

Medical Device Recall Classification

Implementation Report

5-Level Classification System

Generated on August 25, 2025

Data Science Team

Table of Contents

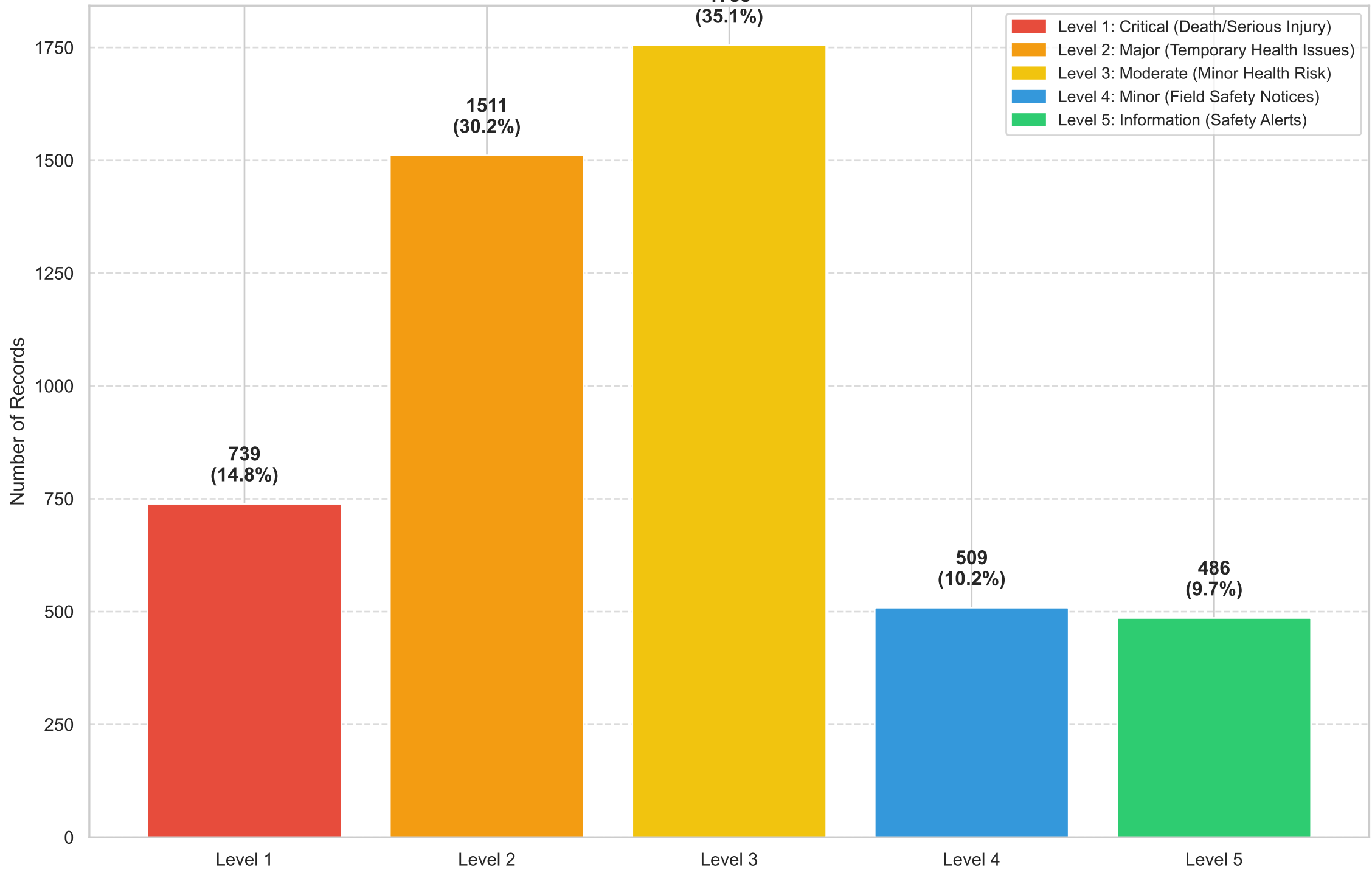
1. Dataset Overview
2. Class Distribution
3. Feature Analysis
4. Model Performance
5. Validation Results
6. Feature Importance
7. Confusion Matrices
8. Per-Class Metrics
9. Challenges & Limitations
10. Next Steps & Recommendations

