

TECHNOLOGIES FOR EMOTION RECOGNITION USING ARTIFICIAL INTELLIGENCE

Xudoyorov Shaxriyor

Termez University of Economics and Service, Faculty of Economics and Information Technologies, Information Systems and Technologies Department, 3rd year student

Scientific supervisor: Ernazarov Mirzohid

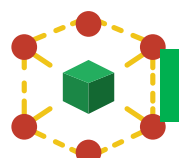
E-mail: xudoyorovshaxriyor57@gmail.com

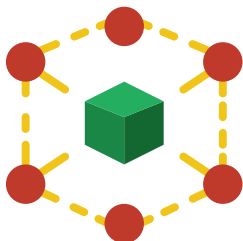
Annotation: This scientific article examines the capabilities of Artificial Intelligence (AI) systems in detecting and analyzing human emotions, namely modern technological advancements in the field of "Affective Computing." The article analyzes algorithms for classifying emotions through facial expressions, speech intonation, text content, and physiological indicators, with particular attention to the role of Deep Learning and Convolutional Neural Networks (CNN). The research thoroughly explores the practical significance of emotion recognition technologies in economics, medicine, education, and security, as well as the ethical issues and data privacy concerns that arise in this process. The article also compares the effectiveness of multimodal emotion recognition systems against traditional methods and provides scientific conclusions on future development trends. This work serves as a theoretical and practical resource for AI specialists and information technology students.

Keywords: Artificial Intelligence, Affective Computing, emotion recognition, neural networks, deep learning, facial expression analysis, speech processing, multimodal systems, machine learning, Natural Language Processing (NLP), biometrics, cognitive psychology, digital economy.

Introduction

Today, artificial intelligence technologies are rapidly penetrating every sphere of human life. As a result of advances in machine learning and neural networks, computers are achieving significant results not only in logical tasks but also in understanding the complex psychological states of human beings. This field, known as Affective Computing, is one of the most promising directions of artificial intelligence, enabling machines to sense, interpret, and respond appropriately to human emotions. The relevance of this technology lies in the fact that when people communicate with each other, they transmit a large part of information (approximately 70-93 percent) through non-verbal channels. This includes facial expressions, body language, voice tone, and eye contact. Therefore, in order to make Human-Computer Interaction (HCI) more natural and effective, AI systems must be able to read these signals.



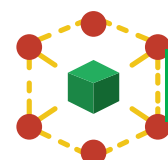


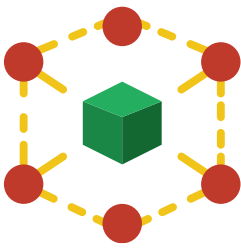
Emotion AI technologies are not only a technological novelty but also an important tool for solving socio-economic problems. For example, in medicine they help with the early detection of signs of depression or autism; in education they help to assess students' level of mastery of lessons through their emotions of interest and boredom; and in marketing they provide unparalleled help in analyzing consumers' true attitudes toward products. However, human emotions are extremely complex and subjective in nature. The same facial expression may mean different things in different cultures or contexts. Therefore, increasing the accuracy of AI in emotion recognition remains one of the most pressing scientific problems today. This article is devoted to a wide-ranging analysis of the technical foundations of emotion recognition, the algorithms used, and the existing obstacles in this field.

In the modern information society, the volume of data is growing in geometric progression. Social networks, video conferences, and online services have created a vast repository of data related to human emotions. Traditional programming methods are not sufficient to process this data. Models based on deep learning are required here, in particular architectures capable of analyzing large volumes of visual and audio data. The purpose of the article is to systematize the current state of emotion recognition technologies, to explain their working mechanisms, and to show the role of this direction in the development of information systems in the context of Uzbekistan. The research also scientifically justifies the advantages of biometric and multimodal approaches used in emotion recognition.

Scientific discussions on emotion recognition technologies are mainly conducted in two directions: psychological modeling and technical algorithmization. From the psychological point of view, researchers mainly rely on the "six basic emotions" model (joy, anger, fear, disgust, sadness, surprise) proposed by Paul Ekman. However, modern AI systems consider this model insufficient, because human emotions do not always fit into these six categories. Robert Plutchik's "wheel of emotions" or Russell's "valence-arousal" model are being discussed as more suitable for technical systems. This is because these models allow emotions to be viewed as a continuous spectrum, which is convenient for regression analysis in neural networks.

