Inclusion of Chatbot System to Real-Time Facial Emotion Recognition

Sandesh Mathpati[1], Shravankumar Arjunagi[2]
[1) 2)] *Department of Artificial intelligence and Data science, Faculty of Engineering and Technology (CO-ED), Sharnbasva University,Kalaburagi, India*

Abstract: - In modern society, social media has become an essential platform for individuals to express views and feelings through various forms of content, including text, messages, and emojis. People are busy with their daily routines and can't find space for each other; most people are stressed and depressed. In modern days, control over emotions by the people is limited; they are depressed over short stress and over-excited for smaller achievement. This paper presents a model designed to convert real-time images into gray scale sketches, utilizing the robust tools offered by Scikit-Image and other Python libraries and use of Python to develop a Chatbot that interacts with the user. This project introduces an algorithm that captures the image of the person with the help of a camera and stores the captured image. Furthermore, the stored image is pre-processed using a convolutional neural network and this is mapped with the pre-stored emoji to display the current emotional state of an individual. Once the emotional state of an individual is known, the user can interact with the Chatbot to express his feelings or thoughts about the emotion. The chatbot converses with the user and tries to normalize if the user is depressed or over-excited.

Keywords*:* Convolutional Neural Network (CNN), Deep Learning, Emotion Recognition, Chatbot, Machine learning.

1. Introduction

Computer vision facial expression recognition (FER) is a critical area for identifying human emotion through facial images. This technology has diverse applications, including computer and human interaction, safety, and medical care. The foundation work began with manual coding techniques, with Ekman and Friesen [1] identifying constants in facial expressions across different cultures, which laid a foundation. As digital image processing techniques evolved, Lyons et al. (1998) applied Gabor wavelets to encode facial expressions, significantly improving feature extraction methods (Lyons et al., 1998) [2]. The 3 major challenges are tackled by Goodfellow et al, [3]. The work of Shan et al., 2009 [4] on Facial expression recognition based on local binary patterns showed the effectiveness, robustness, and computational efficiency of LBP in various FER scenarios.

Facial emotion recognition has gained attention in various fields such as healthcare, education, and social media due to its potential. In this project, FER maps the image with the corresponding emoji. Emojis and avatars are significantly becoming important in online communication, diversifying their usage and becoming an integral part of various fields such as marketing, law, and healthcare. Emojis are not just entertaining pictures but they carry a wide range of emotional and digital interactions. Recently, the study of emojis and avatars has gained importance among academics from various departments including computing, communication, marketing, and behavioral science (Smith et al.)[16]. An emoji or avatar is represented as a symbol representation, written character, idea representation, or smiley embedded within text usually in electronic messages and web pages. Emojis and avatars are a new form of language that effectively conveys ideas and emotions, offering a richer communication medium.

The existing system makes use of deep learning algorithms to train the model to determine the facial expression of an image, the work is explained in the pre-processing stage of methodology of this article, which is understood by studying the use of committee neural networks to improve FER accuracy, combining multiple neural networks to enhance performance by Kulkarni et al. [5]. The field of FER involved when Li and Deng [6] surveyed deep facial expression recognition. Additionally, Zafeiriou et al. [7] reviewed face detection techniques, addressing the challenges and advancements in detecting faces in diverse and uncontrolled environments. Asifullah et al. [8] enhanced the CNN by surveying current difficulties. Furthermore, in this project to recognize facial expressions convolutional neural network (CNN) is used. This model will be trained on a dataset of facial expressions like happy, sad, neutral, and surprised which is obtained for the Kaggel dataset and real-time clicking of pictures with the help of voluntary individuals. The model will recognize the facial expression and map it with the corresponding emotion; these emotions are pre-stored in the database. Along with the emoji which recognizes the facial expression and defines the mood of the human being this project is embedded with the chatbot with organized architecture.

2. Objectives

- To develop a web page using which users can create an account, log into the account and begin the process.
- To create a database where the user account details are stored.
- Accurately classify emotions from facial expressions by real-time facial expression recognition system using CNN.
- To store the captured image and the corresponding emoji in the specified folder.
- Integration of a chatbot to which users can interact and feel better.
- Self-train the chatbot with the user response and store it in the source file.

