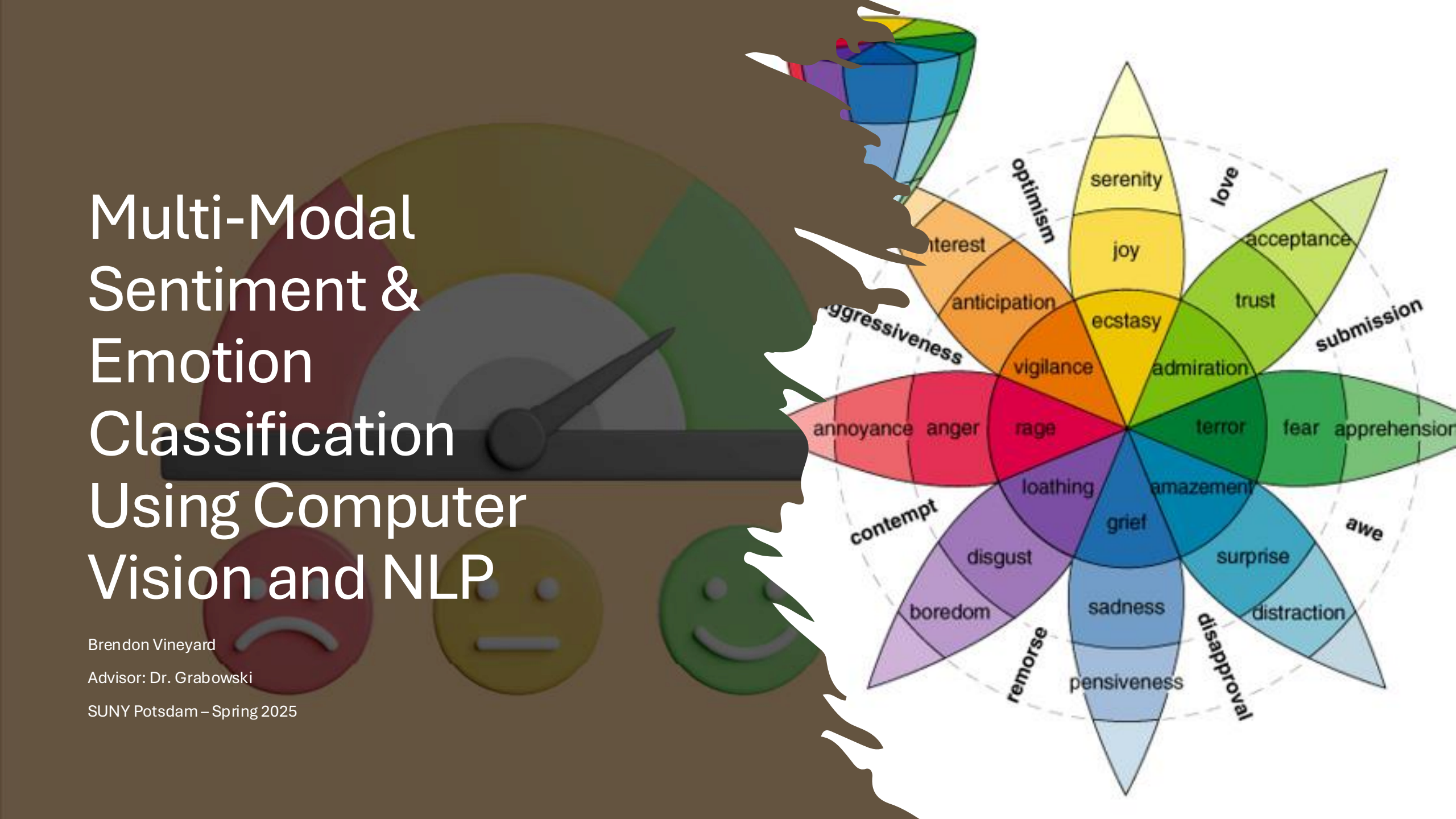
Multi-Modal Sentiment & Emotion Classification Using Computer Vision and NLP

