A Review on Human Behavior Analytics in Surveillance System

Mr. Rohan Chandrakant Patil [1], Dr. Bashirahamad Momin [2]
[1,2] Dept. of Computer Science and EngineeringWalchand College of Engineering, Sangli Maharashtra, India.

Abstract—**Facial Emotion Recognition (FER) has become a crucial topic of study in domains including security, healthcare, and human-computer interaction. A thorough analysis of the most recent techniques, models, and datasets utilized in FER is provided in this review study. Existing surveys were analyzed to identify research gaps and address areas needing improvement. FER datasets are categorized into six emotions sadness, happi- ness, fear, anger, surprise and disgust. The key steps of FER pre-processing, feature extraction and classification are covered in this paper along with a comparison of several deep learning- based models and their performance standards. Challenges such as dataset imbalance, real-time processing, cultural biases, and model scalability are highlighted, along with potential solutions. Additionally, future trends like multi-modal approaches, transfer learning, and reinforcement learning for personalized emotion detection are explored. The findings offer insights into model selection for specific datasets and applications, helping future researchers develop robust face emotion recognition systems. Thisreview aims to facilitate advancements in FER by guiding the creation of scalable and adaptable models capable of performing effectively in real-world surveillance and other applications.**

Index Terms—**Facial Emotion Recognition (FER), Deep Learning, Human-Computer Interaction, Transfer Learning, Surveil- lance Systems.**

I. INTRODUCTION

Human facial expressions are essential visual cues that play a critical role in communicating emotions, allowing individuals to interpret the feelings of others through direct observation, images, or videos. While humans naturally decode these expressions with ease, they pose a significant challenge for machines to accurately interpret. According to psychologist Albert Mehrabian, emotional communication consists of multiple components: only 7% is conveyed through spoken language, while 38% comes from vocal elements such as tone, pitch, and speech rhythm, which can vary widely across cultures. Notably, facial expressions contribute a substantial 55% to emotional communication [1], making them the most significant factor in conveying emotional states. Being able to read someone’s facial expressions gives you important information about their emotions andmental health as well as a glimpse into the actions that are motivated by those feelings. Recognizing these expressions accurately can inform fields like psychology, marketing, security, and education, where understanding emotional responses is critical. However, current deep learning systems often struggle to interpret emotions are inherently complexand exhibit minor variations in expression, facial expressions can be accurately captured and variations in lighting, angles, and individual facial features [2].

Despite significant advancements in computer vision and deep learning, challenges such as these hinder the widespread commercial application of emotion recognition technologies. Many existing systems face limitations in real-time analysis, cultural bias, and the ability to recognize complex emotional states like mixed emotions or micro- expressions. Addressing these challenges requires further research into more robust models that can generalize well across diverse populations and real-world conditions, enhancing the general dependability and accuracy of face emotion detection systems. Developments in this field might have a broad influence because facial expressions are crucial for emotional communication from enhancing human- computer interactions to developing more intuitive assistive technologies and improving mental health assessments [3].Therefore, ongoing review in facial expression recognition is crucial to unlocking its full potential across various industries. Numerous fields of human-computer interaction, including smartphones, efficient computing, intelligent control systems, behavioral and psychological research, pattern recognition,defense, social media analytics, robotics, and more, have seen extensive use of facial emotion recognition (FER). These systems can improve decision-making, user experiences, and offer insightful feedback to refine current technology by deciphering emotions from facial expressions. Deeplearning and computer vision advances are the driving forces behind FER technology, which enables real-time, automated detection, analysis, and interpretation of emotional signals.

The development of FER systems is grounded in foundational psychological studies. The psychologists PaulEkman and Wallace Friesen identified six basic emotions in the 20th century disgust, fear, pleasure, sadness, and surprise through cross-cultural research. These emotions were shown to be universally identifiable despite cultural differences [4]. Later, contempt was added as a seventh basic emotion, expanding the emotional spectrum. This universality in emotional expression has been instrumental in building robust FER systems that can generalize across populations [5]. This step recognizes and extracts face characteristics from pictures or video frames, including the lips, chin, eyes, nose, and eyebrows. To find the face and facial landmarks in a picture, face recognition algorithms like Haar Cascades or methods based on deep learning like MTCNN (Multi-task Cascaded Convolutional Networks) are frequently used. Once facial landmarks are detected, more detailed features are extracted from different regions of the face to capture subtle changesin expression [6].

Despite these advancements, challenges remain in developing systems that can accurately recognize a wide range of complex and subtle emotions, including mixed or masked expressions. The accuracy of systems can also be impacted by facial angles, occlusions, lighting conditions, and cultural differences. Further research is needed to address these limitations, particularly by incorporating more diverse datasets and improving the robustness of models in real-world environments [7]. As FER technology develops further, it has the potential to transform a number of sectors by giving robots the ability to recognize and react to human emotions bringing us closer to more empathetic and human centered deep learning systems.

Emotions are cognitive states or phases that humans experience, often intertwined with moods, attitudes, temperament, personality, disposition, and motivation. These emotions are complex, multi-dimensional phenomena that can be influenced by both internal cognitive processesand external stimuli [8].

They can be broadly categorized into binary sentiments, such as positive and negative, depending on the psychological context or the specific events that trigger them. An individual's behavior, ideas, and interactions are greatly influenced by theseemotional states, which frequently result in dynamic changes throughout time.

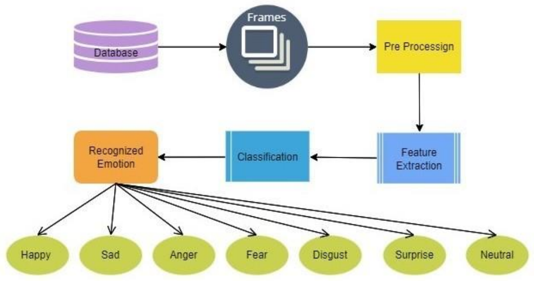

Fig. 1. Facial emotion classification process

Due to this inherent complexity, researchers face significant challenges when studying emotions, especially whenattempting to model or simulate them in human-computer interaction systems. Emotions vary not only from person to person but also across different cultural and situational contexts, and researchers often adopt their own definitionsand assumptions about emotions, It may result in inconsistent results and make it challenging to draw broad generalizations about face emotion identification. The absence of agreed-upondefinitions for emotions, coupled with their subjective nature, presents an ongoing challenge for developing universally accurate systems for emotion detection.

