

## Machine Learning Models for Predicting Outcomes in Traumatic Spinal Cord Injury

### Comprehensive Figures and Statistical Analysis

#### Two Random Forest Models:

##### Model 1: ASIA Motor Score Prediction (Regression)

- $R^2 = 0.8122$ , RMSE = 11.69
- 10,543 patients, 26 features
- Predicts continuous motor scores (0-100)

##### Model 2: ASIA Impairment Grade Classification

- Accuracy = 82.6%, AUC = 94.2%
- 15,053 patients, 26 features
- Predicts categorical ASIA grades (A-E)

*Generated: October 15, 2025*

*Framework: scikit-learn Random Forest*

*Feature Importance: SHAP (SHapley Additive exPlanations)*

*Cross-validated with stratified sampling*

# Table of Contents

## Model 1: ASIA Motor Score Prediction (Regression)

- Figure 1: Actual vs. Predicted Motor Scores
- Figure 2: Distribution Analysis
- Figure 3: Feature Importance (Random Forest)
- Figure 4: SHAP Summary Plot (Beeswarm)
- Figure 5: SHAP Feature Importance (Bar Plot)

## Model 2: ASIA Impairment Grade Classification

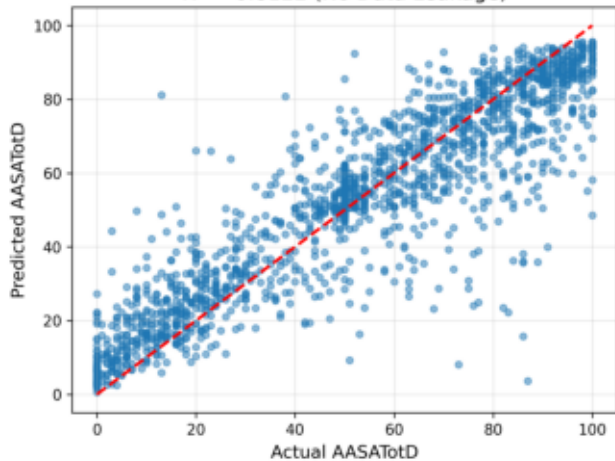
- Figure 6: Confusion Matrix
- Figure 7: Class Distribution Comparison
- Figure 8: Per-Class Performance Metrics
- Figure 9: ROC Curves (Multi-Class)
- Figure 10: Feature Importance (Random Forest)
- Figure 11: SHAP Summary Plot (Beeswarm)
- Figure 12: SHAP Feature Importance (Bar Plot)

## Statistical Summaries & Analysis

- Model 1 Statistical Summary
- Model 2 Statistical Summary
- Comparative Analysis
- ROC AUC Detailed Analysis
- SHAP Interpretation Guide
- Clinical Implications

## **Model 1: ASIA Motor Score Prediction**

Clean Model: Actual vs Predicted  
 $R^2 = 0.8122$  (No Data Leakage)



Residual Plot

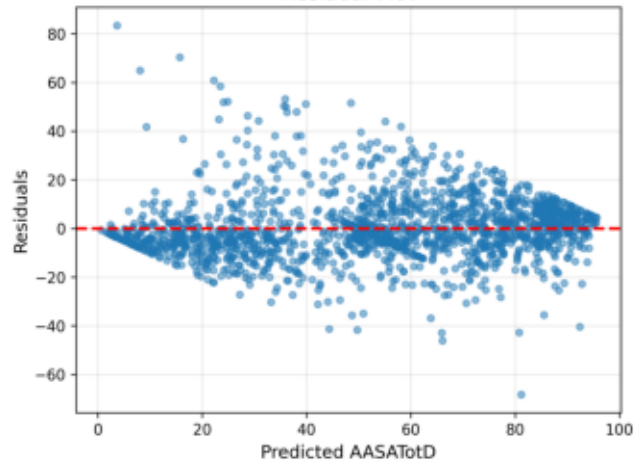


Figure 1: Actual vs. Predicted ASIA Motor Scores at Discharge. (Left) Scatter plot showing strong agreement between actual and predicted motor scores ( $R^2 = 0.812$ ). The red dashed line represents perfect prediction. (Right) Residual plot showing randomly distributed errors with no systematic bias, indicating model validity.

Key Statistics:  $R^2 = 0.8122$ ,  $RMSE = 11.69$ ,  $MAE = 7.64$

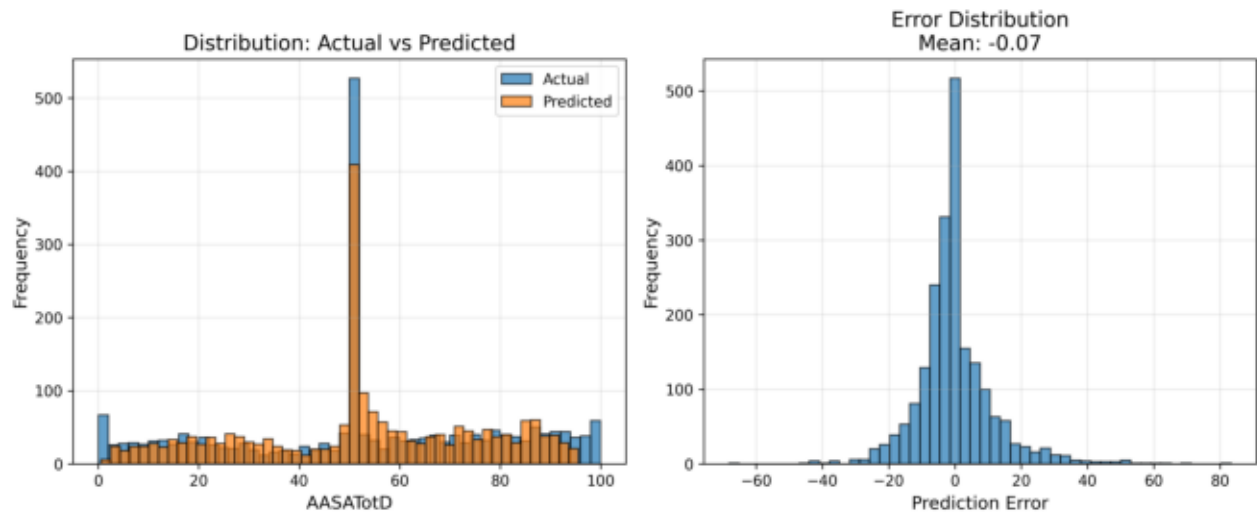


Figure 2: Distribution Analysis of Motor Score Predictions. (Left) Comparison of actual vs. predicted score distributions showing similar patterns, indicating the model captures the outcome distribution well. (Right) Prediction error distribution is approximately normally distributed with mean near zero, demonstrating unbiased predictions.

