Medical Device Recall Classification

Implementation Report

5-Level Classification System

Generated on August 25, 2025

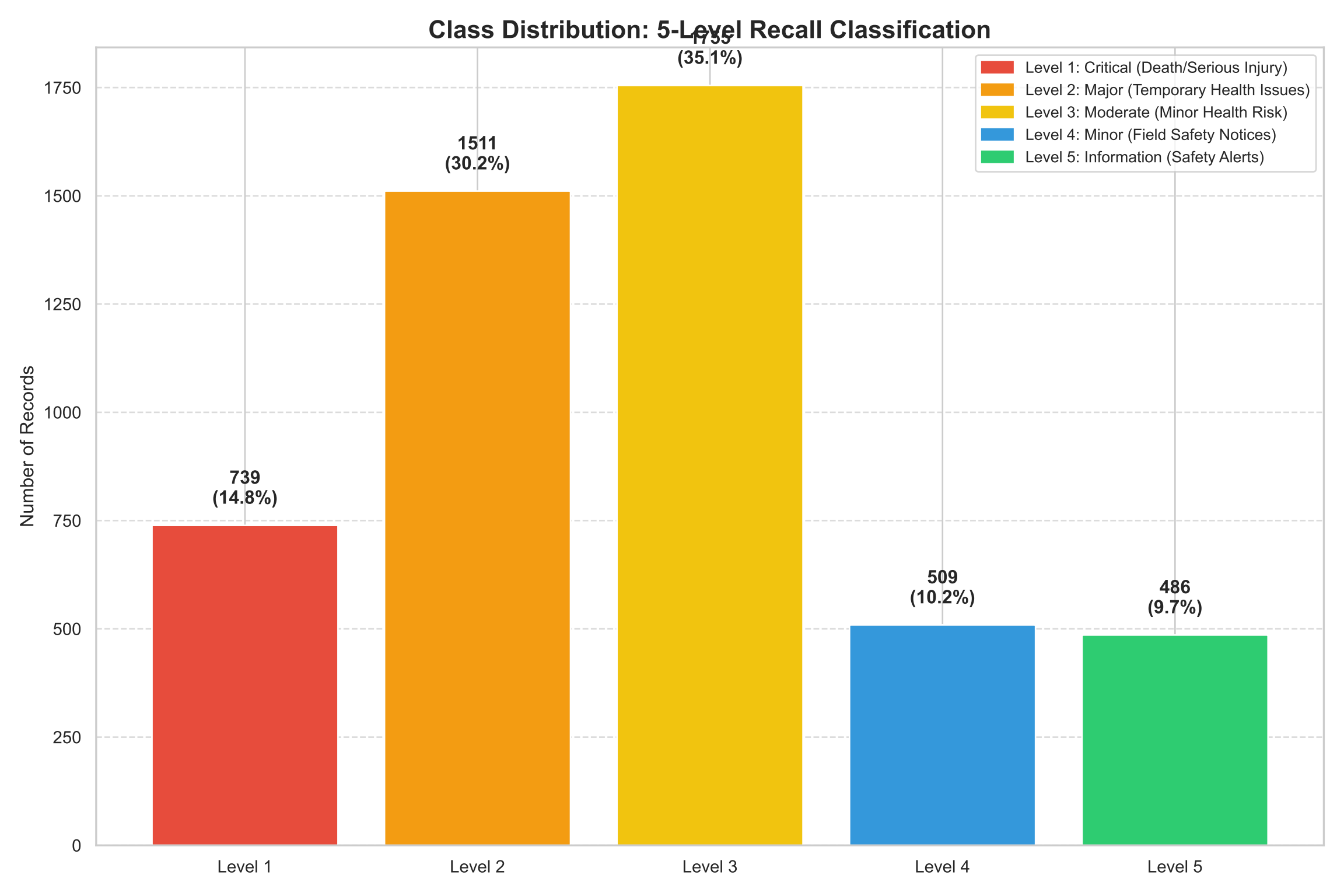
Data Science Team

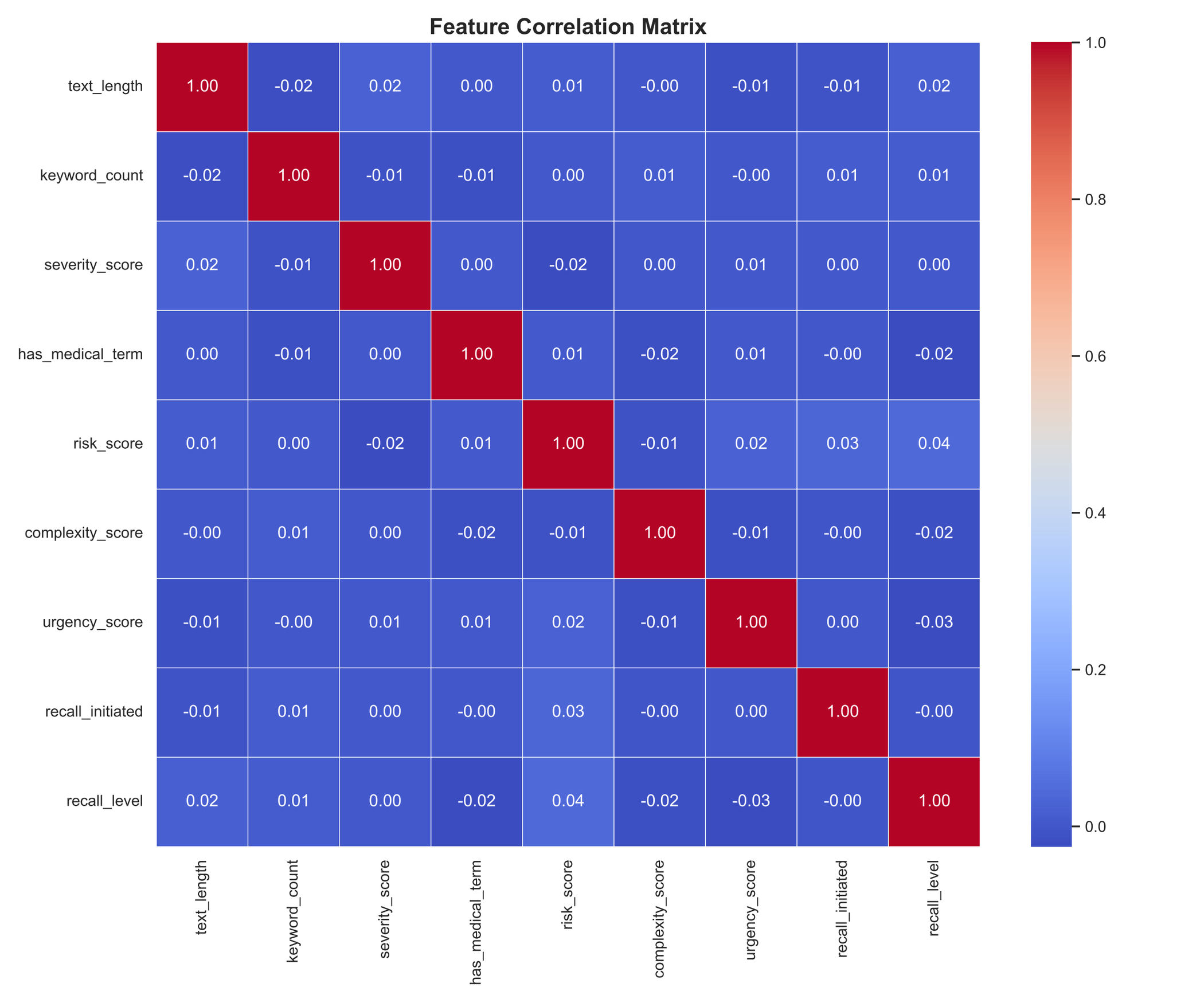