Despite these challenges, emotions are universal to humans and a core set of emotions is recognizable across cultures, which forms the basis for discrete emotion theory [10]. This theory suggests that certain basic emotions are biologically innate and can be identified through specific facial expressions, regardless of cultural background. According to Paul Ekman, people from all walks of life experience the six main emotions of fear, joy, sorrow, surprise, disgust, and anger that shown in figure 1. Ekman's model of basic emotions has been widely validated and is a cornerstone in the making of facial emotion recognition systems which use both facial and vocal cues to identify and classify these emotional states.

Ekman's theory has greatly influenced the design of FER systems which rely heavily on facial expressions and vocal data to categorize emotional states accurately. FER systems areable to identify patterns linked to each of the six fundamental emotions by examining particular facial characteristics, suchas the movements of the lips, eyes, and eyebrows [11]. The inclusion of vocal data, such as tone, pitch and speech rhythm further enhances the accuracy of these systems enabling amore nuanced understanding of emotions in real-time.

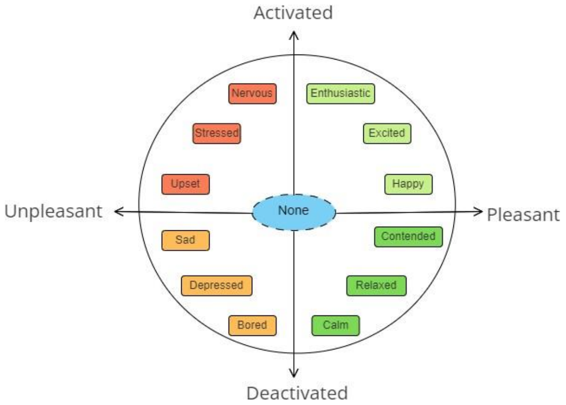

Fig. 2. Plutchik’s Emotion Wheel Model

Alternatively, In Plutchik’s Emotion Wheel as shownin figure 2 offers a more comprehensive model of human emotions. Plutchik expands the range of The eight main emotions joy, fear, anger, sorrow, trust, contempt, surprise,and anticipation are considered fundamental emotions. Hismodel suggests that emotions exist on a spectrum where complex emotions emerge as combinations of the basic ones. For instance, combining joy and trust results in love while fear and surprise combine to create. Plutchik’s model also introduces the idea of emotional intensity where each basic emotion can vary in strength offering a more granular approach to understanding emotional states [12]. Plutchik’s theory emphasizes the interconnectedness of emotions proposing that they are not isolated states but can blend to form more complex emotional experiences. This approach has been used to develop more complex systems for emotion identification and analysis in a number of domains, including affective computing, psychology, and artificial intelligence.

Regarding the recognition of facial emotions (FER), both Ekman’s and Plutchik’s models are critical in shaping how these systems are trained and developed. By focusing on universally recognizable facial cues and combining them with vocal and contextual data, In practical applications, FER systems are becoming more and more adept at identifying a variety of emotions However, many challenges remain particularly in distinguishing between subtle or mixed emotions, where the expression of multiple emotions overlaps, making it harder to classify them with high accuracy.

The cultural variability of emotional expression also complicates FER. While core emotions like happiness or anger are universally recognizable the intensity and way these emotions are expressed can differ significantly across cultures [13]. For instance, certain cultures may promote the repression of specific emotions, like grief or anger, which results in more muted facial expressions, whilst other cultures may promote the more overt exhibition of emotions. FER systems must account for these cultural differences to improve accuracy across diverse populations.

Furthermore, contextual elements like social interactions and environmental circumstances frequently have an impacton emotional states which are not always captured by facial expressions alone. Current research is exploring multi-modal approaches that incorporate contextual information such asbody posture environmental cues and situational background to offer a more comprehensive comprehension of emotional states [14]. Combining these with deep learning techniques FER systems are evolving toward more robust, context-aware models that can detect emotions with greater precision in real-time applications.

The goal of this review paper is to:

- To review all the research paper conducted on Human Behaviour Analytics in Deep Learning, with a focus on Facial Expression Recognition.
- To provide a comparative analysis of research acrossdifferent categories of datasets, including used in surveillance-based emotion recognition.
- To discuss the challenges faced in FER for surveillance systems and propose possible solutions to address these issues.
- To explore future trends that are expected to shape the field of human behaviour analysis in surveillance systems.
- To offer insights into potential real-life applications of Human Behaviour Analytics, particularly in enhancing security monitoring and public safety using surveillance systems.

In this review prior studies that employed small datasetsto examine deep learning models with a focus on research conducted up until August 2024. This section outlines theprogression of these models and highlights key milestones in their development. Next, They give a thorough rundown of the approaches, strategies, and tactics employed in these investigations, addressing the advantages and disadvantages of different datasets as well as their processing methodologies.

Following this, the paper delves into an in-depth analysis of the key concepts, algorithms and techniques that were prominent in publications prior to 2020. This historical context sets the stage for understanding how current approaches have evolved. The discussion addresses not only the technological advances but also challenges related to scalability model interpretability and ethical concerns, such as data privacy and bias in deep learning.

After presenting this technical foundation we are offer a critical analysis of existing research gaps and provide forward looking recommendations. These include emerging trends in deep learning such as transfer learning, explainable DLand the integration of multimodal data to enhance behaviour analysis in complex environments. The recommendations aim to guide future research toward approaches that could significantly increase the scope and depth of knowledge inthis field.

