A Survey on Emotion Recognition

Sara Sarrafan Chaharsoughi
Student, Masters in Applied Computing
Wilfrid Laurier University
Waterloo (ON),Canada

***Abstract*—Emotion Detection is defined as the process of identifying human emotion. Human emotions are mental states of sentiments that occur without conscious effort and are accompanied by physiological changes in facial muscles that result in facial expressions. Non-verbal communication methods such as facial expressions, eye movement, and gestures are used in many applications of human-computer interaction, among them facial emotion is widely used because it conveys the emotional states and feelings of persons. Emotion recognition is difficult due to the lack of a clear pattern between the many emotions on the face, as well as the complexity and variety. Because several critical extracted elements utilized for modelling the face are hand-engineered and rely on prior knowledge, the machine learning system will not attain a high accuracy rate for emotion recognition. Convolutional neural networks (CNN) have been developed in this work for the recognition of facial emotion expression. Facial expressions play a vital role in nonverbal communication which appears due to the internal feelings of a person that reflects on the face.**

Keywords—Emotion recognition, convolutional neural network, machine learning, deep machine learning, image processing

I. Introduction

Emotion is generated by specific events, and recognizing human emotion is a key topic in the research of human-computer interfaces (HCIs) that help people empathize [66][67]. Facial expressions, speech, physiological signs, and language are all examples of how emotions are expressed [68][69]. Facial emotion recognition and detection have very swiftly transitioned to a topic of continuous innovation and development, all due to the gradual removal of the limitations of computer vision with the help of Machine Learning [87]. Algorithms under Machine Learning and Deep Learning generally utilize a large level of computation power; hence the capabilities of algorithmic models match the magnitude of the real-world problems of Image Processing. However, using images to recognize and identify facial expressions and emotions is still a very challenging task as it is difficult to accurately extract emotional features in practice that are useful.

Every key facial element somewhat distinctively changes its values when emotions are expressed by humans. Keeping in contrast, the same features will generally have similar values across images with the same emotions. A general classification of human emotions could be as: happy, sad, anger, surprise, disgust, fear, and neutral. Specific facial expressions are linked to these emotions [70]. Facial expressions, along with other significant clues such as postures, gestures, spoken and vocal expressions, play a vital part in communicating feelings and attitudes [71]. All these emotions are dependent on several minimalistic facial muscle contortions which make our problem fulfilling to solve. Machine Learning and Neural Network algorithms have specifically been good solutions to extract and categorize facial features to further facilitate emotion recognition.

As technology advances, systems will be able to aid human specialists by extracting task-relevant information from data [72]. Automated emotion recognition systems, in particular, have applications in health care, such as detecting psychological discomfort [73], education, such as estimating student involvement [74], and gaming, such as improving the players' experience [75]. Emotion recognition has numerous real-world applications across various fields like autonomous vehicles, security, human-computer interaction, and healthcare. Graphical Processing Units have propelled the capabilities of Facial and Emotion Recognition being small pieces of hardware capable of completing millions of computations in a matter of seconds or minutes. Combining this hardware power with the statistically robust Machine Learning Algorithms help produce efficient solutions for our problem of Emotion Detection. Deep learning is based on the principle of learning a hierarchy of features in which higher-level features are made up of lower-level features. Edge detectors, for example, maybe the lowest-level characteristics, while groupings of edges could be detected at the second level, and filters, at the highest level, could approximate face features [76]. When trained on natural picture data, the lowest-level features in CNNs frequently seem Gabor-like. We focused more on the CNN algorithm as it is a very useful algorithm in this area for the recognition of these facial expressions. In this paper, we study facial emotion recognition by the CNN algorithm to determine one of the seven emotions from a facial image. The machine learning framework should, in theory, focus only on significant aspects of the face and be less sensitive to other areas of the face [91].

The remainder of the paper is as follows. Section II provides Related Works, in Section III we have Emotion Recognition, in Section IV, Facial Emotion Recognition Using Machine Learning Algorithms, and in section V there is the conclusion.

A. Applications Of Emotion Recognition

Emotion recognition is now one of the crux technologies of many applications and frontier areas of research. Each of these applications has unique and specific requirements, all of which are fulfilled by processing the features and detecting emotions. A few of these leading applications in today's world are discussed below: [44].

1) Education

Emotion recognition can be used in the classroom to analyze students' interests and focus in class. Real-time facial recognition is used to detect the faces of students and then recognize the emotion of students during lectures and analyze if the student is showing interest in the lecture or not [88]. The professor is analyzing the emotions during the lecture and improvises the lecture accordingly [45] [46].

Figure 1.1: Representation of facial emotion recognition in the classroom.

2) *Robotics*

Emotion recognition can be used for robotics as well. Robots can be used to help humans in many ways. Human-Robot interaction applications are very helpful for humans. Robots can interact will humans and can copy human emotion by analyzing human facial expressions. This technique can be used further for automated systems [47].

Figure 1.2: Representation of facial emotion recognition used by the robot.

3) *Safety Aid*

Safety Applications can be benefited from facial emotion recognition. Many accidents take place because of drowsiness, fatigue, or the bad mood of the driver. A system is being developed that uses a camera to record drivers and in real-time detects drivers' facial expressions. Using facial expressions, it detects the emotion of the driver and if it detects something wrong in the behavior of the driver it performs safety measures being defined already to the machine. Such as if there is less alertness of the driver during driving the car, the car switches to auto-driving mode [50].

Figure 1.3: Representation of facial emotion recognition used for drivers.

