DEVELOPING A CONVOLUTIONAL NEURAL NETWORK (CNN) MODEL FOR FACIAL EXPRESSION RECOGNITION (FER)

cct College Dublin Computing · IT · Business

Danrlei Martins & Leonardo Diesel, May 2024

ABSTRACT

This Capstone Project focused on developing an accurate Facial Expression Recognition (FER) model by leveraging deep learning techniques, specifically Convolutional Neural Networks (CNNs). The objective was to explore, design, and implement custom architectures and evaluate their performance against existing work. The process involved several stages, such as data preprocessing, data augmentation, architecture design, hyperparameter tuning, and performance assessment using metrics like accuracy and F1-score while utilizing the FER-2013 dataset for training. The resulting FER model exhibited competitive accuracy levels and generalization capabilities, opening up opportunities for real-time implementation and application across various domains.

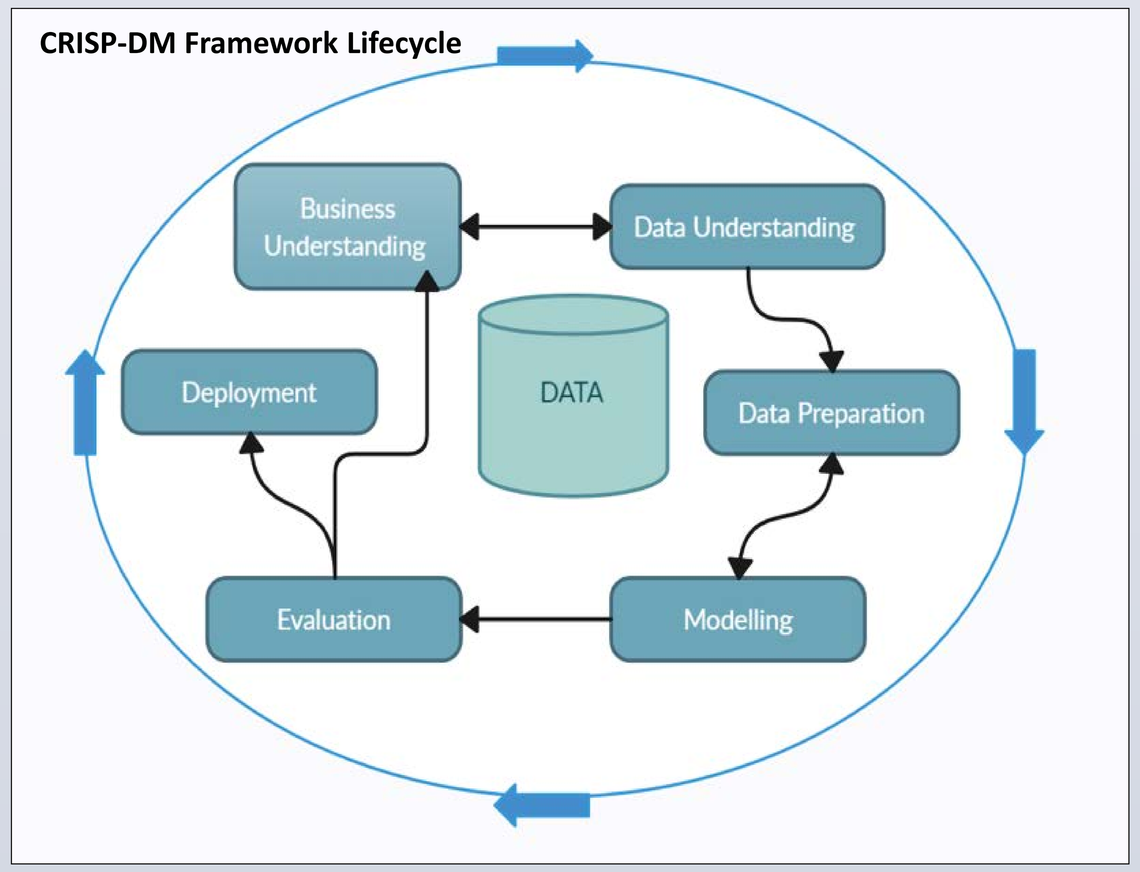
RESEARCH QUESTIONS

The four questions below represent the core focus of the entire study:

- **Q.1.** Is it possible to develop an accurate CNN model for facial expression recognition (FER)?
- **Q.2.** What preprocessing techniques can enhance the performance of the FER model?
- **Q.3** How does our proposed CNN architecture compare to existing models in terms of accuracy and generalization?
- **Q.4** What are the potential real-world applications and deployment considerations for the developed FER model?

CONCEPTUAL FRAMEWORK

The project followed the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which provided a structured approach to data mining and machine learning projects. The six phases included business understanding, data understanding, data preparation, modelling, evaluation, and deployment.



BUSINESS UNDERSTANDING

The FER industry is expanding and has a lot of promise for companies in various sectors. According to Fortune Business Insights (2023), the market value is estimated to reach USD 74.80 billion by 2029.

From enhancing customer service and user experience by measuring satisfaction and engagement levels to improving mental health support by detecting emotional distress to advancing human-robot interactions and emotional AI assistants – real-world applications of FER are vast (Samadiani et al., 2019).

DATA UNDERSTANDING

We trained our machine learning models using the FER-2013 dataset. The dataset was developed by Pierre Luc Carrier and Aaron Courville and was introduced at the International Conference on Machine Learning (ICML) in 2013 (Goodfellow et al., 2013). The dataset contains 35,887 images of facial expressions, each labelled with one of seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral.

