشد.

شکل (۴): فرایند استخراج ویژگی و تشکیل ماتریس ویژگی‌ها.
Fig (4): Process of feature extraction and to organization features matrix

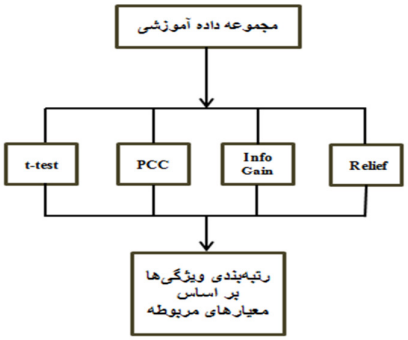
جدول (۱) روش‌های استخراج ویژگی در مطالعات پیشین و روش استخراج ویژگی پیشنهادی را نمایش می‌دهد.

Table (1): Feature extraction methods
جدول (۱): روش‌های استخراج ویژگی

نام‌گذاری روش‌ها	روش‌های استخراج ویژگی
A1	MFCC
A2	MFCC, Pitch
A3	MFCC, Pitch, Energy
A4	MFCC, Pitch, Energy, LPC
PR-B	MFCC, Pitch, PLP, LPC, Formants, DTFT, Energy, ZCR, Min, Max, Mean, Amplitude, STD, Jitter, Shimmer

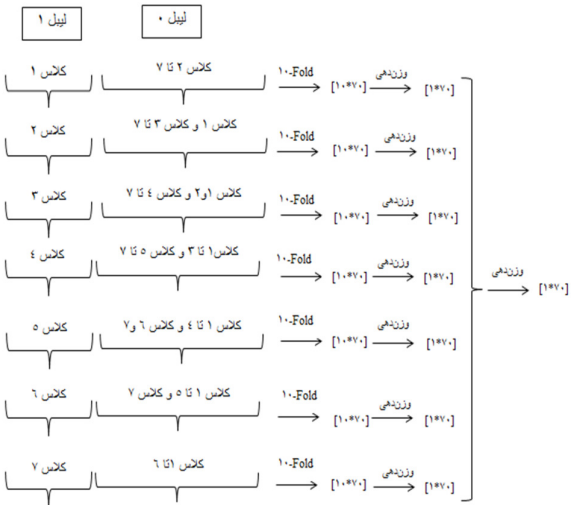
پس از تشکیل بردارهای ویژگی استخراج شده با روش پیشنهادی، این بردارها در ماتریسی تحت عنوان ماتریس ویژگی قرار داده می‌شوند که این ماتریس دارای m سطر (تعداد ویژگی‌ها) و n ستون (تعداد نمونه‌ها) می‌باشد. برای طبقه‌بندی نیاز است که تعدادی از نمونه‌ها به عنوان مجموعه آموزشی و تعدادی از آن‌ها به عنوان مجموعه آزمایشی مجزا گردند، بدین منظور از نسبت ٪۷۵ به ٪۲۵ استفاده می‌شود به گونه‌ای که ٪۷۵ از نمونه‌های موجود در هر یک از حالات احساسی به صورت کاملا تصادفی به عنوان مجموعه آموزشی و ٪۲۵ باقیمانده در هر یک از حالات نیز مجموعه آزمایشی را به خود اختصاص می‌دهند.

- در مرحله دوم انتخاب و رتبه‌بندی ویژگی‌ها صورت می‌گیرد، با توجه به آنکه در ماتریس ویژگی تشکیل شده، تعدادی از ویژگی‌ها دارای ضرایب کم ارزش در تشخیص و طبقه‌بندی می‌باشند، بنابراین در این مطالعه به منظور بهبود عملکرد طبقه‌بند، انتخاب ویژگی‌های مطلوب از میان ویژگی‌های موجود پیشنهاد می‌گردد. روشهای انتخاب ویژگی به کار گرفته شده همانطور که در شکل (۵) نشان داده شده است شامل: آزمون t، ضریب همبستگی پیرسون، میزان بهره اطلاعاتی و الگوریتم ریلف می باشند.