This research draws from a wide range of academic sourcesincluding peer reviewed papers from Scopus, Google Scholar, Elsevier, IEEE and additional scientific datasets. By synthe- sizing insights from these diverse sources The purpose of the study is to present a comprehensive assessment of the stateof behavior analysis powered by deep learning in educational contexts and propose actionable steps for future improvements.

II. DATASETS USED FOR FER SYSTEMS

In order to continuously improve face emotion identification algorithms, facial expression datasets are essential gathering relevant data is to creating automated algorithms that identify emotional groups. Although the accuracy rate for emotionclassification has improved significantly it has yet to reach optimal levels. Taking into account that a person may exhibit awide range of emotions, frequently shifting quickly in a little period of time a comprehensive training dataset is essentialto account for these variations. As the number of emotions to be recognized increases without a large enough and varied training set, neural networks find it harder to discriminate between them correctly. Moreover, training datasets must be sufficiently diverse to avoid biases particularly when dealing with minority classes.

The impact of illnesses or physical disabilities that might result in either temporary or permanent facial paralysis is another important factor to consider, potentially leading to misclassification or even incorrect diagnoses of psychological disorders [1]. The accuracy of emotion classification can vary significantly between different datasets even when the same neural network architecture is used [15].

Currently, there are an enormous number of datasets for emotion detection that contain photographs with varying dimensions, poses, emotions on faces, lighting and topics per image. Usually, these photos are taken in either controlled settings where expressions are simulated or in natural, uncontrolled settings where variations are more extensive. Images captured in controlled environments tend to have limited background variation while those collected in natural settings reflect a broader diversity. The environment in which the data is collected can impact the accuracy of emotion classification, particularly in relation to factors like skin colour or ethnicity. In this review has also shown that cultural norms and societal influences affect how individuals express certain emotions.

For FER systems to achieve robust performance, large and diverse datasets, particularly those captured in natural settings with dynamic conditions, are essential. Since that each person’s facial expressions vary somewhat from another,the quality and diversity of training datasets greatly influenceshow effective these systems are emotions may overlap orindividuals may not visibly express their emotions at all.

In table 1 presents a summary of the most widely used emotion recognition datasets commonly referenced in neural network training literature. These datasets which include both single images and sequences of images or videos depicting specific emotions provide key information such as the envi- ronment used for data collection the number of images, the type of color images, the number of subjects involved and the variety of facial expressions captured. This information is based on a review of existing studies. To effectively analyze both facial and body behavior for emotion detection it's crucial to leverage a comprehensive dataset. The training procedure which depends on a sizable set of labeled instances is crucial to the effectiveness of DL models. In the context of Facial Expression Recognition (FER) numerous datasets have been developed to support researchers in this task. These datasets vary in the number of images and videos they contain aswell as other factors like lighting conditions diversity in the population and variations in facial expressions and poses. Each dataset provides unique challenges and advantages for building accurate models.

III. APPROACHES BASED ON DEEP LEARNING ALGORITHMS

In many artificial intelligence applications including those neural networks are widely utilized in data science, computer vision, machine learning, deep learning, and natural language processing. They efficiently strike a balance between classification accuracy and processing speed and complex architectures that can quickly recognize and categorize patterns by performing the necessary operations to extract certain attributes have been developed as a result of recent advancements. Typically, a neural network goes through threemain stages of operation:

Training phase: By comparing its predictions with the actual ground truth values, In order to enhance performance, thenetwork modifies its settings.

Phase of validation: In this stage, the model's performance is objectively evaluated against a different validation dataset, helping to ensure that the model generalizes well.

Phase of testing: In this stage, input data are sent throughthe various components of the network in order to generatean output value or forecast.

In computer vision applications, neural networks have been shown to be effective like picture categorization particularly in applications like face identification and facial emotion recognition. Beyond their primary use in surveillance systems neural networks are increasingly being applied in medical diagnostics to identify patient conditions [34] and in user interaction applications [35].

As previously mentioned, Deep neural networks have gained popularity in the field of emotion identification dueto their exceptional performance. The domain of computervision is particularly fond of the following kinds of DNNs:

Multi-layer Perceptron: Made up of several fully linkedlayers, MLP is the most basic kind of DNN. It can lessen the significant processing power demands that more intricatedeep learning models sometimes entail.

Convolutional Neural Network: CNNs automatically extractfeatures from input data and are widely utilized in computer vision, facilitating tasks like image classification. They useone or more convolutional layers that carry out convolutional operations using filters, enabling the model to represent the input data at a high level.

Recurrent Neural Network: RNNs are excellent at handling sequential data, including text and time series. Natural language processing (NLP), image captioning, speech recognition, and language translation all often employ them. RNNs include a "memory" mechanism that enables them to use knowledge from past inputs for determining current outputs, and they are distinguished by parameter sharing throughout network levels.

TABLE I
A SUMMARY OF SOME DATASETS

datasets	Description	Emotions
AffectNet [15]	Over 440,000 photos gathered from the internet.	Six fundamental emotions an dneutral.
RAFD DB [16]	Real world 30,000 images.	Six fundamental emotions an dneutral.
SFEW [17]	700 photos including varying head poses, lighting, occlusion, andages.	Six fundamental emotions an dneutral.
FER 2013 [18]	35,887 grayscale photos that were gathered via Google Image Search.	Six fundamental emotions an dneutral.
MMI [19]	2,900 videos annotated with neutral, onset, peak, and offset labels.	Six fundamental emotions an dneutral.
RaFD [20]	8,040 photos including various age, gender, sex, and facial postures.	Six fundamental emotions an dneutral.
CASME II [21]	247 of micro-seque nces expressions.	Regression, Surprise, Disgust, Happy, and others.
Oulu-CASIA [22]	2,880 films taken under three distinct lighting scenarios.	Six fundamental emotions an dneutral.

