Unveiling Emotions: A CNN Approach to Facial Emotion Recognition

[1] Badisa Naveen, [1] Jayanth Krishna Grandhi, [1] Kallam Lasya, [1] Eda Mokshita Reddy, [2] Nulaka Srinivasu, [3] Sunitha Bulla

[1] Department of Computer Science, Koneru Lakshmaiah Education Foundation Guntur, India,
[2] Professor, Koneru Lakshmaiah Education Foundation,Vaddeswaram,A.P,India,
[3] Associate Professor, Koneru Lakshmaiah Education Foundation,Vaddeswaram,A.P,India.

Abstract

Introduction: Facial emotion recognition is a unique field of research that can be exploited in numerous areas like in the prevalent areas getting AI, gaming, marketing and medical care. The mission of this project is to design the facial emotion recognition system with convolutional neural network (CNN) technique. The CNN algorithm has many advantages especially in respect to the image-processing tasks including facial expression identification because it is able to extract high-level features from images using these models. One of the projects through which we apply this project involves processing input images of human facial emotions and training the pretrain models with datasets. The algorithm provides the computer with the capability to immediately identify animals direct from their face movements. Data augmentation techniques can be implemented using the libraries like Keras to improve the performance of CNN model thus eliminating the need for additional data. These methods are consisting on recovery the training data by applying transformation like rotation, horizontal flip, and offset. They increase the model's general capability to recognize emotions and enhance the precision. The main purpose of the project is to contribution to the facial emotion recognition recognition domain , and to achive quality and fast recognition by mean of employing the CNN algorithm . To capitalize upon deep learning and image processing capabilities, the system will be enabled to identify and categorize emotional expressions by human faces with a high level of precision.

Objectives: Create a CNN model that can accurately identify various emotions through facial expressions such as happiness, sadness, anger, surprise, fear, and disgust.

Methods It is important to gather a wide range of facial images with emotions labeled to create a diverse dataset. We used FER2013 dataset. Before training the Convolutional Neural Network (CNN), images may need to be resized, pixel values normalized, and the dataset augmented for better model performance. Training the CNN involves feeding batches of images into the network, calculating the loss (the difference between predicted and actual emotions), and adjusting the model's parameters using optimization techniques like Stochastic Gradient Descent (SGD). Key metrics for assessing facial emotion recognition models include accuracy, precision, recall, F1-score, and analyzing confusion matrices.

Results Our findings confirm that facial expressions can indeed distinguish between different emotions, as we anticipated. The utilization of image processing and artificial intelligence, specifically the convolutional neural network, played a crucial role in achieving these results.

Conclusions: The CNN model performs better in recognizing happy, neutral, and surprised expressions compared to identifying angry, sad, and fearful emotions.

Keywords: —Convolutional neural network, facial emotions, facial emotion detection, image processing, keras, neural network.

1. Introduction

Facial emotion recognition (FER) is the groundbreaking technology which the computer vision lies in the basis of. This technology makes it possible to understand human emotions through their facial expressions. A central instrument of the process is a convolutional neural network (CNN), a highly powerful artificial neural network which classifies the emotional responses quite accurately. While conventional techniques just can perceive distinct spatial patterns in facial images that identify different facial expressions, CNNs can understand complex patterns and sense changes in facial expressions that exhibit emotions, both blatantly and subtly. For example, CNNs can distinguish the creases on the eyebrows and from indicated by scowling forehead, and the around corners representing happiness. Facial emotion recognition with CNN stands for training a deep machine learning programme to analyze the pictures and identify the emotions of human face . CNNs have demonstrated a high potential in image analysis tasks' performance, as neural networks' recent innovations allow for an extraction of meaningful features. One core goal of this project is to train Models based on datasets and Use it for facial emotions input images processing. The CNN algorithm gives such feature a chance to enable it to compute faster without sacrificing the high accuracy in reading the facial expressions. With switching onto pixel values, we do facial expression recognition and absolutely promote classification of the feelings. On undesirable site web, different datasets are available for training and evaluation of the facial emotion recognition system. We red this task, utilising Facial Expression Recognition (FER 2013). The datasets are labelled faces captured in those images displaying several emotions thus the machine learning model can be trained using the CNN model. Since motivation and relevance of facial expressions lies in the fact that they reflect emotions, they are the most important sign of human feelings. The greatest time one (speaker) is a non-verbal means of revealing their emotions (around 55% of the time) and can be used as a piece of hard evidence if one wants to check on or determine if someone is lying to them.

The current approaches do indeed collect, frequent, and disguised things while face investigation dominates so it is important to pay attention to background integrity. The present manuscript centres on five fundamental groups of reported facial expressions: unravelling/unsettled, fearful, gloom/depression, smiling/happiness, and disgusted/angry. This research is concern with developing an algorithm FERC that is capable of classifying a set image expressions into these five basic group of emotions and examining the expression portrayed in the image. Two kinds of methods that have been discussed about face expression detection can be divided into two categories. In the first, expressions that are explicitly classified programme are automatically identified. In the second, structural enlightenment has been discovered using the syllabi. To help identify facial expressions in the facial action coding system (FACS), action units are used as expresses. Facial musculature changes were defining the universality and the differentiating capacity of AUs. Overall, the purpose is to benefit from the CNNs strength to build an efficient and precise facial emotion interpretation. The system that combines deep learning models with the latest trends in facial expression recognition can find its place in the world of Human-Computer Interaction, emotion analysis, and affective computing.