From a technical point of view, Convolutional Neural Networks (CNN) play a leading role in Facial Expression Recognition (FER). However, the main problem being discussed is the quality and diversity of datasets. Many popular datasets (such as FER2013, CK+) are obtained in laboratory conditions and do not take into account factors such as real-life lighting levels, the angle of the face, and obstructions (glasses, masks). This causes a model that has shown 95% accuracy in the laboratory to drop to 60-70% in real practice. Similarly, in Speech Emotion Recognition (SER), in addition to tone and tempo, there is a need to also consider the linguistic content of the speech. For example, the





word "Bravo!" can be used in both a praising and a sarcastic sense. In such cases, the AI system must understand not only the voice frequency but also the context.

Another important topic of discussion is the multimodal approach. Research shows that recognizing emotion through only one channel (for example, only the face or only the voice) increases the probability of error. The most reliable way to recognize emotions is to combine facial expression, speech, text, and even physiological signals (heart rate, skin resistance). However, synchronizing multimodal data and integrating it into a single decision-making model requires significant computational resources. Another issue at the center of discussion is ethics. Analyzing a person's emotional state without their permission can be considered a violation of human rights. Therefore, alongside the development of the technology, the need to develop its legal and moral norms is being actively discussed by scientists.

Main Part

Technological stages of emotion recognition.

The process of recognizing emotions using artificial intelligence usually consists of four main stages:

Data Acquisition: raw data is obtained through cameras, microphones, or sensors.

Pre-processing: noise removal, image normalization, face detection, and centering.

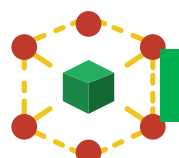
Feature Extraction: extracting key facial points (landmarks), voice frequencies (MFCC), and key words from text.

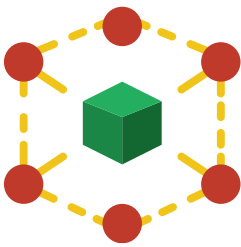
Classification: using machine learning algorithms (SVM, Random Forest, Deep Learning) to assign the extracted features to a specific emotion category.

Analyzing emotions through facial expressions (FER).

The human face is the richest source of emotions. Modern FER systems analyze small muscle movements called "Action Units." For example, raising the eyebrows and opening the mouth indicates the emotion of surprise. CNN (Convolutional Neural Networks) models automatically detect these action units by studying the relationships between pixels in an image. Deep architectures such as VGG-16, ResNet, and Inception can detect even the most subtle changes in the face. However, there is a problem of "cultural characteristics" in face recognition. For example, in some Eastern cultures, people tend to hide their emotions with a more serious facial expression, which can lead the AI to draw incorrect conclusions.

Speech and voice analysis (Speech Emotion Recognition).





In recognizing emotions through speech, the acoustic features of the voice play an important role. These include the fundamental frequency (pitch), intensity (volume), speech rate, and spectral features. Typically, in anger the volume and frequency of the voice rise sharply, while in sadness the speech slows down and becomes monotonous. In modern systems, Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks are used, since they are specialized for analyzing audio signals that change over time. In recent years, the "Transformers" architecture has shown revolutionary results in converting speech to text and simultaneously determining its emotional coloring.

Natural Language Processing and text-based sentiment analysis.

Text analysis is very important in social media comments, e-mails, and customer communications. Here, the AI must understand not only the words in the dictionary but also the context of the sentence. For example, in the sentence "This phone is so bad that I wouldn't even wish it on my enemy," the word "bad" expresses a negative emotion. However, in the sentence "This concert was so dreadful, I'm still in awe," the word "dreadful" may be used in a positive sense. Models such as BERT (Bidirectional Encoder Representations from Transformers) solve such complexities by learning the bidirectional context of the sentence.

Multimodal and physiological approaches.