3 Methodology

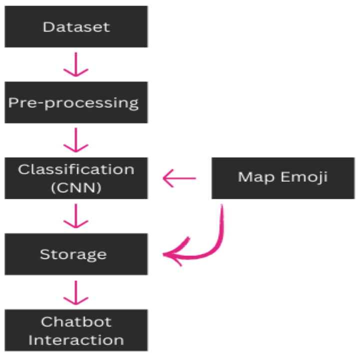

Fig.1. System Architecture

Based on the above flow diagram, the process begins with the webcam capturing an image of the user. This image is then processed by an algorithm that analyses and identifies the user's facial features and expressions. The captured image is then stored in a specified location, allowing the system to identify and reprocess it if needed. The algorithm then determines the user's facial expression, classifying it into predefined categories such as happy, sad, angry, or surprised. Once the expression is identified, it is matched with a corresponding emoji, which provides a visual representation of the user's current emotional state. This emoji is also stored in a specified location for logging and tracking purposes. Chatbot is then integrated into the model. The user can open the chatbot to verify that their current emotion has been correctly identified and stored. This step ensures the accuracy of the emotion recognition process. The user then begins interacting with the chatbot, which is now equipped with the user's emotional data. This enables the chatbot to offer responses based on user feelings, enhancing the overall user experience and making the interaction more meaningful and engaging.

3.1 Dataset

Training and testing of the model is carried out with the dataset of images obtained from the Fer2013 dataset from Kaggle and real-time images captured from the camera involve a careful process of model development and evaluation to accurately recognize facial expressions. This dataset consists of a vast collection of images representing various facial expressions such as happiness, sadness, and surprise. Approximately 28709 training data and 7178 training data have been used and around 200 to 300 captured images are used as training and testing data. The CNN architecture, designed to learn the hidden patterns and features within the images, undergoes training using this enriched dataset. During training, the difference between the actual and predicted emotions or facial expressions was minimized by adjusting the internal parameters of the model.

Fig.2. Dataset to train the model

TABLE.1. Training and testing data values for CNN

Facial expression	Training Data	Testing Data
Rage	4005	972
Disgust	390	105
Fear	5003	999
Jolly	7190	1812
Neutral	5001	1225
Unhappy	4801	1250
Shocked	3201	796

3.2 Pre-processing

In this method, different parts of the face are identified with the help of pixel intensities; the important points of the face are found using cascaded regression trees. The shape of the image is identified and parameter estimation is done. The normal images are converted into grayscale image, which reduces the complexity and to speed up the convergence during training pixel values are normalized to a range of 0-1 or -1 to 1 by dividing by 255.

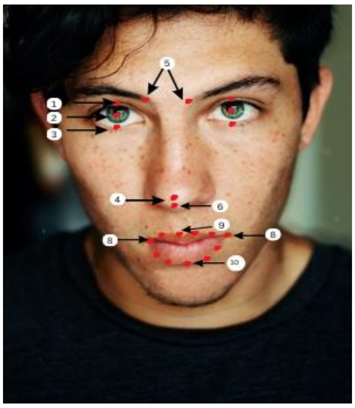

Fig.3. Feature points used for extraction

The ratio of displacement of these 21 points is calculated using pixel coordinates.

TABLE.2. Measurement of facial distance points

Points	Description of distance
1 and 2	Measurement of inner and upper point of the eye’s distance.
2 and 3	Measurement of inner and lower point of the eye’s distance.
4 and 6	Measurement of nose center and points at the inside end of the eyebrows’ distance.
6 and 8	Measurement of nose tip and points at each end of the mouth distance.
6,9, and 10	Measurement of nose tip and points at the center of each lip’s distance.
8	Measurement of the corners of the mouth’s distance.
Other points around mouth	Mouth circumference.

Various methods used

- Eulerian Motion Magnification (EMM)
- Amplitude- Based Eulerian Motion Magnification
- LBP-TOP feature extraction from EMM
- Binary Pattern (LBP)
- Eccentricity features Eccentricity features

3.3 Classification

Upon completing the training phase, the CNN's performance is strictly evaluated using real-time data. Support vector machine, convolutional neural network, and hidden markov model are the algorithm models that help for classification. SVM kernels and other kernels like edge detection kernels, Gaussian kernels, and pooling kernels are used in the classification process. Through rigorous training, the CNN learns to identify slight differences in facial features associated with each emotion, thereby enabling accurate classification in real-time scenarios. To make possible real-time emotion recognition, the model leverages OpenCV's Haar cascade XML files to detect faces in webcam feeds, extracting bounding boxes that encapsulate facial regions. Through careful validation and testing procedures, the model's accuracy, precision, and recall are assessed; ensuring its reliability and effectiveness in real-world applications and the speed and efficiency of classification can be improved by understanding Tejas Pandit et al. [10] and A. Krizhevsky et al. [11].