II. Related Works

The elements of a gesture dynamic are proposed for an emotion recognition system and supervised learning methods are used to evaluate them [6]. The hidden units of a convolutional autoencoder are proposed as a framework for high-level parameters in body movements [7]. Using STIP features, a system for recognizing a person's affective state from face-and-body footage is proposed [8]. The survey on sentiment analysis and deep learning applications is discussed [9][10]. The backpropagation algorithm is used to construct a deep learning system for big data sets [11]. A self-organizing neural architecture was created to identify emotional states from full-body motion patterns [12]. For emotion recognition from video, a model combining CNNs and RNNs has been presented [13]. For the multimodal emoFBVP database, deep learning models with DCNN were created [14]. The results indicate a considerable improvement in accuracy when using a model with hierarchical feature representation for nonverbal emotion recognition [15]. Using promising neural network topologies, a unique design of an artificially intelligent system for emotion recognition is proposed [16]. Emotion Recognition in the Wild (EmotiW) is a system built utilizing a hybrid CNN-RNN architecture that outperforms existing methodologies[17]. A new framework for automatic emotional body gesture identification is being developed to differentiate culture and gender differences [18].

III. Emotion Recognition

Emotion recognition is the main topic of Image Processing. Emotion can be recognized using different techniques. Emotion recognition can be very useful in many areas of research. It can be used to identify users' emotions while the user is using an application to improvise human-computer interaction to maximize the easiness of an application. There are many ways to recognize emotion.

A. *Emotion Recognition Based on Facial Recognition*

Facial expression is the basic way to recognize emotion. As the emotion changes facial expression describes the emotion of humans. Each emotion has a different facial expression containing different characteristics. The main face features to be considered in emotion recognition are eyebrows, eyes, and mouth. Getting features correctly is the main key that will affect the result of emotion recognition. It's the main and most used emotion detection technique being used for human-computer interaction. There are various steps in recognizing facial expressions. The first face is being detected and then after that main features of the face is being extracted and then classification algorithms are being used to recognize emotion. Recognition of seven emotional states based on facial expressions (neutral, pleasure, sadness, surprise, anger, fear, and disgust). As features, six participants' coefficients representing aspects of facial expressions were used. For a three-dimensional facial model, the features were estimated. The characteristics were classified using a k-NN classifier and an MLP neural network [1].

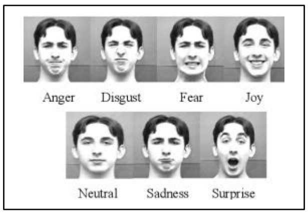

Figure 3.1: Facial emotion expressions.

B. *Emotion Recognition Based on Speech*

Human speech is another way of expressing emotion. Different emotion has different characteristics that can be used for emotion detection. Different emotional speech has different voice pitches, speed of taking and duration of the speech, and many other characteristics. There are many factors that can affect the result of emotion detection using speech. One of them is the noisy environment in the background while recording speech. For speech-based emotion recognition, one needs to control the surrounding environment to minimize errors. Hidden Markov models, time series modeling, cestrum processing, and deep learning techniques are the main computational fundamentals of emotion estimation from voice signals [2].

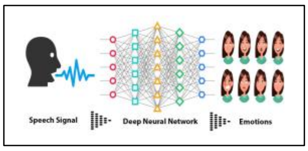

Figure 3.2: Representation of emotion recognition using speech.

C. Emotion Recognition Based on Gestures

Emotion can be identified by observing human gestures. In a different environment, humans have different gestures to express different emotions. Human movement and motions are being identified and evaluated to identify the emotions of humans using human motion posture. There is a limitation of this technique as well. Many human gestures do not have any emotional significance or many human emotions have the same gesture. So, to identify emotion using gestures, we use facial expressions or speech with this technique to get a more accurate result. Emotional body motions can be used as part of what is known as body language. Then create a comprehensive framework for recognizing emotional body gestures automatically. In RGB and 3D, it conducts human detection and dynamic body posture estimation algorithms. Then employ multi-modal techniques to improve emotion identification by combining speech or facial expressions with body gestures [3].

Figure 3.3: Representation of emotion recognition using gesture.

D. Emotion Recognition Based on Physiological Signals

Emotions can be obtained from physiological signals. In different conditions, the human body produces different physiological signals that are being considered for emotion recognition. Emotions can be reflected in physiological signs such as brain electricity, electrocardiogram, pulse, and skin electrical reaction. Emotion detection using physiological signals is more accurate as humans do not have control over their physiological signals, it reflects human's emotional conditions. Electroencephalogram (EEG) signals can be used to recognize emotions. Because EEG signals are noisy, non-linear, and non-stationary, creating an emotion recognition system with high accuracy is a difficult task. The model will be built on the construction of feature maps based on topographic (TOPO-FM) and holographic (HOLO-FM) representations of EEG signal features for emotion recognition. On feature maps, deep learning was used as a feature extractor method, and the retrieved features were then fused together for the classification process to recognize diverse objects [4].

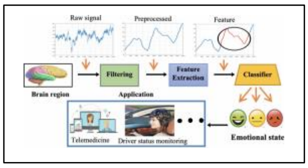

Figure 3.4: Emotion recognition using physiological signals.

IV. Facial Emotion Recognition Using Machine Learning Algorithms

Emotion recognition using machine learning algorithms is the basic method to perform this task. The first face is being detected and then facial expressions are being extracted using different techniques. Facial expressions extracted are then being used by machine learning classifiers to get the output of emotion [90]. There are many different machine learning algorithms to perform this task that are being discussed.

A. Feature Extractors

The face is being detected and then the face is being processed. There are many different techniques to extract important features from the face that are being discussed.

