OPEN ACCESS

International Journal Of Modern Engineering Research (IJMER)

An Automated System for Facial Recognition and Emotion Assessment Utilizing Machine Learning Techniques

MHYMOOD KHAN[1], PATHAN IRUFHAN KHAN[2]

#1 M.Tech Scholar (CSE), Department Of Artificial Intelligence & Data Science,
#Associate Professor, Department of Artificial Intelligence and Machine learning in Kakinada Institute of Engineering and Technology-II, AP, India.

Abstract: *Human emotion recognition using machine learning is a new field that has the potential to improve user experience, lower crime, and target advertising. The ability of today's emotion detection systems to identify human emotions is essential. Applications ranging from security cameras to emotion detection are readily accessible. Machine learning-based emotion detection recognises and deciphers human emotions from text and visual data. In this study, we use convolutional neural networks and natural language processing approaches to create and assess models for emotion detection. Instead of speaking clearly, these human face expressions visually communicate a lot of information. Recognising facial expressions is important for human-machine interaction. Applications for automatic facial expression recognition systems are numerous and include, but are not limited to, comprehending human conduct, identifying mental health issues, and creating artificial human emotions. It is still difficult for computers to recognise facial expressions with a high recognition rate. Geometry and appearance-based methods are two widely used approaches for automatic FER systems in the literature. Pre-processing, face detection, feature extraction, and expression classification are the four steps that typically make up facial expression recognition. The goal of this research is to recognise the seven main human emotions anger, disgust, fear, happiness, sadness, surprise, and neutrality using a variety of deep learning techniques (convolutional neural networks).*

Keywords: *Emotion Detection, Convolution Neural Network (CNN), Face Recognition.*

I. INTRODUCTION

In contemporary technology, human-machine interaction is growing in popularity, and robots now need to understand human movement and emotions [1]. When a machine detects emotion in people, it may understand human conduct and alert the person using it to the subject's emotions, which increases productivity. Strong feelings known as emotions [2] have an influence on a variety of day-to-day tasks, including dealing, comprehending, organising, thinking, and remembering, concentration, inspiration, and much more [3]. The branch of signal processing known as image processing [4] works with pictures as input and output signals. Image processing is essential for the identification of facial expressions. Our facial expressions convey the feelings we are experiencing [5]. In interpersonal communication, facial expressions are essential. Our faces convey nonverbal scientific signals based on our emotions, which is known as facial expression. This period has made automatic facial expression recognition necessary, since it plays a critical role in the domains of robotics and artificial intelligence. Personal identity and access control, videophone and teleconferencing, forensic software, human-computer interaction, automated surveillance, cosmetology, and other applications are among the several applications that are related to this. This research aims to develop an autonomous face expression recognition system. It will classify and analyse human face photos [6] with diverse expressions [7] into seven distinct groups. Emotion recognition is the process of identifying a person's feelings. People vary widely in how well they are able to discern the emotions of others [8]. The application of artificial intelligence and deep learning to assist humans in identifying emotions is a relatively recent field of study. Since ancient times, researchers have been interested in autonomously recognizing emotions [9]. Currently, emotion identification is achieved by analysing social media data, analysing speech in audio recordings, and identifying facial expressions in photos and videos [10, 11]. In the realm of computer science, machine learning is an emerging technique that is very important. That is intended to have a 95% impact over the next three years. A kind of computer learning called "deep learning" [11] uses artificial neural networks, or algorithms modelled after the structure of the human brain. One kind of deep neural network that uses convolution as the mathematical process is called a Convolutional Neural Network (CNN) [12, 13].

II. RELATED WORK

A CNN (Convolution Neural Network) model was suggested by Y. Ma and Cao, G. et al. [14] to classify brain signals and recognize human emotions and expressions from the Electrocardiogram (ECG) dataset. On tests, the system offers an accuracy of about 83%. In order to identify a person's emotions, Shrey Modi and Mohammed Husain Bohara [15] employed a CNNbased emotion detection system that combines a fully connected output layer, max pooling and feature map layers. The ability to identify emotions in real time can be used to predict future terrorist actions in people.

