Comparative Analysis of Pain Recognition Models: Bio-Visual Fusion vs. Graph Neural Networks

Prashant Bahuguna

September 2024

GitHub Repository

1 Introduction

Pain intensity recognition is a crucial component in both clinical and research settings, enabling the assessment of patient discomfort and the efficacy of pain-relief interventions. This study compares two advanced methodologies for person-independent pain intensity recognition: Bio-Visual Fusion and Deep Graph Neural Networks (GNN). The Bio-Visual Fusion approach integrates visual and physiological data (e.g., electromyography (EMG), electrocardiogram (ECG), and skin conductance (SCL)) to predict pain intensity. The GNN approach, on the other hand, utilizes facial landmarks from video data to model pain expressions. These two methods are evaluated and compared on their accuracy and effectiveness in recognizing pain across diverse participants. The goal is to establish a robust, automated system for pain intensity recognition that can be generalized across individuals.

2 Dataset

The dataset used in this study included data from 87 participants subjected to controlled pain stimuli using a thermode device. The pain intensity was calibrated individually for each participant, dividing the range between two reference levels—where pain begins and where it is barely bearable—into three equally spaced intervals. The intensity levels were labeled as BL-1 (Baseline), PA-1, PA-2, PA-3, and PA-4, representing progressively higher pain intensities. Four different stimulation strengths were applied 20 times, resulting in a total of 80 responses per participant.

During the experiments, high-resolution video was captured from three different cameras, alongside sensor data from a Kinect and a biophysiological amplifier. The physiological data included electromyography (EMG) signals from the zygomaticus, corrugator, and trapezius muscles, skin conductance level (SCL), and electrocardiogram (ECG). This multi-modal setup provided comprehensive biopotential and video data, allowing for the assessment of pain intensity from both visual and physiological cues.

3 Methodology

The methodology implemented in this study closely follows the approach outlined in the paper "Bio-Visual Fusion for Person-Independent Recognition of Pain Intensity." The primary focus was on exploring the applicability of early and decision fusion approaches for recognizing pain intensity in a person-independent manner.

3.1 Fusion Approaches

3.1.1 Early Fusion

In this approach, features from multiple modalities (e.g., video and biopotential signals) were directly combined (concatenated) before training the classifier. This method leverages the combined feature space to train a single classifier that learns from both visual and physiological data simultaneously.

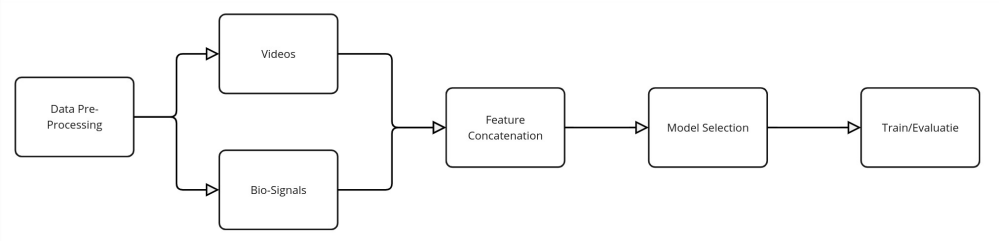


Figure 1: Early Fusion Approach

3.1.2 Late Fusion

This method involves training individual classifiers on each modality separately. The predictions from these classifiers are then combined in a separate fusion step to make a final decision. This can be done using fixed mappings, such as taking the mean or product of the individual predictions, or through trainable mappings like the pseudo-inverse method.

