Nima C Sherpa
Data Science 2025 W
Student No: 0352428

Facial Emotion Recognition With Machine Learning.

Data Problem and Dataset

Even as human emotions from facial expression are inherently challenging to predict, due to emotional ambiguity. This project explores various visualization of emotional relationship, model learned pattern and developing an automated facial emotion recognition system capable of classifying expressions like **anger, sadness, happiness, fear, disgust, neutral, contempt,** and **surprise.** The **Emotion_labled** dataset utilized in this project was uploaded by **AffectNet** dataset from **Kaggle**, consisting of **29,042 RGB images** resized to **96x96** pixels, preprocessed using Principal Component Analysis(PCA) to eliminate low-detail images via the **relFCs** threshold. Diving deeper into image processing and learning its computational performance issue all the images were converted to grayscale reducing **three** time the vector features, **Flattened** and **normalized** each image, saved in a **NPZ (NumPy)** compressed gzip file with its corresponding emotional labels.

Project Goals and Real-World Implications

The initial goal was to create a robust emotion recognition model that can help **search engines** produce accurate more precise outputs and suggest more images of same class, particularly assisting digital illustrators and artists by providing **accurate facial expression references**. Real-World implication extends beyond art, impacting **security systems** through facial recognition, aiding **mental health** by detecting emotional distress in patients, and potentially advancing medical diagnostics by recognizing subtle visual health cues. Image processing and pattern recognition are now a key to interpret the surroundings and maker real-time decisions for operating **Autonomous vehicles** and **Driver Assistance System. ANPR**

(Automatic Number Plate Recognition) is Widely used in law enforcement and traffic monitoring by extracting license plate data from images. It helps track vehicle movements, enforce road rules, and manage toll systems. However, it also raises privacy concerns due to potential surveillance and data misuse. (Wikipedia contributors, 2025)

Analysis Methodology

Initially basic traditional model like SVM (Support Vector Machines), Logistic Regression, Random Forest, and MLP (Neural Network) were explored. SVM was excluded due to very high computational cost as it relies on solving quadratic programming problems. PCA components with different number of components were used to keep most meaningful features and massively improving computation load, before feeding it to the models for test and train. Traditional models faced challenges resulting in low cross-validation score , MLP learning the pattern best amongst these models with **(0.4820) cross-validation score.** To analyze how well the model performed the ebeddings were extracted to learn the confusion matrix, **PCA-unsupervised linear** visualization to compare it with true-emotion and finally some **non-linear** visualization **(t-SNE and U-map) .** Balancing it's local and global structure u-map showed the most promising result without the help of labels but was struggling to differentiate emotions with Suttle differences like sad, surprise and disgust. Therefore, a **Convolutional Neural Network (CNN)** was selected due to its ability to capture spatial patterns in image data. The CNN utilized four CONV2D layers with 'ReLU' activation for finding different patterns in images, Batch Normalization for maintain smooth generalization with no biasness, MaxPooling for keeping the meaningful features and Dropout layers to randomly drop neurons to reduce overfitting. Data augmentation techniques to rotate, zoom, flip and brightness adjustments of images for more data and balance class weight mitigated class imbalance. Alos, **incorporating feedback,** a lower learning rate of 0.0001 was used, resulting in smooth taring curves without deep low plateaus in validation accuracy.

Nima C Sherpa
Data Science 2025 W
Student No: 0352428

Results and Analysis

The CNN model significantly outperformed traditional models, achieving approximately **77% training accuracy** and **63.9% validation accuracy.** CNNs ideally designed to learn from spatial hierarchies in image data, proved to be far more suited for facial expression task compared to traditional flat-vector-based models. The training logs shows steadily increasing accuracy, with no sign of underfitting or overfitting up to the best model at epoch 17. Early stopping captured the optimal weight, and the validation loss was stabilized by early stopping and dropout regulation, resulting in improved model consistently across each epoch. Later analyzing at two confusion matrices MLP confusion Matrix diagonal lines were less prominent, many predictions scattered across unrelated classes and the model confused massively with similar classes like fear and anger. But CNN confusion Matrix showed strong diagonal line, although the same misclassification occurred in fear, sad, and disgust, which are known to have visual overlaps, but the model provided high confidence and clarity in predictions for distinctly expressive emotions. From the classification report **Macro average F1-score = 0.62,** indicating reasonably balanced model, class balancing and image augmentation clearly helped improve recall for underrepresented classers and helped Dense layer of CNN learn better.

Future Work and Limitations

The project’s success lies in creating a CNN that effectively addresses ambiguity and reduces misclassification compared to baseline models. Yet, some confusion remains among subtle emotional classes. Future directions include exploring deeper CNN architectures, transfer learning from pre-trained models and potentially utilizing hybrid models to enhance robustness. Collecting more distinct data that can be easily classified by humans too for ambiguous expressions can be the key to improving confusions for the model. Evaluating the model in real-world settings, such as live facial recognition applications or interactive art tools, will provide further insight into is practical usability and performance. Collaborations with artists or psychologist could refine label accuracy and introduce new motion categories or blends that better reflect human experience. (Minaee et al., 2021)

Nima C Sherpa
Data Science 2025 W
Student No: 0352428

Citation

1. Du, Z., Wu, S., Huang, D., Li, W., & Wang, Y. (2019). Spatio-Temporal Encoder-Decoder Fully convolutional network for Video-Based dimensional emotion Recognition. IEEE Transactions on Affective Computing, 12(3), 565–578. https://doi.org/10.1109/taffc.2019.2940224

2. Minaee, S., Minaei, M., & Abdolrashidi, A. (2021). Deep-Emotion: Facial expression recognition using attentional convolutional network. *Sensors*, *21*(9), 3046. https://doi.org/10.3390/s21093046

3. Wikipedia contributors. (2025, March 30). *Automatic number-plate recognition*. Wikipedia. https://en.wikipedia.org/wiki/Automatic_number-plate_recognition