Realtime Face Detection, Emotion, and Gender Classification using Machine Learning.

Introduction:

Interaction between humans and computers is a common phenomenon and an inherent ability to differentiate between different faces. The advancements in technology trivially benefited from distinct issues in computer vision, modified by sex, hair, and other features [CITATION Noz96 \l 1033]. Face recognition technologies are thus used to improve connectivity to recognize and validate individuals through the characteristics of their faces. It is very much important to understand the facial features and their behavior.

While these attributes and expressions help in classifying human face emotions. The recent improvement in technology has contributed to the usage of Artificial Intelligence because these applications are capable of understanding and understand emotion identification through facial expressions. And it is an opportunity to prove the existence of the latest technological developments for human-computer interaction using Machine Learning and Deep learning [CITATION Moh14 \l 1033]. Recognizing and classifying specific human face Techniques are required, but the technique of deep learning outperforms other techniques by its great capabilities from various datasets and quick computing capabilities. Usually, the Face Recognition process and Classification requires various phases such as pre-processing, identification, guidance, extraction of Appearances, and emotion classification.

Classification of human emotions is defined as a method for defining human emotions by Facial features, vocal words, motions, and actions of the body and other physiological representations Evaluating signs. In today's environment emotion processing and understanding has a variety of Significance range in human-computer interaction, automated tutoring, image, and video Retrieval, adaptive environments, and vehicle alerting systems[CITATION Sho15 \l 1033].
Also, emotion detection plays a vital role for psychiatrists and psychologists identifying various mental health conditions. throughout recent years, scientists and researchers have developed many approaches and methods to recognize emotions from facial characteristics and character recognition. Because of the essence and its scope, it is also a daunting issue in artificial intelligence, computer vision, psychology, and physiology. Researchers and scholars believe that the most important element of understanding human feelings is facial expressions. Although it is difficult to interpret the emotions of humans by using individuals Characteristics of facial expression related to the exposure to ambient stimuli, such as illumination Conditions and Head Movement [CITATION Alr20 \l 1033].

The effectiveness of service robotics relies much on the user experience from a seamless system. Therefore, an AI would be able to retrieve knowledge just as of the expression of the

human, e.g. recognize the emotional state or gender. The use of machine learning (ML) techniques to accurately interpret each of these components has proved to be difficult due to the large sample variation within each task [CITATION Goo15 \l 1033]. This applies to the models being trained under thousands of experiments on millions of parameters [CITATION Amo16 \l 1033]. Also, the classification performance for a face image into one of seven different emotions are 65 percent 5 percent [CITATION Goo15 \l 1033]. The difficulty of this task can be illustrated by attempting to classify the FER-2013 dataset photos in Figure 1 manually into the following classes: "anger," "disgust," "fear," "happy," "sad," "surprise," "neutral." Given these challenges, robot systems designed to assist and solve household chores require precise and computationally efficient facial expressions. In comparison, state-of-the-art approaches are all focused on Convolutional Neural Networks (CNNs) in computer vision activities such as image classification [CITATION Cho \l 1033] and object detection.

Fig. 1: Samples of the FER-2013 emotion dataset [CITATION Goo15 \l 1033]

Literature Review

Almost all of the real-time emotion detection experiments are conducted on static pictures. The amount of facial expression detection work using computer vision technologies has seen a substantial increase in tandem using digital technology advancement as literature studies have been checked during the past two decades. While many applications are related to facial recognition and identification, in most of them, coevolutionary neural networks (CNNs) are not commonly favored for visualizing visual properties [CITATION LeH11 \l 1033]. Centered on the suggested categorization in Fig. 1, Description of the face recognition methods are listed below.

