LITERATURE REVIEW

MACHINE LEARNING BASED EMOTIONAL DETECTOR –

SULTHANA SAFEER

M00849556

SS4343@LIVE.MDX.AC.UK

ABSTRACT

The pursuit of developing advanced technologies for understanding human emotions has led to the emergence of mood detectors which is done with the help of machine learning and artificial intelligence. This review provides a brief analysis of ML based mood detection with chatbot, emphases their methodologies, applications, and challenges. The primary goal is distilling insights from a diverse body of literature, exploring the advancements, trends, and potential future directions in this evolving field.

The literature review systematically categorizes studies based on key themes such as ML Algorithms, data sources, evaluation metrics and application domains. It delves into the methodologies employed, highlighting the variety of approaches used in detecting moods from different modalities, including facial expressions. The review critically evaluates the performance metrics and applications domains. The review critically evaluates the performance metrics utilized to assess the accuracy and effectiveness of these models.

Facial expressions are mirrors of human thoughts and feelings.

Beyond methodological considerations, the review explores the diverse application domains of ML and AI based mood detectors, ranging from healthcare to security human computer interaction. Through a comparative analysis of studies, it identifies common challenges faced by researchers, including data security, cross- model integration and cultural variations. The limitations and gaps in current literature serve as a foundation for proposing future research directions and improvements to current methodologies.

Finally, this study provides a comprehensive overview of the current state of ML and AL-based mood detectors. By combining findings from multiple investigations, this work hopes to add to collective knowledge in the field and encourage future research focused at improving the capabilities and uses of mood detection systems.

Review of Literature.

Introduction: -

People perceive human facial emotions all around them. They are natural signs that assist them in understanding the emotions of everyone in front of them, as well as photographs or videos. These emotions are extremely complicated and difficult for machines to understand, but they are easily understood by people. Mehrabian, a well-known psychologist, discovered via his research that the emotional data that individuals define as emotions is divided into portions. He discovered that language transmits only 7% of overall emotional data, while 38% is transported by our language auxiliary, which varies by culture and includes things like speech rhythm, tone, pitch, and so on. [1]

By measuring these feelings, one may provide maximum user happiness and feedback to help enhance present technologies. This is only possible in the fields of computer vision and deep learning. To develop multiple Facial Emotion Recognition (FER) systems that have been tested for encoding and conveying information from facial images.

Emotions are based on nothing, but mind state or phase perceived by a human, and it’s mostly associated with brain chemicals and neurons, these emotions are often coming with many types depending on Attitude, temper, character, disposition, and motivation. They can also be defined into two different.

Many researchers employ their own set of emotion definitions and assumptions. This complicates the study of human face emotions because all previous studies have significant variation and do not reach a general result. Although all humans have naturally occurred sets of emotions that can be recognized across cultures, the Discrete Emotion Theory states that such feelings are identifiable by an individual's traits. [2]

Different datasets conduct research on various combinations of emotions. For example, very few children's datasets include 'upset' emotions. Those who used it posed while recording. It's tough to capture angry emotions through spontaneous expressions. However, in adult datasets, it is simple to posture for an angry mood. Datasets such as 'RML' have captured emotions in a controlled environment with good lighting conditions. Most datasets consist of movie or television clips that are clipped and used for classification. The classification of emotions varies from dataset to dataset.

FER so called Facial emotion recognition plays a role in computer vision case studies. if we take emotion detection under consideration it, we can separate it into or divide the study into 5,

Database, Deep FER, Classification, Performance Metrics, Applications.

Database is considered by the data intake you’re putting into the machine. Since its facial emotion recognition, we are taking humans facial expressions as the database. Humans can be categorized into 3 types, senior citizens, adults, and kids.

Deep facial emotion recognition usually entails several processes, which, as you indicated, can be divided into pre-processing and feature extraction. Let me clarify each of the following stages:

Pre-processing: - Face Detection: Recognize and locate faces in input images or video frames. This step is critical for isolating the facial region for future study.

Face Alignment: Normalize the detected faces' positions and orientations. This guarantees that facial features are organized in a consistent manner, which is critical for precise analysis.

Image Pre-processing: Improve the quality of the facial image by using techniques like normalization, scaling, and filtering. This step improves the performance of the following processing stages.

Feature Extraction:

Facial Landmarks: Identify major facial features such as the eyes, nose, and mouth. These landmarks serve as reference points for subsequent investigation.

