Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

Enhanced Facial Expression Recognition via Deep Transfer Learning and Augmentation

Akshay Kumar[1], Dr Junaid Babar[2], Muhammad Khalid[3], Sadia Mujtaba[4]
[1]Research Scholar, Department of Computer Science, University of Balochistan, Quetta-Pakistan.
[2]Assistant Professor, Department of Computer Science, University of Balochistan, Quetta-Pakistan.
[3]Lecturer, Department of Computer Science, HITEC University, Taxilla-Pakistan.
[4]Assistant Professor, Department of Computer Science, College for Boys G-15, Islamabad-Pakistan.
akshaykumar210894@gmail.com, junaidbaber@ieee.org, mkmalghani@gmail.com, sadiamujtaba81@gmail.com

DOI: 10.5281/zenodo.10594199

ABSTRACT

Facial Expression is one of the key parts of non-verbal communication. Facial Expression Recognition is the major application of surveillance, automation, health care, and education. Deep learning is important in different fields of computer vision due to its ability to process and analyze large volumes of data, extract features, and correctly classification of images. This research empirically evaluates the performance of a pre-trained model on augmented datasets for facial expression recognition. The study includes preprocessing techniques, data augmentation, and transfer learning using the ResNet50 model. The experiments are conducted on a dataset containing images of three facial expressions: happy, sad, and surprised. The results indicate significant improvements in accuracy as the dataset size and preprocessing techniques increase. In particular, Cubic Support Vector Machine (SVM) and Linear Cubic SVM consistently outperform other classifiers, achieving an impressive accuracy of 99.7% on the augmented dataset. The research demonstrates the potential of data augmentation and preprocessing in enhancing facial expression recognition systems.

Keywords: Facial Expression Recognition, Deep Learning, Pre-Trained, Augmentation, Transfer Leaning, Resnet50

Cite as: Akshay Kumar, Dr Junaid Babar, Muhammad Khalid, & Sadia Mujtaba. (2024). Enhanced Facial Expression Recognition via Deep Transfer Learning and Augmentation. *LC International Journal of STEM, 4*(4), 1–9. https://doi.org/10.5281/zenodo.10594199

INTRODUCTION

Facial expression recognition has gained significant importance in recent years due to the advancements in deep learning and human-computer interaction [1]. Emotions are a crucial aspect of human communication, expressed through various means. Research has shown that non-posed expressions require additional physiological signals like temperature dynamics and heart rate for accurate analysis, but such measurements are often unavailable in real-world scenarios.

Traditionally, the video-based approach has been widely used for expression recognition. Earlier methods relied on hand-crafted features such as LBP, BoW, HoG, and SIFT, which showed promising results on various datasets [2]. The sequence-based approach further enhanced emotion recognition by incorporating temporal information from videos. The challenge of recognizing expressions in unconstrained environments, often referred to as "expression recognition in the wild," has garnered

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

substantial attention. This involves images collected from the internet under different lighting and pose conditions. Incorporating such diverse data, like EmotioNet, into training sets has proven to improve model generalization. Facial expressions play a significant role in human communication, conveying emotions and intentions universally. Studies on emotion recognition [3] have evolved from the initial model of six basic emotions, as defined by Ekman and Friesen, to more complex representations that capture the subtleties of daily affective displays. Despite these advancements, the categorical model based on discrete basic emotions remains a popular approach for facial expression recognition.

Facial expression recognition systems can be categorized into static image-based and dynamic sequence-based methods. Static methods extract spatial information from single images, while dynamic methods consider temporal relationships between consecutive frames in expression sequences. These methods have been extended to incorporate multiple modalities like audio and physiological signals for more accurate recognition. Traditionally, hand-crafted features and shallow learning techniques were common in facial expression recognition. However, the availability of large training datasets and increased processing power has led to the adoption of deep learning methods. Deep neural networks, such as CNNs, have shown remarkable results, outperforming previous techniques and enabling accurate recognition of expressions in challenging real-world scenarios [4]. Despite the success of deep learning, challenges remain, including the need for extensive training data to prevent overfitting and addressing inter-subject variations, pose variations, illumination changes, and occlusions. The evolution of facial expression recognition involves the integration of deep learning techniques, better datasets, and advanced network architectures [5].

