Emotion Recognition Body and Dance

Kriti and Shobha

BACKGROUND

Why recognize emotion from body movement?

- People express emotion not only through face or voice body posture, gestures, rhythm, and motion also carry rich signals.
- Existing systems mostly focus on facial or speech cues, which can fail when: faces are hidden, masked, or poorly lit, speech is unavailable or unreliable, privacy-sensitive settings.
- Our goal is to build a reliable system that identifies emotion purely from human motion, enabling more natural, safer, and context-aware interaction.

PROJECT OVERVIEW

KINEMATIC ACTORS Dataset

- Large, diverse dataset: 1,402 motion sequences, each labeled with one of seven emotions.
- Recorded using 17 body sensors, giving precise joint positions over time.
- Each recording stored in BVH format, containing:
- Skeleton structure (joints and hierarchy)
- Time-varying motion (positions + rotations)

EMNOKINE DATASET

EMOKINE was primarily developed for:

- Human emotion perception studies
- Analysis of expressiveness in dance movements
- Early computational modeling of emotion from body kinematics

It is one of the earliest datasets focusing entirely on movement-based emotion expression.

APPROACH

ML

- Input: 400 to 500 engineered features per sample
- Models: Random Forest, SVM, Gradient Boosting
- Strength: simple, interpretable
- Limitation: cannot capture time-based motion patterns

MLP

- Selected top 150 predictive features using mutual information
- Used Optuna to find best neural network architecture
- Added dropout & regularization to prevent overfitting
- Improved performance over classical ML

LSTM

- Added motion features: velocity, acceleration
- Added data augmentation (noise, Mixup)
- Used feature attention + temporal attention to highlight expressive frames
- Result: more robust understanding of movement dynamics

STGCN

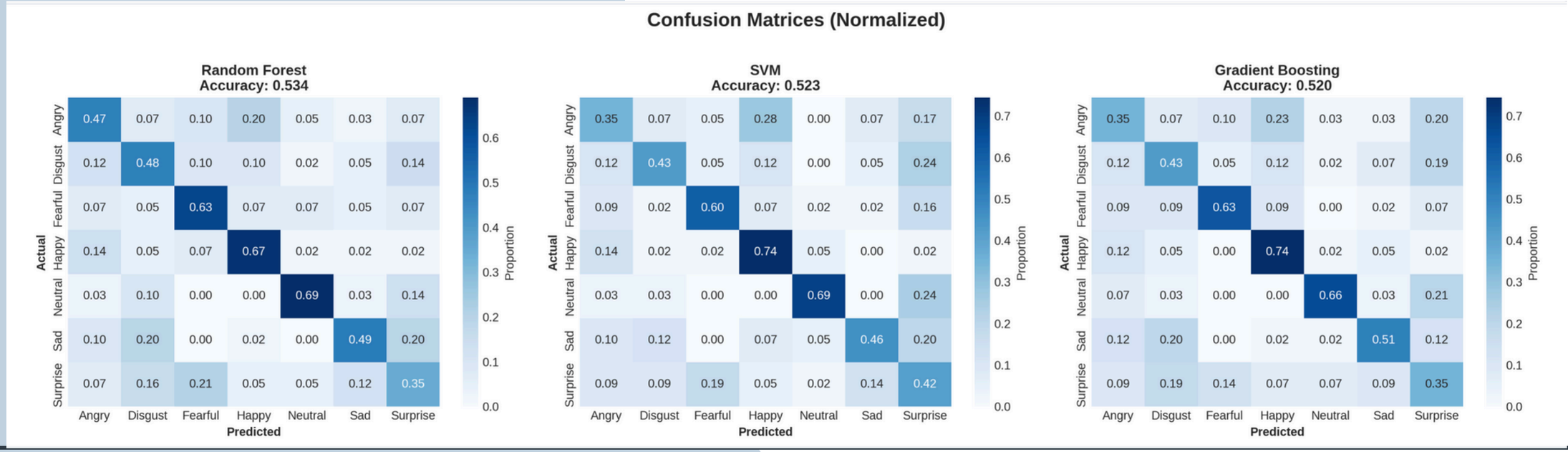
- The human body is a graph: joints (nodes) connected by bones (edges).
- Spatial patterns (which joints move together)
- Temporal patterns (how movement changes over time)
- Produces a highly expressive representation of emotional movement

ML RESULTS

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.534	0.534	0.534	0.532
SVM	0.523	0.540	0.523	0.525
Gradient Boosting	0.520	0.527	0.520	0.520