1. Dataset Overview

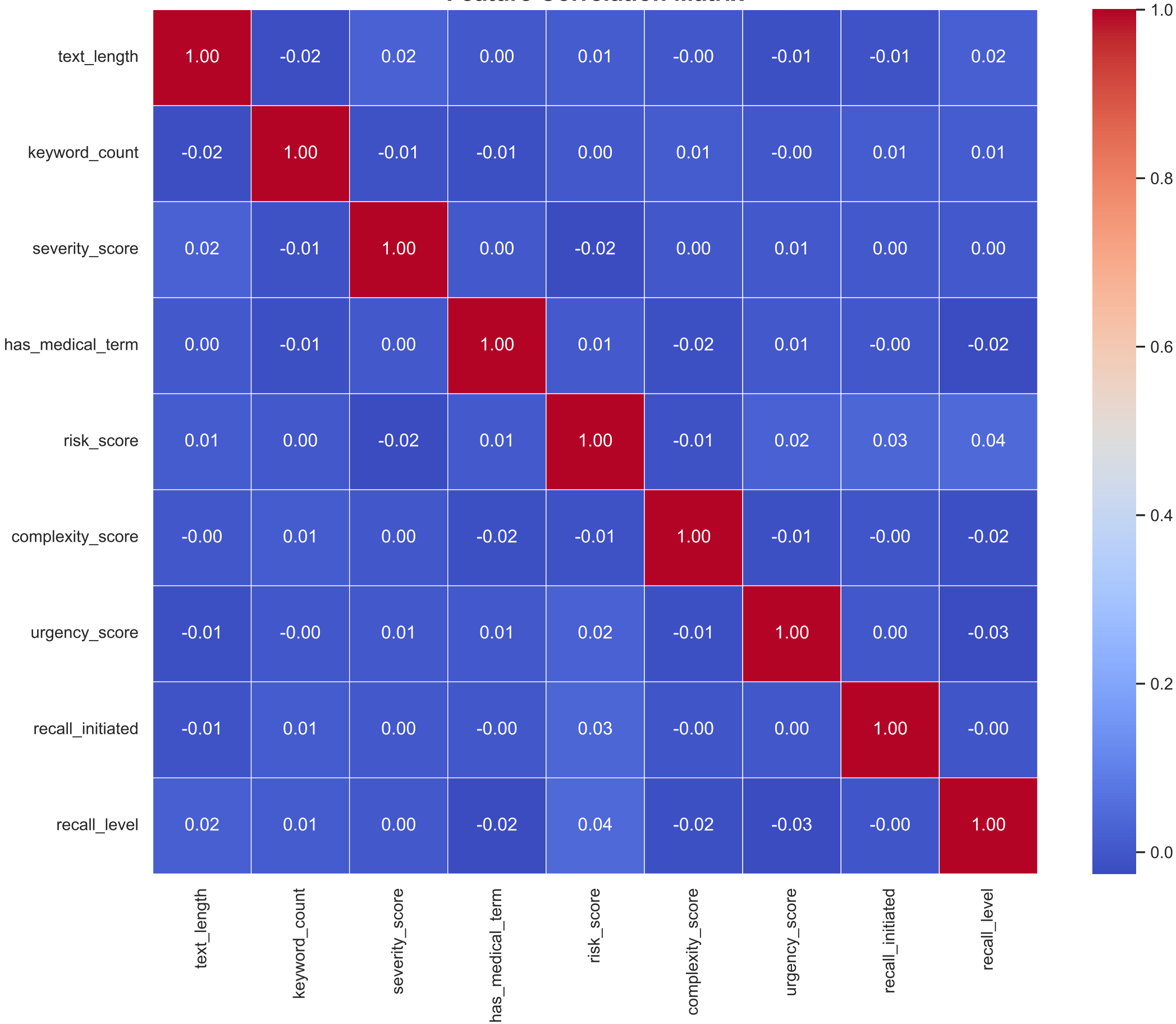
Total Records: 5000
Features: 8
Classification System: 5-Level Recall Severity
Training Set: 4000 records (80.0%)
Validation Set: 1000 records (20.0%)
Stratified Sampling: Yes

Feature	Description	Type
Preprocessing: Feature Standardization		
text_length	Length of recall description	Numeric
keyword_count	Medical keywords detected	Numeric