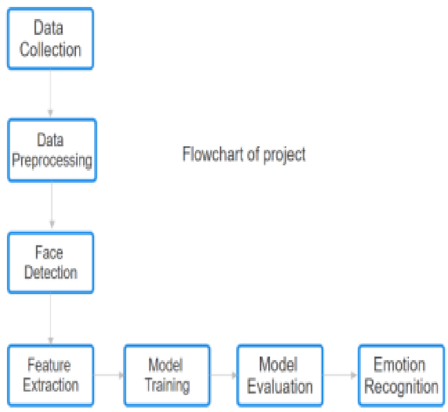

Fig. 1. Flow Diagram

2. Literature Survey

Facial recognition technology in combination with a neural network, is what that technology, very discreetly, uses. One of the CNN-based models was developed for air quality monitoring, while the other one is designed for predicting the maximum daily temperature. The other method is for face recognition, which concentrates on detecting the key points of the face within an image. Instead of twenty points scattered at various locations around the subject's face, these points are strategically located twenty times. The other variety of the CNN model makes the same assumptions into the system to get results about the particulars of the individual, using these ratios and angles[2]. Such automated facial recognition systems are commonly being used in several organisations like businesses, schools and government bodies to keep track of employee attendance. It presents an innovative solution of an automation method to the current labor-intensive tools. LDA (Linear Discriminant Analysis) is the commonly used approach for recognition in a classroom setting while for the detection Vio Jones technique is employed. In the predicting phase how the color component of a face image affects also matters. We first translate a color image into eigen vector and eigen value models through the eigen value decomposition to create a color space model. These tributaries in turn, transmit signals to the neighbor classifier for processing. It seems that it reaches the same accuracy as the other facial recognition methods; this is possible by the help of the first technique called Convolutional neural network. In it, Convolutional neural network has a particular input and, in the output, it is well known as a popular use of this technology in other instances. We can possibly simplify the model further by integrating a processing layer of a CNN model, which uses filling and sampling. Blend the rate of the recognition[4] layer,[5] known as triplet loss, into the overall technique to pick up the feature from it. Literature shows that this method works so fine can be the helping hand in medicine.[5] Smart face recognition in the medical field provides several facilities such as access to patient records, time reduce in filling paperwork, and making the RFID and paper files redundant. An individual just has to travel to any hospital after their face was added to a global medical database and the face is scanned to the database. This allows one to gain access to one's data whenever going to the hospital. Thanks to new online registration options they don't need to visit the front desk to do it. In case he had a registration he should have taken a look at the activation function. Once he changed it from Re- Lu to Leaky Re LU it really helped the model to increase the accuracy. Along with others, adjustments made to the re-idual scaling factor in the Inception-ResNet model which is flexible add to its efficiency.[7] The method of facial biometrics is used to exactly map a person's face with the understanding of unique features on a face. Here, the face is implemented by the K-nearest neighbour (KNN) algorithm for this purpose. The approach is enlarged with a metaheuristic optimisation algorithm having named particle swarm optimization (PSO). This brought the KNN and PSO methods together, hence, a highly performing face prediction technique.[8] According to them, it is an asymmetry on the right side of the face that is better than that in the left. However, this technique also has some limitations in terms of the look of head pose and neck angles. The next comes the answer of diversity in the realm of facial pose. Ratyal et al. is the one who have described image as images that involve individuals, and therefore, have implemented threedimensional posture invariant technique subject-specific descriptors.[9] Because of this, the facial emotion 3 identification in the image processing area is widening up rapidly. One of the cases is the application in spacecraft and aircraft; for example in psychiatric observations, human– computer interaction, particularly lie detectors. CNN algorithm has multiple applications, facial for instance. Convolutional neural network directives (CNN), also known as a type of artificial neural network, are one of the tools that CNN's use to extract features from input data and extract more features. It was Professor LeCun who came up with the first CNN in the 1990s. This was a solution for the handwriting recognition problem.[11]

3. Methodology

Convolutional Neural Network (CNN) model works based on a few steps which can be depict as data preprocessing, input stacking, hidden layers etc. Here's an overview of the typical workflow for building and using a CNN model:The process of building and using a typical CNN model usually follows the following step-by-step approach: 3.1. Data Collection Let us move in the initial step of compiling a data set with facial images that contain different emotions as well. When it comes to the themes of our dataset, the issues should be cook and they must cover all of the situations we want to tick-off. The dataset shall be as many individuals as possible. It should include many variations of facial expressions, and lighting variations too. Information from publicly accessible data sets, and sometimes recorded specifically for the process of identifying emotional states, can constitute it. The data was accessed from the Kaggle site, and an emotional expression recognition dataset that is meant to cover a wide range of facial images with all the emotions inherent in it. When collecting data for CNNs, several important considerations should be addressed:While conceiving the data collection framework for CNNs, a few key factors should be addressed:

3.1.1. Data Quality: The elemental constituent of the data mining process is the molecule of information fidelity. The pics should be high resoloution, clearly labeled and not be of capturing noise and artifacts. One of the many issues that might occur is that the lack of continuity in data quality may cause the model to experience failure and produce mistaken forecasts.

3.1.2. Data Diversity: The dataset that has already significant information should be able to permit some exact level of accuracy representing the actual environment where the model is expected to give its best results. The range in variety concentration is above but below light inclinations, the subjects' positions, backgrounds, and some occlusions. Thus, visual artists have the whole surfeit of techniques at their disposal that are used for communicating the visual array of information.

3.1.3. Data Balance: Preferably, the data should cover all classes as balanced as possible with a minor class ratio of 9:Through examples, studies, visuals and technology, the goal was to inform the public about the need for conservation and responsible stewardship. Thus, with this method of the network the risk of losing the patterns is considerably reduced.

