Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.10 n.48, p. 14-20. July/August., 2024.
DOI: https://doi.org/10.5935/jetia.v10i48.1007

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

REAL-TIME EMOTION RECOGNITION AND CLASSIFICATION FOR DIVERSE SUGGESTIONS USING DEEP LEARNING - A COMPREHENSIVE SURVEY

***Sanika A. Gonjari[1], Rachana N. Pawar[2], Ritika A. Pawar[3], Sharwari H. Kshirsagar[4], Rohini B. Kokare[5]**

[1,2,3,4] Student, VPKBIET Baramati, Pune, Maharashtra, India.
[5] Assistant Professor, AIDS Department, VPKBIET Baramati,Pune, Maharashtra, India.

[1] http://orcid.org/0009-0008-7467-4223, [2] http://orcid.org/0009-0005-7373-1255, [3] http://orcid.org/0009-0007-9199-7705
[4] http://orcid.org/0009-0003-0871-1587, [5] http://orcid.org/0009-0002-1633-1317

Email: *sanikagonjari2003@gmail.com[1], rachananp09@gmail.com[2], ritikap2003@gmail.com[3], kshirsagarsharwari02@gmail.com[4], rohinikokare@gmail.com[5].

ARTICLE INFO

Article History
Received: December 26 th,2023
Revised: July 08th,2024
Accepted: July 8 th,2024
Published: July 18 th,2024

Keywords:
Emotion Recognition,
Diverse Suggestions,
Deep CNN,
Customized Recommendation,
User Experience.

ABSTRACT

At the captivating nexus of technology and human emotions, we harness cutting-edge software to discern individuals' feelings through their facial expressions. This unique capability empowers us to provide customized recommendations for various forms of entertainment and enrichment, such as movies, music, books, and meditation practices. For instance, if a user appears sad, we can offer upbeat music to brighten their mood. Our ultimate objective is to enhance technology comprehension of emotions, enabling it to suggest content that resonates with people's emotional states. Through this integration of facial expressions and diversified suggestions, here aim is to cultivate a more user-friendly and supportive digital environment, fostering feelings of happiness and calmness. Exploring the captivating domain of facial emotion recognition and recommendation systems is the central emphasis of this endeavor. The primary ambition is to construct a framework capable of deciphering user emotions from facial cues. By using deep learning techniques such as the CNN algorithm which can be applied to facial images to process emotion recognition feature vectors with the categorization of individualized content. It attempts to form a balanced blend of advanced tech and boost user engagement, nurturing emotional connections in the digital domain.

I. INTRODUCTION

In recent years, significant improvements in computer vision and machine learning have been made to build systems capable of recognizing and comprehending human emotions from facial expressions in real time. Deep learning is a branch of machine learning that focuses on creating and training deep neural networks, which are artificial neural networks with several layers. These networks are made to automatically recognize and depict intricate patterns and nested characteristics in data. Deep learning algorithms, in contrast to conventional machine learning algorithms, can automatically find and extract pertinent characteristics from raw data, making them extremely successful for a variety of tasks like speech and image recognition, natural language processing, and even gameplay.

Convolutional Neural Networks (CNNs), Convolutional Recurrent Neural Networks (CRNNs), Deep Convolutional Neural Networks (Deep CNNs), and Recommendation Techniques are some of the foundational technologies we examine in this survey study that are essential to contemporary research and applications. Deep CNNs expand the capabilities of CNNs in feature extraction, CRNNs combine the skills of CNNs and RNNs for sequential data processing, and CNNs excel at extracting complicated patterns from data. Techniques for recommendations improve user experience and system performance. This survey offers insights

into the relevance and potential of these technologies and their applications across a range of areas through an in-depth analysis of both.

The human experience is fundamentally shaped by emotions, which have an impact on how we interact, relate to one another, and communicate. The difficult task of facial expression recognition (FER) involves identifying facial expressions in pictures. Accurate classification is challenging because of expression variability and inter-subject variances. An innovative strategy is put out to combat this, combining identity and emotion data to improve FER. Identity and emotion data are extracted separately using deep neural networks, which are well-known for their success in face recognition [1].