Brendon Vineyard

Advisor: Dr. Grabowski

SUNY Potsdam – Spring 2025

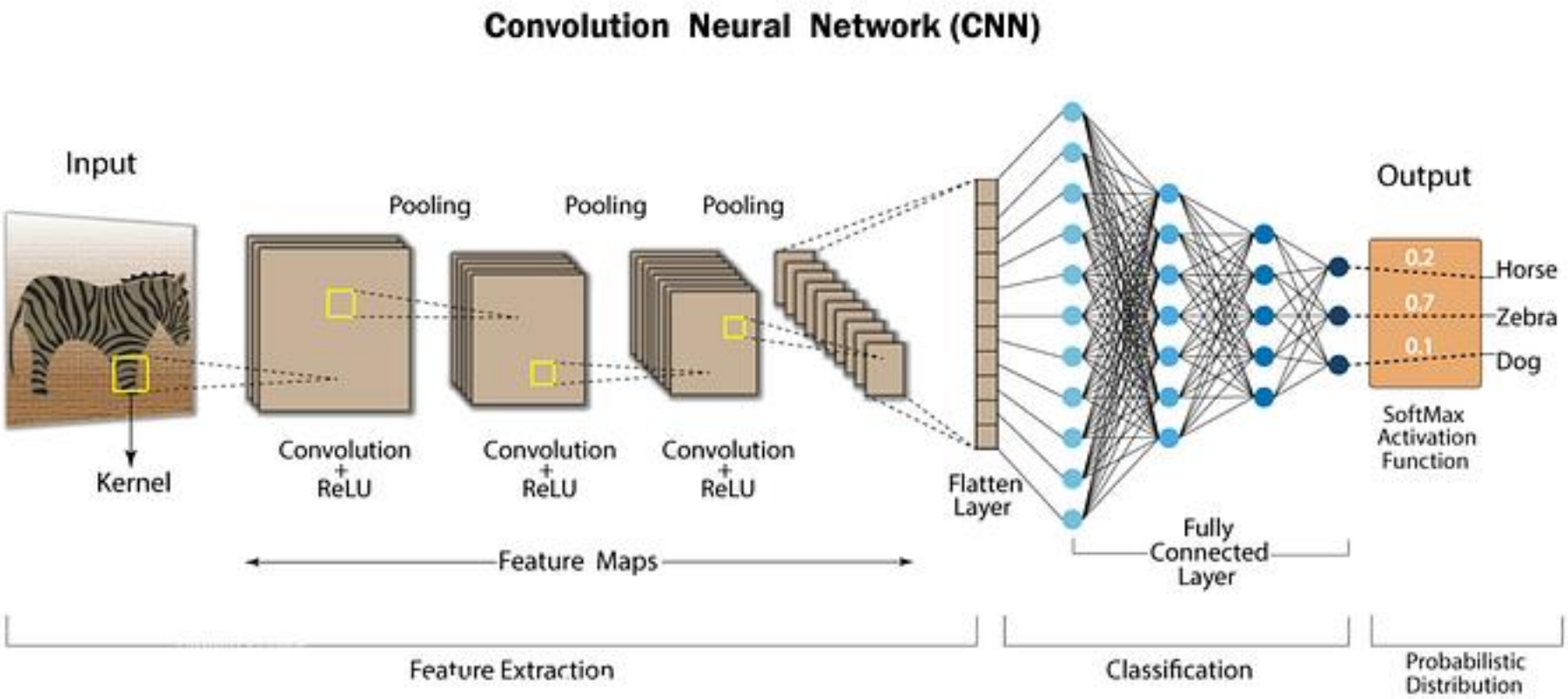


Problem Motivation

- Social media mixes text/images
- Models analyze separately, miss content
- Goal: fuse text and image analysis