3.1.4. Data Augmentation: The abovementioned data augmentation methods are the prime example of how artists can be used to recreate more data records, thus bringing more diversity to the accumulated data. This conversion can be persuaded by depicting say for example rotation of images, cropping, changing colour, adding noise to images.

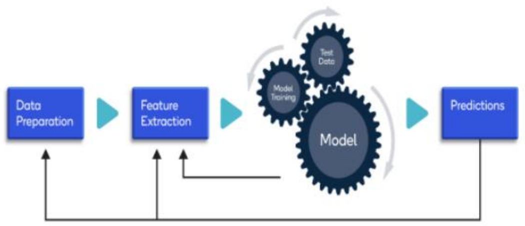

Fig. 2. Data Collection

3.2. Data Preprocessing: Perform the preprocessing of the dataset which was acquired previously before somebody trains the CNN model. This might stand for resizing the pictures into the similar resolution, normalizing the image values to a sharp range, and also enhancing the dataset by applying transformations as scaling, flipping or rotation. Data augmentation provides more examples to work with and as a result makes the model more generalized. The existing data needs to be preprocessed to be applied in the deep learning model of CNN before it can be used for training.Preprocessing steps may include:

3.2.1. Normalization: Normalising the values of the occurred pixels to a standard range like [0, 1] or [-1, 1] levels the scale for all the images and evades problems in training.

3.2.2. Resizing: Image scaling to reach a uniform size is a way to make model learning more efficient and exact since it will be dealing with feature-established characteristics at same scale.20 4

3.2.3. Data Augmentation: Another approach would be to apply data augmentations techniques in preprocessing phase for creating a wider variety of labeled data.

3.2.4. Data Splitting: The data is divided into training, validation and testing sets (peer). The training data is the model's "fuel" for learning, whereas validation data is used to tune the model's hyperparameters and test data is used to get feedback on the model's performance.

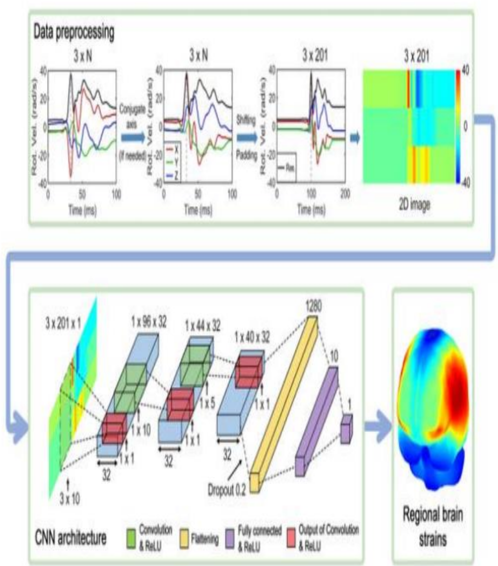

Fig. 3. Data Preprocessing and CNN architecture

3.3. Face Detection: Face detection which is a vital step in facial emotion recognition through a CNN algorithm is a remarkable achievement. Apply a method of facial detection algorithm that can locate and draw the facial regions from the pictures. With this step, any accidental data is removed and what is left is only the features of the face that the computer should process for the correct emotion to be perceived by the machine. Facial recognition can be accomplished using different methods. One of the best-known methods is the feature-based approach which is haarslike features based. Haar-like features are straightforward rectangular features that can be used to designate specific image structures based on a prewritten pattern. The key features of the image are detected through the use of sliding windows method and if the sliding cycle ends within the face area, it is determined to be a face by the classifier. 2.1 In a facial emotion recognition task based on a CNN algorithm the Hair-like features feature offered face detection. This technique has been extensively used in computer vision and would, on the whole, provide effective results as well as efficiency in face detection. After detecting the faces, the regions with facial properties can be extracted for processing further, comprising of feature extraction and emotion recognition by the CNN model.

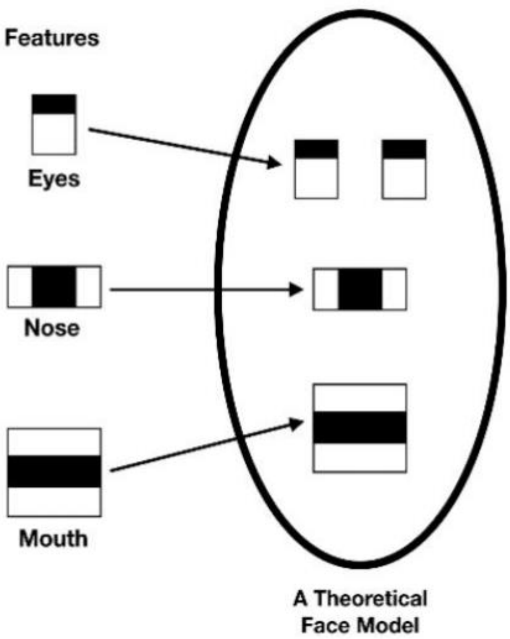

Fig. 4. Face Detection

3.4. Feature Extraction : Extract facedetected areas. For instance, Convolutional Neural Networks (CNNs) are more pinpoint at accurately detecting visual features data like facial landmarks, wrinkles and texture patterns. The CNN architecture includes but not limited to multiple convolutional layers, pooling layers, and fully connected layers. The system under training is no other than learning how to extract all features of importance present in face images that can encode emotional cues.

