

VISHAL CHAUDHARY · DATA ANALYST · DUBLIN, IRELAND

Deep Learning &amp; Computer Vision · Academic Project

# Multimodal Emotion Recognition

ResNet-50 and BERT fusion architecture for 7-class emotion recognition achieving 87% accuracy at 30 FPS real-time inference.

<b>87%</b>	<b>30 FPS</b>	<b>+19pp</b>	<b>7</b>
7-Class Accuracy	Real-Time Speed	vs Unimodal Baseline	Emotion Classes

## TOOLS & TECHNOLOGIES

Python	TensorFlow	Keras	ResNet-50	BERT	OpenCV
SMOTE	Scikit-learn				

Email <a href="mailto:vishal.ch1401@gmail.com">vishal.ch1401@gmail.com</a>	LinkedIn <a href="https://www.linkedin.com/in/vishal111">linkedin.com/in/vishal111</a>	GitHub <a href="https://github.com/chaudhary521">github.com/chaudhary521</a>	Location Dublin, Ireland
--	---	---	--------------------------

## PROBLEM STATEMENT

---

Single-modality emotion recognition systems fail in ambiguous cases where facial expressions and text context send contradictory signals. A person saying "I'm fine" with a distressed facial expression should trigger a different prediction than the same words spoken with a genuinely neutral face. Fusing visual and textual signals intelligently — and knowing which to trust when they conflict — was the core research problem.

## DATASET

---

Combined dataset of facial images and labelled social media posts covering seven emotion categories: joy, sadness, anger, fear, surprise, disgust, and neutral. The dataset contained significant class imbalance — joy and neutral were substantially over-represented while disgust and fear were minority classes. Both image and text modalities were required for every training sample to enable true multimodal fusion training.

## APPROACH & METHODOLOGY

---

Designed a multimodal fusion architecture combining ResNet-50 (pretrained on ImageNet) as the visual encoder for facial expression features, and a BERT-base transformer as the text encoder for sentiment and contextual meaning. An attention mechanism was trained to dynamically weight each modality's confidence contribution per sample — when a face is partially occluded or expression is ambiguous, the text signal gets up-weighted automatically. SMOTE oversampling and synthetic image augmentation addressed the disgust and fear class imbalance. A web-based demo interface was built for non-technical stakeholder testing.

## KEY TECHNICAL HIGHLIGHTS

---

- › ResNet-50 extracts facial feature vectors; BERT encodes text representations — both fed to a cross-modal attention layer.
- › Attention mechanism dynamically up-weights text when facial confidence is low, and vice versa — adaptive fusion.
- › Achieved 87% accuracy on 7-class emotion classification versus 68% (image-only) and 72% (text-only) baselines.
- › SMOTE oversampling and synthetic image augmentation corrected class imbalance for disgust and fear.
- › Real-time inference pipeline built with OpenCV processes live video at 30 FPS.
- › Web demo allows users to upload a photo with a caption and receive per-emotion confidence scores for each modality.

## KEY INSIGHTS & RESULTS

---

Multimodal fusion achieved 87% accuracy — a 19 percentage point improvement over the best unimodal baseline. The attention mechanism successfully corrected ambiguous facial expression predictions using text context in 34% of test cases. Real-time 30 FPS inference was achieved without model quantisation, suggesting headroom for deployment on standard inference hardware.

## BUSINESS IMPACT

---

Demonstrates a generalisable architecture for multimodal classification wherever visual and textual signals co-occur. Immediate application domains include mental health monitoring systems, customer experience analysis in retail and call centre contexts, and human-computer interaction for accessibility tools. The attention-weighting mechanism is particularly valuable for real-world deployment where one modality may be unreliable due to environmental conditions.

---

This case study is part of Vishal Chaudhary's data analytics portfolio. For more projects and contact details visit: [github.com/chaudhary521](https://github.com/chaudhary521)