A fast-expanding study area with a wide range of possible applications is real-time emotion recognition through facial expressions. The effects of this technology are wide-ranging and significant. The technique described in this study uses physiological cues like photoplethysmography (PPG) and the galvanic skin response (GSR) to classify music according to how it affects listeners' emotions [2]. Deep neural networks and regression models are used in this novel method to accurately classify music based on emotions without the usage of sensors, increasing music recommendation services for a more personally tailored user experience [2]. The COVID-19 epidemic has sped up the development of online learning, forcing universities and other organizations all around the world to switch to remote teaching methods [3]. However, there are special difficulties with evaluating the emotions and participation of students when learning online. Online environments frequently lack such involvement, in contrast to conventional in-person classes where teachers may monitor students' responses in real-time. The objective is to empower teachers and improve the overall online learning experience by utilizing lightweight facial recognition models, effective neural network architecture, and real-time video-based classification.[3] The cold-start problem and static user representations are drawbacks of conventional personalized recommendation techniques like collaborative filtering and content-based filtering [4]. In the following sections, we will undertake a thorough examination of the diverse methodologies and technologies that have propelled the advancement of real-time facial emotion identification. Additionally, we will delve into the significant challenges that researchers face in their quest for precise and resilient emotion recognition in real-world situations. As we navigate through this captivating landscape, we will also shed light on intriguing possibilities for future research, where further progress in this field has the potential to revolutionize our interactions with technology and, ultimately, with each other. The proposed system delivers a user-friendly mobile application and shows improved sentiment analysis using an upgraded measure, eventually intending to improve users' emotional states and well-being [5].

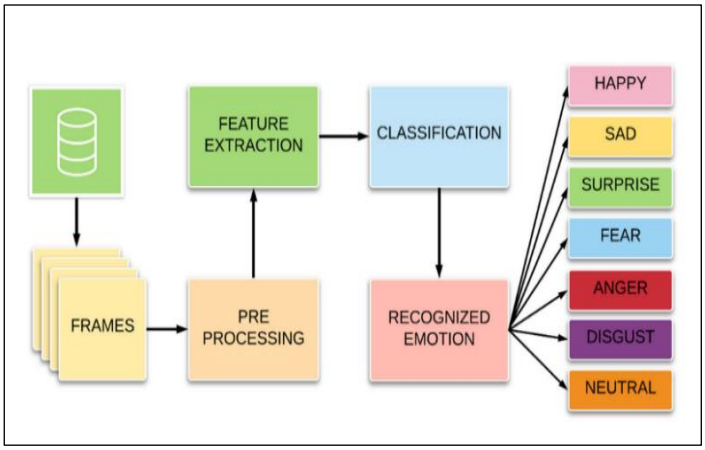

Figure1. Facial Emotion Classification Process.
Source: [6].

II. PLANNED ANALYSIS

II.1 COMPARATIVE ANALYSIS OF THE PLANNED SURVEY

Table 1: Survey Table.

Sr No	Paper	Technique Used	Advantages	Gaps
1.	[3]	Multi-Task Convolutional Neural Networks	Real-time processing is faster with an accuracy of 66.34%.	1. Classified only 3 classes. 2. Low-resolution images not identified.
2.	[2]	Convolutional recurrent neural network, multilayer perceptron	Accuracy 96.26%.	1. RS-BF + RF attained the worst results. 2. More computation power.
3.	[1]	Deep CNN, deep learned Tandem Facial Expression	In comparison with CK+ and FER+, CK+ gives good results.	1. The FER+ database attained the worst results 2. Transfer learning and multi-task learning can improve FER.
4.	[4]	Deep Learning methods and CNN	Comprehensive User Interest Representations, Click-Through Prediction.	Cold-start scenarios, privacy, Diversity in Content Recommendations.
5.	[7]	CNN	Three methods are used for book recommendation with accuracy 100%.	Emotion Recognition less accurate (60%).
6.	[5]	CNN BLSTM-RNN.	BLSTM-RNN gives accuracy 89%	Use advanced recommendation algorithms.

Source: Authors, (2024).