In this research, we will empirically evaluate the performance of pertained model on different augmented dataset for facial expression recognition. This Research is also focused on preprocessing for enhancement of image.

LITERATURE REVIEW

Previous Studies

The paper titled "Hierarchical Deep Learning for Facial Expression Recognition with Fused Appearance and Geometric Features" addressed the increasing relevance of interaction technology in the context of AI advancements. The authors introduced [6] a novel approach grounded in hierarchical deep learning that combines appearance and geometric features. The paper also presents an autoencoder-based technique to generate neutral facial images, enabling the extraction of dynamic emotional features. Evaluation on CK+ and JAFFE datasets demonstrates the method's efficacy, achieving 96.46% accuracy for CK+ and 91.27% for JAFFE through ten-fold cross-validation. This model generate a good accuracy but on small scale dataset.

The authors published [7] an innovative approach using a hybrid deep learning model. The methodology commences with the utilization of two distinct deep Convolutional Neural Networks (CNNs): a spatial CNN that processes static facial images, and a temporal CNN that handles optical flow images. The authors employ a linear Support Vector Machine (SVM) for the classification of facial expressions. Through extensive experimentation across three publicly available video-based facial expression datasets—BAUM-1s, RML, and MMI. The training the two models are more costly in case of computation.

The present paper [8] introduces a novel facial expression recognition method grounded in the CNN model. Acknowledging the significance of the CNN's hierarchical structure, the study zeroes in on the pivotal role played by the activation function—a bedrock that imparts nonlinearity, endowing deep

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

neural networks with genuine artificial intelligence while the Rectified Linear Unit (ReLu) activation function ranks among the most proficient, it is not without its limitations. The issue of neuron necrosis—stemming from the ReLu function's derivative always being zero for negative input values—is addressed in this study. This design aligns with CNN's activation function principles and offers a solution to the aforementioned challenge. The study conducts an extensive analysis and comparison of five widely-used activation functions—namely sigmoid, tanh, ReLu, leaky ReLUs, and softplus–ReLu—alongside the newly-proposed activation function. The experimental evaluation unfolds across two prominent public facial expression databases, namely JAFFE and FER2013. They can't conducted experiments with different models.

This study [9] shows an innovative deep learning framework that seamlessly combines Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) cells, aiming to elevate real-time FER capabilities. The proposed framework encompasses three fundamental components. Firstly, it employs enhanced pre-processing techniques to address illumination variations and retain nuanced edge information in individual images. Secondly, the processed images are subject to separate processing within two distinct CNN architectures, effectively extracting spatial features from each image. Lastly, the spatial feature maps generated by these parallel CNN layers are seamlessly fused and integrated with an LSTM layer.

This study presented [10] a novel framework designed to simultaneously harness spatial features and temporal dynamics for improved FER. The proposed framework operates by sequentially extracting spatial features from individual frames of an expression image sequence, achieved through a deep network. The culmination of this process involves the amalgamation of insights from the fused features via a Bidirectional Long Short-Term Memory (BiLSTM) network. To validate the efficacy of the proposed approach, comprehensive experiments were conducted on three benchmark databases—namely, CK+, Oulu-CASIA, and MMI. The empirical results unequivocally demonstrate the superiority of the presented framework, consistently outperforming state-of-the-art methodologies. They performed experience on small dataset.