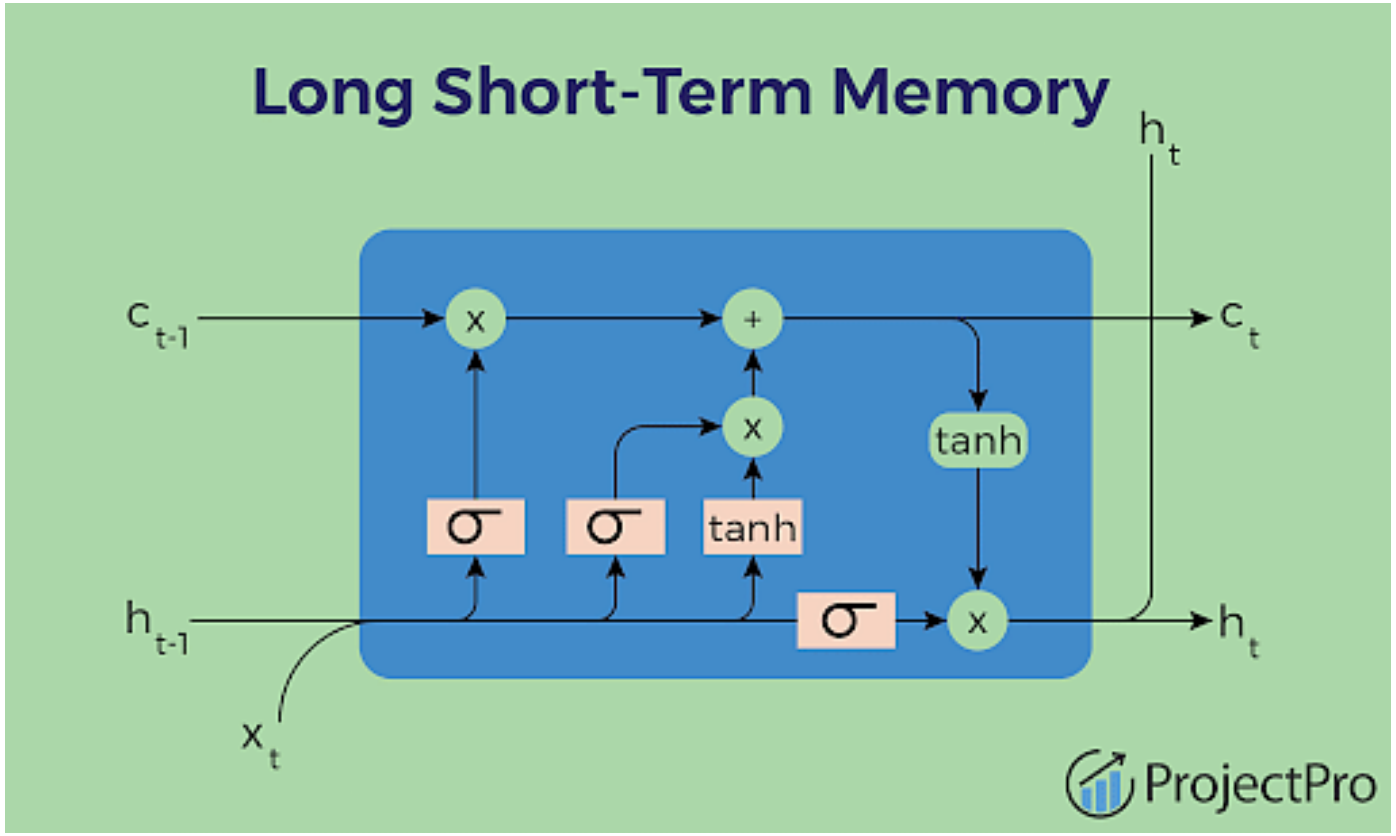
Model Architecture Overview