1. Centered on features: These techniques are used to identify differences in posture, point of view, or Inflammatory disorders. Such techniques are primarily meant for the localization of the ear.
2. Expression-based: This allows the use of a collection of several values with which the

templates or Known Models. Such values would reflect the facial representational variation. They instead use the trained templates for identification.
3. Knowledge-based (rule-based): These approaches use a collection of relationship-defining laws Among the facial features and reflect human intelligence Typical picture. It is used specifically for localizing the ear.
4. Template-matching: Methods of template-matching using similarity with an Input picture and facial features (full-face or facial characteristics) processed for Forecast.

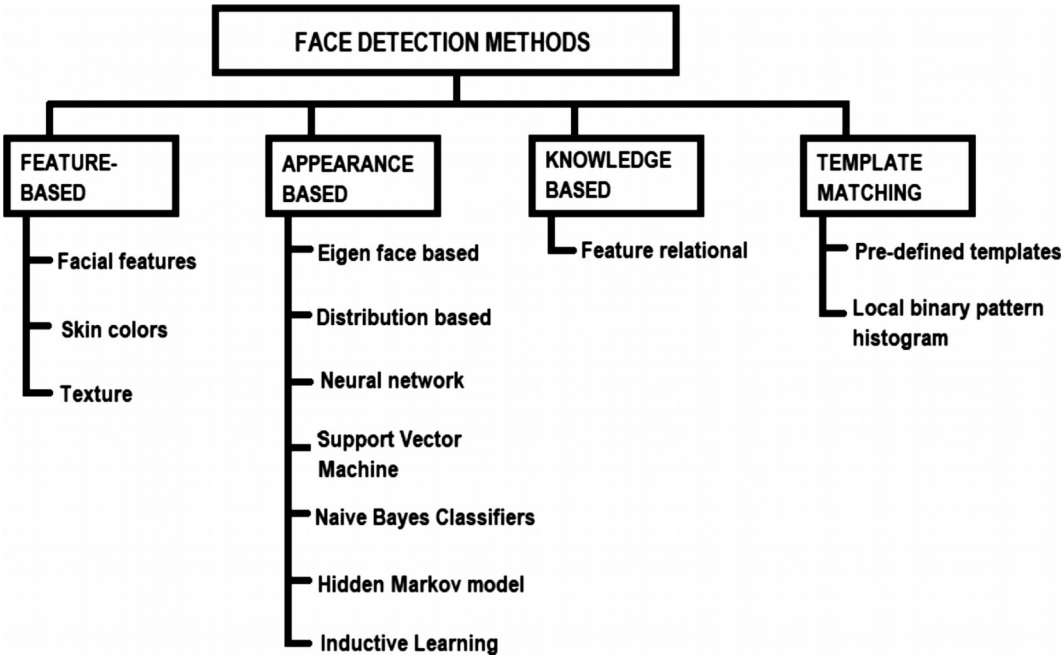

Fig. 1, Description of the face recognition methods

A few of the limitations of a knowledge-based approach is the construction of the correct collection The laws regulating face recognition are complicated. It cannot also identify different faces in various pictures. The methodology which matches templates is easy to use Run but not sufficiently for face recognition. On the other side, the interface is centered in Techniques that display a performance rate of 94 percent even on multi-faceted pictures. The appearance-based approach compared to other approaches, is easy and effective. Taking into account the benefits of feature-based and appearance-based approaches over the remaining approaches, they performed a literature review of the strategies used in each of these methods. Face recognition is in [CITATION Sat16 \l 1033] Usage of the neural network and classification with the help of the support vector Device (SVM). But faces of the same individual are not grouped in this system Right [CITATION Sat16 \l 1033]. The technique for lighting compensation which considers skin color as a function, Accuracy of about 90 percent [CITATION Jai02 \l

1033]. Faces associated with the segmentation of the skin and Consideration are granted to edge detection and satisfactory output Pictures all in color [CITATION Mol14 \l 1033]. The white comparison technique for photographs is used to achieve a 91 percent accuracy[CITATION Che13 \l 1033].
In [CITATION Lea17 \l 1033], the R-CNN system exceeded 96 percent precision. Confrontation recognition Is implemented using Neural Convolution Network (CNN) method prototype on the YTF dataset reaching 88 percent accuracy[CITATION LiS \l 1033]. A 96 percent precision is reached By applying a Histogram of Gradients to CNN on AFL and AFLW datasets Faces (HOG). The form is invariant pose[CITATION Sum12 \l 1033].

