Image based Emotion Recognition System

Ridit Jain, Ashi Jain, Runishka Rao and Lakshit Wasan

[1] BML Munjal University, Gurugram, Haryana, India
lncs@springer.com

Abstract

The "image-based emotion recognition system" addresses a state-of-the art movement in the field of enthusiastic artificial intelligence, utilizing deep learning and image handling methods. This amplifies planning convolutional neural networks (CNNs) and different digital image processing methodologies to analyze facial expressions, passing on exact emotion recognition. Utilizing datasets like FER2013 and UTKFace, the system is prepared to distinguish a range of emotions and demographic features like age and gender. The examination incorporates exploratory information examination, comprehensive preprocessing steps such as normalization and resizing, and advanced appear models. Key revelations outline the system's capability in feeling affirmation, age, and sex expectation, underscoring its potential applications in ranges like personalized promoting, healthcare, and human-computer interaction. The endeavor features the blend of present-day picture getting ready procedures and significant picking up, inferring an essential stroll inside the field of energetic machine intelligence and AI.

1 Introduction

The "Image-based Feeling Acknowledgment Framework" venture sets out on a journey to bridge the gap between human feelings and innovation in the middle, the structure uses convolutional neural networks (CNNs), a viable gadget in picture planning and AI, to unravel and group sentiments based on facial expression. The extension is supported by comprehensive datasets, strikingly FER2013 and UTKFace, which give an assorted extend of facial pictures clarified with passionate states, ages, and sexes.

The methodology of this project is meticulously crafted, beginning with exploratory data analysis to understand the dataset's composition. This is followed by a series of preprocessing steps, including image normalization, resizing, and grayscale conversion, ensuring that the images are optimally prepared for the CNN models. The CNN structures utilized within the extend are outlined to capture the nuanced highlights of facial expressions, making them capable at recognizing a wide cluster of feelings.

The system's ability to accurately predict emotions extends beyond mere facial recognition; it delves into understanding the subtleties of human expressions. This capability has many practical applications, from improving user experience in digital interfaces to providing valuable insights in psychological research. Furthermore, the project explores the prediction of demographic characteristics like age and gender, adding another dimension to its capabilities.

This project is a testament to the potential of machine learning and image processing in understanding and interpreting human emotions. It stands as a significant contribution to the field, paving the way for more empathetic and intuitive human-computer interactions.

2 Literature Review

In a review directed by M.A. Nasri and M.A. Hmani [1], a facial emotion recognition system was created utilizing CNN to distinguish seven feelings/emotions. datasets from the Empathic and AffectNet data sets were utilized with Xception CNN and K-fold cross-validation. The fundamental assessment measure is classification accuracy. Limitation be that as it may, dependence on the EMOTIC dataset may restrict generalizability across various settings. Despite this restriction, the review's methodology joining preprocessing and CNN offers advancement in static picture-based emotion recognition, particularly in data sets, for example, Empathic and AffectNet.

Tengfei Tune, Wenming Zheng, Suyuan Liu, Yuan Zong, Zhen Cui, and Yang Li [2] presented a new methodology using graph embedded convolutional neural network (GECNN) for emotion recognition in view of on EEG.This technique works on the catch of spatial data. Nonetheless, limitations include potential biases associated with specific EEG data information and potential restrictions in the appropriateness of more extensive datasets. The novelty of the paper lies in GECNN, which coordinates graph convolutional layers and consideration components to perceive feelings from picture-based EEG signals.

[3] The paper gives a new understanding technique for emotion recognition using convolutional neural network (CNN) joined with image edge detection. This technique improves interaction by normalizing images of facial expression and featuring their edges, serving to recognize emotions accurately. This technique was approved to make use of blended datasets from Fer-2013 and LFW databases, showing high acknowledgment rate and quicker response speed. There are restrictions in the variety of the datasets utilized, which might influence the model's capacity to sum up to a more extensive facial expressions and backgrounds.

[4] The paper presents a creative way to deal with facial emotion recognition, using a mix of image edge detection and CNN procedures to increase accuracy. The strategy was approved on blended datasets from the Fer-2013 and LFW information bases, showing great acknowledgment rates and altogether faster training. although, limitations connected with the variety of the dataset must be noted, which might affect their capacity to sum up across a more extensive scope of facial recognition and foundations. Also, while edge discovery helps recognition, it may not catch the nuances of more nuanced facial emotions.

[5] They utilize another way to deal with emotion recognition utilizing EEG signals. They changed these signs into pictures and investigated them utilizing different convolutional neural networks (CNNs) like AlexNet, ResNet50, VGG16, and configurable CNNs. This approach accomplished high exactness and shown that configurable CNNs are especially possible. Impediments likewise remember dependence on exact boundaries in signal handling and arrangement, which can influence the vigor and adaptability of results. Furthermore, testing and approval were performed on a solitary dataset, which might restrict the generalizability of the outcomes to various datasets and genuine situations.

[6] The report on facial expression classification using ResNet-50V2 covers information assortment (utilizing the FER2013 dataset), picture preprocessing and naming, arranging CNN and ResNet-50V2 models, preparing the dataset with those models, it covers a five-step procedure: and last investigation and discussion. The oddity of this review is that we test various ages (25, 50, 100, 200) to assess the performance improvement of the model. However, this study faces limitations in achieving high accuracy without incurring significant loss of values. Initial testing with 25 ages yielded a curacy of 55.83 with high misfortune, prompting further testing with additional ages.

[7] The report on Emotion Recognition from Facial Expressions uses a strategy zeroed in on facial emotion recognition using the Beginning model and different gathering techniques, for instance, K-Nearest Neighbors and Random Forests. The interest lies in the utilization of transfer learning. Transfer learning is a strategy for utilizing information from one issue space to tackle another issue area, diminishing computational expenses, and accomplishing precision with less preparation. However, although not expressly referenced in the citation segment, a potential impediment could be the dependence on predefined datasets and classifier models. This can restrict the framework's capacity to adjust to certain situations with different and complex emotional expressions.

[8] This strategy incorporates a series of steps: uses a library of facial expression images, preprocesses the pictures, separates, and chooses features, then, at that point, classifies and recognizes expression / emotion. The specialty of this technique lies in the utilization of the Mini_Xception design, custom fitted to the requirements of the facial emotion recognition task. Although, the report features a significant limitation: the high computational requests of preparing neural networks. This cycle was computationally intensive and required the utilization of GPUs, as ordinary computer training was insufficient to meet the data training requirements of this project.