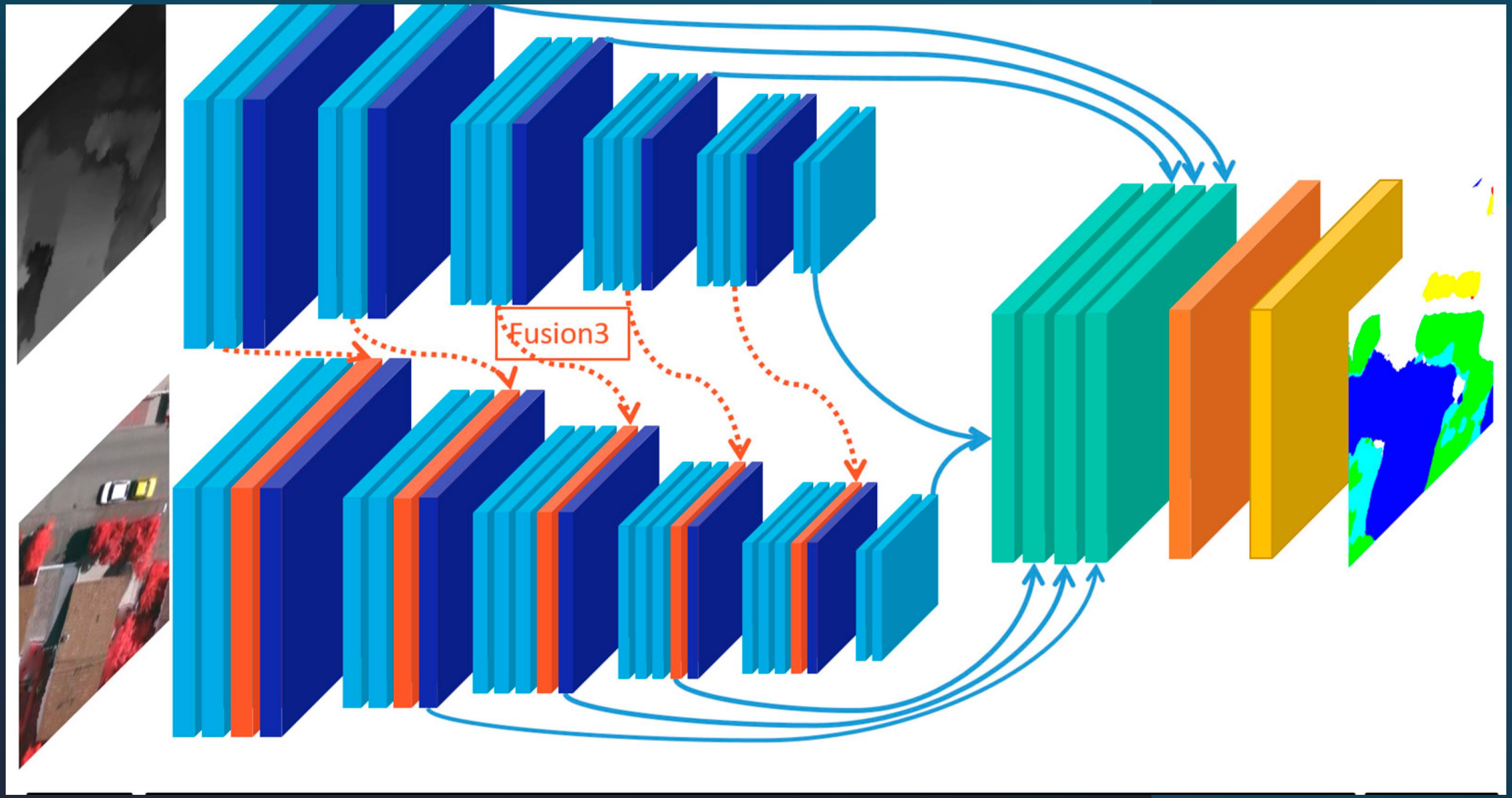
- CNNs trained on 3 separate image datasets for emotion.