Expression Features: Extract face expression-related elements such muscle movement, wrinkles, and other visual clues. Deep learning models, such Convolutional Neural Networks (CNNs), are frequently used to automatically learn and extract expressive characteristics.

Temporal Analysis: Analyze image or video frame sequences to investigate the temporal component of facial expressions. This is especially useful for detecting dynamic changes in emotions over time.

Deep learning models: Use deep neural networks, such as CNNs or Recurrent Neural Networks (RNNs), to automatically learn hierarchical representations of facial traits.

In the context of machine learning (ML)-based emotional detectors, classifications can broadly be categorized into two main approaches: traditional machine learning and deep learning. Let's explore each of these categories:

Traditional Machine Learning:

Feature Engineering: In traditional ML, feature engineering plays a crucial role. Features are manually crafted based on domain knowledge or through extensive analysis of the data. These features could include aspects like facial landmarks, color histograms, texture patterns, or any other relevant information that might contribute to emotional understanding.

Classifiers: Once features are extracted, traditional machine learning algorithms like Support Vector Machines (SVM), Decision Trees, Random Forests, or others can be employed. These algorithms learn to map the extracted features to specific emotional classes based on a training dataset.

Advantages: Traditional ML approaches are often interpretable, and they might require less computational resources compared to deep learning. However, their performance might be limited when dealing with complex and high-dimensional data like images.

Deep Learning:

End-to-End Learning: Deep learning, especially Convolutional Neural Networks (CNNs) and recurrent architectures, allows for end-to-end learning. This means that the model learns hierarchical representations directly from the raw input data, eliminating the need for extensive feature engineering.

Automated Feature Learning: Deep learning models automatically discover and learn relevant features from the data during the training process. This is particularly beneficial when dealing with complex

data like images or sequences, as the model can capture intricate patterns that may be challenging to define manually.

Neural Networks: Deep learning models in this context often involve neural networks with multiple layers, enabling the extraction of hierarchical and abstract features, which can be crucial for understanding subtle nuances in facial expressions and emotions.

Advantages: Deep learning models can achieve state-of-the-art performance in tasks like facial emotion detection, especially when large amounts of labelled data are available. However, they may require more computational resources and data compared to traditional ML approaches.

Performance metrics are essential for evaluating the effectiveness of emotional detectors or any machine learning model. The choice of metrics depends on the specific nature of the task and the characteristics of the dataset. When it comes to performance metrics, there will be accuracy, F1 score, RMSE, Loss and confusion matrix. We solve and find performance metrics using formulas for each of them respectively.

Under Applications, we can take a minimum of 3 as of now i.e., Robotics, automative safety and gaming.

Certainly! Emotional detectors in the context of machine learning find applications in various fields. Here are explanations for three specific applications: robotics, automotive safety, and gaming.

1. Robotics:

Emotional detectors can be integrated into robotics to enable robots to perceive and respond to human emotions. This enhances human-robot interaction by allowing robots to adapt their behavior based on the emotional state of the users. For example, a robot may modify its responses or actions when detecting signs of frustration, happiness, or anxiety in a human interacting with it.

Assistive Robotics: Robots assisting individuals with special needs can adapt their assistance based on the user's emotional state, providing a more personalized and supportive interaction.

Customer Service Robots: Robots in service industries can better understand and respond to customer emotions, improving the overall service experience.

2. Automotive Safety:

Emotional detectors in automotive safety systems can contribute to safer driving experiences. By monitoring the emotional state of the driver, the vehicle can implement safety measures or alerts to prevent accidents caused by distracted or emotionally distressed drivers.

3. Gaming:

Emotional detectors can significantly enhance the gaming experience by enabling games to respond dynamically to the player's emotions. This creates a more immersive and personalized gaming environment where the game adapts to the player's emotional reactions.

These applications demonstrate the versatility of emotional detectors, showcasing their potential to enhance human-machine interactions, improve safety in various domains, and create more engaging and personalized experiences in entertainment. As technology continues to advance, emotional

detection systems are likely to find even more diverse and innovative applications across different industries. [3]

Research on FER has explored various techniques, modalities, SOTA models, and feature combinations to improve accuracy, providing insights for future growth [4,5]. To ensure detailed information for future research, it's important to categorize and thoroughly review the vast amount of work. To better comprehend the work, the writers analyzed multiple surveys and compared. The authors offer a survey that addresses research gaps, problems, and potential answers, distinguishing it from previous investigations.

Furthermore, there is a shortage of comprehensive coverage of cutting-edge models, methodologies, and comparative studies across several categories. The authors propose a new survey for children and adults that compares approaches, state-of-the-art models, and emerging trends to advance FER and Multimodal Emotion Recognition research.

