Data Diffusion

Student: Yazdan Ghanavati, 2ex

1 Introduction

In critical care settings like the Intensive Care Unit (ICU), continuous physiological monitoring generates a wealth of time-series data. Effectively harnessing this data for predictive analytics, such as forecasting the outcome of a Spontaneous Breathing Trial (SBT), can significantly enhance clinical decision-making and improve patient management. This report evaluates and compares three distinct data diffusion and classification strategies designed to predict SBT outcomes from multivariate physiological signals. The core aim is to discern the most effective approach and understand the underlying reasons for their differential performance.

The strategies explored are:

- **Decision-Level Aggregation**: Each physiological parameter is processed by a separate classifier, and their individual predictions are then combined.
- **Feature-Level Aggregation**: Hand-crafted statistical features extracted from all physiological parameters are concatenated and fed into a single, unified classifier.
- **Data-Level Aggregation**: A 1D Convolutional Neural Network (CNN) directly processes the raw multivariate time series, automatically learning features.

A robust patient-aware cross-validation methodology is employed across all approaches to ensure the generalizability and statistical reliability of the results. The Area Under the Receiver Operating Characteristic Curve (AUC) is the primary metric for evaluation, chosen for its resilience to class imbalance and its ability to assess a model's discriminatory power across all possible classification thresholds.

1.1 Data-Level Aggregation: Employing Convolutional Neural Networks for Automated Feature Learning

Moving beyond manually engineered features, the Data-Level Aggregation approach employs a 1D Convolutional Neural Network (CNN). This allows the model to automatically learn relevant temporal features directly from raw multivariate physiological signals. The CNN processes the 8 distinct physiological components (e.g., SpO_2, HR) over 32 time steps, treating each as a channel. Its architecture uses multiple 1D convolutional layers for feature extraction, followed by flattening and dense classification layers, with a sigmoid output for probability scores.

1.2 Potential Enhancements for CNN Performance

While the current CNN architecture provides a foundational model for data-level aggregation, its performance could potentially be further enhanced through several advanced deep learning techniques, which were outside the immediate scope of this laboratory exercise:

- **Increased Model Complexity**: Exploring more convolutional layers, varying filter sizes, or increasing the number of filters could allow the model to capture more intricate feature representations.
- **Regularization Techniques**: Implementing methods like Dropout or L1/L2 regularization is crucial to prevent overfitting, especially given the relatively small dataset size (353 data entries).
- **Batch Normalization**: Adding Batch Normalization layers can stabilize and accelerate the training process.
- **Hyperparameter Tuning**: Systematically optimizing parameters such as learning rate, number of epochs, batch size, kernel sizes, and filter counts could yield significant performance gains.
- **Hybrid Architectures**: Combining CNNs with Recurrent Neural Network (RNN) layers (e.g., LSTMs or GRUs) could allow the model to capture long-range temporal dependencies more effectively.

These advanced techniques represent vital steps in maximizing the predictive power and robustness of deep learning models in practical applications.

2 Comparative Performance Analysis

This project compared the effectiveness of the three distinct aggregation strategies using patient-aware cross-validation (20-30 iterations) and the AUC metric.

2.1 Summary of Mean AUC Scores

Architecture	**Mean AUC Score**
Decision-Level Aggregation	0.646
Feature-Level Aggregation	0.865
Data-Level Aggregation (CNN)	0.815

Table 1: Summary of Mean AUC Scores for Different Aggregation Strategies

2.2 Visualizing Performance Distribution

Figure 1 visually compares the distributions of AUC scores for all three architectures, providing insight into their average performance and consistency.

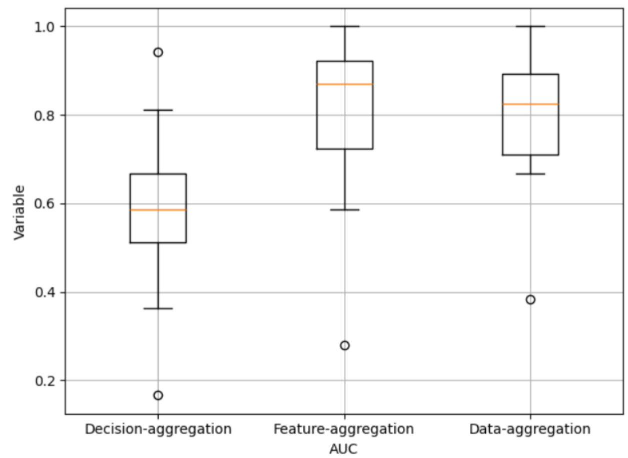

Figure 1: Box Plot Comparison of AUC Distributions for Decision-Level, Feature-Level, and Data-Level Aggregation Architectures