Model Architecture Overview

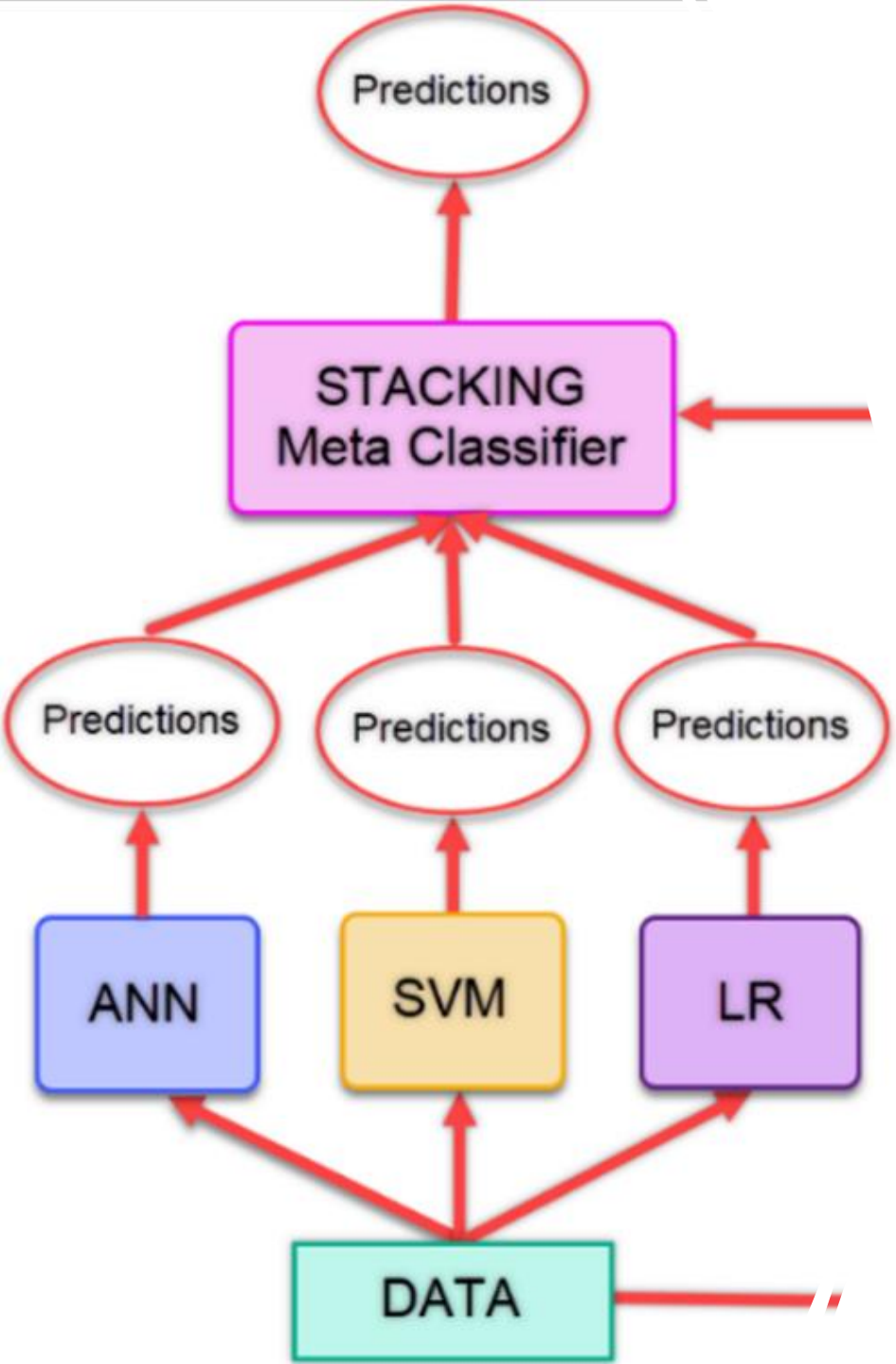
- Text is processed using LSTM for sentiment prediction.




Fusion3

Datasets Used

- FER-2013: Grayscale emotion-labeled faces.
- RAF-DB: Refined facial emotion dataset.
- FER+: Enhanced labeling for FER.
- Seniment140: 1.6 million tweets labeled with sentiment.

