User Experience Evaluation using Machine Learning and Facial Expressions: A Systematic Review

Bianca Lima Santos[1], **Luciano A. Digiampietri**[1]

[1]Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (USP)

{bianca_lima,digiampietri}@usp.br

Abstract. *This systematic review investigates the application of machine learning in facial expressions and emotion recognition within the realm of user experience (UX). The main objective is to identify advances in the state of the art regarding using facial expressions to detect emotions and, consequently, predict or improve user experience. The methodology provided a comprehensive analysis of existing literature, highlighting diverse definitions of UX and their implications for assessing user interactions with systems. Despite efforts to evaluate and enhance UX through various methodologies, few studies focus on predicting UX by integrating emotional states before interaction and user-reported experiences. This gap stems from the absence of a unified UX definition, complicating methodological standardization and result comparability across studies. Many reviewed works emphasize developing recommendation algorithms tailored to music, news, and other content to optimize UX through emotional data. The review identified the challenge of establishing a consistent framework for UX definition across research, revealing varied approaches using different datasets aimed at enhancing recommendations, improving user satisfaction, comparing perceived attitudes, and integrating with established questionnaires.*

1. Introduction

Throughout the day, individuals interact with various items that become integral to their routines, such as office chairs, television remote controls, cell phone applications, and electronic games. The way a user interacts with a given product is referred to as user experience (UX).

According to ISO 9241-210 [ISO 9241-210 2010], UX encompasses a person's responses and perceptions related to the use of elements, including beliefs, emotions, preferences, physical and psychological responses, behaviors, and achievements that occur before, during, and after using a product, system, or service. One critical dimension in evaluating user experience is the emotions felt by the user during the interaction [Bernhaupt 2015, Veriscimo et al. 2021].

Despite its importance, evaluating UX is complex due to the involvement of various factors [ISO 9241-210 2010]. Different approaches exist for UX evaluation, but many do not adequately emphasize the user's emotions when interacting with a product [Veriscimo et al. 2020]. Additionally, cultural and linguistic differences can influence how users interpret questionnaires and other evaluation methods. Thus, UX evaluation requires selecting appropriate methods that comprehensively capture the user experience, including emotions and feelings during interaction with a product or system.

Considering facial expression recognition, automatic emotion recognition is related to classifying emotion archetypes, which are models of facial expressions representing specific emotions [Cowie et al. 2001, Dias et al. 2023]. However, recognizing and understanding the nature of emotions is not an end in itself.

While questionnaires and interviews are widely used for user experience analysis, numerous studies are exploring the automation of UX assessment. Traditional methods, such as surveys and interviews, are costly and can introduce bias and result in inaccurate data regarding users' emotions due to human factors such as exaggeration, embarrassment, and forgetfulness [Koonsanit and Nishiuchi 2020].

The use of artificial intelligence has become increasingly widespread, with techniques and tools emerging in various application areas. Many studies have focused on developing algorithms capable of recognizing emotions, particularly through facial expressions [van Erven and Canedo 2023]. These studies introduce new possibilities for automatically evaluating user experience and recommending personalized content, thus enhancing the overall user experience.

This systematic literature review aims to compile and analyze existing machine learning methods and techniques that utilize the recognition of facial expressions and emotions to predict or enhance user experience. This focus is motivated by the increasing development of digital platforms and the challenging process of evaluating user experience.

In this study, advances in the state of the art regarding the use of facial expressions to detect emotions and, consequently, predict or improve user experience are investigated. The review will observe and present the approaches used for emotion detection, the tools and strategies applied, and the methods for processing emotions. This includes both improving the user experience and making UX predictions for some specific digital system.

2. Methodology

Faced with the challenge of evaluating studies focused on promoting a better user experience in an automated way, this research aims to identify and analyze existing machine learning methods and techniques that use recognition of facial expressions and emotions for prediction or improving the user experience, aiming to identify gaps in the literature and understand the state of the art. To achieve this, the following research questions guided the study:

1. What attributes are used to detect emotions?
2. What are the techniques used to detect emotions?
3. What are the categorized emotions?
4. How is user experience defined?
5. How does the study predict or improve user experience?
6. Which dataset was used?
7. What are the reported limitations and challenges?