Deep Facial Emotion Recognition

The authors provided a detailed description of the stages required for FER. Various strategies can be used for each phase based on the specific situation.

The authors will discuss pre-processing, feature extraction, and various state-of-the-art models in detail, as shown in Figure 13. This section compares and provides insights for current and future FER research.

A. PRE-PROCESSING.

During this stage, authors clean up the dataset by removing noise, compressing it, and removing unnecessary data. The pre-processing of picture or video frames includes the following stages:

Face detection: It identifies faces in photographs and pictures. Face detection is a subset of object class detection that determines the presence of a face in an image.

Normalization is also known as feature scaling. Following this stage, picture characteristics are reduced and normalized while maintaining the identifiable spectrum of feature values. Normalization techniques include Z Normalization, Min-Max Normalization, and Unit Vector Normalization. These methods improve numerical consistency and model development.

Data Augmentation is generating new data from an image using various transformations while keeping the face data intact, allowing for better handling of limited data.

b: The HAAR classifier.

Haar characteristics are often quantified by lowering pixel sizes in groups. Haar Classifier uses Haar-like properties to recognize images. This approach can detect objects of different sizes [6]. Haar classifiers identify the most effective features for face detection during training. It may suggest good detection accuracy with little computation complexity.

c: Adaptive Skin Color.

The adaptive skin-color model detects facial regions using a skin-color model [7]. This method achieves great accuracy by segmenting images based on skin tone. Distinguishing between the face and non-facial regions is straightforward. The sole limitation is that this algorithm is not compatible with varying levels of illumination. While adaptive gamma correction methods can prevent this issue, their considerable processing power makes them unsuitable for real-time application.

d. ADABOOST CONTOUR POINTS

Adaboost's minimal computing power makes it ideal for real-time settings [8]. This approach allows for cascading multiple classifiers. First, it trained the faces and created a robust classifier with high accuracy across all photos.

Next, the new Face is compared to the classifier-built model. Additionally, contour points are used to detect faces. Contour points can improve accuracy and performance by extracting minimal features at the end, resulting in a simpler model.

f: MTCNN.

Multi-task Cascaded Convolutional Networks (MTCNN) is a popular framework for handling computer vision issues, including face detection and alignment [9]. The method involves three critical stages. A convolutional network recognizes facial features and landmarks like eyes, nose, and mouth. There are three steps to MTCNN.

• P-Net- for proposals

• R-Net- for refinements

• O-Net- for outputs

In the initial stage, a shallow CNN is utilized to generate candidate windows. In the second step, a complicated CNN refines the result. Finally, in the third stage, a more complex CNN refines the output and plots facial landmark positions appropriately.

Choosing the right facial detection algorithm for a specific application can be challenging due to the numerous options available. The authors emphasize the importance of choosing the optimal algorithm for real-time applications to ensure efficient processing and data collection.

Data augmentation

Image Data Augmentation is commonly used to enhance detection accuracy for specific jobs. This method [10] is commonly used to achieve accurate results while training deep neural networks. FER training datasets lack sufficient pictures, necessitating data augmentation to achieve optimal accuracy levels. This is similar to the concept of enriching existing training datasets, as explained in.

B. Feature Extraction

1) TEXTUAL FEATURES

The following descriptors use texture-based feature extraction techniques. The Gabor filter, which combines phase and magnitude information, is a widely used texture descriptor for feature extraction. The Gabor filter uses the magnitude feature to restrict information regarding facial image organization [11-15]. LBP features are commonly formed using binary code and can be achieved by thresholding between the center and nearby pixels [16,17]. LBP with Three Orthogonal Planes (TOP) features can be recovered using multi-resolution approaches, as demonstrated in [18]. It can extract non-dynamic appearances from static face pictures utilizing features [19].

GEOMETRIC FEATURES

Geometric feature learning involves extracting discrete geometric elements from photographs. Geometric aspects are simple things composed of geometric elements including lines, points, curves, and surfaces. Corners are an essential feature of items. Complex objects often have unique corner features that set them apart.

Corner detection is a technique for extracting an object's corners [20]. Corners were defined uniquely using the distance and angle between two straight line segments.

FEATURE LEARNING THROUGH DEEP NETWORKS

Deep learning is a popular study topic in computer vision, with impressive results in picture categorization utilizing classification methods [21]. Deep learning employs hierarchical nonlinear transformations and representations to capture high-level concepts. The authors provide a brief overview of image/video-based approaches for recognizing emotions. Recent years have seen the use of four conventional approaches in FER. The approaches include Deep Believe Network, CNN, Deep Autoencoder, and Recurrent Neural Network.