شکل (۵): فرایند انتخاب ویژگی
Fig (5): Process of feature selection

لازم به توضیح است، به دلیل اینکه روش‌های انتخاب ویژگی به کار گرفته شده دو کلاسه می‌باشند در حالی که احساسات موجود در پایگاه داده چند کلاسه است. بدین منظور در این مطالعه بر روی ماتریس آموزشی عملیات جداسازی یک کلاس از سایر کلاس‌ها صورت گرفته به گونه‌ای که کلاس ۱ همان برچسب ۱ و سایر کلاس‌ها دارای برچسب صفر باشند و عملیات رتبه‌بندی ویژگی بر روی ماتریس جدید ایجاد شده که دارای ۲ کلاس (۰ و ۱) است، صورت می‌پذیرد و برای سایر کلاس‌ها نیز به همین طریق عملیات رتبه‌بندی ویژگی انجام می‌شود. این مرحله جداسازی چند کلاسه نامیده می‌شود. برای انتخاب مجموعه ویژگی نهایی، روش وزن‌دهی به ویژگی‌ها ارائه می‌گردد تا ویژگی‌های دارای کمترین وزن به عنوان مجموعه ویژگی انتخابی که دارای بیشترین تاثیر در تشخیص احساسات می‌باشند، معرفی گردند. روش وزن‌دهی به ویژگی‌ها بر اساس جایگاه آن ویژگی در خروجی روش جداسازی یکی در برابر همه می‌باشد.



روش پیشنهادی برای معرفی مجموعه ویژگی نهایی به این صورت است که بعد از اجرای جداسازی دودویی، ماتریسی تعیین کرده که تعداد ستون‌های ماتریس برابر M (تعداد کلاس‌ها) و تعداد سطرهای ماتریس برابر nf (تعداد ویژگی‌ها) می‌باشند. در ستون‌های این ماتریس، اندیس ویژگی‌های انتخاب شده در هر طبقه‌بند قرار گرفته‌اند. در واقع جایگاه هر ویژگی در هر ستون مورد بررسی قرار می‌گیرد، بگونه‌ایی که مجموع جایگاه هر ویژگی بر روی تمام کلاس‌ها تقسیم بر تعداد کل ویژگی‌ها می‌گردد یعنی اگر مجموع جایگاه هر ویژگی برابر N باشد، در نتیجه وزن هر ویژگی برابر با N/nf می‌باشد. در ادامه برای روشن شدن روش وزن‌دهی پیشنهادی مثالی بیان شده است:

فرض کنید عناصر ماتریس A(a_{ij})، بیانگر اندیس ویژگی iام در jامین مرحله از جداسازی‌های دودویی یکی در برابر همه می‌باشد. برای محاسبه وزن ویژگی اول (a_{1j} j =1,2,…,7) به صورت زیر عمل می‌شود:

$$A=\begin{bmatrix} a_{11} & a_{32} & a_{43} & a_{14} & a_{45} & a_{16} & a_{27} \\ a_{31} & a_{12} & a_{33} & a_{24} & a_{35} & a_{26} & a_{17} \\ a_{41} & a_{42} & a_{23} & a_{34} & a_{15} & a_{36} & a_{47} \\ a_{21} & a_{22} & a_{13} & a_{44} & a_{25} & a_{46} & a_{37} \end{bmatrix}$$

مجموع جایگاه ویژگی مورد نظر تقسیم بر تعداد ستون‌ها.

$$a_{1j} \quad \text{وزن ویژگی} = \frac{۱+۲+۴+۱+۳+۱+۲}{۴} = ۳/۵$$

- پس از رتبه‌بندی ویژگی‌ها نوبت به طبقه‌بندی می‌رسد. در صورتیکه که پایگاه داده مد نظر برای تشخیص احساسات دارای دو کلاس باشد می‌توان از طبقه‌بند ماشین بردار پشتیبان دو کلاسه (باینری) استفاده نمود، اما در صورتیکه پایگاه داده مورد نظر دارای احساسات بیش از دو کلاس باشد باید از طبقه‌بندهای چندکلاسه استفاده نمود. بدین منظور طبقه‌بندهای جدیدی بر مبنای ماشین بردار پشتیبان ارائه شده است که تحت عنوان ماشین بردار پشتیبان چندکلاسه (MSVM) به کار برده می‌شود.