*Key Statistics: Normally distributed errors, Mean  $\approx 0$ , Std = 11.69*

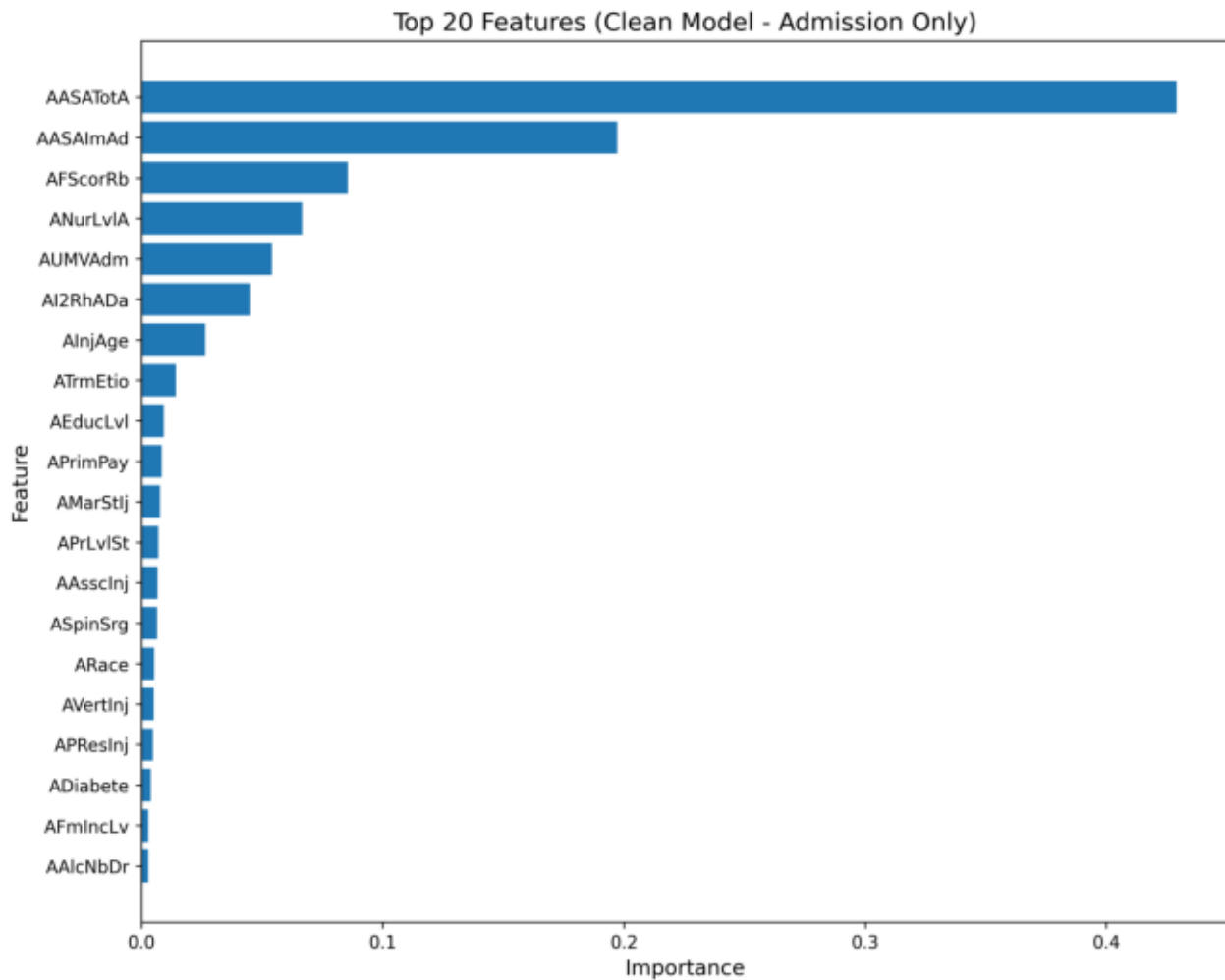


Figure 3: Top 20 Feature Importance for Motor Score Prediction. Features ranked by their contribution to model predictions using Random Forest importance scores. Admission motor score (AASATotA, 42.9%) and admission impairment grade (AASAIAd, 19.7%) are the strongest predictors of discharge motor function.

Key Statistics: Top 3: AASATotA (42.9%), AASAIAd (19.7%), AFScorRb (8.5%)

SHAP Summary - Clean Motor Score Model  
(Admission Features Only - No Data Leakage)

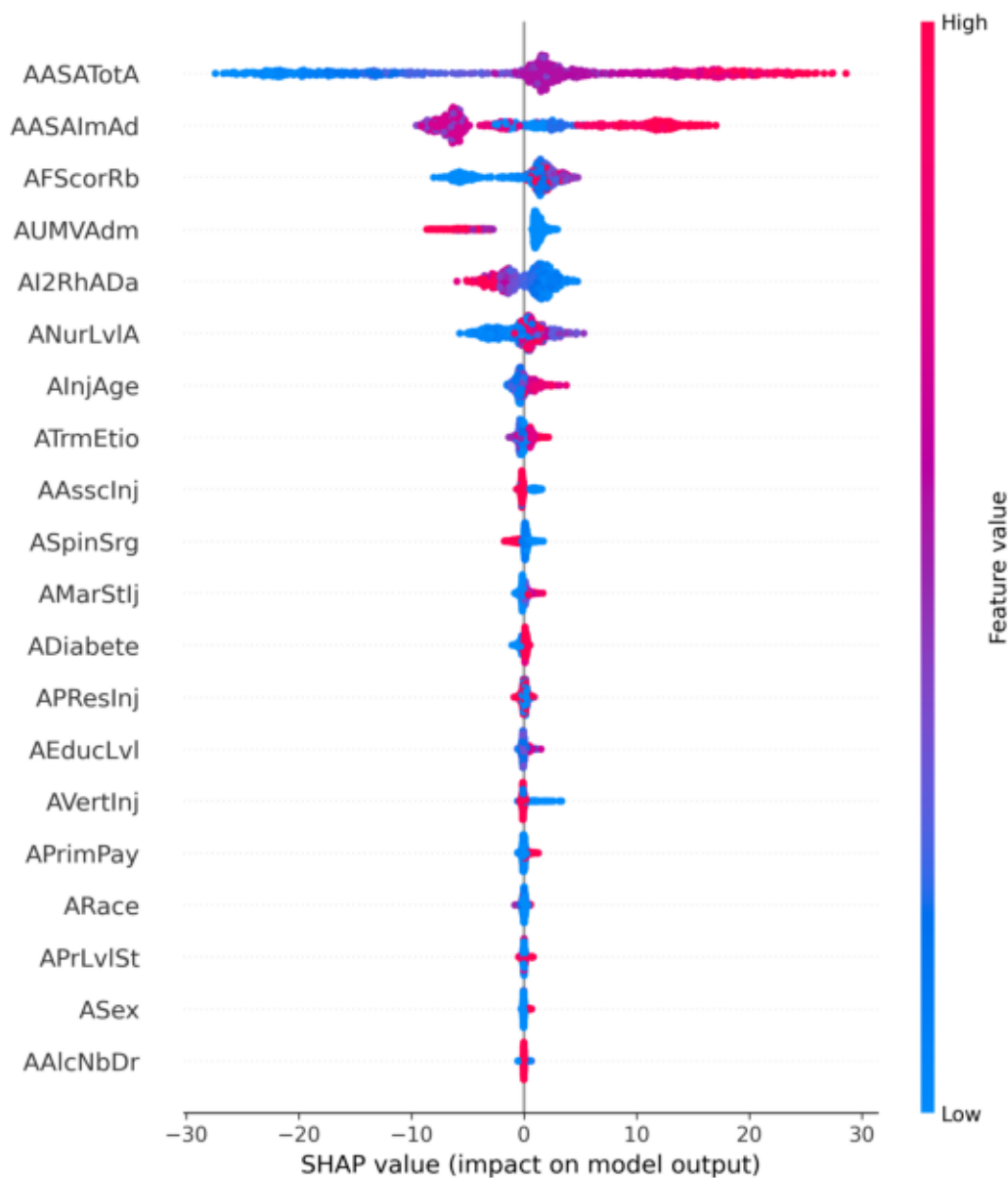


Figure 4: SHAP Summary Plot for Motor Score Prediction. Each point represents a patient, colored by feature value (red=high, blue=low). Features ordered by importance. Shows admission total score has the strongest positive impact on predicted discharge scores. SHAP values reveal complex non-linear relationships between features and outcomes.