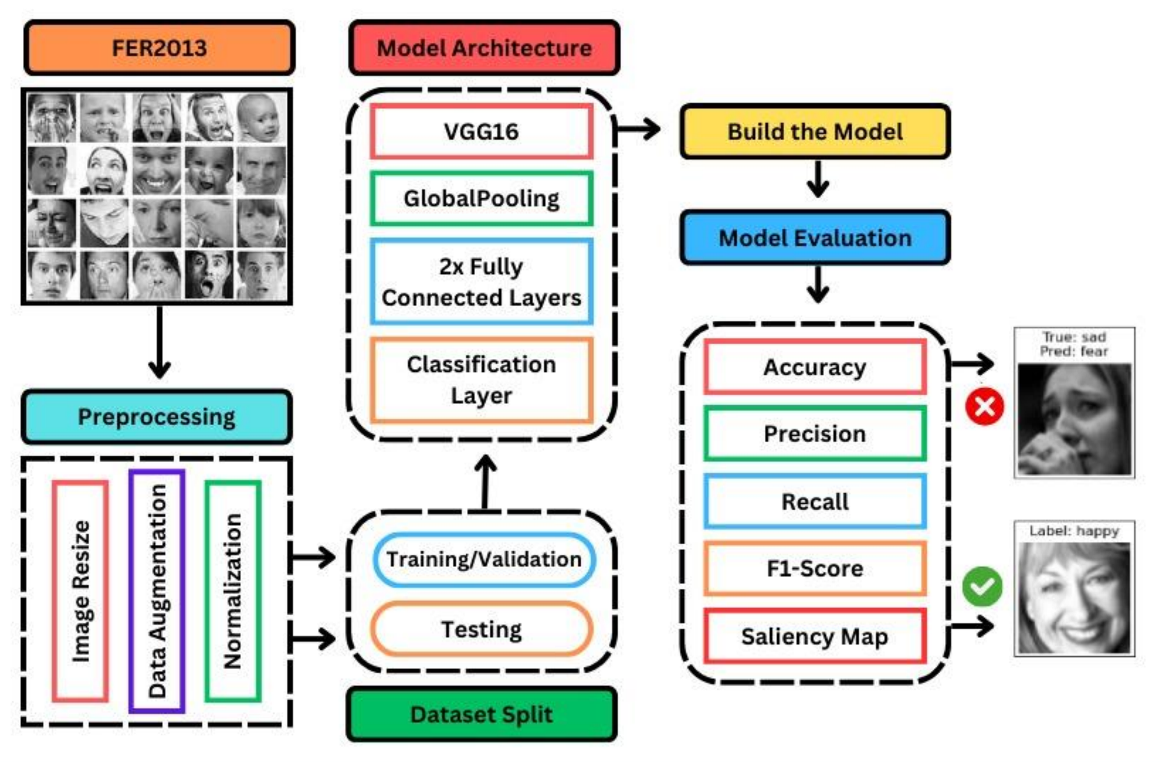
DATA PREPARATION

We employed data augmentation techniques (rotation, zoom, flipping, shifting) to increase diversity and balance classes. Data augmentation is a widely used approach in deep learning models, particularly for image data, as it helps increase the training dataset's diversity and size, reduce overfitting, and improve the model's generalization capabilities (Shorten and Khoshgoftaar, 2019).

The data was split into: Training (22,968 images), validation (5,741 images), and testing (7,178 images) sets.

MODELLING

We chose to work with the VGG16 architecture, a popular CNN model. It was developed by the Visual Geometry Group (VGG) at the University of Oxford and presented at the International Conference on Learning Representations (ICLR) 2015 conference (Simonyan and Zisserman, 2015). We focused on applying transfer learning techniques, leveraging pre-trained models existing knowledge and fine-tuning for enhanced performance.