2.1. Search Strategy and Information Sources

Key scientific sources in the field of computing were utilized.

1. ACM Digital Library;

2. IEEE Xplore Digital Library;
3. ScienceDirect Digital Library.

These platforms were chosen for their extensive publication coverage in the areas of computer science, engineering, and information technologies. These concepts are fundamental for finding the state of the art in the applicability of machine learning to, based on facial expressions, impact the user experience.

The following keywords were used in the search for publications: "machine learning", "artificial intelligence", "deep learning", "user experience", "ux", "facial expression", "emotion detection".

Articles published in English between 2018 and 2024 were included in the review.

3. Inclusion and Exclusion Criteria

Specific inclusion and exclusion criteria were defined to ensure the relevance and quality of the studies included in this systematic review. These criteria ensure that only pertinent studies are considered in the analysis.

3.1. Inclusion Criteria

Only articles that meet all the inclusion criteria below were included:

- (a) Recent studies, from 2018 to 2024;
- (b) Studies with at least six pages;
- (c) Studies that are fully available for access;
- (d) Published studies.

3.2. Exclusion Criteria

Articles that met at least one of the following exclusion criteria were removed from review:

- (a) Studies that have not gone through a peer review process;
- (b) Studies that do not focus on user experience in digital systems;
- (c) Studies that are not empirical, such as reviews, opinions, or comments;
- (d) Studies that do not use facial recognition;
- (e) Duplicate studies;
- (f) Studies that do not present clear performance metrics;
- (g) Studies that do not focus on predicting or improving the user experience.

4. Study Selection

To select the studies, a search string was created using keywords, their synonyms, and acronyms. This string was submitted to the search engines of the selected sources. The initial screening was based on reading the titles and abstracts of the articles. Subsequently, the inclusion and exclusion criteria were applied, resulting in the selection of articles included in this review.

5. Summarization and Synthesis

After selecting the articles, each work was read in full. Based on this reading, detailed summaries of each article were prepared, highlighting the following key aspects:

- Objectives of the study;
- Machine learning methods used;
- Purpose of the methods used;
- Techniques for recognizing facial expressions and emotions;
- Evaluation of results;
- Challenges and limitations identified.

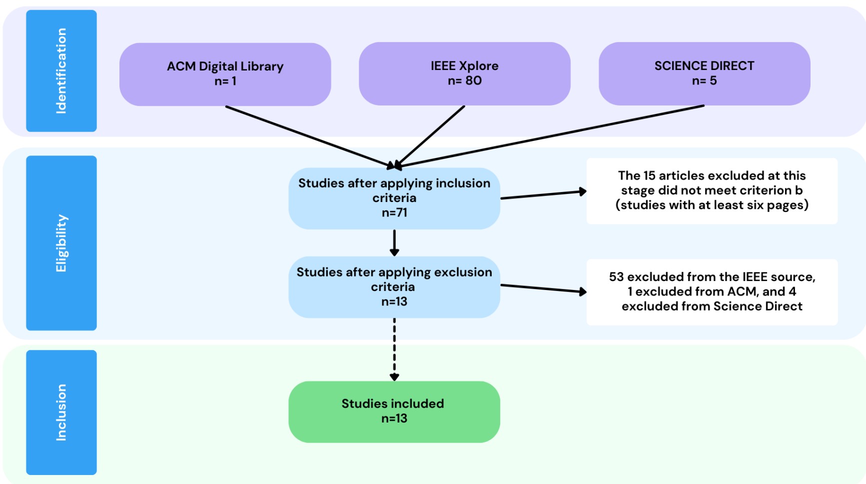
At the end of the readings, a technical report was prepared summarizing the methods identified in the studies. This technical report was used to answer the research questions guiding this study. Furthermore, after implementing the described method, the existing gaps in the literature and the main challenges encountered were highlighted.

