Automatic Detection of Users' Emotional States Using Music Listening Data

Detección Automática de los Estados Emocionales de los Usuarios Mediante Datos de Escucha Musical

DOI: https://doi.org/10.17981/cesta.06.02.2025.04

Scientific research. Date received: October 10, 2025, Date accepted: December 5, 2025.

Gustavo Barraza Mercado
Soluprojects, (Colombia)
gustavo.barraza@soluprojects.com

Johan Mardini
Universidad de la Costa, (Colombia)
jmaidini@cuc.edu.co

Ernesto Esmeral-Romero
Universidad de la Costa, (Colombia)
eesmeral2@cuc.edu.co

How to cite
G. Barraza Mercado, J. Mardini, and E. Esmeral-Romero "Automatic Detection of Users' Emotional States Using Music Listening Data". J. Comput. Electron. Sci.: Theory Appl., vol. 6 no. 2, pp. 37-52, 2025. DOI: https://doi.org/10.17981/cesta.06.02.2025.04

Abstract

Introduction: Music is a natural medium for emotional expression and regulation in everyday life. Recent studies highlight its potential for non-intrusive emotion detection.

Objective: Developing a system to detect users' emotional states from their behavior while listening to music. The goal is to achieve accurate and automated emotional classification.

Method: Music listening data were processed using normalization, feature extraction, and feature selection. Both supervised and unsupervised machine learning algorithms were applied and evaluated.

Results: The proposed system achieved an average classification accuracy of 82.4%, with a precision of 80.9% and a recall of 81.6% across all evaluation scenarios. Feature selection methods, such as Chi-Square and Relief, reduced computation time by approximately 25% while improving model generalization.

Conclusions: Music-listening behavior is an effective source of emotion detection without invasive measurements. The proposed system is compatible with future intelligent and emotion-sensitive applications.

Keywords: Automatic emotion detection, Mood states, Data processing, Emotional intelligence, Music.

Resumen

Introducción: La música es un medio natural para la expresión y la regulación emocionales en la vida cotidiana. Estudios recientes destacan su potencial para la detección no intrusiva de emociones.

Objetivo: Desarrollar un sistema para detectar los estados emocionales de los usuarios a partir de su comportamiento al escuchar música. Se busca lograr una clasificación emocional precisa y automatizada.

Método: Los datos de escucha musical se procesaron mediante técnicas de normalización, extracción y selección de características. Se aplicaron y evaluaron algoritmos de aprendizaje automático, tanto supervisados como no supervisados.

Resultados: El sistema propuesto alcanzó una precisión media de clasificación del 82,4 %, con una precisión del 80,9 % y una recuperación del 81,6 % en todos los escenarios de evaluación. Los métodos de selección de características, como Chi-Square y Relief, redujeron el tiempo de cálculo en aproximadamente un 25 %, al tiempo que mejoraron la generalización del modelo.

Conclusiones: El comportamiento de escucha de música es una fuente eficaz para detectar emociones sin mediciones invasivas. El sistema propuesto es compatible con futuras aplicaciones inteligentes y sensibles a las emociones.

Palabras clave: Detección automática de emociones, Estados de ánimo, Procesamiento de datos, Inteligencia emocional, Música.

INTRODUCTION

Emotions are an essential component of human experience and are expressed through physiological, cognitive, and behavioral patterns, with high interindividual variability. These differences are reflected in intensity, frequency, triggering stimuli, and forms of emotional regulation, directly influencing adaptation to the environment and social interactions. Contemporary research argues that emotions are not fixed states but dynamic, constructed processes that emerge from the interaction between the individual and their context, highlighting the need to approach them from integrative, multidimensional perspectives [1], [2].

From a functional perspective, emotions play an adaptive role by preparing the organism to respond effectively to environmental demands and facilitate appropriate behaviors in different situations. In addition, they play a key social role, as emotional expressions, both verbal and nonverbal, serve as signals that help us anticipate behaviors, regulate interpersonal relationships, and reduce communicative ambiguity. In this sense, moods are conceived as fluctuating affective configurations over which individuals perceive a certain degree of control, although they are conditioned by internal and contextual factors [3].

Music has established itself as one of the most powerful stimuli for evoking, modulating, and regulating emotions. Various studies have demonstrated its ability to influence mood, activate autobiographical memories, and generate meaningful emotional experiences. Recent research confirms that intentional music listening helps regulate emotions and contributes to psychological well-being, reinforcing its status as a natural tool for emotional self-regulation [4], [5].

However, the emotional response to music does not depend solely on its acoustic or structural characteristics, but rather on the interaction among the listener, the situation, and the listening context. Individual factors such as personality, musical training, personal preferences, and the listener's objectives significantly influence emotional experience. Consequently, music is deliberately used to modify affective states, relieve stress, intensify emotions, or evoke memories, underscoring the importance of studying music listening in everyday, uncontrolled contexts [6], [7].

At the same time, advances in artificial intelligence and machine learning have driven the development of systems that automatically detect emotional states from text, facial expressions, and multimodal signals. However, many of these methods have limitations related to their intrusive nature, dependence on experts, or artificial induction of emotions, which can alter the user's original emotional state. In the field of music, online streaming platforms focus on recommendation systems based on consumption patterns, without incorporating explicit mechanisms for dynamic emotional detection or considering the listener's actual affective state [8], [9].

Given this scenario, there remains a gap in the literature on non-intrusive methodologies for inferring emotional states from naturally listened-to music. In response to this need, this paper proposes a methodology for detecting and classifying users' emotional states from their playlists and treats music as key multimedia content for emotional modeling. This approach seeks to contribute to both affective computing and the design of user-centered intelligent systems, offering a contextual, adaptable approach aligned with everyday emotional experience.

RELATED WORKS

Emotion is a complex construct that integrates subjective experiences, physiological responses, and behavioral expressions, enabling individuals to adapt to their environment. Emotions are not mere automatic reactions but processes that emerge from the interaction among context, personal experience, and the neurocognitive structures that underpin them [10]. This multidimensional approach has been widely accepted in contemporary cognitive science, displacing reductionist models that viewed emotions exclusively as biological responses.

