Published by: Iranian Society for Biomedical Engineering / www.isbme.ir

Iranian Journal of Biomedical Engineering

www.ijbme.org / P-ISSN: 2008-5869 / E-ISSN: 8006-9685

Volume 17, Issue 3, Autumn 2023, 249 - 262

Arousal and Valence Classification of Music Emotion using Music and Demographic Features

Talesh Jafadideh, Alireza

Assistant Professor, Biomedical Engineering Group, School of Engineering Science, College of Engineering, University of Tehran, Tehran, Iran

ARTICLE INFO

DOI: 10.22041/ijbme.2024.2025282.1891

Received: 20 March 2024 Revised: 24 May 2024 Accepted: 1 July 2024

KEYWORDS

Music Emotions

Arousal

Valence

Music and Demographic Features

Classification

ABSTRACT

Two of the most prominent human emotions are arousal and valence. In this article, the aim is to answer the question whether predicting arousal and valence emotions arising from listening to music without using physiological signals and only using demographic and musical characteristics can provide appropriate results?. For this purpose, 48 30-second music with very high and very low levels of arousal and valence were selected from the DEAM music collection. Then, each of these music was separately labeled in terms of arousal and valence emotions by 175 Iranian participants with an age range of 14-35 years. These integer labels were from 1 (the lowest rate) to 5 (the highest rate). The root mean square energy, tempo, zero-crossing, spectral flatness, spectral centroid, spectral flux, spectral rolloff, rhythmic Complexity, and chromagram features were extracted from each music. The demographic features were age, gender, education level, economic level, ethnicity, zip code, and the hours of listening to music in each day. Observations related to label 3 (middle rate) were discarded due to the very low number of occurrences of this label compared to other labels, and 8051 observations were used for classification. The entire data was divided into 4 equal, nonoverlapping parts and classified 4 times so that each time one of the parts was used for testing and the rest parts were used for training the model. This process was repeated 10 times and the average results of the test data were calculated for the classification criteria. The arousal and valence emotions were analyzed separately. For classification performance comparison, five different classifiers including neural network, K nearest neighbors, support vector machine, decision tree, and random forest were taken into account. The neural network offered the best classification performance for arousal emotion by 77% accuracy, 90.3% specificity, 77% sensitivity and valence emotion by 79.7% accuracy, 91.2% specificity, 79.7% sensitivity. The results offer that the neural network can be an appropriate classifier for classification of the musical emotions of Iranian society using the music and demographic features.

***Corresponding Author**

Address: Biomedical Engineering Group, School of Engineering Science, College of Engineering, University of Tehran, Tehran, Iran

Postal Code: *14155-6619* **E-Mail:** *alireza.talesh@ut.ac.ir* **Tel:** *+98-21-61112174*

ناشر: انجمن مهندسی پزشکی ایران / www.isbme.ir

نشریه‌ی مهندسی پزشکی زیستی

شاپای چاپی: ۵۸۶۹-۲۰۰۸ / شاپای الکترونیکی: ۹۶۸۵-۸۰۰۶ / www.ijbme.org

دوره: ۱۷، شماره: ۳، پاییز ۱۴۰۲، ۲۴۹ - ۲۶۲

طبقه‌بندی احساسات برانگیختگی و خوشایندی موسیقی با استفاده از ویژگی‌های موسیقی و جمعیتی

طالش جفادیده، علیرضا

استادیار، گروه مهندسی ورزش، دانشکده‌ی علوم مهندسی، دانشکدگان فنی، دانشگاه تهران، تهران، ایران

مشخصات مقاله

شناسه‌ی دیجیتال: 10.22041/ijbme.2024.2025282.1891

ثبت در سامانه: ۱ فروردین ۱۴۰۳ بازنگری: ۴ خرداد ۱۴۰۳ پذیرش: ۱۱ تیر ۱۴۰۳

واژه‌های کلیدی

احساسات موسیقی
برانگیختگی
خوشایندی
ویژگی‌های موسیقی و جمعیتی
طبقه‌بندی

چکیده

دو مورد از برجسته‌ترین احساسات انسانی، برانگیختگی و خوشایندی است. هدف این مقاله پاسخ دادن به سوال «آیا پیش‌بینی احساسات برانگیختگی و خوشایندی حاصل از گوش دادن به موسیقی بدون استفاده از سیگنال‌های فیزیولوژیک و فقط با استفاده از ویژگی‌های جمعیتی و موسیقایی می‌تواند نتایج مناسبی ارائه دهد؟» است. بدین منظور ۴۸ موسیقی ۳۰ ثانیه‌ای با سطوح برانگیختگی و خوشایندی بسیار بالا و بسیار پایین از مجموعه‌ی موسیقی DEAM انتخاب شده و توسط ۱۷۵ شرکت کننده‌ی ایرانی با محدوده‌ی سنی ۱۸-۳۵ سال بر اساس میزان برانگیختگی و خوشایندی (هر کدام از این دو احساس به طور جداگانه) با یکی از اعداد صحیح ۱ (کم‌ترین) تا ۵ (بیش‌ترین) برچسب‌گذاری شده است. ویژگی‌های موسیقایی انرژی، تمپو، تعداد عبور از صفر، صافی طیفی، مرکز طیفی، شار طیفی، پرتاب طیفی، پیچیدگی ریتمیک و ویژگی‌های کروماگرام و ویژگی‌های جمعیتی سن، جنسیت، میزان تحصیلات، سطح اقتصادی، قومیت، منطقه‌ی شهری و تعداد ساعت گوش دادن به موسیقی در روز، از موسیقی‌ها و افراد شرکت کننده استخراج گردیده است. مشاهدات مربوط به برچسب ۳ (متوسط) به دلیل تعداد بسیار کم رخداد این برچسب نسبت به سایر برچسب‌ها کنار گذاشته شده و ۸۰۵۱ مشاهده برای طبقه‌بندی مورد استفاده قرار گرفته است. کل داده‌ها به ۴ بخش مساوی و جدا از هم (بدون هم‌پوشانی) تقسیم شده و طبقه‌بندی ۴ بار صورت گرفته به طوری که در هر بار یکی از بخش‌ها برای تست و سایر بخش‌های باقی‌مانده برای آموزش مدل به کار گرفته شده است. این فرایند ۱۰ بار تکرار شده و متوسط نتایج داده‌های تست برای معیارهای طبقه‌بندی محاسبه گردیده است. هر کدام از احساسات برانگیختگی و خوشایندی به طور جداگانه آنالیز شده است. برای ساختن مدل طبقه‌بند، ۵ طبقه‌بند شبکه‌ی عصبی، k نزدیک‌ترین همسایه، ماشین بردار پشتیبان، درخت تصمیم و جنگل تصافی به کار گرفته شده است. بهترین عمل‌کرد طبقه‌بندی توسط شبکه‌ی عصبی برای برانگیختگی با صحت ۷۷٪، اختصاصیت ۹۰/۳٪ و حساسیت ۷۷٪ و برای خوشایندی با صحت ۷۹/۷٪، اختصاصیت ۹۱/۲٪ و حساسیت ۷۹/۷٪ به دست آمده است. نتایج نشان می‌دهند که شبکه‌ی عصبی می‌تواند یک طبقه‌بند مناسب برای طبقه‌بندی احساسات موسیقایی جامعه‌ی ایرانی بر اساس ویژگی‌های موسیقی و جمعیتی باشد.

***نویسنده‌ی مسئول**

نشانی: گروه مهندسی ورزش، دانشکده‌ی علوم مهندسی، دانشکدگان فنی، دانشگاه تهران، تهران، ایران

کد پستی: ۱۴۱۵۵-۶۶۱۹ **پست الکترونیک:** *alireza.talesh@ut.ac.ir* **تلفن:** ۶۱۱۱۲۱۷۴-۲۱-۹۸+

۱- مقدمه

احساسات در بسیاری از تجربیات روزانه‌ی انسان امروزی نقش تعیین کننده‌ای ایفا کرده و تاثیر قابل توجهی بر زندگی او در راستای شناخت، ادراک و تصمیم‌گیری منطقی دارد [۱]. تشخیص احساسات موسیقی[۱] (MER) یک زمینه‌ی جوان اما به سرعت در حال گسترش است [۲]. تشخیص احساسات موسیقی فرایندی است برای استفاده از رایانه جهت استخراج و تجزیه و تحلیل ویژگی‌های موسیقی، شکل‌گیری روابط نگاشت بین ویژگی‌های موسیقی و فضای احساسات و استفاده از این روابط به منظور تشخیص احساساتی که توسط موسیقی بیان می‌شود [۳]. ویژگی‌های موسیقی اغلب از سیگنال‌های صوتی، نمرات موسیقی نمادین، متون اشعار و حتی ویژگی‌های فیزیولوژیکی مانند الکتروانسفالوگرام[۲] (EEG) استخراج می‌شود. فضای احساسی را می‌توان با تعداد محدودی از دسته‌بندی‌های گسسته مانند دسته‌بندی برانگیختگی[۳] و خوشایندی[۴] یا دسته‌بندی پیوسته نشان داد [۴-۷]. از MER می‌توان به طور گسترده در بسیاری از زمینه‌ها مانند توصیه‌ی موسیقی، بازیابی اطلاعات موسیقی، تجسم موسیقی، آهنگ‌سازی خودکار موسیقی، روان‌درمانی و غیره استفاده کرد. بنابراین MER به یک کانون تحقیقاتی در جامعه‌ی دانشگاهی و صنعتی تبدیل شده است [۳، ۸].