6. Conducting

The search was carried out in the digital libraries *ACM Digital Library*, *IEEE Xplore Digital Library* and *Science Direct*, searching in the title, abstract and keywords fields for articles published between 2018 and 2024, using the following search *string*:

```
("user experience" OR "UX") AND
("artificial intelligence" OR "machine learning" OR "ml") AND
("facial expression" OR "emotion recognition")
```

Figure 1. Systematic review conduction diagram



The search resulted in a total of 86 articles: 1 article from ACM DL, 80 from IEEE, and 5 from Science Direct. After applying the inclusion and exclusion criteria, the set was reduced to 13 articles, with 1 from Science Direct, representing 7.69% of the selected articles, and 12 from IEEE, representing 92.31%. The conduction flow is illustrated in Figure 1.

The most common criterion for excluding articles was the lack of focus on user experience in digital systems. The second predominant criterion was the lack of use of

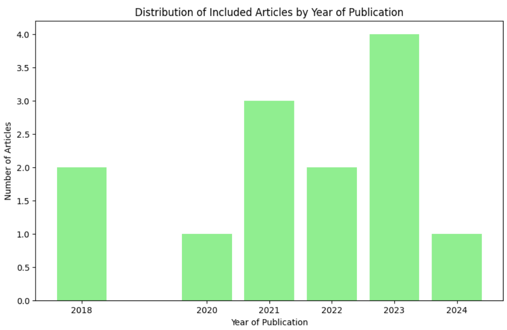


Figure 2. Distribution by year of publication

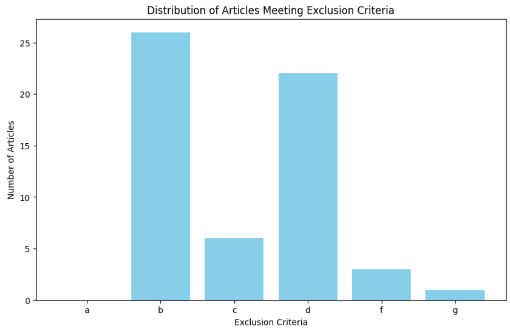


Figure 3. Distribution of exclusion criteria

facial recognition, even with the use of specific keywords for this topic in the search *string*. The representation of the distribution of exclusion criteria met by the articles is shown in Figure 3. For simplicity, in cases where an article met multiple exclusion criteria, only the first identified criterion was considered for this representation.

Figure 2 shows the distribution of selected articles by year of publication. There is a concentration of articles in 2023, with no articles published in 2019.

7. Results

The results were synthesized both qualitatively and quantitatively. The main approaches, tools, and strategies used were identified. We highlighted gaps and challenges in the literature, and analyzed trends and future research directions.

Tables 1 and 2 summarizes the 13 articles selected in the context of this review.

7.1. Emotions

The analysis of the selected articles revealed that 4 out of the 13 works used the list of emotions (“anger”, “disgust”, “fear”, “happiness”, “sadness”, “surprise”, “neutral”) to detect emotions based on users’ facial expressions [Qian et al. 2018, Koonsanit and Nishiuchi 2020, Liu and Lee 2018, Isman et al. 2021].

Only one of the studies did not use any of the aforementioned emotions [Karimah et al. 2024]. This article classified users as “very engaged”, “typically engaged”, and “not engaged”. The remaining studies used at least two of the six emotions: “anger”, “disgust”, “fear”, “happiness”, “sadness”, and “surprise”. The frequent use of the “neutral” emotion was also observed, used to classify the absence of detectable expressions in the user. The frequency with which each emotion was used as a label in the articles is shown in Figure 4.

Approximately 76% of the studies used only the user’s face as an attribute for detecting emotions. Three studies used additional attributes: in [Qian et al. 2018], the user’s voice was also used; in [Koonsanit and Nishiuchi 2020], the user’s gender and age were considered; and in [Kwon et al. 2022], the pose of the user’s head was also relevant. Figure 5 shows the distribution of these attributes.

