

Survey on Emotion Recognition System

S. Sowmiya¹ and J. C. Miraclin Joyce Pamila²

¹PG Scholar, Dept. of CSE, Government College of Technology, Coimbatore – 641013, India

²Professor, Dept. of CSE, Government College of Technology, Coimbatore – 641013, India

E-mail: ¹sowmiya18798@gmail.com, ²miraclin@gct.ac.in

Abstract

Humans are good at making millions of facial movements during conversation. In interpersonal relationships, the human emotion sensing system is vital. At an older age, automatic emotion recognition is a trendy research problem. Emotions are expressed through personas, hand and body gestures, and facial expressions. Emotion recognition via facial expressions is one of the most important fields in the human-machine interface. The strategy of recognizing emotions from facial expressions is known as facial expression analysis. Emotions are automatically perceived by the human brain, and software that can recognise emotions has recently been developed. This technology is constantly improving, and it will ultimately be able to sense emotions as accurately as human brains. The purpose of this work is to present a survey of emotion detection research using various machine learning techniques. It also summarises the benefits, drawbacks, and limitations of current approaches, as well as the concept's evolution in research.

Keywords: Face detection, Emotion recognition, feature extraction, Convolution neural network, Deep learning

1. Introduction

In recent years, facial expression detection systems have advanced substantially. Machine learning, image processing, and human cognition have all advanced as a result of advancements in adjacent arenas. FER seems to be more popular in a variety of applications, including Human-Computer Interaction, Robot control, and driver status surveillance. Emotions are fluctuating phenomena that are elicited by a driving force, making emotion identification a difficult problem in real-world applications. The complexity of the human face, which encompasses colour, orientation, expression, posture, texture, and so on, necessitating the establishment of a facial expression detection system. Recently, the concept of emotion recognition has gained a lot of traction among researchers in the field of smart

systems and Human-Computer Interaction. The facial expression can be categorized into six different basic emotions based on the facial characteristics.

Table 1. Emotions and expressions

Emotion	Facial expression
Anger	Eyebrows lowered and burrowed, focused stare, elevated chin
Joy	A Smile is formed by raising the corner of the mouth.
Surprise	Big eyes, arched brows, and a drooping jaw.
Fear	Scrunched brows, gaping mouth, and bright blue eyes
Sadness	Lip corner muscle relaxer, glazed eyes
Anxiety	Biting of the lips.

The table above depicts the feelings and expressions. Recognition system is vital in achieving long-term human-computer collaboration. Speech, facial expressions, gestures, and thoughts are all aspects of how people convey their feelings. Emotions can be 's absorptive a range of data sources, such as speech, text, and pictures. Emotion analysis is beneficial in a multitude of fields. Judging human emotions is a tricky proposition to their diversity and ambiguity. Different emotions can have the same expression at the same time, and some emotions can be portrayed in a variety of ways. Emotions vary depending on a person's personality, gender, geographic location, ethnicity, culture, and situation, as well as a range of other physiological, social, and individual aspects.



Figure 1. Glimpse of few emotions

1.1 Applications of emotion recognition

Different firms are already using emotion recognition to evaluate customer sentiment toward their products, brands, marketing activities, employees, or in-person interactions. Understanding client emotions is vital for business growth and fostering customer experiences, but the benefits of this technology extend beyond market research and digital advertising.

1.1.1 Recognition of emotions in online applications and interviews:

Emotion recognition can be used to gauge how candidates are feeling during interviews and how they respond to certain topics. This data can be utilised to improve the organisation of future candidates' interviews and streamline the application process. Head orientation/pose analysis can also be used to gauge attentiveness with Sightcorps technology.

1.1.2 Emotional analysis in online learning:

Anonymous emotion detection for online education is a great tool to evaluate and improve the online student journey. Emotional feedback is used to evaluate a school's course materials, teaching techniques, organisation, and layout as students progress through each module in real time. To locate cavities of interest or course hurdles, uses real facial responses and engagement levels to create tweaks.

1.1.3 Emotion recognition in health care:

Health care is one field that is benefiting from this technology, with AI-powered recognition software assisting in determining when patients require medication, assessing their emotional response in clinical trials, and assisting clinicians in determining how to effectively triage their patients.

