IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13587

# VIDEO BASED EMOTION DETECTION USING DEEP LEARNING

**Kamini N. Ahire[1], Kartik J. Mohol[2], Vidhi G. Divekar[3], Pratham Pawar[4], Eknath Raut[5]**

Department of Computer Engineering Universal College of Engineering and Research, Sasewadi, Pune[1-5]

**Abstract:** Social networking platforms have become an essential means for communicating feelings to the entire world due to rapid expansion in the Internet era. Several people use textual content, pictures, audio, and video to express their feelings or viewpoints. Text communication via Web-based networking media, on the other hand, is somewhat overwhelming. Every second, a massive amount of unstructured data is generated on the Internet due to social media platforms. Video emotion analysis is one of the hottest topics in the video understanding community to cognize the emotion in videos for affective computing, video recommendation, and so on. Currently, many studies tend to employ different deep structures to model video contents for this task. In fact, audiences responses (e.g., physiological signals and comments) are also important since they are directly related to video emotional content and can reflect the emotions in videos.

## I. INTRODUCTION

Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us selfdriving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it. Many researchers also think it is the best way to make progress towards human-level AI. Machine Learning algorithms are the programs that can learn the hidden patterns from the data, predict the output, and improve the performance from experiences on their own. By doing a reasonable analysis of machine learning algorithms, it can provide direction reference for subsequent machine learning development, thereby improving the applicability of machine learning algorithms and providing more convenience for the economic development of the industry.

Video-based emotion recognition using deep learning is a Technology that involves using deep neural networks and machine learning techniques to analyse and understand human emotions expressed in videos. This process aims to automatically detect and classify emotions based on the video input.

• **Necessity:**

I. HCI
II. Robotics
III. Interviews

- Background: Deep learning is a vast area of research in recent time world and its applications are very worldwide.
- **Usage** :

I. Feature Extraction
II. Model Training
III. Real time Analysis

- **Application :**

I. Healthcare
II. Education
III. Security

## II. WHAT IS EMOTION DETECTION?

Emotions are an inseparable component of human life. These emotions influence human decision-making and help us communicate to the world in a better way. Emotion detection, also known as emotion recognition, is the process of identifying a person's various feelings or emotions (for example, joy, sadness, or fury). Researchers have been working hard to automate emotion recognition for the past few years.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

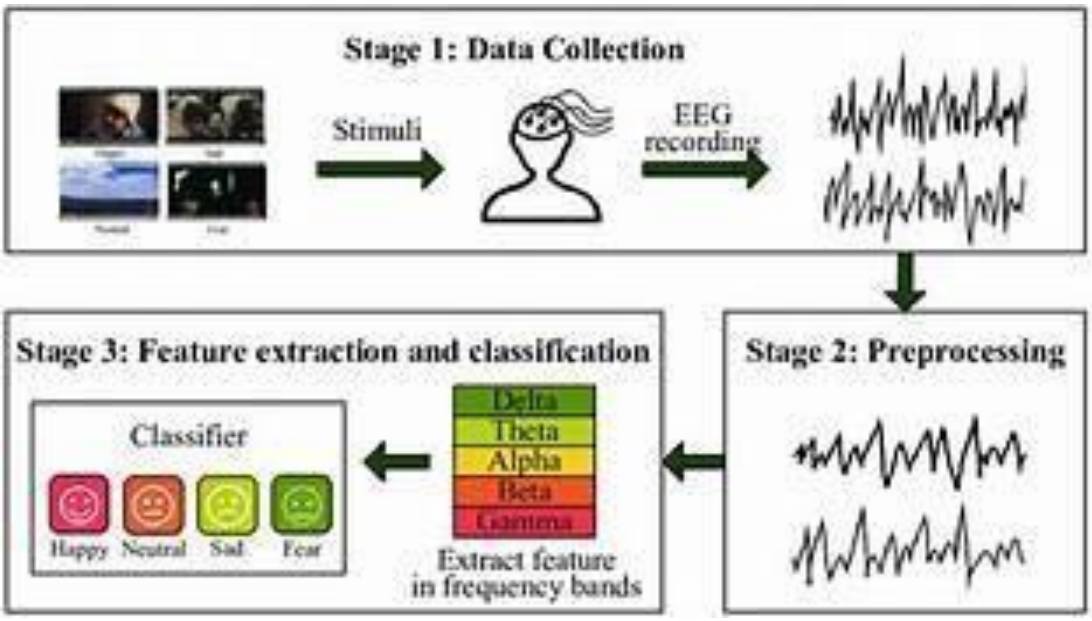
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13587

