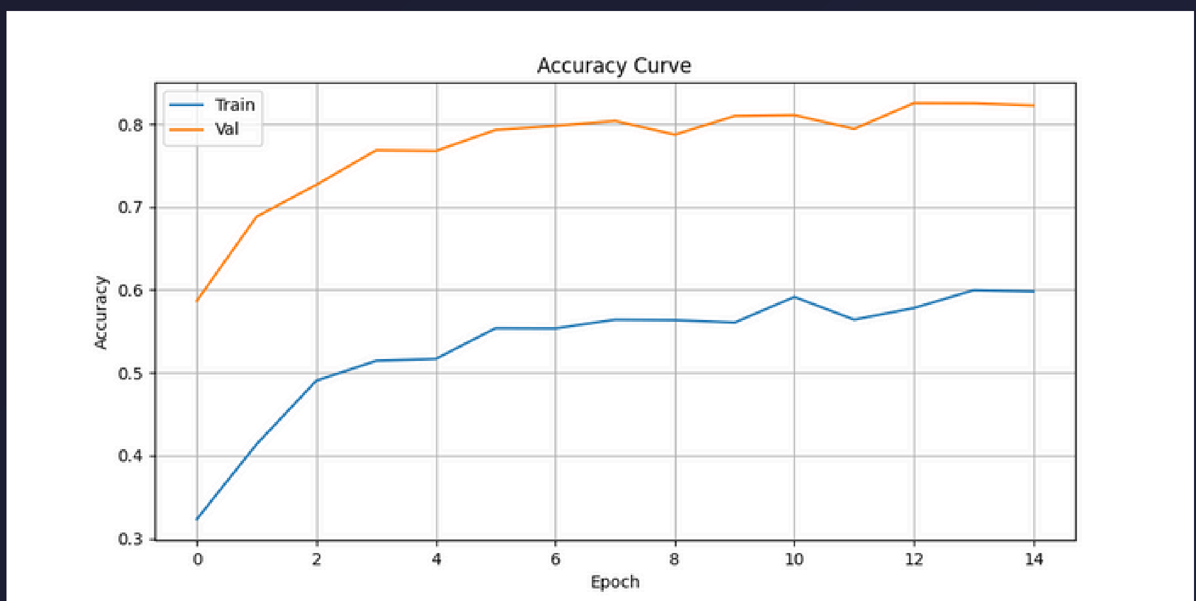
Emotion Recognition CNN



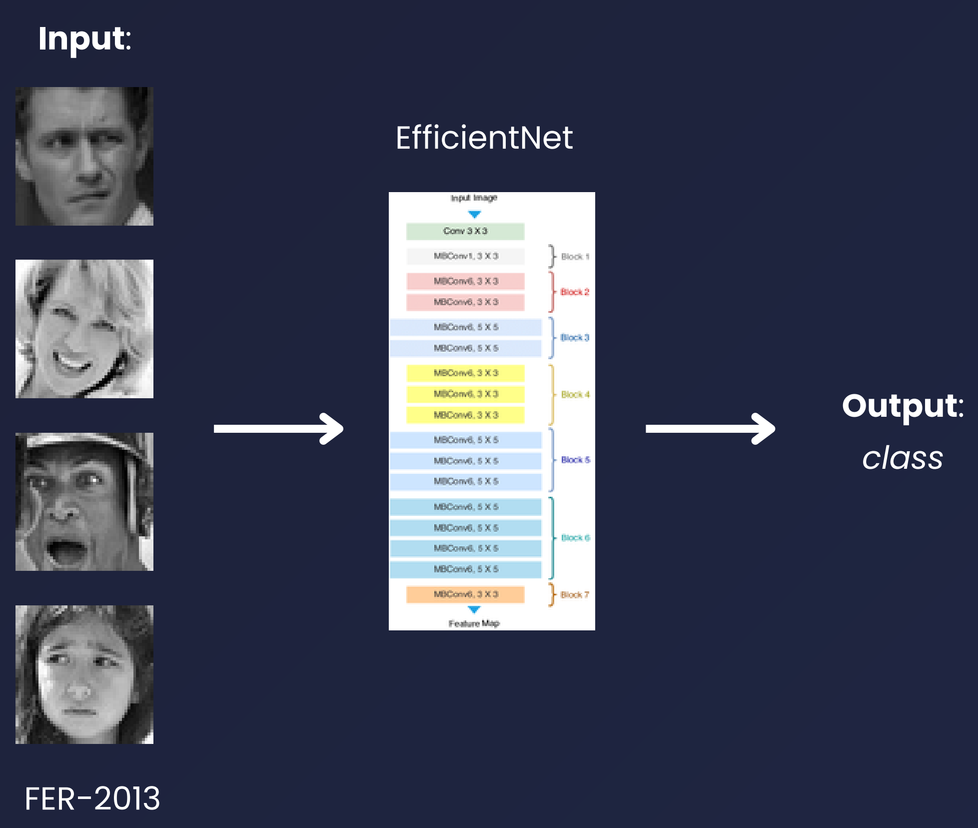
Gap created by mixup forcing the network to learn more robust features.