Combining facial expressions with electroencephalography (EEG) helps enhance the recognition of emotions. EEG is a tool that may be used to assess brain electrical activity, which can provide insight into an individual's underlying emotional state [16]. When producing material, such suggestions or ads, the user's emotional state might be taken into consideration. Emotion detection is a feature that health and wellness apps may use [17] to provide feedback on stress levels and suggest mindfulness or relaxation exercises. In education, the level of student interest in the classroom may be observed. The technologies can identify people who are hostile, irate, or irritated. Then, such information might be used to take action prior to such individuals committing crimes. Artificial intelligence [18] technologies provide criminals feedback on their behaviour and appearance so they can learn to control their emotions [19]. A foundational study on automated picture categorization in general was published by Krizhevsky et al. [20]. This research demonstrates a deep neural network with functionality similar to that of the human visual cortex. Using the CIFAR-10 dataset and a self-developed labelled array of 60,000 images divided into 10 classes, a model for categorising objects from pictures is obtained. Another significant outcome of the research is the visualisation of the filters in the network, which allows the model to be evaluated in terms of how it breaks down images. F. Zhou et al [21] proposed a deep convolutional neural network model used to recognize ships in the movement for Polarimetry Synthetic Aperture Radar (POLSAR) Images. This model utilizes a Faster Region based Convolutional Neural Network (FRCNN) method to find a ship of various sizes. NASA/JPL AIRSAR dataset used by validated the model. Machine learning (ML) is the systematic study of algorithms and statistical models that computer systems use to do tasks without explicit instructions by relying on patterns and inference rather than precise instructions. ML algorithms [22] construct a mathematical model based on samples of data, referred to as ‘‘training data,’’ to generate predictions or judgments without the need for explicit programming Zhang [23]. ML is a sub-field of artificial intelligence (AI). It is the most popular dominant technology today because it helps automate complex problems. Its basic idea is to create models that learn the relevant features of a dataset to make accurate predictions. It is used by most popular apps, such as Facebook and Netflix, to predict which advertisements to show and which TV episodes a user will enjoy [24]. The study by Jung et al. [25] proposed a deep neural network architecture that combines visual and audio features for facial emotion recognition [26]. The authors extracted both visual features from facial images and audio features from speech signals, and fused them using a late fusion approach. The proposed model achieved an accuracy of 91.2% on their dataset, which contained images of facial expressions representing seven different emotions. The study by Kim et al. [27] proposed a 3D convolutional neural network (CNN) architecture for emotion recognition from facial expressions. The 3D CNN captures spatiotemporal information from facial image sequences, allowing for modeling of temporal dynamics in facial expressions. The authors achieved an accuracy of 89.5% on their dataset, which contained video clips of facial expressions representing seven different emotions. The CNN is a type of deep neural network that uses the mathematical function convolution, which can be understood as multiplying two functions. For this, the network uses convolution filters. These filters are matrices, usually square, which are used in convolutional layers [28]. There are several emotion models or approaches to emotion detection [29] such as (1) the basic model (categorical approach), where a small number of basic emotions are defined; (2) the dimensional feeling model (dimensional approach) describing feelings according to more generally, but practically mainly only according to two dimensions – the first from pleasant to unpleasant feelings, and the second from excitement to apathy; (3) the componential model of appreciation, which tries to detect emotions from evaluations or interpretations of a speech, text, or events.

III. Problem Identification

Human facial expressions may be easily categorized into seven basic emotions: neutrality, anger, contempt, surprise, fear, sorrow, and happiness. The activation of particular facial muscles is how we show our emotions on our faces. The signals present in an expression can be both subtle and complex, providing a wealth of information about our state of mind. By utilizing facial emotion recognition, we can assess how content and services impact the audience/users in a convenient and inexpensive way. Retailers can utilize these measures to assess customer interest, for instance [30]. Healthcare providers can improve their service by incorporating more details on patients' emotional wellbeing during treatment. Entertainment producers have the ability to track audience interaction during events in order to consistently produce the content they desire. Humans have a strong ability to understand others' emotions, with babies as young as 14 months being able to distinguish between happiness and sadness [31]. Can computers outperform us in capturing emotional states? In order to respond to the question, we created a deep learning neural network that enables machines to draw conclusions about our emotional statuses. Put simply, we provide them with the ability to perceive things just as we do.

