IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

Enhanced Real-Time Facial Emotion Recognition System Using Google Net

Piyali Hemant Ingale[1], Arpita Angad Lahane[2], Shradha Santosh Tarade[3], Prof. D. A. Gore[4]
Students, Department of Computer Engineering[1,2,3]
Guide, Department of Computer Engineering[4]
Navsahyadri Education Society's Group of Institutions Pune, Maharashtra, India
piyaliingale5555@gmail.com

Abstract: *Facial expressions are vital for conveying non-verbal information in human interactions, and automated facial expression recognition has gained significant interest in human-machine interfaces since the early 1990s. Traditional machine learning methods for expression recognition often rely on complex feature extraction techniques but tend to produce limited results. In this paper, we utilize deep learning advancements to introduce a Google Net-based architecture for automatic facial expression recognition. This approach eliminates the need for manual feature engineering by using Google Net's strong feature extraction capabilities to interpret the semantic information present in facial expressions.*

Keywords: Facial Expressions, Automated Recognition, Deep Learning, Google Net, Convolutional Neural Network(CNN)

I. INTRODUCTION

Recognizing emotions is a crucial aspect of developing socially aware systems, with applications across fields such as healthcare (e.g., mood profiling), education (e.g., personalized tutoring), and security (e.g., surveillance). Speech Emotion Recognition (SER) holds significant potential due to the widespread use of speech-based devices, yet SER models need to generalize effectively across diverse environments to maintain reliable performance.

Traditionally, emotion recognition systems are trained using supervised learning approaches. High-performing models in tasks like computer vision rely on large datasets with thousands of labeled samples, while speech recognition systems (ASR) require hundreds of hours of annotated audio data. In emotion recognition tasks, labels are typically obtained from sensory evaluations conducted by multiple raters, who annotate samples after listening or observing stimuli. However, this data collection process is cognitively demanding, costly, and often limited to small, curated datasets with a restricted number of emotional labels, impacting model generalization.

Currently, most facial expression recognition studies rely on the Facial Action Coding System (FACS), which maps facial muscles to expression spaces. The main objective of this system is to classify human facial movements based on visual appearance. However, this mapping can face certain challenges. For example, the facial gestures associated with expressions can sometimes be deliberately altered, as actors might display expressions they do not actually feel. In one experiment, a patient who is partially paralyzed on one side was asked to smile. When prompted, only one side of the mouth moved, but when the patient heard a joke, both sides of the mouth rose naturally, indicating genuine expression.

This demonstrates that different pathways can communicate an expression, depending on its origin and nature. For computers, numerous possibilities are emerging to enhance their ability to express and recognize expressions. It is now possible to mimic facial action units in digital systems, enabling computers to present graphical faces that foster more natural interactions. In terms of recognition, computers have become capable of identifying basic facial expression categories, including happiness, surprise, anger, and disgust.

Emotion recognition is essential for developing socially aware systems and has impactful applications in fields like healthcare (e.g., mood tracking), education (e.g., personalized tutoring), and security (e.g., surveillance). Speech Emotion Recognition (SER) has significant potential due to the prevalence of speech-enabled devices, but SER models must perform reliably across diverse conditions to ensure robust outcomes.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

II. BACKGROUND AND LITERATURE REVIEW

Ge, Huilin, et al. [1] conducted a study titled "Facial expression recognition based on deep learning," which examines the application of deep learning techniques to enhance facial expression recognition accuracy. The authors explore various deep learning architectures and propose a model that improves upon existing methods by achieving higher recognition rates across diverse datasets. Their work emphasizes the importance of feature extraction in expression recognition, demonstrating that deep learning models can autonomously learn and identify complex expression patterns, which can be crucial for applications in emotion driven human-computer interactions and healthcare.

Bisogni, Carmen, et al. [2] in their paper "Impact of deep learning approaches on facial expression recognition in healthcare industries," investigate how deep learning-based facial expression recognition (FER) systems can be effectively used in healthcare environments. The study assesses multiple deep learning techniques to determine which architectures provide the most accurate and reliable results for healthcare applications, especially in tracking patient emotions and responses. The findings highlight the potential of deep learning models to offer healthcare professionals on model generalizability valuable insights into patient well-being and engagement, suggesting significant implications for improving patient monitoring and mental health assessment.