Table of Contents

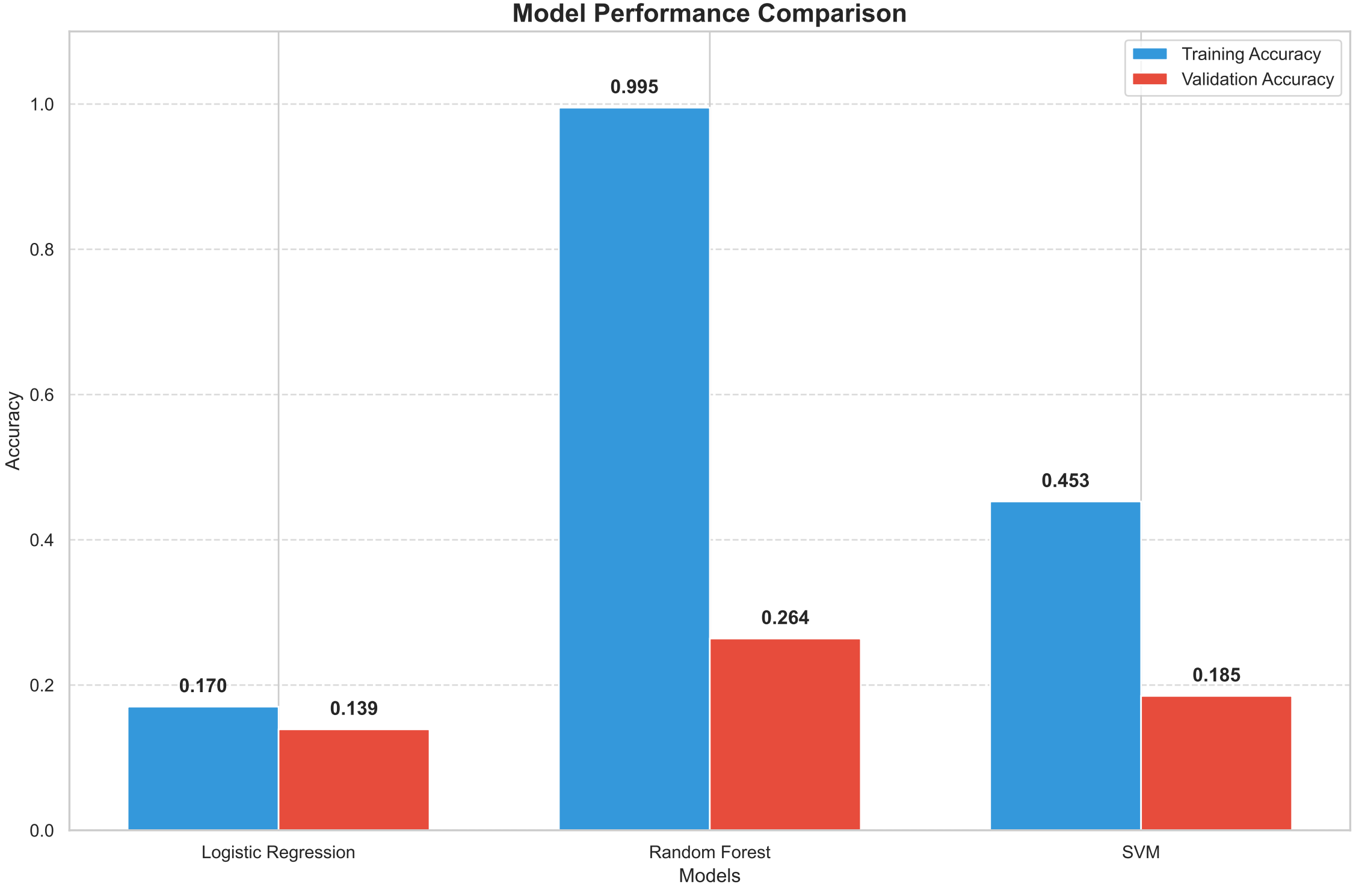
1. Dataset Overview

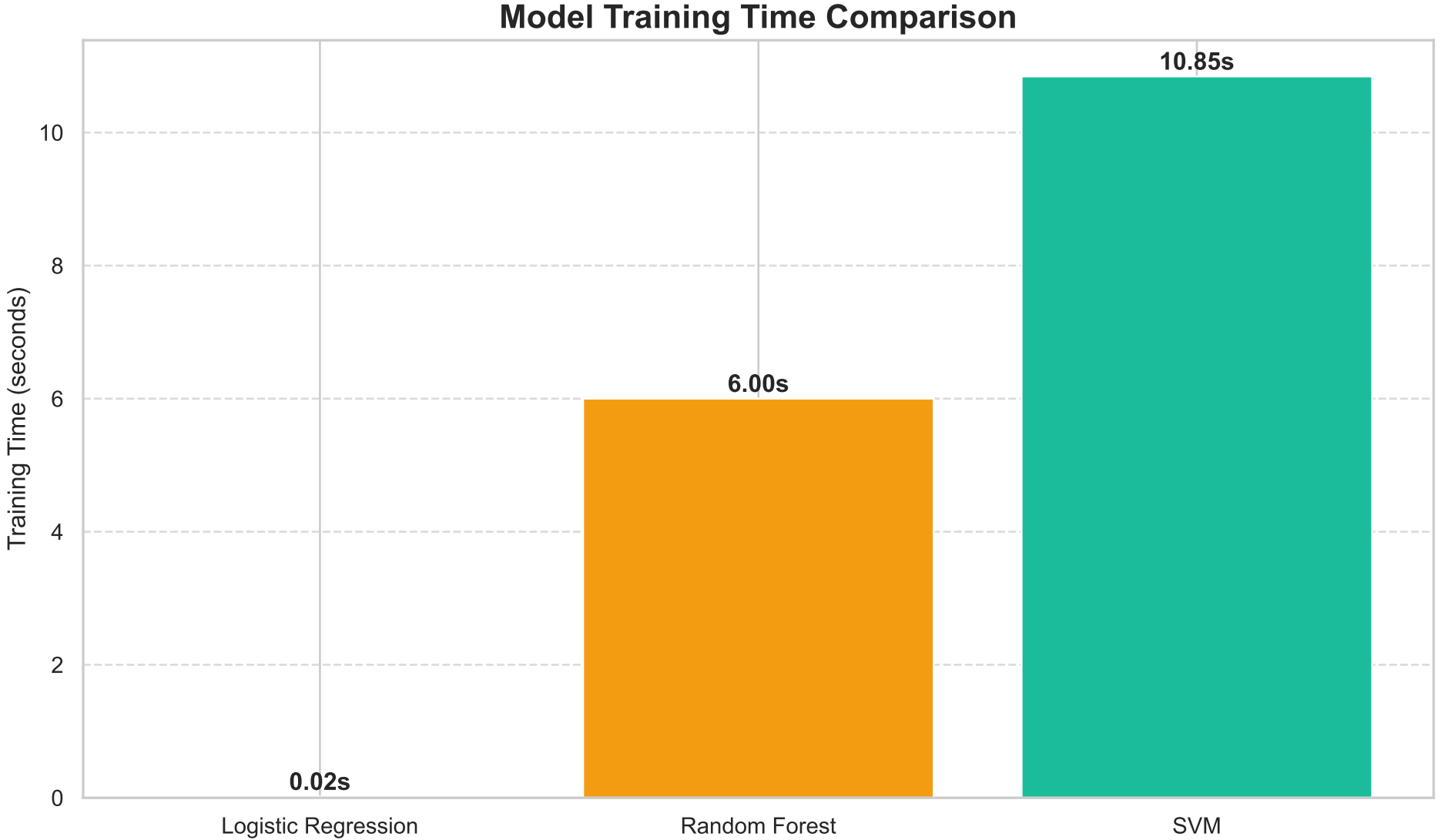
Total Records: 5000
Features: 8
Classification System: 5-Level Recall Severity