The proposed [11] system unfolds across four key components, each contributing to the refined functionality of the FER. The first component focuses on isolating a region of interest through face detection, executed within the context of the input image. Subsequently, the second component introduces a novel deep learning-based convolutional neural network (CNN) architecture for the purpose of feature learning, a pivotal task in classification that facilitates the precise identification of expression types. To further amplify the effectiveness of the system, the third component employs inventive data augmentation techniques, strategically applied to the facial images. These techniques enrich the learning parameters of the CNN model, thereby enhancing the system's overall performance. In the fourth component, a delicate balance is achieved between data augmentation and deep learning features. The empirical validation of the proposed system is extensively demonstrated through experiments conducted on three benchmark databases: KDEF (comprising seven expression classes), GENKI-4k (involving two expression classes), and CK+ (encompassing seven expression classes). The performances achieved on each database are meticulously presented and detailed, showcasing the system's prowess but these are on small datasets. Its need to implements in real life scenarios.

This paper [12] gives a novel strategy for human facial expression recognition, incorporating an enhanced version of the Cat Swarm Optimization (CSO) algorithm named Improved Cat Swarm Optimization (ICSO). Deep Convolutional Neural Network (DCNN) techniques are employed to extract intricate facial features from the input images. ICSO is leveraged to select optimal features that

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

distinctly characterize a person's facial expression. The fusion of DCNN with ICSO substantially enhances the system's retrieval efficacy.

METHODOLOGY

Our proposed approach implements deep transfer learning and data augmentation to enhance facial expression recognition. In first step we use preprocessing for enhancement of images. Then we select a state-of-the-art pre-trained deep neural network. The pre-trained model is fine-tuned by changing its last three layers for our facial expression problem. Extracted features are used for classification.

Dataset

The dataset use for experiments are publically available on kaggle with three facial expressions that are happiness, sadness and surprise. The dataset contain 3689 images of different facial expression gets from different online website. Its contain 1435 image of happy, 1165 of sad 1089 of surprise expression. The sample images from dataset are shown in figure 1.

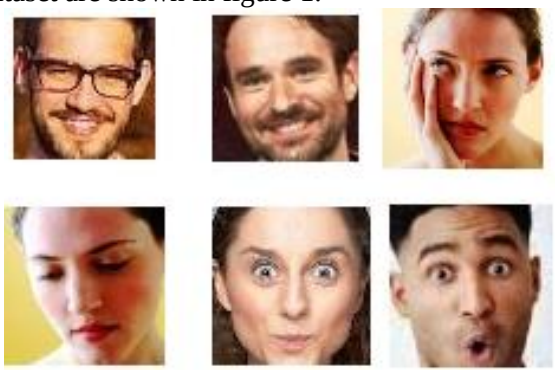

Figure 1: Different Facial Expression Sample Images from Dataset

Preprocessing

Preprocessing are the operations on dataset before training the model. In case of images we can enhancement the image with different preprocessing operations. In our research we apply the Bi-Histogram Equalization on Image for Enhancement. Sample of normal dataset image and preprocessed dataset image are illustrated in figure 2.

Figure 2: Before and after Bi-Histogram Equalization on Facial Expression Images

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

Dataset Augmentation

Data augmentation is a technique commonly used in deep learning to artificially expand the size of a training dataset by applying various transformations [13] to the existing data like rotation, flip-flop etc. The goal of data augmentation is to improve the model's generalization performance by exposing it to a wider variety of examples without actually collecting new data from the real world. This is particularly beneficial when the available dataset is limited, as it helps mitigate over fitting and enhances the model's ability to handle different variations of the same input. In our data augmentation approach, its starts by loading and then performs rotation augmentation by randomly applying a rotation angle within a predefined range to each image. We also do this augmentation with different number of images generated. Table 1 is number of images that are generated and used for experiments.

Name	Happy	Sad	Surprise	Total	No of New Generated images	%age of dataset extension
DS1 (Original)	1435	1165	1089	3689	0	0%
DS2 (Balanced)	1435	1435	1435	4305	616	14.30%
DS3	2000	2000	2000	6000	2311	62.64%
DS4	5000	5000	5000	15000	11,311	306.61%

Fine Tuning of ResNet50

Transfer learning has emerged as a powerful technique in deep learning, allowing pre-trained models to be adapted to specific tasks with limited data. In this approach we use transfer learning and fine-tuning of ResNet50 pre trained model for facial expression recognition. ResNet50 [14] introduced the concept of residual blocks, which address the vanishing gradient problem in deep networks, allowing for the training of very deep models.