The functions of emotions include preparing the organism for action, facilitating decision-making, and regulating social interactions, aspects supported by recent research in psychology and neuroscience. For example, variations in affective states have been found to correlate with patterns of brain activation and autonomic responses that effectively predict adaptive behaviors [11]. This evidence reinforces the idea that emotional regulation is a central process for human well-being and adaptation.

Emotion detection has evolved alongside technological advances, incorporating techniques for processing physiological signals and machine learning algorithms. A recent systematic review shows that emotion recognition has grown into an interdisciplinary field, integrating data from EEG, ECG, body sensors, and behavioral signals to improve the accuracy of inferring affective states in dynamic environments [12]. Multimodal approaches combine multiple sources of information to achieve more robust results than traditional unidimensional methods.

Music has emerged as one of the most effective stimuli for modulating and regulating emotional states, with a growing body of studies examining how listening to music influences individuals' affective experience. A bibliometric analysis of Music Emotion Regulation indicates that research in this field has grown rapidly over the last decade, particularly between 2020 and 2024, reflecting a shift from theoretical approaches to practical applications and data-driven methodologies [13]. Furthermore, most studies on emotional regulation through music focus on listening as the primary means of influencing emotional states, underscoring its relevance in everyday contexts.

In the field of technology, research on emotion detection in musical systems has integrated deep learning and musical feature analysis to infer affective states from sound content. Recent reviews indicate that machine learning-based methods have improved the ability to classify music-induced emotional responses. However, they still face challenges related to the cultural and contextual diversity of listeners [14]. These technological developments enhance the automated detection of affective states in interactive, personalized systems, creating new opportunities to integrate musical experiences and human emotions.

In automatic emotion detection, multiple systems, APIs, and tools have been developed to analyze visual, auditory, and multimodal signals. Affectiva-Affdex is a commercial platform based on computer vision and deep learning for real-time facial expression recognition. Comparative studies have evaluated its performance against other solutions such as Microsoft Azure Face API and Emotient FACET. The results show variations in accuracy and sensitivity across datasets and application contexts [15], [16].

TABLE I. SUMMARY OF RELEVANT RESEARCH WORKS

Tool / Platform	Type	Data Modality	Main Functionality	Primary Use
Affectiva-Affdex	Commercial API	Video (facial expressions)	Detection of basic emotions and affective metrics (valence, engagement) using computer vision and deep learning	Human–computer interaction, user experience studies, affective research
Microsoft Azure Face API	Cloud-based commercial API	Images and video	Facial analysis, emotion detection, and facial attribute recognition	Web and mobile applications, emotional analytics in commercial environments
Emotient FACET	Commercial platform	Facial video	Automatic emotion recognition from microexpressions	Experimental research and behavioral analysis
openSMILE	Open-source library	Audio (speech and music)	Automatic extraction of acoustic features for emotional analysis	Emotion recognition in speech and music, affective computing
OpenEAR	Open-source toolkit	Audio	Extension of openSMILE focused on real-time affective recognition	Emotion research and affective system prototyping
MixedEmotions Toolbox	Open-source toolbox	Text, audio, and video	Multimodal emotion analysis and affective signal fusion	Academic research and multimodal system development

Tool / Platform	Type	Data Modality	Main Functionality	Primary Use
Google Cloud Vision	Cloud-based commercial API	Images and video	Visual analysis with inference of facial attributes and expressions	Large-scale AI applications and digital services
AffectNet	Public dataset	Facial images	Training and validation of emotion recognition models	Deep learning research and evaluation of affective models

Another relevant example is openSMILE, a widely used tool in affective computing for automatically extracting acoustic features from speech and music signals. This library generates feature vectors used in speech and music emotion detection models and has been integrated into numerous multimodal studies due to its flexibility and efficiency [17], [18]. Complementing openSMILE, openEAR extends its capabilities by incorporating specific functionalities for real-time affective recognition. This facilitates experimentation and systematic comparison of emotional classification methods [19].

In multimodal analysis, MixedEmotions Toolbox stands out as a set of open-source modules for emotion analysis from text, audio, and video. This tool enables the integration of multiple data streams in real time, with customizable configurations, which are particularly useful for experimental research and prototyping complex affective systems [20]. These types of toolkits have been widely adopted in academic projects that require multimodal processing and integrated affective prediction. In addition, they facilitate the replicability and comparative evaluation of emotional models [21].

In addition to open-source tools, various commercial cloud APIs, such as Microsoft Azure Cognitive Services and Google Cloud Vision, offer facial recognition and emotion analysis capabilities within their artificial intelligence services. These platforms allow affective states to be inferred from images and video using large-scale trained models and are easily integrated into web and mobile applications. Their adoption has grown in areas such as customer service, smart surveillance, and digital health [22]. However, recent studies warn that these technologies still face challenges related to algorithmic bias and the cultural generalization of models [23].

In addition, the use of affect databases has been key to training and validating emotional recognition systems. AffectNet, comprising more than one million facial images annotated with emotional expressions and valence and activation dimensions, has become one of the most widely used datasets in affective computing. This resource has been widely used in recent studies and cited in high-impact publications. Finally, papers on Affective Computing use it to evaluate deep neural network models for automatic facial emotion detection tasks [24]–[26].

Overall, the state of the art shows significant advances in automatic emotion detection, driven by deep learning, signal processing, and multimodal approaches. Despite the availability of commercial APIs, open libraries, and large-scale databases, challenges remain in model generalization, cultural sensitivity, and the effective fusion of multiple affective sources. These limitations highlight the need for continued research into the design of more robust, interpretable, and ethically responsible systems. Such systems must be capable of representing the complexity of human emotional states and responding to the demands of emerging applications in education, healthcare, and user-centered technologies.

METHODOLOGY

Dataset

Developing an appropriate database for studying emotions is a fundamental task. The material in this database should be audiovisual and collected in a real environment, without resorting to people who are predisposed to express emotion. The database for detecting emotional states has been created as a repository of information for research in this field. It must be composed of data obtained from user interactions.