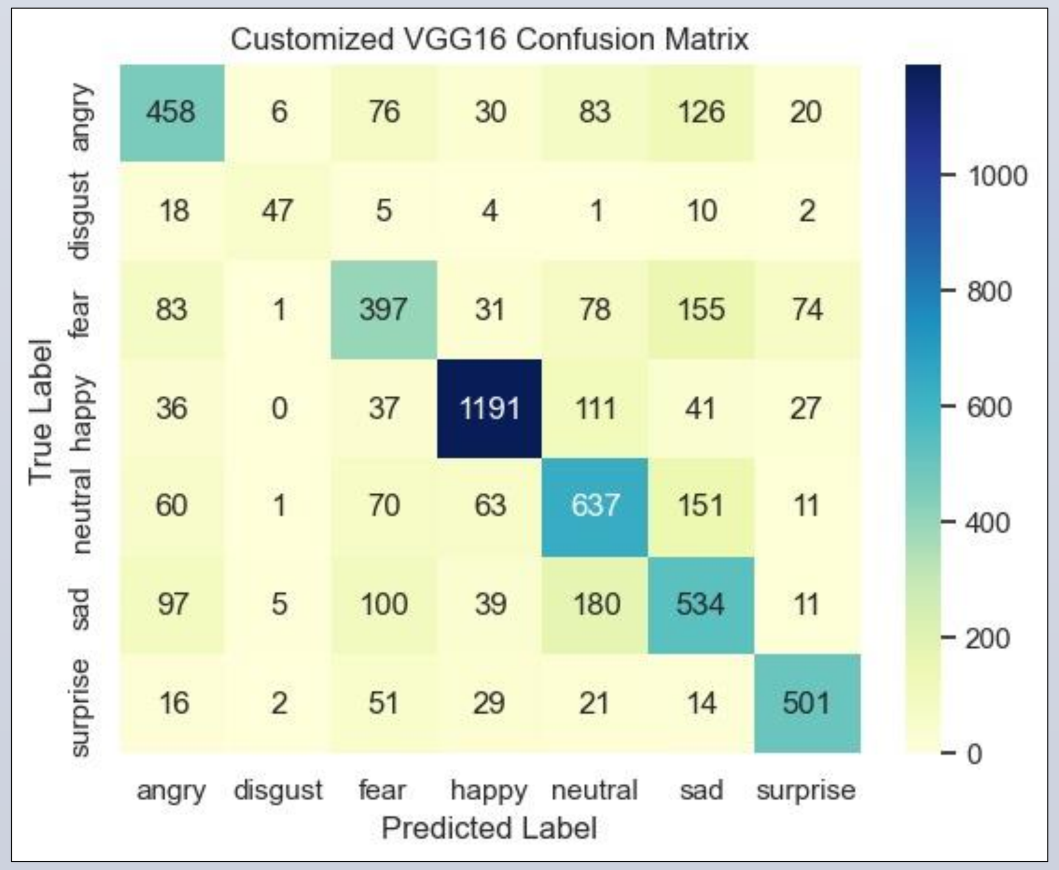
Proposed Custom CNN Model

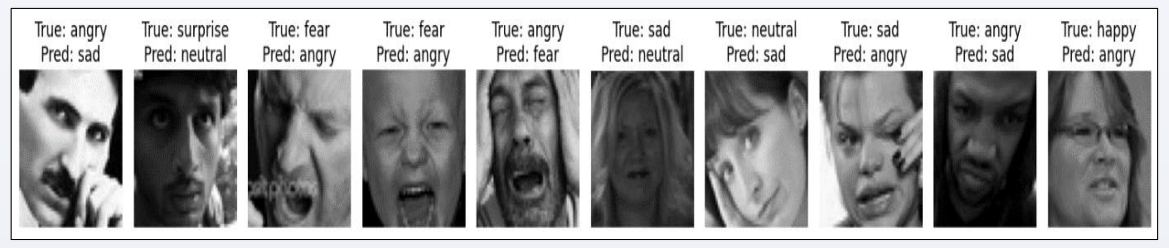
The custom model was built upon the pre-trained VGG16 model as the base, but we made several modifications to the architecture. Firstly, we applied Global Average Pooling (GAP) to the output of the base model, which helped reduce the number of parameters and prevent overfitting.

Additionally, we added two dense layers with 4096 and 1024 units, respectively, along with ReLU activation functions and dropout layers to introduce non-linearity and regularization. The final layer was dense with seven units and a softmax activation function to perform the multi-class classification task.

EVALUATION

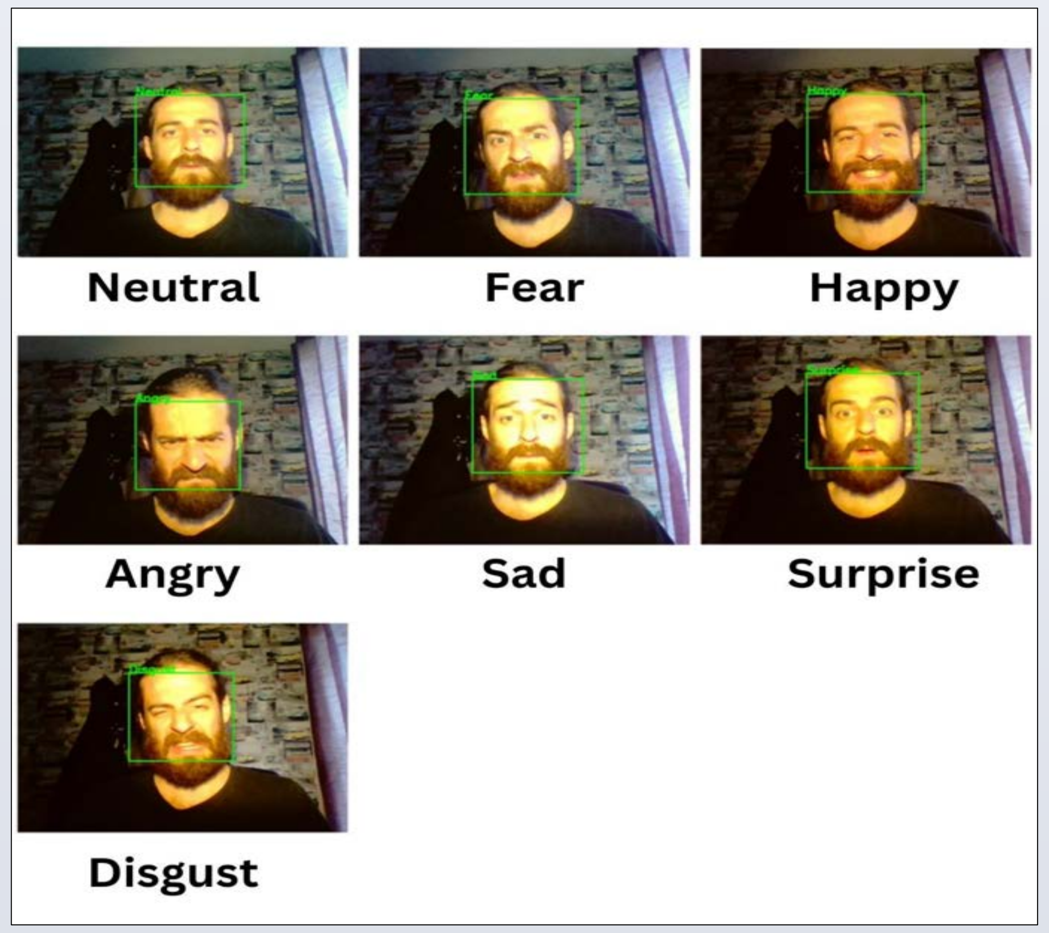
In the evaluation phase, we assessed the performance of our trained models using various metrics, including accuracy, precision, recall, F1-score, and confusion matrices. These metrics provided insights into the models' strengths and weaknesses in accurately classifying facial expressions across the seven categories present in the dataset. The custom model achieved a test accuracy of 67.11%, outperforming the baseline VGG16 model's test accuracy of 61.94%.