Classification

Classification process are applied on feature vector to predict the instance in true class. In this experiments classifiers used are Wide Neural Network (WNN), Narrow Neural Network (NNN), Medium Neural Network (MNN), Bi-Layered Neural Network, Tri-Layered Neural Network, Cubic Support Vector Machine (SVM) and Linear Cubic Support Vector Machine (SVM), Among them Cubic and Linear SVM outperform as compared to other.

DATA ANALYSIS AND RESULTS

Results

We evaluated original and augmented datasets on different classifier for classification of facial expression recognition. Table 2 is showing different classifier accuracy on different augmented datasets. They classifier Wide Neural Network (WNN), Narrow Neural Network (NNN), Medium Neural Network (MNN), Bi-Layered Neural Network, Tri-Layered Neural Network, Cubic Support Vector Machine (SVM) and Linear Cubic Support Vector Machine (SVM) are epically evaluated on Different dataset in which DS1 original, DS2 is Balanced, DS3 is Augmented with 2000 images in each and DS4 is augmented with 5000 images in each. In these case Cubic SVM and Linear SVM outperform in overall accuracy with 99.7%.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

Dataset	WNN	NNN	MNN	BNN	TNN	Cubic SVM	Linear SVM
DS1	95.0%	94.1%	94.2%	93.1%	93.8%	95.4%	95.9%
DS2	96.7%	96.2%	96.6%	96.0%	95.8%	97.1%	97.1%
DS3	97.7%	97.2%	98.0%	97.5%	96.8%	98.3%	98.2%
DS4	99.7%	99.6%	99.7%	99.6%	99.5%	99.7%	99.7%

Table 2: Results of Different classifier on without preprocessing datasets.

In Table 3, results of augmented dataset with Bi-Histogram Equalization preprocessing are illustrated for different classifier. Cubic SVM and Linear SVM also generate highest result in term of accuracy of 99.5% and 99.6%.

Dataset	WNN	NNN	MNN	BNN	TNN	Cubic SVM	Linear SVM
Bi-DS1	96.0%	95.1%	96.3%	94.8%	95.0%	96.5%	96.4%
Bi-DS2	97.9%	97.1%	98.0%	97.0%	96.9%	98.2%	98.2%
Bi-DS3	97.5%	96.5%	97.2%	96.1%	95.8%	97.6%	97.6%
Bi-DS4	99.4%	99.3%	99.3%	99.3%	99.3%	99.6%	99.5%

Table 3: Results of Different classifier on with preprocessing datasets

Table 4 is representing the time of each classifier for DS4 dataset which have 5000 instance in each class. In term of time Medium Neural Network take 37.101 seconds while the accuracy of MNN is 99.6 as compared to Cubic SVM which have accuracy of 99.7% and Classification time is 116.8 seconds.