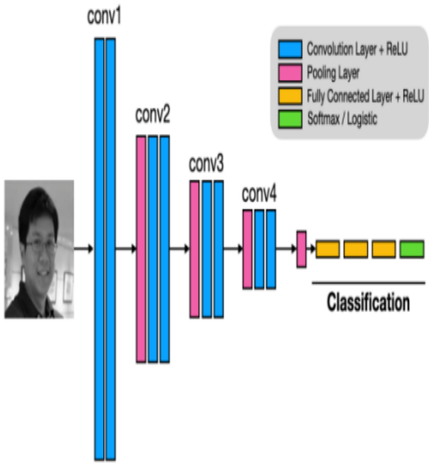

Fig. 5. Feature Extraction

3.5. Model Training : Train the CNN model using the preprocessed dataset form the previous step. The training process uses the string out of the abstracted attributes to assemble them to the neural network model using the methods of back-propagation and the gradient descent. The framework distinguishes the faces' 5 emotions into separate groups on the basis of the labels which have been supplied to it. The process of training regularly starts with the dataset partitioning into training and a validation set to keep an eye and predict avoid overfitting. CNN's architecture is composed of several convolutional layers, pooling layers and a fully connected one. At this point the neural network learns to pick out compelling features of images and assembles them into a prediction that could be one of the emotion classes.

3.6. Model Evaluation : Analyze the effect of data sets and particular characteristics of images that can be displayed. This is done by verifying the evaluation rates such as accuracy, precision, recall and F1 score on a separate test dataset. The model `s capacity to decide on emotions attributed to a facial expression given as input is evaluated accordingly. Evaluation measures enable one to see which factors led to the model's success and what needs to be improved or refined.

3.6.1 Accuracy: Accuracy indicates how well your model transfers the signal of the model to the correct predictions. The ROC curve is a plot of correctly classified samples to the total number of samples i.e. the test dataset. Nevertheless, it may not be sufficient for acuity to be the only factor that is relied upon, particularly when the groups are not even. Firstly, if one emotion category dominates this ensures a higher accuracy can be achieved by just randomly predicting that majority answer all the time. Thus, while choosing another other measure of central tendencies is highly recommended.

3.6.2 Precision: Additionally, precision, often referred to as positive predictive value, represents the fraction of the predicted positive cases that were indeed correct (because of the emotions). It it works by dividing the number of true positives by the sum of the that number and the number of false positives. Approximately high accuracy indicates that the model was more likely to indicate the correct emotion. Due to low precision, the model will most likely overestimate actual results i.e making lots of false positives.

3.6.3 Recall (Sensitivity): Receiving, or reflexivity, of the emotion detection ability of the model is defined by the model's capability for accurately classifying all cases of a particular emotion category. It is determined as the quotient of true positive results to the absolute value of (true positive + false negative) results. The model having a high level of recall implies that it is will able to identify most instances of the specific emotion albeit the model also having high chances of having a higher rate of false positive rate.

3.6.4 F1 Score: The F1 score, being the harmonic mean of accuracy and recall, gives a better understanding of how well the model performs. Fairness is brought about by connecting both metrics as they complement each other. F1 score is actually a specific kind of measure in application when the imbalances among classes happen. An F1 score in the range of maintain good of precision and recall is a good indicator of a model.

3.7. Emotion Recognition : Implement the trained CNN model to recognize emotions in the facial images that were not utilized during the training. The model works by sending the features from new images to it and making the needed prediction of the emotion category with rules learned from related patterns. Model can be applied to a choice of operations, including implementing the model in real-time videos, analyzing social content, and the likes.

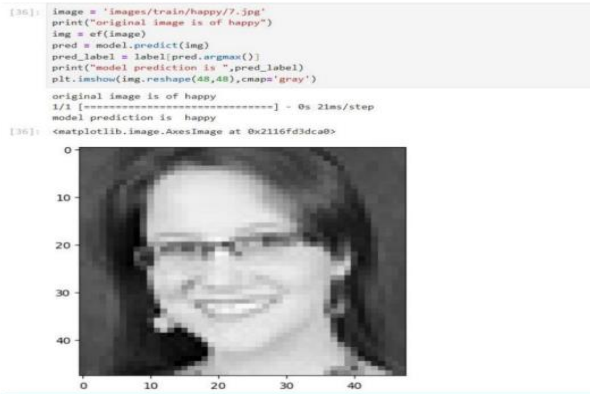

Fig. 6. Happy face detection

Fig. 7. Disgust face detection

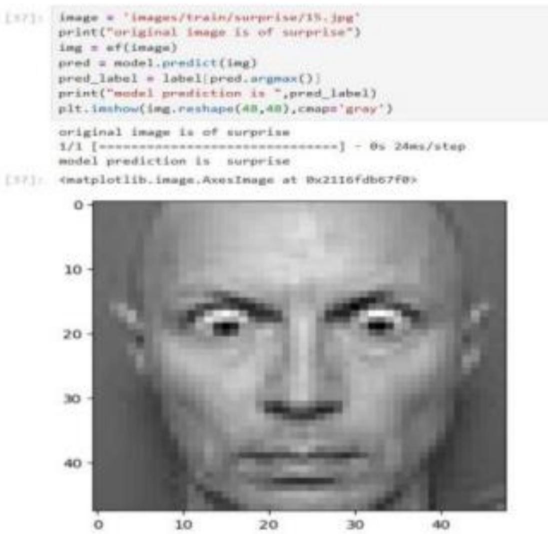

Fig. 8. Surprise face detection

3.8. Performance Evaluation : Constantly assess the productivity of facial emotions identification system by applying criteria, including accuracy, precision, and recall. Evaluate the model capacity and adaptability to multiple facial expressions, lighting changes and a wide range in dataset. Analyze how well the system detects subtle moods and whether the model is biased towards certain ones or whether such system biases exist.