When carrying out an in-depth analysis of strategies for detecting emotions in the articles reviewed, it was observed that approximately 77% of the studies employed one or

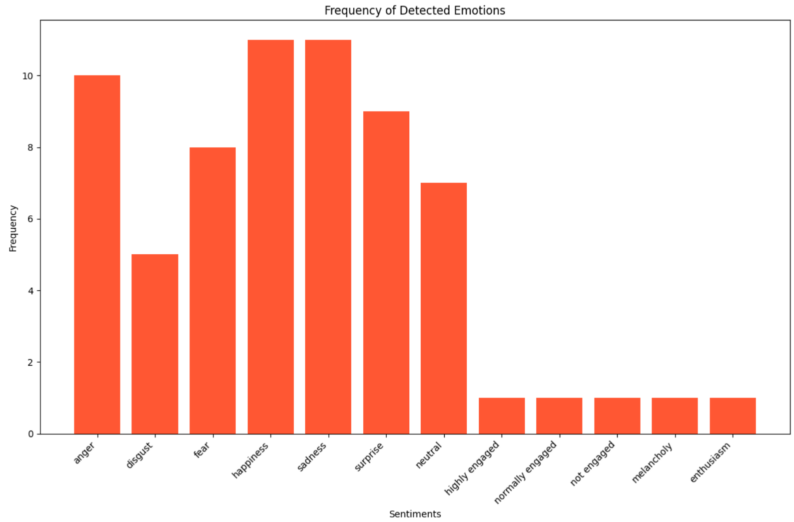


Figure 4. Frequency of emotions detected in the selected articles

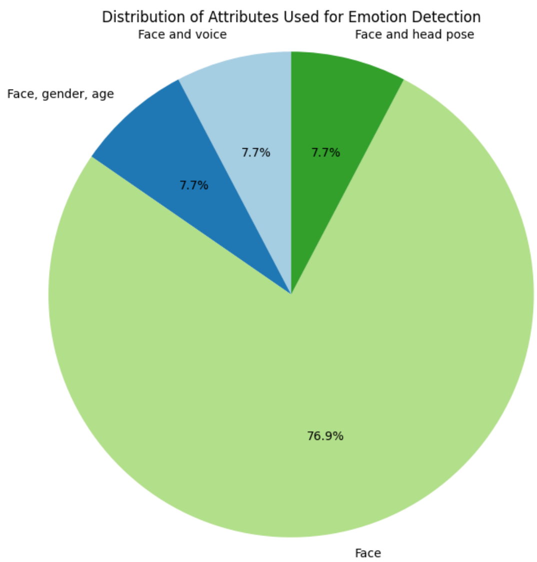


Figure 5. Attributes used to detect emotions

Article	Year	Context	Characteristics for Emotion Detection	Categorized Emotions
[Qian et al. 2018]	2018	Generic	Face and voice	Anger, disgust, fear, happy, sad, surprise and neutral
[Koonsanit and Nishiuchi 2020]	2020	Video	Face, gender, age	Anger, disgust, fear, happy, sad, surprise and neutral
[Afriansyah et al. 2021]	2021	Games	Face	Anger, happy, neutral, sad and surprise
[Liu and Lee 2018]	2018	Generic - suggests use for games	Face	Anger, disgust, fear, happy, sad, surprise and neutral
[Isman et al. 2021]	2021	Generic - suggests use for games	Face	Anger, disgust, fear, happy, sad, surprise and neutral
[Kwon et al. 2022]	2022	Games	Face and head pose	Anger, disgust, fear, happy, sad and surprise
[Chimienti et al. 2022]	2022	Movies	Face	Neutral, sad and happy
[Karimah et al. 2024]	2024	E-learning	Face	Very Engaged, Typically Engaged, and Not Engaged
[Gupta et al. 2023]	2023	Music	Face	Happy, sad, anger, surprise, melancholy, enthusiasm
[Eliyajer et al. 2023]	2023	Music	Face	Sad, anger, fear, happy, neutral, surprise
[Selvi et al. 2023]	2023	Music	Face	Happy, sad, fear and anger
[Ashani Malsha et al. 2021]	2021	News	Face	Quote that categorizes into 7 emotions, but does not list them
[Mabel Rani et al. 2023]	2023	Music	Face	Happiness, sadness, surprise, fear and anger