1.1.4 Emotion analysis in video game testing:

Video games are created with a certain target audience in mind, with the intent of eliciting a specific set of behaviours and feelings from the players. Participants are requested to play the game for a set time period during the testing process, and their feedback is used to optimize the final product. Facial expression detection can let you grasp out what feelings a user is having in real time. This is a terrific complement to verbal feedback because it gives a more in-depth look at the gaming experience.

1.1.5 Automotive industry and emotion recognition:

Emotion recognition technology is also being used in the automotive sector, as car manufacturers all over the world strive to make cars more personal and safe to drive. It's only natural that vendors use AI to assist them comprehend human emotions as part of their quest to develop smart car features. Smart automobiles that use facial emotion detection can inform the driver when he is feeling drowsy, depleting road fatalities.

2. Literature Survey

Rajesh Kumar G A et al., [11] proposed Facial Emotion Analysis using Deep CNN. They emphasized that character traits are mental states of sentiments that arise spontaneously and are accompanied by physiological changes in facial muscles that result in facial expressions. Happiness, sadness, rage, disgust, fear, surprise, and other critical emotions are examples. Facial expressions are important in nonverbal communication because they reveal a person's inner feelings. There has been a lot of research into computerized human emotion. It is, nonetheless, a significant achievement for the human visual system. They designed a new deep Convolution Neural Network (CNN) to give a better technique for anticipating human emotions (Frames by Frames) as well as how emotion intensity varies on a face in this system from low to high degrees of emotion. In their algorithm, they utilize the FEREC-2013 database. The experiment's results are reasonable, and the level of accuracy achieved may encourage researchers to design future models of computer-based emotion recognition systems.

The main idea of the present approach is to determine the standardized parentages of many different emotional states in a face (happiness, sadness, disgust, anger, surprise, and fear). As the resultant emotion, the feeling with the most parentages is placed on a certain face. The initial random weights, the activation function used, the training data, the number of hidden neurons, and the system's network architecture all have a role in determining the performance of a neural network. Convolutional neural networks are fed images directly. Convolutional neural networks are utilised to learn a feature pecking order that may subsequently be used for pneumatic hierarchy.

Wei-feng Liu et al., [12] developed Automatic Viola Jones based on Local Binary Patterns of Local Areas. According to them, automatic Viola Jones is one of the most serious issues in the field of Viola Jones. They create an autonomous Viola Jones technique based on Local Binary Patterns of Local Areas in their research (LLBP). To begin, the eyeballs'

position is fixed using the projection method. The 5 prior knowledge of face structure can then be used to determine the general vicinity of the eyes and mouth. The LBP feature on the neighbouring sites is then used to compute the facial feature for Viola Jones. Finally, they were using the JAFFE facial database to undergo a recognition test, demonstrating the efficacy of the proposed technique. The following are their thoughts on the suggested model: Convolutional Neural Networks have made advancements in a variety of pattern recognition fields during the last decade, from image processing to speech recognition. The major benefit of CNNs is that they reduce the number of parameters in ANNs.

This accomplishment has prompted academics and developers to examine larger models for demanding tasks that were previously impossible to handle with regular ANNs; The most crucial consequence of CNN-solved challenges is that spatially dependent features should not be included. To put it another way, users of facial recognition software don't have to be concerned about the location of the faces in the picture. Notwithstanding the where they exist in the images, the only important factor is that they are identified. CNN's capacity to extract semantic information when input propagates deeper levels is another essential feature. For example, in image classification, the edge may be recognized in the first layers, then lesser features in the second layers, and lastly higher-level characteristics such as faces in the tertiary layers. They enhanced their dexterity..

Zhang et al., [13] stated that facial expression is vital for human communication and applied in real-time application. Extracting properties is a crucial step in visual feature detection. They developed a method for solving the feature extraction challenge by picking salient features that demonstrated a high rate of correct recognition. In the JAFFE database, the suggested system achieved a higher CRR, and it ranks among the best performers in the CK database (Cohn-kanade2), which includes pre-processing, training, and testing stages. They use the nose as a focal point, with the face regions manually trimmed.

There are eight scales and four orientation Gabor filters in use. All sets of patches are extracted in the training stages by shifting patched of series or Gabor image. Gabor filters, which are uniquely developed from statistical information of character structures, extract features from greyscale character images. They used the JAFFE database, which consists of 213 grayscale images of seven different facial expressions (6 basic and 1 neutral). Their FER system achieved a 93 success rate; however, it is not suitable for video-based FER systems.