Training Set: 4000 records (80.0%)
Validation Set: 1000 records (20.0%)
Stratified Sampling: Yes
Preprocessing: Feature Standardization

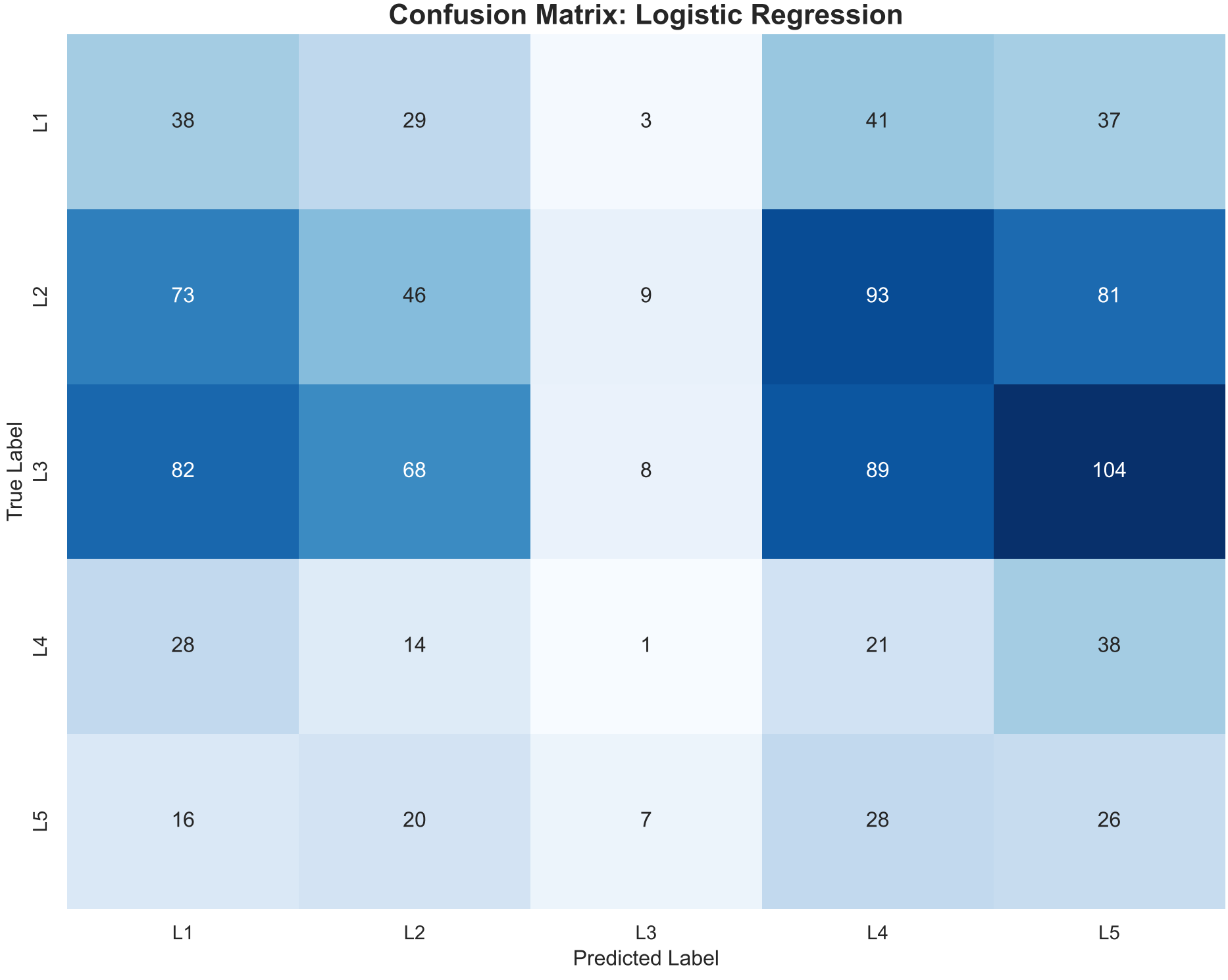
Feature	Description	Type
text_length	Length of recall description	Numeric