1) Face Registration and Representation

Input facial image is being aligned with the similar data provided before. Landmark points are being used to point out important facial features such as eyebrows, eyes, nose, and mouth. This is done on the whole training dataset. After that, it is being processed by three mostly used algorithms i.e., Gabor Feature, Local Binary Pattern, and Histogram of Oriented Gradient. These three are the best-used feature extraction algorithms that provide accurate results. [77]

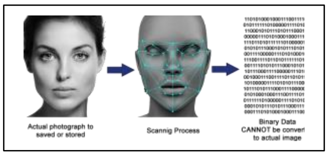

Figure 4.1: Face registration and representation.

2) Gabor Feature

Gabor features can be compared to human visuals because of their orientation representations. By rotating and filtering one primary wavelet, these filters can be created. Among the various important image attributes, such as edge orientation histograms and box filters, they are the best.

For Gabor analysis, the eye centers must first be located before the images can be aligned properly. Transform, rotate, and scale are used to achieve this alignment. This is how 2D images are usually registered. Manual landmark determination is used to perform normalization. This is to avoid any registration scheme misalignment impacts. [78]

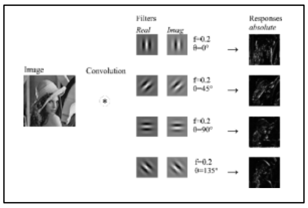

Figure 4.2: Gabor feature.

3) Local Binary Pattern (LBP)

The texture and geometry of a digital picture are described by Local Binary Pattern. This is accomplished by segmenting a picture into tiny sections from which the characteristics are retrieved. These characteristics are binary patterns that characterize the pixels' surroundings in the areas. The acquired features from the areas are combined into a single feature histogram, which serves as an image representation. The similarity of their histograms may then be used to compare images. Face recognition utilizing the Local Binary Pattern approach, according to various studies, produces excellent results in terms of speed and discrimination. The approach appears to be highly resistant to face photos with various facial expressions, lighting circumstances, image rotation, and human aging[80].

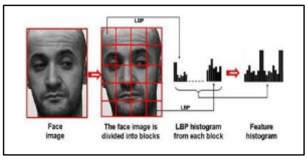

Figure 4.3: Local binary pattern.

The original Local Binary Pattern operator is a useful tool for describing the texture. The image's pixels are labeled by thresholding each pixel's 3x3-neighborhood with the center value and treating the result as a binary integer. After that, the labels' histogram may be utilized as a texture description. A diagram of the basic LBP operator may be seen in Figure 5.4. The operator was then expanded to include neighborhoods of various sizes. Any radius and number of pixels in the neighborhood may be achieved by using a circular neighborhood and bilinearly interpolating the pixel values.

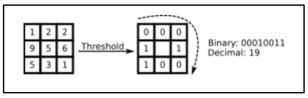

Figure 4.4: Original local binary pattern operator.

4) *Histogram of Oriented Gradients (HOG)*

Histogram of Oriented Gradient (HOG) is a feature extraction descriptor that is utilized in a variety of applications. Geometric and photometric modifications have no effect on the HOG features. HOG is implemented by splitting the picture into small linked sections called cells and producing a histogram of gradient directions or edge orientations for the pixels within each cell. The HOG descriptor is the result of combining these histograms.

There are two major parameters that define the HOG descriptor. The size of cells per row and column is the first parameter. The size of the cell reflects the size of the patch used in the histogram computation. The second parameter pertains to the number of bins orientation, which is mostly utilized to generate the gradient's intervals of angles [83].

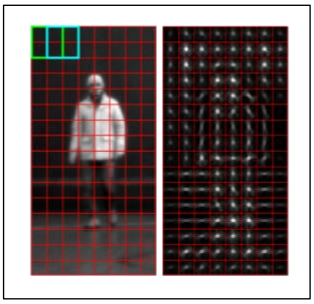

Figure 4.5: Histogram of oriented gradients.

5) *Dimensionality Reduction*

After the features are being extracted it is high-dimensional data and it contains many extra features that are of no use. These redundant features should be removed and high-dimensional data should be converted into low-dimensional data.

Laplacian eigenmaps algorithm can be used to convert high-dimensional data into low-dimensional data. The nearest point in high-dimensional space is converted into a close point in low-dimensional space using this procedure. The extended eigenvector problem is used to solve problems. The first n eigenvectors correspond to the first n eigenvalues in n-dimensional Euclidean space. A spectral regression algorithm is used to map high-dimensional data to find a projection function. High-dimensional data we get from feature extraction algorithm, i.e., Gabor Feature, Local Binary Pattern, and Histogram of Oriented Gradient [77].

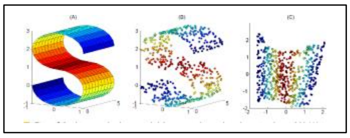

Figure 4.6: Dimensionality reduction.

B. Machine Learning Classifiers

After the face is being processed and important features are being extracted from the face. Machine learning classifiers will be used to predict emotion. There are a variety of classifiers available. Some of them are discussed.

1) *Support Vector Machine (SVM)*

A Support Vector Machine (SVM) was first introduced in 1992 by Boser, Guyon, and Vapnik at COLT-92. The support vector machine (SVM) is a class of supervised learning algorithms that are used for classification and regression [63]. SVMs are part of a family of generalized linear classifiers. Alternatively, Support Vector Machine (SVM) is a classification and regression prediction tool that makes use of machine learning theory to maximize predictive accuracy while automatically avoiding overfitting.

Support Vector machines are systems that employ the hypothesis space of linear functions in a high dimensional feature space and are taught with an optimization theory learning algorithm that applies a statistical learning theory learning bias. The support vector machine was first popular within the NIPS community and is now an active aspect of machine learning research all across the world. When employing pixel maps as input, SVM achieves accuracy similar to advanced neural networks with elaborated features in a handwriting recognition job [64].

It is also utilized for a variety of applications, including handwriting analysis, face analysis, and so on, with a focus on pattern classification and regression-based applications.