severity_score	Automated severity calculation	Numeric
has_medical_term	Contains medical terminology	Binary
risk_score	Calculated health risk score	Numeric
complexity_score	Technical complexity	Numeric
urgency_score	Response urgency indicator	Numeric
recall_initiated	Manufacturer-initiated recall	Binary

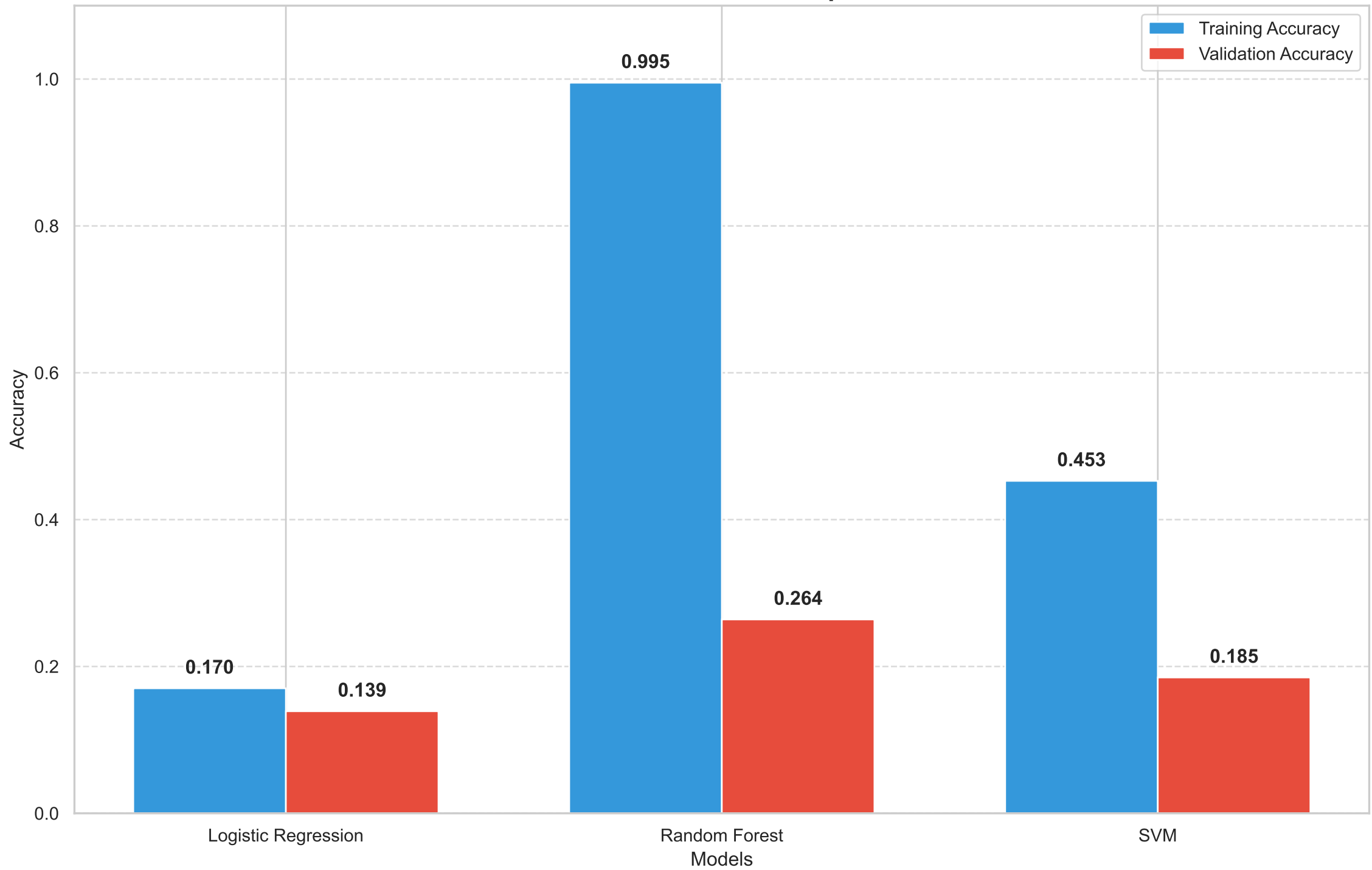
Class Distribution: 5-Level Recall Classification