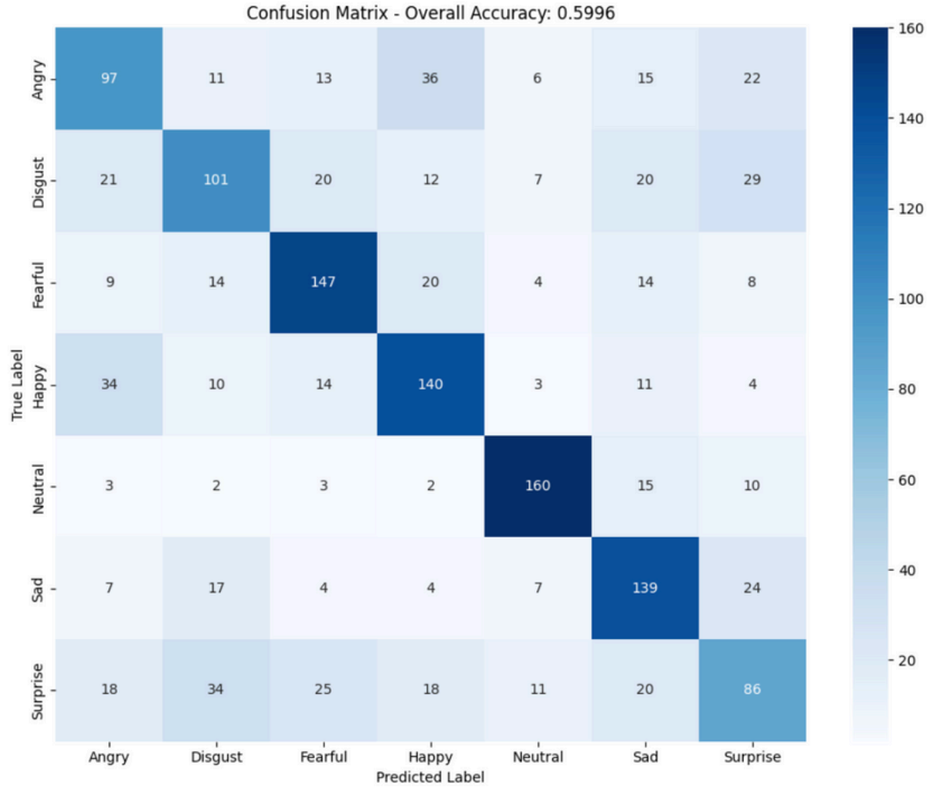
Classical ML Models (SVM, RF, GBM)

- Accuracy: ~52–53%
- Performed similarly
- Random Forest performs slightly better on Sad (0.49), Disgust (0.48) and Fearful (0.63).
- SVM shows best performance for Happy (0.74) and Surprise(0.42).
- Gradient Boosting performs well across classes except surprise and angry on comparison.



MLP RESULTS

```
==================================================
CROSS-VALIDATION RESULTS
==================================================
Mean Accuracy: 0.5996 ± 0.0121
Fold Accuracies: ['0.5773', '0.6103', '0.6103', '0.6000', '0.6000']
```

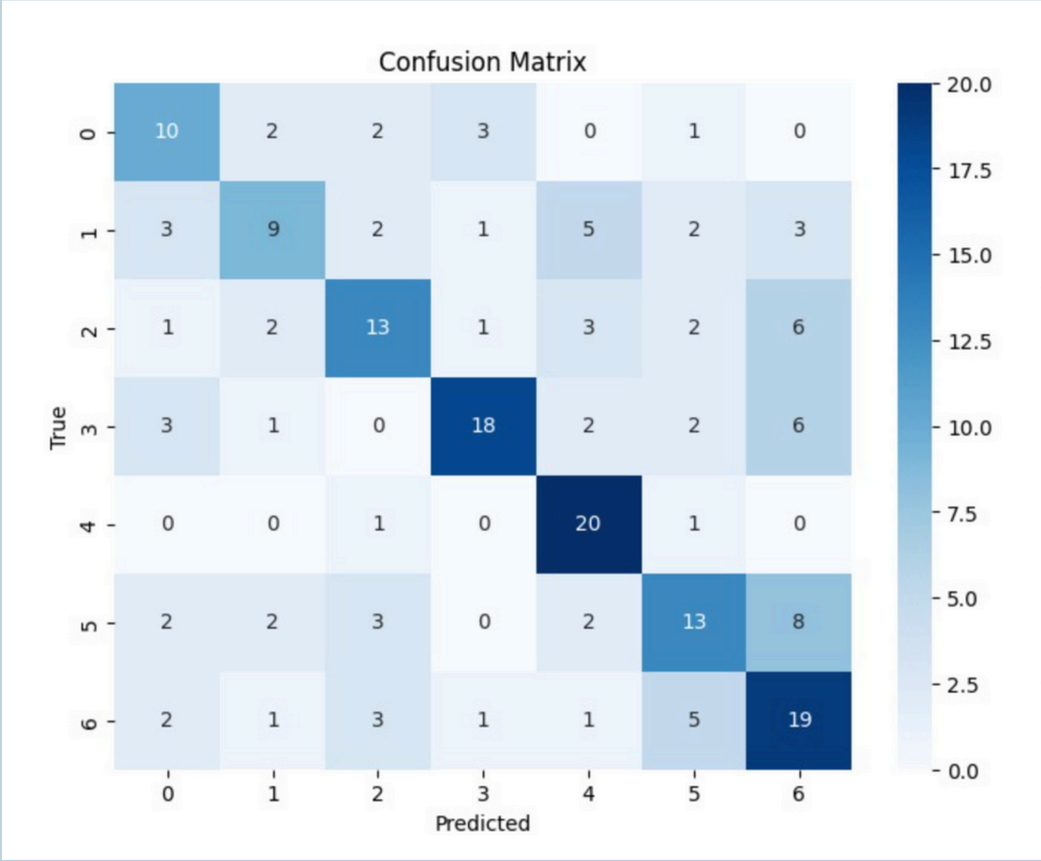



```
Classification Report:
              precision    recall  f1-score   support

       Angry       0.51      0.48      0.50       200
     Disgust       0.53      0.48      0.51       210
     Fearful       0.65      0.68      0.67       216
       Happy       0.60      0.65      0.62       216
     Neutral       0.81      0.82      0.81       195
         Sad       0.59      0.69      0.64       202
    Surprise       0.47      0.41      0.44       212

    accuracy                           0.60      1451
   macro avg       0.60      0.60      0.60      1451
weighted avg       0.59      0.60      0.60      1451
```