[9] The Human Emotion recognition system report gives a methodology that stresses the use of open-source tools and advancements. This is a step into step approach that utilizes a demonstrated visual information base to look at emotion recognition. The uniqueness is the improvement of a test system to mimic the proposed work, because of fact that it is not possible to implement these algorithms in a real environment. This test system permits performing different tests and picturing test results for test analysis. Although, the dependence on simulation rather than true testing is a restriction, which might influence the generalizability and relevance of the outcomes to real-life-situation.

[10] The report on “Emotion Recognition Using Convolutional Neural Networks” uses an approach that involves using a CNN to create a classification model that integrates feature extraction and classification. To reduce the long training time and computational costs, the project applies transfer learning, especially fine-tuning the pre-trained VGG model approach, significantly reducing training time from several weeks down to just a few days and allows the

model to better adapt to variations in local data. effective. However, this approach is limited by the need for large and diverse datasets, long initial training times, and the requirement for high-performance, energy-intensive hardware.

3 Dataset

Dataset for Emotion Recognition

This file contains a 48x48 pixel grayscale image of a face. Faces are automatically registered so that they are centered and occupy approximately the same location in each image. The task was to classify each face into one of seven categories based on the emotion on the face (0 = angry, 1 = disgusted, 2 = afraid, 3 = happy, 4 = high caring, 5 = surprised, 6 = natural). There are 28,709 samples in the training set and 3,589 samples in the test population.

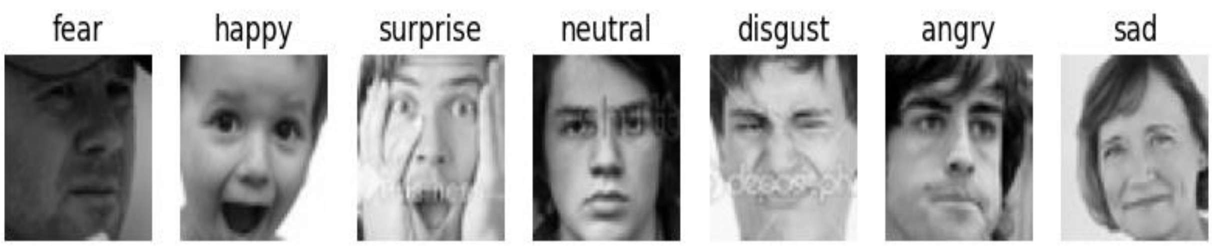


Fig: The types of images with emotion in the dataset

Dataset for Age and Gender Prediction

The UTKFace dataset, accessible on Kaggle, includes more than 20,000 different facial images commented on with age, gender, and identity information. This comprehensive dataset covers a wide range of ages, working with both study of age estimation and regression. It likewise gives orientation data, working with the development of orientation recognition models. Additionally, the inclusion of identity information is helpful for studying the impact of nationality on aging patterns. UTKFace is an important resource for facial analysis and computer vision research due to its diversity and rich annotations.

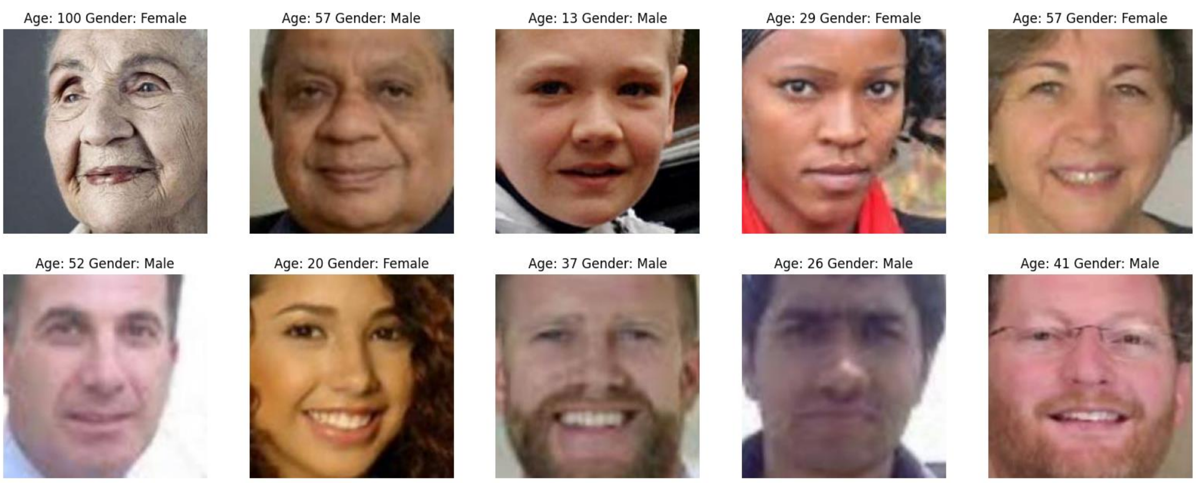


Fig: The types of images with age and gender in the dataset

4 Methodology for Emotion Recognition

4.1 Exploratory Data Analysis

Dataset Extraction- The first dataset contained in archive.zip is coordinated into two primary envelopes: "Train" and "Test". Every one of these envelopes contained subfolders addressing various classes or classifications fundamental for arranging and naming information. After extraction, this various leveled structure remained. The rundown of classes gives a reasonable outline of the dataset's creation. This is significant for training and validation tasks.