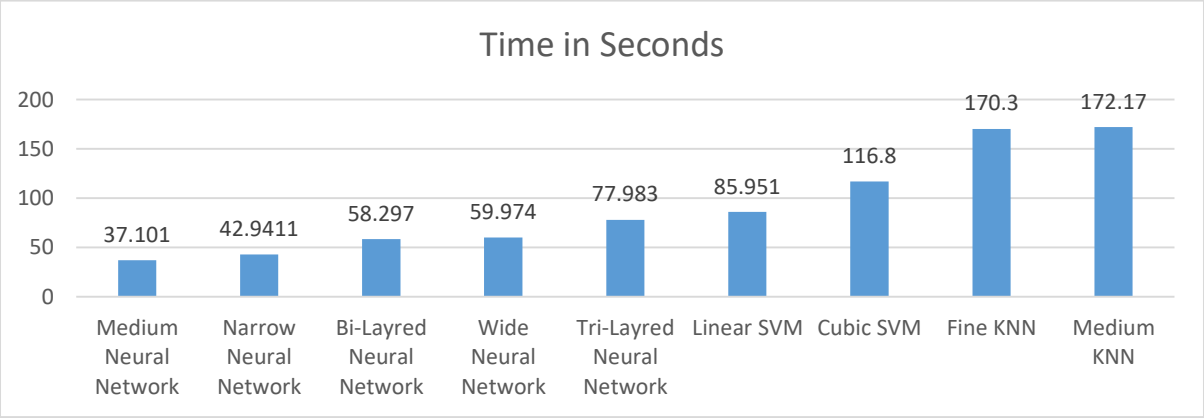

Table 4: Time Graph of different Classifiers on DS4

Table 5 is confusion matrix of Cubic SVM for DS4 dataset. In which accuracy of Happy, sad and surprise class is 99.6%. True Positive and true negative of each class also represent in Table 5.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

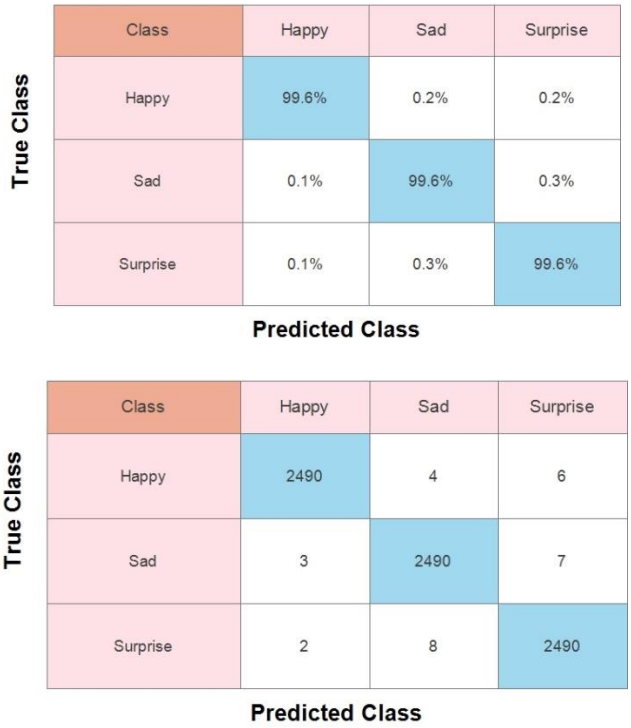

Table 5: Confusion Matrix of accuracy and True Positive Classes of Cubic SVM classifier on DS4

CONCLUSION AND RECOMMENDATIONS

Conclusion

In this research, we present a comprehensive investigation into facial expression recognition using transfer learning with deep learning, data augmentation, and preprocessing techniques. The major finding of this paper is that data augmentation is a powerful tool for enhancing the performance of facial expression recognition systems, especially when dealing with limited datasets. Augmenting the dataset with thousands of additional images significantly improves classification accuracy. Preprocessing techniques, such as Bi-Histogram Equalization, can enhance the quality of facial expression images, leading to improved recognition results. Transfer learning, specifically adapting new layers of the ResNet50 model for our problem, proves effective in capturing complex features for facial expression recognition. Among the classifiers evaluated, the Cubic Support Vector Machine (SVM) and Linear Cubic SVM consistently deliver the highest accuracy, reaching an impressive 99.7% on the augmented dataset. In the future, there is potential to apply customized models to further improve the accuracy of facial expression recognition, either through augmentation or the addition of new images to the dataset.