روش‌های تشخیص احساسات موسیقی عمدتا شامل طبقه‌بندی هیجان و رگرسیون هیجان است [۹]. احساس موسیقی بر اساس مدل حلقه‌ی احساسی هونر[۵] و مدل عاطفی دوبعدی راسل[۶] طبقه‌بندی می‌شود [۱۰]. محققان بر اساس روش پردازش سیگنال صوتی، انرژی، ملودی، هارمونی، ویژگی‌های حوزه‌ی زمان و حوزه‌ی فرکانس و سایر ویژگی‌های موسیقی را استخراج کرده و از طریق روش‌های یادگیری ماشین از جمله ماشین بردار پشتیبان[۷] (SVM)، مدل مخلوط گاوسی، شبکه‌ی عصبی[۸] (NN) و الگوریتم K نزدیک‌ترین همسایه[۹] (KNN) احساسات موسیقی را دسته‌بندی می‌کنند [۹، ۱۱]. یکی دیگر از روش‌های تشخیص احساسات موسیقی، رگرسیون عواطف موسیقی بر اساس مدل احساسات راسل یا تایر[۱۰] است. ویژگی‌های موسیقی، شناختی و هر ویژگی دیگری که مرتبط با تشخیص احساس موسیقی استخراج می‌شود به عنوان ورودی به رگرسیونی مانند رگرسیون خطی چندگانه یا رگرسیون ماشین پشتیبان داده شده تا مقدار برانگیختگی و خوشایندی موسیقی به عنوان خروجی به دست آید [۹، ۱۰، ۱۲، ۱۳].

ژو و ژیا با ترکیب دو رگرسیون ماشین پشتیبان و رگرسیون تکه‌ای k-صفحه‌ای، صحت طبقه‌بندی احساسات موسیقی را نسبت به استفاده از هر کدام ار رگرسورها به طور جداگانه در حدود ۳ تا ۴ درصد و نسبت به استفاده از طبقه‌بند SVM در حدود ۶ درصد بهبود داده‌اند [۱۰]. آگاروال و اوم یک سیستم MER برای تخمین مقدار ارزیابی هر اثر موسیقی ساخته و از روش‌های رگرسیون برای آشکارسازی تغییرات احساسی در موسیقی استفاده کرده‌اند [۱۴]. تورس و هم‌کارانش احساسات موسیقی را از طریق اطلاعات اشعار (لیریک) موسیقی شناسایی کرده‌اند [۱۵]. پانوار و هم‌کارانش با تجزیه و تحلیل اطلاعات کمکی مانند اشعار در موسیقی، احساسات موسیقی را شناسایی کرده‌اند. در این تحقیق هم‌چنین اثرات تشخیص سه مدل طبقه‌بندی پرکاربرد SVM، KNN و مدل مخلوط گاوسی (GMM) مقایسه شده است [۱۶]. بیلال و مورات ۳۴ ویژگی موسیقی از هر فایل موسیقی ترکی را انتخاب کرده و توسط سه طبقه‌بند NN، SVM و KNN حس حاصل از هر موسیقی را طبقه‌بندی نموده‌اند که بهترین صحت برابر با ۷۹/۳٪ و با استفاده از NN به دست آمده است [۱۷]. تاثیر ویژگی‌های موسیقی بر طبقه‌بندی احساسات به طور گسترده توسط سانگ و هم‌کارانش مورد مطالعه قرار گرفته است. در وب‌سایت Last.FM مجموعه‌ی داده‌ای از ۲۹۰۴ آهنگ با برچسب‌های شاد، غمگین، عصبانی و ملایم جمع‌آوری شده است. آن‌ها با استفاده از الگوریتم‌های استاندارد، ویژگی‌های مختلف صدا را استخراج نموده و برای طبقه‌بندی، مجموعه‌ی داده را با استفاده از SVM با کرنل تابع چندجمله‌ای و شعاعی و اعتبارسنجی متقابل ۱۰ افزاره آموزش داده‌اند. آن‌ها گزارش داده‌اند که ویژگی‌های طیفی موسیقی با توجه به نتایج به دست آمده عمل‌کرد بهتری از خود نشان داده است [۱۸]. پاندا و هم‌کارانش ویژگی‌های استاندارد و ملودیک استخراج شده از سیگنال‌های صوتی را برای تشخیص احساسات موسیقی ترکیب کرده‌اند. آن‌ها یک مجموعه‌ی داده‌ی صوتی جدید برای طبقه‌بندی احساسات موسیقی تهیه کرده‌اند. آن‌ها برای هر موسیقی در مجموعه‌ی داده، ۲۵۳ ویژگی استاندارد و ۹۸ ویژگی ملودیک را استخراج نموده و تشخیص احساسات را با استفاده از

---

۱ Music Emotion Recognition
۲ Electroencephalogram
۳ Arousal
۴ Valence
۵ Hevner
۶ Russell
۷ Support Vector Machine
۸ Neural Network
۹ K-Nearest Neighbors
۱۰ Thayer

الگوریتم‌های طبقه‌بندی مختلف انجام داده‌اند. آن‌ها هم‌چنین از روش‌های انتخاب ویژگی مختلف استفاده کرده‌اند. با توجه به نتایج تجربی، مشاهده شده که عمل‌کرد ویژگی‌های ملودیک بهتر از ویژگی‌های استاندارد است. بهترین نتیجه به عنوان ۶۴٪ معیار F با استفاده از روش انتخاب ویژگی ReliefF و طبقه‌بند SVM به دست آمده است [۱۹]. یانگ و هم‌کارانش نیز در کار مروری خود گزارش داده‌اند که استفاده از ویژگی‌های طیفی موسیقی نسبت به ویژگی‌های مربوط به ریتم و دینامیک موسیقی عمل‌کرد بهتری در دسته‌بندی احساسات موسیقی ارائه می‌دهد [۲۰]. هر چند باید به این نکته نیز اشاره کرد که در ۶۰٪ مطالعات مربوط به دسته‌بندی احساسات موسیقی از ویژگی‌های طیفی استفاده شده و معمولا درصد صحت به دست آمده برای حس خوشایندی نسبت به برانگیختگی به طور متوسط کم‌تر بوده است [۶، ۲۰].

هدف این مقاله بررسی این موضوع است که فقط با استفاده از ویژگی‌های موسیقی و ویژگی‌های جمعیتی افراد شرکت کننده‌ی ایرانی با چه صحتی می‌توان احساسات برانگیختگی و خوشایندی حاصل از گوش دادن به یک موسیقی را پیش‌بینی کرد. برای این منظور مراحل زیر انجام شده است.

۱- تعداد ۴۸ موسیقی ۳۰ ثانیه‌ای با سطوح برانگیختگی و خوشایندی بسیار بالا و بسیار پایین از مجموعه‌ی داده‌ی DEAM[۱] انتخاب شده است [۲۱].

۲- هر موسیقی انتخاب شده برای ۱۷۵ شرکت کننده‌ی ایرانی پخش شده و هر شرکت کننده به هر موسیقی دو برچسب، یکی برای برانگیختگی و دیگری برای خوشایندی نسبت داده که شامل اعداد صحیح از ۱ (کم‌ترین) تا ۵ (بیش‌ترین) است.

۳- ویژگی‌های موسیقی و جمعیتی افراد شرکت کننده استخراج گردیده است.

۴- ویژگی‌ها و برچسب‌های شرکت کنندگان برای ارزیابی عمل‌کرد پنج طبقه‌بند پرکاربرد NN، SVM، KNN، درخت تصمیم (DT) و جنگل تصادفی (RF) به کار گرفته شده است تا بهترین طبقه‌بند و بهترین صحت مشخص شود.

هدف نهایی و کاربردی این تحقیق این است که با داشتن ویژگی موسیقایی یک موسیقی و دریافت ویژگی جمعیتی یک فرد ایرانی به عنوان ورودی بتوان پیش‌بینی کرد که این فرد در صورت گوش دادن به این موسیقی چه احساسی از لحاظ برانگیختگی و خوشایندی خواهد داشت. به عبارت دیگر فقط با ویژگی‌های جمعیتی و موسیقی و بدون نیاز به ثبت سیگنال فیزیولوژیکی که سختی‌ها، چالش‌ها و هزینه‌های خاص خود را دارد تا چه صحتی می‌توان احساسات برانگیختگی و خوشایندی حاصل از گوش دادن به موسیقی را برای افراد ایرانی پیش‌بینی کرد. البته که این یک تلاش اولیه و محدود است اما شاید بتواند چشم‌انداز روشنی را برای تحقق این هدف نهایی فراهم آورد.

در ادامه‌ی مقاله درباره‌ی موسیقی‌های مورد استفاده، پروتکل آزمایش و نحوه‌ی جمع‌آوری برچسب موسیقی‌ها از شرکت کنندگان ایرانی توضیحاتی داده شده است. سپس ویژگی‌های استخراج شده از موسیقی و ویژگی‌های جمعیتی شرکت کنندگان معرفی شده است. آماده‌سازی ویژگی و نحوه‌ی ارزیابی عمل‌کرد پنج طبقه‌بند NN، SVM، KNN، DT و RF و پارامترهای بهینه‌ی هر یک گزارش شده است. در بخش بعد نتایج حاصل از طبقه‌بندی گزارش شده و پس از آن بحث روی نتایج و محدودیت‌های کار ارائه شده است. در انتها نیز جمع‌بندی و مشخصات مراجع مورد استفاده آورده شده است.

۲- راه‌اندازی آزمایش

۲-۱- داده‌ی موسیقی مورد استفاده

موسیقی‌های مورد استفاده در این مطالعه از مجموعه‌ی داده‌ی DEAM انتخاب شده است [۲۲، ۲۳]. این مجموعه یک پایگاه داده برای تجزیه و تحلیل احساس موسیقی بوده که دارای یک مجموعه‌ی داده‌ی معیار برای وظیفه[۲] تشخیص احساسات موسیقی است. موسیقی‌های این پایگاه داده از چندین منبع مانند FMA، Jamendo و مجموعه‌ی داده‌ی MedleyDB انتخاب شده است. این پایگاه داده شامل ۱۸۰۲ قطعه‌ی موسیقی (۵۸ آهنگ کامل و ۱۷۴۴ گزیده‌ی ۴۵ ثانیه‌ای) است. گردآورندگان این مجموعه‌ی داده موسیقی‌ها را مجددا نمونه‌برداری کرده تا فرکانس نمونه‌برداری تمام موسیقی‌ها برابر با ۴۴۱۰۰ هرتز شود. نویسندگان DEAM پس از مشاهده‌ی ناپایداری‌های زیاد به دلیل واریانس زیاد در برچسب‌های اختصاص داده شده تصمیم گرفته‌اند که ۱۵ ثانیه‌ی اول قطعه‌های موسیقی را کنار بگذارند. به این دلیل و بر اساس این واقعیت که طول اکثر نمونه‌های موسیقی تنها ۴۵ ثانیه بوده، برای هر نمونه‌ی موسیقی فقط ۳۰ ثانیه انتخاب شده که از ثانیه‌ی ۱۵ شروع و در ثانیه‌ی ۴۵ تمام شده است [۲۱، ۲۳]. هر موسیقی حداقل توسط ۵ نفر به صورت پویا برچسب‌گذاری شده است (برای هر ۰/۵ ثانیه از موسیقی یک برچسب). برچسب‌گذاری برای دو حس برانگیختگی و خوشایندی و از بین اعداد ۱۰- تا ۱۰ صورت گرفته است. برچسب ایستای یک قطعه‌ی موسیقی با متوسط‌گیری از برچسب‌های پویای آن

