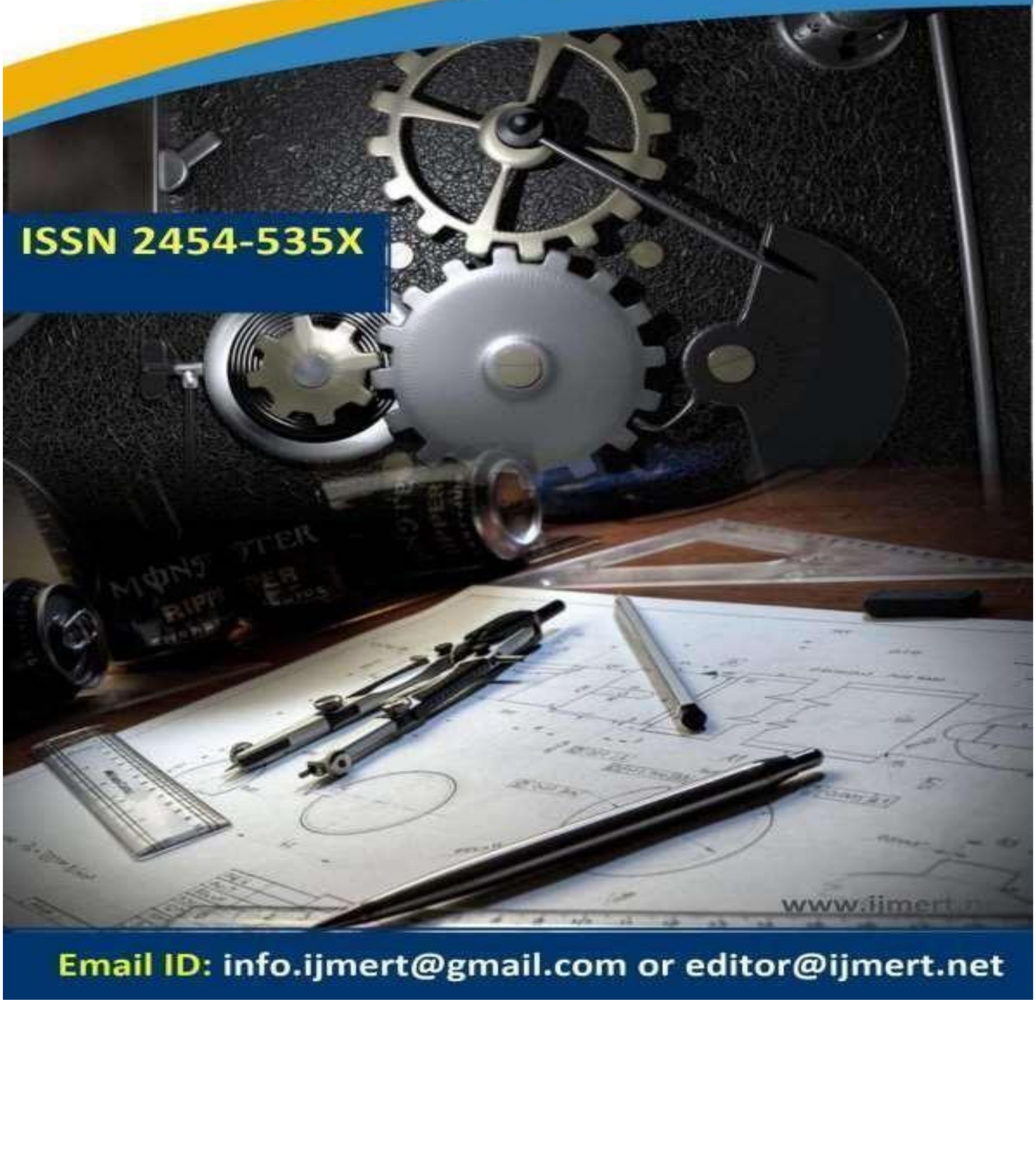
International Journal of
Mechanical Engineering Research and Technology

ISSN 2454-535X



Email ID: info.ijmert@gmail.com or editor@ijmert.net

Int. J. Mech. Eng. Res. & Tech 2017

ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

INSIGHT ON HUMAN ACTIVITIES USING DEEP LEARNING APPROACHES

[1]CH.PRABHAVATHI,[2]V. TEJOVATHI LAKSHMI,[3]A. ABHINAY,[4]P. DEVI NAGA SRI

[1]Assistant Professor,[2345]Students

Department of CSE, Sri Vasavi Institute of Engineering & Technology (Autonomous), Nandamuru