Grabor apps are used with PCA for a comparatively limited dataset [CITATION Sid18 \l 1033] An algorithm based on PCA that uses face recognition functions close to HAAR Was introduced at [CITATION DDa16 \l 1033]. There are two most important relational solutions Image Recognition. They are classed under machine learning and in-depth Techs for research. A comparative analysis of both methods shows the profound Learning is superior as it decreases the role of designing new extractor features for any problem. Deep learning strategies have proliferated with a multitude of Domains, in specific picture classification due to state of the art findings obtained in These domains [4].
In [CITATION Mar15 \l 1033] , they train on the JAFEE data set to identify 6 emotions And reached 87.53 percent accuracy. Multilayer dependent neural network Perceptron Was used for classification with k-nearest neighbors (k-NN) classifier 7 Emotions that demonstrated a substantial increase in precision hitting 96 percent [CITATION Tar17 \l 1033].

Commonly used CNNs for extraction of the functionality involve a series of layers at the end which are completely linked. Completely signed in In a CNN, layers tend to contain most parameters. VGG16[CITATION Sim14 \l 1033] comprises around 90 percent of the total The last completely linked layers have their parameters. The typical process pushes the network out of the input picture to remove global functionality. Modern CNN systems such as Xception [CITATION Cho16 \l 1033] benefit from the convergence of two of CNN's most common theoretical assumptions: the use of residual modules [CITATION HeK16 \l 1033] and the use of depth-wise separable convections [CITATION How17 \l 1033]. Separable from profundity Convolutions also minimize the sum of parameters by splitting the extraction processes and the mixture processes inside the convolution sheet.FER-2013 this model has achieved a 71 percent accuracy [CITATION Goo15 \l 1033] using Around 5 million criteria. 98 percent of all parameters in this design is found in the last linked layers. Using a collection of CNNs, the second-best methods described in [CITATION Goo15 \l 1033] achieved a precision of 66 percent.

[CITATION Las19 \l 1033] suggested a face recognition method utilizing HAAR cascades, standardization, and identification of emotions utilizing CNN on FER 2013 (KNN) for template scanning and Standardized Local Gabor Binary Pattern Histogram Series (ULGBPHS). Using KNN and SVM algorithms, the model used 4 separate machine learning strategies (SVM, KNN, Random Forest and Classification & Regression Trees) and strong precision values, i.e. 70 percent at 106 epochs. This concept can be rendered easier by still utilizing no algorithms in the machine language. [CITATION Kum17 \l 1033] create a stronger proposition. Here, the real-time identification of emotions using CNN with 9 layers to train and categorize 7 specific forms of emotions, providing an accuracy of about 90%.[CITATION Hos18 \l 1033] suggested utilizing a Gabor filter to make the network easy to access Focus on the face because the display orientations fit well with the facial wrinkles and in effect would be an input For CNN. The network focuses on useful features which offer an age-accuracy of 7% and 2 Role accuracy rate. established a model that integrated interrelationships that connect age and gender to attach such architectures to boost overall performance. The weaknesses were the difficulties in splitting the data into folds, training each classifier, cross-validating, and combining the resulting classifiers into a ready-to-test classifier.

Despite the notable performance of conventional methods of facial recognition by the extraction of handcrafted attributes, researchers have concentrated on the deep learning approach in the past decade due to its strong automatic recognition ability. [CITATION Mol16 \l 1033] placed forward deep CNN for FER through many datasets open. The photos had been reduced to 48x 48 pixels after removing the facial landmarks from the details. Then they used the data methodology for the augmentation. The design used consists of two layers of convolution-pooling, then introduce two modules of initiation forms, which include 1x1, 3x3 and 5x5 convolution levels. They give the ability to use the network-in-network approach , which allows the dynamically applied convolution layers to increase local performance, and this technique also makes it possible to reduce the question of overfitting.