II.2 DESCRIPTIVE ANALYSIS OF THE PLANNED SURVEY

In paper [3], in the context of image classification, Multi-Task Convolutional Neural Networks (CNNs) have been employed as a technique to improve real-time processing speed while achieving an accuracy of 66.34%. However, it is important to note that these networks exhibit limitations. Firstly, they are designed to classify only three specific classes, which may restrict their applicability in scenarios with a broader range of categories. Additionally, Multi-Task CNNs struggle to accurately identify objects in low-resolution images, highlighting a weakness in their ability to handle such visual data effectively.

In the Proposed paper [2], the Authors state that the utilization of ANN and Multilayer Perceptron achieved an accuracy of 75.46%. However, it's important to note that the combination of Random Sampling (RS) and Random Forest (RF) produced suboptimal results, indicating the need for further exploration and refinement of this particular approach.

Authors presented Deep CNN and deep learned Tandem Facial Expression analysis techniques [1] show promising results, with the CK+ dataset achieving a high accuracy of 99.31%, but the FER+ dataset lags with an accuracy of 84.3%. The potential for improvement lies in exploring transfer and multi-task learning methods to enhance facial expression recognition in the FER+ dataset.

Deep Learning, including CNN, provides robust user interest representations and improves click-through prediction in recommendation systems [4]. Nevertheless, it struggles with challenges such as cold-start scenarios, privacy concerns, and content diversity in recommendations, necessitating ongoing research for solutions.

In the domain of book recommendation, the employment of Convolutional Neural Networks (CNN) has yielded impressive results [7], achieving a remarkable accuracy rate of 100%. However, it is important to note that these CNN-based methods may not perform as effectively in the realm of Emotion Recognition, where they exhibit a comparatively lower accuracy rate of 60%.

In the context of this study [5], the utilization of CNN and Bidirectional Long Short-Term Memory Recurrent Neural Networks (BLSTM-RNN) has proven advantageous, with BLSTM-RNN achieving a commendable accuracy of 89%. However, it is imperative to consider the integration of more advanced recommendation algorithms, highlighting the potential for further enhancing the overall performance and efficacy of the system [5].

II.3. TECHNICAL ANALYSIS OF THE PLANNED SURVEY

II.3.1. CONVOLUTIONAL NEURAL NETWORKS (CNNS)

CNNs are a subset of deep neural networks that are mostly used for computer vision-related tasks like image classification, segmentation, and object detection. They consist of several layers, such as fully connected, pooling, and convolutional layers. Convolutional layers are particularly important because they can assist the network in learning hierarchical features from images [3].

Multi-Task Learning (MTL)

In the MTL machine learning paradigm, a single model is trained to carry out several tasks concurrently. It is proposed that learning several related tasks simultaneously can result in better performance than training separate models for each task [3].

MT-CNNs

The architecture of CNNs is expanded by MT-CNNs to support multiple tasks. For each task, this is typically accomplished by adding more output layers. An MT-CNN might have additional output branches for tasks like object detection, segmentation, and pose estimation, for instance, if a standard CNN is made for image classification [1]. The "Face Detection 1" unit uses any quick method, such as MTCNN (multi-task CNN), to locate the largest facial region in each t-th video frame. The proposed lightweight CNN is then used by the unit Emotional feature extraction to obtain the emotional features x(t) of the extracted face. This CNN has been trained to classify emotions on static images [3].

II.3.2. Convolutional Recurrent Neural Networks (CRNNs)

The benefits of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are combined to create CRNNs, which can process data that has both spatial and sequential dependencies. They are especially useful for tasks that require understanding local patterns and long-term dependencies, such as those involving time-series data or sequences of images.

Multilayer Perceptron (MLPs)

Multilayer perceptron (MLP) neural networks are a particular kind of feedforward neural network. They consist of a large number of interconnected layers of nodes (neurons), each of which is connected to every node in the layer beneath and the layer above it. Each connection has a corresponding weight, and each node applies an activation function to the weighted sum of its inputs. In the absence of physiological sensors, regression target features are selected from the previously created GSR and PPG features that are extracted from the sample-level inner attention-mechanism-based convolutional recurrent neural network encoder [2]. The correlation model between the selected regression target features and musical features is then created and sent to the smartphone by the MLP regression training on the server [2]. The musical characteristics of the input music signal are fed into MLP-based regressors to automatically generate emotion features for PPG and GSR during the application phase on the smartphone. The combined GSR and PPG features are then fed into segment-level inner attention mechanism-based bidirectional gated recurrent neural networks for emotion-based music classification [2].