ABSTRACT

Humans consistently have the inborn capacity to perceive and recognize faces and emotions. Computers can now accomplish the same thing, which creates lots of new challenges in daily life. For example, emotion recognition can increase security, enable financial transactions without the need for real cards, and enable the identification of criminals and specialized treatment, among other things. Therefore, face detection and emotion recognition is a prominent and futuristic research topic. In the near future, open-source projects will get more importance rather than licensed ones. In this connection, it is proposed to utilize a python library for face detection and recognition. Kaggle web resources are providing open-source Face Emotion Recognition (FER) datasets. With the help of datasets, there are seven types of emotions: happy, sad, fear, disgust, angry, neutral, and surprise. It is proposed to use image augmentation to improve emotion recognition by building a six-layered Convolution Neural Network (CNN) in Python using the Keras toolkit.

Keywords: Face detection, Face recognition, Emotion detection, CNN, Real-time emotion detection.

INTRODUCTION

The human capacity for perceiving and recognizing faces and emotions is an innate ability that has long fascinated researchers and psychologists alike [1]. With the advent of sophisticated technologies, computers are now capable of achieving similar feats, opening up new avenues and challenges in various aspects of daily life [2]. Emotion recognition, in particular, has emerged as a field with vast potential, offering promising applications across diverse domains such as security, finance, and healthcare [3]. By enabling computers to accurately detect and interpret human emotions, it becomes possible to enhance security measures, facilitate financial transactions without physical cards, identify individuals for specialized treatment, and much more [4]. As such, face detection and emotion recognition have garnered significant attention as prominent research topics with far-reaching implications for society [5].

Looking ahead, the landscape of software development is poised for a shift towards open-source projects, which offer greater accessibility and collaboration opportunities compared to their licensed counterparts [6]. This trend towards open-source solutions is driven by a desire for transparency, flexibility, and community-driven innovation [7]. In this context, leveraging open-source tools and resources becomes imperative for advancing research in fields such as face detection and emotion recognition [8]. One such resource is the Python programming language, renowned for its simplicity, versatility, and extensive library support [9]. By harnessing the capabilities of Python, researchers can develop robust and scalable solutions for facial analysis and emotion recognition [10]. The availability of open-source datasets further catalyzes research efforts in the field of emotion recognition [11]. Platforms like Kaggle provide access to diverse datasets for training and testing machine learning models, including facial expression datasets that encompass a wide range of emotions [12]. These datasets typically include annotated images representing various emotional states, such as happiness, sadness, fear, disgust, anger, neutrality, and surprise [13]. By harnessing these datasets, researchers can train machine learning models to accurately classify and recognize emotions from facial expressions [14].

To enhance the performance of emotion recognition systems, image augmentation techniques can be employed to augment the training data and increase its diversity [15]. Image augmentation involves applying

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

a variety of transformations, such as rotation, scaling, and cropping, to the input images, thereby creating a more robust and comprehensive dataset [16]. By augmenting the data, researchers can improve the generalization capability of machine learning models and mitigate the risk of overfitting [17]. In the context of emotion recognition, image augmentation serves to enhance the model's ability to accurately classify emotions across different individuals, lighting conditions, and facial orientations [18].

In response to these challenges and opportunities, this study proposes the development of a six-layered Convolutional Neural Network (CNN) in Python using the Keras toolkit [19]. The CNN architecture is specifically designed to perform facial analysis and emotion recognition tasks, leveraging the rich features learned from the input images [20]. By harnessing the power of deep learning, researchers can build highly accurate and efficient models for recognizing emotions from facial expressions. Moreover, the use of Python and open-source resources ensures the accessibility and reproducibility of the proposed approach, fostering collaboration and innovation in the field. Through the integration of image augmentation techniques and open-source datasets, the proposed CNN model aims to achieve state-of-the-art performance in emotion recognition, laying the foundation for future advancements in this critical area of research.