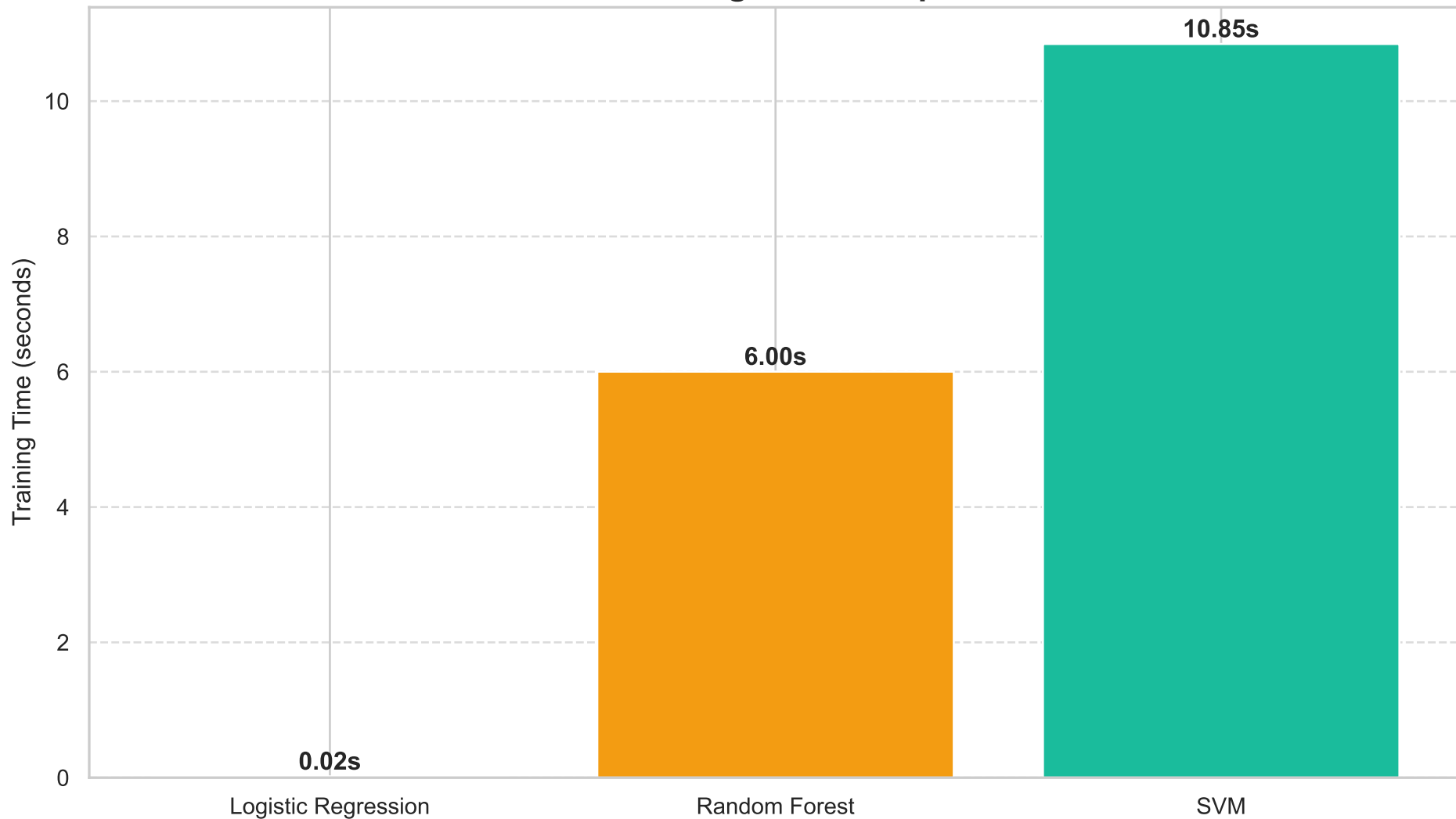
Feature Correlation Matrix



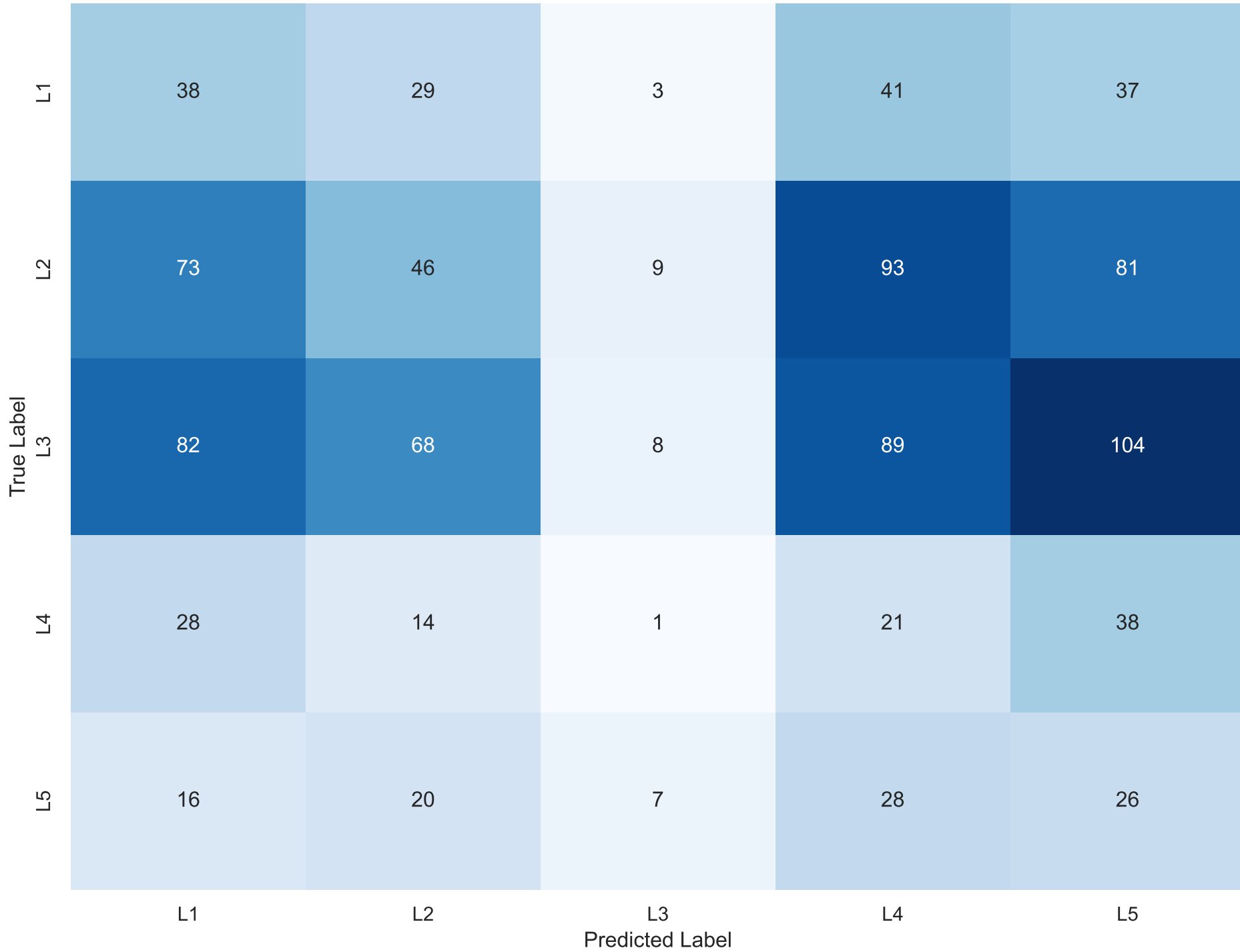
Model Performance Comparison