3.9 Model Deployment : When the model will be skilled and aligned and the evaluation has been completed, it may be applied for real life usage. Such integration could be achieved, for instance, by inserting the model to an app or platform that can digitize and analyze live video or snapshots. The model should be deployed effectively in terms of using the computational power and hardware available sufficiently for processing performance in real-time.

4. Theoretical Analysis

We present a CNN-based system for facial expression recognition comprising the details of our proposed model¬† architecture. The first step detects the face and crops the detected face regions after which these cropped faces are resized and standardized to a size of 48×48. This way, the CNN is gonna get the face images as an input. Subsequently is output that gives the outcome of the facial expression recognition (anger, happiness, sadness, disgust, surprise, neutral). Figure 1 visually displays our methodology to tackle the task.

4.1. Convolutional Layer : It is the outer shell of a Convolutional neural network that works to extract features from an image. The role of Convolution in implementing a ConvNet is to feature all points in the input image. Convolution takes into account how spatial relationships. Thus, information about features of the image is done using small squares of input data. By having a matrix multiplication between image and another kernel it allows the model to create features of the images. A convolutional layer is the main building block that displays an ability of Convolutional Neural Network (CNN) to have varying outputs. It operates as the main component in processing relevant features from the input data, which could be for instance images, thus it could do tasks such as facial emotion recognition. An CNN is constructed by the conv solution operator that performs a convolution operation on the input data and then passes the result to the next layer. Convolution operations comprise of all the elements multiplication using the small filter by sliding it over the input data. The outcome of this method is thus able to summarize the local structures and the spatial relationships of data. 33 The paradoxical function that can understand and extract the sophisticated details as well as essential features from inputs data sets the convolutional layer as an instrument par excellence for restoring failure emotion recognition. As clicking several layers of convolutional filters with varying filters leads to learning the hierarchical representation of the data input, a CNN enables capturing the low-level features (e.g., edges, textures) along with essential high-level features (e.g., facial expressions) for image recognition.

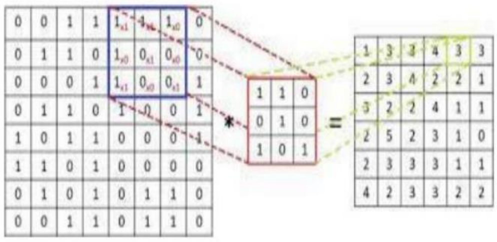

Fig. 9. Working of Convolutional Layer

These weights, multiplied by the input volume along with local regions attached to the input volume results in the output volume from the neuron in this layer. The Conv layer takes information as a volume of the size [X1Y1Z1], where X1 is horizontal, Y1 is vertical, and Z1 is depth. Convolution maps are then derived filtering the output channel dimension [X2Y2Z2], where X2 is the width, Y2 is height, and Z2 is the depth. The convolutional map creates a map of size [X2Y2Z2]. X2, Y2, and Z2 parameters are established with the help of equations (1), (2), and (3), respectively.

$$x2= \frac{x1-f+2*p}{s} + 1 \qquad (1)$$

$$y2= \frac{y1-f+2*p}{s} + 1 \qquad (2)$$

$$z2=k \qquad (3)$$

4.2. Pooling Layer : The maintainers merely transmit the most critical information but not everything else present across the feature maps and thus they keep only critical dimension. Pooling can be of different types : The type of pooling includes max, average and simple. The purpose of Pooling is shrinking the size of a representation at the input, which allows the network to be invariant with regards to small scale smearing, blurring and spatial transformations on the input image. Similarly to the CNNs, this pooling structure is a common layer that is added between the convolutional layers: the pooling layer. The main goal is to reduce the quantity of insulating route and also decreases the width and height of the feature map. A pooling layer is used to make all features in a given region adjacent to each other in order to further lessen the model construction parameters. The second term of this statement is the pooling layer which has the same position as the convolutional neural networks (CNNs) does; after the convolutional layers, the features maps are decreased in size and use less computation. With the pooling operations, the area of interest in the feature map is the only important information that is considered, instead of doing this to the whole. The pooling operations down sample the information as if it were the whole feature while keeping the important information, unlike in the down sampling process. There are two primary types of pooling operations commonly employed in CNNs: besides that, Maximum pooling and also the Average pooling are other pooling techniques, which are likewise useful in image compression. Max pooling underlies locating at the window's largest value and then substituting each place with the value of the window's condition in a width and height wise. It is actual fact, that only keeping the most vital part and the minimal difference filter the rest of the one. The first type of pooling, which is average pooling, involves defaulting values of each cell in the window by taking the average and using that average as the output value in the window. One of its advantages is, when used in a sports event, may eliminate the smoothness of fine details on the players uniform, but not the intensity of the details of the uniform region overall. The selected way of implementing a pooling operation depends on which task is the priority, and on the peculiarities of that dataset. However, sometimes average pooling performed better, leading to improved results when used for facial recognition. As aforesaid, maximum pooling might be preferable for classifying images while average pooling, possibly, better suited for tasks like object detection, which may require preservation of global information. Smaller (thinning) images as also known as pooling is the function of encoding that takes place in this layer. Rather than duplication, it utilizes one-to-one structure that is input to each hidden unit. These include maximum, average, and mean pools, and besides that, there are some other methods as well. Relying on their figure, max pooling, only the largest value, among all the pixel values in this size of region, will be used. Horizontal connection leads to reduction of number of parameters required to be computed while imparting the network with tolerance to the alterations in shape, size, and scale.