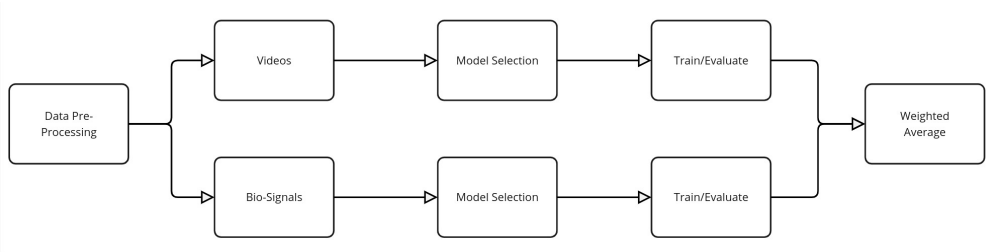


Figure 2: Late Fusion Approach

3.1.3 Hybrid Fusion

In addition to early and decision fusion, a hybrid approach was investigated, combining specific channels at the feature level followed by decision-level fusion with other channels. This combination aims to maximize the strengths of both early and late fusion by optimizing the integration of complementary information from different modalities.

3.2 Classification Techniques

The study employed classifiers such as Support Vector Machines (SVM) and Random Forests to evaluate the effectiveness of the different fusion approaches. Feature selection methods were also applied to reduce the dimensionality of the combined feature space, ensuring that only the most discriminative features were used in the final model.

4 Pre-Processing

The pre-processing for each of the dataset i.e video and biosensors was done seperately.

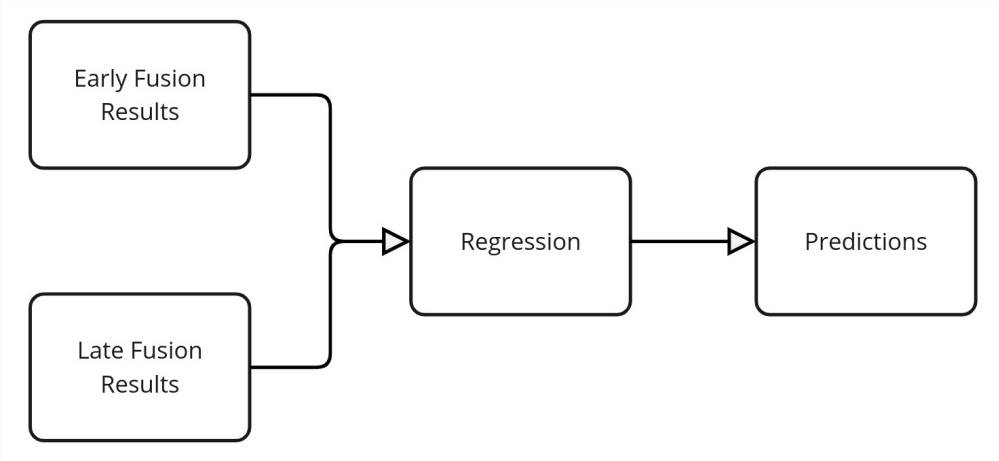


Figure 3: Hybrid Fusion Approach

4.1 Video Pre-Processing

The features from the videos were extracted using the dlib library, these features included key facial and head pose features from the dataset, including:

- **Head Pose Estimation**: Features extracted included fold intensity, eyebrow distance, eye closure, mouth height, yaw, pitch, roll, and translation (x, y, z).
- **Nasolabial Fold Intensity**: Analyzing the deepening of the nasolabial fold as a pain indicator.

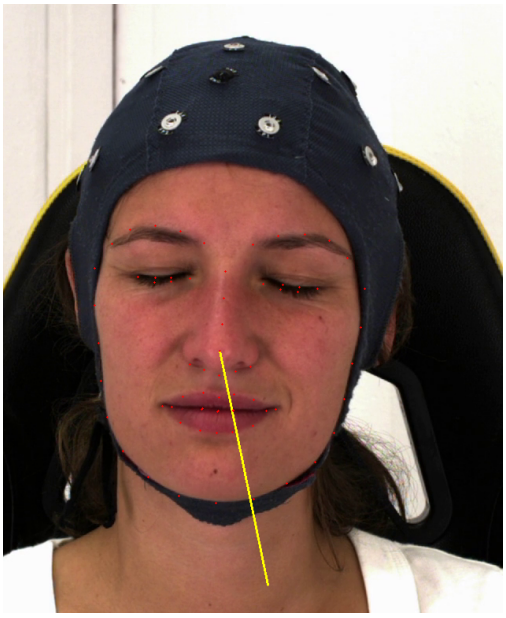
4.1.1 Feature Calculation

After extracting these features, the following steps were performed:

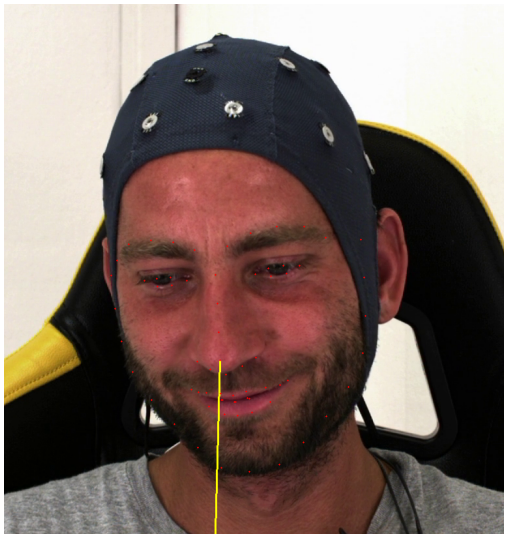
- **Low-Pass Filtering**: A low-pass filter was applied to smooth the signals.
- **Temporal Derivatives**: The first and second temporal derivatives of each smoothed signal were computed.
- **Statistical Parameters Extraction**: Statistical parameters were calculated from the low-pass filtered signal and its derivatives, including - Mean, Median , Standard Deviation, Range, Inter-quartile Range, Inter-decile Range and Median Absolute Deviation

4.1.2 Data Analysis

- **Scree Plot for PCA**: The plot highlights how much variance each component contributes.



(a) Face Sample 1



(b) Face Sample 2

Figure 4: Head Pose Estimation Sample

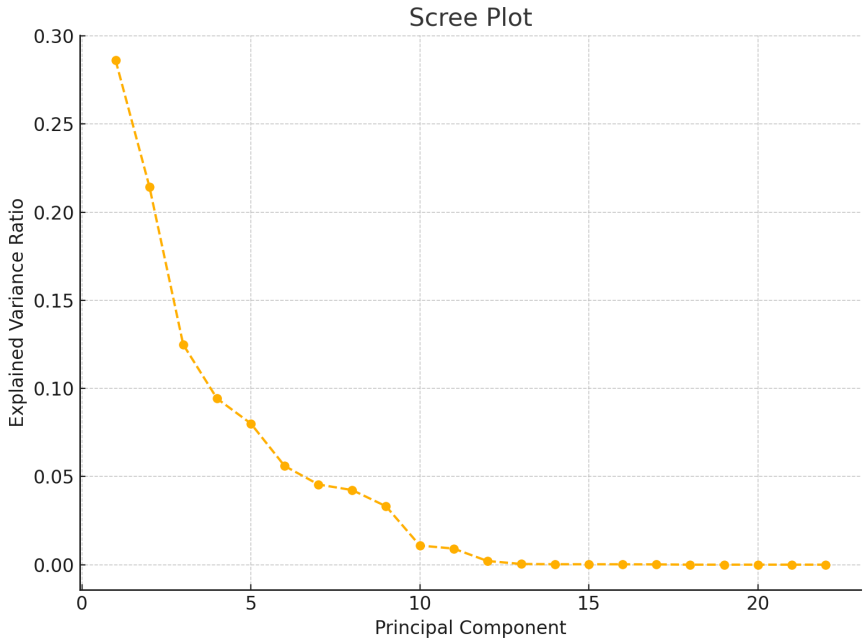


Figure 5: Scree Plot for PCA

- **Mean Distribution**: Shows the distribution of the average distance between the eyebrows, Mouth Height, Eye Closure, and Eyebrows.