However, some physical activities such as heart rate, shivering of hands, sweating, and voice pitch also convey a person's emotional state but emotion detection from text is quite hard. In addition, various ambiguities and new slang or terminologies being introduced with each passing day make emotion detection from text more challenging. Furthermore, emotion detection is not just restricted to identifying the primary psychological conditions (happy, sad, anger); instead, it tends to reach up to 6-scale or 8-scale depending on the emotion model.

**Emotion models/emotion theories**

In English, the word 'emotion' came into existence in the seventeenth century, derived from the French word 'emotion, meaning a physical disturbance. Before the nineteenth century, passion, appetite, and affections were categorized as mental states. In the nineteenth century, the word 'emotion' was considered a psychological term (Dixon 2012). In psychology, complex states of feeling lead to a change in thoughts, actions, behaviour, and personality referred to as emotions. Broadly, psychological or emotion models are classified into two categories: dimensional and categorical.




## III. WHAT IS DEEP LEARNING

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It's achieving results that were not possible before.

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labelled data and neural network architectures that contain many layers.

**How deep learning works:**

Most deep learning methods use neural network architectures, which is why deep learning models are often referred to as deep neural networks.

The term "deep" usually refers to the number of hidden layers in the neural network. Traditional neural networks (4:37) only contain 2-3 hidden layers, while deep networks can have as many as 150.

Deep learning models are trained by using large sets of labelled data and neural network architectures that learn features directly from the data without the need for manual feature extraction. One of the most popular types of deep neural networks is known as convolutional neural networks (CNN or ConvNet). A CNN convolves learned features with input data, and uses 2D convolutional layers, making this architecture well suited to processing 2D data, such as images.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024
DOI: 10.17148/IJARCCE.2024.13587

CNNs eliminate the need for manual feature extraction, so you do not need to identify features used to classify images. The CNN works by extracting features directly from images. The relevant features are not pretrained; they are learned while the network trains on a collection of images. This automated feature extraction makes deep learning models highly accurate for computer vision tasks such as object classification.

Examples of Deep Learning at Work
Deep learning applications are used in industries from automated driving to medical devices.

**Automated Driving:** Automotive researchers are using deep learning to automatically detect objects such as stop signs and traffic lights. In addition, deep learning is used to detect pedestrians, which helps decrease accidents.

**Aerospace and Defense:** Deep learning is used to identify objects from satellites that locate areas of interest, and identify safe or unsafe zones for troops.

**Medical Research:** Cancer researchers are using deep learning to automatically detect cancer cells. Teams at UCLA built an advanced microscope that yields a high-dimensional data set used to train a deep learning application to accurately identify cancer cells.

**Industrial Automation:** Deep learning is helping to improve worker safety around heavy machinery by automatically detecting when people or objects are within an unsafe distance of machines.