[۱] Dataset for Emotional Analysis of Music

[۲] Task

قطعه به دست آمده و در مقیاس ۹ نقطه‌ای (۱-۹) برای برانگیختگی و خوشایندی مقیاس‌بندی شده است. برای این مطالعه، برچسب‌گذاری‌های ایستا در نظر گرفته شده است. در مدل احساسات راسل معمولا ۴ حالت پرکاربرد به صورت برانگیختگی بالا و خوشایندی بالا، برانگیختگی بالا و خوشایندی پایین، برانگیختگی پایین و خوشایندی بالا و برانگیختگی پایین و خوشایندی پایین وجود دارد. در این مطالعه برای هر حالت ۱۲ موسیقی انتخاب شده است. در نتیجه ۲۴ موسیقی برای برانگیختگی بالا و ۲۴ موسیقی برای برانگیختگی پایین به دست آمده است. برای خوشایندی نیز به همین تعداد برای سطح بالا و پایین انتخاب شده است. برای سطح بالا/پایین برچسب‌های با محدوده‌ی مقادیر (۹-۷)/(۳-۱) در نظر گرفته شده است.

۲-۲- پروتکل آزمایش

تعداد ۱۷۵ فرد سالم (۱۰۹ مرد) با میانگین، انحراف معیار و بازه‌ی سنی به ترتیب ۲۹، ۴/۳۷، ۱۸-۳۵ سال در این آزمایش شرکت کرده‌اند. قبل از آزمایش گوش دادن و نمره دادن به موسیقی‌ها بر اساس دو حس خوشایندی و برانگیختگی، هر یک از افراد یک فرم رضایت‌نامه را امضا نموده و یک پرسش‌نامه را تکمیل کرده‌اند. با این پرسش‌نامه، اطلاعات جمعیت‌شناسی از افراد دریافت شده است. سپس مجموعه‌ای از دستورالعمل‌ها به آن‌ها داده شده است تا از پروتکل آزمایش و معنای مقیاس‌های مختلف مورد استفاده برای ارزیابی دو حس خوشایندی و برانگیختگی مطلع شوند. پس از درک دستورالعمل‌ها شرکت کننده برای گوش دادن به موسیقی‌ها و برچسب‌گذاری به اتاق آزمایش هدایت شده است. در ابتدا جهت آشنایی شرکت کننده با سیستم، یک آزمایش تمرینی انجام شده است. در این کارآزمایی ثبت نشده، یک نمونه‌ی موسیقی پخش شده و پس از آن فرد نمره‌دهی دو حس خوشایندی و برانگیختگی حاصل از گوش دادن به قطعه‌ی موسقی را انجام داده است. سپس آزمایش اصلی با پخش ۴۸ قطعه‌ی موسیقی آغاز شده است.

آزمایش با خط پایه به مدت ۲ دقیقه شروع شده که طی آن یک علامت + به شرکت کننده نمایش داده شده و از او خواسته شده تا در این دوره کاری انجام ندهد. سپس ۴۸ موزیک در ۴۸ کارآزمایی ارائه شده که هر کدام شامل مراحل زیر است.

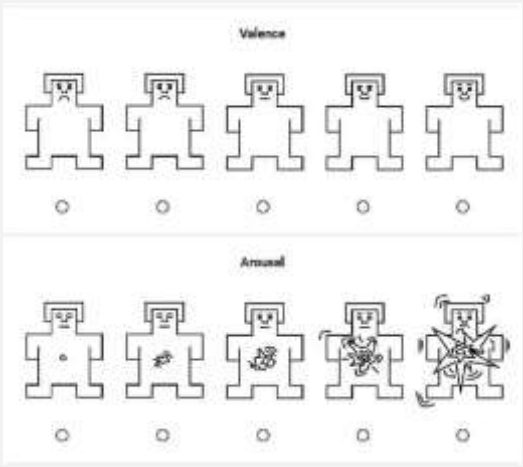
۱- روشن شدن صفحه‌ی نمایش ۲ ثانیه‌ای که شماره‌ی آزمایش فعلی را نشان داده تا افراد از میزان پیش‌رفت آزمایش آگاه شوند

۲- پخش ۳۰ ثانیه موسیقی

۳- خودارزیابی برای برانگیختگی و خوشایندی (۸ ثانیه)

۲-۳- برچسب‌گذاری موسیقی

در پایان هر کارآزمایی، خودارزیابی سطوح برانگیختگی و خوشایندی توسط خود شرکت کنندگان انجام شده است. از آدمک‌های خودارزیابی[۱] (SAM) [۲۴] برای تجسم مقیاس‌ها استفاده شده است (شکل ۱). آدمک‌ها در وسط صفحه نمایش داده شده و اعداد ۱ تا ۵ در زیر آن‌ها با دایره مشخص شده است. شرکت کنندگان موشواره را دقیقا به صورت افقی و درست زیر اعداد حرکت داده و برای نشان دادن سطح خودارزیابی کلیک کرده‌اند. مقیاس خوشایندی از غمگین تا خوشحال و مقیاس برانگیختگی از آرام یا کسل کننده تا تحریک شده یا هیجان زده متغیر است.



شکل (۱)- تصاویر مورد استفاده برای خودارزیابی خوشایندی (بالا) و برانگیختگی (پایین) [۲۵]

۳- استخراج ویژگی و طبقه‌بندی

برچسب‌های اختصاص داده شده به بعضی از موسیقی‌ها توسط تعدادی از شرکت کنندگان ثبت نشده است. هم‌چنین تعداد برچسب‌های مربوط به کلاس ۳ (متوسط) در مقایسه با سایر برچسب‌ها بسیار کم‌تر بوده (در حدود ۳۰۰) که باعث کاهش عمل‌کرد طبقه‌بندی شده است. از این رو مشاهدات بدون برچسب و با برچسب ۳ کنار گذاشته شده است. در نهایت ماتریس ویژگی دارای ۸۰۵۱ سطر (مشاهده) با برچسب‌های {۱،۲،۴،۵} برای هر یک از دو حس برانگیختگی و خوشایندی شده است. تعداد ستون‌های این ماتریس برابر با تعداد ویژگی یعنی ۳۹ بوده که شامل ۳۲ ویژگی موسیقی و ۷ ویژگی جمعیتی است. در ادامه این ۳۹ ویژگی معرفی شده است. این ماتریس ویژگی وارد فرایند طبقه‌بندی شده که در ادامه این فرایند نیز توضیح داده شده است.

[۱] Self-Assessment Manikins

۳-۱- استخراج ویژگی از موسیقی

در این مطالعه ویژگی‌های موسیقایی ریشه‌ی میانگین مربعات انرژی، تمپو، تعداد عبور از صفر[۱]، صافی طیفی[۲]، مرکز طیفی[۳]، شار طیفی[۴]، پرتاب طیفی[۵]، پیچیدگی ریتمیک و ویژگی‌های کروماگرام[۶] از موسیقی‌ها استخراج گردیده است. این ویژگی‌ها بر اساس کدهای تحت نرم‌افزار MATLAB با عنوان LabROSA-coversongID و پارامترهای پیش‌فرض آن محاسبه شده است [۲۶، ۲۷]. در ادامه اطلاعات بیش‌تری برای هر ویژگی فراهم شده است.

۳-۱-۱- ریشه‌ی میانگین مربعات انرژی

این ویژگی بر اساس تمام نمونه‌ها در یک فریم محاسبه می‌شود. دامنه‌ی کلی یک سیگنال با انرژی آن مطابقت دارد. برای سیگنال‌های صوتی، این ویژگی به طور کلی برابر با میزان بلندی سیگنال است زیرا با انرژی بالاتر، صدا بلندتر می‌باشد. در مقایسه با پوش دامنه حساسیت کم‌تری نسبت به موارد پرت دارد. این ویژگی در تقسیم‌بندی صدا و وظایف طبقه‌بندی ژانر موسیقی مفید است.

۳-۱-۲- تمپو

در موسیقی، تمپو به عنوان ضرب در دقیقه شناخته می‌شود که به سرعت پخش یک قطعه‌ی موسیقی اشاره دارد. هر چقدر تمپو بیش‌تر باشد سرعت اجرای موسیقی بیش‌تر است.

۳-۱-۳- تعداد عبور از صفر

این ویژگی بیان‌گر تعداد دفعاتی است که یک شکل موج از محور افقی زمان عبور می‌کند. این ویژگی اساسا در تشخیص صداهای ضربی در مقابل صداهای همراه با زیر و بمی، تخمین صدای تک‌صدایی، تصمیم‌گیری قسمت‌های صدا/بی‌صدا برای سیگنال‌های گفتاری و ... مورد استفاده قرار گرفته است.

۳-۱-۴- صافی طیفی

صافی طیفی یا ضریب تونالیته که به عنوان آنتروپی وینر نیز شناخته می‌شود معیاری است که در پردازش سیگنال دیجیتال برای مشخص کردن طیف صوتی به کار می‌رود. صافی طیفی معمولا بر حسب دسی‌بل اندازه‌گیری شده و روشی را برای تعیین کمیت یک صدا ارائه داده که مشخص شود چقدر شبیه یک تن خالص و نه شبیه نویز است زیرا مسطح بودن زیاد صافی طیفی نشان می‌دهد که طیف در تمام باندهای طیفی دارای قدرت مشابهی بوده و نمودار طیف نسبتا مسطح و صاف به نظر می‌رسد که این شبیه به نویز سفید و نه شبیه به یک تن خالص می‌باشد. صافی طیفی با تقسیم میانگین هندسی طیف توان بر میانگین حسابی طیف توان محاسبه می‌شود [۲۸].