II.3.3. Facial Expression Recognition Using Deep CNN

A Deep Convolutional Neural Network (CNN) is a specific kind of neural network that is used for processing data that is grid-like, like images. Its specialty is automatically extracting hierarchical features from input images.Facial expression recognition uses facial images to quickly extract pertinent features. The computer vision task of Facial Expression Recognition (FER) uses Deep Convolutional Neural Networks (CNNs) to automatically categorize facial expressions in images or videos into predefined emotion categories (such as happiness, sadness, anger, etc.).

Using a special kind of neural network called a CNN, this method automatically extracts pertinent facial features from facial images. For processing grid-like data, such as images, CNNs, a particular class of deep learning models, were developed. Pooling, convolutional, and fully connected layers are among the many layers that make them up.

Tandem Facial Expression Model

An ensemble learning strategy that combines several sub-models, each of which focuses on a different aspect of facial expressions, is known as a tandem facial expression model. By simultaneously taking into account several aspects of the facial expression, this method seeks to enhance performance. Faces in

tandem Models are particularly helpful in situations where there are multiple sources of variability in facial expressions, and accurately capturing various aspects of expression is essential. They have uses in areas like affective computing, human-computer interaction, and facial expression analysis.

Two Convolutional Neural Network

The model is composed of two convolutional neural networks. The one on the left shows how the Deep ID network finds identity features. The proper deep residual network is trained using facial expression databases. Following independent training, the identity feature and the profoundly learned emotion feature are combined as the TFE features [1] and fed to the resulting fully connected layers. Finally, they restrict joint learning on the newly combined network to the facial expression database.

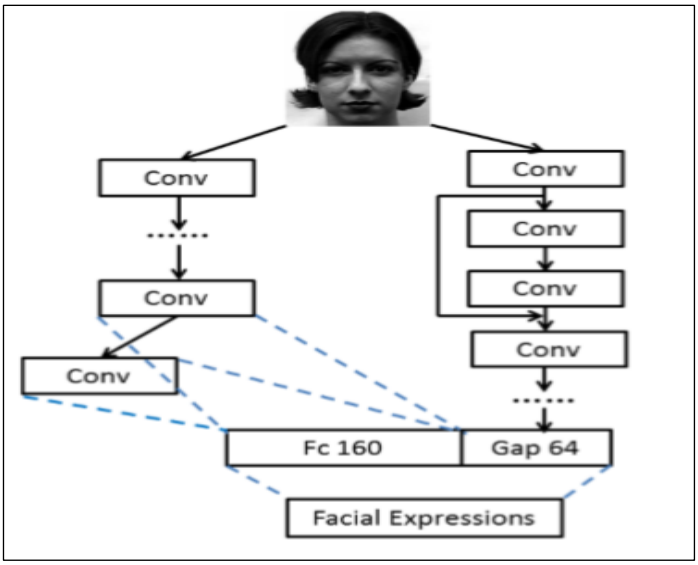

Figure 2: Two Convolutional Neural Network.
Source: [1].

II.3.4. Recommendation Techniques

Collaborative Filtering (CF)

Collaborative Filtering is a well-known recommendation method that, by compiling user preferences, automatically predicts a user's interests. It can also be divided into user-based and item-based techniques that use information about user-item interactions to produce suggestions.

Learning-Based Methods (BPR, DIN, DIEN)

Bayesian Personalised Ranking (BPR) is a technique that uses matrix factorization to optimize the ranking of items uniquely. It seeks to identify latent variables that represent user preferences. The deep learning-based recommendation model DIN (Deep Interest Network) takes users' changing interests into account while making recommendations. It models user interests using attention techniques. Deep Interest Evolution Network (DIEN) is a development of DIN, DIEN takes sequential behaviour and the development of user interests over time into account. It is intended to record persistent dependencies in user behaviour.

Multi-Interest User Representation (MUIR)

MUIR refers to methods designed to effectively represent users who have a variety of interests. It acknowledges that user interests might vary and that preferences can shift over time. Models like DIN and DIEN serve as illustrations of MUIR-based methodologies.