Sailunaz et al., [14] proposed Emotion detection from text and speech. Their survey examines emotion detection research efforts, including emotion models, emotion datasets, emotion detection algorithms, their features, drawbacks, and some potential future prospects. They concentrated on studying research findings and interpreting emotions through text and speech. They looked into several feature sets that have been employed in previous approaches. They outlined key accomplishments in this field and suggested possible extensions for a better end. They discovered that several emotion models can be used to categorize emotions. An appreciating emotion model must be chosen to identify and evaluate emotion from text from any source. It should specify the range of emotions that are appropriate for a given situation. They looked into the evolution of emotion. This study gives a detailed portrayal of emotion detection and the tools used to do it.

Yamazoe et al., [15] proposed a Body Mounted camera system. They envisioned a body-mounted system that would record the user's experience as audio/visual data. Two cameras and a microphone are included in their proposed methodology (head detection and wide-angle). The user's head motion is captured by the head detection, while the user's frontal view photos are captured by the wide-angle color. Their solution addresses the disadvantages of head-mounted cameras in terms of ease of recording audio data, images in the user's field of view, and head gestures such as nodding and shaking to check the effectiveness of the proposed system.

They first used their approach to quantify or evaluate the accuracy of feature extraction by extracting the rotation center and nose position. They employed a Random Sample Consensus method (RANSAC). They designed a wearable device for capturing audio and visual information related to user experiences by using the proposed system. Audible information, emotions, and visuals in the user's field of view, or readily captured in sequences. The system's downside is that the prototype's size and weight are still insufficient, and the system's very simplistic structure means that it lacks global positional capability.

Joseph et al., [16] demonstrated facial emotion detection using modified eye map–mouth map algorithm on an enhanced image and classification with TensorFlow. They stated that Emotion detection by facial expression is a burgeoning subject of research. When a guy interacts with a computer, facial expression detection is also useful for identifying his behavior. They proposed The recognition of facial expressions in relation to changes in facial geometry. First, a discrete wavelet transform and fuzzy combination are used to improve the image. After discovering the landmarks, the modified eye map and mouth map

technique is used to find the facial geometry. Finally, using a neural network and the TensorFlow central processing unit version, the size and angle of the created triangles are determined and categorized. The following are some of the approaches to the proposed work:

1. Image improvement with DWT and fuzzy, which is utilized to improve a dull image.
2. Modified mouth map for mouth detection, which, when compared to state-of-the-art approaches, is a more efficient way of recognizing a person's mouth. Eye geometry is constructed using a modified eye map that is used once the eye is detected using the Viola-Jones technique.
3. Tensorflow is used to classify emotions and validate the categorized findings.
4. They used three distinct databases to evaluate the proposed system: Karolinska Directed Emotional Faces (KDEF), Oulu-CASIA NIR-VIS facial expression (Oulu-CASIA), and expanded Cohn–Kanade (CK+).

Other emotions can be identified using the geometry of the face in relation to the neutral mood. The observed geometry shows the feasibility of the proposed improved eye map–mouth map technique. The variances in the geometry vary according to a person's emotion. The use of several datasets demonstrates that the technique may be used to detect geometry in other databases as well. The algorithm's settings were fine-tuned to get the best results for detecting emotion. Their findings suggest that the proposed algorithm is gender agnostic and may be applied to any gender.

Vimaleswaran et al., [17] proposed E – Therapy Improvement Monitoring Platform for Depression using Facial Emotion Detection of Youth. They devised a depression prediction model that incorporated a feed-forward neural network model with the Depression Scale to determine the severity of depression. Their venture is to calculate the improvement level based on observed depression levels, then visualize it in a center console for the therapist and patient to track emotional and behavioral problems. Their primary goal is to provide a viable treatment alternative for teenagers who are depressed. Clients and therapists will be able to participate in an online intervention period and track the client's level of depression progress and innovation using a smartphone app that is recommended as part of the solution.