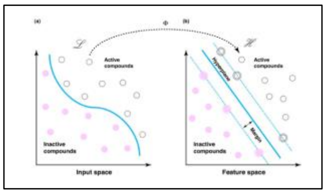

Figure 4.6: Support vector machine.

2) *Random Forest*

The random forest classifier is made up of a number of tree classifiers, each of which is constructed using a random vector sampled separately from the input vector, and each tree casts a unit vote for the most popular class to categorize an input vector [65]. To create a tree, the random forest classifier utilized in this work uses randomly selected characteristics or a mixture of features at each node. For each feature/feature combination chosen, bagging, a method of generating a training dataset by randomly drawing with replacement N samples, where N is the size of the original training set [65], was employed.

Any instances (pixels) are categorized by selecting the class with the highest number of votes from all tree predictors in the forest. The design of a decision tree necessitated the selection of an attribute selection measure as well as a pruning mechanism. There are several techniques for selecting characteristics for decision tree induction, and the majority of them provide a quality measure directly to the attribute. The Information Gain Ratio criteria (Quinlan) and the Gini Index are the most commonly utilized attribute selection metrics in decision tree induction [65].

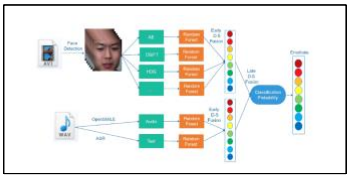

Figure 4.7: Random forest.

3) *K-Nearest Neighbor (KNN)*

The KNN (k Nearest Neighbors) algorithm is a non-parametric, instance-based, or lazy approach that has been recognized as one of the most straightforward methods in data mining and machine learning [59][60][61]. The KNN method works on the idea that the most comparable samples in the same class have a high probability. In general, the KNN method determines the k nearest neighbors of a query in the training dataset and then predicts the query based on the main class in the k nearest neighbors. As a result, it was recently chosen as one of the top ten algorithms in data mining [62].

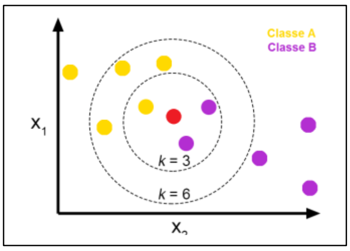

Figure 4.8: k-nearest neighbor.

4) *Artificial Neural Networks (ANN)*

An Artificial Neuron is essentially a biological neuron engineering technique. It's a gadget with a lot of inputs and just one output. An ANN is made up of a large number of small processing units that are linked and stacked together [57][58].

Similar to real neurons, artificial neural networks have artificial neurons that accept inputs from other components or artificial neurons, and when the inputs are weighted and combined, the result is turned into the output via a transfer function. A Sigmoid, hyperbolic tangent function or a step might be used as the transfer function [57].

Basically, computers are adept at calculations in that they accept inputs, analyze them, and then return a result based on calculations performed at specific Algorithms that are coded in software, but ANN enhance their own rules; the more judgments they make, the better the decisions may become [57].

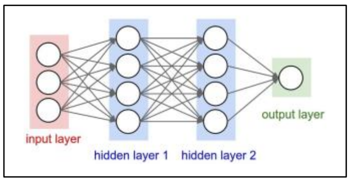

Figure 4.9: Artificial neural network.

5) *Decision Tree*

A decision tree is a graphical representation of such mappings. A tree is either a leaf node with a class label or a structure made up of a test node linked to two or more subtrees. A test node computes some outcomes depending on an instance's attribute values, with each potential outcome connected with one of the subtrees. Starting at the root node of the tree, an instance is categorized. If this node is a test, the instance's conclusion is determined and the process is resumed using the appropriate subtree. When a leaf is discovered, its label indicates the anticipated class of the instance.

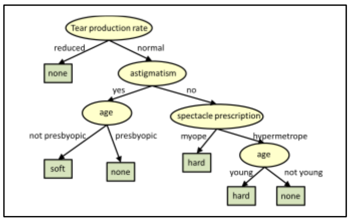

Figure 4.10: Decision tree.

C. *Deep Machine Learning Algorithm*

Deep learning (sometimes referred to as deep machine learning, deep structured learning, hierarchical learning, or DL) is a sort of machine learning that learns through a sequence of algorithms. that tries to model high-level abstractions in data using Considering complicated structures or using model architectures. Otherwise, it's made up of a lot of non-linear transformations [19]. Deep neural networks, convolutional deep neural networks, deep belief networks, and recurrent neural networks have all been used in fields like computer vision, automatic speech recognition, natural language processing, to obtain state-of-the-art outcomes on numerous tasks, researchers combined auditory recognition with bioinformatics. Deep learning has also been referred to as a rebranding of neural networks or a buzzword.[20]. Deep learning is a class of machine learning algorithms that use for:

1. feature extraction and transformation, utilize a cascade of multiple layers of nonlinear processing units. The output from the preceding layer is used as input for each subsequent layer. Pattern analysis (unsupervised) and classification are examples of applications for the methods, which can be supervised or unsupervised (supervised) [21].

2. They are based on the (unsupervised) learning of many levels of data features or representations to build a hierarchical representation, higher-level features are generated from lower-level features.

3. They are a subset of the larger machine learning area of data representation learning.

4. They acquire a hierarchy of concepts by learning several levels of representations that correspond to various levels of abstraction [22].

