# AN OVERVIEW OF THE MACHINE'S "UNDERSTANDING" OF HUMAN EMOTIONS

# UNA PANORAMICA SULLA "COMPRENSIONE" DELLE EMOZIONI UMANE DA PARTE DELLA MACCHINA

**Umberto Bilotti**
**Università degli Studi di Salerno**
**ubilotti@unisa.it**

**Michele Domenico Todino**
**Università degli Studi di Salerno**
**mtodino@unisa.it**

**Maurizio Sibilio**
**Università degli Studi di Salerno**
**msibilio@unisa.it**

**Double Blind Peer Review**

**Citazione**
Bilotti, U., Todino, M.D., & Sibilio, M. (2024). An overview of the machine's "understanding" of human emotions. *Giornale Italiano di Educazione alla Salute, Sport e Didattica Inclusiva*, 8(2), Edizioni Universitarie Romane.

**Doi:**
https://doi.org/10.32043/gsd.v8i2.1138



**gsdjournal.it**
**ISSN: 2532-3296**
**ISBN 978-88-7730-493-3**

**ABSTRACT**
The use in the educational context of Artificial Intelligence-based technologies developed to automate the process of intepreting the human emotial state, must be the subject of reflection by the pedagogical community. In this work, the potential levels of machine understanding of human emotions are assessed, and the issues of applicability are noted in order to favour the creation of tools that are consistent with a simplex didactic action and in line with the ideas on emotive intelligence.

L'utilizzo in ambito educativo di tecnologie basate sull'Intelligenza Artificiale, sviluppate per automatizzare il processo di interpretazione dello stato emotivo umano, deve essere oggetto di riflessione da parte della comunità pedagogica. In questo lavoro si valutano i potenziali livelli di comprensione delle emozioni umane da parte delle macchine e si rilevano le problematiche di applicabilità per favorire la creazione di strumenti coerenti con un'azione didattica semplessa ed in linea con le idee sull'intelligenza emotiva.

**KEYWORDS**
Artificial Intelligence, Emotion Recognition, Sentiment Analysis, Seamless Learning
Intelligenza Artificiale, Riconoscimento delle Emozioni, Analisi dei sentimenti, Apprendimento senza soluzione di continuità

**Received 30/04/2024**
**Accepted 15/06/2024**
**Published 24/06/2024**

## Introduction

A high level of understanding of the emotional state is certainly a desirable characteristic for building the teacher-student and student-student relationship. The reason for the first case is well explained in Rivoltella's work (Rivoltella, 2014) as a relevant component for the teacher's ability to 'hold' the class. The second scenario, however, may be attributed to the success of peer-to-peer learning strategies that provide an appropriate balance between competence and self-recognition in the lived experience (Bonaiuti, Calvani and Picci. 2014). As a result, the shared reliance on the affective component of interclass relationships increases the teacher's responsibility for continuous and systemic observation of possible and multimodal related manifestations. One potential application of analyzing various biometric traits, such as facial expression, voice, gesticulation style, and body posture, is automatic emotion recognition. These physical and behavioral cues enable the interpretation of emotions expressed by individuals, facilitating communication, understanding, and empathy. This technology finds applications in social avatars, empathetic virtual assistants, and personalized e-learning platforms, among other uses (Bilotti et al. 2024). As we will reflect at several points in this contribution, the correct detection of the student's emotional state is not a definitive goal. On the contrary, it directs the teacher's action to pursue the principles expressed by Universal Design for Learning, which are opposed to the conception of the 'average student'. For each student, in fact, the learning process occurs in a unique way because it depends on variables that operate in continuous resolution to the context in which they are immersed and to their personal interests. Consequently, the competences to be developed cannot derive from a single type of content, delivered in a single form, and proposed in pursuit of a single purpose (Savia 2016). In this paper, after introducing the main classes of biometric methods related to emotion recognition, we will try to explain what the levels of understanding of human emotion by machines can be, finally highlighting the problems arising from their application in the educational-didactic context.

## 1. Classification and emulation

The understanding of emotional phenomena by the machine and in particular by Artificial Intelligence (AI) is tied to two main tasks: the first, which has a longer tradition in the literature, is that of classification, the other, which has gained particular importance with the extensive use of Generative Artificial Intelligence (GAI), is that of the synthetic creation of text or multimedia files that can express or trigger an emotional phenomenon of the human beings.