Given that most existing emotional databases consist solely of recordings or multimedia content that do not apply to our research, we have included descriptive information about emotional states to enable the identification and categorization of each state.

The initial database (training database) contains information resulting from research conducted using a test described in the following section. The responses from this test allowed us to gather information about the specific qualities and technical skills of each staff member surveyed.

In this way, the training dataset was formed as follows:

Responses obtained through testing = training data.

Users who use the application = actual data.

A. Description of the database

A review of the state of the art in existing databases revealed that none meet the requirements and focus on detecting the emotions essential to testing this system, including anger, fear, sadness, and happiness.

To perform the tests, the database was divided into two parts. First, a database containing user data, initially entered by users, for detecting emotional states. The second contains an assessment and categorization based on the emotional context.

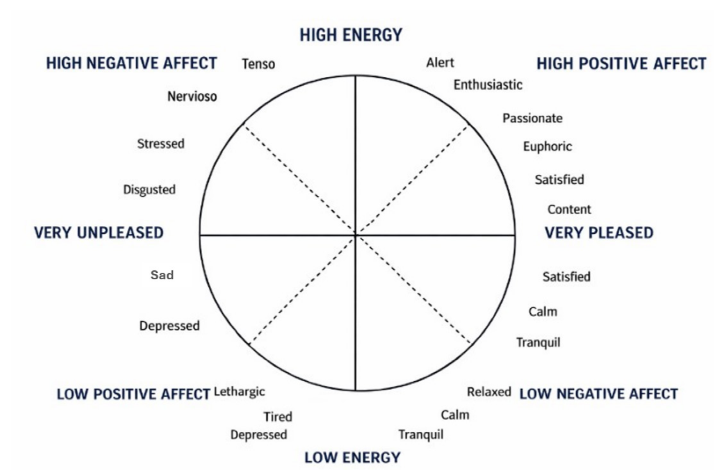

Figure 1. Continuous representation for emotional state detection

Having a database that includes the context of all emotions makes it easier to compare their characteristics. To verify whether the emotions were expressed correctly, the database underwent a subjective evaluation process in which the two databases were reviewed separately: the first, consisting of user data, was called the "neutral database" because it contains each user's emotions. Fifteen people participated in this process. These people were evaluated in the same scenario. The results of this evaluation are shown in Figure 1.

The second phase was conducted with a group of people who spontaneously rated each song's representation and emotional state according to their criteria. We will refer to this database as the "Test database" and later call it the "training database." Ten people participated in this process, who were evaluated under the same scenario as our neutral database. The results are also shown in Figure 1. This allowed us to demonstrate that all emotions are correctly identified at least 70% of the time, using a "training database" developed by a group of 10 people.

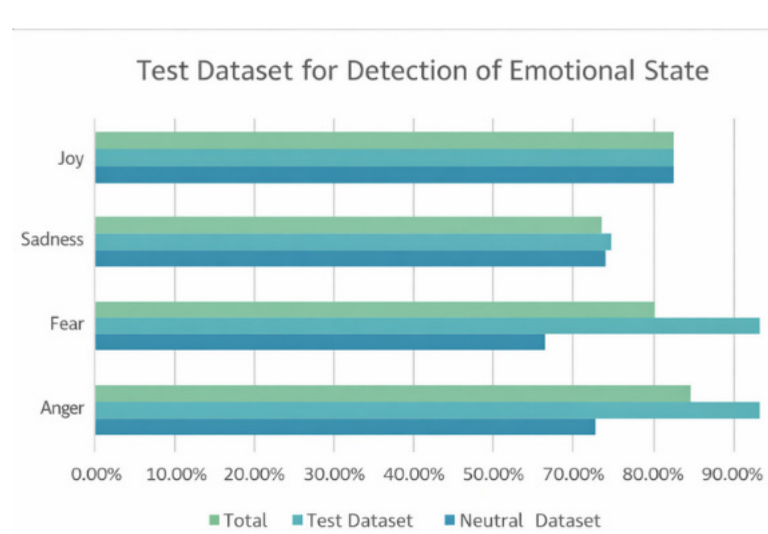

Figure 2. Result of the subjective evaluation of the database

B. Definition of metrics for comparing emotional states, for the creation of a training database

First, note that the database consists of emotional states: the total set, the Test Database, and the Neutral Database. This representation allows us to visualize the relative proportion of each basis in the system's evaluation scheme, providing a clear context for the analysis and validation of experimental results (see Figure 1).

The search for and identification of each emotional state are based on functional and non-functional parameters, as well as other fundamental performance and development characteristics, which will help our system to be more accurate. Therefore, it is necessary to establish metrics for comparing different emotional states. At this stage, the aim is to define metrics for comparing emotional states and to establish criteria for evaluating them and their representations. To do this, the different needs that may arise must be considered. Likewise, the weighting level for each criterion must be established.

C. Assessment of emotional states based on each song

For the comparison of emotional states, the evaluation is based on the following elements: speed, variation, range, breathing, intensity, articulation, and voice quality. These were taken from research on databases that implemented some form of emotion detection in multimedia content.

• *Speed*: This refers to the speed of the song or the pace at which the music is performed, which can be determined by the genre or rhythm of the song.

• *Variation*: A (musical) variation is a composition characterized by containing a musical theme that is imitated in other sub-themes or variations, which retain the same harmonic pattern as the original theme, and each part is associated with the other. The melodic patterns and tempos of each variation differ; this can range from significant variation to no variation at all.

• *Range:* The range of a song refers to the tonal distance between the lowest and highest notes.

• *Breathing*: is based on inhaling and exhaling air; the latter determines our rhetorical and singing ability, but this depends directly on the former. Depending on the style of music, breathing can range from rhythmic to irregular.

• *Intensity*: Intensity in music is the quality that differentiates a soft sound from a loud one. It depends on the sound's intensity and the distance to the receiver. It is one of the four essential qualities of articulated sound, along with pitch, duration, and timbre.