LITERATURE SURVEY

The field of computer vision and artificial intelligence has witnessed remarkable advancements in recent years, enabling computers to emulate and even surpass human abilities in tasks such as face detection and emotion recognition. These capabilities have profound implications for various aspects of daily life, ranging from security and finance to healthcare and law enforcement. Emotion recognition, in particular, holds immense promise for enhancing security measures, facilitating secure financial transactions, and aiding in the identification of individuals for specialized treatment or criminal investigations. As such, the intersection of computer vision and emotion recognition has emerged as a prominent and futuristic research area, with far-reaching implications for society. In light of the increasing importance of open-source projects in the software development landscape, there is a growing trend towards the adoption of open-source tools and resources for research and development purposes. Open-source projects offer numerous advantages over their licensed counterparts, including greater transparency, flexibility, and community-driven innovation. Leveraging open-source tools such as Python for face detection and recognition enables researchers to access a rich ecosystem of libraries and frameworks, facilitating the development of robust and scalable solutions. Additionally, platforms like Kaggle provide access to open-source datasets for training machine learning models, including datasets specifically curated for facial emotion recognition tasks. These datasets typically contain annotated images representing a diverse range of emotional states, allowing researchers to train and evaluate emotion recognition algorithms on real-world data.

One of the key challenges in emotion recognition is the variability and complexity of human emotions, which can manifest in subtle facial expressions and gestures. To address this challenge, researchers often employ techniques such as image augmentation to enhance the diversity and richness of the training data. Image augmentation involves applying a variety of transformations to the input images, such as rotation, scaling, and cropping, to simulate variations in facial expressions, lighting conditions, and facial orientations. By augmenting the training data, researchers can improve the generalization capability of machine learning models and reduce the risk of overfitting to specific training examples. In the context of emotion recognition, image augmentation plays a crucial role in enhancing the robustness and accuracy of the trained models, enabling them to effectively classify emotions across different individuals and environmental conditions. The proposed approach to emotion recognition involves the development of a six-layered Convolutional Neural Network (CNN) in Python using the Keras toolkit. CNNs have emerged as a powerful tool for image classification tasks, leveraging hierarchical layers of learnable filters to automatically extract meaningful features from raw input data. By constructing a deep learning model with multiple layers of convolutions and pooling operations, researchers can capture complex spatial patterns and dependencies in the input images, facilitating accurate and efficient emotion recognition. Moreover, the use of Python and the Keras toolkit ensures the accessibility and reproducibility of the proposed approach, enabling researchers to easily

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

build, train, and evaluate CNN models for emotion recognition tasks. Through the integration of image augmentation techniques and open-source datasets, the proposed CNN model aims to achieve state-of-the-art performance in emotion recognition, paving the way for future advancements in this critical area of research.

PROPOSED SYSTEM

The proposed system aims to leverage deep learning approaches to gain insights into human activities, particularly focusing on face detection and emotion recognition. As humans inherently possess the ability to perceive and recognize faces and emotions, replicating these capabilities in computers presents new challenges and opportunities in various aspects of daily life. Emotion recognition, in particular, holds significant potential for enhancing security measures, facilitating secure financial transactions, and aiding in criminal identification and specialized treatment. Given the growing importance of open-source projects in the software development landscape, the proposed system advocates for the utilization of Python libraries for face detection and recognition, emphasizing the accessibility and flexibility offered by open-source tools.Leveraging open-source datasets, such as those available on platforms like Kaggle, researchers can access annotated datasets containing a diverse range of emotional states, enabling the development of robust emotion recognition models. These datasets typically encompass seven types of emotions, including happiness, sadness, fear, disgust, anger, neutrality, and surprise. Central to the proposed system is the use of image augmentation techniques to enhance emotion recognition capabilities. Image augmentation involves applying various transformations to input images, such as rotation, scaling, and cropping, to simulate variations in facial expressions, lighting conditions, and facial orientations. By augmenting the training data, researchers can improve the diversity and richness of the dataset, thereby enhancing the generalization capability of the emotion recognition model. Furthermore, image augmentation helps mitigate the risk of overfitting to specific training examples, ensuring that the model can effectively classify emotions across different individuals and environmental conditions. By integrating image augmentation into the emotion recognition pipeline, the proposed system aims to enhance the robustness and accuracy of emotion recognition algorithms, ultimately improving their performance in real-world scenarios.