Table 1. Summary of selected articles - Part 1

more Convolutional Neural Networks (CNN) techniques (Figure 6). The preference for this class of neural networks is a consequence of their notorious specialization in processing visual data. CNNs use a robust and adapted architecture for extracting hierarchical and spatial features from images, making them exceptionally efficient in classification and visual pattern recognition tasks, as is the case with the problem of classifying emotions given facial expressions.

The study by [Qian et al. 2018] exemplifies this trend by employing three distinct CNN-based architectures for emotion classification: Google-Net, Residual-Net, and VGG-Net. Each of these architectures was selected for their respective feature extraction capabilities.

Similarly, [Koonsanit and Nishiuchi 2020] used the Facial-expression-keras approach, demonstrating the versatility and effectiveness of CNNs in capturing and interpreting human facial expressions. Furthermore, [Kwon et al. 2022] adopted the EfficientNet-B0 architecture, known for its efficiency in terms of performance and resource consumption, while [Chimienti et al. 2022] opted for ResNet50, a CNN architec-

Article	Dataset for Detecting Emotions	Technique(s) Used for Detecting Emotions	Focus of the study
[Qian et al. 2018]	Own Dataset	CNN (VGG-Net, Google-Net, Residual-Net)	UX Improvement
[Koonsanit and Nishiuchi 2020]	Own Dataset	CNN (Facial-expression-keras)	UX Prediction
[Afriansyah et al. 2021]	Indonesia Mixed Emotion Dataset (IMED)	KNN	UX Prediction
[Liu and Lee 2018]	FER2013	CNN-SVM	UX Prediction
[Isman et al. 2021]	FER2013	CNN	UX Prediction
[Kwon et al. 2022]	WIDER (face detection), AFLW2000 (head pose detection)	CNN (EfficientNet-B0)	UX Prediction
[Chimienti et al. 2022]	FER2013	CNN (ResNet50)	UX Prediction
[Karimah et al. 2024]	Own Dataset	Random Forest	UX Prediction
[Gupta et al. 2023]	Own Dataset	CNN	UX Improvement
[Eliyajer et al. 2023]	FER2013	CNN	UX Improvement
[Selvi et al. 2023]	FER2013	CNN	UX Improvement
[Ashani Malsha et al. 2021]	FER2013	CNN	UX Improvement
[Mabel Rani et al. 2023]	Kaggle dataset - unspecified	CNN (Modified Convolution Neural Network)	UX Improvement

Table 2. Summary of selected articles - Part 2

ture that has improved significantly in accuracy and performance in image classification and complex pattern recognition tasks.

In [Karimah et al. 2024], Random Forest was the only method used to classify facial expressions. However, within the context of the article, the focus was not on classifying basic human emotions (“anger”, “disgust”, “fear”, “happiness”, “sadness”, and “surprise”) but rather on categorizing students into three engagement levels: very engaged, normally engaged, and not engaged.

7.2. User Experience

User experience (UX) definitions vary among the selected articles, reflecting the diverse approaches and contexts of UX. UX is perceived, in general, as an interaction between the user and the system, where emotional and functional aspects play crucial roles. However, the methods for measuring UX vary according to the definition adopted in each study.

The paper [Liu and Lee 2018] focuses on the development of a facial emotion classifier, suggesting that UX is defined by the ability to detect and respond to users’ emotions. Although this approach provides information about the user’s emotional state, it can be considered limited as it reduces UX to a simple emotional classification, without considering other contextual and interactive factors.