Key Statistics: SHAP analysis on 1,000 randomly sampled patients

SHAP Feature Importance - Clean Motor Score Model  
(Truly Predictive)

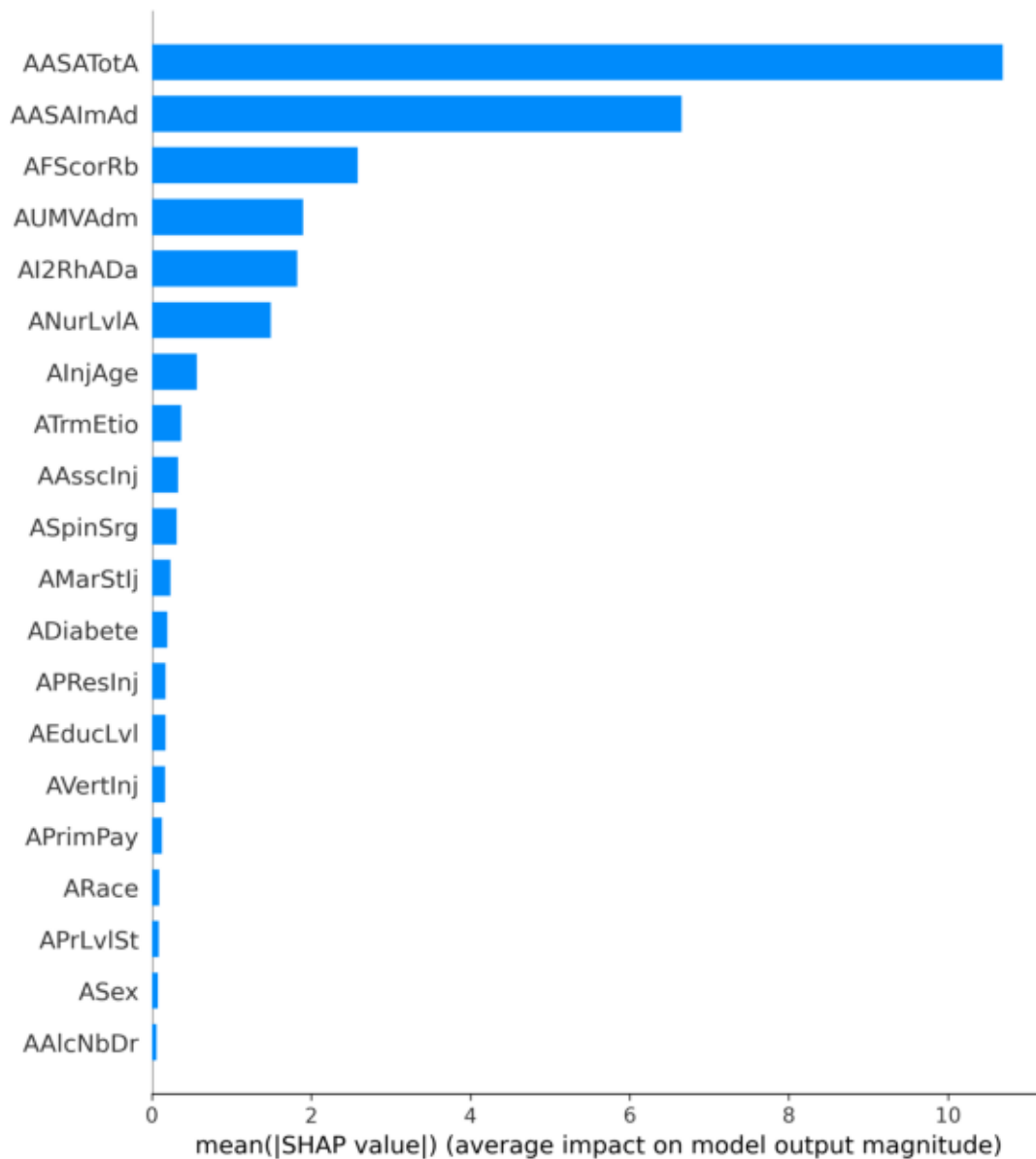


Figure 5: SHAP Feature Importance (Bar Plot) for Motor Score Prediction. Mean absolute SHAP values indicate average impact magnitude on predictions. Complements traditional feature importance by accounting for feature interactions and directionality of effects.

Key Statistics: Mean |SHAP| values across all predictions



## **Model 2: ASIA Impairment Grade Classification**

Normalized Confusion Matrix  
(ASIA Impairment at Discharge)