- Identifying faces in the environment (for example, in a picture; this stage is also known as face detection),
- The dog ran quickly after the ball. Identifying facial characteristics within the identified face area (such as recognizing facial components' shapes or detailing skin texture in a facial zone; this process is known as facial feature extraction).
- The next step is to develop a detailed plan for the project. Examining the movement of facial characteristics and/or alterations in facial features and categorizing it into facial expression categories such as smile or frown, emotional categories like happiness or anger, and attitude categories like liking or ambivalence, etc. This process is known as facial expression interpretation.
- Multiple projects have been completed in this area, and our objective is to not only create an Automatic Facial Expression Recognition System but also enhance its accuracy in comparison to existing systems.

IV. Methodology

Preprocessing, as employed in the approach, refers to fundamental image manipulation where the input and output are both intensity pictures. Reduce the noise, Convert the Image to Binary/Grayscale, Pixel Brightness Transformation, and Geometric Transformation make up the majority of the preprocessing stages that are completed. A computer technique called Face Registration is used widely to identify faces in digital photos. During this stage of face registration, faces are initially identified within the image using a specific group of landmark points referred to as "face localization" or "face detection". Identifying specific regions, points, landmarks, or curves/contours in a 2-D or 3D image is a crucial part of facial feature extraction for face recognition [32]. During this feature extraction stage, a numeric feature vector is created based on the registered image that was generated. Some typical characteristics that can be identified include Lips, Eyes, Eyebrows, Nose tip. In the third stage of classification, the algorithm tries to categorize the provided faces displaying one of the seven primary emotions. In machine learning, it is a typical task to study and create algorithms that can learn from data and make predictions. These algorithms operate by creating data-based forecasts or choices, by constructing a mathematical model using input data. Typically, the final model is constructed using data gathered from various datasets. Three sets of data are typically utilized at various points in the model creation process. The model is first trained on a dataset, which consists of examples used to adjust the parameters (such as weights of connections between neurons in artificial neural networks) of the model. The training dataset is used to train the model with a supervised learning technique, such as gradient descent or stochastic gradient descent. In actuality, the training dataset typically includes pairs of an input vector and its corresponding answer vector or scalar, commonly referred to as the target. The existing model is utilized with the training data set to generate an outcome [33], which is subsequently evaluated against the desired outcome for every input vector in the training dataset. The parameters of the model are modified based on the comparison result and the specific learning algorithm in use. The process of model fitting may involve choosing variables as well as determining parameter values. Next, the model that has been fitted is utilized to forecast the outcomes for the data points in another dataset known as the validation dataset. The validation dataset allows for an impartial assessment of a model trained on the training dataset when adjusting the model's hyperparameters, such as the number of hidden units in a neural network [34]. Using validation datasets for regularization involves stopping training once validation error increases, indicating overfitting to the training dataset. The validation dataset's error can fluctuate during training, resulting in multiple local minima and complicating the simple procedure in practice. This issue has resulted in the development of numerous ad-hoc guidelines for determining the onset of overfitting. The test dataset is primarily utilized to impartially assess the performance of the final model trained on the training dataset.

Step 1: Gathering a dataset of pictures. In the FER2013 database, there are 35887 pre-cropped grayscale images of faces, each with a resolution of 48 by 48 pixels, and classified into one of 7 emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral.
Step 2: Image pre-processing.
Step 3: Detection of a face from each image. Step 4: The face that has been cropped is changed to grayscale pictures.
Step 5: The pipeline guarantees that each image can be inputted as a (1, 48, 48) NumPy array to the input layer.
Step 6: The Convolution2D layer receives the NumPy array as input.

Step 7: Convolution The Convolution2D layer takes the NumPy array as input.
Step 8: The MaxPooling2D pooling method uses (2, 2) windows on the feature map to retain only the highest pixel value.
Step 9: Throughout the training process, the pixel values undergo both Forward propagation and Backward propagation in the Neural network. Step 10: The Softmax function is represented as a probability for each emotion category. The model can display the specific breakdown of emotion probabilities in the facial expressions.