1) Convolutional Neural Network (CNN)

CNN is an important area of deep learning, a neural feedforward network based on biological receptive field mechanisms [23], [24]. CNN does not need to manually extract features. CNN's design is influenced by visual perception. [25] CNN has been widely used in image analysis and speech recognition in recent years. However, using CNNs for card classification remains difficult [26]. Several CNN architectures have been introduced in the last decade. From 1989 to the present, various changes have been made to the CNN architecture. Such changes include structural reformulation, regularization, and parameter optimization. On the contrary, note that the significant improvement in CNN performance is mainly due to the reorganization of processing units and the development of new blocks. In particular, the latest developments in the CNN architecture for the use of network depth have been carried out. With enough training data, CNN can learn data-driven, highly representative, layered hierarchical image features. [27]. Convolutional neural networks have recently demonstrated excellent image classification performance in large-scale visual recognition challenges. One of the functions of a CNN is to reduce images into a format that is easier to handle while preserving elements that are important for accurate prediction. This is crucial for creating a building that is not just beautiful but also functional.

a) Usage of CNN

Oquab M [28] showed how to use a limited amount of training data to effectively transfer image representations learned from CNNs to other visual recognition tasks in large annotated datasets. Gatys LA [29] used an image representation derived from a neural convolutional network optimized for object recognition. This makes the expanded image information explicit. The results provide new insights into depth imaging for learning neural convolutional networks and show their potential in advanced image synthesis and manipulation. Milletari F [30] uses convolutional neural networks (CNNs) to solve problems in the areas of computer vision and medical image analysis and proposes volume, full convolution, and 3D image segmentation techniques based on neural networks. bottom. Zbontar J [31] shows how to extract depth information from a modified image pair. It approached the problem by learning a similarity measure for small image fields using a convolutional neural network. Ma L [32] suggested using CNN for image question answering (QA) tasks. The CNNs proposed to provide an end-to-end framework with a convolutional architecture for learning not only the representation of images and questions but also the intermodal interactions to generate answers. Pathak D [33] suggested a way to learn high-density pixel labels from tags at the image level. Each image-level label applies restrictions to the output markup of the CNN classifier.

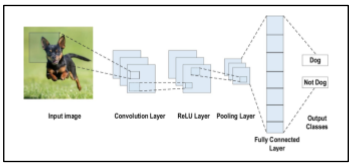

Figure 4.11: Structure of a CNN.

b) CNN Architecture

The general structure of a CNN generally consists of an input layer, a convolution layer (Conv), a pooling layer (Pool), a fully connected layer (FC), and an output layer, as shown in Figure 4.12.

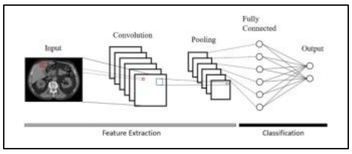

Figure 4.12: Structure of a typical CNN model.

As shown in figure 4.13, Conv is used to extract features from the input image. The input image is extracted by the Dot Product operation using a convolution kernel consisting of weight matrices to produce a labeled image.

A CNN is a deep learning system that takes an image as input, assigns relevance (learnable weights and biases) to various aspects/objects in the image, and can distinguish between them. A CNN requires substantially less preparation than conventional classification techniques. The architecture of a CNN is inspired by the arrangement of the visual cortex [34] and is akin to the connectivity network of neurons in the human brain.

A simple CNN is a series of layers and each layer that uses a differentiable function to transform one volume of activations to another. CNN architectures are built with three types of layers: Convolutional Layer, Pooling Layer, and Fully Connected Layer (exactly like in regular Neural Networks). These layers will be stacked to form the full CNN architecture.

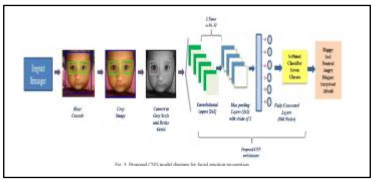

Figure 4.13: CNN architecture.

c) Emotion Detection

Emotion is also a significant part of intelligence. The challenge is that we want to develop a system that can discriminate between emotions, which forces scholars to use their own emotional cognitive system. To detect emotion in the voice, bionics and biology have been used [37]. They use the physical anatomy of the human ear to create appropriate features for human ear perceived qualities, such as the MFCC and Lyon cochlear model [38], or to develop the model to increase recognition performance.

d) Face detection using CNN

L.Tan et al. used MTCNN to detect photos [39]. MTCNN is a face detection approach based on CNN. It employs three cascaded CNNs for rapid and accurate face detection as well as face detection learning (five facial landmarks detection, i.e., two eyes, two mouth corners, and nose). It creates a picture pyramid based on the input photos, which then feeds into a three-stage cascaded architecture. In the first stage, candidate regions are created, then refined in the second and third stages. The third stage produces the final detection findings as well as the associated facial landmark position [40].

V. Conclusion

This paper outlines the application of Facial Emotion Recognition that is being used. It also focuses on the algorithms that are being used for Facial Emotion Recognition includes machine learning and deep learning algorithms. After surveying multiple research papers and analyzing results, we can also conclude that deep learning algorithm i.e., Convolutional Neural Network (CNN), works much better for Facial Emotion Recognition and provides more accurate results as compared to other algorithms.

Acknowledgment

We would like to appreciate Dr. Safaa Bedawi whose guidance was very helpful in researching existing Emotion Recognition Techniques. We are also grateful to people pioneering the research in Emotion Recognition to give us a chance to explore the latest technology and best methods.

References

[1] Tarnowski, Paweł & Kołodziej, Marcin & Majkowski, Andrzej & Rak, Remigiusz, "Emotion Recognition Using Facial Expressions", International Conference on Computational Science, 12-14 June, 2017.

[2] Drakopoulos, Georgios & Pikramenos, George & Spyrou, Evaggelos & Perantonis, Stavros, "Emotion Recognition From Speech: A Survey", International Conference on Web Information Systems and Technologies, 2019.