۳-۱-۵- مرکز طیفی

مرکز طیفی معیاری است که از آن در پردازش سیگنال دیجیتال برای مشخص کردن یک طیف استفاده می‌شود. این معیار نشان می‌دهد که مرکز جرم طیف در کجا قرار دارد. این معیار از نظر ادراکی، ارتباط قوی با تاثیر روشنایی صدا دارد به طوری که صداهای با مرکز بالاتر (مانند ترومپت) نسبت به صداهای با مرکز پایین‌تر (مانند توبا) روشن‌تر درک می‌شوند.

۳-۱-۶- شار طیفی

شار طیفی یک معیار برای تشخیص سیگنال‌هایی که طیف آن‌ها به کندی تغییر می‌کند از سیگنال‌هایی که طیف آن‌ها به سرعت تغییر می‌کند است. برای سیگنال‌های کلاس آهسته این معیار مقدار کم‌تری دارد و برای سیگنال‌های کلاس سریع مقدار این معیار بزرگ‌تر است. یکی از مزیت‌های بالقوه‌ی شار طیفی نسبت به پوشش یا مشتق آن این است که شار طیفی به اطلاعات ریتمیک حساس است که با تغییرات گام[۷] منتقل می‌شود حتی زمانی که با تغییرات دامنه همراه نباشد [۲۹].

۳-۱-۷- پرتاب طیفی

پرتاب طیفی فرکانسی است که در زیر آن درصد مشخصی از کل انرژی طیفی (در این مطالعه ۹۵٪) قرار دارد. از نقطه‌ی پرتاب طیفی برای تمایز بین گفتار صدادار و بدون صدا، تمایز گفتار/موسیقی، طبقه‌بندی ژانر موسیقی، تشخیص صحنه‌ی آکوستیک و طبقه‌بندی حالت موسیقی استفاده می‌شود [۳۰].

۳-۱-۸- پیچیدگی ریتمیک

این ویژگی برابر با انحراف معیار مشتق (دستور diff در نرم‌افزار MATLAB) سیگنال صوتی است. هر چقدر تغییرات سیگنال بیش‌تر باشد پیچیدگی آن نیز بیش‌تر بوده و مقدار این ویژگی بیش‌تر خواهد بود.

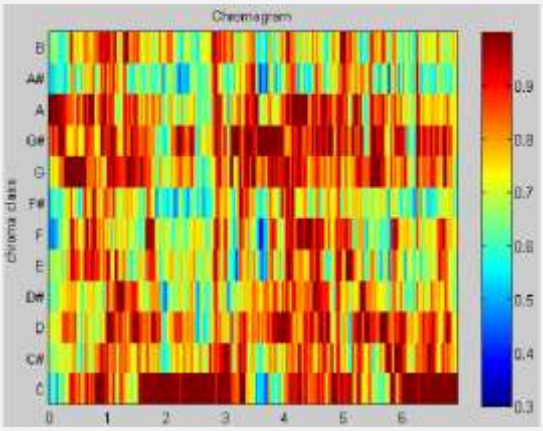
---

۱ Zero Crossing
۲ Spectral Flatness
۳ Spectral Centroid
۴ Spectral Flux
۵ Spectral Rolloff
۶ Chromagram
۷ Pitch

۳-۱-۹- کروماگرام

کروماگرام به عنوان کل اطلاعات صوتی طیفی نگاشت شده در یک اکتاو تعریف می‌شود. هر اکتاو به ۱۲ کلاس گام تقسیم می‌شود. تصویر کروماگرام نشان دهنده‌ی توزیع انرژی محتوای فرکانس سیگنال در ۱۲ کلاس گام است (شکل ۲) [۳۱]. این ۱۲ گام به صورت زیر است.

{C, C♯, D, D♯, E, F, F♯, G, G♯, A, A♯, B}



شکل (۲)- یک نمونه‌ی کروماگرام، محور عمودی بیان‌گر ۱۲ گام و محور افقی بیان‌گر شماره‌ی فریم‌های مختلف زمانی داده‌ی صوتی بوده که در این‌جا برای ۶ فریم نشان داده شده است

یکی از ویژگی‌های اصلی کروماگرام این است که ویژگی‌های هارمونیک و ملودیک موسیقی را به تصویر می‌کشد در حالی که این ویژگی‌ها در برابر تغییرات صدا و ساز قوی هستند. هدف ویژگی‌های کروما نمایش محتوای هارمونیک (مانند کلیدها، آکوردها) یک پنجره‌ی صوتی کوتاه‌مدت است.

داده‌ی صوتی به پنجره‌های زمانی کوتاه تقسیم شده و برای هر پنجره کروما برای ۱۲ پیچ محاسبه گردیده و با کنار هم قرار دادن کرومای پنجره‌های مختلف و رسم آن، کروماگرام ساخته شده است. در این مطالعه زمان ضرب‌های فایل صوتی پیدا شده و پنجره‌ی زمانی در واقع بین زمان دو ضرب متوالی تعریف شده است. یک ضرب به عنوان یک نبض منظم و تکرار شونده تعریف شده که زیربنای یک الگوی موسیقی است. ریتم در موسیقی با یک توالی تکراری از ضربات با استرس و بدون استرس مشخص می‌شود. با توجه به این تعریف ضرب و بنیادی بودن آن تصمیم گرفته شده است تا از زمان ضرب برای تعریف پنجره‌ی زمانی استفاده گردد. سطر و ستون ماتریس نهایی کروماگرام به ترتیب مربوط به گام و زمان بوده که یک نمونه از تصویر مربوط به این ماتریس در شکل (۲) نشان داده شده است. این ماتریس بر بزرگ‌ترین مقدار خود تقسیم شده است تا مقادیر ماتریس بین صفر و ۱ قرار گیرد. دو نوع ویژگی برای استخراج از کروماگرام تعریف شده که هر نوع برای هر گام جداگانه محاسبه گردیده است. در نتیجه از کروماگرام در مجموع ۲۴ ویژگی استخراج شده است. نوع اول مجموع کروما و نوع دوم مرکز جرم زمانی کروما بوده که هر کدام برای هر گام جداگانه محاسبه شده است.

۳-۱-۱۰- مجموع کروما

برای هر سطر (گام) ماتریس، مجموع مقادیر کرومای آن سطر که بیش‌تر از ۰/۸ بوده محاسبه شده است. مقدار آستانه برابر با ۰/۸ در نظر گرفته شده تا صرفا لحظاتی (ستون‌هایی) در محاسبه‌ی انرژی گام درگیر شوند که بیش‌ترین انرژی را نسبت به کل ماتریس کروماگرام دارند (رنگ قرمز در شکل ۲) تا از این طریق مشخص شود که برای موسیقی پخش شده کدام گام بیش‌ترین انرژی را داشته است.

۳-۱-۱۱- مرکز جرم زمانی کروما

برای هر سطر (گام) میانه‌ی لحظاتی که شدت کروما بیش‌تر از ۰/۸ بوده محاسبه شده تا مشخص شود در طول مدت ۳۰ ثانیه پخش موسیقی، به طور متوسط تمرکز انرژی در چه زمانی بوده است. دلیل استفاده از میانه حساس نبودن به داده‌ی پرت است.

۳-۲- استخراج ویژگی‌های جمعیتی از افراد

در این مطالعه ویژگی‌های جمعیتی شامل سن، جنسیت، میزان تحصیلات، سطح اقتصادی، قومیت، منطقه‌ی شهری و تعداد ساعت گوش دادن به موسیقی در روز از افراد استخراج شده که در ادامه اطلاعات بیش‌تری برای هر ویژگی ارائه شده است.

۳-۲-۱- سن

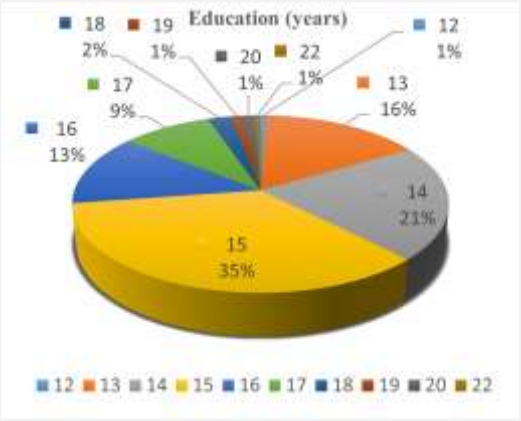
میانگین، انحراف معیار و محدوده‌ی سنی افراد شرکت کننده به ترتیب برابر با ۲۹، ۴/۳۷ و ۱۸-۳۵ سال است.

۳-۲-۲- جنسیت

در مجموع تعداد ۱۷۵ نفر شامل ۱۰۹ مرد و ۶۶ زن در این مطالعه شرکت کرده‌اند.

۳-۲-۳- میزان تحصیلات

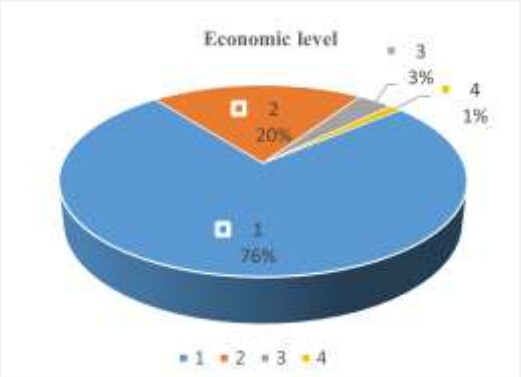
تعداد سال تحصیلی شرکت کنندگان در بازه‌ی ۱۲ تا ۲۲ سال است که حدود ۷۵٪ از آن‌ها دارای ۱۳-۱۷ سال سابقه‌ی تحصیل بوده‌اند. تعداد سال تحصیلی بر حسب درصد تعداد شرکت کنندگان در شکل (۳) نشان داده شده است.