**Electronics:** Deep learning is being used in automated hearing and speech translation. For example, home assistance devices that respond to your voice and know your preferences are powered by deep learning applications.

| Sr. No | Name | Accuracy-AffectNet Dataset | Result | Number of Emotion detected |
|---|---|---|---|---|
| 1 | FER in the wild via Deep attentive Center loss(2021) | 65% | DACL outperformed baseline methods | Neu, hap, sad, ang,fear,disg |
| 2 | Efficient feature learning with Wilde Ensemblebased CNN (2020) | 59.3% | Reached human level FER on AffectNet but accuracy was less. | Ne, ha, Sa, Fe,Dis,Ang, Con |
| 3 | EMOCA: Emotion driven monocular face capture and animation | 69% | EMOCA outperforms existing 3DMM based face reconstruction . | Ne, ha, Sa, Fe,Dis,Ang, Con |
| 4 | Relative uncertainty learning for FER | 56.65+/- 0.13 | Help FER model achieve state-of-art performance. | Ha, Sad, Fe, Dis, neu |

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13587

| Paper Name | Dataset Name | Sample / Videos | Type | Accuracy | Link |
|---|---|---|---|---|---|
| Video-based Emotion Recognition using Aggregated Features and Spatio-temporal Information | FER2013 | 35889 images split into 25889 for training, 5000 for validation, and 5000 for the test | Images | Accuracy of 48.01% on the validation set for emotion recognition. | https://www.kaggle.com/datasets/msambare/fer2013 |
| Automatic Emotion Detection as a Teaching Aid in Online Knowledge Assessment(2021) | AffectNet | 0.4 million images manually labeled for the presence of eight (neutral, happy, angry, sad, fear, surprise, disgust, contempt) facial expressions | Images | 67.70% | https://paperswithcode.com/dataset/affectnet |

| Paper Name | Dataset Name | Sample / Videos | Type | Accuracy | Link |
|---|---|---|---|---|---|
| Analysing Affective Behavior in the First ABAW 2020 Competition | Aff-Wild2 | 564 videos of around 2.8M frames with 554 subjects (326 of which are male and 228 female) | Videos | 62.34% | https://ibug.doc.ic.ac.uk/resourc es/aff-wild2/ |
| Emotion Recognition on large video dataset based on Convolutional Feature Extractor and Recurrent Neural Network(2020) | RECOLA (Remote COLaborative and Affective interactions) | 46 Recordings. The database consists of 9.5 hours of audio, visual, and physiological recordings | Audio / Videos | 70% | 1) https://diuf.unifr.ch/main/diva/r ecola/ (2) https://www.researchgate.net/publication/261121552_Introducing_the_RECOLA_multimodal_corpus_of_remote_collaborative_and_affective_interactions |
| Real time emotion recognition in video stream, using BCNN and F-CNN | CK+ (Extended CohnKanade) | 593 videos each of 30 frames per second (FPS) | Videos | 58.14% | https://paperswithcode.com/data set/ck |

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13587

## IV. MODELS USED IN EMOTION DETECTION

**What is K-Nearest Neighbors Algorithm?**

K-Nearest Neighbours is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining, and intrusion detection. It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data). We are given some prior data (also called training data), which classifies coordinates into groups identified by an attribute.

**Advantages of the KNN Algorithm:**

1. **Easy to implement** as the complexity of the algorithm is not that high.

2. **Adapts Easily** – As per the working of the KNN algorithm it stores all the data in memory storage and hence whenever a new example or data point is added then the algorithm adjusts itself as per that new example and has its contribution to the future predictions as well.

3. **Few Hyper parameters** – The only parameters which are required in the training of a KNN algorithm are the value of k and the choice of the distance metric which we would like to choose from our evaluation metric.

**Disadvantages of the KNN Algorithm**

1. **Does not scale** – As we have heard about this that the KNN algorithm is also considered a Lazy Algorithm. The main significance of this term is that this takes lots of computing power as well as data storage. This makes this algorithm both time-consuming and resource exhausting.

2. **Curse of Dimensionality** – There is a term known as the peaking phenomenon according to this the KNN algorithm is affected by the curse of dimensionality which implies the algorithm faces a hard time classifying the data points properly when the dimensionality is too high.