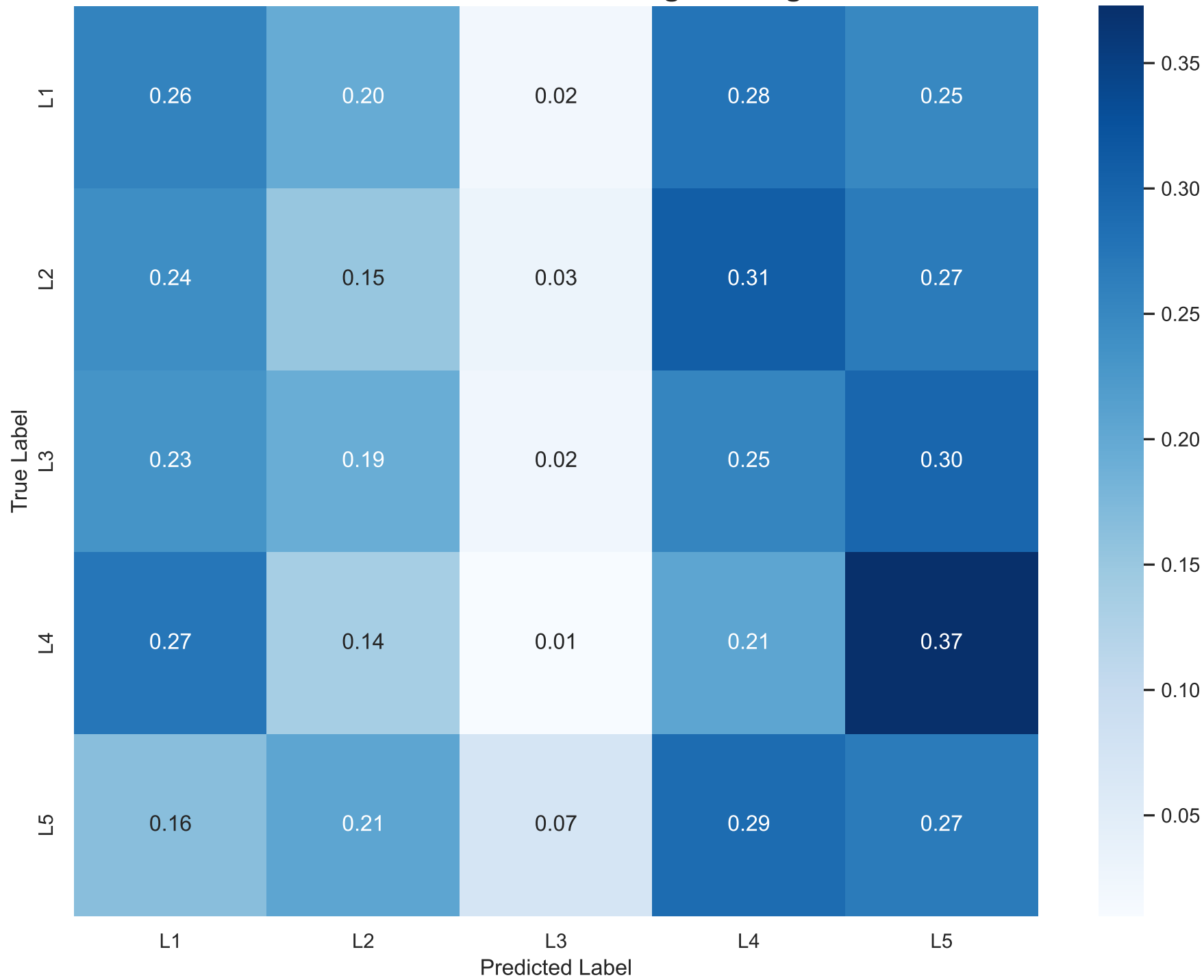
Model Training Time Comparison



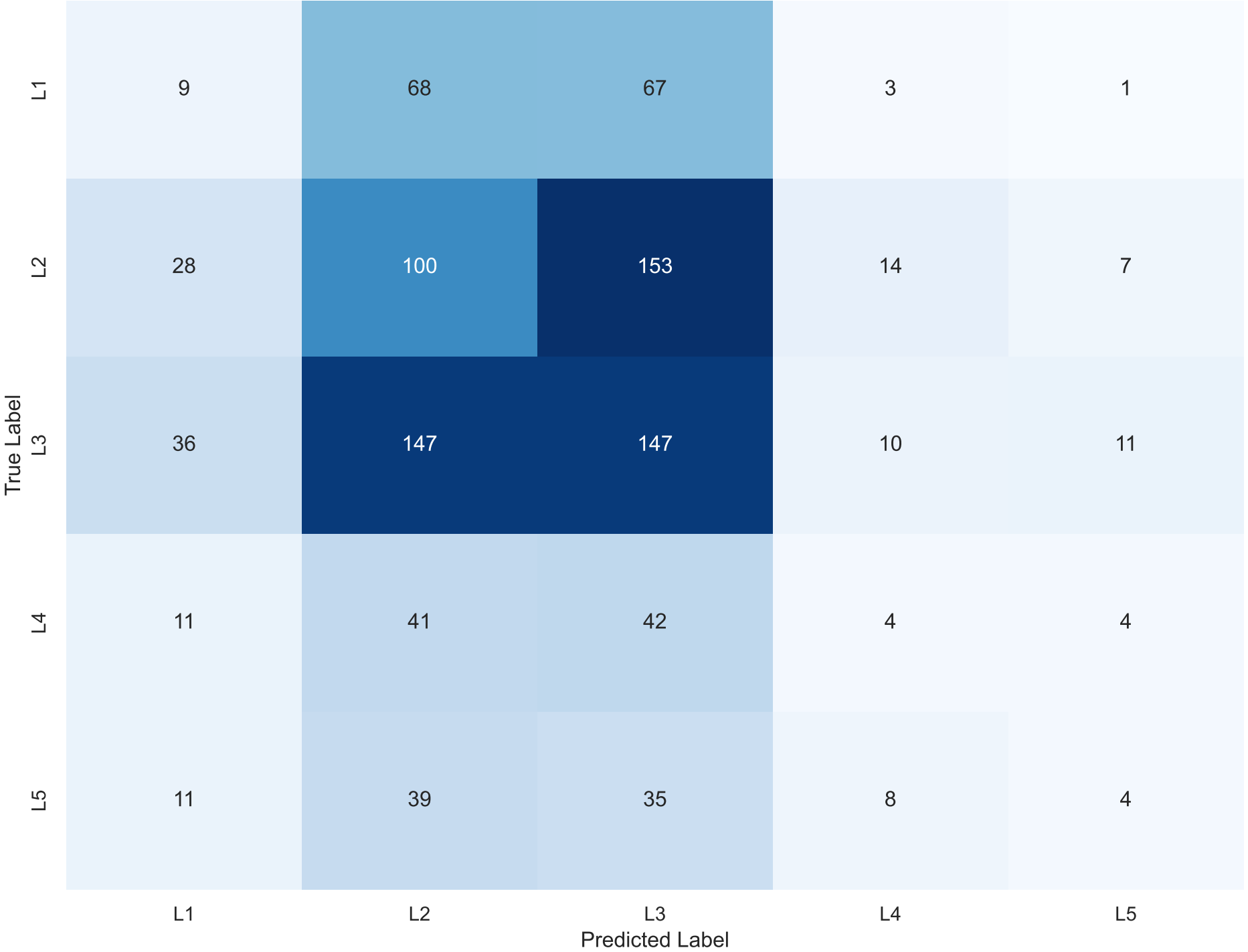