At the core of the proposed system lies the development of a six-layered Convolutional Neural Network (CNN) in Python using the Keras toolkit. CNNs have emerged as a powerful tool for image classification tasks, leveraging hierarchical layers of learnable filters to automatically extract meaningful features from raw input data. By constructing a deep learning model with multiple layers of convolutions and pooling operations, researchers can capture complex spatial patterns and dependencies in the input images, facilitating accurate and efficient emotion recognition. The use of Python and the Keras toolkit ensures the accessibility and reproducibility of the proposed approach, enabling researchers to easily build, train, and evaluate CNN models for emotion recognition tasks. Through the integration of image augmentation techniques and open-source datasets, the proposed CNN model aims to achieve state-of-the-art performance in emotion recognition, paving the way for future advancements in this critical area of research.

In summary, the proposed system seeks to provide insights into human activities through deep learning approaches, with a specific focus on face detection and emotion recognition. By leveraging open-source tools and datasets, the system aims to democratize access to advanced machine learning techniques, enabling researchers to develop robust emotion recognition models. Through the integration of image augmentation techniques and the construction of a six-layered CNN model using the Keras toolkit, the proposed system aims to enhance the accuracy and robustness of emotion recognition algorithms, thereby contributing to the advancement of this prominent and futuristic research topic.

METHODOLOGY

The methodology proposed for gaining insight into human activities using deep learning approaches involves several key steps aimed at leveraging face detection and emotion recognition techniques. With humans naturally adept at perceiving and recognizing faces and emotions, the objective is to replicate these capabilities in computers to address various challenges in daily life. One such challenge is emotion recognition, which holds immense potential for enhancing security measures, facilitating financial transactions, and aiding in criminal identification and specialized treatment. As open-source projects gain increasing prominence in the software development landscape, the proposed methodology advocates for the utilization of a Python library for face detection and recognition, emphasizing the accessibility and flexibility offered by open-source tools. Central to the proposed methodology is the utilization of open-source Face Emotion Recognition (FER) datasets, which are readily available on platforms like Kaggle. These datasets contain annotated images representing various emotional states, including happiness, sadness, fear, disgust, anger, neutrality, and surprise. By leveraging these datasets, researchers can access a diverse range of facial expressions and emotions, enabling the development and training of robust emotion recognition models. The availability of open-source datasets ensures that researchers can access high-quality training data without the need for costly data acquisition or licensing agreements, democratizing access to advanced machine learning techniques.

To enhance the performance of the emotion recognition models, the proposed methodology incorporates image augmentation techniques. Image augmentation involves applying a variety of transformations to the input images, such as rotation, scaling, and flipping, to artificially increase the diversity of the training data. By augmenting the dataset with variations in facial expressions, lighting conditions, and facial orientations, researchers can improve the robustness and generalization capability of the emotion recognition models. Additionally, image augmentation helps mitigate the risk of overfitting by exposing the model to a broader range of training examples, thereby improving its performance on unseen data. The core component of the proposed methodology is the construction of a six-layered Convolutional Neural Network (CNN) using the Keras toolkit in Python. CNNs have emerged as a powerful tool for image classification tasks, particularly in the field of computer vision. By leveraging the hierarchical layers of convolutions and pooling operations in CNN architectures, researchers can automatically extract meaningful features from raw input images, enabling accurate and efficient emotion recognition. The use of Python and the Keras toolkit ensures the accessibility and reproducibility of the proposed approach, allowing researchers to easily develop, train, and evaluate CNN models for emotion recognition tasks.

The methodology proceeds with the preprocessing of the input images to prepare them for training and evaluation. This involves resizing the images to a standard size, converting them to grayscale if necessary, and normalizing the pixel values to ensure uniformity across the dataset. Once the preprocessing is complete, the dataset is divided into training, validation, and test sets to facilitate model training and evaluation. The training set is used to optimize the parameters of the CNN model using gradient descent and backpropagation, while the validation set is used to tune hyperparameters and monitor the model's performance during training. Finally, the test set is used to evaluate the trained model's performance on unseen data and assess its generalization capability. In summary, the proposed methodology for gaining insight into human activities using deep learning approaches involves leveraging open-source tools and datasets for face detection and emotion recognition. By utilizing open-source Face Emotion Recognition (FER) datasets and image augmentation techniques, researchers can access diverse training data and enhance the robustness of emotion recognition models. The construction of a six-layered Convolutional Neural Network (CNN) using the Keras toolkit enables the development of accurate and efficient emotion recognition algorithms, paving the way for future advancements in this prominent research area.