3. **Prone to Overfitting** – As the algorithm is affected due to the curse of dimensionality it is prone to the problem of overfitting as well. Hence generally feature selection as well dimensionality education techniques are applied to deal with this problem.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13587

- **Classification  Regression**

Examples of some popular supervised learning algorithms are Simple Linear regression, Decision Tree, Logistic Regression, KNN algorithm, etc.

**What is Recurrent Neural Network (RNN)?**

Recurrent Neural Network (RNN) is a type of Neural Network where the output from the previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a Hidden Layer. The main and most important feature of RNN is its Hidden state, which remembers some information about a sequence. The state is also referred to as Memory State since it remembers the previous input to the network. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks.

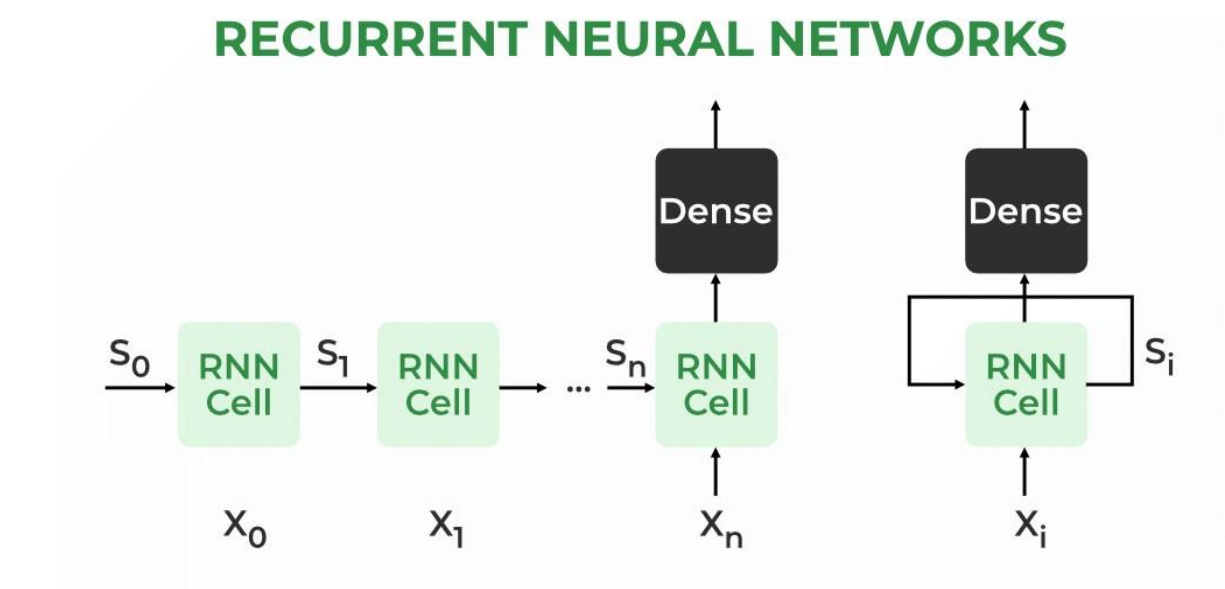
- **Architecture of Recurrent Neural Network:**

RNNs have the same input and output architecture as any other deep neural architecture. However, differences arise in the way information flows from input to output. Unlike Deep neural networks where we have different weight matrices for each Dense network in RNN, the weight across the network remains the same. It calculates state hidden state $H_i$ for every input $X_i$. By using the following formulas: $h= \sigma(UX + Wh_{-1} + B)$

$Y = O(Vh + C)$ Hence

$Y = f(X, h, W, U, V, B, C)$

Here S is the State matrix which has element si as the state of the network at timestep i
The parameters in the network are W, U, V, c, b which are shared across timestep




**Advantages of Recurrent Neural Network:**

1. An RNN remembers each and every piece of information through time. It is useful in time series prediction only because of the feature to remember previous inputs as well. This is called Long Short Term Memory.
2. Recurrent neural networks are even used with convolutional layers to extend the effective pixel neighborhood.