GEMEP FERA [23]	289 picture sequences.	Anger, Fear, Sadness, Relief, Happy.
Human3.6M [24]	A large-scale dataset for human pose estimation and behavior recognition, featuring 3.6 million 3D human poses captured from multiple viewpoints.	Walking, Sitting, Discussing, Eating, Talking.
MPII Human Pose Dataset[25]	Almost 40,000 persons labeled across approximately 25,000 photos from YouTube recordings of regular people.	Walking, Jumping, Sitting, Running, Dancing.
COCO [26]	Part of the COCO dataset for multi-person keypoint detection, includeshuman body joint annotations in various real-world scenes.	Running, Jumping, Sitting, Playing, Riding, Talking.
Kinetics 700 [27]	A large-scale dataset containing around 650,000 video clips labeledwith 700 human activities.	Running, Dancing, Playing, Cooking, Fighting, Laughing.
PoseTrack [28]	Dataset for multi-person pose estimation and tracking, featuring annotated body joints and movements in video sequences.	Walking, Running, Dancing, Interacting.
ActRec-Classroom [29]	A dataset for classroom behavior analysis with 5,000+ images, including student behavior categories like paying attention or writing.	Listening, Reading, Writing, Raising Hand, Fatigue.
AVA Dataset [30]	Annotates actions in real-world scenes, focusing on spatiotemporal localization of human actions in long-form videos.	Walking, Talking, Hugging, Clapping, Sitting.
PIE [31]	Focuses on human behavior analysis under varying pose, lighting, and facial expressions, ideal for real-world applications.	Happy, Sad, Neutral, Talking.
J-HMDB [32]	A smaller but highly curated dataset with detailed pose and bodypart annotations in various actions from real-world video clips.	Walking, Climbing, Fighting, Kicking, Running.
Penn Action [33]	Includes 2,326 video sequences of 15 different activities with detailedhuman joint and action annotations.	Walking, Running, Jumping, Kicking, Dancing.

In table 2 summarizes, a number of DNN-based architectures have shown notable progress in emotion identification.

TABLE II
ARCHITECTURE TYPES AND MODELS

Architecture	Type
CNN	ResNet12, ResNet34, ResNet56, 2D-ResNet, EmoResNet, VGG14, VGG16, VGG17, VGG-M, GoogleNet, LeNet, YOLOv3, EfficientNet,AlexNet, CAER-Net, CAER-Net-S, MTCNN
GAN	GAN, 2k GAN
RNN	LSTM, EmoNet

Convolutional Neural Networks (CNNs) are the most used DNN architecture in facial emotion recognition (FER) systems shown in figure 3. A sequence of convolutional and pooling procedures make up a typical CNN architecture, then a Soft- Max function for multiclass classification and multiple fully connected (FC) layers.

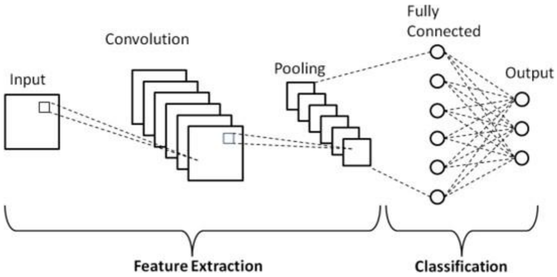

Fig. 3. CNN Architecture

There are several benefits of using a CNN for FER systems. First, CNNs perform at very high levels. They also eliminate the need for human feature extraction because the learning process is done automatically on the training data. The capacity to transfer learning, which enables later models to be constructed upon the fundamental layers of previously trained CNNs, is one of the most prominent advantages [36].

Because information from one task may be transferred to another transfer learning is especially beneficial as By doing away with the requirement to acquire new training data for every job, it cuts down on processing time [37].Consequently, compared to building a network from scratch,It is usually faster and often takes a smaller dataset to usea network that has been trained with transferable learning.The majority of pre-trained networks were developed using portions of the ImageNet collection which contains a wide variety of picture data [38].

Google’s Inception network is one of the most well-known CNN-based designs for a variety of picture categorization applications [39]. Known for its complex architecture the inception network has evolved continually in terms ofspeed and accuracy, resulting in several versions from V1 (also referred to as GoogLeNet) to V4 [40]. Due to the impressive performance of ResNet a hybrid version knownas Inception-ResNet has also been developed [41]. The foundation of Inception networks is the Inception module,which integrates convolutional, pooling, and concatenationoperations. The distinctive feature of the Inception module is its ability to execute convolutional operations at the same level with several filters of varying sizes. This design choice allows the model to become wider rather than deeper, effectively mitigating the risk of overfitting. Additionally, the architecture replaces fully connected layers with average pooling at the network's end, significantly reducing the number of unnecessary parameters. Throughout its evolution each version of the inception network has aimed to enhance computational efficiency while minimizing the number of parameters achieving improvements that also contributed to reducing the error rate. As a result, various research use different variants of the Inception network for fine-tuning,emotion recognition, transfer learning and feature extraction [42].

The Visual Geometry Group (VGG) convolutional neural network is another noteworthy design that performs well in emotion recognition. [43]. The VGG model has other versions such as VGG16 and VGG19 which function based on the same fundamental principles but differ mainly in their level of detail. The network's depth rises as the model moves from basic to more sophisticated iterations adding more convolutional layers in succession to the initial layers. Despite the large size of the VGG architecture, which requires substantial time for parameter training, it has yielded promising results in various applications. Consequently, different VGG variants have been employed in numerous studies focusing on emotion recognition and related tasks [44].

AlexNet [45] and LeNet [46] feature similar architectural designs but AlexNet is distinguished by its significantly greater number of stacked convolutional layers, while LeNet incorporates a pooling layer immediately following certain convolutional layers. LeNet is notable for being one of the pioneering models that introduced convolutional neural networks (CNNs). By using some of the channels from the previous layer for each filter in the convolutional layers, LeNet breaks the network's symmetry and lowers the computational cost. Average pooling is used by the subsampling layers. Although designed for low-resolution images, LeNet's performance was limited by the computational power available at the time, resulting in less impactful outcomes. Both AlexNet and LeNet have been utilized in [47] this to assess techniques for transfer learning applications and face expression recognition [48].