• *Articulation*: in music, refers to the way in which the transition from one sound to another or on the same note is produced. It is the set of elements that define the different possibilities of connection between the notes that make up a melody or, by extension, the chords that make up a succession of chords in a passage or in a homophonic composition.

• *Voice quality:* Voice quality refers to how clearly the message you want to convey is heard. A good voice helps listeners relax and enjoy listening. Poor-quality voice can hinder communication and frustrate the listener.

D. Classification of the criteria to be evaluated

Comparing emotional states (see Figure 1) according to the criteria outlined above requires defining the measurement value for each factor. To facilitate measurement, scores of 1-5 have been assigned based on the level of compliance. Each of the points to be evaluated in the previous section was used as a basis for this. The aim was to establish a relationship between a song and its emotional state, as shown in Table II.

Table II. Classification of Evaluation Criteria (Training Dataset)

Speech Rate	
Level	**Options**
5	Very fast
4	Fast
3	Slightly fast
2	Normal
1	Slow

Breathing	
Level	**Options**
5	Excited
4	Agitated
3	Normal
2	Irregular
1	Rhythmic

Variability	
Level	**Options**
5	Very high
4	High
3	Normal
2	Slight
1	Low / Minimal variability

Articulation	
Level	**Options**
4	Precise
3	Slow
2	Normal
1	Tense

Voice Quality	
Level	**Options**
4	Chest-based (thoracic)
3	Strident
2	Resonant
1	Irregular

Pitch Range	
Level	**Options**
3	Wide
2	Moderate
1	Narrow

Intensity	
Level	**Options**
3	High
2	Normal
1	Low

E. Relationship between emotions and their representation

TABLE III. WEIGHTING OF EVALUATED CRITERIA.

Criterion	Anger	Happiness	Sadness	Fear
Speed	3	4	1	5
Variation	5	4	2	5
Range	3	3	1	3
Breathing	1	1	1	2
Intensity	3	3	1	2
Articulation	1	2	3	4
Voice Quality	4	3	2	1
Total Criterion	20	20	11	22

The criteria defined for the “training database” were based on Table II for recognizing emotional states, which enabled the generation of Table III, which contains the weights of the criteria evaluated for the four main states used in the tests: Anger, Happiness, Sadness, and Fear.

Similarly, a qualitative representation of the relationship between emotions and the criteria applied was generated (see Table IV), which contains the information needed to classify a song based on its emotional representation.

TABLE IV. RELATIONSHIP BETWEEN EMOTIONS AND THEIR REPRESENTATION (APPLIED CRITERIA).

Representations/ Emotions	Anger	Happiness	Sadness	Fear
Speed	Slightly accelerated	Accelerated or slowed	Paced	Very accelerated
Variation	Very high	High	Slightly low	Very high
Range	Wide	Wide	Narrow	Wide
Breathing	Rhythmic	Rhythmic	Rhythmic	Irregular
Intensity	High	High	Low	Normal
Articulation	Tense	Normal	Paced	Precise
Voice Quality	Chest-based	Strident	Resonant	Irregular

F. Data collection for the construction of the test dataset

To create the test dataset, a list of users was generated based on their age range [18-30] years. To collect the data, a "survey" was used, administered via a specially designed web platform, which enabled data collection for each attribute (characteristic) presented in Table V. The data collected were stored in a database (MySQL) for later processing. The complete process is shown in Figure 3.

TABLE V. DATA COLLECTED.

Sex	Full Playback	Satisfied
Song Listened	Pause	Training State
Genre	Playback Forward	Current Emotional State
Number of Plays	Playback Backward	Would Buy App
Volume	Age	Playback Date

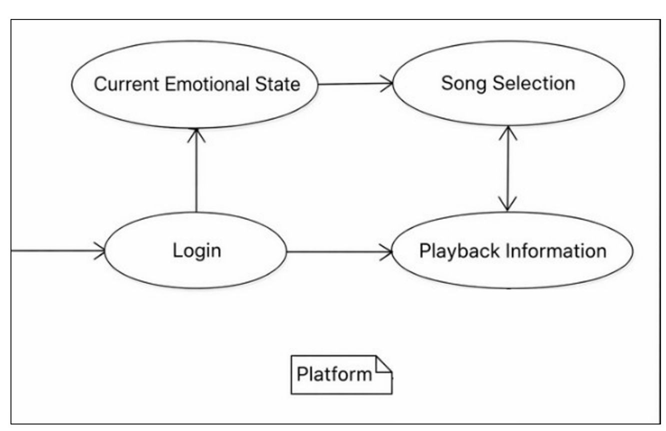

Figure 3. Web platform data collection process.

PROCESSING OF EMOTIONAL STATE DATASETS

A. Data normalization

For the training dataset, 14 of the 15 characteristics shown in Table V were selected, and the C4.5 and GHSOM algorithms were applied to perform feature classification. Data normalization prevents any feature from contributing more than another to the distance measurement [27], [28]. Four normalization techniques (var, range, and logistics) were used and are described below.

Var: normalizes the variance of the variable to unity and the mean to zero. This is a simple linear transformation, as shown in Formula 1. Where and σ are, respectively, the mean and standard deviation of variable x. This is equivalent to expressing variable x as the distance between the number of standard deviations and its mean.

$$\hat{x} = \frac{x - \bar{x}}{\sigma} \qquad \textit{(1)}$$

Range: scale the values of the variable between [0, 1] with a simple linear transformation, as shown in Figure N. The parameters of the transformation are the minimum value and the range (max(x) - min(x) of the variable, as shown in Formula 2. If the transformation is applied to new data outside the minimum and maximum ranges, the transformed values will also be outside the [0, 1] range.

$$x' = \frac{x - min(x)}{(x) - min(x)} \qquad \textbf{\textit{(2)}}$$

Log: a useful logarithmic transformation when the variable's values are distributed exponentially, with too many very small values and fewer large values. This transformation is a good way to obtain higher resolution at the lower end of the component vector. What is done is a nonlinear transformation (see Formula 3), where *ln* is the natural logarithm, which generates non-negative values.

$$x' = ln(x - (x) + 1) \qquad \textbf{\textit{(3)}}$$

Logistic: this normalization method ensures that all values, from minus infinity to plus infinity, fall within the [0, 1] range. The transformation is roughly linear in the middle of the range (around the mean), and exhibits smooth nonlinearity at both ends, ensuring that all values fall within the range. See Formula 4.

$$x' = \frac{1}{\left(1+e^{-\hat{x}}\right)} \qquad \textbf{\textit{(4)}}$$

B. Selection and classification of characteristics

The algorithms used have been implemented with criteria for sorting attributes.