A collection of 327 files was placed in a folder and each file underwent processing to generate. set of features. Upon picking up the file, the filename was examined in order to extract the. label for emotions shown in figure 1. The feelings tag was added to a set of tags that will make up our multi-class target feature. The photo underwent processing to identify faces [35] and predict features. The characteristics obtained from each document were added to a list, which was then transformed into a NumPy array sized 327*68*2. Additionally, our target classes were stored as a NumPy array. The identical procedure was utilized for the Rafid database. After developing the feature set and target variable, we applied Support Vector technology. Devices that can anticipate feelings. The Support Vector Machines (SVM) and Logistic Regression algorithms were implemented using the Sk learn machine library [36]. The approach employed for multiclass classification was. One-VsRest applies to all algorithms. Logistic regression algorithm was adjusted for the regularization penalties "l1" and "l2". We adjusted the linear kernel to rbf and poly to observe the differences in outcomes. Cross-validation was implemented in conjunction with SVM to eliminate any potential biases in the datasets. At first, the dataset was split into 70% for training and 30% for testing. We tested various other divisions like 80:20 and 70:30. A 70:30 split appeared more attractive because we assumed that each class would be evenly represented in the test set. We started by testing with 4 splits for the cross-validation score. For better outcomes, we selected the values 5 and 10, commonly used in 28-fold cross-validation. Random Forest Classifier [37] and Decision Trees were tested on our dataset as well however, they showed lower accuracy compared to other algorithms in our experiment. Therefore, we opted to proceed with SVM and Logistic Regression.

The One-Vs-Rest approach is used for all algorithms. The logistic regression algorithm was adjusted for the penalties "l1" and "l2". We adjusted the linear kernel to rbf and poly as well to observe the differences in outcomes. Cross-validation method was implemented in conjunction with SVM [38] to eliminate any biases within the databases. At first, the dataset was split into 70% for training and 30% for testing. We experimented with various other divisions like 80:20 and 70:30. A 70:30 ratio appeared more attractive because we assumed that each class would be evenly represented in the test set. We started by testing the cross-validation score with 4 splits. In order to enhance the outcomes, we opted for the standard values 5 and 10 for 28-cross-validation. The Random Forest Classifier and Decision Trees were tested on our dataset but showed lower accuracy compared to other algorithms. As a result, we chose to move forward with SVM and Logistic Regression. Emotion labels in the dataset: 0: - 4593 images- Angry.

1: -547 images- Disgust
2: -5121 images- Fear
3: -8989 images- Happy
4: -6077 images- Sad
5: -4002 images- Surprise
6: -6198 images- Neutral

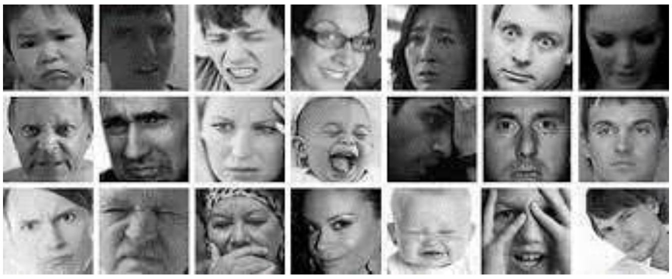


Figure 1: The Emotion

4.1 Haar Features

A Haar feature is like a Kernel, commonly used for edge detection. All human faces have common features such as the eye area being darker than the upper cheek region, and the nose area being lighter than the eye region. Through these distinctive characteristics, such as their position and dimensions, we can identify a face

Here are a few Haar features, based on which we can determine if a face is present or not. A Haar feature indicates that a black area is denoted by +1 while a white area is denoted by -1. It employs a 24X24 window in order to analyse an image. Calculation of each feature involves subtracting the sum of pixels in the white rectangle from the sum in the black rectangle. Different sizes and positions of kernels are considered to generate numerous features. The calculation of each feature requires determining the sum of pixels in the white and black rectangles. A 24X24 window will contain over 160,000 Haar features, demonstrating a significant amount. In order to address this issue, they implemented the use of integral images. It reduces the computation of pixel sums, regardless of the pixel count, to a task that only requires four pixels.