With convolutional layers, the YOLOv3 [49] architecture substitutes independent logistic classifiers for the conventional SoftMax activation functions. This allows for predictions to be made at three different scales, enhancing the model's accuracy in object detection. The YOLOv3 face detection model has been utilized for feature extraction in [50] this.

Another notable CNN is EfficientNet [51], It has been enhanced to get a high degree of precision. Compound scaling is a technique used in this model to efficiently expand the model size. EfficientNet achieves balanced and Using a predetermined set of coefficients to scale each dimension uniformly improves efficient performance instead than altering the width, depth or resolution randomly.

Convolutional neural networks (CNNs) as shown in table 3 are often the neural network design most frequently utilized for emotion recognition. CNNs have continually shown notable results in both practical applications and theoretical model development, whether they are combined with other types of networks or utilized individually for extraction of features fol- lowed by classification. Moreover, this architecture enables the creation of functional solutions for real-time implementation.

TABLE III
CNN DESIGN UTILIZED IN FER SYSTEMS

Architecture	Emotions Detected	Accuracy	Dataset Used
CNN [37]	7	99.69%	CK+
VGG16 [44]	6	76.2%	FER-2013
AlexNet, GoogLeNet, LeNet [45]	8	99.93%, 98.58%	Multi-PIE, CK+
SVM [52]	7	94.69%	BU4D
ResNet50 [53]	9	85%	Caltech-256
Two-level CNN [54]	5	96%, 85%	Caltech faces, NIST
ResNet50 [55]	6	92.78%	FER-2013

2D CNN–LSTM [56]	7	79.9%	RAF-DB
CNN–RNN [57]	7	78.4%, 3.9%	MMI, CK+
MTCNN [58]	8	60.99%, 99.22%	FER 2013, iSPL

Deep neural networks, which are being developed to sim- ulate intricate human cognitive functions are becoming more and more reliant on generative adversarial networks (GANs) for facial emotion recognition (FER) systems.

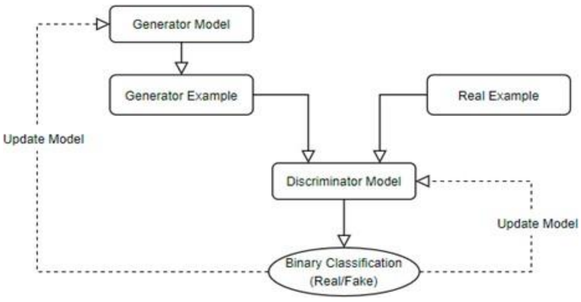

Fig. 4. GAN Architecture

Researchers are exploring GANs' potential to enhance the capabilities of neural networks, particularly their ability toapproach human-like thought processes. For example, in the realm of computer vision GANs not only attempt to replicate images based on training data but also train themselves to create entirely new highly realistic images [59].

In GAN architecture shown in figure 4, the generative model creates outputs based on given input and these are evaluated by a discriminator model which identifies whether the results are real or generated by the network. GANs arealso flexible in that they can impose relational inductive biases in the data. For instance, in facial recognition tasks, facial landmarks are modeled as graphs, allowing the systemto make inferences about facial attributes and identity [60].

Recurrent neural networks (RNNs) are also commonly usedin facial emotion recognition (FER) systems. shown in figure 5, The key distinction between RNNs and traditional neural networks lies in their recurrent layers, where connections between neurons form cycles. In FER, RNNs are often utilized for processing sequences of images. In these cases, each imagein the sequence can depend on the context established by priorimages enabling more effective emotion recognition. RNNs use forward propagation and store data that may be required later. If the model's prediction is incorrect adjustments are made using the learning rate. With each backpropagation cycle, the model's accuracy improves incrementally.

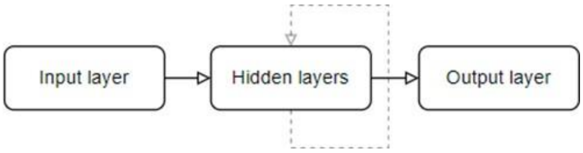

Fig. 5. RNN Architecture

Facial Emotion Recognition (FER) systems are capable of recognizing and detecting human emotions, but because each person experiences emotions differently, they are not always 100% accurate. Interpreting the context is also crucial for understanding human emotions, which remains a challenging task for AI-based systems. However, the process of recogniz- ing facial emotions enables differentiation between allies and adversaries between potential and real threats, serving as a vital source of information in social interactions. From this standpoint the significance of FER systems becomes clear.As interpersonal relationships deepen, perceiving the emotionsof others becomes critical for effective communication. Addi- tionally, In human-computer interaction, automated emotion recognition is crucial, helping to reduce some of the artificial aspects of such interactions and improving communication.

IV. RESEARCH CHALLENGES AND OPEN ISSUES

We analysis various papers research gaps in existing methods for group tracking in video monitoring systems, particularly when comparing learning-based approaches. One key issue is the extended inference time of learning-basedmethods, which can take 1 to 4.5 times longer than the total running time of a video, making them impractical for real-time applications. Current methods often fail to maintain accurate tracking when objects are blocked from view [63].

Need for representative datasets tailored to specific environments, considering factors like age, gender, race, culture, clothing, and climate, which can restrict the model's applicability across diverse settings. Model aims to maintain low computational complexity, there are concerns aboutits processing speed and efficiency, especially for real-time applications, which must be addressed to enable practical deployment in surveillance systems [64].

The scalability and adaptability of the model are also not addressed, particularly how it can be extended to different surveillance environments or handling varying numbers ofconsumer products, a key factor for practical deployment.The paper does not consider user privacy and data protection crucial aspects when using video surveillance for consumer product detection in today's technology-driven landscape [15].