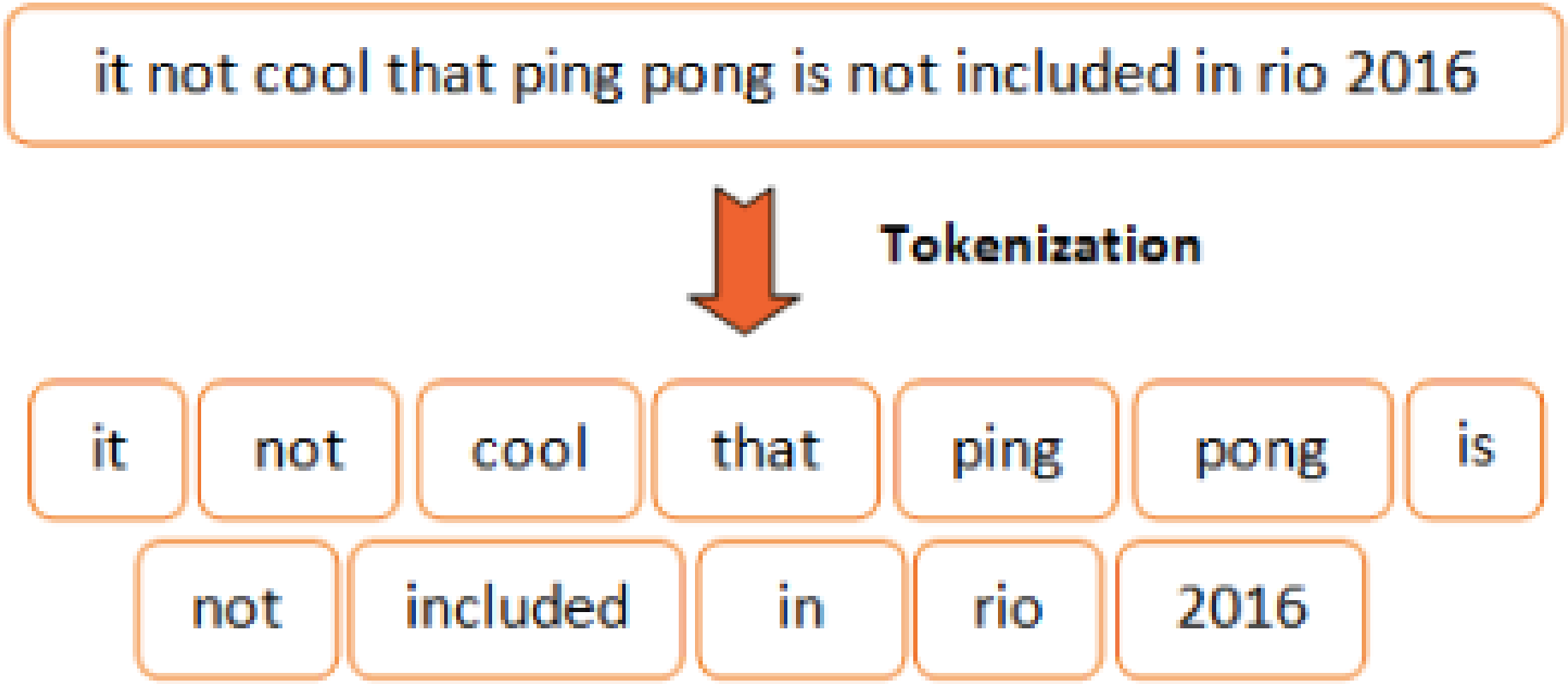


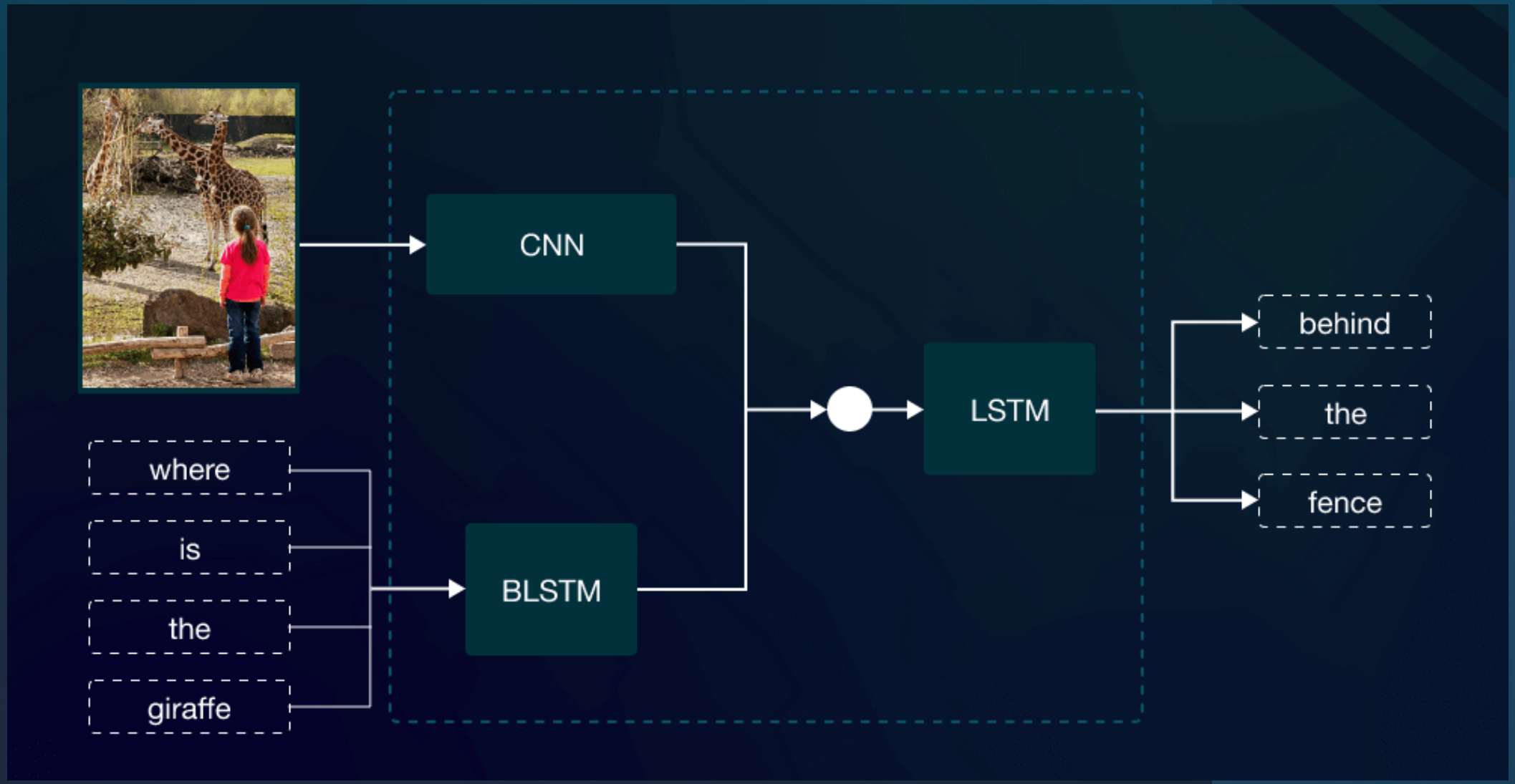
Fusion Strategy

- Images processed by 3 CNNs
- Softmax outputs stacked, classified

Sentiment

- Text tokenized, passed to BiLSTM



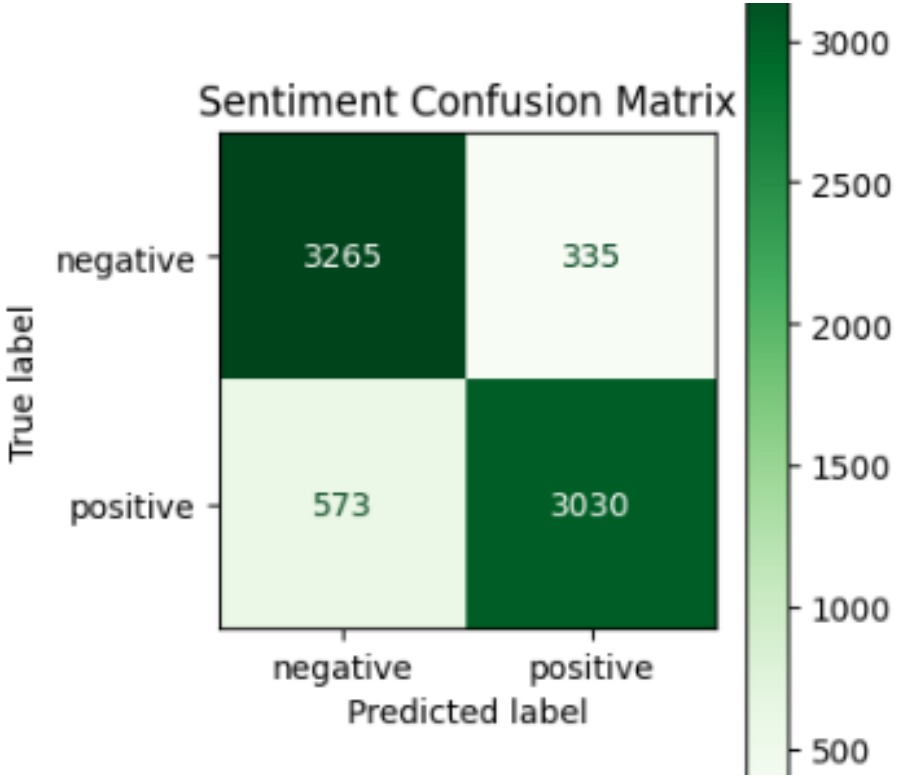

CNN
LSTM
BLSTM
where
is
the
giraffe
behind
the
fence

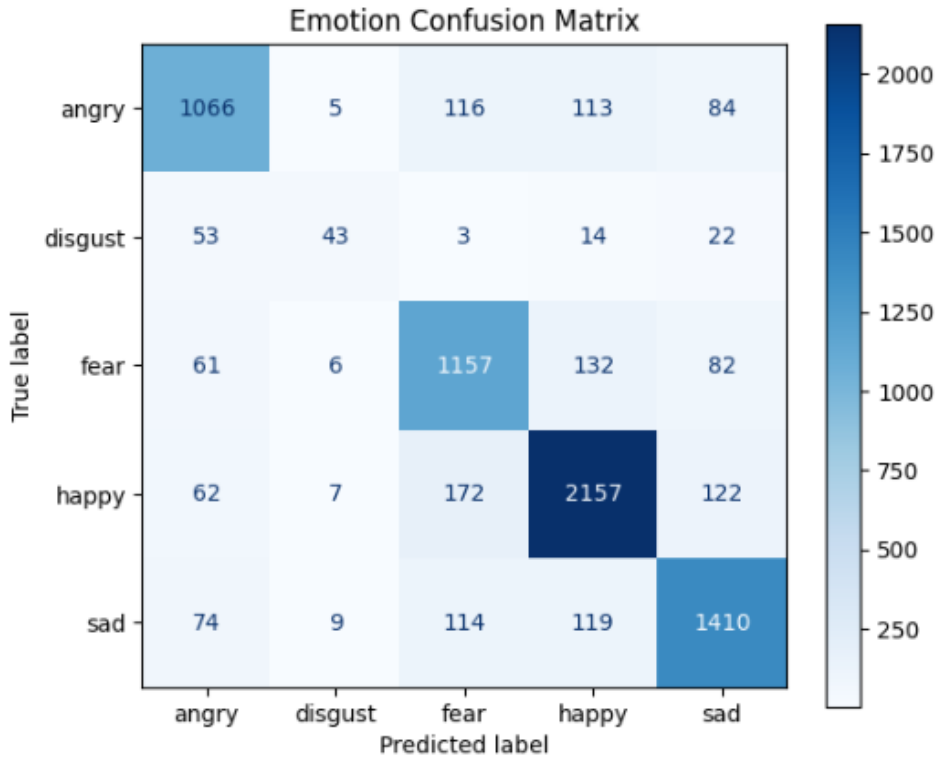
Model Improvements

- Replaced meta-classifier with MLP
- Added Dropout and BatchNorm
- Filtered inconsistent emotion classes
- Aligned to 5-class emotion labels
- Used pretraining and freezing options

Final Performance

- Emotion Accuracy: up to 95.00%