In contrast, [Karimah et al. 2024] uses the User Experience Questionnaire (UEQ) to collect responses from users in six distinct dimensions: attractiveness, perspicuity, efficiency, reliability, stimulation, and novelty. This work takes a comprehensive approach to measuring UX, integrating emotional and functional aspects of interaction with the system. Furthermore, the study employs facial recognition methods to assess users’ emotional engagement during interactions. Data collected from facial expressions is combined with UEQ responses, providing detailed and robust UX analysis to improve accuracy in capturing users’ feelings and reactions in real time. This method seeks to ensure that both subjective and objective aspects of the user experience are considered and analyzed.

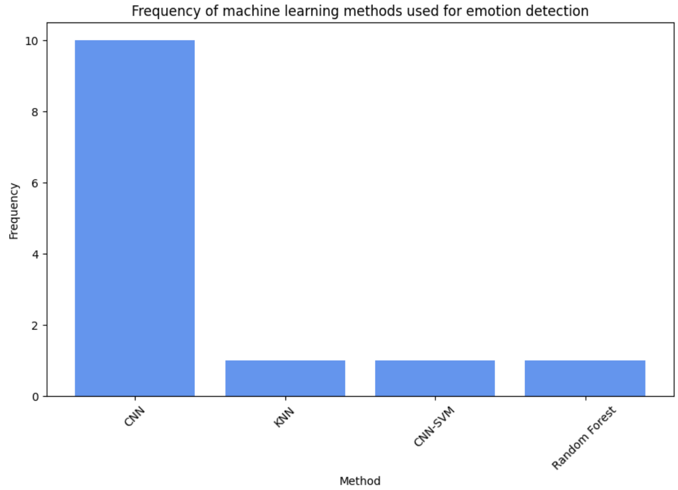


Figure 6. Emotions detection methods

In the article [Chimienti et al. 2022], questionnaires were also used so that users could determine their satisfaction with the system after interacting with it. Users tested the system through tasks and completed questionnaires (Expressing Mixed Emotions, SUS, and SUPR-Q) after each interaction to evaluate usability and user experience. The collection of data provided by users as form responses, together with facial recognition, allowed the system to provide personalized recommendations, increasing user satisfaction and engagement.

In the [Koonsanit and Nishiuchi 2020] study, the final accuracy is determined based on the prediction of user experience, using features extracted from emotion detection, as well as age and gender. The evaluation is compared with the answers to a form that contains a single question rating the experience perceived by the user: “From 1 to 5, what is the chance of you recommending the film you watched to a friend or colleague?”.

7.3. Datasets

The datasets used reveal significant variation in terms of data sources, reflecting the diversity of approaches to detect facial emotions. A predominance of the FER2013 dataset was noted, used in six of the studies analyzed [Liu and Lee 2018, Isman et al. 2021, Chimienti et al. 2022, Eliyajer et al. 2023, Selvi et al. 2023, Ashani Malsha et al. 2021]. The distribution of the different datasets used can be observed in Figure 7. FER2013 is a widely recognized dataset containing images labeled for different emotions, which explains its popularity due to its availability and relevance to the task of emotional recognition.

In addition to FER2013, other datasets were mentioned, including the Indonesia Mixed Emotion Dataset (IMED), used in [Afriansyah et al. 2021], and a Kaggle dataset, whose details were not provided in [Mabel Rani et al. 2023].

The frequency of use of proprietary datasets in four of the articles analyzed highlights the tendency to customize data to meet specific needs of experiments [Qian et al. 2018, Koonsanit and Nishiuchi 2020, Karimah et al. 2024, Gupta et al. 2023]. Although this allows for precise adaptation to the objectives of each research, it also makes it difficult to replicate and independently validate the results.