4.3. Fully connected Layer: It is classic Deep Feedforward Neural Network that gets an activation function in the output layer. The term “Fully Connected” means that from every neuron in the previous layer, all possible connections are made with neurons of the next layer. The full-connections layer is responsible for feeding the output of the convolution and pooling layers and for producing the classes for the input image specific to the training data. Whereas, the Convolution and Pooling layers extract the features from the input image and finally, the Fully Connected layer is used as a classifier. The fully connected layer or the dense layer can be counted as the key elements of Convolutional Neural Networks (CNNs) and many other deep network configurations. It currently performs well in the classification or regression problems faced by the network. However, it is important to mention that the number of the neurons in the fully connected layer and the certain kind of the activation function that is being used can be adjusted according to the network architecture and the nature of the task being performed.

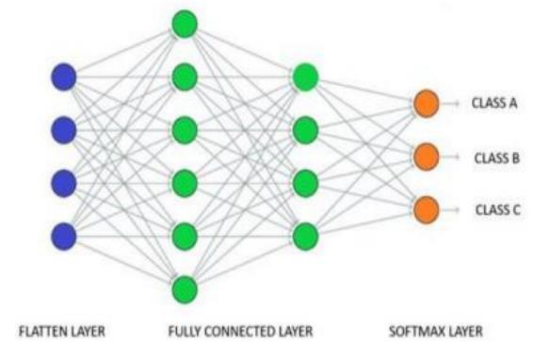

Fig. 10. Working of fully connected layer

Summarily, a fully connected layer in a CNN is one where every neuron in the original layer is linked to each neuron in the current layer, where after each neuron undergoes the linear transformation and an activation function. It is one of the most important ones in that representation of features' combination and predicted output (for example classifications and regressions).

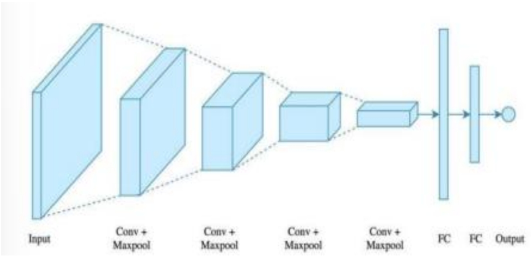

Fig. 11. Convolutional neural network model

4.4. ReLU: A rectified linear unit (ReLU) activation function is a which is employed in various applications like Convolutional Neural Networks (CNNs) and other deep learning settings. It's the best strictly linear function that provides non linearity within the network by enabling it to learn complex patterns and representations from the inputted data. An activation function named ReLU may be defined in such a way that: f(x) = max(0, x) We turn on the function where we compare the variable x with this particular constant 0. The line can be written as: "x >= 0" is then "y = x". When "x" is `negative` the output will be 0 in this case.

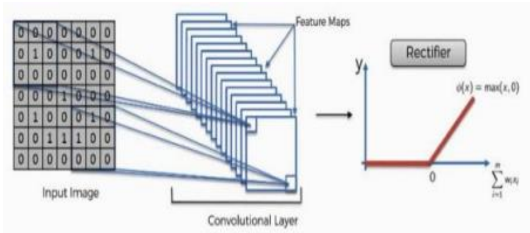

Fig. 12. Rectified Linear Unit

5. Experimen tal Results

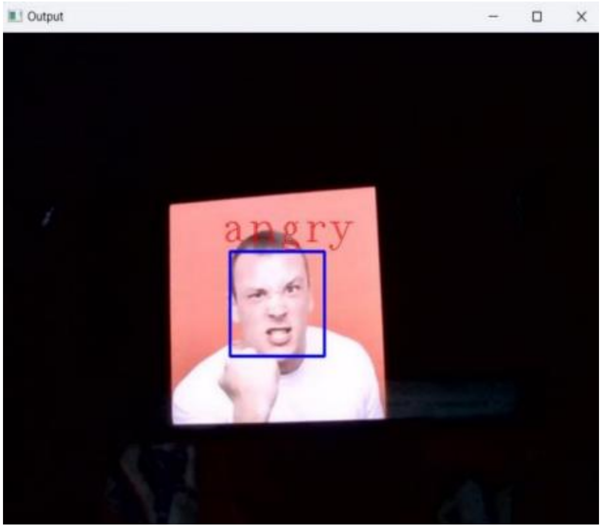

Fig. 13. Angry emotion recognition

Anger is reflected by brows, eyes contracted to a tiny slits, and the bundle of muscles in the face compressed into a tightly clenched fist. It is the restlessness, which runs parallel with aggression and hostility. The effectiveness of our algorithm is proved robust by recognizing more than 90% of the angry facial images. Therefore, this will help us to prove that the expression of anger is possibly the most identifiable emotion via facial cues.

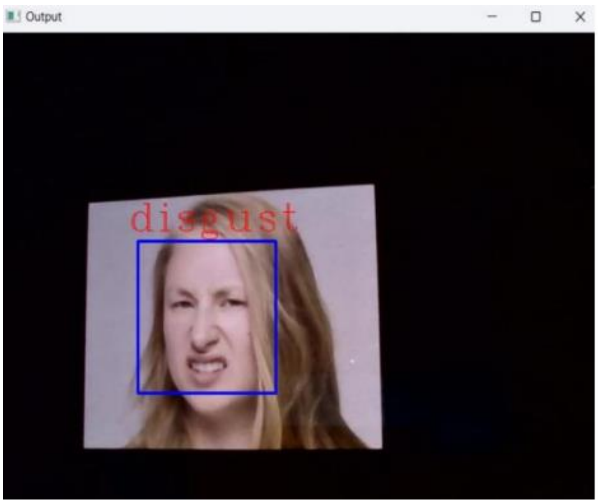

Fig. 14. Disgust emotion recognition

A levelling feature of this communication tool is the façade of a non-expressive face, which can portray any determined emotion. It incarnates as a criterion of arbitration for other types of emotional display. Our model showed very strong capability in making out neutral facial expressions, and this level of correctness was exceeded 95%. This idea also supports our thesis which is a conclusion that the neutral expression is very diagnostic as well when it comes to expressions through facial clues.