شکل (۳)- سال تحصیلی بر حسب درصد شرکت کنندگان

۳-۲-۴- سطح اقتصادی

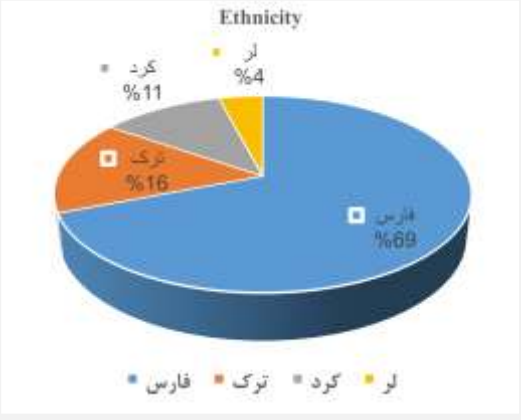
سطح اقتصادی بر اساس میزان درآمد کم‌تر از ۱۰، بین ۱۰ تا ۲۰، بین ۲۰ تا ۳۰ و بیش‌تر از ۳۰ میلیون تومان به چهار سطح ۱، ۲، ۳ و ۴ تقسیم شده است. در شکل (۴) سطح اقتصادی بر حسب درصد تعداد شرکت کنندگان نشان داده شده است.



شکل (۴)- سطح اقتصادی بر حسب درصد شرکت کنندگان

۳-۲-۵- قومیت

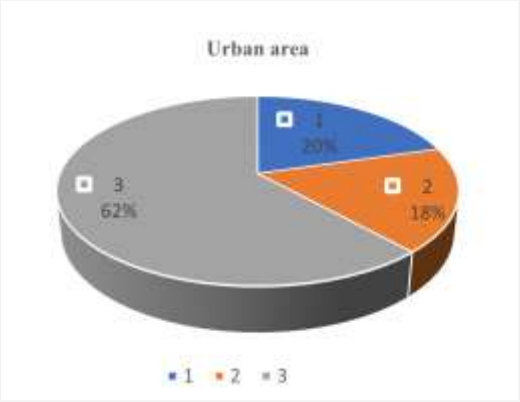
شرکت کنندگان در این آزمایش از چهار قومیت فارس، ترک، کرد و لر بوده‌اند (شکل ۵).



شکل (۵)- درصد حضور شرکت کنندگان بر اساس قومیت

۳-۲-۶- منطقه‌ی شهری

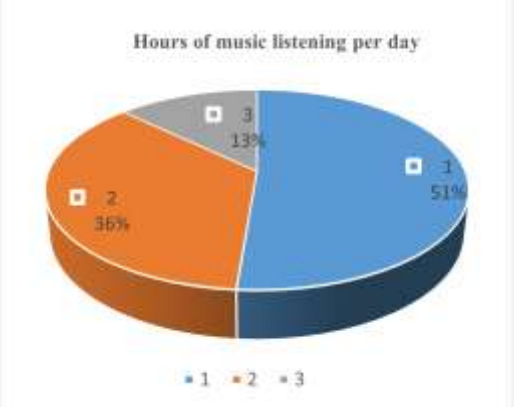
محل سکونت اکثر شرکت کنندگان در زمان آزمایش شهر تهران بوده است. بر اساس این که فرد ساکن کدام منطقه بوده سه برچسب ۱ تا ۳ به ویژگی مربوط به منطقه‌ی شهری اختصاص داده شده است. بر این اساس منطقه‌ی ۱-۳ دارای برچسب ۱، منطقه‌ی ۴-۱۰ دارای برچسب ۲ و منطقه‌ی ۱۰-۲۲ و یا سایر شهرها دارای برچسب ۳ است. با اختصاص این برچسب‌ها به جای شماره‌ی مناطق، نتایج طبقه‌بندی بهتری حاصل شده به همین دلیل چنین دسته‌بندی اتخاذ گردیده است. درصد حضور شرکت کنندگان در این سه دسته‌ی مختلف مناطق شهری در شکل (۶) نشان داده شده است.



شکل (۶)- درصد حضور افراد شرکت کننده در سه دسته‌ی مختلف مناطق شهری

۳-۲-۷- تعداد ساعت گوش دادن به موسیقی در روز

تعداد ساعت‌های گوش دادن به موسیقی به سه دسته‌ی کوچک‌تر یا مساوی با ۱ ساعت (دسته‌ی ۱)، بین ۱ تا ۳ ساعت (دسته‌ی ۲) و بیش‌تر از ۳ ساعت (دسته‌ی ۳) تقسیم‌بندی شده که جزئیات آن در شکل (۷) قابل مشاهده است.



شکل (۷)- ویژگی تعداد ساعت گوش دادن به موسیقی در روز بر اساس درصد حضور شرکت کنندگان

۳-۳- طبقه‌بندی احساسات موسیقی

قبل از شروع فرایند طبقه‌بندی باید هر ستون ماتریس ویژگی با استفاده از رابطه‌ی (۱) نرمالیزه شود تا مقادیر هر ستون بین صفر و ۱ قرار گیرد.

$$x_{i,j,n} = \frac{x_{i,j} - min(x_j)}{max(x_j) - min(x_j)} \quad (۱)$$

در این رابطه $x_{i,j}$ درایه‌ی i-ام از ستون j-ام ماتریس ویژگی، min و max توابع محاسبه‌ی کمینه و بیشینه‌ی مطلق مقادیر ستون j-ام و $x_{i,j,n}$ مقدار درایه‌ی i-ام از ستون j-ام ماتریس ویژگی بعد از نرمال‌سازی است. از آن‌جا که دامنه‌ی تغییرات و مقیاس اندازه‌گیری هر یک از ویژگی‌های مشاهدات با یک‌دیگر متفاوت هستند، نرمال‌سازی اهمیت زیادی پیدا می‌کند زیرا این کار باعث می‌شود که تابع فاصله به سمت ویژگی یا متغیر با مقیاس بزرگ‌تر منحرف نشود. برای کاهش بعد نیز چندین روش و هم‌چنین آنالیز مولفه‌های اساسی بررسی شده که تاثیری در بهبود عمل‌کرد بهینه‌ی به دست آمده در این تحقیق نداشته و به همین دلیل در این‌جا به آن‌ها اشاره نشده است.

در این مطالعه به منظور طبقه‌بندی احساسات از پنج طبقه‌بند back propagation NN، SVM، KNN، DT و RF استفاده شده است. این پنج طبقه‌بند از پرکاربردترین طبقه‌بندها در زمینه‌ی طبقه‌بندی احساسات هستند [۳۲]. طبقه‌بندی و ارزیابی طبقه‌بندها برای هر یک از دو حس برانگیختگی و خوشایندی به طور جداگانه صورت گرفته است. تمام مراحل طبقه‌بندی مانند تنظیم فراپارامتر، آموزش و آزمایش مدل‌ها با استفاده از نرم‌افزار MATLAB R2023b انجام شده است.

کل داده‌ها به ۴ بخش مساوی و جدا از هم (بدون هم‌پوشانی) تقسیم شده و فرایند طبقه‌بندی ۴ بار صورت گرفته به طوری که در هر بار یکی از بخش‌ها برای تست و سایر بخش‌های باقی‌مانده برای آموزش مدل به کار گرفته شده است. با این کار هر یک از مشاهدات یک بار به عنوان تست در نظر گرفته شده است. برای هر طبقه‌بند، با استفاده از داده‌ی آموزش و روش اعتبارسنجی متقابل ۱۰ افزاره، مدل ساخته شده است. عمل‌کرد مدل بهینه‌ی ساخته شده، با استفاده از داده‌ی تست مورد ارزیابی قرار گرفته است. در هر بار آموزش، مدل از ابتدا و بدون استفاده از اطلاعات آموزش‌های قبلی مورد آموزش قرار گرفته است. نتایج طبقه‌بندی ۴ بخش تست کنار هم قرار داده شده است تا نتایج طبقه‌بندی کل داده‌ی مورد استفاده به دست آید.

با استفاده از این نتایج کل، معیارهای طبقه‌بندی محاسبه شده است. این فرایند ۱۰ بار تکرار شده و متوسط معیارهای طبقه‌بندی در بخش نتایج گزارش شده است.

۳-۴- معیارهای ارزیابی طبقه‌بندها

از سه معیار طبقه‌بندی محبوب شامل صحت[۱]، حساسیت[۲] و اختصاصیت[۳] برای ارزیابی عمل‌کرد طبقه‌بندی استفاده شده است. صحت یعنی مدل تا چه اندازه خروجی را درست پیش‌بینی کرده است. حساسیت توانایی مدل را برای شناسایی صحیح موارد مثبت ارزیابی می‌کند. اختصاصیت نیز ظرفیت مدل را برای تشخیص صحیح موارد منفی ارزیابی می‌کند. این پارامترها با استفاده از روابط (۲) تا (۴) محاسبه می‌شوند.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (۲)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (۳)$$

$$Specificity = \frac{TN}{TN + FP} \quad (۴)$$

در این روابط TP مثبت درست، FP مثبت اشتباه، TN منفی درست و FN منفی اشتباه است. در نهایت از ماتریس درهم‌ریختگی[۴] (CM) برای برجسته کردن بیش‌تر و نشان دادن بهتر عمل‌کرد طبقه‌بندی برای برچسب‌های مختلف استفاده شده است. این ماتریس تفکیک دقیقی از پیش‌بینی‌های طبقه‌بندی کننده برای هر برچسب را ارائه داده و تعداد TP، TN، FP و FN را نشان می‌دهد. این دانه‌بندی و ریزریز کردن نتایج امکان تحلیل عمیق عمل‌کرد مدل را فراهم می‌کند.