Umer, Saiyed, et al. [3] in "Facial expression recognition with trade-offs between data augmentation and deep learning features" analyze the balance between data augmentation techniques and feature extraction in enhancing FER systems. The authors focus on the advantages of data augmentation for increasing the diversity of training data and its influence. By evaluating the impact of different augmentation strategies, they conclude that a combination of data augmentation and deep learning feature extraction results in improved recognition accuracy and robustness, especially when dealing with limited datasets.

III. PROBLEM STATEMENT

Our goal is to develop a robust and automated face detection system that analyzes facial expressions and extracts meaningful insights.

This involves creating datasets for model training, designing optimized classifiers, and learning facial descriptors.

We propose a model capable of recognizing six universal facial expressions—anger, happiness, sadness, surprise, disgust, and fear—across cultures.

The system will detect faces, identify unique characteristics, and make weighted predictions of an individual's expression.

IV. PROPOSED SYSTEM

The proposed system for automated facial expression recognition focuses on accurately classifying human expressions using video input. As shown in the diagram, the system begins by processing an input video, from which training samples are generated. These samples undergo a preprocessing phase that prepares the data for analysis, typically involving steps like normalization and resizing. After preprocessing, feature extraction is performed to identify distinct facial characteristics relevant to expressions. For testing, new samples are processed through the feature extraction module and fed into a classification model, which categorizes the expressions into seven primary categories: Happy, Neutral, Angry, Disgust, Fear, Sad, and Surprise. This structured methodology aims to enhance the accuracy and efficiency of real-time expression recognition, which can be applied in various domains such as psychology, healthcare, and human computer interaction.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

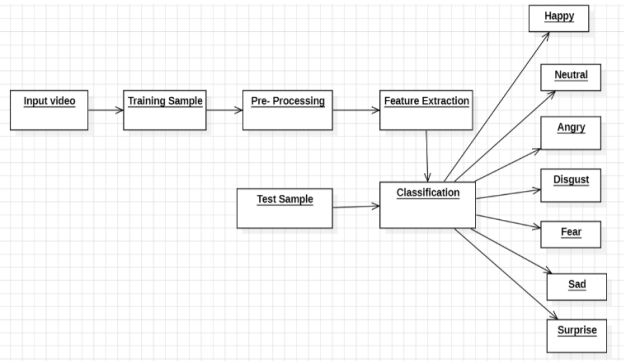

Fig. System Architecture

V. METHODOLOGY

The methodology for expression recognition from video input involves a systematic pipeline that uses Google Net for training and evaluating the model's ability to identify distinct expressions. The process starts with an Input Video, which serves as the primary data source and is divided into two subsets: Training Samples and Test Samples. The Training Samples train the model, while the Test Samples are used to evaluate its performance on unseen data.

After selecting Training Samples, they undergo a Pre-Processing phase. Pre-processing enhances data quality and standardizes it for consistency across all samples. This includes resizing frames to match Google Net's input requirements, normalizing pixel values, and removing noise. Additional steps may involve adjusting lighting, aligning faces to ensure consistent feature positioning, and correcting for any frame misalignments. These steps are essential to minimize non-contributive variations, allowing Google Net to focus on relevant aspects of facial expressions.

Next, the pre-processed data moves to the Feature Extraction phase. Google Net's deep layers are leveraged to extract key features that differentiate various expressions. This involves analyzing facial landmarks (like eyes, mouth, and eyebrows) and tracking changes in these landmarks over time to capture expression-related patterns. The resulting features create a condensed representation of the video frames, preserving only the critical details for expression classification. The extracted features are then input into the Classification model, which is based on Google Net's architecture. During training, the model learns to associate specific feature patterns with expressions, adjusting its layers and weights to improve classification accuracy. Google Net's deep learning capabilities enable it to identify complex patterns in the data, enhancing its ability to detect subtle differences across expressions.