Image Analysis- An image from a dataset is stacked and changed over from BGR to RGB variety space, yet naturally OpenCV loads and shows it in BGR design. It then, at that point, yields the picture properties (aspects and information type) and draws a histogram for each variety channel (red, green, blue) to get a definite investigation of the picture's variety power dissemination

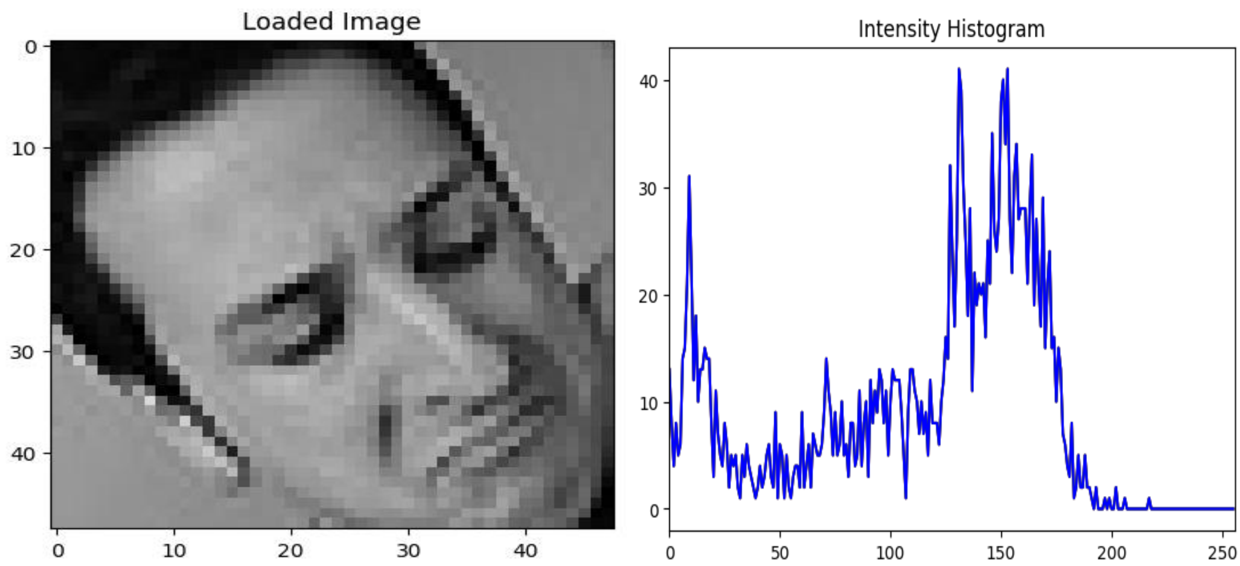


Fig: Loading an image and then plotting the histogram for one channel

Dataset distribution analysis- This step centers around dissecting and envisioning the distribution of emotions in both the training and testing datasets. This code counts the quantity of images for every emotion classification in the preparation and testing directories. These numbers are put away in a word reference and switched over completely to a panda's information outline for simpler analysis. This step includes making pie graphs for both the preparation and testing sets. Each pie diagram outwardly addresses the extent of every emotion classification in the dataset. The pie diagram gives a natural comprehension of the structure of the informational collection and features the general recurrence of every feeling in both the preparation and test sets. This is significant for surveying the equilibrium and portrayal of the informational index.

	category	count
0	fear	1024
1	happy	1774
2	surprise	831
3	neutral	1233
4	disgust	111
5	angry	958
6	sad	1247

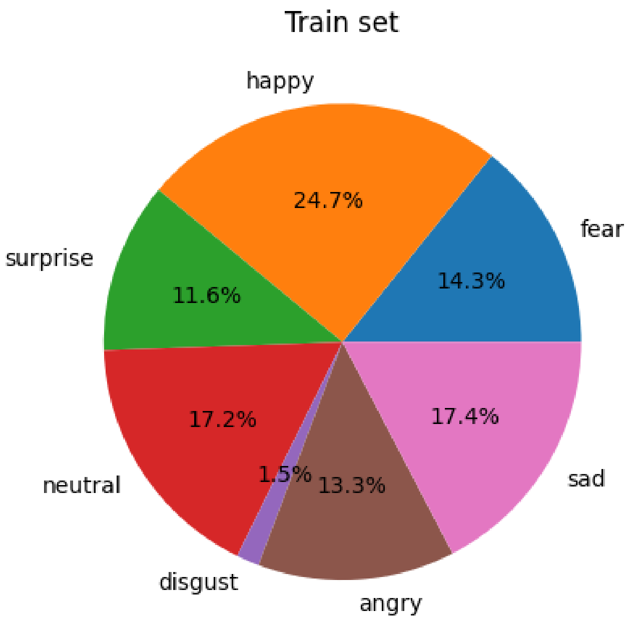


Fig: Count of no. of images for each emotion and then plotting pie chart for different emotions in the train dataset

	category	count
0	fear	4097
1	happy	7215
2	surprise	3171
3	neutral	4965
4	disgust	436
5	angry	3995
6	sad	4830

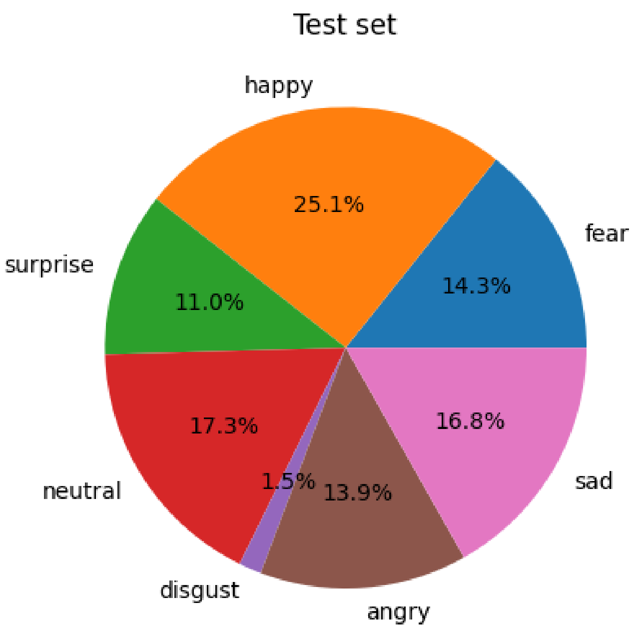


Fig: Count of no. of images for each emotion and then plotting pie chart for different emotions in the test dataset

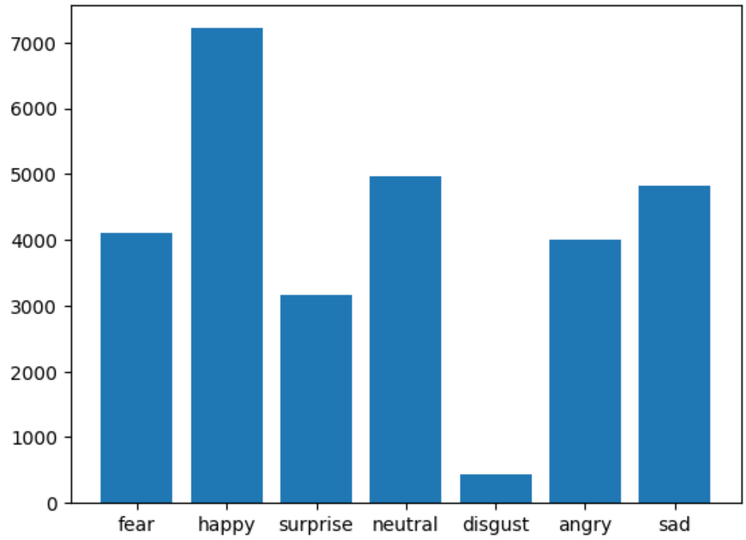


Fig: Bar Chart which takes input as dictionary and extracting the labels and their corresponding counts

4.2 Image Pre Processing

Image Histogram- Histogram equalization has become a common comparison technique in image processing due to its efficiency and simplicity. It is one of the complex methods of adjusting the dynamic range and contrast of the image by modifying the image to use a histogram to obtain the desired image.