keyword_count	Medical keywords detected	Numeric
severity_score	Automated severity calculation	Numeric
has_medical_term	Contains medical terminology	Binary
risk_score	Calculated health risk score	Numeric
complexity_score	Technical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

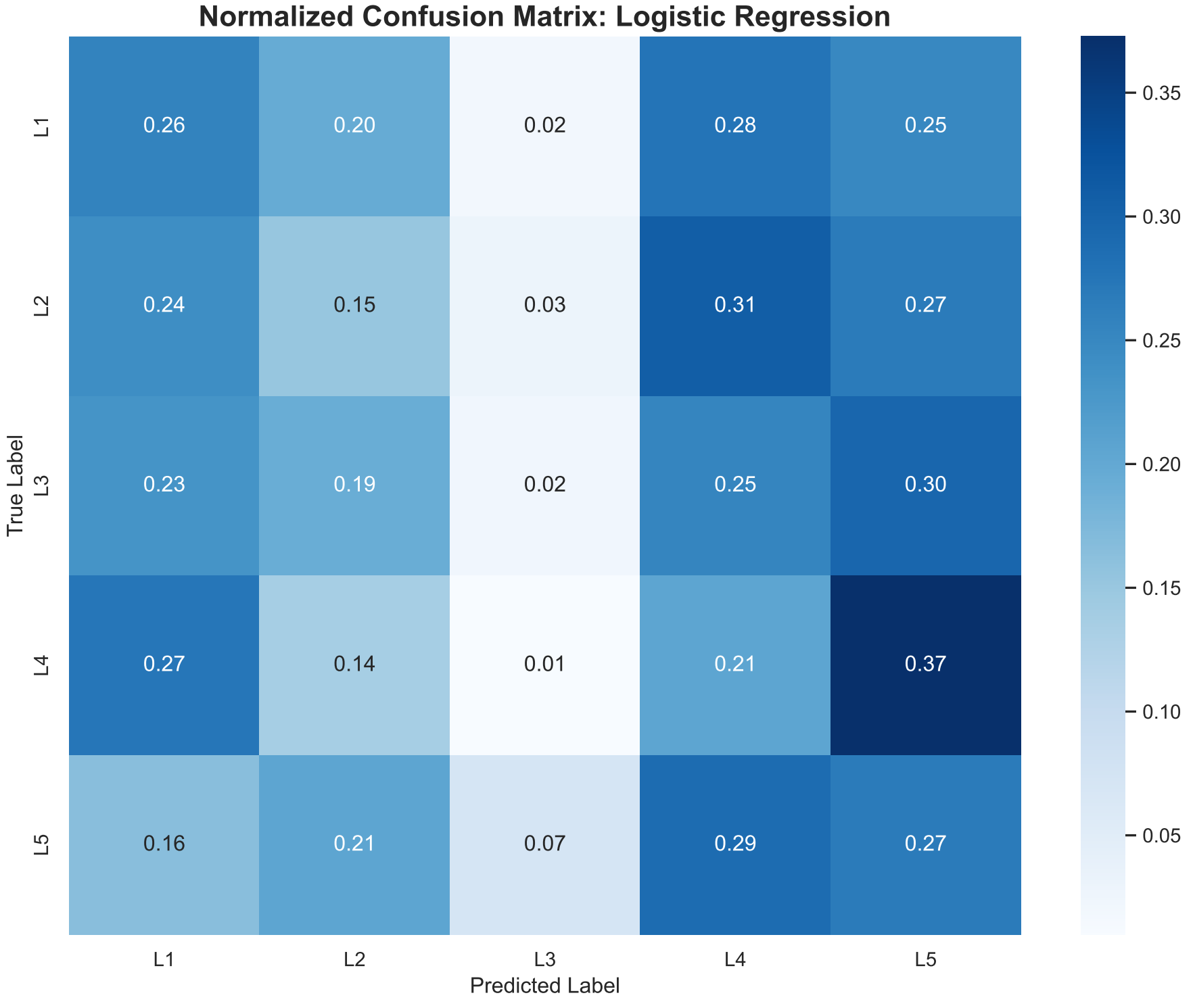

Class Distribution: 5-Level Recall Classification
Level 1: Critical (Death/Serious Injury)