3.4 Emojis Mapping

Fig.4. Emojis used for mapping

Emojis are created and stored in the database. These emojis represent the emotional state of the person; the widely used emojis for this model are angry, happy, disgusted, fearful, sad, neutral, and surprised. Once the model is trained it is mapped with the emoji and displays the accurate emotion of the individual.

3.5 Storage

Emojis are created and stored in the database. These emojis represent the emotional state of the person; the widely used emojis for this model are angry, happy, disgusted, fearful, sad, neutral, and surprised. Once the model is trained it is mapped with the emoji and displays the accurate emotion of the individual.

3.6 Evaluation

Data Quality Metrics: Each attribute in the dataset was tested for completeness by calculating the percentage of missing value and any attribute with less than 5% missing value was removed and the model showed 95% of completeness. The accuracy of the model is shown as 93% which is evaluated by considering random sampling of 500 units against a benchmark (Goodfellow et al.)[3].

Model performance metrics: The ratio of the precision, recall, and F1-score is calculated. For instance, precision is calculated for a particular emotion happy (number of images of emotion happy/images of all the emotions) and recall is calculated for a particular emotion happy (number of images of correct emotion happy/total number of images of emotion happy). F1-score is used to evaluate the balance between precision and recall.

TABLE.3. Model performance by precision, recall and F1-score.

Emotion	Precision	Recall	F1-score
Happy	94%	92%	93%
Sad	90%	88%	89%
Angry	92%	90%	89.5%
Surprise	92%	91%	91.5%
Neutral	91%	89%	90%

3.7 Chatbot Interaction

Now, let's talk about the interaction part. This is the advanced phase of the project and various types of chatbots, different techniques used, and the importance of chatbots are understood by Adamopoulou et al. [12] and Nuruzzaman et al. [13]. The images that are captured with their associated emotions can be displayed within a chatbot interface. Recognition of speech (Jurafsky et al. [14]) is an important aspect in the chatbot and this makes interaction more lively and interactive. The value the chatbot can create in conversation (Shawar et al. [15]) is incredible. So, imagine chatting with a friendly bot, and it shows you a picture of yourself smiling, tagged with the 'happy' emotion. It's like having a little digital friend.