Figure 6: Mean Distribution Plot

4.2 Bio-Signals Pre-processing

The biosignals, including electromyography (EMG) and electrocardiogram (ECG), underwent the following pre-processing steps:

- **Filtering**: A Butterworth bandpass filter was used to reduce noise and minimize the effects of trends in the signals.
- **Noise Reduction for EMG**: An additional noise reduction procedure based on Empirical Mode Decomposition (EMD) was applied to the EMG signals.
- **Feature Extraction**: Signal amplitude and frequency features such as peak height, peak difference, mean absolute difference, Fourier coefficients, and bandwidth were extracted. Additional features were derived based on entropy (approximate and sample entropy), stationarity, and statistical moments.
- **Windowing**: All features were computed on a window of 5.5 seconds, resulting in a total of 131 features extracted from the biosignals.

4.2.1 Plots and Visualizations

- **EMG Histograms**: Histograms for all the EMG-related features.

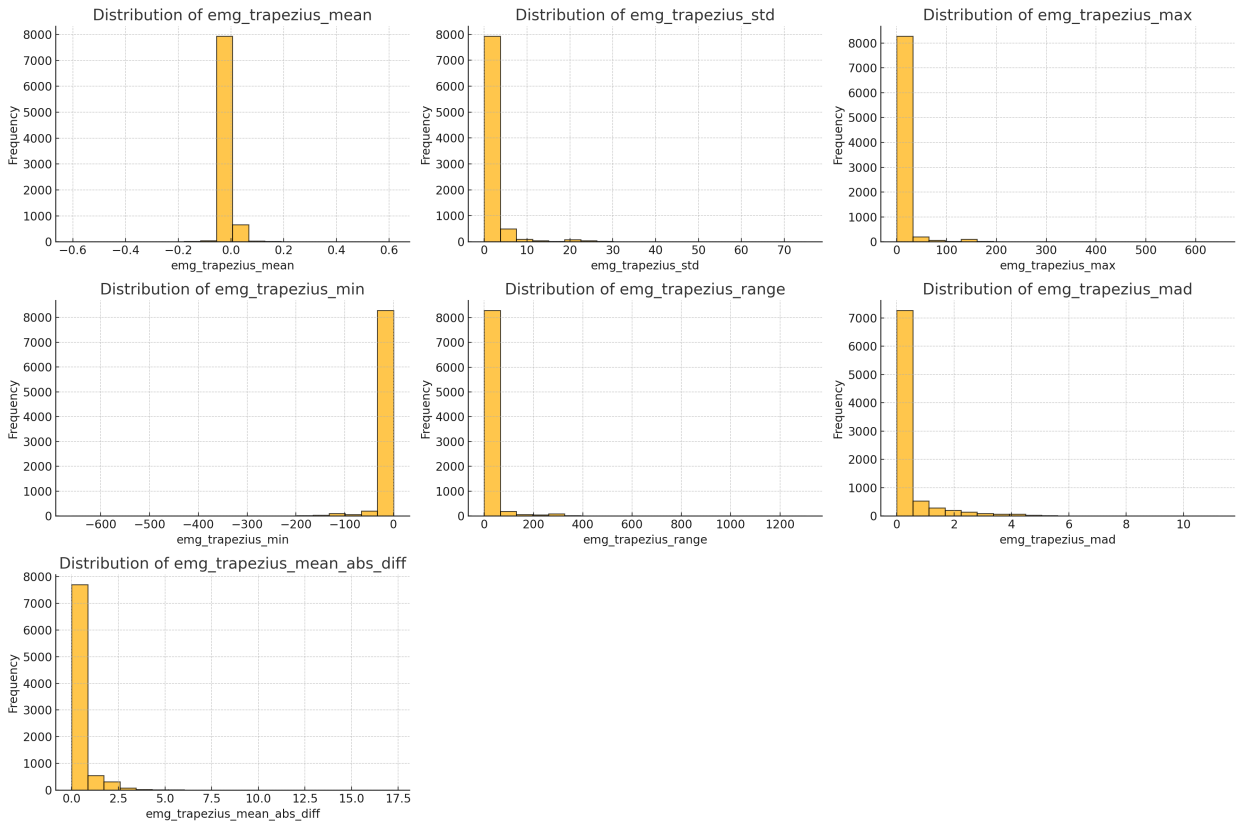


Figure 7: EMG Histograms

- **SCL Histograms**: Histograms for all the SCL-related features.