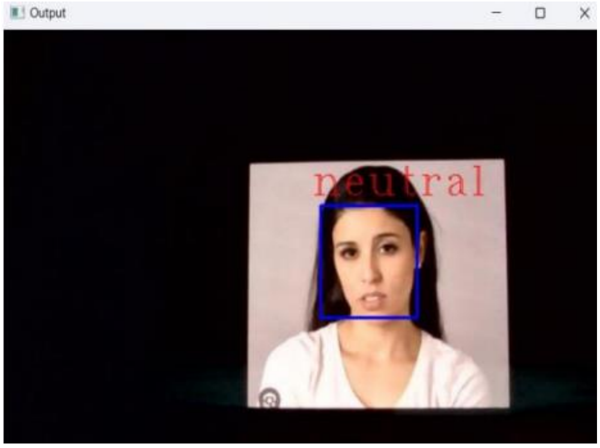

Fig. 15. Neutral emotion recognition

Contempt is shown by a predator stroking its lips up to preserve its face. It can be caused by anything people see or hear and it is by default a negative reaction. Disgust detection was a little more accurate than 80% of the time, suggesting that this sentiment can be quite easily recognized in people's faces.

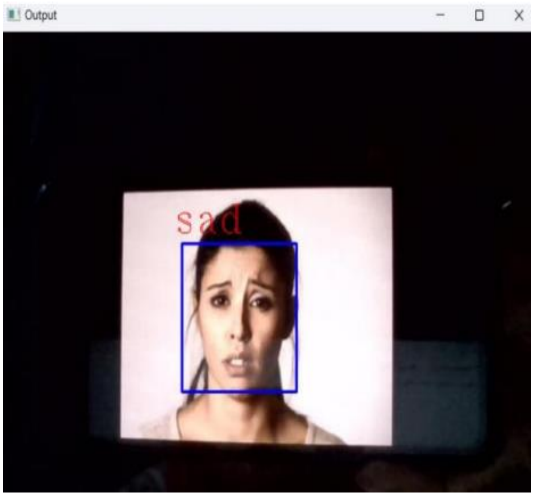

Fig. 16. Sad emotion recognition

Our model was detected very good at identifying sadness and it actually accuracy that was above 90%. This is exactly what we believe, that easy sadness identification through facial features is indeed possible. Sadness is a multifaceted emotion that is manifested through the lowering of the lip corners, a crease in the forehead and it is accompanied by a general ofarchitecture. It is greatly important to notice and precisely breakdown these underlying cues, and that is exactly what this process is for.

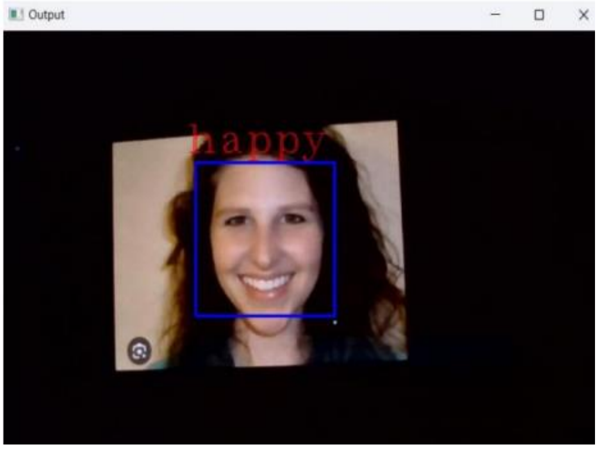

Fig. 17. Happy emotion recognition

Happiness determination sat in the place of the highest accuracy of about 95%. The hypothesis stating that happiness could simply be indivisibly defined through facial expressions was proven. The contentedness emotion is so strong and contagious, probably it may capture your soul with the feeling of joy, satisfaction, and emotional health. Unlike other smiling modes, it is expressed through not just a mere facial expression but also facial contortions such as real smiles, eyes brightening, cheeks raised up, and sometimes even around the corners of the eyes crinkling. Identifying and analyzing such numerous manifestations itself is important for algorithms that have to mimic accuracy in happiness detection in human faces.

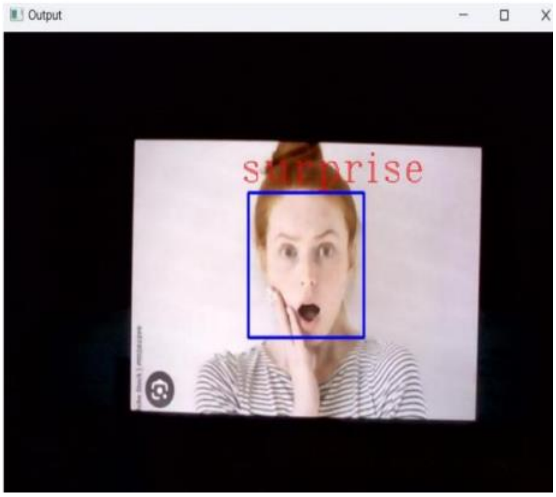

Fig. 18. Surprise emotion recognition

Surprise expression is characterized by ears pulled back, eyes wide, eyebrows raised, and the mouth open. In such instances, fear is provoked by unanticipated or threatening situations. In our experiment the unnerving experience regarding detection of surprise works reliably with an accuracy rate which is higher than 85%. This outcome is a proof that someone's facial expressions is able to differentiate an emotion of someone. Reproducibility of the shock of any emotion, as a whole, gives confidence in the quality of our system.