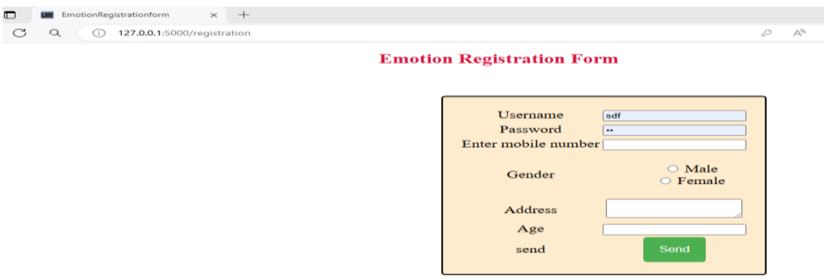

Fig.5. User registration to create chat account

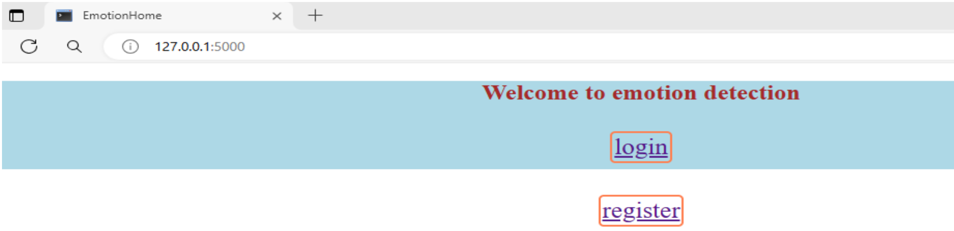

Fig.6. User login to interact with chatbot

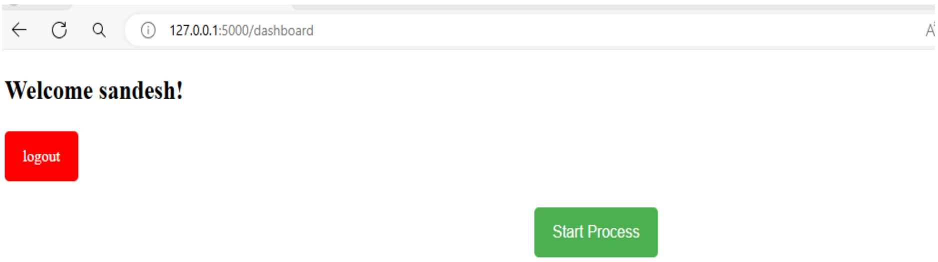

Fig.7. User dashboard to start process

Chatbot evaluation metrics

User satisfaction and interaction time are calculated by conducting a survey of 100 users who interacted with the Chabot. Data are recorded manually and graphs are plotted for user satisfaction and interaction time against user names.

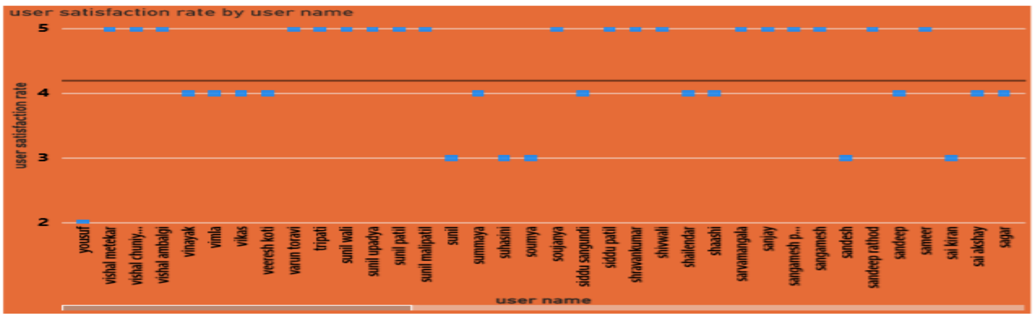

Fig.8. The average user satisfaction score was 4.2 out of 5-point Likert scale.

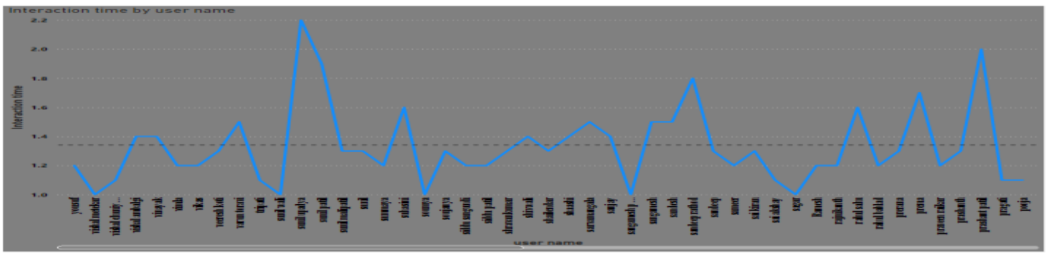

Fig.9. The average interaction time was 1.35 seconds

4 Result

Fig.10. Capture the image and save it

The above figure (Fig.10) is captured by the model using a webcam and it is stored in the specified location as random.PNG file.

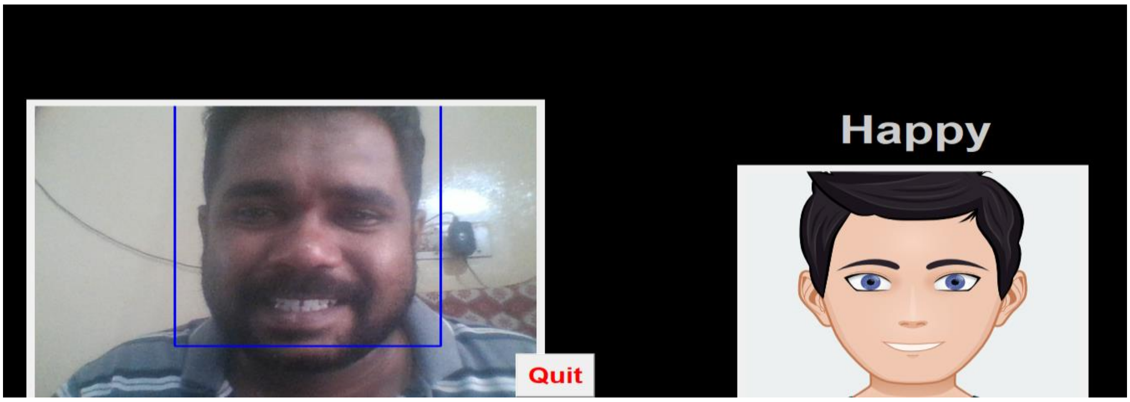

Fig.11. Match the image with the emoji (happy)

The captured image in Fig.10 is used to identify the emotion; the model displays the emotion of the image as "happy". This displayed emotion (Fig.11) is stored in the specified location as happy.PNG.