Fig: Original image and equalized image after histogram equalization

Sobel filter- The Sobel filter is an edge recognition channel normally used in image processing. It applies a convolution operation to the image to highlight edges by highlighting sudden changes in intensity. It uses two 3x3 kernels to approximate gradients in the horizontal and vertical directions, which are then combined to detect edges.

Fig: Original image and Filtered image after Sobel filtering

Frequency Filter- A frequency filter, in image processing or signal processing, is a method of selectively modifying or extracting specific frequency components from a signal or image. It works in the frequency areas, permitting clients to improve or suppress specific frequency ranges, which can be valuable for like noise reduction, sharpening, or feature extraction.

Fig: Original image and Filtered image after Frequency filtering

Data Augmentation- Data augmentation is important in computer vision-based projects, where makes the dataset better by applying changes to existing images. By increasing the size of the data set and introducing variability, it improves the generalizability and reliability of the model. Then we convert augmented “disgusting” image data into NumPy arrays and display them in grid format for visual inspection, ensuring that augmentation techniques are appropriate to maintain data quality.

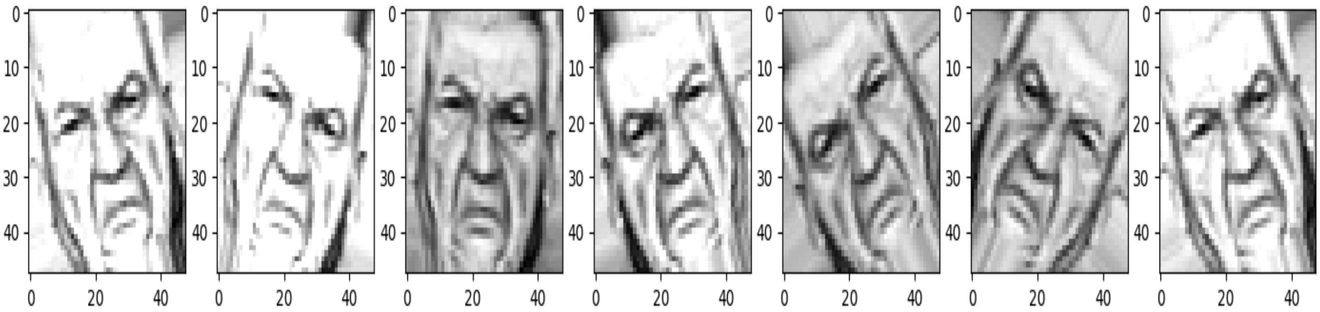


Fig: Images of emotion disgust by Data Augmentation

Resizing images- to a consistent size is crucial for deep learning because neural networks require uniform input dimensions. This ensures that the model can process all images in a consistent manner, making it easier to design and train the network. The image should be resized to (256*256) as CNN trains images of that size.

Image normalization- involves scaling pixel values to a standardized range, typically [0, 1] or [-1, 1]. This process helps the model converge faster during training by reducing the impact of varying pixel value ranges in different images. Normalization also aids in better weight initialization and gradient flow, ultimately improving the model's ability to learn meaningful features from the data.

Gray Scaling- Grayscale in digital image processing is the transformation of a variety image into a grayscale or highly contrasting image, where every pixel just holds luminance data addressing the grayscale. This works on the image by eliminating variety data while protecting underlying subtleties. The images in the dataset are as of now in grayscale design, so no gray scaling is performed unequivocally.

4.3 Model Building

Convolutional Neural Network (CNN)

Convolutional neural networks (CNN) are a type of artificial neural network mostly used for emotion recognition systems. CNNs are good at learning from raw pixel data(information), effectively capturing spatial hierarchies and strongly handling variations in different emotional expressions. Unlike regular(traditional) machine learning models that require manual feature engineering, CNNs automatically learn applicable features, making them ideally suited for recognizing emotions/feelings in diverse facial expressions and human complexity.

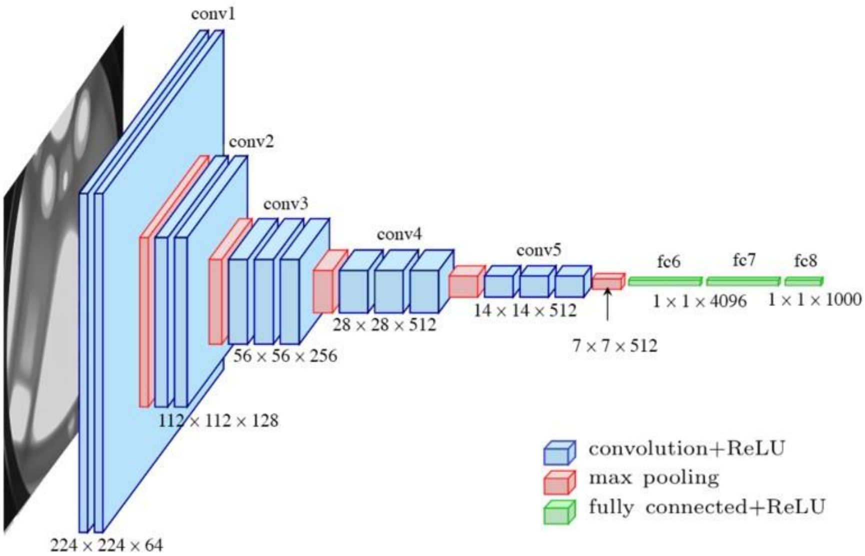


Fig: General CNN Architecture

Model Architecture

Convolutional Layers: This architecture starts with Conv2D layers. The first layer uses 64 filters for basic feature detection, followed by a layer of 128 filters for more complex models and a layer of 512 filters for further refinement. Together, these layers learn a hierarchical representation of data, giving the network the ability to expertly recognize complex features. Batch normalization layer: After the convolution layer, batch normalization normalizes the layer outputs, thereby improving training efficiency and model performance.