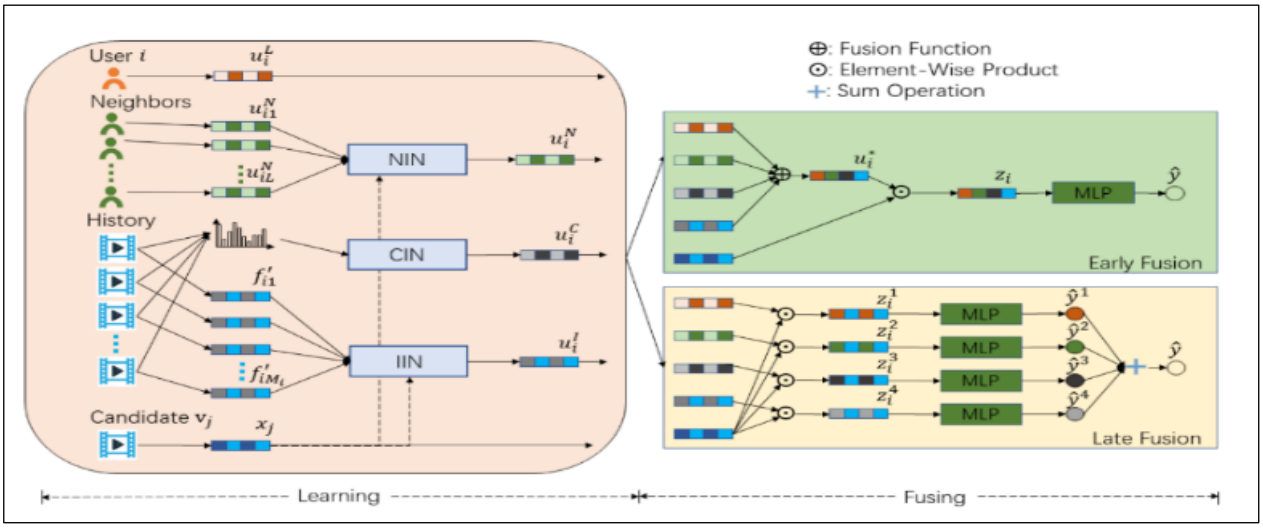

Figure 3: The framework of our Multiple User Interest Representations (MUIR) for video recommendation.
Source: [4].

Performance Metrics (AUC, Precision, Recall, NDCG)

AUC: Area Under the Receiver Operating Characteristic Curve, or AUC, is a widely used statistic for assessing how well binary classification algorithms work. It measures how well the model prioritizes positive items over bad ones in recommendation systems.

Precision: Precision estimates the proportion of items among the suggested items that are actually relevant. It emphasizes the veracity of the suggested things.
Recall: Recall quantifies the percentage of pertinent items that are effectively suggested. It highlights the coverage of important topics.
NDCG: By taking into account the position of pertinent items in the recommendation list and adjusting their importance depending on that position, NDCG (Normalised Discounted Cumulative Gain) examines the ranking quality of recommended items.

Interaction Methods for User and Item

Element-Wise Product: Multiplying the corresponding elements of the user and item representations results in interaction through element-wise products. It records interactions between features in pairs.
Concatenation: Concatenation is the combining of user and item representations, typically along a single axis. By doing this, a collective representation is produced that can be recommended.

II.3.5. Facial Expression Recognition-Enhanced Real-Time Book Recommendation System

Recommendation System

The paper also describes conventional book recommendation techniques and offers the idea of a recommendation system, specifically for books. It refers to recommendations made through collaborative filtering, knowledge-based recommendations, and content-based recommendations. It draws attention to the drawbacks of conventional approaches, including their propensity to concentrate on particular sectors and the cold start issue brought on by collaborative filtering. The research suggests a novel approach that uses facial expression recognition to learn user preferences in real time to overcome these restrictions. The real-time and authentic user data provided by this method enhances the tailoring of book recommendations while promoting human-machine connection.