در این مطالعه طبقه‌بند ماشین بردار پشتیبان چندکلاسه از نوع یکی در مقابل همه استفاده شده است. روش کار بدین صورت است که به تعداد کلاس‌ها، M طبقه‌بند کننده (به تعداد احساسات) ایجاد می‌شود و هر طبقه بندی کننده یک کلاس را از بقیه‌ی **M-1** کلاس متمایز می‌سازد. به عنوان مثال در اولین طبقه‌بند کلاس ۱ به عنوان طبقه یک و سایر کلاس‌ها به عنوان طبقه صفر در نظر گرفته و طبقه‌بند با نمونه‌های آموزشی، آموزش داده می‌شود. سپس خروجی طبقه‌بند را بر روی نمونه‌های آزمایشی به دست آورده و نمونه‌های آزمایشی که به عنوان طبقه ۱ در خروجی ظاهر گشته از سایر نمونه‌ها متمایز می‌شوند. سپس سایر نمونه‌ها که در طبقه صفر قرار گرفته‌اند به عنوان ورودی دومین طبقه‌بند در نظر گرفته می‌شوند و به همین ترتیب چرخه تا آخرین طبقه‌بند ادامه می‌یابد تا در نهایت خروجی حاصل از M طبقه‌بند به دست آمده و با برچسب نمونه‌های آزمایشی مقایسه گردد تا میزان درستی تشخیص طبقه‌بند[۲۲] بر روی M کلاس موجود محاسبه شود.

در ادامه ویژگی دارای کم‌ترین رتبه از مجموعه ویژگی حذف گردیده و فرایند طبقه‌بندی تکرار می‌شود. این عملیات تا رسیدن به تعداد ۱ ویژگی ادامه می‌یابد. سپس از میان مجموعه ویژگی‌های موجود،

مجموعه‌ای که دارای بیشترین میزان تشخیص و کم‌ترین تعداد ویژگی است به عنوان مجموعه ویژگی نهایی معرفی می‌گردد.

۴- تحلیل نتایج

داده‌های مورد استفاده برای انجام آزمایش‌ها، سیگنال‌های گفتار ۱۶ گوینده شامل ۱۰ مرد و ۶ زن است که از پایگاه داده زبان آلمانی دانشگاه برلین (Emo-DB) استخراج شده‌اند [۳۱] و مراحل سیستم بازشناسایی احساسات که شامل استخراج ویژگی، انتخاب ویژگی و طبقه‌بند می‌باشد بر روی این سیگنال‌ها پیاده‌سازی شده است و نتایج ذیل حاصل شده‌اند.

لازم به توضیح است که کلیه مراحل انتخاب ویژگی و طبقه‌بندی در داخل تابع اعتبار سنجی ضربدری K-Fold با K=10 قرار می‌گیرد.

جدول (۲) میزان دقت طبقه‌بند چند کلاسه را برای هر یک از روش‌های استخراج ویژگی، بدون بکارگیری روش‌های انتخاب ویژگی را نمایش می‌دهد.

Table (2): Feature extraction methods PR-B, A1, A2, A3 and A4 with the MSVM classification

جدول (۲): روش‌های های استخراج ویژگی PR-B، A1، A2، A3 و A4 با طبقه‌بند MSVM

روش‌های استخراج ویژگی	دقت طبقه‌بندی (٪)
روش PR-B	**82/23**
روش **A1**	**81/25**
روش **A2**	**82/42**
روش **A3**	**79/88**
روش **A4**	**74/61**

با توجه به جدول (۲) مشاهده می‌شود که روش A2 با طبقه‌بند MSVM شبیه سازی شده، دارای بیشترین دقت طبقه‌بندی ۸۲/۴۲٪ می‌باشد و تعدادی از ویژگی‌ها در روش پیشنهادی نه تنها موجب بهبود عملکرد سیستم نشده بلکه برخی از آنها عملکرد سیستم را تضعیف کرده‌اند. بنابراین باید عملیات کاهش ابعاد داده و انتخاب ویژگی‌های برتر برای بهبود عملکرد سیستم صورت پذیرد. در نتیجه پس از استخراج ویژگی توسط روش پیشنهادی **PR-B** حال نوبت به مراحل انتخاب ویژگی و محاسبه دقت طبقه‌بندی توسط طبقه‌بند چندکلاسه MSVM می‌رسد. بدین منظور تمام مراحل انتخاب ویژگی و طبقه‌بندی در داخل تابع اعتبار سنجی K-Fold بیرونی با K=10 قرار می‌گیرد.