## 2. Emotion Detection

The problem of classifying an emotional state is strongly linked to the type of data analyzed: biometric techniques for emotion detection can analyze images or videos of facial expressions (Li et al.  2024), audio files with voice recordings (Suganya and Charles. 2019) neurological signals (Xie, Zhou, and Sun. 2021)., or specific body postures or movements (Noroozi F. et al. 2021); in addition, there are methods for emotion detection from text (Cahyani et al. 2022). Depending on the type of data, different classifications have been identified and the task then required of the machine is to correctly associate each individual element with one of the predetermined classes of emotion. A long-standing multidisciplinary reflection on the human emotional spectrum has proposed several classification models, although the one commonly used is *the discrete approach*, which involves a finite number of classes: from the simple partitions into positive, neutral and negative emotions of EEG signals to the more complex ones of five, six or eight *basic emotions* typical of image, audio and video datasets. Given the strong interest in commercial and security applications, the last decade has been characterised by a "race" to find the most efficient method in terms of accuracy and computational time ratio. Some work, such as that of (Lévêque, L. et al. 2022), has shown that there are now methods capable of solving this classification problem with comparable or equal performance to human beings. It should be highlighted, nonetheless, that comprehending a person's emotional state is a process that cannot be boiled down to the categorization of a large collection of phenomenologies. In fact as Ekman P. (1999) explains, it is possible to establish characteristics that can distinguish basic emotions from other affective phenomena, but no one of these characteristics, such as in our case phisiological signals, realises a univocal correspondence with them.

Figure 1: Sample images of CK+, a dataset used to do Facial Expression Recognition.

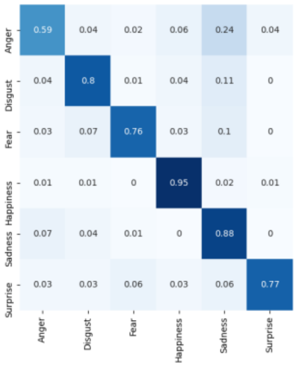


Figure 2: Example of a confusion matrix used to calculate the accuracy of an Emotion Recognition method.

## 3. Emotion feature's Approximation

The phenomena related to emotional states are both a product of our personal experience flow and an adaptive behavioural inheritance from our evolutionary history. This principle is in fact the necessary condition for a common denominator of all human productions that succeed in preserving emotional potential. Throughout human history, artistic achievements have played a crucial role in representing and expressing important ideas in a variety of worldwide civilizations. These visual representations provide a unique viewpoint on how different civilizations understand and communicate the complexities of their reality through visual media. Gaining insight into the expressive potential that each of these pieces of art has becomes an essential interpretive tool that opens up access to deep levels of human experience, and when expertly performed, their reinterpretation could serve as a means of emotional experimentation (Bilotti et al. 2023). The landscape of Generative Artificial Intelligence (GAI) applications has expanded significantly, offering a plethora of tools that can generate text, audio, image, and video files in response to textual prompts. This advancement has democratized content creation to a considerable extent, enabling users to generate diverse media types with relative ease. However, while certain types of synthetic data, such as text and occasionally images, may be intuitive and accessible to a broad audience, the production of other forms of synthetic data remains a complex endeavor. These data types often require specialized skills or access to powerful computing resources that are beyond the reach of standard laboratory or classroom settings, particularly in primary and secondary education environments. Moreover, the requests processed by GAI applications are becoming increasingly specific and nuanced over time. Simple requests, such as asking for a joke or a scary story, or requesting an image designed to evoke surprise in a specific audience, implicitly demand elements that are statistically proven to elicit certain emotions. This places a premium on the ability of GAI systems to understand context and tailor their outputs accordingly. As GAI technology continues to evolve, there is a growing emphasis on enhancing its capability to generate content that not only meets the basic criteria of the request but also resonates emotionally with the intended audience. This involves leveraging advanced algorithms and machine learning techniques to analyze patterns in human behavior and preferences, enabling GAI systems to produce content that is engaging, immersive, and emotionally resonant.

Figure 3: Images generated by DALL-E 2 through a simple sequence of prompts 'create image of the same child (happy, bored, confused, focused, surprised)'.