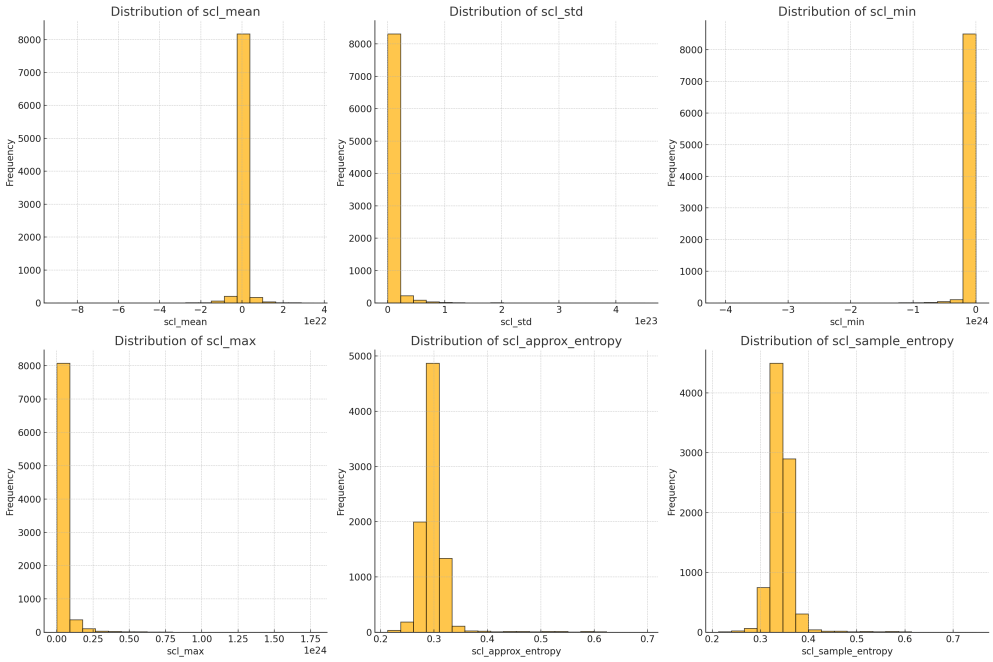


Figure 8: SCL Histograms

- **Features Correlation with Pain Intensity**: Shows the top 10 features that are most correlated with pain intensity based on the dataset. These features have been selected based on their correlation coefficients, providing insights into which biosensor readings are most closely related to pain levels.

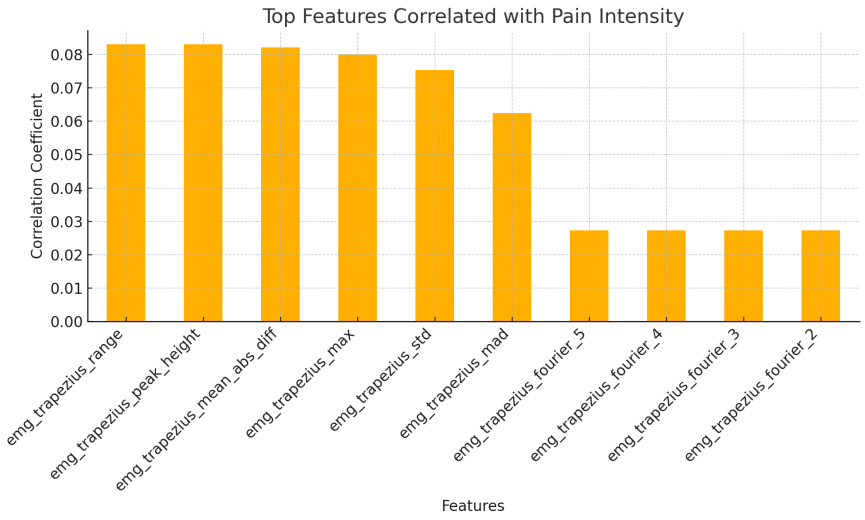


Figure 9: PCA for Biosensors

5 Hyperparameters

The following hyperparameters were used in the classification models:

Support Vector Machine (SVM):

- **Kernel:** Linear/RBF
- **Regularization (C):** Controls penalty for misclassification 1.3
- **Outputs:** Softmax applied to convert outputs into probabilities

Random Forest:

- **Number of Trees:** 300-500
- **Split Criterion:** Gini index (impurity-based splitting)
- **Depth:** 50-60

Pseudo-Inverse (for Decision Fusion):

- Least-squares optimal linear mapping applied using a holdout set

Feature selection was performed using Hybrid Sequential Floating Forward Selection (SFFS) to optimize the feature set before classification.

6 Metrics

6.1 My results vs "Bio-Visual Fusion for Person-Independent Recognition of Pain Intensity"

The results of my experiments are largely in line with the findings reported in the original paper, as demonstrated in the accuracy comparison of various model variants.These discrepancies can be attributed to the challenges in perfectly replicating all methods and the inherent variability in datasets, preprocessing steps, and experimental setups. Nonetheless, the results demonstrate comparable performance.

Method	**My Results**	**Bio-Visual-Fusion**
Early Fusion SVM	0.66	0.658
Early Fusion RF	0.67	0.789
Video	0.68	0.719
Late Fusion (Weighted Average)	0.66	0.766 (mean)

Table 1: Comparison of Results

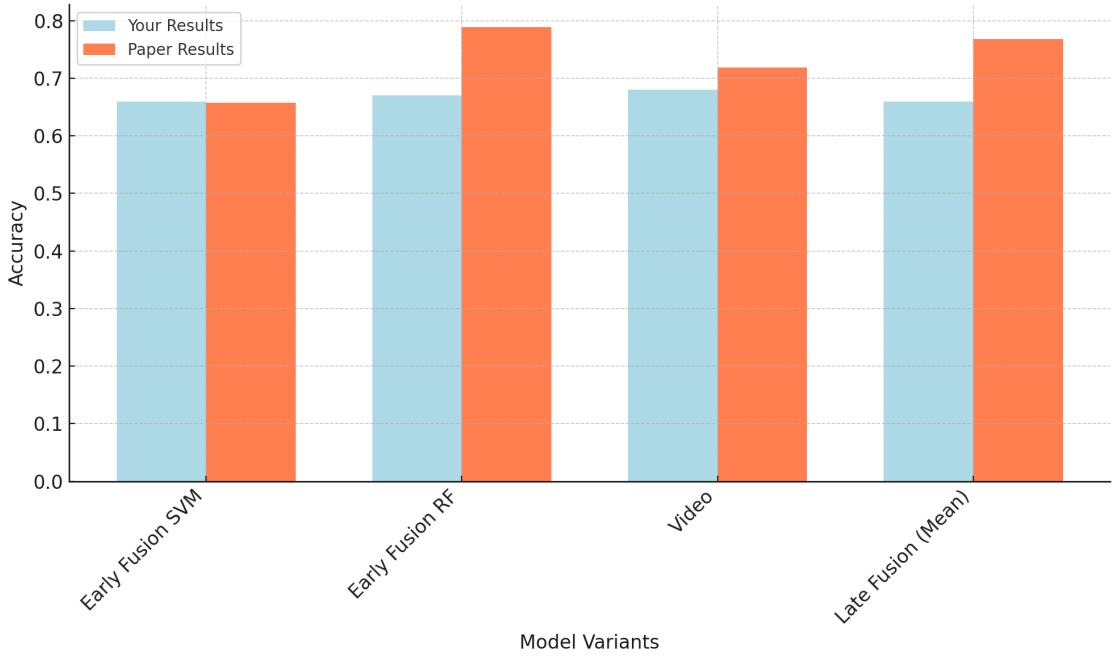


Figure 10: Comparison of model performance for Bio-Visual Fusion and GNN approaches.