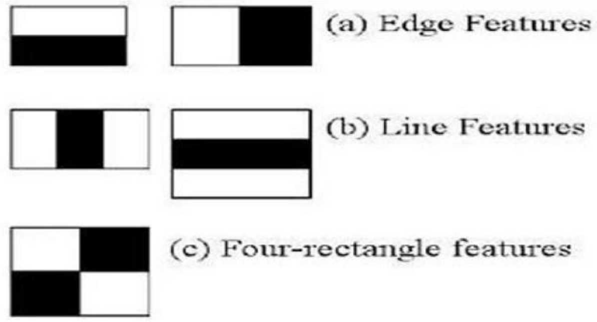


Figure 2: The Haar Features

4.2 Cascading

The second function is designed to identify the nose bridge; however, it is not necessary for detecting upper lips since they have a consistent feature. Therefore, it can be effortlessly removed. Adaboost helps us identify the relevant features out of over 160,000 options. Once all the characteristics are located, a weighted score is assigned to them in order to determine if a specific window contains a face or not. F(x) = a1f1(x) + a2f2(x) + a3f3(x) + a4f4(x) + a5f5(x) + …. F(x) is strong classifier and f(x) is weak classifier. The weak classifier always outputs a binary value, either 0 or 1. If the characteristic exists, the value will be 1; otherwise, it will be 0. Typically, a strong classifier is created using around 2500 classifiers. Chosen attributes are deemed satisfactory [39] if they outperform randomly guessing, meaning they must identify over half of cases. Cascading is shown in figure 3 a more compact and effective classifier. Non facing areas can be easily and complex cells, according to Hubel & Wiesel (1959, 1962). CNN [40] models have different variations, but typically include convolutional and pooling layers that are organized into modules. These modules are followed by one or more fully connected layers, as seen in a typical feedforward neural network. Stacking modules on top of each other is a common practice for creating deeper models. It shows a standard CNN structure for a simple image classification assignment. An input image goes through the network, then goes through various layers of convolution and pooling. Afterwards, inputs from these processes are passed to one or multiple fully connected layers. In the end, the class label is outputted by the final fully connected layer. Although this base architecture is widely used, there have been numerous proposed changes in recent years aimed at enhancing image classification accuracy or lowering computation expenses. CNNs and ANNs in general employ learning algorithms to modify their free parameters in order to achieve the intended network output. Backpropagation is the algorithm most frequently utilized for this task. Backpropagation is used to calculate the gradient of a given objective function to decide on the necessary adjustments to the parameters of a network, in order to decrease errors
(DCNN)

4.4 Model Implementation

In computer vision, deep learning is a popular technique. We chose CNN layers as the building blocks upon which to build our model design. Since everyone loves Mr. Bean, we decided to use his photo to show how photos are entered into the model [41]. A common structure of a convolutional neural network consists of an input layer, several convolutional layers, several dense layers (also known as fully-connected layers), and an output layer. These layers are arranged one after the other in a straight line. In Keras, the model is initiated with Sequential () and additional layers are included to construct the architecture shown in figure 5.