MaxPooling2D Layer: Max pooling reduces the spatial size of the object map, contributing to improving computational efficiency and making the model somewhat invariant to changes in scale and direction.

Dropout layer: To avoid overfitting, dropout randomly disables a portion of the input units during training, thereby improving the model's robustness to noise variations and data.

Flatten Layer: It transforms 2D feature maps into 1D vectors, preparing data for further processing in fully connected layers.

Dense layer: After convolution, the architecture switches to a dense layer. The Flatten layer converts feature maps into 1D vectors. Next are two dense layers, one with 256 neurons for learning high-level representations and the other with 512 neurons for further refinement. The final layer with 7 neurons predicts lessons, making it suitable for tasks such as emotion recognition.

```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 48, 48, 64)        640

batch_normalization (Batch   (None, 48, 48, 64)        256
Normalization)

activation (Activation)      (None, 48, 48, 64)        0

max_pooling2d (MaxPooling2   (None, 24, 24, 64)        0
D)
dropout (Dropout)            (None, 24, 24, 64)        0
conv2d_1 (Conv2D)            (None, 24, 24, 128)       204928

batch_normalization_1 (Bat   (None, 24, 24, 128)       512
chNormalization)

activation_1 (Activation)    (None, 24, 24, 128)       0

max_pooling2d_1 (MaxPoolin   (None, 12, 12, 128)       0
g2D)

dropout_1 (Dropout)          (None, 12, 12, 128)       0

conv2d_2 (Conv2D)            (None, 12, 12, 512)       590336

batch_normalization_2 (Bat   (None, 12, 12, 512)       2048
chNormalization)

activation_2 (Activation)    (None, 12, 12, 512)       0

max_pooling2d_2 (MaxPoolin   (None, 6, 6, 512)         0
g2D)
dropout_2 (Dropout)          (None, 6, 6, 512)         0

conv2d_3 (Conv2D)            (None, 6, 6, 512)         2359808

batch_normalization_3 (Bat   (None, 6, 6, 512)         2048
chNormalization)

activation_3 (Activation)    (None, 6, 6, 512)         0

max_pooling2d_3 (MaxPoolin   (None, 3, 3, 512)         0
g2D)

dropout_3 (Dropout)          (None, 3, 3, 512)         0

flatten (Flatten)            (None, 4608)              0
dense (Dense)                (None, 256)               1179904

batch_normalization_4 (Bat   (None, 256)               1024
chNormalization)

activation_4 (Activation)    (None, 256)               0

dropout_4 (Dropout)          (None, 256)               0

dense_1 (Dense)              (None, 512)               131584
batch_normalization_5 (Bat   (None, 512)               2048
chNormalization)

activation_5 (Activation)    (None, 512)               0
dropout_5 (Dropout)          (None, 512)               0
dense_2 (Dense)              (None, 7)                 3591
=================================================================
Total params: 4478727 (17.08 MB)
Trainable params: 4474759 (17.07 MB)
Non-trainable params: 3968 (15.50 KB)
```

Fig: CNN Architecture

5 Methodology for Age and Gender Prediction

5.1 Exploratory Data Analysis

Dataset Extraction- The process of collecting and configuring the UTKFace dataset includes the following steps: First, the dataset is downloaded from Kaggle using the Kaggle API for easy access. The downloaded zip file containing the dataset is then extracted to the local environment. To facilitate dataset management, a home directory was created for efficient access to the data files.

Labelling and Categorization- The resulting images were labeled and classified. Each image is carefully labeled with information regarding the subject's age and gender. To streamline data management and analysis, a structured database was built. This block of data allows for organized storage and retrieval of relevant information related to each image, making it easier to perform downstream tasks such as training a machine learning model or performing statistical analysis.

Visualization- The data set was inspected visually. This includes showing example images, creating visualizations to understand age distribution, and examining gender distribution. These visualizations give knowledge about key highlights of the dataset.

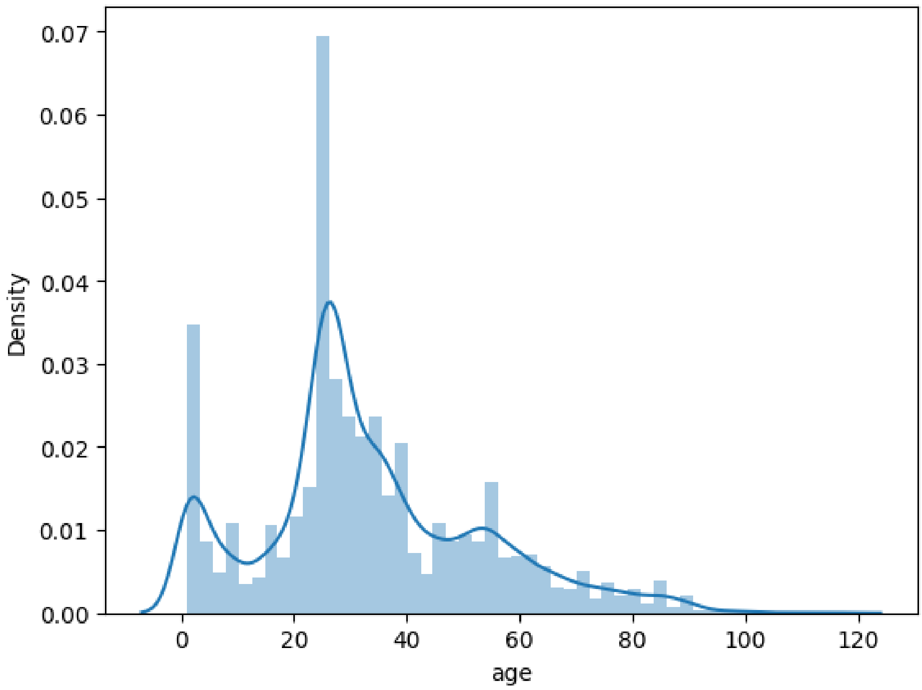


Fig: Age Distplot

The distplot in Seaborn serves as a visual representation of the distribution of image counts per age in the dataset. This graphical depiction combines a histogram with a Kernel Density Estimate (KDE) plot, providing a holistic perspective on the distribution of ages. In the histogram, each bar corresponds to a particular age range, showcasing either the count or density of images within that range. The accompanying KDE plot offers a smoothed estimation of the probability density. Together, these elements facilitate a distinction in expression understanding of the overall trend and concentration of images across various age groups in the dataset.