6.2 My results vs. "Deep Graph Neural Network for Video-Based Facial Pain Expression Assessment"

Method	My Results	GNN-Paper
AUs + SVM	N/A	0.669
CM + SVM	N/A	0.714
CM + DGCNN (Video)	N/A	0.732
Early Fusion SVM	0.66	N/A
Early Fusion RF	0.67	N/A
Video	0.68	0.732
Late Fusion (Weighted Average)	0.66	N/A

Table 2: Comparison of Results

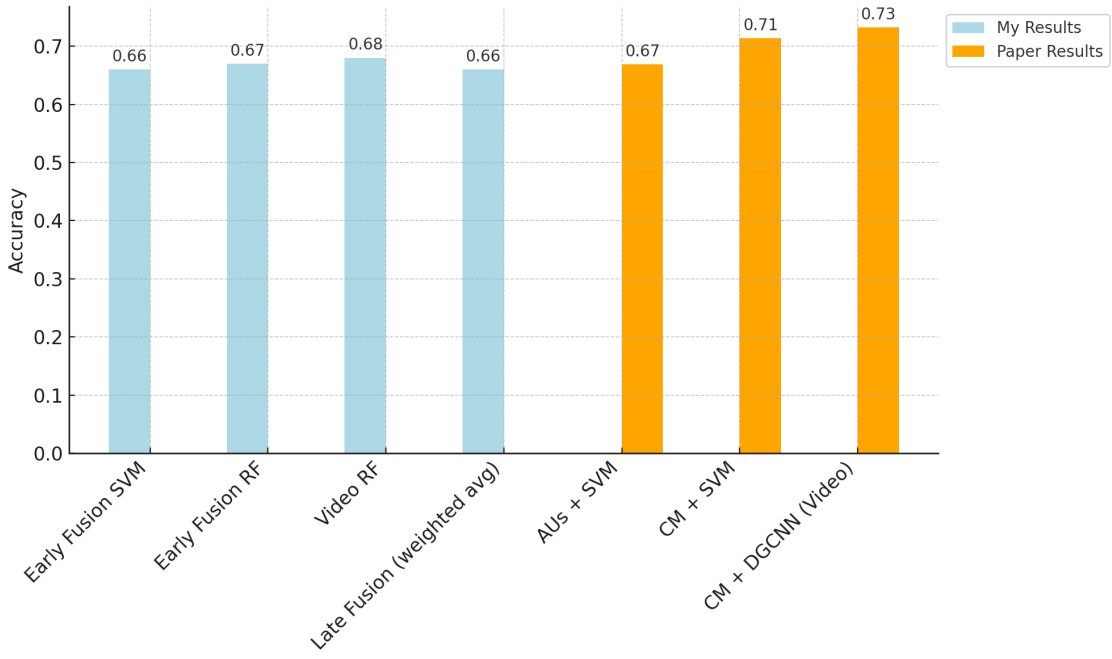


Figure 11: Comparing Result with GNN paper

7 Results and Discussion

The experimental results demonstrate a comparative analysis between Bio-Visual Fusion and GNN-based models. The Bio-Visual Fusion approach, which combines visual and biosignal data, showed better generalization across participants, especially for higher pain intensities. However, there were discrepancies between my results and the results reported in the original papers. These discrepancies can be attributed to dataset variations, missing data, and preprocessing steps that may have differed between the implementations.

The accuracy for the Bio-Visual Fusion approach using Random Forest classifiers was slightly lower than expected, which suggests potential improvements in feature selection and model tuning. On the other hand, the GNN approach was consistent in recognizing pain expressions based solely on facial landmarks, but it struggled with certain pain intensities where biosignals provided crucial additional information.

It is also worth noting that missing biosignal data, particularly from the zygomaticus and corrugator muscles, may have impacted the overall accuracy of the Bio-Visual Fusion model. Addressing this missing data in future work would likely lead to improvements in model performance.