Finally, the model is evaluated using the Test Samples, which undergo the same pre-processing and feature extraction steps. Google Net classifies each sample into one of the following expression categories: Happy, Neutral, Angry, Disgust, Fear, Sad, or Surprise. The classification outcomes are then used to assess the model's accuracy and generalization performance, ensuring robust expression recognition across diverse faces and conditions.

Algorithm Google Net

The Google Net architecture, also known as Inception V1, introduced in 2014 by Google researchers in collaboration with several universities, is detailed in the influential paper "Going Deeper with Convolutions." This architecture distinguished itself as the winning model in the ILSVRC 2014 image classification challenge by achieving a substantial reduction in error rates compared to its predecessors, including Alex Net (ILSVRC 2012 winner) and ZF-Net (ILSVRC 2013 winner), as well as outperforming VGG, the 2014 runner-up. The key innovations in Google Net include 1×1 convolutions and global average pooling, which collectively enhance its performance and efficiency in image classification tasks.

Google Net's architecture represents a significant evolution from earlier models by implementing techniques like 1×1 convolutions and global average pooling, which enable the development of deeper and more sophisticated networks. A defining feature of the Inception architecture is its use of 1×1 convolutions to reduce the number of parameters,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

effectively lowering computational demands while allowing the network to become deeper and more capable of recognizing intricate data patterns.

For Automated Facial Expression Recognition, these architectural advancements are crucial in accurately identifying and classifying expressions. By capturing subtle variations in facial expressions, the model can reliably distinguish between expressions such as happiness, sadness, anger, and surprise. This enhanced expression recognition supports various applications in fields such as human computer interaction, security, and behavioral analysis, where understanding nuanced facial expressions is essential.

Auxiliary Classifier for Training:

The Google Net architecture includes auxiliary classifier branches used exclusively during training to improve learning stability. Each branch consists of a 5×5 average pooling layer with a stride of 3, followed by 1×1 convolutions with 128 filters. This is succeeded by two fully connected layers, producing outputs of 1024 and 1000 neurons, respectively, and ending with a softmax layer for classification. The loss from these auxiliary branches is weighted at 0.3 and combined with the main loss. This approach helps prevent the vanishing gradient problem and adds regularization, enhancing model robustness during training. For automated facial expression recognition, these auxiliary classifiers support the model in learning subtle distinctions between expressions, even for less prominent facial features or nuanced expressions.

This setup improves the model's ability to generalize across diverse facial variations, supporting more accurate classification of expressions such as happiness, sadness, anger, and surprise. Ultimately, this architectural element strengthens the reliability and depth of expression analysis, making Google Net well-suited for applications in emotion recognition, human-computer interaction, and behavioral analysis.

Mathematical Model

Set Theory: S={s, e, X, Y} Where: **s = Start of the Program:**

Authenticate the user and initialize the facial expression recognition program.

Capture the video input from the user along with any queries related to expression analysis.

e = End of the Program:

Display the analyzed facial expressions on the user's screen (monitor or mobile) in real-time or at the end of the analysis.

Log out the user and end the session.

X = Input of the Program:

The input consists of video data capturing the user's facial expressions.

Y = Output of the Program:

The output is the recognized facial expressions, categorized by emotions such as happiness, sadness, anger, or surprise.

This setup describes a streamlined approach for automated facial expression recognition using Google Net, delivering precise expression analysis directly to the user.

Finally we display the doctor's prescription on the screen (monitor or mobile).