Confusion Matrix: Logistic Regression



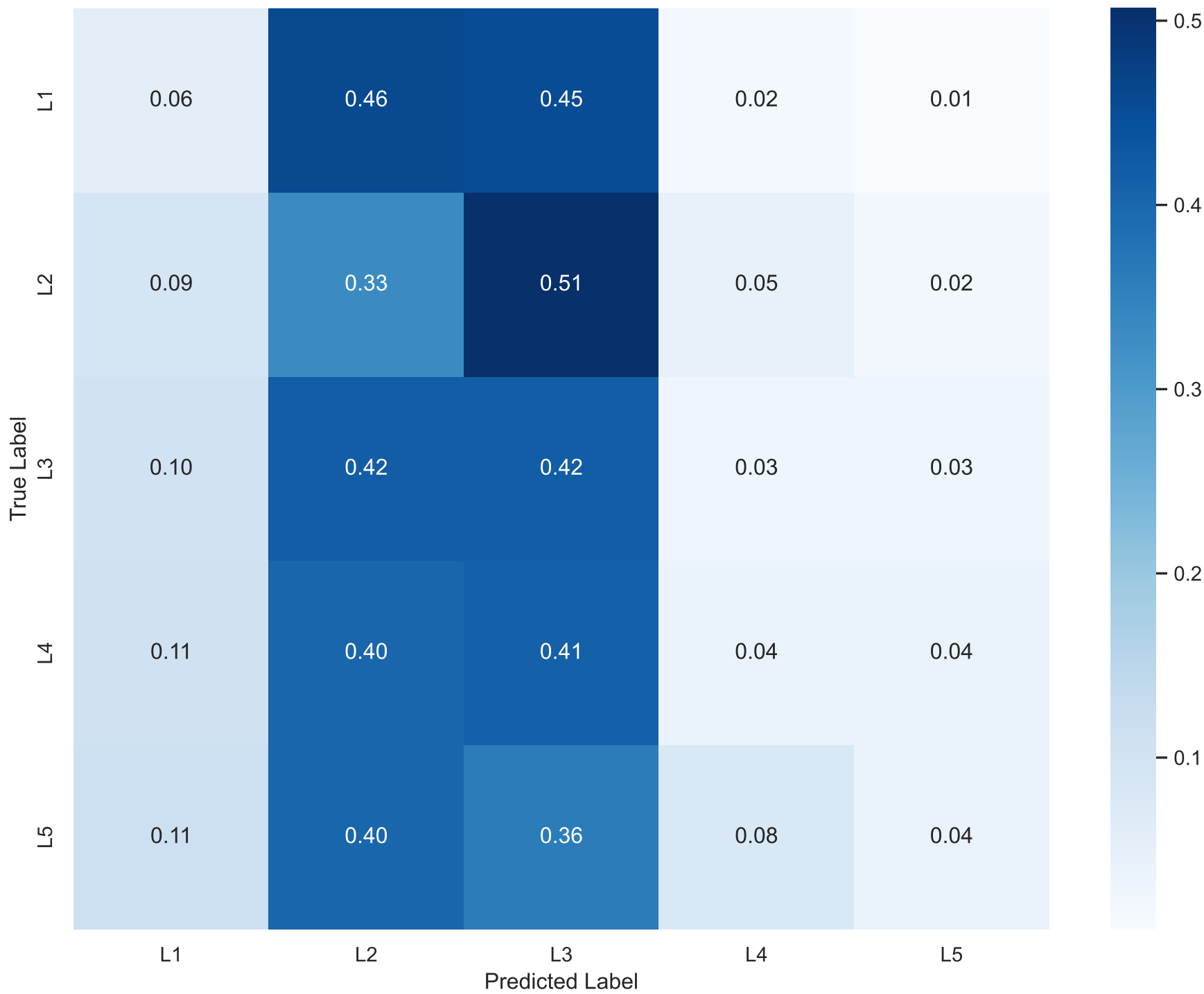
Normalized Confusion Matrix: Logistic Regression