Recommendation

The potential for the implementation of customized models seems encouraging for further research in this area. These models have the potential to improve facial emotion identification accuracy even more by adding new instances to the dataset or by extending the augmentation approaches e.g Generative adversarial networks. This work opens the door for the development of more complex and advanced techniques, which will ultimately aid in the ongoing enhancement of facial expression recognition software.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

REFERENCES

[1] Kuo, C. M., Lai, S. H., & Sarkis, M. (2018). A compact deep learning model for robust facial expression recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 2121-2129).
[2] Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. IEEE transactions on affective computing, 13(3), 1195-1215.
[3] Fathallah, A., Abdi, L., & Douik, A. (2017, October). Facial expression recognition via deep learning. In 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA) (pp. 745-750). IEEE.
[4] Zhao, X., Shi, X., & Zhang, S. (2015). Facial expression recognition via deep learning. IETE technical review, 32(5), 347-355.
[5] Abdullah, S. M. S., & Abdulazeez, A. M. (2021). Facial expression recognition based on deep learning convolution neural network: A review. Journal of Soft Computing and Data Mining, 2(1), 53-65.
[6] Kim, J. H., Kim, B. G., Roy, P. P., & Jeong, D. M. (2019). Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. IEEE access, 7, 41273-41285.
[7] Zhang, S., Pan, X., Cui, Y., Zhao, X., & Liu, L. (2019). Learning affective video features for facial expression recognition via hybrid deep learning. IEEE Access, 7, 32297-32304.
[8] Wang, Y., Li, Y., Song, Y., & Rong, X. (2020). The influence of the activation function in a convolution neural network model of facial expression recognition. Applied Sciences, 10(5), 1897.
[9] Rajan, S., Chenniappan, P., Devaraj, S., & Madian, N. (2020). Novel deep learning model for facial expression recognition based on maximum boosted CNN and LSTM. IET Image Processing, 14(7), 1373-1381.
[10] Liang, D., Liang, H., Yu, Z., & Zhang, Y. (2020). Deep convolutional BiLSTM fusion network for facial expression recognition. The Visual Computer, 36, 499-508.
[11] Umer, S., Rout, R. K., & Pero, C. (2021). Facial expression recognition with trade-offs between data augmentation and deep learning features. J Ambient Intell Humaniz Comput, 1–15.
[12] Sikkandar, H., & Thiyagarajan, R. (2021). Deep learning based facial expression recognition using improved Cat Swarm Optimization. Journal of Ambient Intelligence and Humanized Computing, 12, 3037-3053.
[13] DeVries, T., & Taylor, G. W. (2017). Dataset augmentation in feature space. arXiv preprint arXiv:1702.05538.
[14] Mukti, I. Z., & Biswas, D. (2019, December). Transfer learning based plant diseases detection using ResNet50. In 2019 4th International conference on electrical information and communication technology (EICT) (pp. 1-6). IEEE.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

AUTHORS PROFILE

Akshay Kumar is a highly dedicated and skilled professional with a robust educational background and significant expertise in the field of computer science. He completed his Bachelor's in Computer Science from Usman Institute of Technology, affiliated with Hamdard University Karachi. Akshay currently serves as a Database Administrator in the Colleges of Higher & Technical Education in Balochistan. His keen research interests lie in the realms of Facial Expression Recognition, Natural Language Processing, and Social Media Analytics. Akshay is passionate about exploring the intersection of technology and human interaction, contributing valuable insights to these cutting-edge domains.

Junaid Baber received the MS and Ph.D. in computer science from the Asian Institute of Technology, Thailand. He spent one year as a research scientist at the national institute of informatics, Tokyo. Currently, he is working as a faculty member at University of Balochistan, Quetta. His research interests lie in machine learning, high performance computing, and data analytics.
Affiliation: Dept of CS\&IT, University of Balochistan, Quetta, Pakistan

Muhammad Khalid is a highly motivated and skilled professional with a strong educational background and extensive experience in the field of computer science. He is currently in the research stage of his Ph.D. in Computer Science at HITEC University, Taxila, Pakistan. Khalid completed his Master of Science in Computer Science at the University of Balochistan, and he also holds a Bachelor of Science in Computer Science from the Institute of Southern Punjab. He is currently working as a Lecturer in Computer Science at HITEC University, Taxila. His research interest is Deep Learning for Surveillance and Natural Language Processing.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0)