One significant gap is the limited use of temporal information, as in HAR studies have focused on traditional machine learning algorithms without fully utilizing thepotential of deep learning to model time sequences effectively.Additionally, the variability of activities and residents in smart home environments poses a challenge, with existingmodels often failing to account for the diversity in human behavior, including differences in age, health conditions, and interactions with pets [65].

The study identifies several critical concerns related to the detection of depression using video analysis, focusing on the issues of false positives, false negatives, and the importance of reliable data annotation. One major risk is the occurrenceof false positives, where individuals may be mistakenly identified as depressed, potentially causing unnecessary distress. Conversely, false negatives represent missed opportunities for early intervention, where individuals in need of support are overlooked. Both scenarios highlight the need for improved accuracy in detection systems [66].

V. FUTURE DIRECTIONS

There are a number of exciting avenues for developing deception detection with contemporary deep learning (DL) methods. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), two sophisticated models that can automatically uncover intricate patterns from unprocessed data, may be integrated in future studies. While most of the existing literature relies on Adopting deep learning methodologies can provide deeper insights into the interactionbetween verbal and nonverbal clues in deception scenarios, in contrast to typical machine learning techniques like decision trees and support vector machines (SVMs). Combining these methods may result in detection systems that are more reliable and accurate in fields like behavioral video analysis and natural language processing.

Furthermore, Super-Resolution techniques can enhance the effectiveness of deception detection systems by improving the quality of surveillance footage. Traditional interpolation methods like bilinear or bi-cubic interpolation often fail to reconstruct the lost information accurately. In contrast, models such as SRCNN (Super-Resolution Convolutional Neural Network) and VDSR (Very Deep Super-Resolution) offer superior performance by learning high-level representations of low-resolution inputs which could significantly benefit small or blurry visual data in surveillance.

Transfer learning, which enables models trained on huge, popular data sets to change to smaller, domain-specific datasets, is another significant development in the area. In Facial Emotion Recognition (FER) systems, transfer learning ensures that pre-trained models can be fine tuned to identify emotions with minimal data requirements significantly reducing the time and resources needed to train models from scratch. This approach provides a workable alternative forreal-time emotion identification as it is particularly applicable in situations when labeled data is limited.

In addition, Deep Reinforcement Learning presents an innovative direction for emotion behavior analysis. Deep Reinforcement Learning enables AI agents to interact with their environment, receive feedback through rewards,and optimize their actions over time. This capability can be extended to facial emotion classification systems by continuously learning from diverse scenarios and emotional states. DRL can help detect personalized emotional responses that vary across individuals, thus improving the overall performance of emotion recognition in real-world applications.

These future directions collectively highlight the need for multi-disciplinary approaches that incorporate deep learning algorithms, advanced vision systems, and reinforcement learn-ing. With further exploration, these innovations can drive significant progress in deception detection and human behavior analysis, paving the way for more reliable, scalable and adaptive surveillance systems.

VI. CONCLUSION

Due to its numerous uses in fields including education, surveillance, and human-computer interaction, facial emotion recognition, or FER, has attracted a lot of interest lately. However, the research in this domain still faces challenges including the need for comprehensive datasets, improved model accuracy and better handling of complex emotional states. By emphasizing important elements including datasets, pre-processing techniques, feature extraction methods, and classification methodologies, this study sought to present a comprehensive picture of the current status of FER.

The paper conducted an in-depth analysis of the existing surveys and datasets categorizing them into groups suchas sadness, happiness, fear, anger, surprise and disgust to emphasize the diversity required for FER systems to generalize well across different populations. It was observed that a major gap exists in well-balanced datasets for children, adult people indicating a potential area for future work. Through a comparison of deep learning models, their architectural differences and benchmark accuracies the study offers insights that will help guide researchers in selecting appropriate models based on specific datasets and applications.

The study examined outstanding issues such real-time processing limits in addition to reviewing the advantages and disadvantages of current approaches. interpretability of the model, dataset imbalance and cultural biases in emotion recognition. Possible solutions were suggested to overcome these challenges, along with a discussion of future trends in FER, including transfer learning, multi-modal analysis andreinforcement learning techniques.

The development of FER systems will be greatly aided by developments in architectures for deep learning and the acces- sible availability of more varied datasets as the field develops.By addressing the identified research gaps and open issues future studies can develop more robust, adaptable, and scalableFER solutions, especially in surveillance environments where the need for accurate emotion recognition is paramount.

REFERENCES

[1] Mehrabian, A., & Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. Journal of consulting psy- chology, 31(3), 248.

[2] Lumley, M. A., Cohen, J. L., Borszcz, G. S., Cano, A., Radcliffe, A. M., Porter, L. S., ... & Keefe, F. J. (2011). Pain and emotion: a biopsychoso- cial review of recent research. Journal of clinical psychology, 67(9), 942-968.

[3] Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. Sensors, 20(3), 592.

[4] Ekman, P., Matsumoto, D., & Friesen, W. V. (1997). Facial expression in affective disorders. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS), 2, 331-342.

[5] Eliot, J. A., & Hirumi, A. (2019). Emotion theory in education research practice: an interdisciplinary critical literature review. Educational tech- nology research and development, 67, 1065-1084.

[6] Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross- cultural recognition of basic emotions through nonverbal emotional vocalizations. Proceedings of the National Academy of Sciences, 107(6),2408-2412.

[7] Spezialetti, M., Placidi, G., & Rossi, S. (2020). Emotion recognition for human-robot interaction: Recent advances and future perspectives. Fron-tiers in Robotics and AI, 7, 532279.

[8] Ramis, S., Buades, J. M., & Perales, F. J. (2020). Using a social robotto evaluate facial expressions in the wild. Sensors, 20(23), 6716.

[9] Mahmood, M. R., Abdulrazzaq, M. B., Zeebaree, S., Ibrahim, A. K., Zebari, R. R., & Dino, H. I. (2021). Classification techniques' performance evaluation for facial expression recognition. Indonesian Journal of Electrical Engineering and Computer Science, 21(2), 176- 1184.