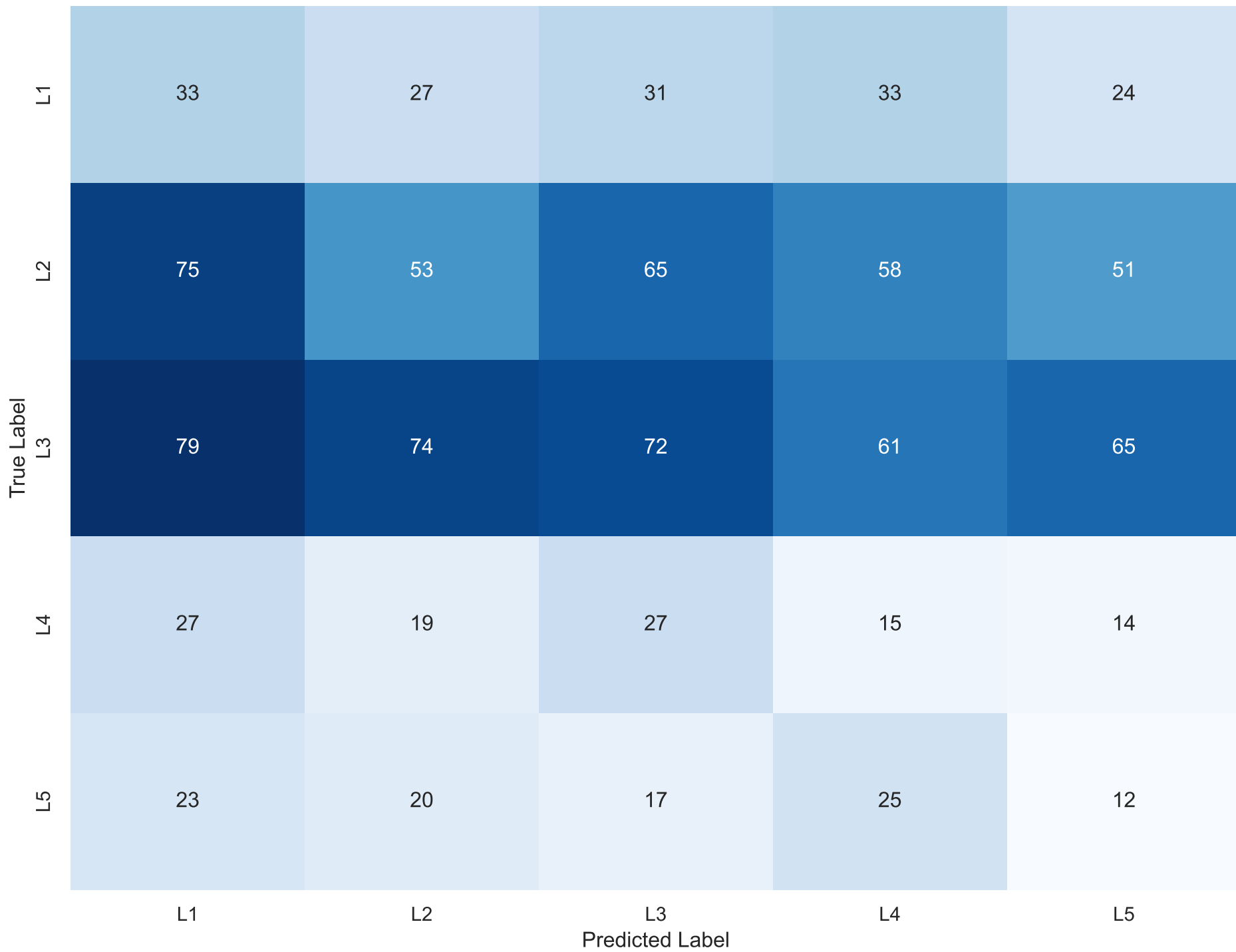
Confusion Matrix: Random Forest



Normalized Confusion Matrix: Random Forest



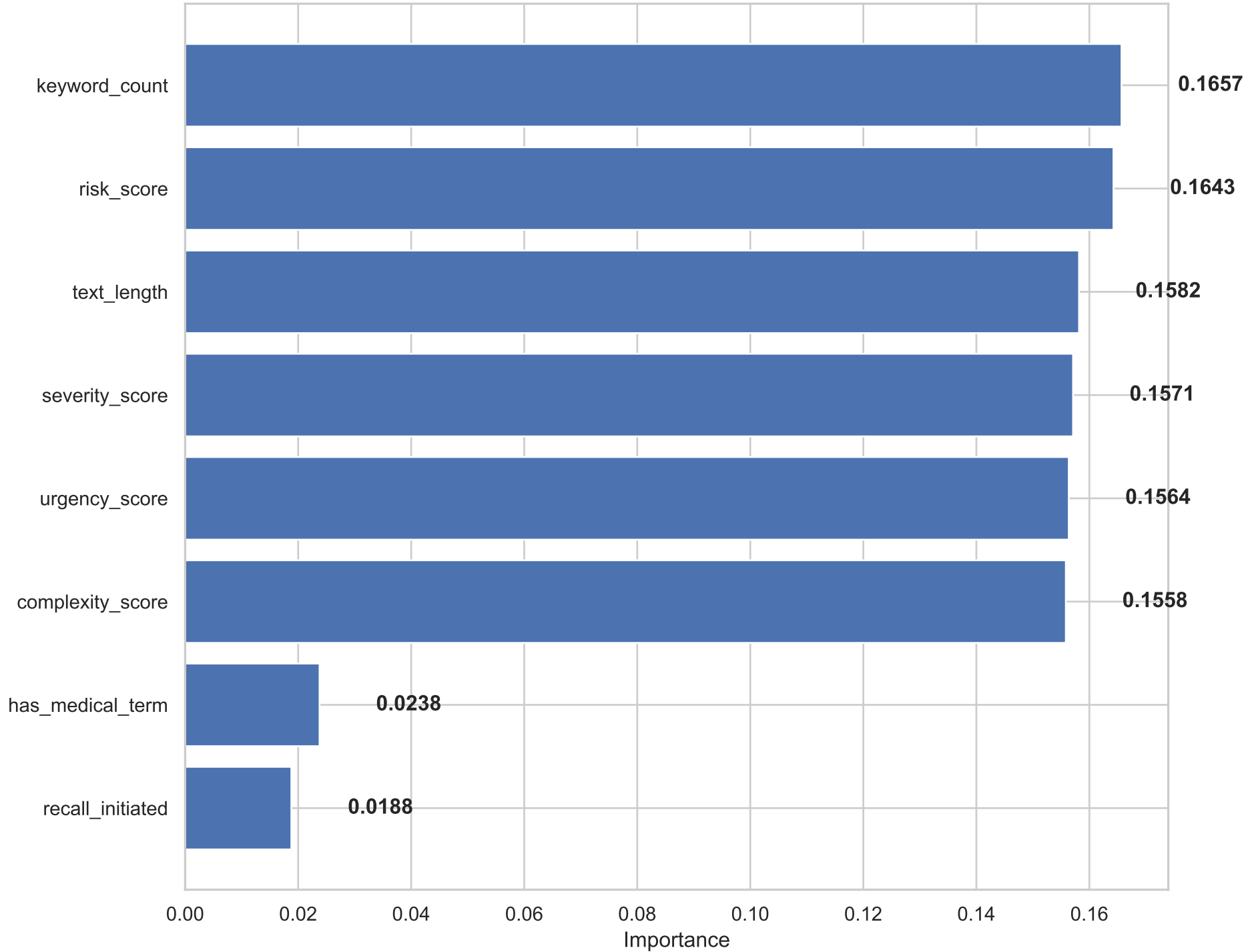
Confusion Matrix: SVM



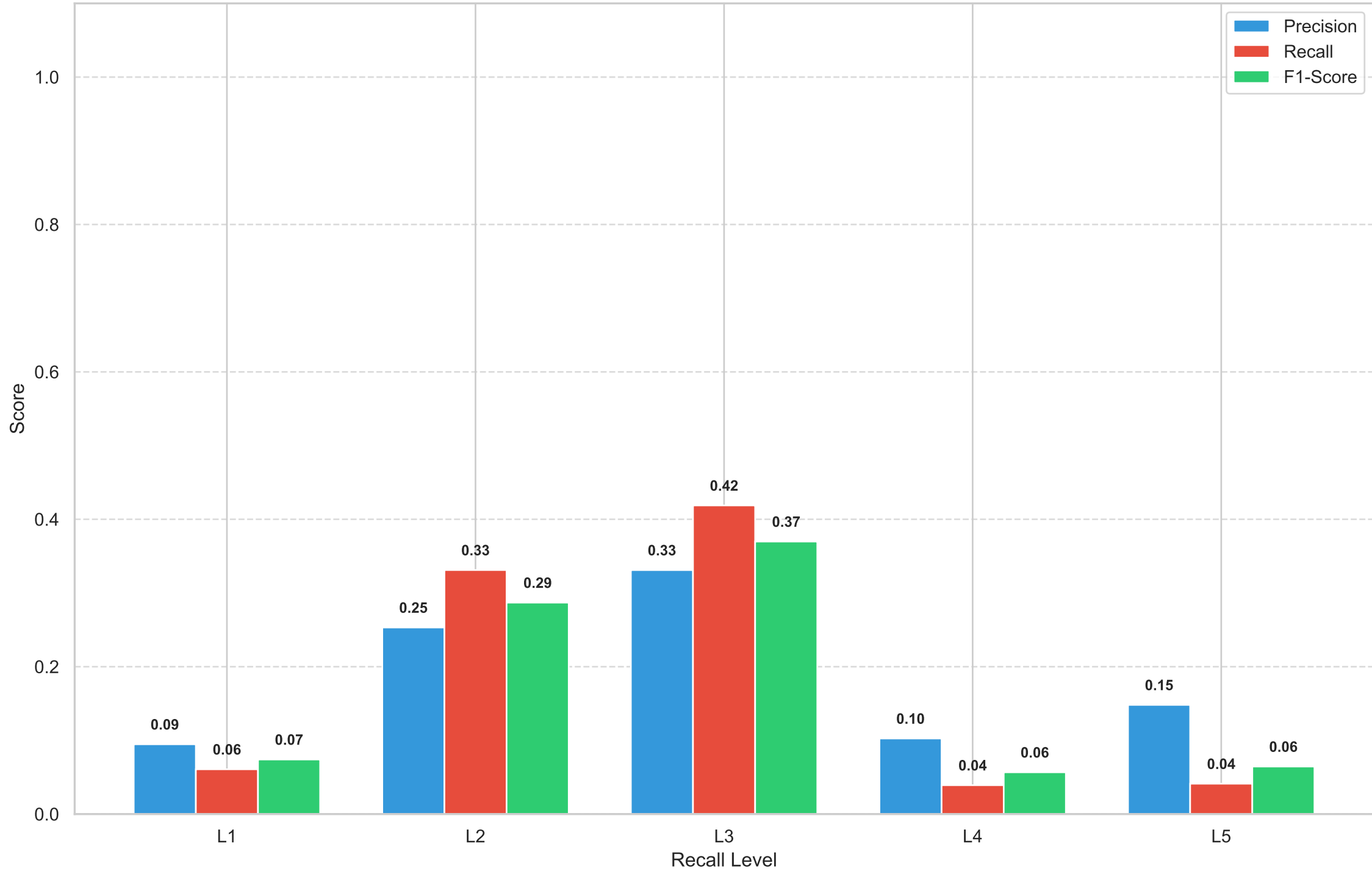
Normalized Confusion Matrix: SVM



Feature Importance (Random Forest)



Performance Metrics by Class (Random Forest)



Challenges & Limitations

1. Class Imbalance:

- Unequal distribution of samples across recall levels affects model training
- Balanced class weights were used to mitigate this issue

2. Feature Interpretability:

- Some features have complex relationships that are difficult to interpret
- Trade-off between model complexity and interpretability

3. Validation Limitations:

- No external validation dataset was available
- Cross-validation provides an estimate of performance, but real-world testing is needed

4. Multi-Class Complexity:

- 5-level classification is inherently more difficult than binary classification
- Some classes show confusion with adjacent levels (e.g., Level 2 vs Level 3)

5. Feature Engineering Opportunities:

- Current features could be enhanced with domain-specific medical knowledge
- Text-based features from recall descriptions are not fully utilized

Next Steps & Recommendations

1. Model Improvements:

- Hyperparameter optimization using grid search or Bayesian methods