Chi-Square: is a discretization method that has proven capable of eliminating redundant and/or irrelevant attributes.

Relief: is based on the nearest neighboring technique, which assigns weight to each attribute. The weight of each attribute is adjusted based on its ability to distinguish among the class variable's values.

Therefore, focusing on the var method for data normalization and using the attribute selection algorithms described above, we conducted a simulation in which we applied these selection methods and obtained the following results (Table VI).

TABLE VI. ATTRIBUTE SELECTION APPLYING THE CHI-SQUARE AND RELIF ALGORITHMS.

Chi-Square					Relief		
Nro.	**Orden Inicial**	**Nro.**	**Orden Final**	**Nro.**	**Orden Inicial**	**Nro.**	**Orden Final**
1	Gender	14	Training Status	1	Gender	14	Training Status
2	Song Listened	12	Satisfied	2	Song Listened	3	Genre
3	Genre	9	Playback Re-wind	3	Genre	9	Playback Re-wind
4	Playback Count	6	Complete Play-back	4	Playback Count	1	Gender
5	Volume	3	Genre	5	Volume	8	Playback For-ward
6	Complete Play-back	8	Playback For-ward	6	Complete Playback	7	Pause
7	Pause	1	Gender	7	Pause	6	Complete Play-back
8	Playback Forward	7	Pause	8	Playback For-ward	12	Satisfied
9	Playback Rewind	11	App Purchase	9	Playback Re-wind	2	Song Listened
10	Age	4	Playback Count	10	Age	11	App Purchase
11	App Purchase	5	Volume	11	App Purchase	5	Volume
12	Satisfied	2	Song Listened	12	Satisfied	4	Playback Count
13	Current State	10	Age	13	Current State	10	Age
14	Training Status			14	Training Sta-tus		
Evaluator attribute			Status	Evaluator attribute			Status
Selected Attributes: 14,12,9,6,3,8,1,7,11,4,5,2,10: 13				**Selected Attributes**: 14,3,9,1,8,7,6,12,2,11,5,4,10: 13			

Once the attribute rankings were obtained, the common attributes across algorithms were identified. In addition, the attributes are chosen according to their order of relevance, since, as can be seen in Table VI, those in the last position are not as relevant as the first five selected.

The learning algorithms were chosen because they represent different types of classifiers and are frequently used in comparative studies. It should be noted that these algorithms have not been widely implemented for the detection of emotional states, and part of this research work is to determine which of these algorithms provides the best results in detecting emotional states based on the songs listened to

C4.5: It is an algorithm based on decision trees with training data. In addition to being a fast, robust, easy-to-use, and easy-to-understand classifier, it is one of the most popular methods.

GHSOM: It consists of a multi-layered hierarchical structure, where each layer is composed of several independent growing SOMs. Starting from a higher-level map, each map, like the growth grid or grid model, is expanded to represent a collection of data at a specific level of detail.

C. Components of the system for detecting emotional states and their relationships

Figure 4 shows the system components and their interactions.

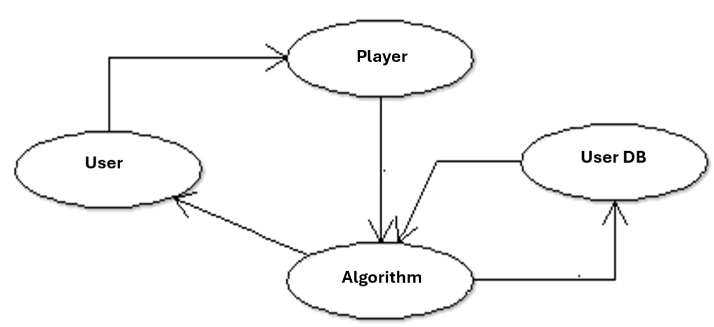

Figure 4. Components and their interaction in emotional state detection systems

Below is a brief description of the system's components.

- User

This person represents the user who uses the emotion detection system through a media player and is responsible for selecting the songs they want to listen to, which will be used in the following stages.

- Player

The media player is how we collect user information. It provides us with important data, such as which song is playing, whether the volume is increasing or decreasing, and whether the song is fast-forwarded.

- User Database

It represents the knowledge base on which the entire study of emotional state detection is modeled: the user's playback history database.

- Algorithms

This refers to the system whose objective is to detect emotional states through interaction with all its components. It is responsible for analyzing and processing communication actions with other components to determine the user's emotional state.

D. Methods implemented by the system for detecting emotional states.

Algorithms play a key role in the system developed for detecting emotional states. The algorithm's processes are described in detail below.

FEATURE PROCESSING FOR DETECTING EMOTIONAL STATES.

Based on an analysis of the most reliable methods for detecting users' emotional states, we concluded that no single method could guarantee an accurate assessment of a user's mood. We will therefore develop a system that takes the best of each method to obtain the desired result. Therefore, the process was divided into stages to clarify how the characteristics were processed (see Figure 5).

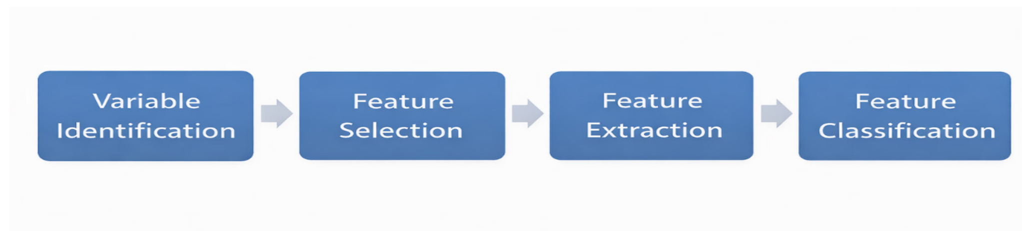

Figure 5. Feature processing for emotional state detection

A. Identification of variables for detecting emotions.

Once the variables were selected, we analyzed the many variables available in a test environment, focusing on detecting emotional states. So, we can consider what to look for when identifying the most suitable variables, as shown in Figure 6.