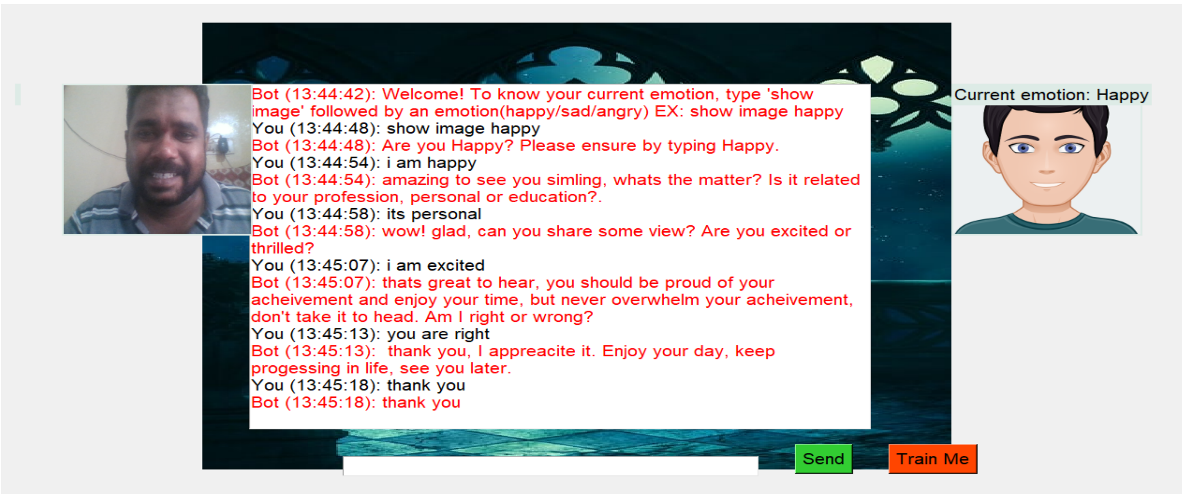

Fig.12. Interaction of user with the chatbot

The above figure (Fig.12) shows the interaction of user with the chatbot. Users can specify his/her emotions by typing "show image" and the emotion they are in, if the current emotion of the user is displayed on the screen based on the previously captured and matched emotion then the user can interact with the chatbot to normalize or motivate himself/herself.

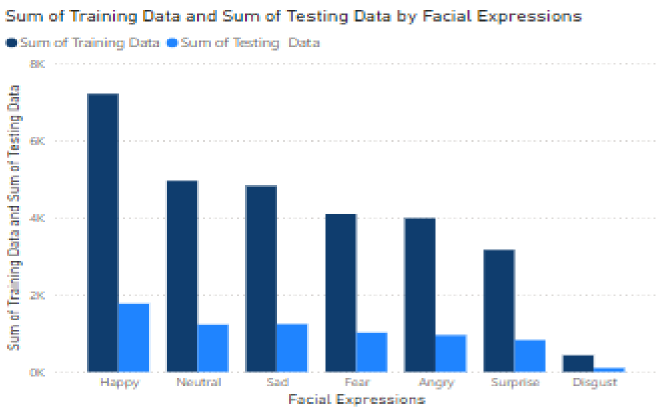

Fig.13. Facial expression vs training and testing data.

CNN use **edge detection kernel, Gaussian kernel, and pooling kernel.** The accuracy is checked for various kernels and tabulated in TABLE 4.

TABLE.4. Accuracy for different kernels used

Sl.No	Kernels	Accuracy
1	Edge detection	76.50%
2	Gaussian	74.14%
3	MaxPooling	80.25%

The model is tested using **convolutional neural network, support vector machine, and hidden markov model** and the accuracy is tabulates in TABLE 5.