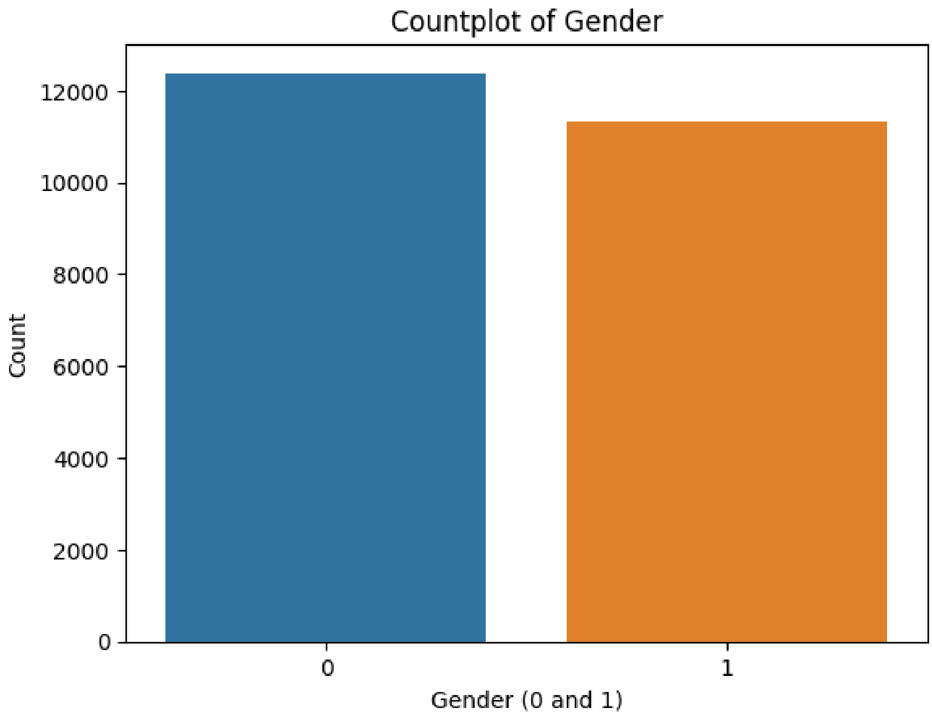


Fig: Gender Count plot

The utilization of the Seaborn count plot serves the purpose of visually portraying the distribution of gender labels in the dataset. On the x-axis, there are representations of gender labels (0 and 1), while the y-axis showcases the frequency of occurrences for each gender category. The visual presentation gives a clear and summarized overview of how gender is separated in the data. Including the title "Gender Count Table" gives extra context, and the axis labels contribute to the overall clarity of the visualization. This type of plotting is good for assessing the balance or asymmetry in the representation of different genders in a data set.

5.2 Image Pre Processing

Resizing Image- As part of the feature extraction process, all images were uniformly resized to 128 x 128 pixels. The main goal of this resizing process is to establish a consistent and standardized input size for each image in the dataset. In addition to ensuring compatibility for subsequent processing steps, normalizing image sizes to 128 x 128 pixels allows the model to learn features in a unified spatial context. The task of resizing optimizes the overall uniformity of the dataset, thereby increasing optimal performance in further stages of model training and evaluation.

Normalization: This involves scaling the pixel values to fall within the range [0, 1], thus establishing a standard frame. Standard for photography. Standardization is especially important during the training phase because it ensures consistency and facilitates convergence of machine learning models. Through this preprocessing step, the learning process is optimized and the model's ability to generalize to samples from the data set is improved.

Gray Scaling: In the field of predicting age and gender from images, the basic preprocessing step involves a gray scale. This converts colour images to grayscale, an operation that streamlines the data, limiting computational complexity while preserving important facial features. Switching to grayscale removes colour details, redirecting the model's attention to structural aspects necessary for accurate age and sex estimation. In our method, the strategic use of grayscale appears to be an indispensable aspect of data preprocessing, which improves model performance.

5.3 Model Building

Model Architecture

Convolutional Layers: The architecture starts with Conv2D layers. The first layer uses 32 filters and is followed by another Conv2D layer with 64 filters. These classes capture the basic characteristics of the input data.

MaxPooling2D Layers: MaxPooling2D layers reduce spatial dimensionality and improve computational efficiency.

Flatten Layer: It transforms 2D feature maps into 1D vectors for further processing

Dense layer: After convolution, the architecture switches to a dense layer. Next are two dense layers, one with 256 neurons and the other with 512 neurons. These classes teach high-level representations.

Output layer: The final output consists of two separate dense layers, one for predicting gender and another for predicting age. Each output layer has a single neuron.

Model: "model"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 128, 128, 1)]	0	[]
conv2d (Conv2D)	(None, 126, 126, 32)	320	['input_1[0][0]']
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0	['conv2d[0][0]']
conv2d_1 (Conv2D)	(None, 61, 61, 64)	18496	['max_pooling2d[0][0]']
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0	['conv2d_1[0][0]']
conv2d_2 (Conv2D)	(None, 28, 28, 128)	73856	['max_pooling2d_1[0][0]']
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 128)	0	['conv2d_2[0][0]']
conv2d_3 (Conv2D)	(None, 12, 12, 256)	295168	['max_pooling2d_2[0][0]']
max_pooling2d_3 (MaxPooling2D)	(None, 6, 6, 256)	0	['conv2d_3[0][0]']
flatten (Flatten)	(None, 9216)	0	['max_pooling2d_3[0][0]']
dense (Dense)	(None, 256)	2359552	['flatten[0][0]']
dense_1 (Dense)	(None, 256)	2359552	['flatten[0][0]']
dropout (Dropout)	(None, 256)	0	['dense[0][0]']
dropout_1 (Dropout)	(None, 256)	0	['dense_1[0][0]']
gender_out (Dense)	(None, 1)	257	['dropout[0][0]']
age_out (Dense)	(None, 1)	257	['dropout_1[0][0]']

Total params: 5107458 (19.48 MB)
Trainable params: 5107458 (19.48 MB)
Non-trainable params: 0 (0.00 Byte)

Fig: CNN Architecture

6 Model Building for Song Detection

Model Architecture

Convolutional Layers: The model starts with Conv2D layers. The first layer uses 32 filters, followed by the second Conv2D layer with 64 filters. These classes are responsible for capturing the basic characteristics of the input data.