**Disadvantages of Recurrent Neural Network:**

1. Gradient vanishing and exploding problems.
2. Training an RNN is a very difficult task.
3. It cannot process very long sequences if using tanh or relu as an activation function.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024
DOI: 10.17148/IJARCCE.2024.13587

**Applications of Recurrent Neural Network:**

1. Language Modelling and Generating Text
2. Speech Recognition
3. Machine Translation
4. Image Recognition, Face detection
5. Time series Forecasting

## V. CONCLUSION

- In conclusion, our project on emotion detection using video showcases the exciting possibilities of emotional intelligence. We identified key challenges, such as the need for video datasets, according to that we are creating our own dataset, the complexities of feature engineering from video frames, and the computational demands associated with real-time processing.

- Till now we have done with literature survey of 18 papers and Also train and test the data by using images. As well as we have working on our own dataset for Emotion recognition system.

- As we continue our work, we remain committed to better understand and respond to the human emotions, ultimately creating more empathetic and responsive systems which will beneficial

## REFERENCES

[1]. Handrich, Sebastian; Dinges, Laslo; Saxen, Frerk; Al-Hamadi, Ayoub; Wachmuth, Sven (2019). [IEEE 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) - Kuala Lumpur, Malaysia (2019.9.17-2019.9.19)] 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) - Simultaneous Prediction of Valence / Arousal and Emotion Categories in Real-time. , (), 176–180. doi:10.1109/ICSIPA45851.2019.8977743Stevens, K. L. (2004, December 6). Online Marketing. Paper presented at the 7th Annual Conference on Business Management: Industry Trends. doi:10.1024/LR.1205.208

[2]. Kollias, Dimitrios; Nicolaou, Mihalis A.; Kotsia, Irene; Zhao, Guoying; Zafeiriou, Stefanos (2017). [IEEE 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Honolulu, HI, USA (2017.7.21-2017.7.26)] 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Recognition of Affect in the Wild Using Deep Neural Networks. , (), 1972– 1979. doi:10.1109/CVPRW.2017.247

[3]. Kollias, Dimitrios; Zafeiriou, Stefanos P. (2020). Exploiting multi-CNN features in CNNRNN based Dimensional Emotion Recognition on the OMG in-the-wild Dataset. IEEE Transactions on Affective Computing, (), 1–1. doi:10.1109/TAFFC.2020.3014171

[4]. 4. Zafeiriou, Stefanos; Kollias, Dimitrios; Nicolaou, Mihalis A.; Papaioannou, Athanasios; Zhao, Guoying; Kotsia, Irene (2017). [IEEE 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Honolulu, HI, USA (2017.7.21-2017.7.26)] 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Aff-Wild: Valence and Arousal 'In-the-Wild' Challenge. , (), 1980–1987. doi:10.1109/CVPRW.2017.248

[5]. 5. Kollias, Dimitrios; Nicolaou, Mihalis A.; Kotsia, Irene; Zhao, Guoying; Zafeiriou, Stefanos (2017). [IEEE 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Honolulu, HI, USA (2017.7.21-2017.7.26)] 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) - Recognition of Affect in the Wild Using Deep Neural Networks. , (), 1972– 1979. doi:10.1109/CVPRW.2017.247

[6]. 6. Kollias, Dimitrios; Zafeiriou, Stefanos P. (2020). Exploiting multi-CNN features in CNN-RNN based Dimensional Emotion Recognition on the OMG in-the-wild Dataset. IEEE Transactions on Affective Computing, (), 1–1. doi:10.1109/TAFFC.2020.3014171

[7]. 7. Dimitrios Kollias;Attila Schulc;Elnar Hajiyev;Stefanos Zafeiriou; (2020). Analysing Affective Behavior in the First ABAW 2020 Competition . 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), (), – . doi:10.1109/fg47880.2020.00126.