[CITATION Mol161 \l 1033] To provide a clearer description of the emotions, examined the effect of pre-processing data before training the network. Data rise, rotation adjustment, cropping, 32x32 pixel down sampling, and strength normalization are the measures that were implemented. At the evaluation stage, the better weight obtained during the preparation period is used. This methodology was applied in three open databases: CK+, JAFFE, BU-3DFE. Scholars discover that it is more efficient to integrate both of these pre-processing measures than to execute them individually.

Such strategies of pre-processing, often applied by [CITATION Moh17 \l 1033]. They recommend a CNN novel for identifying facial AUs. They use two convolution layers for the network, one is preceded by two completely linked layers showing the amount of AUs enabled

and a max pooling. For the issue of facial occlusion [CITATION LiY18 \l 1033] present a new CNN process, first of all, the data inserted into the VGGNet network, then implement the CNN methodology with the ACNN mechanism of focus. This design educated and checked FED-RO, RAF-DB, and AffectNet in three-wide databases. (Yolcu et al,2019) suggested identification of the important sections of the face. They utilized three similarly built CNNs each detecting a part of the face including the nose, the ears, and the mouth. Before they put the images into CNN, they go through the crop stage and key-point facial recognition. The prominent face obtained in conjunction with the raw image was integrated into CNN 's second method to detect facial emotion. Researchers clarify that this approach is more successful than using raw images or facial iconisation alone (see Figure 2).

(Agrawal, 2019) performs a analysis using the FER2013 database on the variability in the effect of the CNN parameters on recognition levels. First, all images are all defined at 64x64 pixels, allowing for a variation in size and number of filters as well as the chosen type of optimizer (adam, SGD, ada delta) on a simple CNN consisting of two successive layers of convolution, the second layer plays the role of complete pooling, then a softmax. According to these experiments, investigators are designing two new CNN models with an average precision of 65.23 percent and 65.77 percent, the peculiarity of both models being that they don't have fully connected dropout layers, meaning that the same filter size remains within the network. (Jain, D.,2019) suggest a novel deep CNN containing two residual blocks, both of which contain four-component blocks After the pre-processing period, these model trains run on JAFFE and CK+ databases allowing the images to be cropped and normalized.

Experiments of facial expression modulation during emotional state (Kim,2019) suggest a spatio-temporal architect with a mixture of CNN and LSTM. Second, CNN measures the spatial features of facial expression in all emotional state frames followed by an extra LSTM to maintain the maximum spectrum of those spatial characteristics. And (Z. Yu, 2018) Introduce a novel architecture called Spatio-Temporal Convolutional with Nested LSTM (STC-NLSTM), this architecture focuses on three deep learning sub-networks, including 3DCNN for spatio-temporal extraction features, followed by temporal T-LSTM to maintain temporal dynamics, and then the C-LSTM for multi-level scaling.

According to these experiments, investigators are designing two new CNN models with an average precision of 65.23 percent and 65.77 percent, the peculiarity of both models being that they don't have fully connected dropout layers, meaning that the same filter size remains within the network. (Jain, D.,2019) suggest a novel deep CNN containing two residual blocks, both of which contain four-component blocks After the pre-processing period, these model trains run on JAFFE and CK+ databases allowing the images to be cropped and normalized.