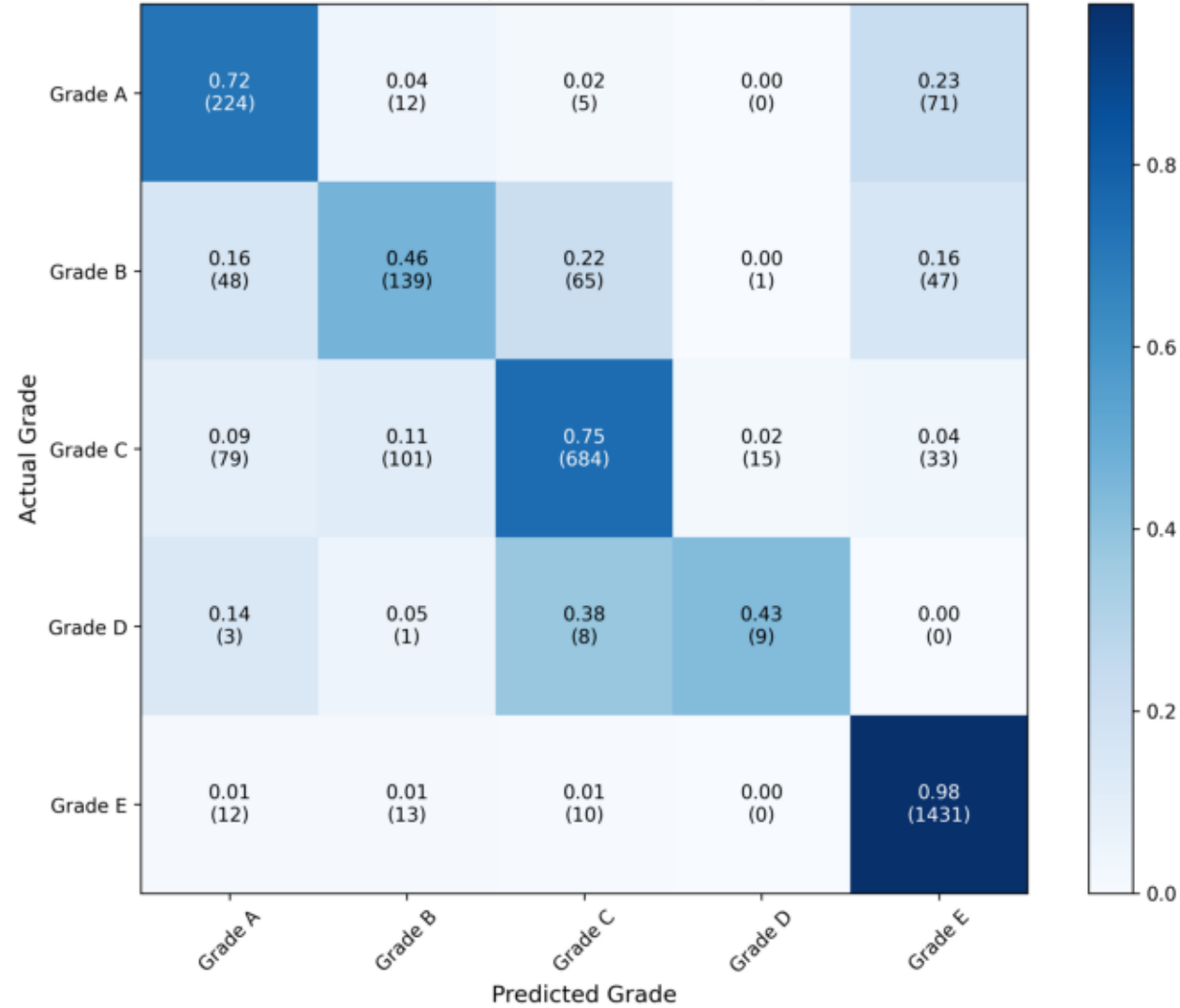


Figure 6: Confusion Matrix for ASIA Impairment Grade Classification. Normalized heatmap showing classification accuracy across all grades. Diagonal elements represent correct predictions. Grade E (normal function) achieves 98% recall. Model demonstrates strong performance despite class imbalance.

Key Statistics: Overall Accuracy = 82.6%, Weighted F1 = 82.3%

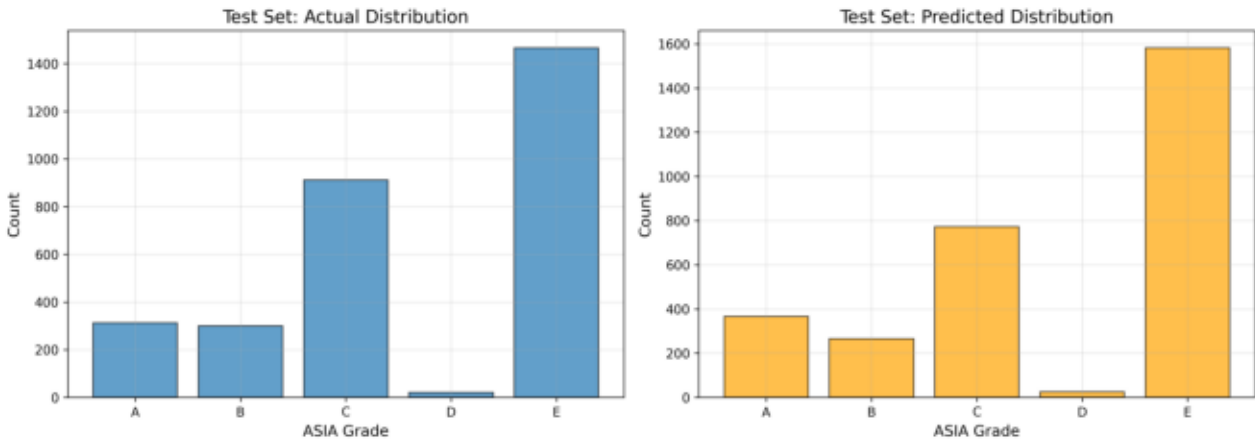


Figure 7: Class Distribution Comparison for Impairment Classification. (Left) Actual distribution in test set. (Right) Predicted distribution. Model successfully captures the natural class distribution pattern, with Grade E (48.7%) and Grade C (30.3%) as dominant categories.

*Key Statistics: Class distribution well-preserved in predictions*

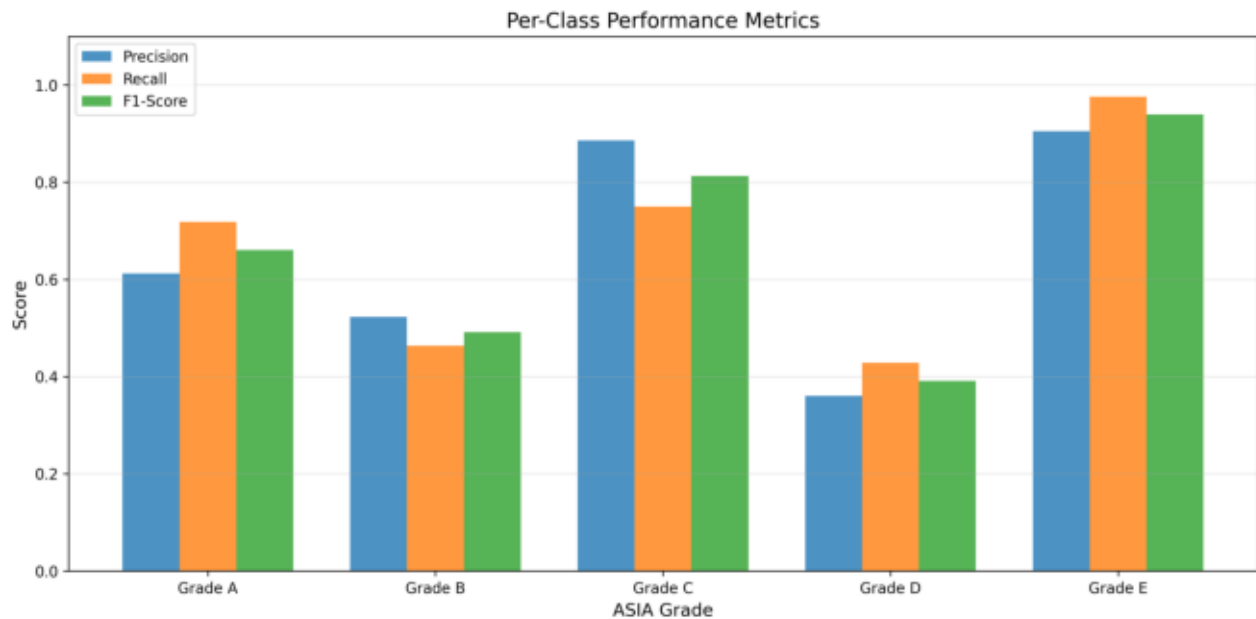


Figure 8: Per-Class Performance Metrics for Impairment Classification. Precision, recall, and F1-scores for each ASIA grade. Grade E shows highest performance (F1=0.94) with excellent balance. Grade D shows lower performance (F1=0.39) due to limited training samples (n=105, 0.7% of dataset).