The focus of this thesis is to predict depression using the recorded video from the therapy session, calculate remission levels, and conclude with a statistical representation once the therapy session is finished. Initiation, first, second, third, and finally decision are the five

steps of the proposed solution. The patient's depression level is predicted using a model built by Their nonintrusive multilayer neural network-based architecture for predicting Depression levels by observing facial expressions with the Facial Action Coding System. The algorithm continues on to the decision step after forecasting the depression level, when the improvement level is decided. Finally, a statistical representation of the current session's depression level and/or improvement level will be provided for each session. Psychotherapists and teenagers who need to endure depression therapy can use the provided solution e-therapy improvement monitoring. This solution will assist therapists in understanding the state of their patients as well as making decisions about the therapy they deliver. Patients can also track their progress over time by attending each session, which aids them in determining if their condition has improved.

Lakshmi D et al., [18] proposed Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders. They used modified Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) feature descriptors to propose a unique feature descriptor for facial expression recognition in this paper. To locate the eye, nose, and mouth region using Viola-Jones technique, first use viola-Jones face detection to detect the face region. Then use Butterworth high pass filter to enhance the found region. Second, the features of the identified eye, nose, and mouth regions are extracted using the proposed modified HOG and LBP feature descriptors.

Deep Stacked AutoEncoders is used to concatenate and reduce the dimensionality of the retrieved features from these three sections. Finally, for classification and recognition, a multi-class Support Vector Machine is used. The proposed improved feature descriptors can effectively distinguish emotions on the basis of experimental results. Feature extraction extracted spatial and texture information utilizing modified HOG and LBP feature descriptors in their research. These features are combined, and Deep Stacked Autoencoders (DSAE) are used to circumvent the curse of dimensionality restriction. Finally, the human facial expressions are classified using a multi-class Support Vector Machine (SVM) classifier. The appearance of the face regions with respect to the neutral expression can also be used to recognize other human expressions and emotions, according to this work.

Rani and Pooja [19] proposed Emotion Detection of Autistic Children using Image Processing. They tried with facial expressions to discern autistic children's emotional states. Their research is focused on four distinct emotions. The four emotions are sad, happy, neutral, and furious. To detect autistic children's emotions, image processing and machine

learning approaches are applied. The traits are extracted from the faces of autistic toddlers using a local binary pattern. Machine learning techniques are used to classify emotions. In the classification process, support vector machines and neural networks are two machine learning classifiers that are used. To extract features from autistic photographs, they used texture feature extraction of photos utilising local binary patterns. They employed ANN to classify the data. Artificial Neural Network Implementation as a Classifier With 256 neurons in the input layer, 20 neurons in the hidden layer, and 2 neurons in the output layer, with sim as an activation function, a two-layer feed forward network (multi-layer perception) is built. The Levenberg Marquart nonlinear optimization algorithm was used to train the network. With a combination of steepest descent and the gauss Newton Method, it is one of the quickest back propagation techniques. The technique operates like a steepest descent method when the current solution is distant from the correct one: sluggish but certain to converge. The experiment achieved different performances, and the overall accuracy was 90% .

Dagar et al ., [20] developed an Automatic emotion detection model from facial expressions. They used Gabor feature extraction and neural networks to suggest a frame from live streaming in their framework. Finally, they determined face expressions independently, then channeled the processed feature vector through the previously acquired pattern classifiers. The first goal of their research is to use anatomical grasp to recognize emotions. The facial action coding system is used to depict face behavior (FACS).

From an anatomical standpoint, FACS ties appearance changes with muscle function. Initially, they took live streaming footage and extracted frames from it, attempting to detect those frames in which they were confronted. They used the Gabor approach to learn about neural networks and coupled it with other methods. They utilized MLP . during the test phase they have picked any random image from the dataset and use pattern classification to discover the particular mood of the image.

3. Conclusion

In this paper, literature review for the emotion detection based on facial features by using various machine learning, deep learning classification algorithms has been presented. From this study, a clarity that convolutional neural network gives the better results for emotion detection based on facial features has been obtained. In future work, these methods would be proposed for emotion detection and classification, and also to suggest music for corresponding emotion.