MaxPooling2DLayers: The MaxPooling2D layer follows convolutional layers. They reduce dimensionality, thereby improving computational efficiency while retaining the necessary information.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 48, 48, 32)        320
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 48, 48, 64)        18496
_________________________________________________________________
batch_normalization (BatchNo (None, 48, 48, 64)        256
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 24, 24, 64)        0
_________________________________________________________________
dropout (Dropout)            (None, 24, 24, 64)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 24, 24, 128)       73856
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 22, 22, 256)       295168
_________________________________________________________________
batch_normalization_1 (Batch (None, 22, 22, 256)       1024
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 11, 11, 256)       0
_________________________________________________________________
dropout_1 (Dropout)          (None, 11, 11, 256)       0
_________________________________________________________________
flatten (Flatten)            (None, 30976)             0
...
Total params: 32,116,743
Trainable params: 32,116,103
Non-trainable params: 640
```

Fig: CNN Architecture

Flatten Layer: The flattened layer transforms 2D feature maps into 1D vectors, preparing the data for further processing in dense layers.

Dense Layers: After convolution, the architecture switches to a dense layer. Next are two dense layers: one with 256 neurons and the other with 512 neurons. These layers learn high-level representations from flattened features.

Output Layers: The final output consists of two separate dense layers. One is dedicated to predicting gender, while the other predicts age. Each output layer consists of a single neuron

7 Experimental Result

1: We focused on improving the dataset and preparing it for training without using traditional digital image processing techniques. Instead, we use data augmentation to increase the size of the dataset and introduce variability. We already had grayscale conversion in our dataset to simplify the presentation and resized images to ensure a consistent size. Additionally, normalization was performed to scale the pixel values to achieve more efficient neural network training. Performing these preprocessing steps improved the robustness and generality of the model without resorting to complex image processing techniques. The accuracy came out to be 66.4%.

2: We have carried out advanced digital image handling strategies to further develop emotion recognition. Gaussian Blur is applied to smooth the image, reduce noise, and highlight key features. Histogram equalization is used to improve image contrast, bringing out subtle details related to emotions. By utilizing these techniques, we expect to make a superior model with capacity to separate and foresee feelings precisely. These methods help us to support image processing to improve the quality of input data, ultimately leading to more accurate and robust emotion predicted results. The accuracy came out to be 74.4%.

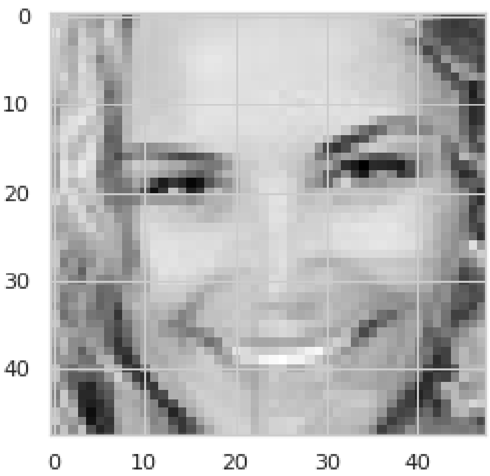
Fig: Input image and predicted emotion

3: We applied a multifaceted approach to effectively predict emotions. We used the Sobel filter to highlight edges and gradations in the image, focusing on facial features essential for emotion recognition. Additionally, frequency filters were used to target specific frequency components in the image, thereby enhancing relevant emotional signals. To supplement these picture handling methods, we incorporated a convolutional neural network (CNN) model, utilizing its profound gaining capacities to gain and concentrate complex examples from images. Preprocessed images. This comprehensive approach allows us to achieve exceptional accuracy in predicting individual emotions by combining advanced image processing and deep learning methods. The accuracy came out to be 80.6%.

4: We applied a systematic approach to predict both age and gender of individuals from images. Initially, we performed grayscale conversion to simplify image representation and reduce computational complexity. We then apply normalization techniques to normalize the pixel values, ensuring consistent input to our neural network. To accommodate different image sizes, we have resized the images accordingly. Finally, we harness the force of a convolutional neural network (CNN) model, permitting it to learn and extract complex age-and gender related designs from preprocessed pictures. This comprehensive pipeline, combining preprocessing and deep learning, allows us to obtain accurate predictions of age and gender from facial images. For gender the accuracy is 94% and for age the accuracy is 84%.

Fig: Input image and predicted age and gender

5: We've have used a holistic approach to predict the right songs for everyone based on their emotions. We start the process by converting the input images to grayscale, thereby simplifying their presentation, and reducing computational complexity. Next, we apply a normalization technique to normalize the pixel values, ensuring image uniformity. To handle different image sizes, we resized them appropriately. Importantly, we integrated a convolutional neural network (CNN) model, which can learn and extract complex emotional patterns from preprocessed images. This CNN demonstration is instrumental in combining feelings with significant music proposals, subsequently encouraging a more personalized and sincerely resounding music determination encounter. This comprehensive preprocessing and profound learning handle contributed to effectively foreseeing tunes based on individuals' feelings, subsequently moving forward their music tuning in travel. The accuracy is coming out to be 94.6%.