۴- نتایج و بحث

در این مطالعه طبقه‌بندی احساسات موسیقی از لحاظ دو حس برانگیختگی و خوشایندی مورد بررسی و ارزیابی قرار گرفته است. برای این منظور ۴۸ موسیقی با سطوح برانگیختگی بسیار بالا و پایین و سطوح خوشایندی بسیار بالا و پایین از مجموعه‌ی داده‌ی DEAM انتخاب شده و توسط ۱۷۵ شرکت کننده‌ی ایرانی به هر موسیقی دو برچسب یکی برای سطح برانگیختگی و دیگری برای سطح خوشایندی اختصاص داده شده است. هر برچسب اختصاص داده شده شامل یکی از چهار مقدار ۱ (بسیار کم)، ۲ (کم)، ۴ (زیاد) و ۵ (بسیار زیاد) بوده و مشاهدات با برچسب ۳ (متوسط) به دلیل رخداد بسیار کم کنار گذاشته

[۱] Accuracy
[۲] Sensitivity
[۳] Specificity
[۴] Confusion Matrix

شده است. در نهایت ۸۰۵۱ مشاهده‌ی برچسب‌دار به دست آمده که هر مشاهده دارای ۳۹ ویژگی شامل ۳۲ ویژگی موسیقی و ۷ ویژگی جمعیت‌شناسی است. این ۸۰۵۱ مشاهده به پنج طبقه‌بند پرکاربر در زمینه‌ی تشخیص احساسات شامل back propagation NN، SVM، KNN، DT و RF داده شده است تا عمل‌کرد آن‌ها در پیش‌بینی برچسب‌ها مشخص شود. سوال مطرح شده برای انجام این پژوهش این بوده که با داشتن فقط ویژگی موسیقی و جمعیت‌شناسی و بدون استفاده از ویژگی‌های حاصل از سیگنال‌های فیزیولوژیکی که مستلزم زمان، هزینه، انرژی و سروکار داشتن با چالش‌های مربوط به خود است تا چه میزان می‌توان در پیش‌بینی احساسات برانگیختگی و خوشایندی حاصل از گوش دادن به موسیقی برای یک فرد ایرانی موفق بود؟ البته پاسخ جامع و مطمئن به این سوال به طوری که بتواند در مرحله‌ی عمل و کاربرد با آرامش خاطر بالا مورد استفاده قرار گیرد نیازمند در نظر گرفتن شرایط و پارامترهای زیاد و استفاده از تعداد افراد و موسیقی بیش‌تر می‌باشد اما امید است که این مطالعه و نتایج حاصل از آن بتواند چشم‌انداز امیدبخشی برای پاسخ دادن به این سوال و انگیزه‌ی لازم برای انجام تحقیقات بیش‌تر فراهم کند.

در این بخش پارامترهای بهینه‌ی طبقه‌بندها و نتایج معیارهای طبقه‌بندی یعنی صحت، حساسیت و اختصاصیت و ماتریس درهم‌ریختگی برای پنج طبقه‌بند back propagation NN، SVM، KNN، DT و RF گزارش شده است. در ادامه نتایج حاصل از صرفا استفاده از ویژگی‌های موسیقی و صرفا استفاده از ویژگی‌های جمعیت‌شناسی ارائه شده و در آخر محدودیت‌ها و پیشنهادات برای کارهای آینده مطرح شده است.

هر کدام از دو حس برانگیختگی و خوشایندی دارای چهار کلاس ۱، ۲، ۴، و ۵ است. نتایج معیارهای طبقه‌بندی صحت، حساسیت و اختصاصیت برای هر یک از دو حس و برای پنج طبقه‌بند مورد استفاده در این مطالعه در جدول (۱) گزارش شده است. بهترین و بدترین عمل‌کرد برای این سه معیار و برای هر دو حس به ترتیب برای NN و SVM است. مقادیر صحت، حساسیت و اختصاصیت برای حس برانگیختگی با NN به ترتیب برابر با ۷۷٪، ۷۷٪ و ۹۰/۳٪ و برای حس خوشایندی با NN به ترتیب برابر با ۷۹/۷٪، ۷۹/۷٪ و ۹۱/۲٪ به دست آمده است. سطح عمل‌کرد طبقه‌بند SVM در هر سه معیار تقریبا ۲۰٪ تا ۳۰٪ پایین‌تر از NN بوده است. نزدیک‌ترین عمل‌کرد به NN توسط RF با حدود ۱٪ تا ۲٪ کم‌تر ارائه شده است. دو طبقه‌بند KNN و DT نیز عمل‌کرد نزدیک به هم و در حدود ۳٪ تا ۵٪ کم‌تر نسبت به NN ارائه کرده‌اند. نتایج دو معیار حساسیت و اختصاصیت نشان می‌دهد که طبقه‌بندها (به جز SVM) عمل‌کرد نزدیکی در پیش‌بینی کلاس مثبت (مطلوب) و منفی (نامطلوب) داشته و یک تعادل مناسب بین پیش‌بینی کلاس مثبت به عنوان مثبت و پیش‌بینی کلاس منفی به عنوان منفی برقرار بوده است.

جدول (۱)- نتایج سه معیار صحت، حساسیت و اختصاصیت بر حسب درصد برای طبقه‌بندی هر یک از دو حس برانگیختگی و خوشایندی با پنج طبقه‌بند NN، SVM، KNN، DT و RF

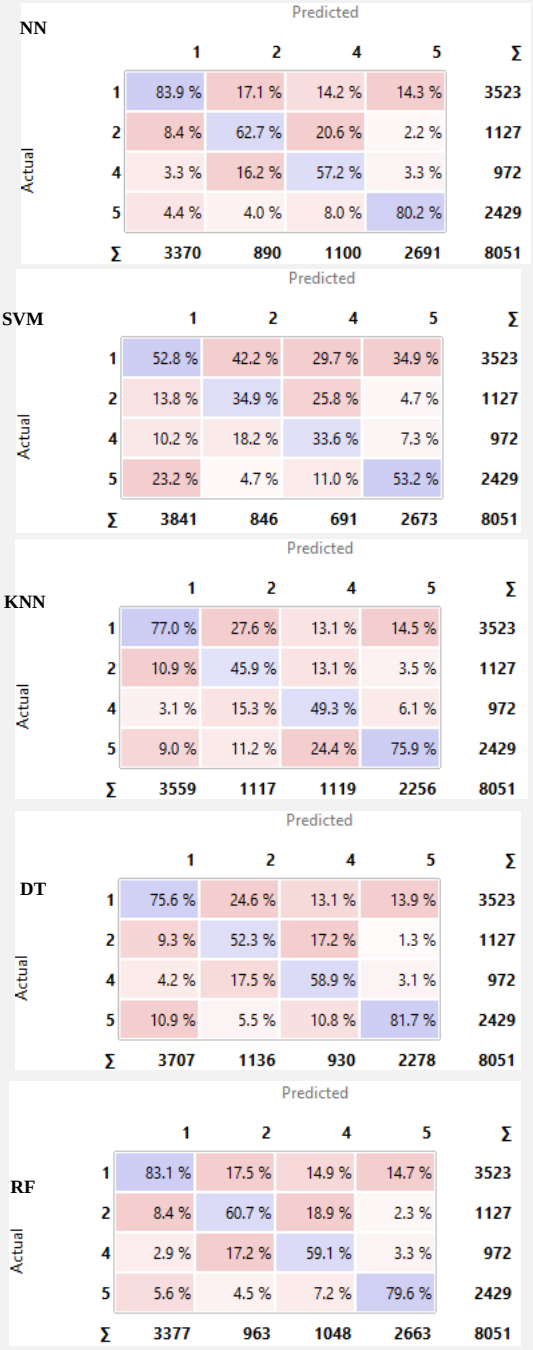
		صحت	حساسیت	اختصاصیت
برانگیختگی	NN	۷۷	۷۷	۹۰/۳
	SVM	۴۶/۳	۴۶/۳	۷۶/۲
	KNN	۷۲/۴	۷۲/۴	۸۷/۳
	DT	۷۲/۱	۷۲/۱	۸۷/۱
	RF	۷۵/۸	۷۵/۸	۹۰
خوشایندی	NN	۷۹/۷	۷۹/۷	۹۱/۲
	SVM	۴۶/۸	۴۶/۸	۷۵/۹
	KNN	۷۶/۳	۷۶/۳	۸۹/۲
	DT	۷۶	۷۶	۸۹/۲
	RF	۷۸/۹	۷۸/۹	۹۱

فرایند چهار بخشی کردن داده و محاسبه‌ی معیارهای طبقه‌بندی ۱۰ بار تکرار شده و متوسط معیارها برای ۱۰ بار تکرار به عنوان نتایج طبقه‌بندی گزارش شده است. به منظور بررسی اعتبار و معنی‌دار بودن صحت طبقه‌بندی، خروجی صحت طبقه‌بندها برای این ۱۰ بار تکرار با روش غیرپارامتری آزمون ویلکاکسون رتبه‌ی علامت‌دار[1] [۳۳] مورد مقایسه قرار گرفته است. برای هر دو حس خوشایندی و برانگیختگی، روش NN به طور معنی‌داری از تمام طبقه‌بندها بهتر بوده (all-p<۰/۰۰۵ و all-z>۲/۸) و روش SVM به طور معنی‌داری از تمام طبقه‌بندها ضعیف‌تر عمل کرده است (all-p<۰/۰۰۵ و all-z<-۲/۸). روش RF نیز نسبت به تمام طبقه‌بندها به جز NN عمل‌کرد معنی‌دار بهتری داشته است (all-p<۰/۰۰۵ و all-z>۲/۸). هم‌چنین تفاوت معنی‌داری میان دو روش KNN و DT مشاهده نشده است (p>۰/۱۶۷ و z<۱/۳۸).