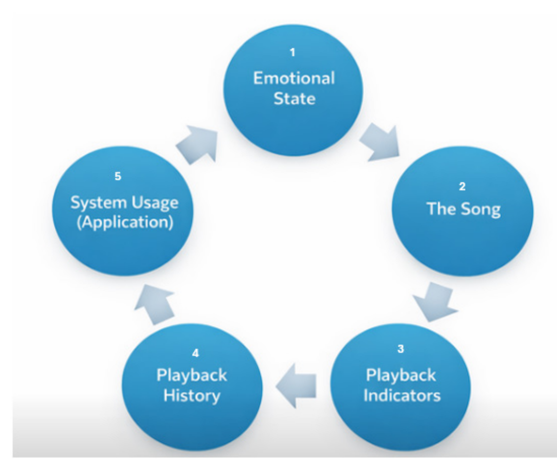

Figure 6. Variables for recognizing emotional states

Each stage is summarized below:

• ***Emotional State***: a starting point is established for the user, determining their emotional state at the time of using the application (this will be used for algorithm learning).

• ***The song:*** this is a very important factor, as it relates the emotional state to multimedia content, in this case, an audio file. This will later be linked to an emotional state.

• ***Playback Indicators***: these refer to how the user interacts with the application (algorithm) and uses it; for example, whether they turn the volume of a song up or down, whether they listen to it in its entirety or not, whether they change songs, etc.

• ***Playback History***: one of the most critical points is the use of the system. This is because if the user uses it consistently, the likelihood of making a more accurate judgment of their emotional state will approach the expected result.

Once the variables used to create the training dataset were analyzed and identified, it was determined that the best for identifying emotional states were those listed in Table V.

EXPERIMENTS AND RESULTS

To process the “TRAINING” and “TEST” datasets and evaluate the results using 10-fold cross-validation as an alternative dataset for the simulation system in a laboratory environment.

The best results were selected based on the metrics used to evaluate the efficiency of the implemented methodology, both in test laboratories and in real environments. We will compare each scenario, highlighting the best results obtained and describing their contribution to the methodology.

Each scenario is listed by importance, from test scenarios without attributing selection to scenarios in which different attribute selection techniques were applied to each classification algorithm. Each scenario is subdivided into several scenarios identified by the letter [E].

The metrics used to compare the different techniques are presented below:

• **Precision**: identifies the proportion of predicted positive cases that are correct

• **Accuracy**: provides the total number of correct predictions.

• **Recall (positive recognition)**: indicates the proportion of positive cases correctly identified (i.e., the emotional state we want to identify).

A. Scenario [E1]

C4.5 and GHSOM algorithms applied (TRAIN AND TEST)

All features (13) were used with the C4.5 and GHSOM classification techniques in the TRAIN and TEST datasets.

Table VII. *Comparison of C4.5 and GHSOM, without feature selection, with scenario (Train, Test)*

	SCENARIO		FEATURES	PRECISION	ACCURACY	RECALL
NO SELECTION OF FEATURES	**TRAIN**	**C4.5**	13	0.938	0,9377	0,9192
		GHSOM	13	0.893	0,8928	0,4876
	TEST	**C4.5**	13	0.938	0,9377	0,9192
		GHSOM	13	0.907	0,9074	0,4993

The first simulation scenario (E1) presents the datasets used for training and testing with the maximum number of features (13), where no attribute selection technique was applied, and the training scenarios for classification techniques show better performance than GHSON. In the Test scenario, the results show that C4.5 remains the best-performing algorithm across the applied metrics (see Table VII). The simulation results show that the C4.5 algorithm without feature selection outperforms GHSOM, achieving 93.80% precision, 93.77% accuracy, and 91.92% recall in both training and testing, demonstrating consistent performance and adequate generalization.

B. Scenario [E2]

[E2.1] Chi-Square y C4.5

The simulation was performed by varying the number of features in the Test dataset. The results show that C4.5 maintains stable performance with all features. In other words, the C4.5 classification algorithm remains stable regardless of the number of features selected. Therefore, our best result would be with 13 features. Figure 7 presents the results of the metrics obtained using 13 features, achieving a precision of 94.60%, an accuracy of 94.61%, and a recall of 94.04%. Note that reducing the number of features does not significantly affect the model's performance.

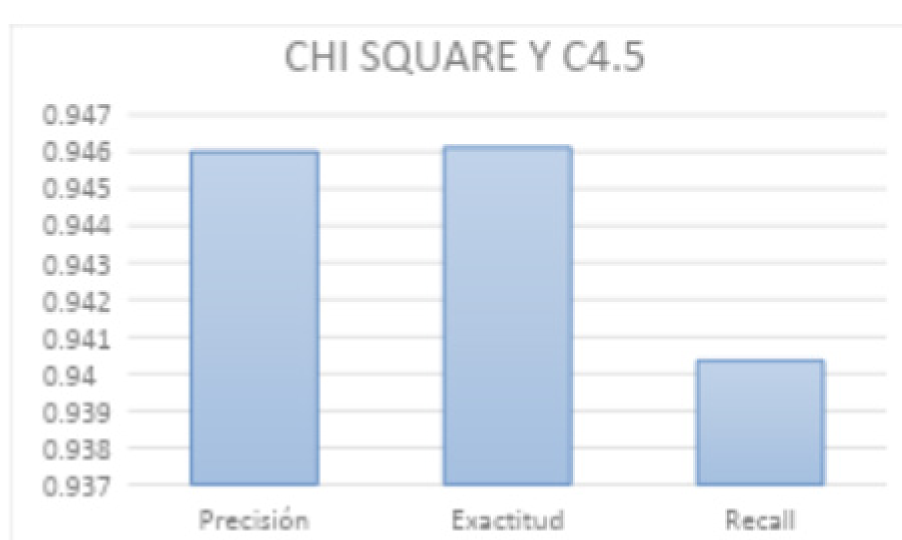

Figure 7. Variation in Chi-Square and C4.5 characteristics (13 characteristics)