References

- [1] Verma, Rohit, and Mohamed-Yahia Dabbagh. "Fast facial expression recognition based on local binary patterns." 2013 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, 2013.
- [2] N. Sebe, M. S. Lew, Y. Sun, I. Cohen, T. Gevers, and T. S. Huang. Authentic facial expression analysis. *Image and Vision Computing*, 25(12):1856–1863, 2007
- [3] B. Hasani, M. M. Arzani, M. Fathy, and K. Raahemifar. Facial expression recognition with discriminatory graphical models. In 2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS), pp. 1– 7, Dec 2016
- [4] B. Hasani and M. H. Mahoor. Spatio-temporal facial expression recognition using convolutional neural networks and conditional random fields. *arXiv preprint arXiv:1703.06995*, 2017
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012
- [6] Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.
- [7] Mollahosseini, B. Hasani, M. J. Salvador, H. Abdollahi, D. Chan, and M. H. Mahoor. Facial expression recognition from world wild web. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016
- [8] Graves, A. r. Mohamed, and G. Hinton, Speech recognition with deep recurrent neural networks. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6645-6649, 2013
- [9] Sanin, C. Sanderson, M. T. Harandi, and B. C. Lovell, Spatiotemporal covariance descriptors for action and gesture recognition. In *IEEE Workshop on Applications of Computer Vision*, 2013
- [10] S. Jain, C. Hu, and J. K. Aggarwal, Facial expression recognition with temporal modeling of shapes. In *Proc. IEEE International Conference on Computer Vision Workshops*, pp. 1642-1649, 2011.
- [11] Kumar, GA Rajesh, Ravi Kant Kumar, and Goutam Sanyal. "Facial emotion analysis using deep convolution neural network." 2017 International Conference on Signal Processing and Communication (ICSPC). IEEE, 2017.

- [12] Liu, Wei-feng, Shu-juan Li, and Yan-jiang Wang. "Automatic facial expression recognition based on local binary patterns of local areas." 2009 WASE International Conference on Information Engineering. Vol. 1. IEEE, 2009.
- [13] T. Zhang, W. Zheng, Z. Cui, Y. Zong and Y. Li, SpatialTemporal Recurrent Neural Network for Emotion Recognition, IEEE Transactions on Cybernetics, arXiv:1705.04515, Issue: 99, pp. 1-9, 2018
- [14] Sailunaz, Kashfia, et al. "Emotion detection from text and speech: a survey." Social Network Analysis and Mining 8.1 (2018): 1-26.
- [15] Yamazoe, Hirotake, et al. "A body-mounted camera system for head-pose estimation and user-view image synthesis." Image and Vision Computing 25.12 (2007): 1848-1855.
- [16] Joseph, Allen, and P. Geetha. "Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow." The Visual Computer 36.3 (2020): 529-539.
- [17] Vimalaswaran, Brindahini, and Gayashini Shyanka Ratnayake. "E-Therapy Improvement Monitoring Platform for Depression using Facial Emotion Detection of Youth." 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, 2021.
- [18] Lakshmi, D., and R. Ponnusamy. "Facial emotion recognition using modified HOG and LBP features with deep stacked autoencoders." Microprocessors and Microsystems 82 (2021): 103834.
- [19] Rani, Pooja. "Emotion detection of autistic children using image processing." 2019 Fifth International Conference on Image Information Processing (ICIIP). IEEE, 2019.
- [20] Dagar, Debishree, et al. "Automatic emotion detection model from facial expression." 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT). IEEE, 2016.
- [21] F. Visin, K. Kastner, K. Cho, M. Matteucci, et al., Renet: A recurrent neural network based alternative to convolutional networks. arXiv preprint arXiv:1505.00393, 2015
- [22] F. Visin, K. Kastner, A. Courville, Y. Bengio, et al., ReSeg: A Recurrent Neural Network for Object Segmentation. arXiv preprint arXiv:1511.07053, 2015
- [23] P. Khorrami , T. L. Paine , K. Brady, C. Dagli , T. S. Huang, How Deep Neural Networks can Improve Emotion Recognition on Video Data, IEEE Conference on Image Processing (ICIP) , 2016

Author's biography

S. Sowmiya received the B.E degree in Computer Science and Engineering from the EASA college of engineering and technology, Coimbatore. The M.E degree in Computer Science and Engineering from Government College of Technology, Coimbatore. Affective computing, machine learning, and image processing are among her research interests.

J. C. Miraclin Joyce Pamila is Professor & Head of the Department of Department of Computer Science and Engineering, Government College of Technology, Coimbatore. She teaches and guides students at both under graduate and Post graduate levels. She is a life member of ISTE and currently takes up research in the area of Machine Learning, Data Analytics and Natural Language Processing.