Results

EfficientNet-B0 with fine-tuning significantly outperformed earlier ResNet trials, both in accuracy, validation loss and speed.

Project Objectives

-> Build a CNN model to classify emotions from images using EfficientNet-B0
-> Overcome FER-2013 dataset limitations (small 48×48 grayscale images, class imbalance) using data augmentation, unfreezing pretrained layers, and optimizing training
-> Achieve >80% test accuracy via architecture tuning and evaluation

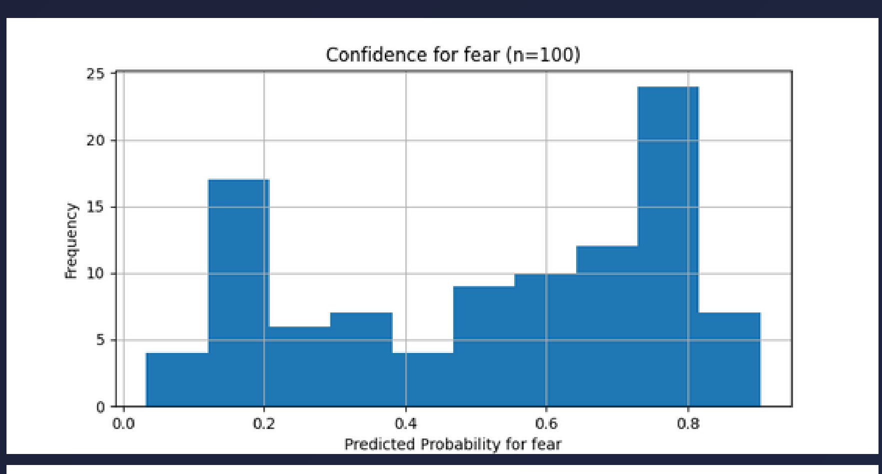


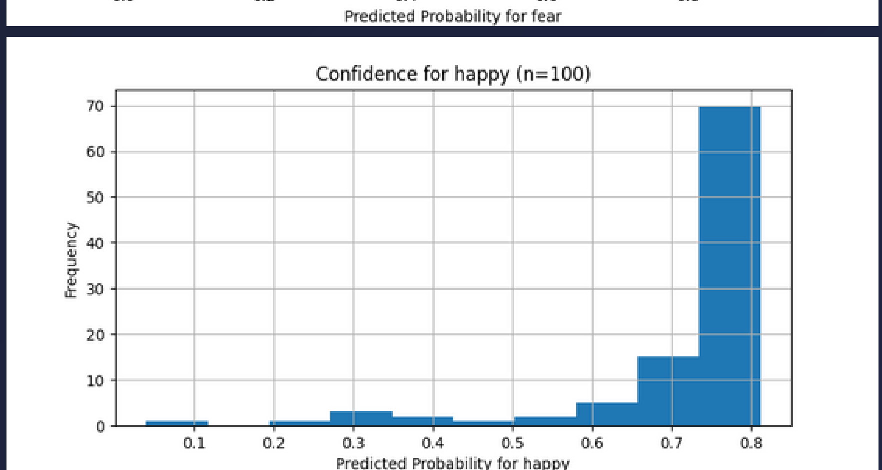
Training Strategy

- Gradual unfreezing: progressively training more layers of the pretrained backbone to leverage the learned features
- Data augmentation: applied mixup augmentation and standard flips/rotations to add to the training set
- Regularization: used label smoothing, dropout, and weight decay to prevent overfitting
- Class imbalance handling: applied class weights in the loss to compensate for underrepresented classes
- Batch size: reduced from 64 to 32 to improve gradient stability

Challenges Faced

- Early Overfitting: frozen backbone led to high training accuracy but poor validation performance
- NaN Loss: instabilities that came from large class weights for rare classes
- Class Imbalance: classes like ‘disgust’ and ‘fear’ are severely underrepresented (500/600 samples vs ~5000 for others), which limits recall
- Narrow Augmentation: initial augmentations were insufficient to cover data variability
- Fixes: unfreezing additional layers gradually; enhancing augmentations; applying sanity checks on class weights; using different learning rates and regularization





Conclusion

- Fine-tuned EfficientNet-B0 achieved robust emotion recognition, surpassing previous efforts while using ResNet
- Regularization techniques (dropout, label smoothing, mixup) stabilized training and improved generalization
- Interpretability analyses revealed model biases: frequent classes like ‘happy’ achieved higher confidence, while rare classes like ‘fear’ remained challenging.
- This shows the importance of architecture choice and data handling for this task

Val Loss: 0.9295 Acc: 0.8223

	precision	recall	f1-score	support
0	0.6935	0.7818	0.7350	958
1	0.6074	0.7387	0.6667	111
2	0.7652	0.7129	0.7381	1024
3	0.9535	0.9126	0.9326	1774
accuracy			0.8223	3867
macro avg	0.7549	0.7865	0.7681	3867
weighted avg	0.8293	0.8223	0.8245	3867

Future Work

-> integrating AffectNet dataset for broader generalization
-> improving performance with ensemble methods
-> integrating with an object detection system for contextual emotion analysis