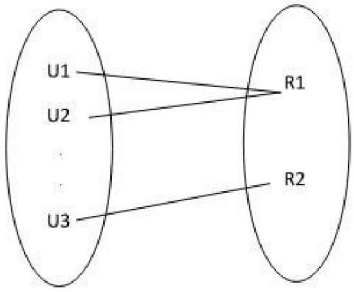

Figure No 5.2: Mathematical Model

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

X Y ϵ U
Let U be the Set of System.
U = {User, I, C, H, A, D, R}
Where User, I, C, H, A, D, and R are the elements of the set, representing key components of the system:
User: The person interacting with the facial expression recognition system.
I: Input data, which consists of video data capturing facial expressions.
C: Camera (or other video capture device) used to record the user's facial expressions.
H: Hardware resources required to process and analyze the video input, such as a GPU-enabled system.
A: Application, which may be a web-based or mobile platform where the facial recognition system operates.
D: Display module, where the captured and analyzed expression data is shown to the user.
R: Result or output, representing the recognized facial expressions, such as happiness, sadness, anger, or surprise.
This set-based representation helps encapsulate the core elements involved in automated facial expression recognition using the Google Net-based model, ensuring a clear understanding of each component's role in the process.
Above mathematical model is NP-Complete=Failures and Success conditions.
Failures:
Huge database can lead to more time consumption to get the information.
Hardware failure.
Software failure.
Success:
Search the required information from available in Datasets.
User gets result very fast according to their needs This Problem is NP- Complete.

VI. APPLICATINS

The main application of this project is to measure pulses and calculate billing and overload monitoring.
This system can be implementing as power theft monitoring .
It can be used in automation system also, to control electrical equipment from long distance.

VII. FUTURE SCOPE

The future scope of this automated facial expression recognition system holds exciting potential for advancements and applications across multiple domains. Firstly, the system can be extended to recognize more subtle and complex expressions, such as mixed emotions, by refining the model and incorporating additional data. This could improve the accuracy and depth of emotional analysis in diverse real-world scenarios. Additionally, integrating the system with wearable devices or smartphones could enable real-time emotion tracking for applications in mental health monitoring, enhancing accessibility for users and professionals alike.
Future work could also focus on adapting the system for cross-cultural and age-related variations in facial expressions to improve its universality and inclusiveness. Moreover, incorporating advanced natural language processing (NLP) and multimodal data (such as voice tone or body language) could allow the system to provide a more comprehensive analysis of human emotions and intent. Finally, by leveraging advancements in artificial intelligence and edge computing, the system could be optimized for faster, more efficient processing, making it viable for deployment in areas such as autonomous vehicles, customer service, education, and healthcare for empathetic and responsive interactions.

VIII. CONCLUSION

In conclusion, the proposed system for facial expression recognition using deep learning techniques represents a significant advancement in emotion analysis. By leveraging a structured methodology that includes training, pre-processing, feature extraction, and classification, the system is capable of accurately identifying a wide range of facial expressions, including happiness, sadness, fear, anger, surprise, disgust, and neutrality. Utilizing a robust model like Google Net enhances the precision and efficiency of expression detection, making the system suitable for real-time

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

applications. This solution holds promising implications for diverse fields, including mental health assessment, human-computer interaction, and behavioral studies, providing a reliable tool for interpreting human emotions. With further development, the system could be refined to achieve higher accuracy and adaptability, making it an invaluable resource in fields that require precise and timely emotion recognition.

REFERENCES

[1]. Ge, Huilin, et al. "Facial expression in deep learning." Computer Methods and Programs in Biomedicine 215 (2022): 106621.

[2]. Bisogni, Carmen, et al. "Impact of deep learning approaches on facial expression recognition in healthcare industries." IEEE Transactions on Industrial Informatics 18.8 (2022): 5619-5627.

[3]. Umer, Saiyed, et al. "Facial expression recognition with trade-offs between data augmentation and deep learning features." Journal of Ambient Intelligence and Humanized Computing (2022): 1-15.

[4]. Ahmed, Zeyad AT, et al. "[Retracted] Facial Features Detection System To Identify Children With Autism Spectrum Disorder: Deep Learning Models." Computational and Mathematical Methods in Medicine 2022.1 (2022): 3941049.

[5]. Savchenko, Andrey V., Lyudmila V. Savchenko, and Ilya Makarov. "Classifying emotions and engagement in online learning based on a single facial expression recognition neural network." IEEE Transactions on Affective Computing 13.4 (2022): 2132-2143.

[6]. Zhu, Armando, et al. "Cross-task multi-branch vision transformer for facial expression and mask wearing classification." arXiv preprint arXiv:2404.14606 (2024).