Facial Emotion Recognition Algorithm

A face recognition algorithm plus an expression recognition algorithm make up the facial expression recognition system. For face detection in real-time video streams recorded by a camera, the face recognition method uses OpenCV's Haar-Adaboost algorithm [7]. Preprocessing video frames and applying image equalization parameters are involved. A trained mini-Xception convolutional neural network is used by the expression recognition algorithm to extract facial information and categorize expressions. The network picks up characteristics linked to several emotions, including sadness, happiness, fear, disgust, surprise, scare, and anger. As a result of recognition, the system chooses the expression class with the highest likelihood. Notably, the research describes the experimental procedure, which includes book suggestions based on user preferences and user preference detection by facial expression recognition. Experiments are conducted to assess the system's precision and success rates in predicting users' preferred book types and recommending books.

Convolutional Neural Networks (CNNs)

The article uses convolutional neural networks (CNNs), including the mini-Xception framework, to recognize facial expressions. To make it easier to classify emotions, CNNs are employed to extract features from input grayscale photos.

Haar-Adaboost Algorithm: The Haar-Adaboost technique, which is a part of OpenCV, is used to find faces in real-time video streams collected from a camera.
User Preference Inference: The study uses facial expression recognition to determine user preferences from expressions like "Happy" or "Surprised." This method improves book suggestions.
Recommendation Success Rate Calculation: The success rate of book recommendations is calculated in the study to assess how well the suggested recommendation system works.
Real-time User Monitoring: As users browse books, the system is actively watching their facial expressions to determine their preferences.
Emotion Categorization: For proper recognition and advice, emotions are divided into fundamental categories as Sad, Happy, Fear, Disgust, Surprise, and Angry.
Image preprocessing: Preprocessing techniques, including image equalization, are used to enhance the quality of the input photos for identifying facial expressions.

II.3.6. Knowledge-Based Recommendation System (KBRS)

This research study [5] presents a comprehensive Knowledge-Based Recommendation System (KBRS) that focuses on tracking emotional wellness and providing personalized message recommendations. The procedure is divided into several parts, starting with subjective evaluations to set the eSM2 metric parameters, which are critical for assessing users' preferences and the emotional content of utterances gathered from social media. These trials consider a range of user profiles, such as those with acute stress and mild to moderate depression levels, to develop a trustworthy sentiment meter.

The study divides phrases into depressive, stress-related, or non-depressive/non-stress categories using convolutional neural networks (CNN) and bidirectional long short-term memory recurrent neural networks (BLSTM-RNN). The emotional health monitoring system can identify users' emotional states from their social media posts thanks to these algorithms. We introduce the eSM2 sentiment analysis metric, which includes adjustment variables based on a range of user profile characteristics, including age, gender, geography, and educational attainment.

The recommendation engine in the KBRS architecture creates personalized messages for those who are stressed out or depressed using emotional data, user profiles, and ontologies [5]. The system uses several machine learning and deep learning approaches to accomplish this, making sure that users only view messages that are acceptable in terms of emotion and context. The study examines how user preferences and context could be taken into account by ontologies to improve suggestion accuracy. The study component as a whole combines sentiment analysis, machine learning, deep learning, and ontological techniques to construct a new KBRS that addresses emotional well-being through personalized messages.

•**Sentiment Analysis:** The paper employs sentiment analysis techniques to determine the emotional content of sentences extracted from social media. The eSM2 sentiment metric is introduced to quantify sentiment intensity.

•**Machine Learning:** Machine learning algorithms, including CNN and BLSTM-RNN, are utilized to classify sentences into depressive, stress-related, or non-depressive/non-stress categories.

•**Deep Learning:** Deep learning architecture, such as CNN and BLSTM-RNN, is applied for character-level representation and classification of sentences, enhancing the accuracy of emotional state detection.

•**Ontologies:** Ontological techniques are used to improve recommendation accuracy by considering user preferences, context, and other factors. The paper relies on ontologies to create a knowledge-based recommendation system.

•**User Profile Analysis:** The paper incorporates user profile analysis, considering factors like age, gender, location, and educational level, to customize recommendations and sentiment analysis.