[3] Fatemeh Noroozi, Ciprian Adrian Corneanu, Dorota Kaminska, Tomasz Sapi ´ nski, Sergio Escalera, and Gholamreza Anbarjafari, "Survey on Emotional Body Gesture Recognition", Journal of IEEE Transactions on Affective Computing, Volume 1, 2018.

[4] Ante Topic, Mladen Russo, "Emotion Recognition based on EEG Feature Maps through Deep Learning Network", Engineering Science and Technology, an International Journal, Volume 24, Issue 6, 2021.

[5] Shu L, Xie J, Yang M, Li Z, Li Z, Liao D, Xu X, Yang X, "A Review of Emotion Recognition Using Physiological Signals", Sensors, 2018.

[6] J. Arunnehru, M. Kalaiselvi Geetha. "Automatic Human Emotion Recognition in Surveillance Video", Intelligent Techniques in Signal Processing for Multimedia Security, Springer-Verlag, 2017.

[7] D.Holden, J.Saito, T.Komura. "A Deep Learning Framework for Character Motion Synthesis and Editing" SIGGRAPH '16 Technical Paper, July 24 - 28, Anaheim, CA,2016.

[8] H. Gunes, C.Shan, Sh.Chen, Y.Tian. "Bodily Expression for Automatic Affect Recognition. Emotion Recognition: A Pattern Analysis Approach" Published by John Wiley & Sons, Inc.,2015.

[9] L.Zhang, Sh.Wang, B.Liu. "Deep Learning for Sentiment Analysis",2018.

[10] H. Brock. "Deep learning - Accelerating Next Generation Performance Analysis Systems" 12th Conference of the International Sports Engineering Association, Brisbane, Queensland, Australia, pp. 26–29. 2018.

[11] Y.LeCun, Y. Bengio, Geoffrey Hinton. "Deep learning", Nature, Vol. 521, pp. 436-444,2015.

[12] N. Elfaramawy, Pablo Barros, German I. Parisi, Stefan Wermter. "Emotion Recognition from Body Expressions with a Neural Network Architecture" Session 6: Algorithms and Learning, Bielefeld, Germany,2017.

[13] P. Khorrami, Tom Le Paine, Kevin Brady, Charlie Dagli, Thomas S. Huang. "How Deep Neural Networks Can Improve Emotion Recognition on Video Data", 2017.

[14] H.Ranganathan, Sh.Chakraborty, S. Panchanathan, "Multimodal Emotion Recognition using Deep Learning Architectures", 2017.

[15] P. Barros, D. Jirak, C. Weber, S. Wermter. "Multimodal emotional state recognition using sequence dependent deep hierarchical features" Neural Networks. 72, pp. 140–151, 2015.

[16] E. Correa, A. Jonker, M. Ozo, R. Stolk. "Emotion Recognition using Deep Convolutional Neural Networks",2016.

[17] S. Ebrahimi, V. Michalski, Kishore Konda, Roland Memisevic, Christopher Pal."Recurrent Neural Networks for Emotion Recognition in Video" ICMI 2015, Seattle, WA, USA.,2016.

[18] F. Noroozi, C. Adrian Corneanu, D. Kami´nska, T. Sapi´ nski, S. Escalera, and G. Anbarjafari "Survey on Emotional Body Gesture Recognition" Journal of IEEE Transactions on Affective Computing, 2015.

[19] R. Collobert, "Deep Learning for Efficient Discriminative Parsing". videolectures.net. Ca. 7:45., May 6, 2011.

[20] G. Lee. "Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts". IEEE Spectrum, 20 October 2014.

[21] P. Glauner, "Comparison of Training Methods for Deep Neural Networks", arXiv:1504.06825,2015.

[22] S. Hyun Ah, and S.Y, Lee. "Hierarchical Representation Using NMF." Neural Information Processing. Springer Berlin Heidelberg, 2013.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015.

[24] F. B. Zhou, L. H. Zou, X. J. Liu, and F. Y. Meng, "Micro landform classification method of grid DEM based on convolutional neural network", *Geomatics Inf. Sci. Wuhan Univ.*, 2018.

[25] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[26] Luo, L. Liu, J. Yin, et al., "Deep learning of graphs with Ngram convolutional neural networks". IEEE Trans. Knowl. Data Eng. 29(10), 1–1, 2017.

[27] S. H.-Chang, H.R. Roth, M. Gao, et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset. M. " 2016.

[28] M Oquab, L. Bottou, I. Laptev, et al., "Learning and transferring mid-level image representations using convolutional neural networks ",2014.

[29] L.A. Gatys, A.S. Ecker, M. Bethge, "Image Style Transfer Using Convolutional Neural Networks"[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society (CVPR 2016, Las Vegas, pp. 2414–2423, 2016.

[30] F. Milletari, N. Navab, S.A. Ahmadi, V-Net: "Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,", pp. 565–571,2016.

[31] J. Zbontar, Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches".J. Mach. Learn. Res. 17(1–32), 2,2016.

[32] L. Ma, Z. Lu, H. Li, "Learning to answer questions from image using convolutional neural network", AAAI 3(7), 16 ,2016.

[33] D. Pathak, P. Krahenbuhl, T. Darrell, Constrained convolutional neural networks for weakly supervised Segmentation, IEEE International Conference on Computer Vision. IEEE Computer Society (ICCV2015, Santiago, pp. 1796–1804, 2015.

[34] A. Amini, A. Soleimany, "MIT deep learning open access course" 6.S191, available online: http://introtodeeplearning.com, 2020.

[35] T. Lindeberg, "Scale invariant feature transform," Scholarpedia, vol. 7, no. 5, p. 10491, May 2012.

[36] D. M. Hawkins, "The problem of overfitting," J. Chem. Inf. Comput. Sci., vol. 44, no. 1, pp. 1–12, Jan. 2004.