2.3 Interpreting Performance Gaps

- **Decision-Level Aggregation (Mean AUC: 0.646)**: This approach performed lowest as individual classifiers failed to capture crucial interdependencies between physiological parameters, leading to inconsistent predictions.
- **Feature-Level Aggregation (Mean AUC: 0.865)**: Achieving the highest mean AUC, this strategy leveraged well-designed hand-crafted features. Concatenating these features allowed a single classifier to learn complex interactions among them, proving highly effective and consistent.
- **Data-Level Aggregation (CNN) (Mean AUC: 0.815)**: While strong, the CNN did not surpass feature-level aggregation. This is likely due to the limited dataset size (353 entries), which can restrict deep learning models from fully learning robust features from scratch. Hand-crafted features, derived from domain expertise, often offer more stable representations with scarcer data, where deep learning might struggle to fully optimize.

3 Understanding Model Discrimination: The ROC Curve

AUC is a robust metric for binary classifiers, especially with imbalanced classes, measuring a model's ability to distinguish between positive and negative cases across all thresholds. The Receiver Operating Characteristic (ROC) curve plots True Positive Rate (Sensitivity) against False Positive Rate ($1-$ Specificity) for various threshold settings. A perfect classifier reaches the top-left corner, while random guessing forms a diagonal line (AUC $= 0.50$).

Figure 2 presents a representative ROC curve for the Data-Level Aggregation (CNN) model.

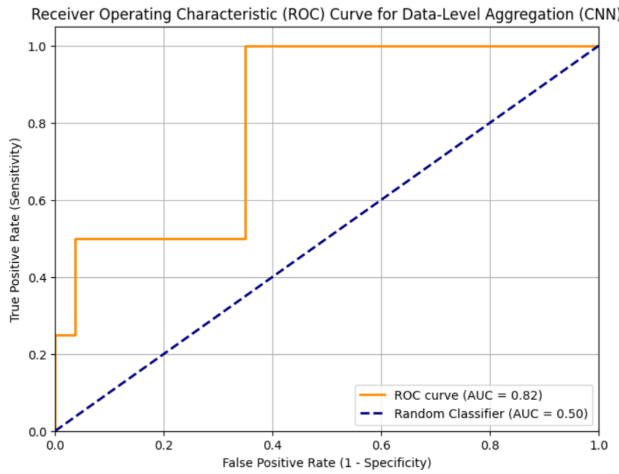

Figure 2: Representative Receiver Operating Characteristic (ROC) Curve for the Data-Level Aggregation (CNN) Classifier on a Test Split (AUC = 0.82).

3.1 Interpretation of the CNN's ROC Curve

The CNN's ROC curve, with an AUC of 0.82, confirms strong discriminative power. Its distinct stepped or staircase shape is observed due to:

- **Limited Distinct Probability Outputs**: The model generates a relatively small number of unique probability scores, causing abrupt changes in TPR and FPR as the threshold crosses these values.
- **Small Test Set Size**: With fewer samples in the test set, discrete $(\mathrm{FPR}, \mathrm{TPR})$ pairs lead to a less smooth curve, where each step represents a threshold change affecting a few samples.

The sharp vertical segments indicate significant TPR gains with minimal FPR increases, while flat horizontal segments suggest substantial FPR increases are needed for further sensitivity gains. For instance, the large step from approximately (0.05, 0.5) to (0.35, 0.5) implies a considerable rise in false alarms to increase sensitivity beyond 50%. This stepped pattern accurately reflects the model's behavior and prediction distribution given the dataset constraints.

4 Conclusion

This laboratory successfully evaluated three data aggregation architectures for SBT outcome prediction. Feature-level aggregation emerged as the most effective (**Mean AUC: 0.865**), leveraging well-designed hand-crafted features to learn complex interactions across physiological signals.

While the data-level CNN demonstrated strong performance (**Mean AUC: 0.815**) by automatically extracting features, its results were slightly outmatched by feature-level aggregation. This is primarily attributed to the limited dataset size, where hand-crafted features often provide a more stable and efficient representation compared to a deep network learning from scratch.

In conclusion, for this dataset, integrating expert-derived features proved optimal. However, the CNN approach offers significant potential for future development with larger datasets, promising fully automated and more sophisticated feature extraction.