[10] Bhatti, Y. K., Jamil, A., Nida, N., Yousaf, M. H., Viriri, S., & Velastin, S. A. (2021). Facial expression recognition of instructor using deep features and extreme learning machine. Computational Intelligence and Neuroscience, 2021(1), 5570870.

[11] Shen, P., Wan, D., & Li, J. (2023). How human–computer interaction perception affects consumer well-being in the context of online retail: From the perspective of autonomy. Nankai Business Review Interna- tional, 14(1), 102-127.

[12] Mohsin, M. A., & Beltiukov, A. (2019, May). Summarizing emotions from text using Plutchik's wheel of emotions. In 7th scientific conference on information technologies for intelligent decision making support (ITIDS 2019) (pp. 291-294). Atlantis Press.

[13] Hudlicka, E. (2017). Computational modeling of cognition–emotion interactions: Theoretical and practical relevance for behavioral health-

care. In Emotions and affect in human factors and human-computer interaction (pp. 383-436). Academic Press.

[14] He, Z., Li, Z., Yang, F., Wang, L., Li, J., Zhou, C., & Pan, J. (2020). Advances in multimodal emotion recognition based on brain–computer interfaces. Brain sciences, 10(10), 687.

[15] Sui, J., Jiang, R., Bustillo, J., & Calhoun, V. (2020). Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. Biological psychiatry, 88(11), 818- 828.

[16] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). Affectnet: A dataset for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 10(1), 18-31.

[17] Li, S., Deng, W., & Du, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2852-2861).

[18] Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. J. (2006, April). A 3D facial expression dataset for facial behavior research. In 7th internationalconference on automatic face and gesture recognition (FGR06) (pp. 211- 216). IEEE.

[19] Gursesli, M. C., Lombardi, S., Duradoni, M., Bocchi, L., Guazzini,A., & Lanata, A. (2024). Facial emotion recognition (FER) through custom lightweight CNN model: performance evaluation in public datasets. IEEE Access.

[20] Yang, C., Guan, Y., Liu, T., Zhang, B., & Li, W. (2024). MMI-ML: Maximize Mutual Information between Different Views for Few-Shot Remote Sensing Image Classification. IEEE Geoscience and Remote Sensing Letters.

[21] Mohammadzadeh, A., Tehrani-Doost, M., & Yaghoobi, E. (2024). Va- lidity study of an emotional face-dataset in Iranian community. Iranian Journal of Psychiatry, 1-9.

[22] Chauhan, A., & Jain, S. (2024). FMeAR: FACS Driven Ensemble Model for Micro-Expression Action Unit Recognition. SN Computer Science, 5(5), 598.

[23] Zhang, H., Huang, B., & Tian, G. (2020). Facial expression recognition based on deep convolution long short-term memory networks of double-channel weighted mixture. Pattern Recognition Letters, 131, 128-134.

[24] Valstar, M. F., Jiang, B., Mehu, M., Pantic, M., & Scherer, K. (2011, March). The first facial expression recognition and analysis challenge. In 2011 IEEE international conference on automatic face & gesture recognition (FG) (pp. 921-926). IEEE.

[25] Ionescu, C., Papava, D., Olaru, V., & Sminchisescu, C. (2014). Human3. 6M. Ieee Transactions on Pattern Analysis and Machine intelligence, 1.

[26] Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In Proceedings of theIEEE/CVF conference on computer vision and pattern recognition (pp. 5693-5703).

[27] Puri, D. (2019, September). COCO dataset stuff segmentation challenge. In 2019 5th international conference on computing, communication, control and automation (ICCUBEA) (pp. 1-5). IEEE.

[28] Smaira, L., Carreira, J., Noland, E., Clancy, E., Wu, A., & Zisser- man, A. (2020). A short note on the kinetics-700-2020 human action dataset. arXiv preprint arXiv:2010.10864.

[29] Doering, A., Chen, D., Zhang, S., Schiele, B., & Gall, J. (2022). Posetrack21: A dataset for person search, multi-object tracking and multi-person pose tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20963-20972).

[30] Yang, F., Wang, T., & Wang, X. (2023, September). Student Classroom Behavior Detection Based on YOLOv7+ BRA and Multi-model Fusion. In International Conference on Image and Graphics (pp. 41-52). Cham: Springer Nature Switzerland.

[31] Li, A., Thotakuri, M., Ross, D. A., Carreira, J., Vostrikov, A., & Zisserman, A. (2020). The ava-kinetics localized human actions video dataset. arXiv preprint arXiv:2005.00214.

[32] Rasouli, A., Kotseruba, I., Kunic, T., & Tsotsos, J. K. (2019). Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 6262-6271).

[33] Sahoo, S. P., & Ari, S. (2019). On an algorithm for human actionrecognition. Expert Systems with Applications, 115, 524-534.

[34] Luvizon, D. C., Picard, D., & Tabia, H. (2020). Multi-task deep learning for real-time 3D human pose estimation and action recognition.

IEEE transactions on pattern analysis and machine intelligence, 43(8), 2752- 2764.

[35] Habib, S., Ahmad, M., Haq, Y. U., Sana, R., Muneer, A., Waseem, M., ... & Dev, S. (2024). Advancing Taxonomic Classification through Deep Learning: A Robust Artificial Intelligence Framework for Species Identification Using Natural Images. IEEE Access.

[36]] Alhasson, H. F., Alsaheel, G. M., Alsalamah, A. A., Alharbi, N. S., Alhujilan, J. M., & Alharbi, S. S. (2024). Integration of machine learning bi-modal engagement emotion detection model to self-reporting for educational satisfaction measurement. International Journal of Information Technology, 1-15.

[37] Phung, & Rhee,. (2019). A High-Accuracy Model Average Ensem-ble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. Applied Sciences. 9. 4500. 10.3390/app9214500.

[38] Pereira, R., Mendes, C., Ribeiro, J., Ribeiro, R., Miragaia, R., Rodrigues, N., ... & Pereira, A. (2024). Systematic Review of Emotion Detection with Computer Vision and Deep Learning. Sensors, 24(11), 3484.