## 4. Some applications in the educational context

In the past  years, there has been a surge in educational research experimenting with AI-based tools. Many of the high-performance approaches for Emotion Recognition (ER) that we have seen in the literature are either Machine Learning or Deep Learning techniques.  The study of Ma et al. (2018) suggests an application that enables instructors and online learning environments to ascertain students' learning emotions in real time, increase the efficacy of lessons, and boost student engagement and communication. A special feature of the proposed development is the system of generating timely reminders or encouragement when depressed and thus improving learning outcomes. An example of an application of interest, as it is designed for a category of students with special educational needs, is proposed in the work of Campitiello et al. (2022). An ER approach is utilised in the to teach kids with ASD how to identify and mimic facial expressions that are typically associated with one of seven emotions (six fundamental emotions and the neutral emotive state). Several applications have been developed to counter or manage negative emotions, particularly anger or anxiety, at different ages, but many of these do not have an AI component.  The work of Barron-Estrada M.L. and Zatarain-Cabada R. (2019) proposed a comprehensive framework involving experts from different scientific-disciplinary fields. Starting from the answer to a particular task proposed by the teacher, a Deep Learning technique analyses the text and tries to estimate the emotional state of the student who made it. Then the results are interpreted by a pedagogical agent who returns feedback to the student and updates a model with those considerations. Finally, the teacher, through the updated model, can propose a new task in a more consistent way to the student's emotional state.

## 5. Applicability Issues

### 5.1 Distance Learning and 2D platforms

From the various review works such as that of Saxena et al. (2020) it is possible to identify emotion detection methods in the literature with high levels of accuracy for each case introduced in the previous section. In particular, Facial Expression Recognition methods are efficient in both image and video analysis. However, due to the controlled conditions of light and face position of the various datasets used, the most consistent application scenario in the educational-didactic context is distance learning through 2D platforms. In addition to the strong constraint on the mode of the lesson, one must therefore consider all the widely discussed problems with 2D platforms for teaching. The experimentation with Distance Learning and e-

Learning that started in the last decade, for reasons of necessity, took on a global character during the pandemic period. As a matter of fact, one of the positive 'collateral' aspects solved one of the main difficulties of research in this field, namely the availability of the sample. The students and lecturers who populated the thousands of online meetings held daily definitively confirmed not only the potential but also, and most forcefully, the limitations of the 2D platforms (Di Tore et al. 2022). In contrast to the problems put forward by students regarding the level of concentration and the state of isolation, it is possible to realise a virtual or hybrid learning environment capable of fostering both peer collaboration and the inclusive paradigm (Haslam, Madsen, and Nielsen 2021)(Pace 2021).

### 5.2 Back to face-to-face Learning

Once the case of the collective video call has been ruled out, it is required to investigate the technically more challenging application case involving the face-to-face teaching mode that is still chosen by the majority of students. Unlike many other research-proposed aims, the use of a tool like video recording of lectures in the classroom might be beneficial in order to help with this observation process. Already from the work of Mangione and Rosa (2017), it appears from Tab. 4 that the main choices of subjects to be recorded in the classroom, are the students and the teacher-student interactions, regardless of the type of school grade. The videos collected are therefore considered by the teachers themselves, dense with information that is effective in updating their didactic action. The subject is further explored in the work of Todino and Sibilio (2019) up to the actual implementation of technological equipment for the classroom environment fit for purpose. The laboratory they suggest includes five high-definition cameras that can rotate 360°. Four of these cameras are attached to brackets, while one is mounted on special glasses worn by the teacher. With this camera setup, it becomes feasible to capture multiple perspectives—egocentric, allocentric, and heterocentric—of the teaching and learning activities occurring in the classroom. So, if we want to respect the type of video recording and framing suggested by the teachers themselves, one should prefer emotion detection methods such as that of Bisogni et al. (2023) that are also effective at greater distances and thus robust with respect to low resolution faces such as those that could be extracted from the overall classroom shot. These observations were considered in the work of Gao J. et al. (2023), where videos of several lectures held in the same university classroom are recorded. From the videos, image sequences are derived, and individual parts of the images containing the faces of the students themselves and their facial expressions are evaluated. In addition to the proposed application, it is interesting to observe how the dataset used to test the method was constructed. The emotions considered in this case are, in fact, neither a subset nor an extension of the set of basic emotions described by Ekman. In this specific case, the most consistent emotional states to be analysed in relation to a teaching action are neutral, focus, boredom, happiness, and confusion. Many works in the literature have already pointed out that certain facial expressions or variations in tone do not uniquely correspond to a certain emotion. In other words, a certain person may express a certain emotion differently from another. Moreover, some people tend to hold back or limit the expressive effects resulting from a given emotion. It must therefore be considered that these particular cases may occur with higher frequency in an environment such as the classroom where the student is inevitably under constant evaluation.