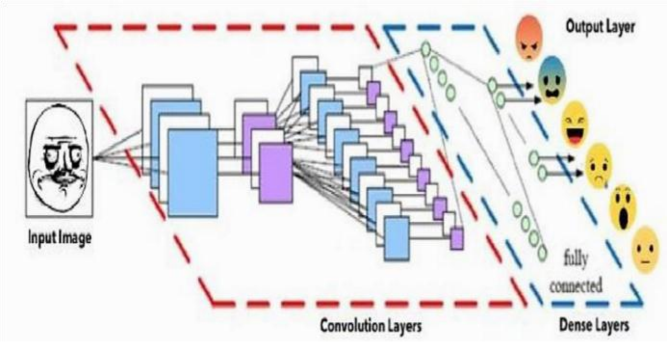
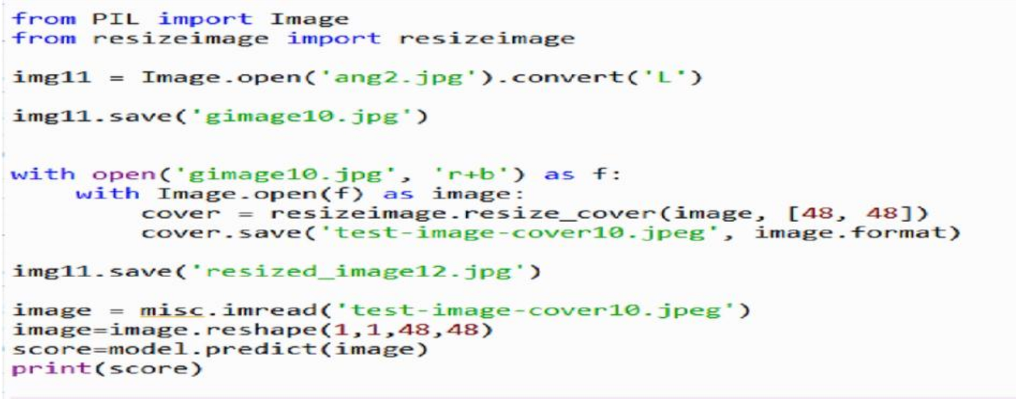
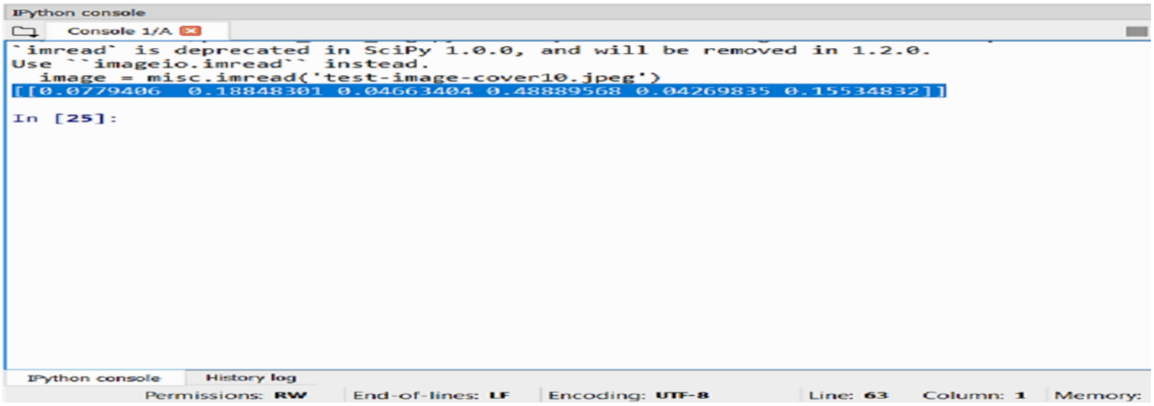


Figure 5: The Model Implementation



```
from PIL import Image
from resizeimage import resizeimage

img11 = Image.open('ang2.jpg').convert('L')

img11.save('gimage10.jpg')


with open('gimage10.jpg', 'r+b') as f:
    with Image.open(f) as image:
        cover = resizeimage.resize_cover(image, [48, 48])
        cover.save('test-image-cover10.jpeg', image.format)

img11.save('resized_image12.jpg')

image = misc.imread('test-image-cover10.jpeg')
image=image.reshape(1,1,48,48)
score=model.predict(image)
print(score)
```



```
`imread` is deprecated in SciPy 1.0.0, and will be removed in 1.2.0.
Use ``imageio.imread`` instead.
  image = misc.imread('test-image-cover10.jpeg')
[[0.0779406  0.18848301 0.04663404 0.48889568 0.04269835 0.15534832]]

In [25]:
```

```
Permissions: RW    End-of-lines: LF    Encoding: UTF-8    Line: 63    Column: 1    Memory:
```