	index	name	artist	mood
0	207	Geno - 2000 Remaster	Dexys Midnight Runners	Happy
1	260	ID8	Dimitri Vangelis & Wyman	Happy
2	590	Two Princes	Spin Doctors	Happy
3	652	Wonderful Together (feat. Envy Monroe)	Blasterjaxx	Happy
4	624	We Are Your Friends - Justice Vs Simian	Simian	Happy

Fig: Input image and predicted song based on the emotion

Comparison Analysis

	Accuracy
Prediction of emotion using pre-processing	66.4%
Prediction of emotion using gaussian blur and equalized histogram	74.4%
Prediction of emotion using filter and CNN model	80.6%
Prediction of age and gender	84% and 94%
Prediction of song suggestion based on emotion	94.6%

Table: Prediction of different techniques with their accuracy

Deployment

We have used the Streamlit platform to host our code and create an automatic, and user-friendly interface. Our application has two different features. First, it provides live emotion prediction by analyzing facial expressions via the device's camera feed, providing real-time emotional feedback to the user. Second, it goes beyond emotion recognition by estimating a user's age and gender and using that information to recommend music tracks tailored to their current mood. This approach combined increases user experience, engagement, and personalization, making the application a flexible tool suitable and useful for various situations such as entertainment, mental health, and well-being.

Detected Emotion: sad

Predicted Age: 24

Predicted Gender: Male

Suggested Music:

	index	name	artist	mood
0	161	Eden - Hunted Version	Anna Calvi	Sad
1	655	Words I Heard	Julia Holter	Sad
2	21	All Mirrors	Angel Olsen	Sad
3	330	Love Is A Losing Game	Amy Winehouse	Sad
4	165	Emotions in Motion	Jaakko Eino Kalevi	Sad

Detected Emotion: happy

Predicted Age: 18

Predicted Gender: Male

Suggested Music:

	index	name	artist	mood
0	389	On My Way (feat. Jordan Kaahn)	Robbie Rivera	Happy
1	557	The Joker	Steve Miller Band	Happy
2	124	Dance Little Sister	Terence Trent D'Arby	Happy
3	335	Mad About You	Hooverphonic	Happy
4	652	Wonderful Together (feat. Envy Monroe)	Blasterjaxx	Happy

Fig: Input image and predicted emotion, age, gender, and song based on the emotion

Here we have used camera to take picture of the user and predicted what is the emotion of the user,age ,gender and the song reccomendation based on his emotion so that it can help the user to calm.

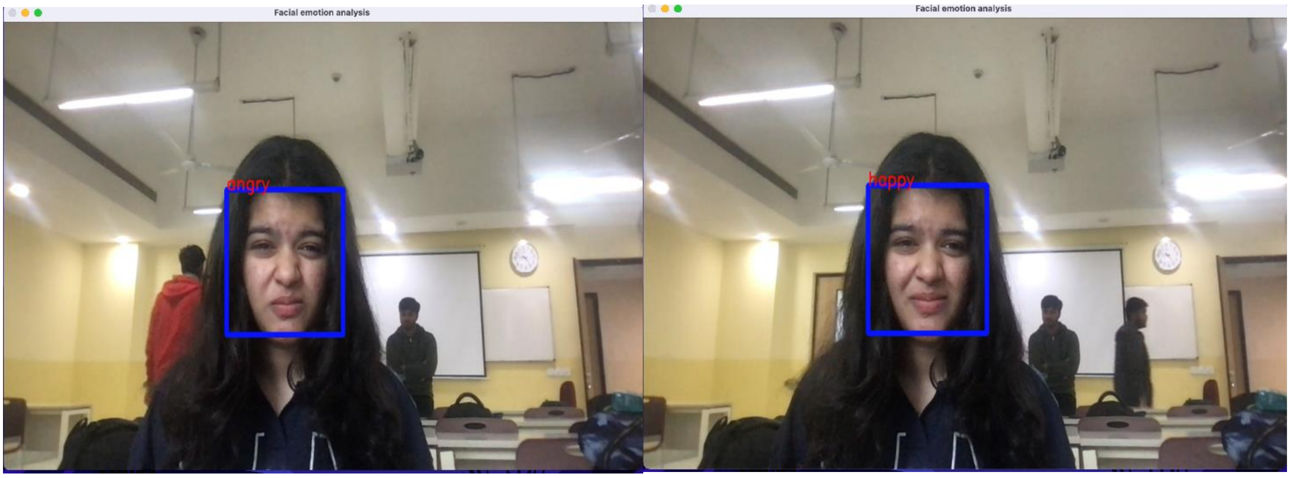


Fig: Input live video and predicted emotion

Here we have used camera to take video of the user and predicted what is the emotion of the user.

8 Conclusion

Image-based emotion recognition system" denotes a significant achievement in the field of artificial intelligence and emotion recognition. This project shows the successful implementation of advanced image handling approach and convolutional neural network (CNN) to carefully differentiate, and group human feelings based on facial expression. The utilization of extensive datasets, for example, FER2013 and UTKFace gave a strong groundwork to display preparation and approval, guaranteeing a different portrayal of feelings, age, and orientation.

Throughout the project, a range of innovative approaches were used, including data augmentation, grayscale transformation and normalization, to improve the quality of the dataset and prepare it for model training more effectively. CNN models with complex design have exhibited outstandingly phenomenal capacity to catch the subtleties of human expression, giving high exactness in recognizing feeling, age prediction, and gender.

The implications of this project are profound, offering potential applications in fields as diverse as marketing, healthcare, and entertainment. In marketing, the system can enable more targeted and emotionally resonant advertising. In healthcare, this can facilitate psychological assessment and patient care. In entertainment, this can deliver a more engaging and personalized user experience.

In short, the “image-based emotion recognition system” is a revolutionary effort that not only advances the field of emotional AI but also opens new paths for empathy technology. It features the staggering cooperative energy of image processing and AI in decoding the complex language of human feelings, setting another norm for future exploration and applications in this field.

9 References

A. Shoufan. (2023). *Exploring Students' Perceptions of ChatGPT: Thematic Analysis and Follow-Up Survey*, 38805-38818.

al, T. W. (May 2023). n IEEE/CAA Journal of Automatica Sinica. *A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development,*.

Cheng, K. L. (2023). *potential Use of Artificial Intelligence in Infectious Disease: Take ChatGPT as an Example.*, 1130–1135 .

Cooper, G. ((2023). J Sci Educ Technol . *Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence.*, 444–452 .

Lancaster. ((2023). *T. Artificial intelligence, text generation tools and ChatGPT – does digital watermarking offer a solution?*

Leslie, D. ((2023).). . AI and Ethics, . *Does the sun rise for ChatGPT? Scientific discovery in the age of generative AI*, 1-6.

M. Abdullah, A. M. (2022). Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS). *"ChatGPT: Fundamentals, Applications and Social Impacts*, 1-8.

M. Gupta, C. A. (2023). n IEEE Access, . *From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy,*.

Shoufan, A. (2023). *Exploring Students' Perceptions of ChatGPT: Thematic Analysis and Follow-Up Survey*.

Tlili, A. S. (2023). Smart Learn. *What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education.*, 10, 15.