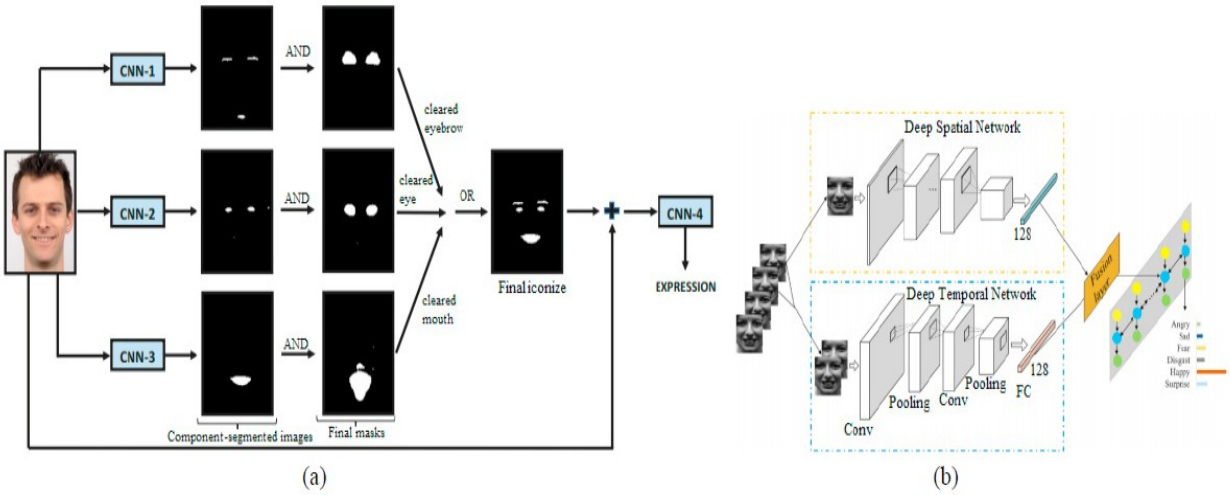

Fig. 2. Different deep learning methods from literature

(Liyanage, C.,2000) developed an algorithm based on rules to improve emotion recognition by using both audio and video data to handle emotions. They concluded that the encoding of video information is superior to audio awareness, while bi-modal approach provides favored effects over either visual or audio forms alone. (Chen, L., 1998) found that a number of the perplexities or overlaps between classes were totally unrelated to the two modalities of the same speaker's information sources. They proposed a rule-based algorithm that uses the adjunct property that can have higher recognition rates.

(Yoshitomi, Y. 2000) developed a multimodal framework, integrating both visual and audio information. They also noticed, in accordance with it, the thermal information gathered by the infrared sensor. This is done because the infra-red images, as in the case of normal cameras, are not susceptible to an effect on lighting. (Duan, Y., 2018) investigated the recognition of facial and vocal stimuli by the integration of the facial and vocal modalities. The system performs better particularly when it is tailor-made for each person.

Looking at the recognition in videos and images, there are more visible hints on the face than the amazing results recorded on various audio-visual labeling levels in the phrase (Ngiam, J., 2011). Sparse Restricted Boltzmann Machines (RBMs) were used for cross-modal computation, shared representation analysis, and multimodal fusion on the CUAVE and AVLetters datasets. TPCA findings hitting an identity rate of 43 per cent total data validity. A local binary invariant rotation descriptor, which first classifies into the binary rotational sequence, was given in Duan et al.[10].

References

D, D., Hudait, Tripathy,, H., & Das, M. (2016). automatic facial detection model from facial expressions. *ICACCCT*.

Learned-Miller., E., Huaizu, , & Jiang, . (2017). Face detection with the faster R-CNN. *IEEE International Conference on Automatic Face & Gesture Recognition*.

Li, S., Liao, S., Lei,, Z., & Dong,. (n.d.). Learning face representation from scratch. *arXiv preprint arXiv:1411.7923 (2014).*, (p. 2014).

Moldoveanu, F., & A., H. A. (2014). Human face detection from images, based on skin color. *International Conference on System Theory, Control and Computing (ICSTCC)* (pp. 532-537). IEEE.

Saty, M., Ludwiczuk, B., Brandon, & Amos. (2016). *Openface: A general-purpose face recognition library with mobile applications*. CMU School of Computer Science.

Alreshidi, Rahman, A., & Ullah, M. (2020). Facial Emotion Recognition Using Hybrid Features. *Informatics* , 6.