MLP Model

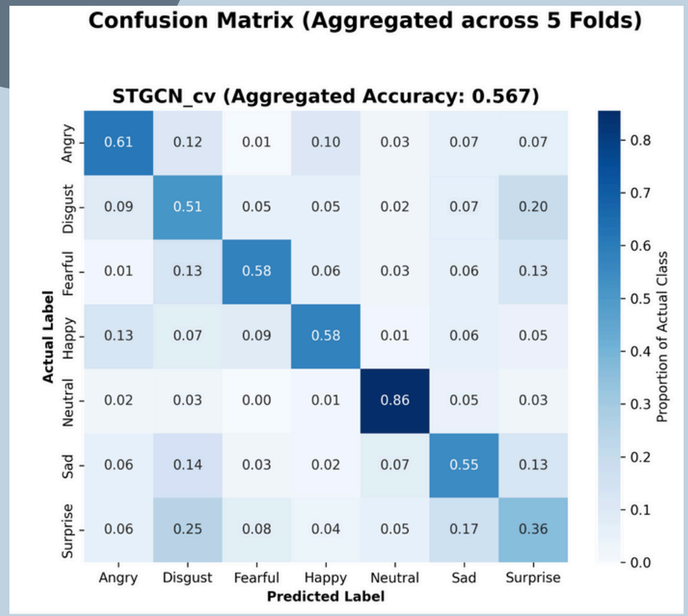
- Accuracy: ~60%
- The MLP shows stable and consistent performance across folds, with very small variance showing that the model generalizes reasonably well.
- The classes that show strong performance in prediction are Neutral (160 correct), Fearful (147 correct), Happy (140 correct), Sad (139 correct). Angry and Surprise still show lowest performance.
- MLP shows highest performance for Neutral emotion (0.80-0.82) across F1, Precision and Recall. Apart from Neutral emotion, MLP shows decent results (0.60-0.62) for Happy, Fearful and Sad classes. Lowest performance is visible among the classes such as Angry, Disgust and Surprise.

LSTM RESULTS



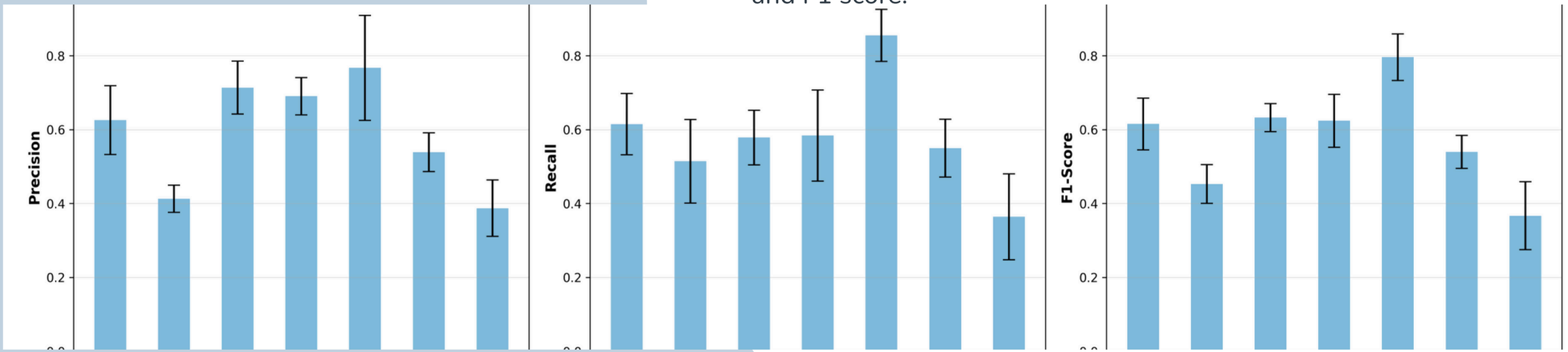
- Final accuracy: ~59% after tuning
- The diagonal values (top-left to bottom-right) represent correct predictions.
- Neutral has the strongest diagonal value (20 correct), showing it is recognized very reliably.
- Happy also has strong correct predictions (19 correct).
- Angry, Disgust and Fearful and Sad have moderate correctness.
- Surprise shows more confusion, meaning the model struggles with this emotion.