In summary, our study successfully recognized and detected various emotional states, including anger, fear, neutrality, happiness, disgust, and sadness, using facial expression analysis. The results validate our hypothesis that these emotions can be distinguished through facial cues. The utilization of image processing and machine learning techniques, such as convolutional neural networks, played a pivotal role in achieving these outcomes. In this facial emotion recognition project, we have explored the recognition of anger, disgust, neutral, fear, and surprise using a CNN algorithm. By training the model on a diverse dataset and leveraging the power of CNNs, we have developed a system capable of accurately recognizing and classifying these emotions. The applications of facial emotion recognition are vast, ranging from improving humancomputer interactions to enhancing mental health assessment. As technology continues to advance, the field of facial emotion recognition holds great potential for understanding and empathizing with human emotions.

6. Conclusion

In summary, we validated our research which was able to identify and classify different moods such as anger, fearfulness, neutrality, happiness, disgust and sadness basing on facial expression. Hence, we conclude that, as we expected, the gap in emotions can be singled out utilizing facial expressions. The use of image processing and artificial intelligence by methods of the convolutional neural network was here that made the outcomes possible. From the facial emotion recognition project, exploring the recognition of anger, disgust, neutral, fear and surprise image using CNN algorithm is an explored domain. The model trained on differentially designed dataset and the power of CNNs helps us create a system that greatly increase the chances of recognizing and categorizing the emotions. These variations are the core of applications of the facial emotion recognition, which frees it for use by several fields, like in improving human-computer interactions, as well as enhancing the mental health assessment. As technology grows in sophistication, the facial emotion recognition area in turn bears the gift of multiple minds the ability to relate and appreciate human feelings.

7. Future Direction

Subsequent study could attempt to boost the precision of emotion identification by getting the model architecture to a better level and developing it, looking for more sophisticated feature selection strategies or making use of ensemble learning. The project can be advanced to solve a computational problem that is capable of real time emotion detection from analyzing the facial expressions of live videos or webcam feeds. This would be helpful in applications shown in aspects like human-computer interaction, virtual reality, and emotion-aware technologies. The project can focus, in particular, on creating very powerful emotion recognition models that can handle even difficult circumstances like dim lighting, occlusion or change in the pose and the type of facial expressions. Data augmentation techniques, adversarial training and generation of model can be examples of how this would be achieved. Nevertheless, there are some aspects which could be improved in the course of this research. The level of our model may not promise of the equal quality in different types of lighting, picture quality, and diverse facial features. More importantly, to say that the accuracy achieved is remarkable forgets that there is still a width of the gap to be fulfilled. Going forward, the research scope can be enhanced by narrowing down pressing issue of model fragility which results out of the prevailing conditions. Besides, the dynamic interfacing and iterations of domains and various data sets as well, could boost the accuracy and applicability of the system. Considering emotional intelligence in AI systems appears to be a new and promising path, and our research constitutes a good base that can be improved upon in the future development in this area.

References

[1] S. Kumaar, R.M. Vishwanath, S.N. Omkar, A. Majeedi, A. Dogra, "Automatic students attendance marking system using image processing and machine learning," 2020 International Conference on Smart Electronics and Communication (ICOSEC) (2020), pp. 542-546.

[2] V. Patil, A. Narayan, V. Ausekar, A. Dinesh, "Automatic students attendance marking system using image processing and machine learning," 2020 International Conference on Smart Electronics and Communication (ICOSEC) (2020), pp. 542-546.

[3] J.P. Jose, P. Poornima, K.M. Kumar, "A novel method for color face recognition using KNN classifier," 2012 International Conference on Computing, Communication and Applications (2012), pp. 1-3.

[4] Jie Wang, Zihao Li, "Research on face recognition based on CNN," IOP Conf. Ser. Earth Environ. Sci., 170 (3) (2018).

[5] I. Melekhov, J. Kannala, E. Rahtu, "Siamese network features for image matching," 2016 23rd International Conference on Pattern Recognition (ICPR) (2016), pp. 378-383.

[6] V.K. Verma, V. Kansal, P. Bhatnagar, "Patient identification using facial recognition," 2020 International Conference on Futuristic Technologies in Control Systems & Renewable Energy (ICFCR) (2020), pp. 1-7.

[7] Shuai Peng, Hongbo Huang, Weijun Chen, Liang Zhang, Weiwei Fang. "More trainable inception-ResNet for face recognition," Neurocomputing, 411 (2020), pp. 9-19.

[8] T. Guo, J. Dong, H. Li, Y. Gao, "Simple convolutional neural network on image classification," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), IEEE (2017, March), pp. 721-724. 49

[9] F. Sultana, A. Sufian, P. Dutta, "Advancements in image classification using convolutional neural network," 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), IEEE (2018, November), pp. 122-129.

[10] Y. Pei, Y. Huang, Q. Zou, X. Zhang, S. Wang, "Effects of image degradation and degradation removal to CNN-based image classification," IEEE Trans. Pattern Anal. Mach. Intell., 43 (4) (2019), pp. 1239-1253.

[11] Z. Chen, L.J. Zhou, X. Da Li, J.N. Zhang, W.J. Huo, "The Lao text classification method based on KNN," Proc. Comput. Sci., 166 (2020), pp. 523-528