جدول (۳) نتایج حاصل از به کارگیری روش‌های انتخاب ویژگی ذکر شده (میزان دقت و تعداد ویژگی انتخابی) را توسط طبقه‌بند MSVM با تابع اعتبار سنجی **10-Fold** و همچنین با ۲ مرتبه تکرار فرآیند نمایش می‌دهد.

Table (3): The results of feature selection methods by MSVM classification

جدول (۳): نتایج حاصل از به کارگیری روش‌های انتخاب ویژگی توسط طبقه-بند MSVM

روش‌های انتخاب ویژگی	طبقه‌بندی	
	Acc(%)	NF
PCC	**92/97**	**8**
t-test	**91/69**	**7**
Info-Gain	**91/60**	**10**
Relief	**93/94**	**7**

جدول (۴) مجموعه ویژگی نهایی توسط روش انتخاب ویژگی برتر (روش Relief) را به همراه اندیس و نام ویژگی نمایش می‌دهد.

Table (4): The final feature set with Relief method and classified by MSVM

جدول (۴): مجموعه ویژگی نهایی توسط روش Relief با طبقه‌بند MSVM

ویژگی	اندیس ویژگی
zc	4
Lpc 2	31
cepstrom 1	57
cepstrom 3	59
lpc 3	32
plp 5	26
plp 4	25

همانطور که از جدول (۴) مشخص است اولین رتبه به ویژگی نرخ عبور از صفر تعلق دارد و ضرایب LPC و PLP نیز از موثر ترین ویژگی‌ها می‌باشند. حال روش پیشنهادی با سایر روش‌های موجود در پژوهش‌های پیشین مورد مقایسه قرار می‌گیرد، جدول (۵) مربوط به مقایسه روش پیشنهادی با پژوهش‌های پیشین می‌باشد.

مطابق جدول (۵) روش پیشنهادی با طبقه‌بند MSVM دارای بیشترین دقت و کم‌ترین تعداد ویژگی نسبت به روش‌های ارائه شده در مطالعات پیشین است. مرجع [۳۲] از روش استخراج ویژگی A1 (MFCC) استفاده کرده است و بدون استفاده از روش انتخاب ویژگی و با طبقه-بندهای متفاوت ماشین بردار پشتیبان، K نزدیکترین همسایه، مدل مخفی مارکف و مدل ترکیبی گوسین به ترتیب به دقت‌های ۸۵/۴۰٪، ۸۶/۷۰٪، ۸۹/۹۰٪ و ۸۶/۱۰٪ دست یافته است که طبقه‌بند مدل مخفی مارکف نسبت به سایر طبقه‌بندها دارای نتیجه بهتری است.

مرجع [۳۳] با روش استخراج ویژگی A2 (MFCC+Pitch) و بدون انتخاب ویژگی توسط طبقه‌بند مدل مخفی مارکف به بیشترین دقت ۶۷/۸۰٪ دست یافته است. مرجع [۳۴] با روش استخراج ویژگی A3 (MFCC+Pitch+Energy) و با روش انتخاب ویژگی بهره اطلاعاتی توسط طبقه‌بند شبکه عصبی به میزان دقت ۹۲/۴۰٪ دست یافته است.

مرجع [۳۵] با روش استخراج ویژگی A4 (MFCC+Pitch+Energy+LPC) و بدون انتخاب ویژگی توسط طبقه‌بند SVM به میزان دقت ۸۲/۵۰٪ دست یافته است.

Table (5): Results compare of the proposed method with previous studies

جدول (۵): مقایسه نتایج حاصل از روش پیشنهادی با مطالعات پیشین

پایگاه داده برلین	استخراج ویژگی + انتخاب ویژگی + طبقه‌بندی	تعداد ویژگی	دقت طبقه-بندی (٪)
روش پیشنهادی	روش PR-B + Relief + MSVM	۷	93/94
نتیجه بیان شده در مرجع [۳۲]	روش A1+بدون انتخاب ویژگی+SVM	13	85/40
	روش A1+بدون انتخاب ویژگی+KNN	13	86/70
	روش A1+بدون انتخاب ویژگی+HMM	13	89/90
	روش A1+بدون انتخاب ویژگی+NN	13	86/10
نتیجه بیان شده در مرجع [۳۳]	روش A2+بدون انتخاب ویژگی+KNN	14	53/17
	روش A2+بدون انتخاب ویژگی+GMM	14	66/00
	روش A2+بدون انتخاب ویژگی+HMM	14	67/83
نتیجه بیان شده در مرجع [۳۴]	روش A3+Info Gain+ NN	10	92/40
نتیجه بیان شده در مرجع [۳۵]	روش A4+بدون انتخاب ویژگی+SVM	15	82/50