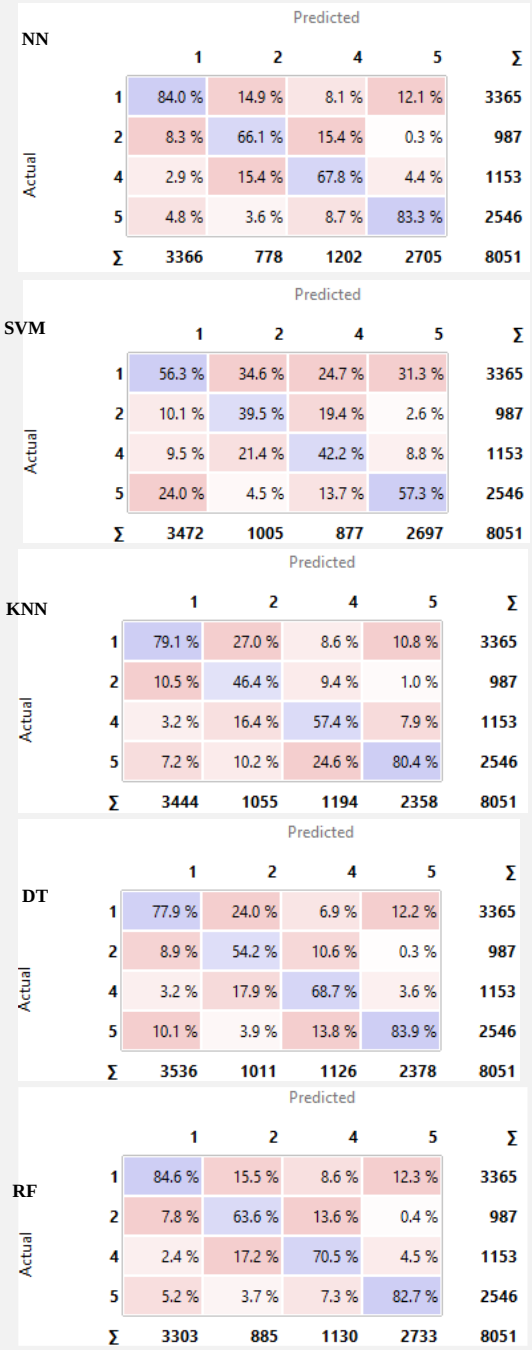
نتایج ماتریس درهم‌ریختگی برای دو حس برانگیختگی و خوشایندی در شکل‌های (۸) و (۹) ارائه شده است. نتایج ماتریس درهم‌ریختگی نیز بیان‌گر عمل‌کرد بهتر NN نسبت به سایر طبقه‌بندها است. هم‌چنین نتایج ماتریس درهم‌ریختگی، نتایج معیارهای صحت، حساسیت و اختصاصیت مبنی بر

[1] Wilcoxon Signed-Rank Test

نزدیکی عمل‌کرد RF به NN، نزدیکی عمل‌کرد DT و KNN و عمل‌کرد بسیار ضعیف SVM را تایید می‌کند. برای هر دو حس برانگیختگی و خوشایندی، دو کلاس ۱ و ۵ چند برابر دو کلاس ۲ و ۴ از طرف افراد انخاب شده و به همین دلیل یک عدم تعادل در برچسب‌های اختصاص داده شده به موسیقی‌ها وجود دارد. این امر موجب شده است تا تمام طبقه‌بندها در فرایند آموزش، تعداد بیش‌تری از دو برچسب ۱ و ۵ را مشاهده کرده و فرایند یادگیری برای این دو کلاس بهتر از دو کلاس ۲ و ۴ انجام شود. در نتیجه در مرحله‌ی تست عمل‌کرد بهتری برای دو کلاس ۱ و ۵ نسبت به دو کلاس ۲ و ۴ مشاهده شده است.



شکل (۸)- نتایج ماتریس درهم‌ریختگی برای حس برانگیختگی برای طبقه‌بندهای NN، SVM، KNN، DT و RF



شکل (۹)- نتایج ماتریس درهم‌ریختگی برای حس خوشایندی برای طبقه‌بندهای NN، SVM، KNN، DT و RF

در هر بار آموزش، مدل جدیدی ساخته شده و در نتیجه پارامترهای مدل تغییر می‌کند. به منظور فراهم کردن اطلاعات پارامترهای طبقه‌بندها برای خوانندگان علاقه‌مند به استفاده از آن‌ها در تحقیقات مشابه، کل داده‌ها با روش اعتبارسنجی متقابل ۱۰ افرازه برای آموزش مدل مورد استفاده قرار گرفته و بهترین مقادیر پارامترها به دست آمده است. پارامترهای بهینه

برای RF تعداد درخت ۱۰، برای DT تعداد برگه‌های ۲ و عمق درخت ۵، برای SVM با کرنل شعاعی مقدار گاما ۰/۰۱ و پارامتر رگولاریزیشن ۱، برای KNN تعداد همسایه‌های ۵ و بر اساس معیار فاصله‌ی اقلیدسی و برای شبکه‌ی عصبی دولایه با ۱۰ نود در هر لایه و تابع فعال‌ساز لجستیکی می‌باشد.

ممکن است این سوال مطرح شود که هر کدام از دو نوع ویژگی موسیقایی و جمعیت‌شناسی به تنهایی چقدر در تولید نتایج جدول (۱) تاثیر داشته و به عبارت دیگر آیا استفاده از هر کدام از دو نوع ویژگی می‌تواند نتایج نزدیک به نتایج گزارش شده در جدول (۱) یا بهتر از آن را فراهم کند یا بهترین نتیجه با استفاده‌ی توامان از هر دو نوع ویژگی به دست می‌آید که در جدول (۱) گزارش شده است. برای پاسخ به این سوال، طبقه‌بندی احساسات خوشایندی و برانگیختگی برای هر یک از دو نوع ویژگی موسیقایی و جمعیت‌شناسی انجام شده و نتایج آن در جدول (۲) ارائه شده است. الگوهای کلی حاکم بر تمام نتایج نشان می‌دهد که دو طبقه‌بند RF و NN عمل‌کرد نزدیک به هم داشته و فقط در مورد معیار اختصاصی بودن و حالتی که تنها ویژگی جمعیت‌شناسی مورد استفاده قرار گیرد RF عمل‌کرد بهتری نسبت به NN داشته است. طبقه‌بند SVM نیز بدترین عمل‌کرد را ارائه داده است. طبقه‌بند DT عمل‌کرد بهتری نسبت به KNN در مورد معیارهای صحت و حساسیت ارائه داده اما در مورد معیار اختصاصی بودن عمل‌کرد این دو برعکس بوده و KNN تقریبا ۶٪ بهتر عمل کرده است. در کل عمل‌کرد طبقه‌بندها برای استفاده از فقط ویژگی موسیقی ۱۵٪-۲۰٪ و برای استفاده از فقط ویژگی جمعیت‌شناسی بیش از ۱۵٪-۳۰٪ نسبت به حالتی که از تمام ویژگی‌ها استفاده شده کاهش پیدا کرده است. صحت طبقه‌بندها زمانی که از کل ویژگی‌ها استفاده شده در حدود ۸۰٪ بوده اما وقتی که فقط از ویژگی موسیقی استفاده شده به حدود ۶۰٪ رسیده و برای استفاده فقط از ویژگی جمعیت‌شناسی به زیر ۵۰٪ رسیده است. این نتایج نشان می‌دهد که ویژگی‌های موسیقی مورد استفاده در این مطالعه نسبت به ویژگی‌های جمعیت‌شناسی استفاده شده تاثیر بیش‌تری در به دست آوردن صحت بالاتر در پیش‌بینی درست کلاس احساسات موسیقی داشته‌اند. البته این موضوع به این معنی نیست که ویژگی‌های جمعیت‌شناسی همواره عمل‌کرد ضعیف‌تری نسبت به ویژگی‌های موسیقایی در پیش‌بینی دو احساس خوشایندی و برانگیختگی دارند بلکه نشان می‌دهد که ویژگی‌های جمعیتی مورد استفاده در این مطالعه نسبت به ویژگی‌های موسیقایی استفاده شده در این مقاله عمل‌کرد ضعیف‌تری داشته‌اند. اگر بتوان ویژگی‌های جمعیتی بیش‌تر و با توزیع یک‌نواخت‌تر نسبت به توزیع‌های نشان داده شده در شکل‌های (۳) تا (۷) را تهیه کرد این احتمال وجود دارد که ویژگی‌های جمعیتی نیز عمل‌کرد مناسبی ارائه دهند زیرا با توزیع بهتر ویژگی‌ها، مدل طبقه‌بندی بهتری برای ویژگی‌های جمعیتی ساخته خواهد شد. هم‌چنین نتایج نشان می‌دهد که استفاده‌ی توامان از دو نوع ویژگی یاد شده اثرات مثبت و هم‌افزا در بهبود صحت طبقه‌بندی دارد.

جدول (۲)- نتایج سه معیار صحت، حساسیت و اختصاصیت بر حسب درصد برای طبقه‌بندی هر یک از دو حس برانگیختگی و خوشایندی با طبقه‌بندهای NN، SVM، KNN، DT و RF برای دو حالت فقط استفاده از ویژگی موسیقی و فقط استفاده از ویژگی جمعیت‌شناسی

			صحت	حساسیت	اختصاصیت
ویژگی موسیقی	برانگیختگی	NN	۶۱/۴	۶۱/۴	۷۷/۳
		SVM	۳۶/۱	۳۶/۱	۷۱/۹
		KNN	۵۰/۴	۵۰/۴	۸۴/۳
		DT	۶۱/۳	۶۱/۳	۷۷/۵
		RF	۶۱/۴	۶۱/۴	۷۷/۷
	خوشایندی	NN	۶۳	۶۳	۷۹/۵
		SVM	۲۷/۵	۲۷/۵	۷۷/۴
		KNN	۵۶	۵۶	۸۶/۵
		DT	۶۳/۳	۶۳/۳	۸۰
		RF	۶۳/۳	۶۳/۳	۸۰
ویژگی جمعیت‌شناسی	برانگیختگی	NN	۴۸/۷	۴۸/۷	۶۶/۴
		SVM	۲۸/۶	۲۸/۶	۶۹/۴
		KNN	۴۴/۸	۴۴/۸	۷۶/۷
		DT	۴۹/۲	۴۹/۲	۷۱/۳
		RF	۴۸/۹	۴۸/۹	۷۲/۶
	خوشایندی	NN	۴۷/۲	۴۷/۲	۶۸/۵
		SVM	۲۹/۱	۲۹/۱	۶۷/۱
		KNN	۴۱/۶	۴۱/۶	۷۷/۱
		DT	۴۷/۳	۴۷/۳	۷۲/۳
		RF	۴۶/۶	۴۶/۶	۷۳/۲

۴-۱- محدودیت‌ها و کارهای آینده

در این مطالعه توزیع ویژگی‌های جمعیتی مخصوصا ویژگی‌های تعداد سال تحصیلی، سطح اقتصادی و قومیت (مطابق شکل‌های ۳ تا ۵) غیریک‌نواخت بوده است. این امر ممکن است موجب آموزش نامطلوب ویژگی‌های جمعیتی توسط مدل و در نتیجه باعث کاهش تاثیر آن‌ها بر عمل‌کرد طبقه‌بندی احساسات برانگیختگی و خوشایندی شده باشد. به منظور بررسی اثر یک‌نواخت‌سازی، برای ویژگی تعداد سال تحصیلات مقادیر ۱۲