Let's talk more about CNN and RNN

CNN is an enhancement over Artificial Neural Network (ANN) [22]. CNN has many different applications. A study [126] found that applying neurons with similar parameters to patches of the previous layer at different areas results in translational invariance. This is one of the primary computational models based on neural networks and coordinated image changes.

Recurrent neural network (RNN)

RNN models contain temporal information and are better suited for predicting sequential data with flexible durations. RNNs use recurrent edges to train a deep neural network in a single feed-forward process, sharing the same parameters across all phases. The RNN is built using the BPTT method [158] and has a chain-like structure with four repeating modules. Long-short-term memory (LSTM) is a standard RNN used to detect inclination fading and blast problems in RNNs. Three gates manage the cell state in LSTM (as shown in [23].

There will be challenges and open issues if we think into consideration. FER has become a competitive subject in recent years.

Numerous research has successfully recognized emotions using facial expression recognition analysis. However, certain issues and concerns must be addressed. This section discusses the obstacles and issues that FER has experienced. The authors analyzed survey papers to identify issues and provide possible remedies.

Occlusion is the most prevalent stumbling block in FER.

The authors discovered that existing research, including JAFFE and CK+ datasets without occlusion, is already publicly available.

Several datasets lack natural facial occlusion. There is a necessity to construct datasets with occlusion.

Although time-consuming and difficult, it is a necessary evil. FER datasets require manual occlusion. There has been insufficient training, and testing in occluded datasets remains a substantial challenge [24]. However, gathering spontaneous emotion information under occlusion might be challenging. The selection of the obstructed location, occlusion level, kind, and preparation materials are crucial.

Conclusion

The review aims to provide a comprehensive for FER, a topic of growing interest among scholars trying to provide all significant components of FER. The authors provided a brief overview of current FER approaches and models for several dataset categories. This paper reviewed existing surveys in FER, identifying gaps and addressing shortcomings. FER statistics are divided into three categories: kids, adults, and senior citizens to provide a comprehensive understanding of its reach. Creating a new database for children is a priority due to a lack of well-balanced datasets.

REFERENCE: -

[1] A. Mehrabian and S. R. Ferris, ‘‘Inference of attitudes from nonverbal communication in two channels,’’ J. Consulting Psychol., vol. 31, no. 3, pp. 248–252, 1967, doi: 10.1037/h0024648.

[2] E. Hudlicka, ‘‘Computational modeling of cognition–emotion interac tions: Theoretical and practical relevance for behavioral healthcare,’’ in Emotions and Affect in Human Factors and Human-Computer Interaction. New York, NY, USA: Academic, 2017, pp. 383–436, doi: 10.1016/B978-0-12-801851-4.00016-1

[3] 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,” *IEEE Access*, vol. 9, pp. 165806–165840, 2021, doi: https://doi.org/10.1109/access.2021.3131733.

[4] P. J. Bota, C. Wang, A. L. N. Fred, and H. P. Da Silva, ‘‘A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals,’’ IEEE Access, vol. 7, pp. 140990–141020, 2019, doi: 10.1109/ACCESS.2019.2944001.

[5] T. Kundu and C. Saravanan, ‘‘Advancements and recent trends in emotion recognition using facial image analysis and machine learn ing models,’’ in Proc. Int. Conf. Electr., Electron., Commun., Com put., Optim. Techn. (ICEECCOT), Dec. 2017, pp. 1–6, doi: 10.1109

[6] L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, ‘‘Human face detection algorithm via Haar cascade classifier combined with three additional classifiers,’’ in Proc. 13th IEEE Int. Conf. Electron. Meas. Instrum. (ICEMI), Oct. 2017, pp. 483–487, doi: 10.1109/ICEMI.2017.8265863.

[7] C.-S. Wang, Y.-C. Jeung, L.-B. Luo, J. Wang, and J.-W. Chong, ‘‘Real-time face recognition using adaptive skin-color model,’’ in Proc. Int. Conf. Inf. Sci. Appl. (ICISA), Apr. 2011, pp. 1–6, doi: 10.1109/ICISA.2011.5772343.

[8] K. S. Kumar, S. Prasad, V. B. Semwal, and R. C. Tripathi, ‘‘Real time face recognition using AdaBoost improved fast PCA algorithm,’’ Int. J. Artif. Intell. Appl., vol. 2, no. 3, pp. 45–58, Jul. 2011, doi: 10.5121/IJAIA.2011.2305.