II.4. DATASET ANALYSIS OF THE PLANNED SURVEY

MovieLens-10M

The MovieLens-10M dataset holds significant importance in the realm of movie recommendation research. In order to enhance its efficacy, researchers have incorporated 10,380 movie trailers sourced from YouTube and have utilized Convolutional Neural Networks (CNNs) to generate 4000-dimensional visual descriptions for each movie. The dataset has been split into two sections, with 80% of the dataset designated for training and the remaining 20% for testing a click-through prediction system. Ratings have been transformed into binary decisions, with a rating of 5 being labelled as "yes" and all other ratings being labelled as "no." Users with less than five "yes" ratings have been excluded from the evaluation process. In essence, the dataset has been enriched with trailers and visuals, a model has been trained to predict clicks, and the evaluation has been focused on users with substantial positive ratings.[4]

MicroVideo-1.7M

The MicroVideo-1.7M dataset is utilized for the purpose of examining interactions with brief micro-videos. It comprises a total of 12,737,619 user interactions performed by 10,986 users on 1,704,880 micro-videos. Each micro-video is characterized by a 512-number depiction based on its cover image, and is assigned to a specific category. The dataset is bifurcated into two distinct segments: one for training and the other for testing. Significantly, there are no videos that are present in both sets, rendering it appropriate for analysing scenarios where new items are introduced. This dataset is primarily employed for predicting whether users will click on micro-videos, with each interaction being labelled as either "yes" or "no."[4].

Table 2: Statistics Of Two Datasets.

	MovieLens-10M		MicroVideo-1,7M	
	Training	test	Training	test
#User	51,001	46,692	10,986	10,986
#Video	10,380	9,830	984,983	719,897
#Category	19	19	512	512
#Positive	1,216,527	268,917	1,754,457	646,933
#Negative	5,781,1953	1,398,369	7,215,852	3,120,375
Positive Ratio	0,22%	0,06%	0,02%	0,01%

Source: [4].

DEFSS

The Developmental Emotional Faces Stimulus collection (DEFSS), which contains a child, teen, and adult face, attempts to provide a standardized collection of emotional stimuli that have been validated by participants across a wide range of ages. The collection consists of 404 validated facial photos of persons aged 8 to 30 displaying five different emotional states: happy, mad, fearful, sad, and neutral. Additionally, the DEFSS contains a neutral emotion that contrasts positive and negative feelings across age groups.

LFW

The 5,749 online identities that make up the 13,233 face images in the LFW (Labelled Faces in the Wild) dataset. In order to compare performance, LFW advises using splits they randomly generated (6,000 pairs) with 10-fold cross validation [6]. The LFW database is used in this study only as a stand-alone testing dataset to evaluate the effectiveness of our identity features created using the CASIA-WebFace dataset [1].

CK+

The CK+ database contains 327 image sequences with labelled facial expressions. Each image sequence only contains an expression label in the last frame. In order to gather more images for training, they typically selected the final three frames of each sequence for training or validation purposes. Additionally, the first frame from each of the 327 labelled sequences would be chosen to represent the "neutral" expression. Thus, this dataset can provide 1308 total images with 8 labels for facial expressions. For testing, they employ the 10-fold cross validation testing protocol and the CK+ database [1].

FER+

This dataset comes from the Face Expression Recognition Challenge of the 2013 Representation Learning Workshop [8]. 28,709 training 48 x 48 face images are included. The 3,589 images that make up the test set contain a total of 7 different facial expressions: anger, disgust, fear, happiness, sadness, and surprise. Due to its noisy labels, this dataset is labelled again using services that rely on crowdsourcing [9]. In this study, a majority vote is used to determine the new set of labels for our experiments.

Figure 4: Ten representative face images whose prediction is corrected by joint learning method.
Source: [1].