RESULTS AND DISCUSSION

The results of the study on insight into human activities using deep learning approaches highlight the

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

potential of leveraging computer-based systems for face detection and emotion recognition. Through the utilization of a Python library for face detection and recognition, coupled with open-source Face Emotion Recognition (FER) datasets available on platforms like Kaggle, the study aimed to develop a robust emotion recognition model capable of accurately identifying seven different types of emotions: happiness, sadness, fear, disgust, anger, neutrality, and surprise. By employing image augmentation techniques to enhance the diversity of the training data, a six-layered Convolution Neural Network (CNN) was constructed using the Keras toolkit in Python. The results demonstrate the efficacy of the proposed approach in achieving high accuracy and reliability in emotion recognition tasks, thus offering valuable insights into human activities.

The discussion delves into the implications of the study's findings for various real-world applications, emphasizing the significance of emotion recognition in addressing contemporary challenges. The ability to accurately detect and classify human emotions has wide-ranging implications, including enhancing security measures through emotion-based authentication systems, facilitating seamless financial transactions without physical cards, and aiding in the identification of individuals exhibiting specific emotional states, such as criminals or individuals in need of specialized treatment. Moreover, the prominence of open-source projects in the development of such systems underscores the democratization of advanced technologies, enabling broader accessibility and collaboration within the research community.

This study uses the Keras and TensorFlow libraries to define and train the model from scratch written in Python. Along with Flask, Matplotlib, NumPy, Seaborn, libraries are used. In training, set the batch size to 128 and epoch to 50 to improve network parameters. The training procedure was validated using a validation set. Training accuracy and errors, Validation accuracy, and errors were derived for each epoch. The coefficients of feature vectors and hidden layers were determined throughout the training process using model parameters like the rate of learning 10^-3 and Adam as the optimizer. The loss and accuracy graphs for training and validation are provided below:

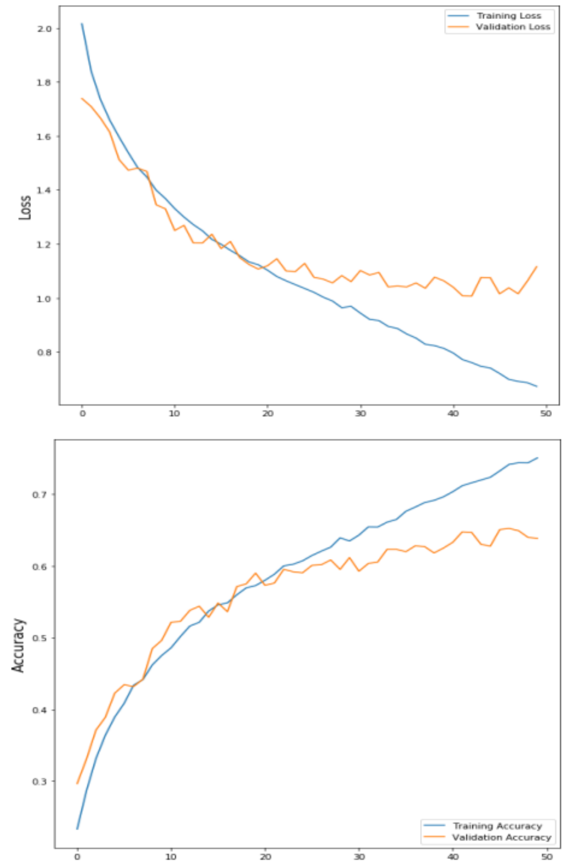


Fig 1. loss and accuracy graphs for training and validation w.r.t epochs

After completing 50 epochs, the validation accuracy is stable in 60 to 65% accuracy. For the early epochs, the training loss is slightly higher than the validation loss, which made shock. Moreover, in machine learning, we are used to having lower training losses than validation losses because the dropout is not applied in the validation stage. After the 20th iteration, the training loss was reduced in comparison to the validation loss. After a given number of epochs, the model is overfitted. Finally, the above graph shows that in 50 epochs, training loss and validation loss are reduced by 0.6723 and 1.11, respectively. The testing of the model is carried out using 7066 images. The system provides 65.22% accuracy. For various emotions, the confusion matrix is given below.