Amodei, Dario, Ananthanarayanan, S., & Rishita . (2016). Deep speech 2: End-to-end speech recognition in english and mandarin. *International conference on machine learning*, (pp. 173-182).

Cai, J. C. (2018). Facial expression recognition method based on sparse batch normalization CNN. . *37th Chinese Control Conference (CCC) (pp. 9608-9613). IEEE*.

Chen, , Zhen-Xue, Cheng-Yun Liu,, & Chang,, F.-L. (2013). Fast face detection algorithm based on improved skin-color model. *Arabian Journal for Science and Engineering*.

Chollet, & François. (2016). Chollet, F. (2016). Xception: deep learning with depthwise separable convolutions. *arXiv:1610.02357*.

Chollet, F. (2016). Xception: deep learning with depthwise separable convolutions. *arXiv:1610.02357*.

Goodfellow, J, I., Erhan, D., Carrier, L. P., & C, A. (2015). Challenges in representation learning: A report on three machine learning contests. *Neural Networks*, 59-63.

He, K. Z. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*.

Hosseini, S. L. (2018). Age and gender classification using wide convolutional neural network and Gabor filter. *International Workshop on Advanced Image Technology (IWAIT)* .

Howard, A. G. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. . *arXiv preprint arXiv:1704.04861*.

Jain, A. K., Hsu, Rein-Lien, & Mottaleb, M. A. (2002). Face detection in color images. *IEEE transactions on pattern analysis and machine intelligence* .

Kumar, G. R. (2017). Facial emotion analysis using deep convolution neural network. *International Conference on Signal Processing and Communication (ICSPC)*.

Lasri, I. S. (2019). Facial Emotion Recognition of Students using Convolutional Neural Network. . *Third International Conference on Intelligent Computing in Data Sciences*.

Le, & Hoang, T. (2011). Applying artificial neural networks for face recognition.

Li, Y. Z. (2018). Occlusion aware facial expression recognition using cnn with attention mechanism. *IEEE Transactions on Image Processing*, *28(5)*, *2439-2450*.

Marciniak,, Tomasz, & Chmielewska,, A. (2015). Influence of low resolution of images on reliability of face detection and recognition. *Multimedia Tools and Applications*.

Mohammadi, F. G., & Abadeh, M. S. (2014). Colony Based Feature Selection Algorithm Colony Based Feature Selection Algorithm. *Engineering Applications of AI*, 35-43.

Mohammadpour, M. K. (2017). Facial emotion recognition using deep convolutional networks. *IEEE 4th international conference on knowledge-based engineering and innovation (KBEI)* .

Mollahosseini, A. C. (2016). *In 2016 IEEE Winter conference on applications of computer vision (WACV) (pp. 1-10). IEEE*.

Mollahosseini, A. C. (2016). Going deeper in facial expression recognition using deep neural networks. *IEEE Winter conference on applications of computer vision (WACV) (pp. 1-10). IEEE*.

Nozaki, K., & HisaoIshibuchi . (1996). Adaptive fuzzy rule based Classification systems. *IEEE Transactions on Fuzzy systems*, *Vol.4*, *No.3*.

Shojaeilangari, Seyedehsamaneh, Yau, W.-Y., Nandakumar, K., Li, J., & Teoh, K. E. (2015). Robust representation and recognition of facial emotions using extreme sparse learning. *IEEE Transactions on Image Processing*, 2140-2152.

Siddharth , Pandey, Pathak, Manjusha, & Ram, A. (2018). Construing the big data based on taxonomy, analytics and approaches. *Iran Journal of Computer Science*.

Simonyan, K. &. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sumam, A. A., Chandra, L., & Paul,. (2012). Face recognition using principal component analysis method. *International Journal of Advanced Research in Computer Engineering & Technology* .

Szegedy, C. V. (2016). Rethinking the inception architecture for computer vision. . *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Tarnowski, P. K. (2017). Emotion recognition using facial expressions. *ICCS*, (pp. 1175-1184).