TABLE.5. Comparison between models

Sl.No	Model	Accuracy
1	CNN	87.25%
2	SVM	81.45%
3	HMM	76.60%

5 Discussion

The proposed model effectively combines emotion identification with interactive responses, using advanced computer vision and artificial intelligence techniques. The system can use CNNs to recognize minute differences in facial features that lead to high precision of emotional classification. Emoji database integration helps bring these identified emotions visually hence increasing engagement and user experience.

Introducing a chatbot in this system provides intelligent and context-aware answers based on the identified emotional state which enriches the interaction between users. This way also not only detects but also supports an expressive manner which is good for a user's mental health as well as their emotional experiences. Chatbots have been found out to be very important because they assist people emotionally by having meaningful conversations.

The methodology of the project such as real-time image capture, emotion mapping, and chatbot interaction presents a holistic way of addressing emotion recognition complexity. Diverse datasets used as well as testing with real world scenarios ensure applicability of the system in different cases.

6 Conclusion

A convolutional neural network (CNNs) is used to recognize real-time facial expressions and translate them into emojis with the help of the fusion of computer vision and emotional intelligence. In this attempt, this model has been trained on a diverse dataset encompassing seven distinct facial expressions, including sad, neutral, happy, surprised, fearful, angry, and disturbed. Through rigorous training, the CNN learns to identify slight differences in facial features associated with each emotion, thereby enabling accurate classification in real-time scenarios. The mapping of recognized emotions to corresponding emojis or avatars adds a layer of visual representation, enhancing user engagement and comprehension. To make possible real-time emotion recognition, the model leverages OpenCV's Haar cascade xml files to detect faces in webcam feeds, extracting bounding boxes that encapsulate facial regions. These bounding boxes are then fed into the trained CNN for classification, enabling the model to infer the emotional state of individuals. Furthermore, the integration of a chatbot system complements this emotion-aware technology by enabling intelligent interactions based on users' current emotional states. With the help of machine learning, natural language processing (NLP), and AI techniques, the chatbot can easily understand and respond to user queries, providing its responses as per user query and the detected emotion. This multi-modal approach goes beyond traditional text-based communication, incorporating voice, images, and other sensory inputs to create more immersive and interactive user experiences. Through flawless integration of machine learning and AI capabilities, the system illustrates the convergence of cutting-edge technologies to enhance human-computer interaction and emotional understanding.

References

[1] Ekman, P, & Friesen, W. V. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17(2), 124-129, 1971.

[2] M. Lyons, S. Akamatsu, M. Kamachi, and J. Global. Coding facial expressions with gabor wavelets. In: Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on, pp. 200–205, Apr 1998.

[3] Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., ... & Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. International Conference on Neural Information Processing (pp. 117-124). Springer, Berlin, Heidelberg.

[4] C. Shan, S. Gong, and P. W. MccOwan. Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, 27(6), pp. 803–816, 2009.

[5] S.S. Kulkarni, N.P. Reddy, and S. Hariharan. Facial expression (mood) recognition from facial images using committee neural networks. BioMed Eng OnLine, 16, 08 2009.

[6] Li, S., & Deng, W. Deep facial expression recognition: A survey. IEEE Transactions on Affective Computing, 2020.

[7] Zafeiriou, S., Zhang, C., & Zhang, Z. A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding, 138, 1-24, 2017.

[8] Asifullah Khan, AnabiaSohail, Umme Zahoora, and Aqsa Saeed. A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53, 12 2020.

[9] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. arXiv:1707.01083, 2017.[10] Tejas Pandit, Akshay

Kapoor, Rishi Shah, and Rushi Bhuva. Understanding inception network architecture for image classification. 02 2020.
[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. NIPS, 1, p. 4, 2012.
[12] Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology. In: Artificial Intelligence Applications and Innovations (AIAI 2020), pp. 373-383. Springer, Cham.
[13] Nuruzzaman, M., & Hussain, O. K. (2018). A Survey on Chatbot Implementation in Customer Service Industry through Deep Neural Networks. In: 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE), pp. 54-61.
[14] Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing (3rd ed.). Pearson.
[15] Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they Really Useful? LDV Forum, 22(1), 29-49.
[16] Smith, J., Johnson, R., Brown, L., & Davis, K. (2023). The Influence of Emojis and Avatars on Digital Communication: A Multidisciplinary Approach. Journal of Digital Communication, 12(3), 45-67
[17] Bickmore, T. W., & Picard, R. W. (2005). Establishing and Maintaining Long-Term Human-Computer Relationships. ACM Transactions on Computer-Human Interaction (TOCHI), 12(2), 293-327