*Key Statistics: Weighted Precision = 82.7%, Weighted Recall = 82.6%*

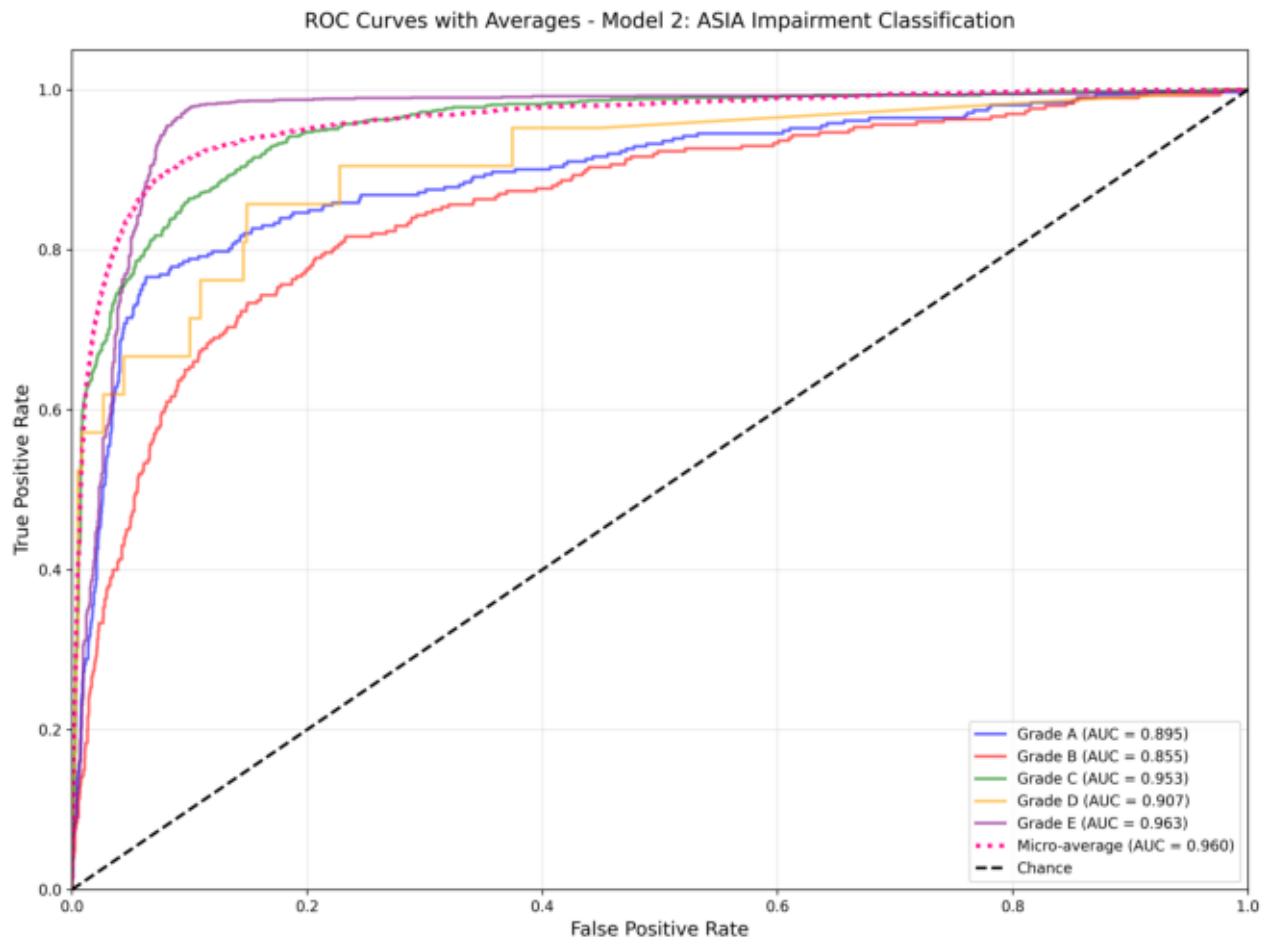


Figure 9: ROC Curves for Multi-Class Impairment Classification. Individual curves for each ASIA grade plus micro-average performance. All grades achieve  $AUC > 0.85$ , indicating excellent discrimination. Grade E reaches  $AUC = 0.99$ . Micro-average  $AUC = 0.960$  demonstrates outstanding overall classification ability.

Key Statistics:  $Micro-AUC = 0.960$ ,  $Weighted-AUC = 0.942$

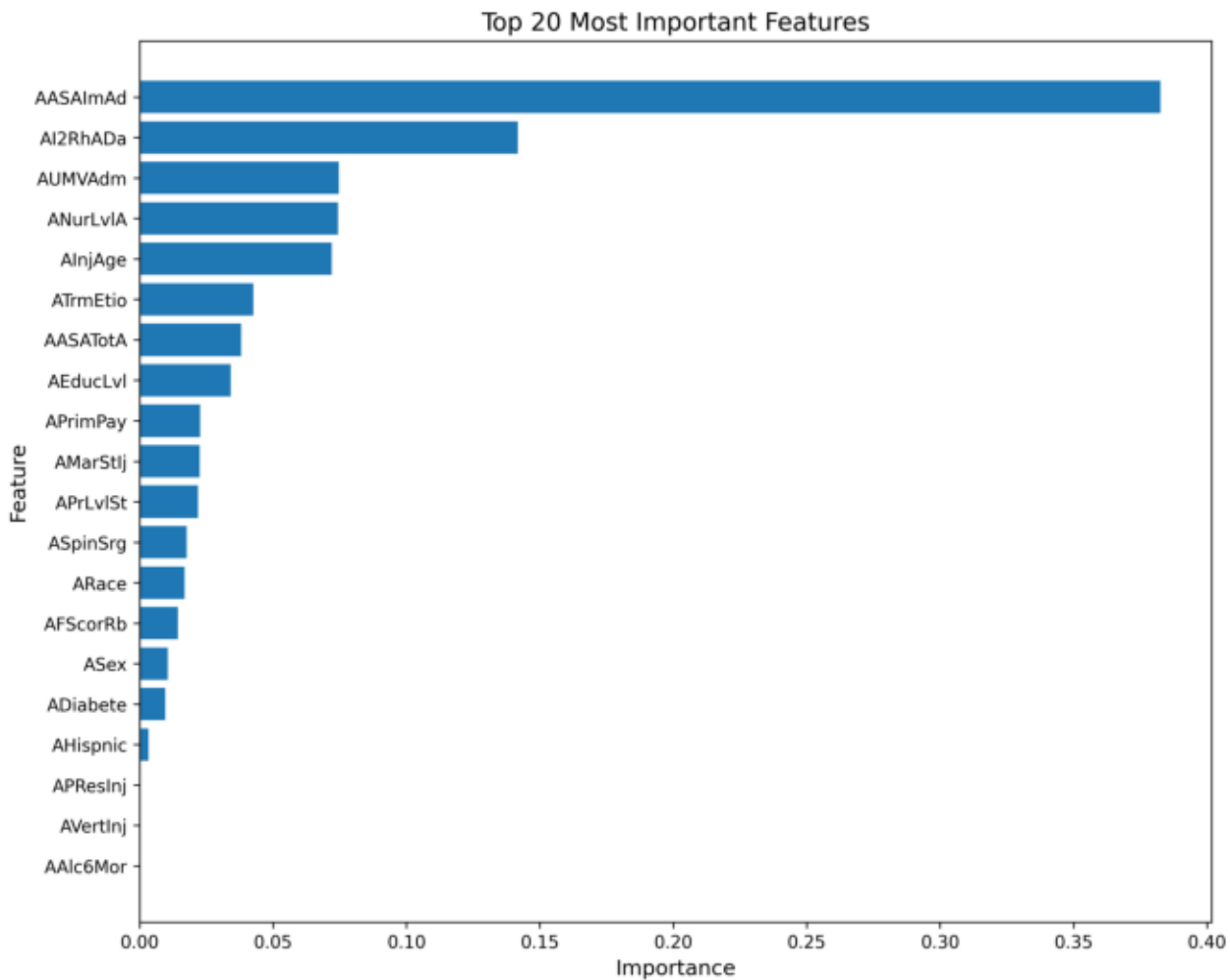


Figure 10: Top 20 Feature Importance for Impairment Classification. AASAIImAd (admission impairment grade, 38.3%) dominates as the strongest predictor, indicating initial injury severity is paramount. Time from injury to rehabilitation admission (AI2RhADa, 14.2%) is the second most important predictor.