## 6. Use's Issues

### 6.1 Efficient use of the proposed tools

The proposed tools for supporting a more emotionally aware teaching action of students, due to the observations made, can be adapted to various teaching modes. However, once these tools are implemented, the scenario that unfolds may not justify the cost-benefit ratio. The teachers' knowledge of every single flow of students' emotions in real time might not be sustainable according to the simplex approach, which instead calls for a continuous ongoing adaptation of the didactic action (Sibilio 2012). Even a choice suggested by common sense and weighted on the 'average temperature' of the class may not be representative of the latter as it could be the case that a change in the mode of interaction or content or activity proposal may satisfy one part of the class and be ineffective on the remaining part. The realisation that this average may not be representative of the class justifies it once again and returns the responsibility for choosing the type of intervention to the teacher. In order to fully exploit the potential of these tools, it is therefore necessary for the data collected to be able to return information, not only during synchronous interaction. It therefore follows from this reflection that the adoption of these tools cannot disregard the seamless condition that releases the learning process from a specific place and time.

### 6.2 A Recursive Strategy for Seamless Learning

In addition to the issues concerning the choice of the data to be analysed, the AI method to be used, the set of emotions to be classified, and the configuration of the data collection devices, it is necessary to devise an educational framework that is not a form of adaptation to a new technology but rather a new application of a pedagogical theory that inherently promotes technology. As Trentin explained, the main issue is to understand how to exploit the potential of the new environments around us to foster increasingly active, interactive and seamless teaching-learning experiences. The realisation of hybrid learning environments cannot be achieved through the simple mechanical combination of real and virtual components, but rather by fostering a mutualistic relationship between them (Trentin 2019). In the same way as the work done during school hours needs to be matched with work at home, so the detection of a certain emotion during a lesson needs to be matched with a different activity or mode. Later on, by initiating a recording from a computer or smartphone, it will be feasible to identify emotional input during the stand-alone research once again. The teacher will next modify a new lesson and put an iterative adaptation method into practice using the data gathered in this second phase.

**Conclusions**

Through an examination of the main categories of technologies leveraging AI to capture data on human emotional states, efforts were made to pinpoint the crucial aspects of their deployment and utilization within educational settings. Following this analysis, various elements concerning the configuration of the learning environment came to light. These encompass both physical factors, such as selecting facial expressions as the data to be analyzed or determining the optimal camera framing within the classroom, and theoretical considerations, such as identifying the spectrum of emotions discernible during teaching activities or crafting activities conducive to seamless learning experiences. To thoroughly evaluate the cost-effectiveness of integrating tools like those mentioned above, it is imperative to conduct experimentation. This experimentation should strive to strike a balance between preserving students' preferred learning modalities while also facilitating a multifaceted utilization of the collected information. By engaging in such experimentation, educators can gain valuable insights into the efficacy and feasibility of incorporating AI-based emotional analysis tools into pedagogical practices. Ultimately, the successful integration of AI-based emotional analysis tools into educational environments hinges on a comprehensive understanding of both the technological capabilities and the pedagogical implications involved. By carefully navigating these complexities and engaging in thoughtful experimentation and reflection, educators can harness the potential of these tools to foster more engaging, adaptive, and effective teaching and learning experiences for all students.

## Author contributions

The article is the result of the scientific collaboration of the authors. However, the attribution of scientific responsibility is as follows: Umberto Bilotti is the author of the sections 1, 2, 3, 4, 5, 6; Michele Domenico Todino is the co-author of paragraph "Introduction" ; Maurizio Sibilio is the co-author of paragraph "Conclusions" and scientific coordinator. Development and testing was conducted at Teaching Learning Center for Education and Inclusive Technologies – Elisa Frauenfelder of the Department of Philosophical Humanities and Education (DISUFF) of the University of Salerno.

## References


Barron-Estrada, M. L., Zatarain-Cabada, R., & Bustillos, R. O. (2019). Emotion Recognition for Education using Sentiment Analysis. Res. Comput. Sci., 148(5), 71-80.

Bilotti, U., Bisogni, C., De Marsico, M., & Tramonte, S. (2024). Multimodal Emotion Recognition via Convolutional Neural Networks: Comparison of different strategies on two multimodal datasets. *Engineering Applications of Artificial Intelligence*, *130*, 107708.

Bilotti, U., Campitiello, L., Todino, M. D., & Sibilio, M. (2023). Emulation and understanding the emotion according to Generative Artificial Intelligence-Case study of emotional component extracted from visual artworks. Journal of Inclusive Methodology and Technology in Learning and Teaching, 3(4).

Bisogni, C., Cimmino, L., De Marsico, M., Hao, F., & Narducci, F. (2023). Emotion recognition at a distance: The robustness of machine learning based on hand-crafted facial features vs deep learning models. Image and Vision Computing, 136, 104724.