۵- بحث و نتیجه‌گیری

در این مقاله برای تشخیص احساسات از گفتار ابتدا روش‌های استخراج ویژگی اعمال و پس از آن برای داشتن عملکرد بهتر سیستم، روش‌های رتبه بندی انتخاب ویژگی مورد بررسی قرار گرفته و در نهایت میزان درستی هر روش توسط طبقه‌بند ماشین بردار پشتیبان چندگانه حاصل گردید. نتایج نشان می‌دهد که روش انتخاب ویژگی رلیف به همراه طبقه‌بند چندکلاسه ماشین بردار پشتیبان دارای بیشترین میزان دقت طبقه‌بندی برای تشخیص احساسات مورد نظر با نرخ تشخیص ٪ ۹۴/۹۳ و ارائه مجموعه‌ای از ویژگی‌های برتر با کم‌ترین تعداد ویژگی می‌باشد. مجموعه ویژگی نهایی نشان می‌دهد که ویژگی‌های نرخ عبور از صفر، ضرایب پیشگویی خطی و ضرایب پیشگویی خطی ادراکی به همراه ضرایب کپستروم در تشخیص احساسات نقش موثرتری نسبت به سایر ویژگی‌ها ایفا می‌نمایند.

کاربرد عملی این پژوهش در تشخیص احساسات از گفتار، در برنامه‌های کاربردی کامپیوتر، ابزار تشخیص برای درمانگران، مراکز پاسخگویی خودکار، ارتباطات تلفن همراه می‌باشد.

پی‌نوشت:

1. Mel Frequency Cepstral Coefficients
2. Linear Prediction Cepstral Coefficients
3. Perceptual linear predictive Coefficients
4. Student's t-distribution
5. Danish Emotional Speech database
6. Bayes classifier
7. Mean of the Long Spectrum
8. Data-Driven Dimensional Emotion Classification
9. Subband based Cepstarl Parameter
10. Teager Energy operator
11. Formant
12. Zero Crossing
13. Cepstral Coefficients
14. Pitch
15. Jitter
16. Shimmer
17. Pearson Correlation Coefficient
18. T-test
19. Relief
20. Information Gain
21. One Against All
22. Accuracy

References

[1] O. Shaughnessy, "Speech communication human and machine ", Reading: Addison- Wesley,1987.

[2] D.J. France, et. al., "Acoustical properties of speech as indicators of depression and suicidal risk," IEEE Trans. Biomedical Engineering, Vol. 47, No. 7, pp. 829-837,2000.

[3] N. Dehak, "Modeling Prosodic Features With Joint Factor Analysis for Speaker Verification", IEEE Transactions on Audio, Speech, and Language Processing 15, 2095–2103. 2007.

[4] X. Helander, Nurminen, "On the importance of prosody on speaker identity," Dans les actes de Eurospeech. 2007.

[5] E. Shriberg, "Modeling prosodic feature sequences for speaker recognition", speech communication 46(3-4), pp. 455–472, 2005.

[6] B. Tischer,"Acoustic correlates of perceived emotional stress.", 1995.

[7] V. Petrushin, "Emotion recognition in speech signal: experimental study, development and application", Proceedings of the ICSLP 2000, Beijing, China, pp. 16-20, October 2000.

[8] D. Ververidis, C. Kotropoulos, "Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition" , signal processing, vol. 88, pp. 2956-2970, 2008.

[9] E. Bozkurt, E. Erzin, "Improving automatic emotion recognition from speech signals." In 10th annual conference of the international speech communication association (interspeech), Brighton, UK, Sept ,pp. 324–327, 2009.

[10] B. Yang, M. Lugger, " Emotion recognition from speech signals using new harmony features" ,signal processing 90, pp. 1415-1423, 2010.