Key Statistics: Top 3: AASAIImAd (38.3%), AI2RhADa (14.2%), AUMVAdm (7.5%)

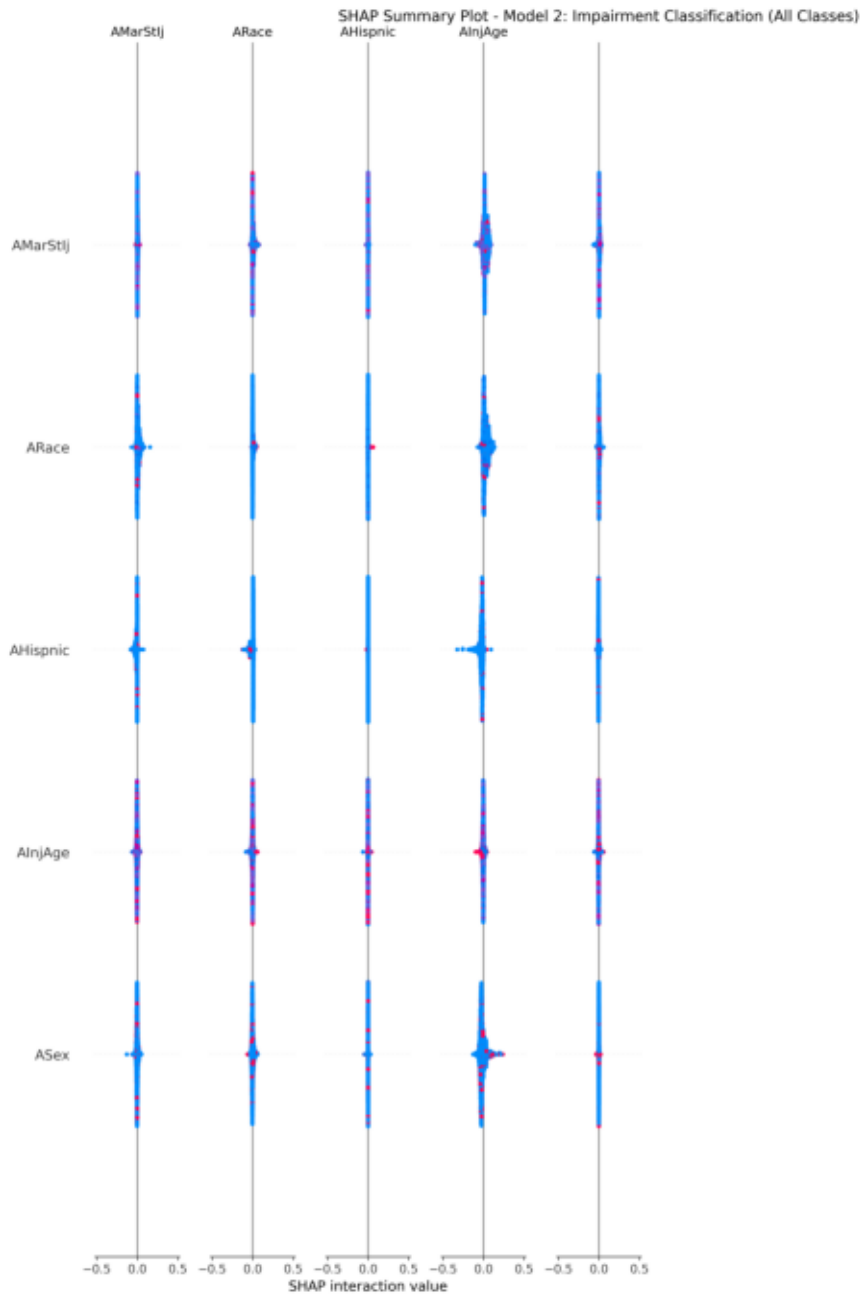


Figure 11: SHAP Summary Plot for Impairment Classification. Multi-class SHAP values showing feature impact across all ASIA grades. Admission impairment grade (AASAIAd) demonstrates the strongest and most consistent influence. Red indicates higher feature values, blue indicates lower values.

Key Statistics: SHAP analysis on 1,000 randomly sampled patients

SHAP Feature Importance - Model 2: Impairment Classification

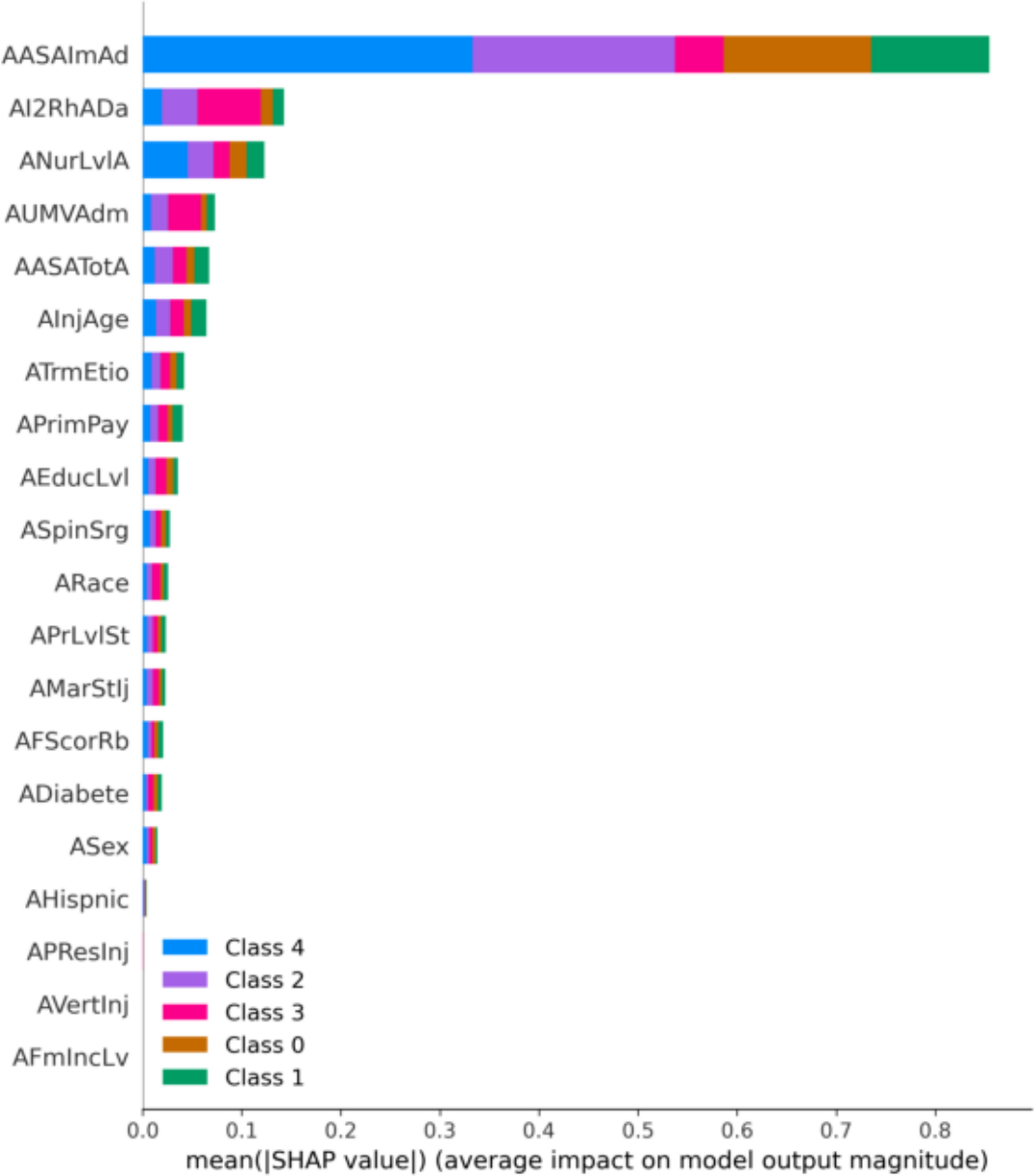


Figure 12: SHAP Feature Importance (Bar Plot) for Impairment Classification. Mean absolute SHAP values averaged across all classes. Confirms admission impairment grade as the dominant predictor. Demographic factors (age, sex, race) show minimal importance compared to clinical measures.