- Sentiment Accuracy: 82%-95% range
- Achieved with joint model combining FER, RAF, FER+, and Sentiment140.





 Emotion: ANGRY (True: ANGRY)
 Sentiment: NEGATIVE (True: NEGATIVE)

 Caption: oooh thats pretty unique id show you my teddy bear but i dont have an iphone

 Emotion: SAD (True: SAD)
 Sentiment: POSITIVE (True: POSITIVE)

 Caption: cara oooh followers thanks guys xoxo

 Emotion: FEAR (True: FEAR)
 Sentiment: POSITIVE (True: POSITIVE)

 Caption: its ok sometimes life gets in the way of living life dont feel bad and i hope that things are ok now

 Emotion: HAPPY (True: HAPPY)
 Sentiment: NEGATIVE (True: NEGATIVE)

 Caption: wow with no source for organic bean or broccoli sprouts i think that donut needs to be chocolate covered

Testing and Gradio App

[insert link]

Challenges Faced

- Emotion label mismatches across datasets
- Balancing sentiment and emotion learning
- Batch size vs GPU limits
- Overfitting on smaller datasets

Key Takeaways

- Fusion improves emotion understanding
- Pretraining and ensembles boost accuracy
- Small tweaks improve performance
- Applications: sentiment, health, marketing

Capstone Project CIS 480 – Spring 2025

Author: Brendon Vineyard

Email: vineyabn207@potsdam.edu