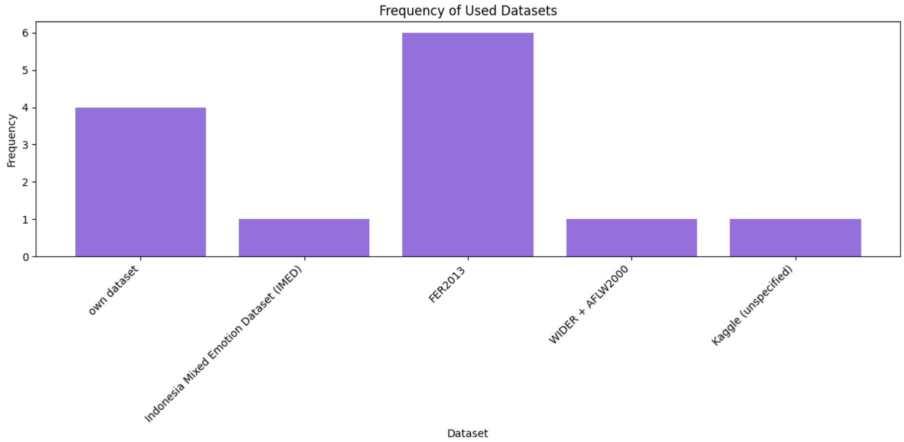


Figure 7. Datasets

7.4. Challenges and Limitations

While reading the articles, several significant limitations and challenges in implementing facial emotion recognition systems for evaluating user experience (UX) emerged.

The main limitation reported among the articles is the influence of lighting conditions on the accuracy of emotion recognition [Gupta et al. 2023]. Images taken with low light intensity are less sharp, while very high light intensity can distort facial expressions, making it difficult to accurately identify emotions. Furthermore, there is a notable difference between results obtained in controlled laboratory environments and those in practical environments, such as users' homes [Kwon et al. 2022]. Proximity of user's face to the camera is also a reported significant impact factor [Kwon et al. 2022].

Imbalance of facial expression classes was also a reported challenge [Kwon et al. 2022]. Certain expressions, such as happiness and neutrality, are displayed more frequently, while expressions such as disgust and fear are less common, resulting in an imbalance that can affect the accuracy of recognition algorithms. Furthermore, some articles consider only a limited number of emotions, as in [Chimienti et al. 2022], which used only happiness, sadness, and neutrality.

There is also high latency in cases of real-time processing that can impact emotion detection [Mabel Rani et al. 2023]. Other difficulties reported are: variability in facial expressions due to cultural, ethnic, and personal differences, as well as the complexity and subjectivity of human emotions [Gupta et al. 2023, Mabel Rani et al. 2023].

8. Conclusions

The methodology defined to conduct this systematic review provided a comprehensive and detailed analysis of the methods and techniques existing in the literature on the application of machine learning in facial expressions and emotion recognition in the context of user experience.

This review showed that the different definitions of UX reflect the complexity and multifaceted nature of the user experience and the challenge of building a tool to diagnose what a user's experience was like when interacting with a system.

This review identified several studies to evaluate and improve user experience. However few studies focused on applying these methods to predict UX, taking into account attributes such as the user's emotional state before interacting with the system, as the user declares to have been his experience with the system. This gap can be attributed to the lack of a consensual definition of UX, which varies significantly between articles. The diversity of definitions found makes it difficult to standardize methodologies and results.

In the context of this review, it was observed that many studies concentrate on developing recommendation algorithms for music, news, or other content to enhance UX. These algorithms utilize emotional data to personalize recommendations, aiming to boost user satisfaction and engagement.

The review underscored the significant challenge of establishing a consistent pattern for defining user experience. Each article approaches the term differently, whether to enhance recommendations, improve user happiness, compare perceived attitudes, or evaluate alongside established questionnaires in the field.

References

Afriansyah, Y., Nugrahaeni, R. A., and Prasasti, A. L. (2021). Facial expression classification for user experience testing using k-nearest neighbor. In *2021 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pages 63–68.

Ashani Malsha, N. P. G., Dulmini Heshani, K., Ransara, R. K., Ayesh Bandara, D. M. D. D., Suriyaa Kumari, P. K., and Anuththara Kuruppu, T. (2021). Automated sinhala news platform based on machine learning and deep learning. In *2021 3rd International Conference on Advancements in Computing (ICAC)*, pages 134–139.