[39] Yu, C., Zhang, D., Zou, W., & Li, M. (2024). Joint Training on Multiple Datasets With Inconsistent Labeling Criteria for Facial Expression Recognition. IEEE Transactions on Affective Computing.

[40] Ali, Md. Forhad & Khatun, Mehenag & Turzo, Nakib. (2020). Facial Emotion Detection Using Neural Network. International Journal of Scientific and Engineering Research. 11. 1318-1325.

[41] Li, C., & Zhang, C. (2024). Toward a Deeper understanding: RetNet viewed through convolution. Pattern Recognition, 110625.

[42] Melinte, D. O., & Vladareanu, L. (2020). Facial Expressions Recog- nition for Human-Robot Interaction Using Deep Convolutional Neural Networks with Rectified Adam Optimizer. Sensors (Basel, Switzer- land), 20(8), 2393. https://doi.org/10.3390/s20082393

[43] Banerjee, A., & Banik, D. (2024). Resnet based hybrid convolution LSTM for hyperspectral image classification. Multimedia Tools and Applications, 83(15), 45059-45070.

[44] Jiddah, S. M., Kuter, B., & Yurtkan, K. (2024). Stress Detection through Compound Facial Expressions Using Neural Networks. The Eurasia Proceedings of Educational and Social Sciences, 1-9.

[45] Zhang, Y., Wei, X. S., Zhou, B., & Wu, J. (2021, May). Bag of tricks forlong-tailed visual recognition with deep convolutional neural networks. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 4, pp. 3447-3455).

[46] Oh, G., Ryu, J., Jeong, E., Yang, J. H., Hwang, S., Lee, S., & Lim, S. (2021). Drer: Deep learning–based driver’s real emotion recognizer. Sen-sors, 21(6), 2166.

[47] Setiawan, W., Ghofur, A., Rachman, F. H., & Rulaningtyas, R. (2021). Deep convolutional neural network alexnet and squeezenet for maize leaf diseases image classification. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control.

[48] Shuang, Y., & Gang, W. (2020, November). Research on Meter Image Recognition Based on Improved LeNet-5. In Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering (pp. 676-680).

[49] Shin, D. H., Chung, K., & Park, R. C. (2019). Detection of emotion using multi-block deep learning in a self-management interview app. Applied Sciences, 9(22), 4830.

[50] Cˆırneanu, A. L., Popescu, D., & Iordache, D. (2023). New trendsin emotion recognition using image analysis by neural networks, a systematic review. Sensors, 23(16), 7092.

[51] Adarsh, P., Rathi, P., & Kumar, M. (2020, March). YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. In 2020 6th international conference on advanced computing and communication systems (ICACCS) (pp. 687-694). IEEE.

[52] Rathod, M., Dalvi, C., Kaur, K., Patil, S., Gite, S., Kamat, P., ... & Gabralla, L. A. (2022). Kids’ emotion recognition using various deep-learning models with explainable ai. Sensors, 22(20), 8066.

[53] Wang, C. C., Chiu, C. T., & Chang, J. Y. (2023). Efficientnet-elite: Extremely lightweight and efficient cnn models for edge devices by network candidate search. Journal of Signal Processing Systems, 95(5), 657-669.

[54] Kim, J. C., Kim, M. H., Suh, H. E., Naseem, M. T., & Lee, C. S. (2022). Hybrid approach for

facial expression recognition using convolutional neural networks and SVM. Applied Sciences, 12(11), 5493.

[55] Manzoor, A., Ahmad, W., Ehatisham-ul-Haq, M., Hannan, A., Khan,
M. A., Ashraf, M. U., ... & Alfakeeh, A. S. (2020). Inferring emotion tags from object images using convolutional neural network. Applied Sciences, 10(15), 5333.

[56] Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). SN Applied Sciences, 2(3), 446.

[57] Subudhiray, S., Palo, H. K., & Das, N. (2023). Effective recognition of facial emotions using dual transfer learned feature vectors and support vector machine. International Journal of Information Technology, 15(1),301-313.

[58] Cˆırneanu, A. L., Popescu, D., & Iordache, D. (2023). New trendsin emotion recognition using image analysis by neural networks, a systematic review. Sensors, 23(16), 7092.

[59] Fan, J., Wang, S., Yang, P., & Yang, Y. (2020, July). Multi-view facial expression recognition based on multitask learning and generative adversarial network. In 2020 IEEE 18th International Conference on Industrial Informatics (INDIN) (Vol. 1, pp. 573-578). IEEE.

[60] Yang, H., Zhu, K., Huang, D., Li, H., Wang, Y., & Chen, L. (2021). Intensity enhancement via GAN for multimodal face expression recog- nition. Neurocomputing, 454, 124-134.

[61] Li, C., Bao, Z., Li, L., & Zhao, Z. (2020). Exploring temporal repre- sentations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition. Information Processing & Manage- ment, 57(3), 102185.

[62] Atif, M., & Franzoni, V. (2022). Tell me more: automating emojis classi- fication for better accessibility and emotional context recognition. FutureInternet, 14(5), 142.

[63] Yoon, H., Choi, D., & Chung, Y. D. (2023). Group Tracking for Video Monitoring Systems: A Spatio-Temporal Query Processing Ap- proach. IEEE Access, 11, 19969-19987.

[64] Kalyta, O., Barmak, O., Radiuk, P., & Krak, I. (2023). Facial emotion recognition for photo and video surveillance based on machine learning and visual analytics. Applied Sciences, 13(17), 9890.

[65] Liciotti, D., Bernardini, M., Romeo, L., & Frontoni, E. (2020). A sequential deep learning application for recognising human activities in smart homes. Neurocomputing, 396, 501-513.

[66] Min, K., Yoon, J., Kang, M., Lee, D., Park, E., & Han, J. (2023). De- tecting depression on video logs using audiovisual features. Humanities and Social Sciences Communications, 10(1), 1-8.