[E2.2] Chi-Square y GHSOM

TABLE VIII. METRICS APPLIED TO THE TEST DATASET WITH CHI-SQUARE AND GHSOM, SIMULATION SCENARIO [E2.2]

METHOD	FEATURES	PRECISION	ACCURACY	RECALL
GHSON (CHI SQUARE) DATASET TEST	5	0.946	0,9461	0,9404
	6	0.946	0,9461	0,9404
	7	0.946	0,9461	0,9404
	8	0.946	0,9461	0,9404
	9	0.94	0,9397	0,9404
	10	0.946	0,9461	0,9404
	11	0.943	0,9431	0,9371
	12	0.935	0,9348	0,9345
	13	0.911	0,9105	0,8984

For this test scenario (E2.2), the Chi-Square technique was used, with the number of characteristics in the test dataset varied. Then GHSOM was applied, yielding better results with 10 characteristics. This gave the best results across the following metrics (Precision, Accuracy, and Recall), as shown in Table VIII. However, the best performance of the GHSOM model was achieved with 10 features, yielding precision of 94.60%, accuracy of 94.61%, and recall of 94.04%, representing increases of 3.5% in precision, 3.56% in accuracy, and 4.2% in Recall.

C. Scenario 3

For these experiments, feature variation was achieved using attribute selection (Relief) and the C4.5 classification algorithm.

[E3.1] Relief y C4.5

TABLE IX. METRICS APPLIED TO THE C4.5 TEST DATASET WITH RELIEF, WITH VARIATION IN CHARACTERISTICS

METHOD	FEATURES	PRECISION	ACCURACY	RECALL
C4.5 (RELIEF) DATASET TEST	5	0.946	0,9461	0,9404
	6	0.946	0,9461	0,9404
	7	0.946	0,9461	0,9404
	8	0.946	0,9461	0,9404
	9	0.946	0,9461	0,9404
	10	0.946	0,9461	0,9404
	11	0.946	0,9461	0,9404
	12	0.946	0,9461	0,9404
	13	0.946	0,9461	0,9404

The test scenario [E3.1] implemented the Relief technique, varying the number of features in the TEST dataset. The results we obtained using the C4.5 classification technique were stable across all features. This shows that the feature selection algorithm's results remain stable even with fewer features. Therefore, the best result is obtained with 13 features (see Table IX). That is, with 13 features, an accuracy of 94.6%, a precision of 94.61%, and a recall of 94.04% were achieved.

[E3.2] Relief Y GHSOM.

TABLE X. METRICS APPLIED TO RELIEF AND GHSOM TEST DATASETS, WITH VARIATION IN CHARACTERISTICS, SCENARIO [E3.2]

METHOD	FEATURES	PRECISION	ACCURACY	RECALL
GHSOM (RELIEF) DATASET TEST	5	0.946	0,9461	0,4997
	6	0.946	0,9461	0,4997
	7	0.946	0,9461	0,4997
	8	0.946	0,9461	0,4997
	9	0.946	0,9461	0,4997
	10	0.944	0,9440	0,5009
	11	0.946	0,9461	0,4997
	12	0.939	0,9387	0,4987
	13	0.93	0,9299	0,4953

In the test scenario (E3.2), Relief was used, with the number of features in the TEST dataset varied. GHSOM presented the best result with 11 features (see Table X). In addition, it shows more stable results for characteristics 9, 8, 7, 6, and 5. Similarly, with 13 characteristics, a Precision of 93.00%, an accuracy of 92.99%, and a recall of 49.53% were obtained. Note that the best result for GHSOM was achieved with 11 features, yielding 94.60% accuracy, 94.61% precision, and 49.97% recall.

CONCLUSIONS

The development of this study led to two key findings:

The results of scenario E1 show that the C4.5 algorithm consistently outperforms GHSOM when using all 13 features, achieving a training and testing accuracy of 93.80%, an accuracy of 93.77%, and a recall of 91.92%, while GHSOM performed worse, with a maximum accuracy of 90.74% and a recall of around 49%. When incorporating the Chi-Square technique (E2), C4.5 maintained stable performance, achieving 94.60% precision, 94.61% accuracy, and 94.04% recall with 13 features, with no significant variations when reducing dimensionality. In contrast, GHSOM showed a substantial improvement when using 10 features, achieving the same metric values while increasing precision by 3.5%, accuracy by 3.56%, and recall by 4.2% compared to using all features.

In scenarios using the Relief technique, C4.5 (E3.1) maintained consistent results across attribute counts, achieving precision of 94.60%, accuracy of 94.61%, and recall of 94.04%, confirming its robustness and stability across varying dimensionality. GHSOM (E3.2), on the other hand, performed best with 11 features, achieving an accuracy of 94.60%, precision of 94.61%, and a recall of 49.97%; However, when using 13 features, the recall dropped to 49.53%, showing that although GHSOM can achieve competitive precision and accuracy metrics, its limited positive recognition capacity places it below C4.5 for emotional state identification.

Overall, the results confirm that feature selection is key to improving the performance of dimensionality-sensitive algorithms such as GHSOM, achieving increases of more than 3% in precision and accuracy, while C4.5 maintains stable performance with minimal loss, regardless of the number of attributes.

Likewise, the methodology based on systematic scenario evaluation and 10-fold cross-validation proved effective in identifying optimal configurations, highlighting C4.5 as the most robust algorithm and GHSOM as a viable alternative when supported by appropriate feature selection techniques.

AUTHOR CONTRIBUTION

The authors' contributions to this article are as follows:

Gustavo Barraza: Research, data analysis, visualization, writing, and editing.