DEPLOYMENT

After successfully training and evaluating our FER model, we explored the deployment of our solution in a real-world scenario. We developed a real-time FER system using OpenCV, a widely used computer vision library. The system utilized the trained custom model and OpenCV's pre-trained Haar Cascade Classifier for object detection to locate and extract facial regions from the video feed. This deployment showcased the practical application and potential impact of our FER model, demonstrating its effectiveness in real-time scenarios across various domains.



CONCLUSIONS

A summarised answer to the research questions are presented below:

- **Q.1.** Yes, this project demonstrates the development of an accurate CNN model for FER using the FER-2013 dataset. The custom CNN architecture, leveraging transfer learning and hyperparameter tuning, achieved a competitive test accuracy of 67.11%.
- **Q.2.** Data augmentation techniques such as rotation, zooming, flipping, and shifting were applied to increase the diversity and balance of the training data, mitigating class imbalance issues. This preprocessing step significantly improved the model's performance and generalization capabilities.
- **Q.3** The custom model achieves competitive or superior accuracy compared to existing work on the FER-2013 dataset.
- **Q.4** The developed FER model has practical applications across various domains, including healthcare, education, customer behavior analysis, and advertising. It was successfully deployed in a real-time facial expression recognition system using OpenCV, showcasing its potential for integration into broader applications involving human-machine interaction and emotional AI.

REFERENCES

Fortune Business Insights (2023). Global Emotion Detection and Recognition Market Size, Share & Growth Analysis By Type, By Application with Regional Forecast, 2022-2029. [online] Available at: https://www.fortunebusinessinsights.com/industry-reports/emotiondetection-and-recognition-market-101326 [Accessed 5 Mar. 2024].

Goodfellow, I.J., Erhan, D., Luc Carrier, P., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J. and Ionescu, R. (2013). Challenges in representation learning: A report on three machine learning contests. ICONIP 2013: Neural Information Processing, [online] 64, pp.117–124. doi:https://doi.org/10.48550/arXiv.1307.0414.

Samadiani, Huang, Cai, Luo, Chi, Xiang and He (2019). A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data. Sensors, [online] 19(8), p.1863. doi:https://doi.org/10.3390/s19081863.

Shorten, C. and Khoshgoftaar, T.M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, [online] 6(1). doi:https://doi.org/10.1186/s40537-019-0197-0.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science: Computer Vision and Pattern Recognition. [online] doi:https://doi.org/10.48550/arXiv.1409.1556.