و ۱۳ و نیز مقادیر ۱۷ تا ۲۲ سال یکی شده، برای ویژگی سطح اقتصادی مقادیر ۲ تا ۴ یکی شده و برای ویژگی قومیت نیز دو قوم ترک و لر یک گروه در نظر گرفته شده است. نتایج طبقه‌بندی برای تمام طبقه‌بندها به جز SVM به مقدار ۰/۱٪ تا ۰/۲٪ و برای SVM به میزان ۱/۴٪ بهبود پیدا کرده است. هر چند تغییرات اعمال شده به منظور یک‌نواخت کردن توزیع ویژگی‌های جمعیتی تاثیر چندانی بر عمل‌کرد طبقه‌بندی نداشته که البته آن هم می‌تواند به دلیل محدود بودن تغییرات اعمال شده باشد به طوری که هم‌چنان توزیع ویژگی‌ها از توزیع یک‌نواخت مناسب فاصله داشته است، اما برای مطالعات آینده و به منظور استفاده‌های عملی و میدانی از موضوع این مقاله (پیش‌بینی احساسات موسیقی بر اساس ویژگی‌های موسیقایی و جمعیتی) پیشنهاد می‌شود که مدل‌های طبقه‌بندی با استفاده از مشاهداتی که دارای توزیع یک‌نواخت برای ویژگی‌های مورد استفاده هستند آموزش داده شوند تا تمام مقادیر یک ویژگی به خوبی توسط مدل مشاهده گردد. در این مطالعه از ۴۸ موسیقی و ۱۷۵ شرکت کننده استفاده شده است. برای کارهای آینده می‌توان از تعداد افراد و تعداد موسیقی‌های بیش‌تری استفاده کرد. هم‌چنین می‌توان تعداد ویژگی‌های استخراج شده‌ی موسیقی و جمعیت‌شناسی را افزایش داد و از ویژگی‌های مناسب‌تری استفاده نمود. در این مطالعه از موسیقی‌های مجموعه‌ی DEAM که دارای برچسب بوده استفاده شده است. برای کارهای آینده می‌توان یک مجموعه‌ی داده‌ی دارای برچسب برای موسیقی ایرانی تهیه کرد و از آن استفاده نمود. برای استخراج ویژگی از موسیقی شاید شبکه‌های عمیق و ترانسفورمرها بتوانند ویژگی‌های مناسبی به منظور بهبود طبقه‌بندی فراهم سازند. این ایده می‌تواند یکی از عناوین کارهای آینده باشد. این مطالعه بر پیش‌بینی احساسات برانگیختگی و خوشایندی حاصل از گوش دادن به موسیقی متمرکز بوده است. برای کارهای آینده می‌توان این مطالعه را برای تصویر، متن و ویدئو نیز انجام داد.

۵- نتیجه‌گیری

در این مطالعه ۴۸ موسیقی ۳۰ ثانیه‌ای با سطوح برانگیختگی و خوشایندی بسیار بالا و بسیار پایین از مجموعه‌ی داده‌ی DEAM انتخاب شده است. در ادامه به هر موسیقی توسط هر یک از ۱۷۵ شرکت کننده‌ی ایرانی دو برچسب یکی برای حس برانگیختگی و دیگری برای حس خوشایندی اختصاص داده شده است. مقادیر برچسب‌ها از ۱ (کم‌ترین) تا ۵ (بیش‌ترین) برای میزان برانگیختگی و خوشایندی متغیر است. ویژگی‌های موسیقی و ویژگی‌های جمعیتی افراد شرکت کننده استخراج گردیده و توسط پنج طبقه‌بند back propagation NN، SVM، KNN، DT و RF برای پیش‌بینی برچسب قطعه‌های موسیقی برای هر فرد مورد استفاده قرار گرفته است. بهترین نتایج از لحاظ صحت، حساسیت و اختصاصیت توسط شبکه‌ی عصبی NN به دست آمده که صحت و حساسیت تقریبا ۸۰٪ و اختصاصیت تقریبا ۹۰٪ بوده است. این نتایج نشان می‌دهد که چشم‌انداز امیدبخشی برای پیش‌بینی میزان دو حس برانگیختگی و خوشایندی حاصل از گوش دادن به موسیقی برای جامعه‌ی ایرانی وجود دارد و برای استفاده‌های کاربردی و جمعیت‌های با تعداد بسیار بیش‌تر افراد می‌توان امیدوار بود که برای رسیدن به این میزان از پیش‌بینی فقط ویژگی‌های موسیقی و جمعیت‌شناسی کافی بوده و نیازی به ثبت سیگنال‌های فیزیولوژیکی و استخراج ویژگی‌از آن‌ها وجود ندارد که این امر به نوبه خود در بسیاری از کاربردهای مبتنی بر تشخیص احساسات می‌تواند بسیار مفید واقع شود.

۶- مراجع

[1] Cui, X., Wu, Y., Wu, J., You, Z., Xiahou, J., & Ouyang, M. (2022). A review: Music-emotion recognition and analysis based on EEG signals. Frontiers in Neuroinformatics, 16, 997282.

[2] Aljanaki, A., Yang, Y. H., & Soleymani, M. (2017). Developing a benchmark for emotional analysis of music. PloS one, 12(3), e0173392.

[3] Han, D., Kong, Y., Han, J., & Wang, G. (2022). A survey of music emotion recognition. Frontiers of Computer Science, 16(6), 166335.

[4] Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., ... & Turnbull, D. (2010, August). Music emotion recognition: A state of the art review. In Proc. ismir (Vol. 86, pp. 937-952).

[5] Hu, X., & Yang, Y. H. (2017). Cross-dataset and cross-cultural music mood prediction: A case on western and chinese pop songs. IEEE Transactions on Affective Computing, 8(2), 228-240.

[6] Panda, R., Malheiro, R., & Paiva, R. P. (2018). Novel audio features for music emotion recognition. IEEE Transactions on Affective Computing, 11(4), 614-626.

[7] Panda, R., Malheiro, R., & Paiva, R. P. (2020). Audio features for music emotion recognition: a survey. IEEE Transactions on Affective Computing, 14(1), 68-88.

[8] Gómez-Cañón, J. S., Cano, E., Eerola, T., Herrera, P., Hu, X., Yang, Y. H., & Gómez, E. (2021). Music emotion recognition: Toward new, robust standards in personalized and context-sensitive applications. IEEE Signal Processing Magazine, 38(6), 106-114.

[9] Lin, Y. C., Yang, Y. H., & Chen, H. H. (2011). Exploiting online music tags for music emotion classification. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 7(1), 1-16.
[10] Xia, Y., & Xu, F. (2022). Study on music emotion recognition based on the machine learning model clustering algorithm. Mathematical Problems in Engineering, 2022.
[11] Zhang, K., & Sun, S. (2013). Web music emotion recognition based on higher effective gene expression programming. Neurocomputing, 105, 100-106.
[12] Agarwal, G., & Om, H. (2021). An efficient supervised framework for music mood recognition using autoencoder-based optimised support vector regression model. IET Signal Processing, 15(2), 98-121.
[13] Han, B. J., Rho, S., Dannenberg, R. B., & Hwang, E. (2009, October). SMERS: Music Emotion Recognition Using Support Vector Regression. In ISMIR (pp. 651-656).
[14] Agarwal, G., & Om, H. (2021). An efficient supervised framework for music mood recognition using autoencoder-based optimised support vector regression model. IET Signal Processing, 15(2), 98-121.
[15] Torres, D. A., Turnbull, D., Barrington, L., & Lanckriet, G. R. (2007, September). Identifying Words that are Musically Meaningful. In ISMIR (Vol. 7, pp. 405-410).
[16] Panwar, S., Rad, P., Choo, K. K. R., & Roopaei, M. (2019). Are you emotional or depressed? Learning about your emotional state from your music using machine learning. The Journal of Supercomputing, 75, 2986-3009.
[17] ER, M. B., & ESİN, E. M. (2021). Music emotion recognition with machine learning based on audio features. Computer Science, 6(3), 133-144.
[18] Song, Y., Dixon, S., & Pearce, M. (2012, June). A survey of music recommendation systems and future perspectives. In 9th international symposium on computer music modeling and retrieval (Vol. 4, pp. 395-410).
[19] Panda, R., Rocha, B., & Paiva, R. P. (2015). Music emotion recognition with standard and melodic audio features. Applied Artificial Intelligence, 29(4), 313-334.
[20] Yang, X., Dong, Y., & Li, J. (2018). Review of data features-based music emotion recognition methods. Multimedia systems, 24, 365-389.
[21] Soleymani, M., Aljanaki, A., & Yang, Y. H. (2016). DEAM: Mediaeval database for emotional analysis in music. Geneva, Switzerland.
[22] https://cvml.unige.ch/databases/DEAM/.
[23] Takashima, N., Li, F., Grzegorzek, M., & Shirahama, K. (2023). Embedding-based music emotion recognition using composite loss. IEEE Access.
[24] Morris, J. D. (1995). Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response. Journal of advertising research, 35(6), 63-68.
[25] Stevens, F., Murphy, D. T., & Smith, S. L. (2017, September). Soundscape categorisation and the self-assessment manikin. In Proceedings of the 20th International Conference on Digital Audio Effects.
[26] Ellis, D. P., & Poliner, G. E. (2007, April). Identifyingcover songs' with chroma features and dynamic programming beat tracking. In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07 (Vol. 4, pp. IV-1429). IEEE.
[27] http://labrosa.ee.columbia.edu/projects/coversongs
[28] Dubnov, S. (2004). Generalization of spectral flatness measure for non-gaussian linear processes. IEEE Signal Processing Letters, 11(8), 698-701.
[29] https://www.mathworks.com/help/audio/ug/spectral-descriptors.html.
[30] Birajdar, G. K., & Patil, M. D. (2020). Speech/music classification using visual and spectral chromagram features. Journal of Ambient Intelligence and Humanized Computing, 11(1), 329-347.
[31] Weineck, K., Wen, O. X., & Henry, M. J. (2022). Neural synchronization is strongest to the spectral flux of slow music and depends on familiarity and beat salience. Elife, 11, e75515.
[32] Khare, S. K., Blanes-Vidal, V., Nadimi, E. S., & Acharya, U. R. (2023). Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. Information Fusion, 102019.
[33] Wilcoxon, F. (1992). Individual comparisons by ranking methods. In Breakthroughs in statistics: Methodology and distribution (pp. 196-202). New York, NY: Springer New York.