Bonaiuti, G., Calvani, A., & Picci, P. (2012). Tutorship e video annotazione: il punto di vista degli insegnanti. *Italian Journal of Educational Research*, 246-258.

Cahyani, D. E., Wibawa, A. P., Prasetya, D. D., Gumilar, L., Akhbar, F., & Triyulinar, E. R. (2022, September). Emotion Detection in Text Using Convolutional Neural Network. In 2022 International Conference on Electrical and Information Technology (IEIT) (pp. 372-376). IEEE.

Campitiello, L., Todino, M. D., Di Tore, P. A., & Di Tore, S. (2022). Emorobot: an open-source robot to promote the development of social skills in children with autism. ITALIAN JOURNAL OF HEALTH EDUCATION, SPORT AND INCLUSIVE DIDACTICS, 6(1).

Dhanasekar, V., Preethi, Y., Vishali, S., & IR, P. J. (2021, September). A chatbot to promote students mental health through emotion recognition. In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1412-1416). IEEE.

Di Tore, S., Campitiello, L., Todino, M. D., Iannaccone, A., & Sibilio, M. (2022). Education in the metaverse: amidst the virtual and reality. ITALIAN JOURNAL OF HEALTH EDUCATION, SPORT AND INCLUSIVE DIDACTICS, 6(3).

Ekman, P. (1999). Basic emotions. Handbook of cognition and emotion, 98(45-60), 16.

Gao, J., Wei, Y., Wang, K., & Shi, Y. (2023, April). Automatic Academic Emotion Recognition and Its Evolution Analysis in Classroom. In 2023 5th International Conference on Computer Science and Technologies in Education (CSTE) (pp. 42-48). IEEE.

Haslam, C. R., Madsen, S., & Nielsen, J. A. (2021). Problem-based learning during the COVID-19 pandemic: can project groups save the day?. Communications of the Association for Information Systems, 48(1), 21.

Lévêque, L., Villoteau, F., Sampaio, E. V., Perreira Da Silva, M., & Le Callet, P. (2022). Comparing the robustness of humans and deep neural networks on facial expression recognition. Electronics, 11(23), 4030.

Li, C., Xie, L., Shao, X., Pan, H., & Wang, Z. (2024). A multimodal shared network with a cross-modal distribution constraint for continuous emotion recognition. Engineering Applications of Artificial Intelligence, 133, 108413.

Ma, C., Sun, C., Song, D., Li, X., & Xu, H. (2018, August). A deep learning approach for online learning emotion recognition. In 2018 13th International Conference on Computer Science & Education (ICCSE) (pp. 1-5). IEEE.

Mangione G. R. and Rosa A. (2017). Professional vision and peer to peer in newly qualified teachers' path. the use of video for the analysis of the teacher's practice. Form@ re-Open Journal per la formazione in rete, 17(1):120–143, 2017

Noroozi, F., Corneanu, C. A., Kamińska, D., Sapiński, T., Escalera, S., & Anbarjafari, G. (2018). Survey on emotional body gesture recognition. IEEE transactions on affective computing, 12(2), 505-523.

Pace, E. M. (2021). Physically dispersed but virtually reunited: stories of inclusion during lockdown. ITALIAN JOURNAL OF HEALTH EDUCATION, SPORT AND INCLUSIVE DIDACTICS, 5(1).

Rivoltella, P. C. (2014). *La previsione. Neuroscienze, apprendimento, didattica*. La scuola.

Savia, G. (2016). Universal Design for Learning: La Progettazione Universale per l'Apprendimento per una didattica inclusiva. Edizioni Centro Studi Erickson.

Sibilio, M. (2012). La dimensione semplessa dell'agire didattico. Traiettorie non lineari nella ricerca. Nuovi scenari interdisciplinari.

Suganya, S., & Charles, E. Y. A. (2019, September). Speech emotion recognition using deep learning on audio recordings. In 2019 19th International Conference on Advances in ICT for Emerging Regions (ICTer) (Vol. 250, pp. 1-6). IEEE.

Todino, M. D., & Sibilio, M. (2019). Design and implementation of a new mobile video analysis laboratory at the University of Salerno. Research on Education and Media, 11(1), 4-8.

Trentin, G. (2019). Apprendimento senza soluzione di continuità negli spazi ibridi dell'infosfera. Professionalità Studi, 4(2), 8-25.

Xie, Z., Zhou, M., & Sun, H. (2021, October). A novel solution for EEG-based emotion recognition. In 2021 IEEE 21st International Conference on Communication Technology (ICCT) (pp. 1134-1138). IEEE.