Johan Mardini: Results analysis, experiment design, data analysis, visualization, writing, and editing.

Ernesto Esmeral-Romero: Experiment design, writing, and editing.

The authors participated in the review of the results and approved the final version of the article.

CONFLICT OF INTERESTS

The authors declare that they have no interests or financial relationships that could have influenced this work.

REFERENCES

[1] J. J. Gross, "Emotion regulation: Current status and future prospects," Psychological Inquiry, vol. 26, no. 1, pp. 1–26, 2024.

[2] L. F. Barrett, "The theory of constructed emotion: Recent advances," Annual Review of Psychology, vol. 75, pp. 1–25, 2024.

3] A. K. Scherer, "Dynamic emotion processes in social interaction," Emotion Review, vol. 15, no. 2, pp. 123–135, 2023.

[4] D. M. Bogt et al., "Music as a tool for emotion regulation: A systematic review," Behavioral Sciences, vol. 14, no. 9, 2024.

[5] P. N. Juslin, "From everyday emotions to aesthetic emotions: Musical affect revisited," Psychology of Music, vol. 52, no. 1, pp. 3–21, 2024.

[6] A. Gabrielsson, Strong Experiences with Music, 2nd ed. Oxford, UK: Oxford Univ. Press, 2023.

[7] S. Eerola and T. Vuoskoski, "Music and emotion: Individual differences," Music Perception, vol. 41, no. 1, pp. 1–15, 2023.

[8] X. Chang et al., "Music emotion recognition using deep learning models," IEEE Access, vol. 12, pp. 45521–45534, 2024.

[9] Y. Li et al., "Rethinking emotion bias in music emotion recognition," IEEE Transactions on Affective Computing, early access, 2025.

[10] R. A. García-Hernández et al., "A systematic literature review of modalities, trends, and limitations in emotion recognition, affective computing, and sentiment analysis," Appl. Sci., vol. 14, no. 16, Art. no. 7165, 2024.

[11] Z. Yang et al., "Music tempo modulates emotional states as revealed through EEG insights," Sci. Rep., vol. 15, Art. no. 8276, 2025.

[12] "Development and application of emotion recognition technology a systematic literature review," BMC Psychology, vol. 12, Art. no. 95, 2024.

[13] X. Lu and S. Song, "Music-based emotion regulation: a bibliometric systematic review (2000–2024)," Front. Psychol., vol. 16, 2025.

[14] "Review on music emotion analysis using machine learning: technologies, methods, datasets, and challenges," Discover Appl. Sci., vol. 7, Art. no. 692, 2025.

[15] R. A. Calvo, S. D'Mello, J. Gratch y A. Kappas, "The Oxford handbook of affective computing," Oxford Univ. Press, 2020.

[16] A. Mollahosseini, D. Chan y M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," IEEE Trans. Affective Computing, vol. 11, no. 1, pp. 1–12, 2020.

[17] F. Eyben et al., "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," IEEE Trans. Affective Computing, vol. 7, no. 2, pp. 190–202, 2020.

[18] S. Ramakrishnan y I. S. M. Khalil, "Speech emotion recognition using openSMILE features and deep learning," IEEE Access, vol. 8, pp. 213–225, 2020.

[19] F. Eyben, M. Wöllmer y B. Schuller, "openEAR—Introducing the Munich open-source emotion and affect recognition toolkit," IEEE Trans. Affective Computing, vol. 11, no. 4, pp. 576–581, 2021.

[20] P. Poria, E. Cambria y A. Gelbukh, "Emotion recognition in multimodal sentiment analysis," IEEE Intelligent Systems, vol. 35, no. 2, pp. 34–42, 2020.

[21] R. Soleymani et al., "Multimodal emotion recognition in the wild," IEEE Trans. Affective Computing, vol. 13, no. 2, pp. 1–15, 2022.

[22] M. Hassani et al., "Cloud-based emotion recognition systems: A comparative study," IEEE Access, vol. 9, pp. 118345–118360, 2021.

[9] Fortune Business Insights, "Emotion detection and recognition market size, share & trends," Tech. Rep., 2023.

[23] A. Mollahosseini, B. Hasani y M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," IEEE Trans. Affective Computing, vol. 10, no. 1, pp. 18–31, 2020.

[24] Z. Li et al., "Deep facial expression recognition: A survey," IEEE Trans. Affective Computing, vol. 13, no. 3, pp. 1195–1215, 2022.

[25] S. Li y W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," IEEE Trans. Image Processing, vol. 31, pp. 1–14, 2023.

[26] D. Salcedo, C. Guerrero, K. Saeed, J. Mardini, L. Calderon-Benavides, C. Henriquez, and A. Mendoza, "Machine learning algorithms application in COVID-19 disease: a systematic literature review and future directions," Electronics, vol. 11, no. 23, p. 4015, 2022.

[27] C. Henríquez, J. Guzmán, and D. Salcedo, "Opinion mining based on the Spanish adaptation of ANEW on opinions about hotels," Natural Language Processing, vol. 56, pp. 25–32, 2016.

Author 1: is a systems engineer specializing in software development, with a master's degree in systems and computer engineering. His research interests include predictive modeling, artificial intelligence, and software automation. https://orcid.org/0009-0006-4911-9194.

Author 2: is a Colombian Electronic engineer, specialist in convergent networks and full-time professor in the Electronic Engineering Program at the Universidad de la Costa (CUC) in Barranquilla. Linked to the GIECUC research group. His academic and professional career focuses on applied research in areas such as artificial intelligence, machine learning, neural networks, computer vision, educational robotics, cardiovascular disease, and intrusion detection systems. He holds a master's degree in systems and computing engineering with and focus on research and is a PhD. in Information and Communication Technologies. https://orcid.org/0000-0001-6609-1687.

Author 3: professional in Electronic Engineering, master's degree in information technology governance, and Specialist in Business Management, with the ability to lead information and communication technology areas and projects. Currently a professor at Universidad de la Costa (Colombia). https://orcid.org/0000-0002-7526-8349.