0 = Angry
1 = Disgust
2 = Fearful
3 = Happy
4 = Neutral
5 = Sad
6 = Surprise

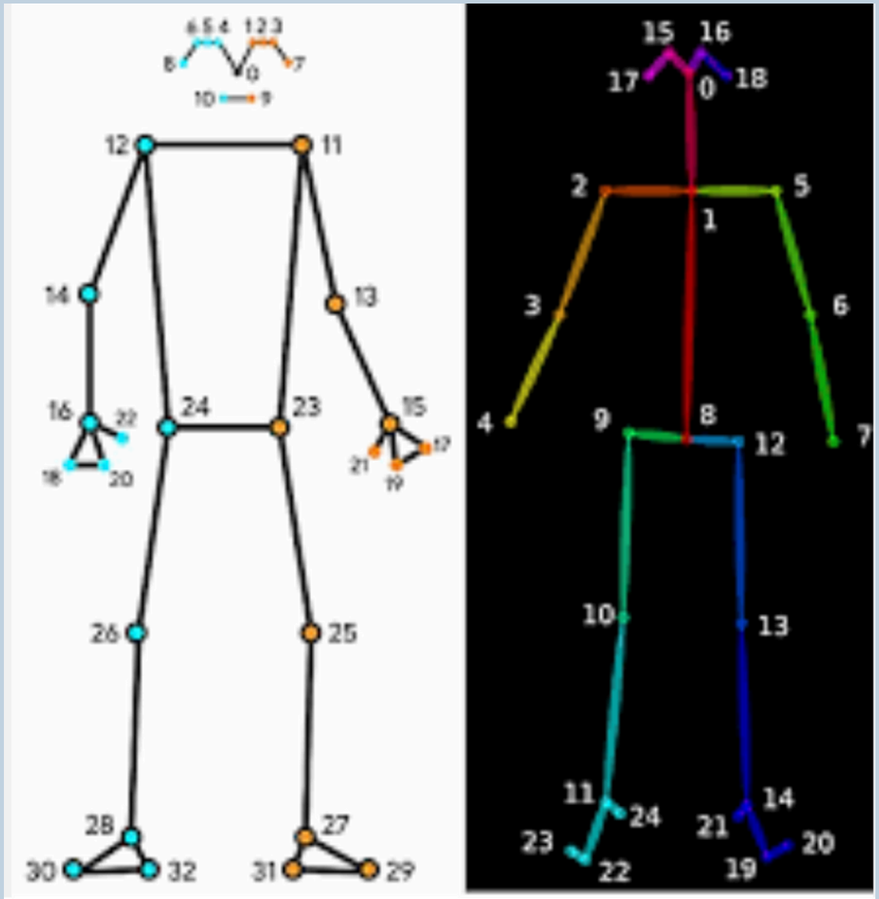


ST-GCN MODEL

- Accuracy: ~56–57%
- Best at capturing spatial movement patterns
- High precision/recall for Neutral, Fearful, Happy
- Neutral has the highest recall, precision and F1 score. This means that neutral movements are most stable and consistent across participants.
- Fearful, Happy, Angry have decent precision (0.60 to 0.70) and F1 score (0.62 approx)
- Disgust and Surprise have shown least performance across recall, precision and F1 score.



USE CASES AND IMPACT



Who benefits?

- Healthcare: track emotional wellbeing, detect distress
- Human–Robot Interaction: robots adapt to users' emotions
- VR/AR: systems adjust difficulty or content based on body cues
- Performance Arts: analyze expressive movement in dance & acting
- Elderly Care / Assistive Tech: detect discomfort or emergencies
- Retail & Customer Interaction: understand body-based engagement

Impact:

- Works even when face/voice are unavailable
- Enables more natural, empathetic, and accessible systems
- Provides richer insights from subtle body-language cues

Thank You