The highest peak of emotion recognition is multimodal systems. A person may lie or control their facial expression, but it is difficult to deceive their physiological indicators (heart rate, blood pressure, electrical conductivity of the skin). Data obtained through wearable devices helps AI to determine the true level of a person's stress or excitement. For example, in determining the level of fatigue of drivers, both facial expression (closing of the eyes) and the strength of gripping the steering wheel (physiological pressure) are analyzed together.

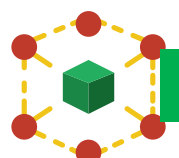
Areas of practical application: Economics and Marketing: Companies analyze the facial expressions of customers watching advertising clips to determine which frame generated more interest. This is called "Neuromarketing."

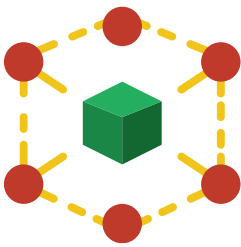
Healthcare: In diagnosing mental illnesses, AI monitors changes in the patient's voice and face over time and warns the doctor of the risk of depression.

Education: Smart systems in distance learning that detect when a student is not understanding a lesson (through boredom or confusion) and recommend simplifying the material.

Security: Identifying suspicious or overly agitated individuals at airports or mass events.

Problems and limitations.





No matter how advanced the technology, there is a big difference between "understanding emotions" and "recognizing emotions." AI can recognize the biological expression of emotion, but it does not understand its underlying cause and meaning. Likewise, the problem of "bias" in datasets remains relevant. If a neural network has been trained only on images of Europeans, it may make significant errors in recognizing the emotions of Asians or Africans.

Conclusion

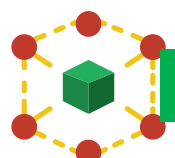
Emotion recognition technologies based on artificial intelligence are taking human-machine communication to a new level. Today, these systems have come a long way from simple facial expression recognition to complex multimodal analyses. Advances in affective computing are creating unprecedented opportunities for increasing economic efficiency, improving the quality of social services, and protecting human health. However, the wide-scale implementation of this technology depends not only on technical perfection but also on how correctly the ethical and legal norms are formed.

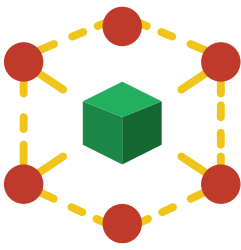
Research shows that in the future AI systems will not only recognize a person's current emotion but also have the ability to predict their emotional state. This in turn requires a deeper integration of cognitive psychology and information technology. In Uzbekistan, too, when implementing digital economy and AI strategies, it is relevant to use emotion recognition technologies and create models adapted to local conditions (taking into account the phonetic and cultural features of the Uzbek language).

In conclusion, artificial intelligence that can recognize emotions is not just an interesting technology, but a powerful tool that helps humanity better understand itself and create a safer, more convenient, and more effective digital environment. In the future, by ensuring algorithm transparency and data integrity, this technology can be usefully applied to all spheres of society.

References

1. Picard, R. W. (1997). *Affective Computing*. MIT Press.
2. Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Times Books.
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
4. Pantic, M., & Rothkrantz, L. J. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
5. Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.





6. Sariyanidi, E., Gunes, H., & Cavallaro, A. (2015). Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
7. Schuller, B. W. (2018). Speech Emotion Recognition: Two Decades in a Nutshell, Benchmarks, and Ongoing Trends. *Communications of the ACM*.
8. Devlin, J., et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint*.
9. Zadeh, A., et al. (2017). Tensor Fusion Network for Multimodal Sentiment Analysis. *EMNLP*.
10. Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*.
11. Resolution PQ-4996 of the President of the Republic of Uzbekistan dated February 17, 2021 "On measures to create conditions for accelerating the introduction of artificial intelligence technologies".
12. Vinciarelli, A., et al. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*.
13. Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*.
14. Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*.
15. Tao, J., & Tan, T. (2005). Affective computing: A review. *International Conference on Affective Computing and Intelligent Interaction*.
16. Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*.
17. Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*.
18. Grave, E., et al. (2018). Learning Word Vectors for 157 Languages. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