- Ensemble methods combining multiple models for better performance
- Explore transformer-based models for text processing

2. Feature Enhancement:

- Incorporate text embedding techniques for recall descriptions
- Add domain-specific medical features based on expert knowledge
- Explore manufacturer history and reputation as features

3. Validation Protocol:

- Implement expert validation for critical classifications (Level 1)
- Create feedback loop for model improvement based on expert corrections
- Test on temporally separated data to evaluate real-world performance

4. Deployment Strategy:

- Develop confidence scoring mechanism for classifications
- Implement model monitoring for performance drift
- Create interpretability tools for regulatory stakeholders

5. Periodic Retraining:

- Schedule regular model retraining with new recall data
- Monitor feature distribution changes over time
- Update feature engineering pipeline as needed

Model Performance Summary

Metric	Logistic Regression	Random Forest	SVM
Training Accuracy	0.1703	0.9950	0.4527
Validation Accuracy	0.1390	0.2640	0.1850
Training Time (s)	0.02	6.00	10.85
Level 1 Recall	0.2568	0.0608	0.2230
Level 2 Recall	0.1523	0.3311	0.1755
Level 3 Recall	0.0228	0.4188	0.2051
Level 4 Recall	0.2059	0.0392	0.1471
Level 5 Recall	0.2680	0.0412	0.1237
Macro F1-Score	0.1360	0.1704	0.1669

End of Report