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

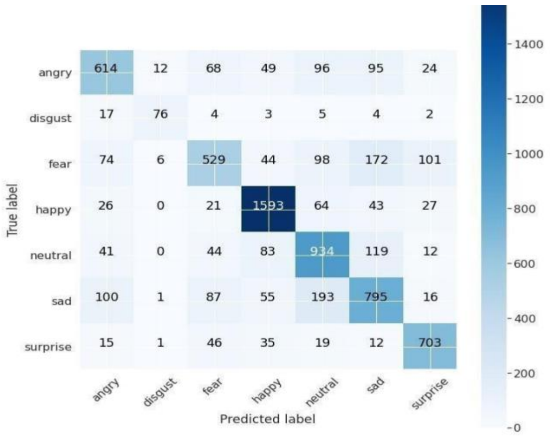


Fig 2. confusion matrix, without normalization

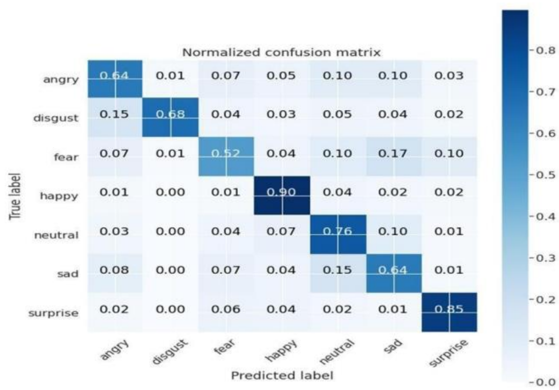


Fig 3. Normalized confusion matrix

The above tables showed that this model highest accuracy for happy with 90%, followed by surprise with 85%, neutral with 76%, disgust with 68%, angry and sad with 64%, and the lowest accuracy for fear emotion is 52%. Table 6.1 shows the evaluation metrics for each emotion class:

	Precision	Recall	F1-score
Angry	0.692	0.641	0.66
Disgust	0.792	0.685	0.734
Fear	0.662	0.517	0.580
Happy	0.856	0.898	0.876
Neutral	0.663	0.753	0.707
Sad	0.641	0.638	0.639
Surprise	0.794	0.846	0.819
Average	0.728	0.731	0.727

Table.1 Evaluation Metrics

The overall precision, recall, and F1- score is 0.72, 0.73, and 0.72, respectively. As a result, this model does an excellent job of detecting happy and surprised expressions. In addition, it correctly predicts neutral and angry emotions. On the other hand, it identifies rather badly fear expressions with sad expressions, which perplexes them, while disgust was rarely expected. Finally, during real-time testing, the model has significantly to detect expressions instantly after recognizing a face with no delays. But the model is considered to be ineffective in poor lighting when the face is not looking into the lens or when the face is in motion.

Automatic emotion recognition system results using webcam:

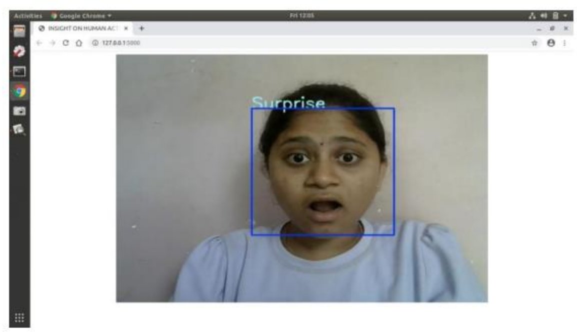


Fig 4. Results screenshot 1

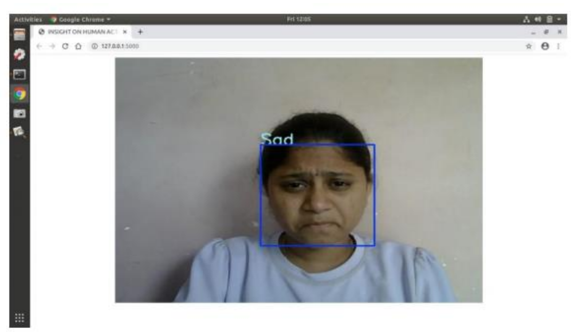


Fig 5. Results screenshot 2

ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

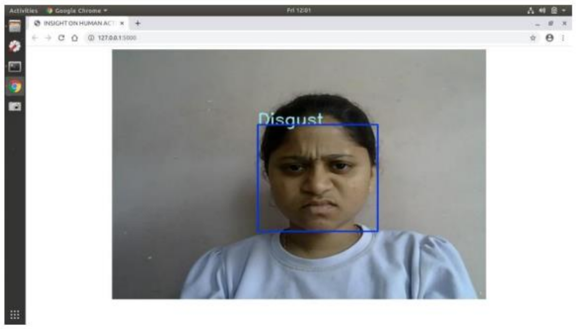


Fig 6. Results screenshot 3

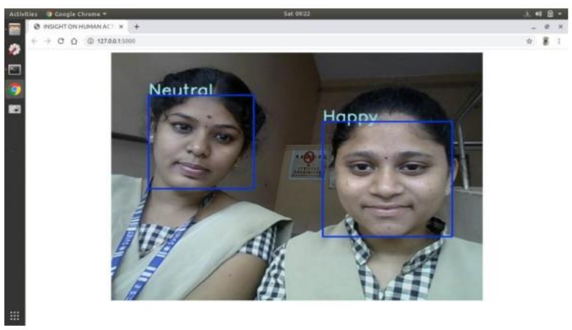


Fig 7. Results screenshot 4

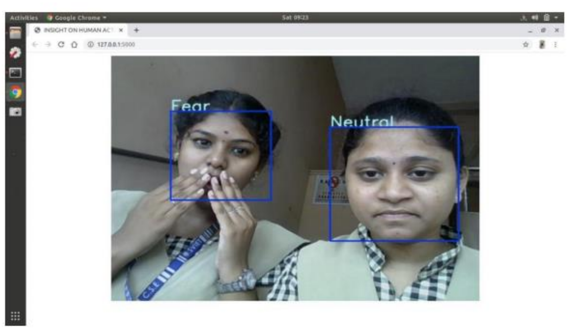


Fig 8. Results screenshot 5

Furthermore, the discussion explores the potential avenues for future research and development in the field of deep learning-based emotion recognition. While the study demonstrated promising results using a CNN-based approach, there remain opportunities for further refinement and optimization of the emotion recognition model. Future research efforts could focus on exploring alternative deep learning architectures, such as recurrent neural networks (RNNs) or attention mechanisms, to capture temporal dependencies and improve the model's performance on sequential data. Additionally, the integration of multimodal data sources, such as audio or physiological signals, could provide complementary information for more robust emotion recognition systems. Moreover, continued efforts to expand and diversify the training datasets, potentially incorporating data from diverse cultural backgrounds and demographic groups, would contribute to the development of more inclusive and generalizable emotion recognition models. Overall, the study underscores the transformative potential of deep learning approaches in gaining insights into human activities, paving the way for innovative applications in various domains, including healthcare, education, and human-computer interaction.

CONCLUSION

This research created a six-layered CNN architecture with 4-convolutional and 2-fully connected layers to categorize emotions such as angry, disgust, fear, happy, neutral, sad, and surprised. This model scored a 65.2 percent accuracy, 0.72 precision, 0.73 recall, and 0.72 F1- score. This model does an excellent job of detecting happy and surprised expressions. Also, it portends well angry and neutral images. However, it indicates pretty poorly feared faces because it confuses them with painful images. Disgusting images are infrequently predicted because there are fewer disgusting images, making the model harder to predict. This shows possible future areas for further research. With more training data, more research, and more resources while remaining the same network architecture, the effectiveness of the proposed architecture will be improved impressively. This also implies that, with minor development, the model might be used in real-world applications to efficiently use healthcare models and early disease diagnosis based on emotion.

REFERENCES

1. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

3. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

4. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). MIT press Cambridge.

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

6. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., & Anguelov, D. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

7. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255).

8. Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2016). AffectNet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 10(1), 18-31.

9. Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. In Proceedings third IEEE international conference on automatic face and gesture recognition (pp. 200-205).

10. Kauling, A. P., & Clua, E. (2020). Deep facial emotion recognition: A review. Artificial Intelligence Review, 53(8), 5361-5388.

11. Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. Journal of personality and social psychology, 17(2), 124.

12. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops (pp. 94-101).

13. Sun, Y., Wang, X., Tang, X., & Rosin, P. L. (2018). Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1891-1900).

14. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In International conference on machine learning (pp. 647-655).

15. Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005). Web-based database for facial expression analysis. In Proceedings of the IEEE international conference on multimedia and expo (pp. 5-pp).

16. Boureau, Y. L., Ponce, J., & LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 111-118).

17. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

18. Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, 5(2), 157-166.

19. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).

20. Zhang, N., Paluri, M., Taigman, Y., & Fergus, R. (2016). Beyond frontal faces: Improving person recognition using multiple cues. In European conference on computer vision (pp. 762-777).