Bernhaupt, R. (2015). *User Experience Evaluation Methods in the Games Development Life Cycle*, pages 1–8. Springer International Publishing, Cham.

Chimienti, M., Danzi, I., Gattulli, V., Impedovo, D., Pirlo, G., and Veneto, D. (2022). Behavioral analysis for user satisfaction. In *2022 IEEE Eighth International Conference on Multimedia Big Data (BigMM)*, pages 113–119.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80.

Dias, B., Ivamoto, V., and Lima, C. (2023). Detecting faces in specific scenarios: Systematic literature review. In *Anais do XX Encontro Nacional de Inteligência Artificial e Computacional*, pages 540–554, Porto Alegre, RS, Brasil. SBC.

Eliyajer, G., Natarajan, B., Bhuvaneswari, R., and Elakkiya, R. (2023). A novel approach for song recommendation system using deep neural networks. In *2023 IEEE World Conference on Applied Intelligence and Computing (AIC)*, pages 382–387.

Gupta, M., Venkatesh, S. N., Suraskar, R., Praveena, K., Jegadesan, S., and Suneetha, S. (2023). Enhancing music recommendations with emotional insight: A facial expression approach in ai. In *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 450–457.

Isman, F. A., Prasasti, A. L., and Nugrahaeni, R. A. (2021). Expression classification for user experience testing using convolutional neural network. In *2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, pages 1–6.

ISO 9241-210 (2010). Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. Standard ISO 9241-210:2010, International Organization for Standardization.

Karimah, S. N., Phan, H., Miftakhurrokhmat, and Hasegawa, S. (2024). Design principle of an automatic engagement estimation system in a synchronous distance learning practice. *IEEE Access*, 12:25598–25611.

Koonsanit, K. and Nishiuchi, N. (2020). Classification of user satisfaction using facial expression recognition and machine learning. In *2020 IEEE REGION 10 CONFERENCE (TENCON)*, pages 561–566.

Kwon, S., Ahn, J., Choi, H., Jeon, J., Kim, D., Kim, H., and Kang, S. (2022). Analytical framework for facial expression on game experience test. *IEEE Access*, 10:104486–104497.

Liu, X. and Lee, K. (2018). Optimized facial emotion recognition technique for assessing user experience. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pages 1–9.

Mabel Rani, A. J., S, M., Jothi Swaroopan, N. M., and Hari Kumar, K. (2023). Face emotion based music recommendation system using modified convolution neural network. In *2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, pages 1–6.

Qian, Y., Lu, J., Miao, Y., Ji, W., Jin, R., and Song, E. (2018). Aiem: Ai-enabled affective experience management. *Future Generation Computer Systems*, 89:438–445.

Selvi, A. S., S, A., Hariharan, M., and P, A. (2023). Emotune: Deep emotion detection and music recommendation system using mobilenetv3. In *2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, pages 1–6.

van Erven, R. C. G. a. S. and Canedo, E. D. (2023). Measurement of user's satisfaction of digital products through emotion recognition. In *Proceedings of the XXII Brazilian Symposium on Software Quality*, SBQS '23, page 62–71, New York, NY, USA. Association for Computing Machinery.

Veriscimo, E. d. S., Bernardes-Júnior, J. L., and Digiampietri, L. A. (2020). Evaluating User Experience in 3D Interaction: a Systematic Review. In *Proceedings of the XVI Brazilian Symposium on Information Systems - SBSI*, SBSI '20, New York, NY, USA. Association for Computing Machinery.

Veriscimo, E. d. S., Bernardes-Júnior, J. L., and Digiampietri, L. A. (2021). Facial emotion recognition in ux evaluation: A systematic review. In Kurosu, M., editor, *Human-Computer Interaction. Theory, Methods and Tools*, pages 521–534, Cham. Springer International Publishing.