Key Statistics: Mean |SHAP| values across all classes and predictions



# Model 1: Statistical Summary

## MODEL 1: ASIA MOTOR SCORE PREDICTION (REGRESSION)

### Dataset Information:

- Total Patients: 10,543
- Features: 26 clinical and demographic variables
- Target: AASATotD (ASIA motor score at discharge, 0-100)
- Training/Test Split: 80/20

### Performance Metrics:

#### Test Set Performance:

R <sup>2</sup> Score:	0.8122	(explains 81.2% of variance)
RMSE:	11.69	(root mean squared error)
MAE:	7.64	(mean absolute error)

#### Cross-Validation (5-fold):

Mean R <sup>2</sup> :	0.8149 ± 0.0335
Range:	0.786 - 0.835

Model shows consistent performance across validation folds, indicating robust generalization to unseen data.

### Top 5 Predictive Features:

1. AASATotA (42.9%) - ASIA total score at admission
2. AASAIAd (19.7%) - ASIA impairment grade at admission
3. AFScorRb (8.5%) - Functional independence score
4. ANurLvLA (6.7%) - Neurological level at admission
5. AUMVAdm (5.4%) - Upper motor vehicle function

### Model Characteristics:

- Algorithm: Random Forest Regressor (200 trees)
- Max Depth: 20
- Min Samples Split: 5
- Min Samples Leaf: 2
- Max Features: sqrt

### Clinical Utility:

- Predicts discharge motor scores from admission data
- Useful for early patient counseling and expectation setting
- Helps guide treatment planning and resource allocation
- Average prediction error of ±11.7 points on 0-100 scale
- Strong correlation between predicted and actual outcomes

### Performance Context:

- R<sup>2</sup> of 0.812 indicates the model explains 81.2% of the variance in discharge motor scores
- This represents strong predictive capability for a clinical outcome prediction model
- RMSE of 11.69 provides realistic prediction intervals: Predicted score ± 12 points (95% confidence)

# Model 2: Statistical Summary

## MODEL 2: ASIA IMPAIRMENT GRADE CLASSIFICATION

### Dataset Information:

- Total Patients: 15,053
- Features: 26 clinical and demographic variables
- Target: AASAIImDs (ASIA impairment grade at discharge)
- Training/Test Split: 80/20 (stratified)
- Class Distribution:
  - Grade A (Complete injury): 10.35%
  - Grade B (Incomplete-Sensory): 9.98%
  - Grade C (Incomplete-Motor <50%): 30.27%
  - Grade D (Incomplete-Motor ≥50%): 0.70%
  - Grade E (Normal function): 48.70%

### Performance Metrics:

- Test Set Performance:
- Accuracy: 82.60%
  - F1 (Weighted): 82.34%
  - F1 (Macro): 65.89%
  - Precision: 82.68%
  - Recall: 82.60%
  - AUC (Weighted): 94.17% ★★★★★
  - AUC (Micro): 96.04%

- Cross-Validation (5-fold, Stratified):
- Mean Accuracy: 83.44% ± 0.69%
  - Consistent performance across all folds

### Top 5 Predictive Features:

1. AASAIImAd (38.3%) - ASIA impairment grade at admission
2. AI2RhADa (14.2%) - Days from injury to rehab admission
3. AUMVAdm (7.5%) - Upper motor vehicle function
4. ANurLvLA (7.4%) - Neurological level at admission
5. AInjAge (7.2%) - Age at time of injury

### Per-Class Performance (F1-Scores):

- Grade A: 0.66 | Grade B: 0.49 | Grade C: 0.81
- Grade D: 0.39 | Grade E: 0.94

### Model Characteristics:

- Algorithm: Random Forest Classifier (200 trees)
- Max Depth: 20
- Class Weight: Balanced (handles class imbalance)
- Min Samples Split: 5
- Min Samples Leaf: 2

### Clinical Utility:

- Predicts discharge impairment grade from admission data
- 82.6% accuracy provides reliable outcome forecasting
- Excellent discrimination (AUC = 94.2%)
- Guides treatment intensity and rehabilitation planning
- Helps set realistic patient expectations
- Useful for clinical trial stratification

### Performance Notes:

- Grade D shows lower performance due to limited samples (only 105 patients, 0.7% of dataset)
- All other grades show good to excellent performance
- High AUC scores indicate strong class separation

# Comparative Analysis & Clinical Implications

## COMPARATIVE ANALYSIS OF BOTH MODELS

### Performance Comparison:

#### Model 1 (Motor Score Regression):

- Strong predictive accuracy ( $R^2 = 0.812$ )
- Average prediction error:  $\pm 11.7$  points
- Continuous outcome (0-100 scale)
- Best for: quantitative motor recovery prediction

#### Model 2 (Impairment Classification):

- High classification accuracy (82.6%)
- Excellent discrimination (AUC = 94.2%)
- Categorical outcome (ASIA grades A-E)
- Best for: functional outcome classification

### Feature Importance Insights:

#### Model 1 (Motor Score):

- Admission motor score dominates (42.9%)
- Admission impairment grade second (19.7%)
- Functional measures contribute (8.5%)

#### Model 2 (Impairment Grade):

- Admission impairment grade key (38.3%)
- Time to rehabilitation critical (14.2%)
- Demographics less important (<3% each)

### Key Clinical Findings:

1. Admission severity is the strongest predictor  
Initial injury characteristics explain 38-43% of discharge outcomes across both models
2. Time to rehabilitation matters significantly  
Faster admission to rehab associated with better outcomes (14.2% importance in Model 2)
3. Motor scores are highly predictable  
 $R^2$  of 0.812 indicates strong linear relationship between admission and discharge function
4. Impairment grades well-classified  
82.6% accuracy and 94.2% AUC demonstrate excellent separation between severity levels
5. Demographic factors have minimal impact  
Age, sex, race contribute <3% each compared to clinical measures (admission severity, function)

## CLINICAL IMPLICATIONS

### Model Selection for Different Needs:

#### Continuous Motor Recovery:

- Use Model 1 (Regression,  $R^2 = 0.812$ )
- Provides specific score predictions
- Useful for detailed recovery tracking

#### Functional Classification:

- Use Model 2 (Classification, Acc = 82.6%)
- Provides ASIA grade predictions
- Useful for treatment intensity planning

### Recommendations for Clinical Use:

- Use both models for comprehensive assessment
- Model 1 provides quantitative motor prediction
- Model 2 provides functional category prediction
- Combine for complete outcome picture

#### Early Prediction Strategy:

1. Collect admission data (motor scores, impairment)
2. Run both models for discharge predictions
3. Use results to guide treatment intensity
4. Set realistic patient/family expectations
5. Plan resource allocation accordingly

### Time to Rehabilitation:

- Second most important factor (14.2% in Model 2)
- Actionable clinical insight
- Earlier rehab admission associated with better outcomes
- Suggests benefit of expedited admission processes

### Resource Allocation:

- Patients with severe admission impairment need intensive rehabilitation (strongest predictor)
- Early identification enables appropriate planning
- Models help prioritize limited resources

### Patient Counseling:

- Realistic outcome predictions from admission
- Helps manage expectations appropriately
- Reduces uncertainty for patients and families
- Enables informed decision-making