[37] M. Drolet, Ricarda I. Schubotz, Julia Fischer: "Explicit authenticity and stimulus features interact to modulate BOLD response induced by emotional speech", Cogn Affect Behav Neurosci 13,2013.

[38] L. Caponetti, C. Alessandro Buscicchio, and G. Castellano: "Biologically inspired emotion recognition from speech", EURASIP Journal on Advances in Signal Processing, 2011.

[39] L. Tan, K. Zhang, K. Wang, X. Zeng, X. Peng, and Y. Qiao, “Group emotion recognition with individual facial emotion CNNs and global image based CNNs,” in Proc. 19th ACM Int. Conf. Multimodal Interact., Nov. 2017.

[40] K.Zhang, Z. Zhang, Z. Li, and Y. Qiao. “Joint face detection and alignment using multitask cascaded convolutional networks” IEEE Signal Processing Letters 23, 10, 2016.

[41] Krizhevsky A, Sutskever I, Hinton GE. “Imagenet classification with deep convolutional neural networks. Commun “ACM. 2017.

[42] LeCun Y, Jackel LD, Bottou L, Cortes C, Denker JS, Drucker H, Guyon I, Muller UA, Sackinger E, Simard P, et al. “Learning algorithms for classification: a comparison on handwritten digit recognition. Neural Netw Stat Mech Perspect”, 2015.

[43] S. Shah, “A comprehensive guide to convolutional neural networks”, available online: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-theeli5-way-3bd2b1164a53, 2018.

[44] C.Vinola and K.Vimaladevi, "A Survey on Human Emotion Recognition Approaches, Databases and Applications", Electronic Letters on Computer Vision and Image Analysis, December 2015.

[45] J.F. Grafsgaard, J.B. Wiggins, K.E. Boyer, E.N. Wiebe, and J.C. Lester, "Automatically Recognizing Facial Expression: Predicting Engagement and Frustration", International Conference on Educational Data Mining, 2013.

[46] J. Whitehill, Z. Serpell, Y. Lin, A. Foster, and J. R. Movellan, " The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions", IEEE Transactions on Affective Computing, Volume 5, 2014.

[47] B. Raducanu and F. Dornaika, "Dynamic Facial Expression Recognition using Laplacian Eigen Maps-based Manifold Learning", IEEE International Conference on Robotics and Automation, May 2010.

[48] L.Zhang, A.Hossain, M.Jiang, "Intelligent Facial Action and Emotion Recognition for Humanoid Robots", International Joint Conference on Neural Networks, July 2014.

[49] Urvashi Agrawal and Shubhangi Giripunje, "Emotion and Gesture Recognition with Soft Computing Tool for Drivers Assistance System in Human Centered Transportation", International Conference on Systems, Man, and Cybernetics, 2013.

[50] J.Chen and Q. Ji, "Drowsy Driver Posture, Facial, and Eye Monitoring Methods", Handbook of IntelligentVehicles, 2012.

[51] Charlotte Jacobé de Naurois, Christophe Bourdin, Anca Stratulat, Emmanuelle Diaz, Jean-Louis Vercher, “Detection and Prediction of Driver Drowsiness Using Artificial Neural Network Models”, Accident Analysis & Prevention, Volume 126, 2019.

[52] Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas, and H Sebastian Seung. Digital selection and analog amplification coexist in a cortex-inspired silicon circuit. Nature 405, 6789 947, 2000.

[53] Zuo, Z.; Li, J.; Wei, B.; Yang, L.; Fei, C.; Naik, N. Adaptive Activation Function Generation Through Fuzzy Inference for Grooming Text Categorisation. In Proceedings of the 2019 IEEE International Conference on Fuzzy Systems, New Orleans, LA, USA, 23–26 June 2019.

[54] Lohani, H.K.; Dhanalakshmi, S.; Hemalatha, V. Performance Analysis of Extreme Learning Machine Variants with Varying Intermediate Nodes and Different Activation Functions. In Cognitive Informatics and Soft Computing; Springer: Singapore, 2019; pp. 613–623.

[55] Yu, D., et al. Mixed pooling for convolutional neural networks. in International conference on rough sets and knowledge technology. 2. Springer,2014.

[56] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, “Going deeper with convolutions”, Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[57] Ajith Abraham, “Artificial Neural Networks”, Stillwater, OK, USA, 2005.

[58] Prof. Leslie Smith, “ An Introduction to Neural Networks”, University of Stirling., 1996,98,2001,2003.

[59] Zhang, S. Jin, Z. Zhu, X. Zhang, J., “Missing data analysis: A kernel-based multi-imputation approach”, Gavrilova, M.L., Tan, C.J.K. (eds.) Transactions on Computational Science III. LNCS, vol. 5300, pp. 122–142. Springer, Heidelberg 2009.

[60] Zhu, X., Zhang, L., Huang, Z.: A sparse embedding and least variance encoding approach to hashing. IEEE Transactions on Image Processing 23(9), 3737–3750, 2014.

[61] Zhu, X., Zhang, S., Jin, Z., Zhang, Z., Xu, Z.: Missing value estimation for mixed-attribute data sets. IEEE Transactions on Knowledge and Data Engineering 23(1), 110–121, 2011.

[62] Zhu, X., Huang, Z., Shen, H.T., Zhao, X.: Linear cross-modal hashing for efficient multimedia search. In: ACM Multimedia, pp. 143–152, 2013.

[63] Vapnik, V., Estimation of Dependencies Based on Empirical Data. Empirical Inference Science: Afterword of 2006, Springer, 2006.

[64] Nello Cristianini and John Shawe-Taylor, “An Introduction to Support Vector Machines and Other Kernel-based Learning Methods”, Cambridge University Press, 2010.