8 Conclusion and Future Work

In conclusion, both the Bio-Visual Fusion and Deep Graph Neural Networks (GNN) methods demonstrated robustness in recognizing pain intensity across a diverse population. The Bio-Visual Fusion method, which leverages multi-modal data (visual and biosignal), provided slightly better accuracy, particularly for higher pain intensities. However, the GNN method excelled in scenarios where facial expressions were the primary indicators of pain.

One of the main challenges encountered in this study was the incomplete dataset, with missing EMG data from the zygomaticus and corrugator muscles, which limited the effectiveness of the Bio-Visual Fusion model. Additionally, discrepancies between the preprocessing steps used in this study and the referenced papers likely contributed to performance differences.

Future work should focus on addressing the limitations of the current dataset by incorporating complete biosignal data and exploring advanced feature selection methods to improve the performance of both models. Furthermore, investigating more sophisticated fusion techniques, such as deep learning-based fusion models, may yield better results. Incorporating additional physiological data modalities could also enhance the robustness and generalizability of the pain recognition system.

Overall, the findings from this study provide a solid foundation for the development of automated, person-independent pain intensity recognition systems.

9 Appendix

In this section we have supplementary material that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the problem or it is information that is too cumbersome to be included in the body of the paper.

9.1 Requirements

- **Python version**: Python 3.8 or higher
- **Required Libraries**:
  - `numpy`, `pandas`, `scikit-learn`, `matplotlib`, `seaborn`
  - `dlib`, `opencv-python`, `scipy`

9.2 Models used

In this study, the following models were used to evaluate the performance of different fusion approaches:

1. **Support Vector Machine (SVM)**: A classifier that finds the optimal hyperplane which maximizes the margin between different classes. It was used for both early and decision fusion approaches to classify pain intensity.

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0$$

2. **Random Forest**: An ensemble learning method that uses multiple decision trees to improve classification accuracy. It was applied to both video and biosignal data to assess the robustness of the hybrid fusion approach.

   For classification:

$$\hat{y} = \text{mode}\left(T_1(\mathbf{x}), T_2(\mathbf{x}), \ldots, T_B(\mathbf{x})\right)$$

   For regression:

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(\mathbf{x})$$

3. **Regression Model**: A regression analysis was performed to assess the relationship between different features and pain intensity levels, particularly for continuous pain assessment.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

9.3 Project Structure

The following is the project structure showing the directory organization and corresponding scripts used for the pain intensity recognition project.

```
Pre-Processing/
- biosignal/
  - Biosignals_final.py   # Pre-processing script for biosignals data
- video/
  - Complete Statistical Parameters.py   # Script for extracting statistical parameters
  - first second derivative.py           # Script for calculating derivatives
  - low pass filter.py                   # Low pass filter application
  - video_process.py                     # Main video processing script
Data/
- Early Fusion/
  - merged.csv       # Merged dataset for early fusion
- Late Fusion/
  - biosignals.csv   # Biosignals data for late fusion
  - video.csv        # Video data for late fusion
Utils/
- utils.py             # Additional utility functions for the project
main.py                  # Main script to run the project
definitions.py           # Definitions and configurations used across the project
__init__.py              # Initialization script for the package
README.md                # Project overview
requirements.txt         # Dependencies for the project
```

9.4 References

1. M. Kächele, P. Werner, A. Al-Hamadi, G. Palm, S. Walter, and F. Schwenker. "Bio-Visual Fusion for Person-Independent Recognition of Pain Intensity." *Proceedings of the 12th International Conference on Multiple Classifier Systems (MCS 2015)*, Springer, pp. 220-230, 2015.

2. S. Patania, G. Boccignone, S. Buršić, A. D'Amelio, and R. Lanzarotti. "Deep Graph Neural Network for Video-Based Facial Pain Expression Assessment." *Proceedings of the 37th ACM Symposium on Applied Computing (SAC'22)*, ACM, pp. 585-590, 2022.