[11] B. Yang, M. Lugger, " Emotion recognition from speech signals using new harmony features" ,signal processing 90, pp. 1415-1423, 2010.
[12] E. Albornoz, et. al. , "Spoken Emotion recognition using hierarchical classifier", Computer speech and language , pp. 556-570, 2011.
[13] A. Hassan, R. Damper, "Classification of emotional speech using 3DEC hierarchical classifier.", speech communication 54, pp.903-916. 2012.
[14] K. Kishore, P. Satish, "Emotion recognition in speech using MFCC and wavelet features." ,IEEE International Advance Computing Conference (IACC), 2012.
[15] M. Bhargavaa, T. Polzehl, "Improving Automatic Emotion Recognition from speech using Rhythm and Temporal feature." School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Bhubaneswar **751013**,India., Quality and Usability Lab, T labs/ TU Berlin, Germany, **7** Mar **2013**.
[16] R. Asadi, H. Fell, 'Improving the accuracy of speech emotion recognition using acoustic landmarks and Teager energy operator features", The Journal of the Acoustical Society of America. ,Vol.137, April 2015.
[17] H. Harb, L. Chen, "Voice-based gender identification in multimedia application", Jornal of Intelligent Informatin Systems, vol.24, issue.2-3, pp.179-198, March 2005.
[18] S. McGilloway, R.Cowie, "Approaching automatic recognition of emotion from voice", A rough benchmark. In ISCA workshop onspeech and emotion, Belfast, 2000.
[19] S. Biersack, V. Kemp, "Exploring the influence of vocal emotion expression on communicative effectiveness", phonetica, Vol. 62, pp. 106-119, 2005.
[20] T. Waaramaa, P. Alku, A. Laukkanen, "The role of F3 in the vocal expression of emotion", logopedics phoniatrics Vocology, Vol. 31, pp. 153-156, 2006.
[21] J. Makhoul, "Linear prediction: A tutorial review," Pro-ceedings of the IEEE, pp. 561-580, 1975.
[22] Shikha Gupta, Jafreezal Jaafar, "FEATURE EXTRACTION USING MFCC," Universiti Tecknologi PETRONAS, CIS Dept, Perak, Malaysia, August 2013.
[23] M. Farrús, J. Hernando, P. Ejarque, "Jitter and Shimmer Measurements for Speaker Recognition," TALP Research Center, Department of Signal Theory and Communications Universitat Politècnica de Catalunya, Barcelona, Spain.
[24] J. Jarmulak, S. Craw, "Genetic algorithms for feature selection and weighting", Appears in Proceedings of the IJCAI'99 workshop on Automating the Construction of Case Based Reasoners, 1999.
[25] Kabir, Md. M., Islam, Md. M., Murase, K., "A new wrapper feature selection approach using neural network", Neurocomputing, pp. 3273-3283, 2010.
[26] R. Kohavi, G. John, "Wrappers for Feature Subset Selection.", Artificial Intelligence, pp.273- 324, 1997.
[27] K. Pearson, "Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia," Phil Trans. R Soc Lond Series A, pp. 253-318.
[28] R.P.L. DURGABAI, "Feature Selection using Relief Algorithm.", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 10, Oct. 2014.
[29] Andrew W. Moore "Information Gain", School of Computer Science, Carnegie Mellon University, 2003.
[30] L. Wang. "Support Vector Machines: Theory and Applications.", ISBN: 3540243887, Springer-Verlag, New-York, Inc. 2005.
[31] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., Weiss, B. "A database of German emotional speech." In Interspeech, 2005.
[32] Sara Motamed, "Speech Emotion Recognition Based on Learning Automata in Fuzzy Petri-net." Journal of mathematics and computer science, pp. 173-185, 2014.
[33] B. Rahul Lanjewar, D.S. Chaudhari, "Comparative analysis of speech emotion recognition system using different classifiers on Berlin emotinal speech database." International Journal of Electrical and Electronics., Vol. 3, Issue 5, pp. 145-156, Dec 2013.
[34] Mayank Bhargavaa, T. Polzehl, "Improving Automatic Emotion Recognition from speech using Rhythm and Temporal feature." School of Electrical Sciences, Indian Institute of Technology Bhubaneswa, 2012.
[35] Peipei Shen, Z. Changjun., "Automatic Speech Emotion Recognition Using Support Vector Machine." International Conference on Electronic & Mechanical Engineering and Information Technology, 2011.