VI. Conclusion

It can be challenging for AI systems to accurately identify emotions since different people may exhibit [2] the same feeling in various ways. Emotions may manifest as subtle changes in body language or facial expressions. Because of this, it may be difficult for AI systems to accurately identify emotions. This research presents a robust facial expression recognition method that aligns behavioural features with physiological biometric features to build a long-lasting face recognition model. The recognition method is based on the geometrical structures of the human face associated [3] with emotions such as sorrow, fear, anger, surprise, and contempt. The behavioural part of the method deals with the underlying attitude of different expressions as a foundation for identification. The [4] property bases in genetic algorithmic genes are

References

[1]. Huang, D.; Guan, C.; Ang, K.K.; Zhang, H.; Pan, Y. Asymmetric spatial pattern for EEG-based emotion detection. In Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, 10–15 June 2012; pp. 1–7. 2

[2]. Y. Perwej, “Unsupervised Feature Learning for Text Pattern Analysis with Emotional Data Collection: A Novel System for Big Data Analytics”, IEEE International Conference on Advanced computing Technologies & Applications (ICACTA'22), SCOPUS, IEEE No: #54488 ISBN No Xplore: 978-1-6654-9515-8, Coimbatore, India, 4-5 March 2022, DOI:10.1109/ICACTA54488.2022.9753501 Cui, Y.; Wang, S.; Zhao, R. Machine learningbased student emotion recognition for business English class. Int. J. Emerg. Technol. Learn. 2021, 16, 94–107.

[3]. K. Tai, "The application of digital image processing technology in glass bottle crack detection system[J]", Acta Technica CSAV (Ceskoslovensk Akademie Ved), vol. 62, no. 1, pp. 381-390, 2017

[4]. Saurabh Sahu, Km Divya, Dr. Neeta Rastogi, Puneet Kumar Yadav, Y. Perwej, "Sentimental Analysis on Web Scraping Using Machine Learning Method" , Journal of Information and Computational Science (JOICS), ISSN: 15487741, Volume 12, Issue 8, Pages 24-29, 2022, DOI: 10.12733/JICS.2022/V12I08.535569.67004

[5]. J Chen, X Yao, Huang Fen* et al., "N status monitoring model in winter wheat based on image processing[J]", Transactions of the Chinese Society of Agricultural Engineering, vol. 32, no. 4, pp. 163-170, 2016

[6]. Dawar Husain, Y. Perwej, Satendra Kumar Vishwakarma, Prof. (Dr.) Shishir Rastogi, Vaishali Singh, N. Akhtar, "Implementation and Statistical Analysis of De-noising Techniques for Standard Image", International Journal of Multidisciplinary Education Research (IJMER), ISSN:2277-7881, Volume 11, Issue10 (4), Pages 69-78, 2022, DOI: 10.IJMER/2022/11.10.72

[7]. Schoneveld, L.; Othmani, A.; Abdelkawy, H. Leveraging recent advances in deep learning for audio-visual emotion recognition. Pattern Recognit. Lett. 2021, 146, 1–7

[8]. Sun, Q.; Liang, L.; Dang, X.; Chen, Y. Deep learning-based dimensional emotion recognition combining the attention mechanism and global second-order feature representations. Comput. Electr. Eng. 2022, 104, 108469

[9]. Y. Perwej, F. Parwej, A. Perwej, "Copyright Protection of Digital Images Using Robust Watermarking Based on Joint DLT and DWT", International Journal of Scientific & Engineering Research (IJSER), France, ISSN 2229-5518, Volume 3, Issue 6, Pages 1-9, June 2012

[10]. Y. Perwej, "An Evaluation of Deep Learning Miniature Concerning in Soft Computing", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), ISSN (Online): 2278-1021, Volume 4, Issue 2, Pages 10 - 16, 2015, DOI: 10.17148/IJARCCE.2015.4203

[11]. Kajal, Neha Singh, N. Akhtar, Ms. Sana Rabbani, Y.f Perwej, Susheel Kumar, "Using Emerging Deep Convolutional Neural Networks (DCNN) Learning Techniques for Detecting Phony News", International Journal of Scientific

[12]. Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN: 2456-3307, Volume 10, Issue 1, Pages 122-137, 2024, DOI: 10.32628/CSEIT2410113