[65] Breiman L Random forests—random features. Technical Report 567, Statistics Department, University of California, Berkeley, 1999.

[66] Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; Taylor, J.G. Emotion recognition in human-computer interaction. IEEE Signal Process. Mag.,18, 32–80,2001.

[67] Busso, C.; Deng, Z.; Yildirim, S.; Bulut, M.; Lee, C.M.; Kazemzadeh, A.; Lee, S.; Neumann, U.; Narayanan, S. Analysis of emotion recognition using facial expressions, speech, and multimodal information. In Proceedings of the 6th International Conference on Multimodal Interfaces, State College, PA, USA, 14–15; ACM: New York, NY, USA, 2004; pp. 205–211, October 2004.

[68] Ayadi, M.; Kamel, M.S.; Karray, F. Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognit., 44, 572–587, 2011.

[69] Wu, C.H.; Chuang, Z.J.; Lin, Y.C. Emotion recognition from text using semantic labels and separable mixture models. ACM Trans. Asian Lang. Inf. Process. TALIP 5, 165–183, 2006.

[70] Ekman, Paul. Are there basic emotions? Psychological Review, 550–553,1992.

[71] Mehrabian, Albert. Silent messages. Wadsworth, 1971.

[72] Gratch, Jonathan and Lucas, Gale M and King, Aisha Aisha and Morency, Louis-Philippe. It’s only a computer: the impact of human-agent interaction in clinical interviews. Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, 85–92, 2014.

[73] Scherer, Stefan and Stratou, Giota and Mahmoud, Mohamed and Boberg, Jill and Gratch, Jonathan and Rizzo, Alessandro and Morency, Louis-Philippe. Automatic behavior descriptors for psychological disorder analysis. Automatic Face and Gesture Recognition (FG), 10th IEEE International Conference and Workshops on. IEEE, 1–8,2013.

[74] Shan, Caifeng and Gong, Shaogang and McOwan, Peter W. Facial expression recognition based on local binary patterns: A comprehensive study. Image Vision Comput., 27 (6), 803–816,2009.

[75] Shaker, Noor, and Asteriadis, Stylianos and Yannakakis, Georgios N and Karpouzis, Kostas. A game-based corpus for analyzing the interplay between game context and player experience. Affective Computing and Intelligent Interaction, Springer. 547–556, 2011.

[76] Khorrami, Pooya, and Paine, Tom Le and Brady, Kevin and Dagli, Charlie and Huang, Thomas S. How deep neural networks can improve emotion recognition on video data. arXiv preprint arXiv:1602.07377, 2016.

[77] Mehta, Dhwani, Mohammad F.H. Siddiqui, and Ahmad Y. Javaid, "Recognition of Emotion Intensities Using Machine Learning Algorithms: A Comparative Study", Sensors, 2019.

[78] Hassan, Masoud & Hussein, Haval & Eesa, Adel & Mustafa, and Ramadhan, "Face Recognition Based on Gabor Feature Extraction Followed by FastICA and LDA", Computers, Materials, and Continua, 2021.

[79] Ahonen, Timo & Hadid, Abdenour & Pietikäinen, Matti, "Face Recognition with Local Binary Patterns", 8th European Conference on Computer Vision, 2004.

[80] O. S. Kulkarni, S. M. Deokar, A. K. Chaudhari, S. S. Patankar and J. V. Kulkarni, "Real Time Face Recognition Using LBP Features", 2017

International Conference on Computing, Communication, Control and Automation (ICCUBEA), 2017.

[81] Md. Abdur Rahim, Md. Najmul Hossain, Tanzillah Wahid, & Md. Shafiul Azam, "Face Recognition using Local Binary Patterns (LBP)", Global Journal of Computer Science and Technology Graphics & Vision, Volume 13, Issue 4, 2013.

[82] Muhammet Fatih Aslan, Akif Durdu, Kadir Sabanci, Meryem Afife Mutluer, "CNN and HOG based comparison study for complete occlusion handling in human tracking", Measurement, Volume 158, 2020.

[83] Deniz, Oscar & Bueno, Gloria & Salido, Jesús & De la Torre, Fernando, "Face recognition using Histograms of Oriented Gradients", Pattern Recognition Letters, 2011.

[84] Bouchra Nassih, Aouatif Amine, Mohammed Ngadi, Nabil Hmina, "DCT and HOG Feature Sets Combined with BPNN for Efficient Face Classification", Second International Conference on Intelligent Computing in Data Sciences, 2018.

[85] Chen, Jie & Shan, Shiguang & Yang, Peng & Yan, Shengye & Chen, Xilin, and Gao, Wen., "Novel Face Detection Method Based on Gabor Features", Advances in Biometric Person Authentication, 2004.

[86] Lin-Lin Huang, Akinobu Shimizu, and Hidefumi Kobatake, "Robust face detection using Gabor filter features", Pattern Recognition Letters, Volume 26, Issue 11, 2005.

[87] Shan, L.; Deng, W. Deep facial expression recognition: A survey. IEEE Trans. Affect. Comput. 2020.

[88] . Eleyan, A.M.A.; Akdemir, B. Facial expression recognition with dynamic cascaded classifier. Neural Comput. Appl, 32, 6295–6309, . 2020.

[89] Amil Khanzada, Charles Bai, and Ferhat Turker Celepcikay. Facial expression recognition with deep learning. arXiv preprint arXiv:2004.11823, 2020

[90] Yousif Khaireddin and Zhuofa Chen. Facial emotion recognition: State of the art performance on fer2013. arXiv preprint arXiv:2105.03588, 2021

[91] Shervin Minaee, Mehdi Minaei, and Amirali Abdolrashidi. Deep-emotion: Facial expression recognition using attentional convolutional network. Sensors, 21(9):3046, 2021.