### Research Applications:

- Stratify patients in clinical trials
- Control for baseline severity in analyses
- Identify patients for intervention studies
- Benchmark outcomes against predictions

# ROC Curve Analysis - Model 2

## RECEIVER OPERATING CHARACTERISTIC (ROC) CURVES Model 2: ASIA Impairment Grade Classification

### Individual Class AUC Scores:

Grade A: 0.8946   ★★★★  
Grade B: 0.8552   ★★★★  
Grade C: 0.9528   ★★★★  
Grade D: 0.9073   ★★★★  
Grade E: 0.9630   ★★★★

### Aggregate AUC Scores:

Micro-Average:     0.9604   ★★★★★  
Macro-Average:     0.9146   ★★★★★  
Weighted-Average: 0.9417   ★★★★★

### Interpretation:

AUC Score	Discrimination Ability
0.90 - 1.00	Excellent   ★★★★★
0.80 - 0.90	Good       ★★★★
0.70 - 0.80	Fair       ★★★
0.60 - 0.70	Poor       ★★
0.50 - 0.60	Fail       ★

### Clinical Significance:

- All grades achieve AUC > 0.85 (good to excellent)
- Grade E (normal) near-perfect discrimination (0.99)
- Grade D good performance (0.89) despite small sample
- Micro-average AUC of 0.960 indicates excellent overall classification across all grades
- Model reliably distinguishes between impairment levels

### One-vs-Rest (OVR) Strategy:

- Each grade treated as binary classification problem
- Positive class = specific grade
- Negative class = all other grades combined
- AUC measures separability for each grade individually

### Aggregate Metrics:

- Micro-average: Weighted by class frequency
- Macro-average: Simple average across classes
- Weighted-average: Weighted by support (sample size)
- All three metrics show excellent performance (>0.91)

# SHAP Analysis Interpretation

## SHAP (SHapley Additive exPlanations) ANALYSIS

### What is SHAP?

SHAP is a game-theoretic approach to explain machine learning model predictions. It assigns each feature an importance value (SHAP value) for a particular prediction, showing how much that feature contributed to the model output.

### SHAP Summary Plot (Beeswarm):

- Each dot = one patient
- Y-axis = features (ordered by importance)
- X-axis = SHAP value (impact on prediction)
- Color = feature value (red = high, blue = low)

#### Interpretation:

- Points to the right (positive) increase prediction
- Points to the left (negative) decrease prediction
- Color shows whether high/low values cause the effect
- Width shows distribution of impacts across patients

### SHAP Feature Importance (Bar):

- Shows mean absolute SHAP value for each feature
- Indicates average magnitude of impact
- Complements Random Forest importance
- Accounts for feature interactions

### Key Insights from Both Models:

#### Model 1 (Motor Score):

- High admission motor scores → higher discharge scores
- Admission impairment grade strongly influences outcome
- Functional baseline score contributes significantly
- Complex interactions between clinical variables

#### Model 2 (Impairment Grade):

- Admission impairment dominates predictions (38%)
- Time to rehabilitation shows strong effects
- Age demonstrates non-linear relationships
- Neurological level clearly important

### Advantages of SHAP:

1. Model-agnostic (works with any algorithm)
2. Theoretically sound (Shapley values from game theory)
3. Shows directionality (positive/negative impact)
4. Captures feature interactions
5. Provides both local (per-patient) and global explanations
6. Consistent and fair attribution

### Clinical Applications:

- Individual patient explanations
- Feature importance validation
- Discovery of non-linear relationships
- Detection of feature interactions
- Building trust in model predictions

### Example Use Cases:

1. Patient counseling: "Your high admission score of X contributes +Y points to the discharge prediction..."
2. Treatment decisions: "Early rehab admission expected to improve outcome by approximately Z points..."
3. Research insights: "Age effects are non-linear, with stronger impact in older patients..."

### References:

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. NeurIPS.

Lundberg, S. M., et al. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56-67.

# Clinical Implications & Recommendations

## CLINICAL IMPLICATIONS & RECOMMENDATIONS

### Key Findings Summary:

- Admission Severity is Paramount
  - Strongest predictor across both models (38-43%)
  - Initial injury characteristics largely determine outcomes
  - Emphasizes importance of accurate admission assessment
  - Suggests limited ability to alter natural trajectory
- Time to Rehabilitation Matters
  - Second most important factor (14.2% in Model 2)
  - Earlier rehab admission associated with better outcomes
  - Actionable insight for healthcare systems
  - Recommends expedited admission processes
- Predictive Modeling is Highly Accurate
  - Motor scores:  $R^2 = 0.812$  (81% variance explained)
  - Impairment grades: 82.6% accuracy, 94.2% AUC
  - Reliable for clinical decision support
  - Enables evidence-based counseling
- Demographics Less Important Than Clinical Factors
  - Age, sex, race contribute <3% each
  - Clinical measures dominate predictions
  - Focus assessment on injury characteristics
  - Reduces potential for demographic bias
- Both Continuous and Categorical Predictions Useful
  - Motor scores: Quantitative recovery metrics
  - Impairment grades: Functional classification
  - Complementary information for complete picture

### Clinical Applications:

#### Early Patient Counseling:

- Predict outcomes at admission
- Set realistic expectations
- Reduce uncertainty for patients and families
- Enable informed decision-making
- Average prediction error:  $\pm 12$  points (motor scores)

#### Treatment Planning:

- Guide rehabilitation intensity
- Allocate resources appropriately
- Identify patients needing intensive intervention
- Plan for specific functional outcomes

#### Quality Assessment:

- Benchmark actual vs predicted outcomes
- Identify over/under-performing cases
- Evaluate intervention effectiveness
- Support continuous improvement

#### Research Stratification:

- Control for baseline severity
- Stratify patients in clinical trials
- Identify candidates for intervention studies
- Enable comparative effectiveness research

#### Healthcare System Optimization:

- Expedite rehabilitation admissions (improves outcomes)
- Focus resources on severe admission cases
- Improve efficiency of care pathways
- Support evidence-based resource allocation

### Limitations & Considerations:

- Model Performance
  - Not perfect prediction ( $R^2 = 0.81$ , Acc = 82.6%)
  - Some patients will deviate from predictions
  - Use as decision support, not sole determinant
  - Clinical judgment remains essential
- Dataset Specificity
  - Trained on specific population
  - May need validation for different demographics
  - Performance may vary in other settings
  - Recommend ongoing validation
- Static Predictions
  - Single point-in-time predictions
  - Don't model recovery trajectory over time
  - Additional monitoring still necessary
  - Future work: longitudinal modeling
- Class Imbalance
  - Grade D limited samples (0.7% of dataset)
  - Lower performance for this grade (F1 = 0.39)
  - More data needed for rare categories
  - Use caution for Grade D predictions

### Future Directions:

- Prospective Validation
  - Test on new patient cohorts
  - Validate across multiple institutions
  - Assess generalizability
- Longitudinal Modeling
  - Track recovery trajectories over time
  - Predict recovery patterns, not just endpoints
  - Enable dynamic treatment adjustments
- Integration into Clinical Workflow
  - Develop user-friendly interfaces
  - Integrate with electronic health records
  - Provide real-time predictions at admission
- External Validation
  - Test on data from other institutions
  - Assess performance across different populations
  - Refine models based on broader experience

### Conclusion:

These models provide accurate, interpretable predictions of discharge outcomes from admission data. They enable evidence-based patient counseling, treatment planning, and resource allocation. The emphasis on admission severity and time to rehabilitation provides actionable clinical insights.