III. CONCLUSIONS

By skillfully interpreting complex emotional data from facial expressions, this system introduces a powerful fusion of cutting-edge deep learning paradigms, most notably Convolutional Neural Networks (CNNs) and Convolutional Recurrent Neural Networks (CRNNs), to perform real-time emotion recognition. Facial expression recognition accuracy is improved by using pre-processing methods such as image equalization. Benchmarks like MovieLens-10M, MicroVideo-1.7M, DEFSS, LFW, CK+, and FER+ are among the carefully chosen datasets by the system, demonstrating its dedication to representative and diverse samples. It performs a thorough comparative analysis of deep learning-based models such as Deep Interest Network (DIN) and Deep Interest Evolution Network (DIEN), Collaborative Filtering (CF), and Bayesian Personalized Ranking (BPR) in the field of recommendation approaches. This analysis uses performance indicators such as AUC, Precision, Recall, and NDCG for a rigorous evaluation methodology to show how well various approaches solve the cold start issue and react to changing user interests. This novel technical paradigm seeks to build emotionally intelligent digital environments by fusing state-of-the-art deep learning with sophisticated evaluation methods and careful examination of datasets. This survey study aims to provide a thorough analysis of the relationship between emotions and technology, ultimately stimulating more investigation and creativity in this area to rethink the way we fundamentally engage with one another and with technology.

IV. AUTHOR'S CONTRIBUTION

Conceptualization: Sanika A. Gonjari, Ritika A. Pawar, Rohini B. Naik
Methodology: Sanika A. Gonjari, Rachana N. Pawar, Ritika A. Pawar, Sharwari H. Kshirsagar, Rohini B. Kokare.
Investigation: Sanika A. Gonjari, Rachana N. Pawar, Ritika A. Pawar, Sharwari H. Kshirsagar, Rohini B. Kokare.
Discussion of results: Sanika A. Gonjari, Rachana N. Pawar, Ritika A. Pawar, Sharwari H. Kshirsagar, Rohini B. Kokare.
Writing-Original draft: Rachana N. Pawar, Sharwari H. Kshirsagar.
Visualization, Writing, Editing: Sanika A. Gonjari, Ritika A. Pawar, Rachana N. Pawar, Sharwari H. Kshirsagar.
Resources:
Supervision: Rohini B. Kokare.
Approval of the final text: Rohini B. Kokare.

V. REFERENCES

[1] M. Li, H. Xu, X. Huang, Z. Song, X. Liu and X. Li, "Facial Expression Recognition with Identity and Emotion Joint Learning," in IEEE Transactions on Affective Computing, vol. 12, no. 2, pp. 544-550, 1 April-June 2021, doi:10.1109/TAFFC.2018.2880201.

[2] H. -G. Kim, G. Y. Lee and M. -S. Kim, "Dual-Function Integrated Emotion-Based Music Classification System Using Features From Physiological Signals," in IEEE Transactions on Consumer Electronics, vol. 67, no. 4, pp. 341-349, Nov. 2021, doi:10.1109/TCE.2021.3120445.

[3] A. V. Savchenko, L. V. Savchenko and I. Makarov, "Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network," in IEEE Transactions on Affective Computing, vol. 13, no. 4, pp. 2132-2143, 1 Oct. Dec.2022,doi:10.1109/TAFFC.2022.3188390.

[4] X. Chen, D. Liu, Z. Xiong and Z. -J. Zha, "Learning and Fusing Multiple User Interest Representations for Micro-Video and Movie Recommendations," in IEEE Transactions on Multimedia, vol. 23, pp. 484496, 2021, doi:10.1109/TMM.2020.2978618.

[5] R. L. Rosa, G. M. Schwartz, W. V. Ruggiero and D. Z. Rodr´ıguez, "A Knowledge-Based Recommendation System That Includes Sentiment Analysis and Deep Learning," in IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2124-2135, April 2019, doi:10.1109/TII.2018.2867174.

[6] C. Dalvi, M. Rathod, S. Patil, S. Gite and K. Kotecha, "A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions," in IEEE Access, vol. 9, pp. 165806-165840, 2021, doi: 10.1109/ACCESS.2021.3131733

[7] Y. Zhao and J. Zeng, "Library Intelligent Book Recommendation System Using Facial Expression Recognition," 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI), Kitakyushu, Japan, 2020, pp. 55-58, doi: 10.1109/IIAI-AAI50415.2020.00021.

[8] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M.Mirza, B. Hamner, W. Cukierski,Y. Tang, D. Thaler, and D. H. Lee, "Challenges in representation learning: A report on three machine learning contests," Neural Networks, vol. 64, p. 59, 2015.

[9] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in ACM International Conference on Multimodal Interaction, 2016, pp. 279–283.