[9] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, ‘‘Joint face detection and alignment using multi-task cascaded convolutional networks,’’ IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342

[10] C. Shorten and T. M. Khoshgoftaar, ‘‘A survey on image data augmentation for deep learning,’’ J. Big Data, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/S40537-019-0197-0.

[11] S. Bashyal and G. K. Venayagamoorthy, ‘‘Recognition of facial expressions using Gabor wavelets and learning vector quantization,’’ Eng. Appl. Artif. Intell., vol. 21, no. 7, pp. 1056–1064, Oct. 2008, doi: 10.1016/J.ENGAPPAI.2007.11.010.

[12] E. Owusu, Y. Zhan, and Q. R. Mao, ‘‘A neural-AdaBoost based facial expression recognition system,’’ Expert Syst. Appl., vol. 41, no. 7, pp. 3383–3390, Jun. 2014, doi: 10.1016/J.ESWA.2013.11.041.

[13] L. Zhang, D. Tjondronegoro, and V. Chandran, ‘‘Random Gabor based templates for facial expression recognition in images with facial occlusion,’’ Neurocomputing, vol. 145, pp. 451–464, Dec. 2014, doi: 10.1016/J.NEUCOM.2014.05.008.

[14] A. Hernandez-Matamoros, A. Bonarini, E. Escamilla-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, ‘‘A facial expression recognition with automatic segmentation of face regions,’’ in Proc. Int. Conf. Intell. Softw. Methodol., Tools, Techn., in Communications in Computer and Information Science, vol. 532, 2015, pp. 529–540, doi: 10.1007/978-3-319-22689-7_41.

[15] G. P. Hegde, M. Seetha, and N. Hegde, ‘‘Kernel locality pre serving symmetrical weighted Fisher discriminant analysis based subspace approach for expression recognition,’’ Eng. Sci. Technol., Int. J., vol. 19, no. 3, pp. 1321–1333, Sep. 2016, doi: 10.1016/ J.JESTCH.2016.03.005

[16] S. L. Happy and A. Routray, ‘‘Automatic facial expression recogni tion using features of salient facial patches,’’ IEEE Trans. Affective Comput., vol. 6, no. 1, pp. 1–12, Jan. 2015, doi: 10.1109/TAFFC.2014. 2386334.

[17] M. J. Cossetin, J. C. Nievola, and A. L. Koerich, ‘‘Facial expression recog nition using a pairwise feature selection and classification approach,’’ in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2016, pp. 5149–5155, doi: 10.1109/IJCNN.2016.7727879.

[18] G. Zhao and M. Pietikäinen, ‘‘Boosted multi-resolution spatiotemporal descriptors for facial expression recognition,’’ Pattern Recognit. Lett., vol. 30, no. 12, pp. 1117–1127, Sep. 2009, doi: 10.1016/J.PATREC.2009.03.018.

[19] Y. Ji and K. Idrissi, ‘‘Automatic facial expression recognition based on spatiotemporal descriptors,’’ Pattern Recognit. Lett., vol. 33, no. 10, pp. 1373–1380, Jul. 2012, doi: 10.1016/J.PATREC.2012.03.006.

[20] J. Wang and W. Zhang, ‘‘A survey of corner detection methods,’’ in Proc. ICEEA, vol. 139, 2018, pp. 214–219, doi: 10.2991/iceea-18.2018.47.

[21] M. Xin and Y. Wang, ‘‘Research on image classification model based on deep convolution neural network,’’ EURASIP J. Image Video Process., vol. 2019, no. 1, pp. 1–11, Feb. 2019, doi: 10.1186/S13640-019-0417-8.

[22] R. Walecki, O. Rudovic, V. Pavlovic, B. Schuller, and M. Pantic, ‘‘Deep structured learning for facial action unit intensity estimation,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 5709–5718, doi: 10.1109/CVPR.2017.605.

[23] P. J. Werbos, ‘‘Backpropagation through time: What it does and how to do it,’’ Proc. IEEE, vol. 78, no. 10, pp. 1550–1560, Oct. 1990, doi: 10.1109/5.58337

[24] L. Zhang, B. Verma, D. Tjondronegoro, and V. Chandran, ‘‘Facial expression analysis under partial occlusion,’’ ACM Comput. Surv., vol. 51, no. 2, pp. 1–49, Jun. 2018, doi: 10.1145/3158369.