Level 2: Major (Temporary Health Issues)
Level 3: Moderate (Minor Health Risk)
Level 4: Minor (Field Safety Notices)
Level 5: Information (Safety Alerts)
Number of Records
0
250
500
750
1000
1250
1500
1750
739
(14.8%)
1511
(30.2%)
1753
(35.1%)
509
(10.2%)
486
(9.7%)
Level 1
Level 2
Level 3
Level 4
Level 5

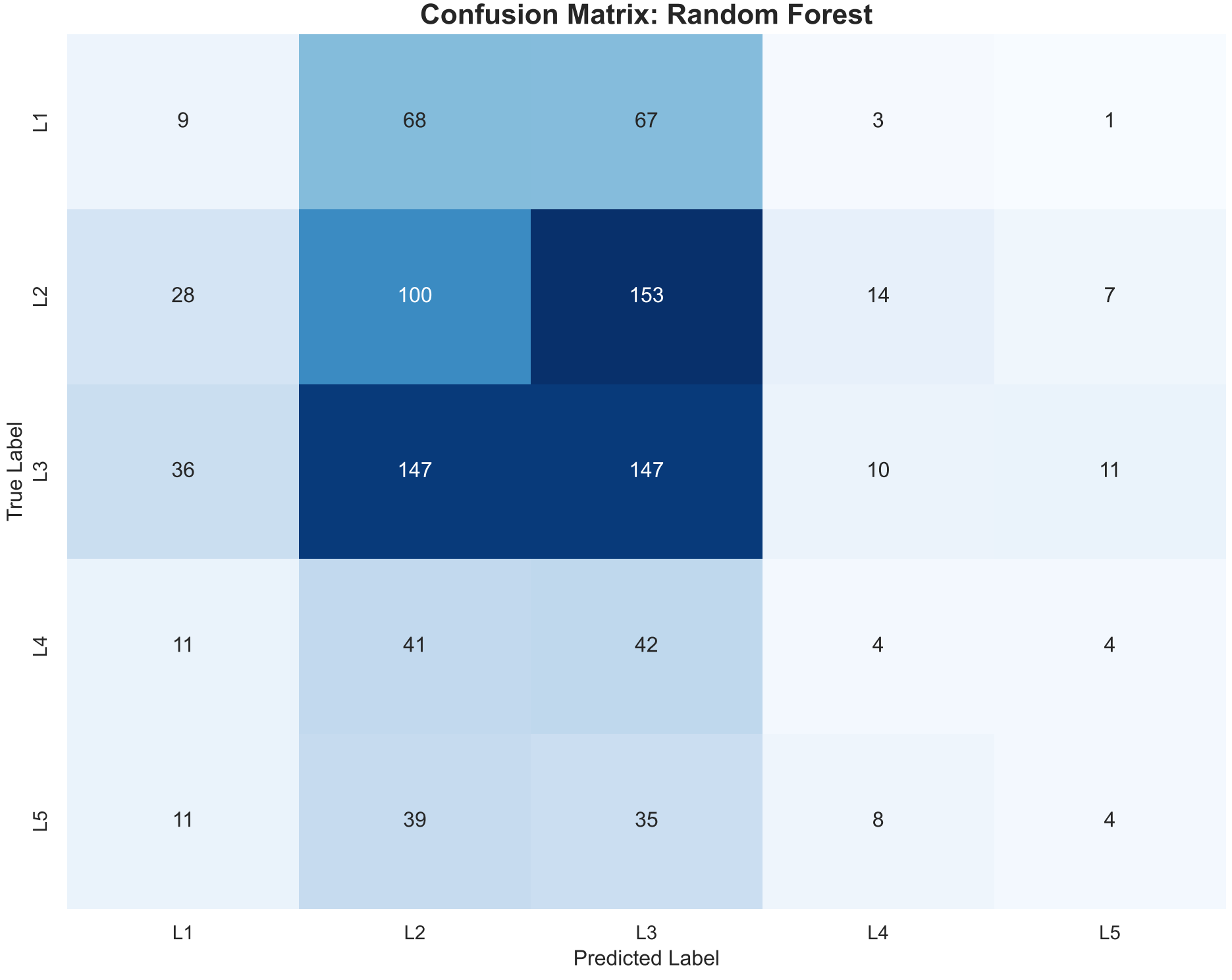

Feature Correlation Matrix
text_length
keyword_count
severity_score
has_medical_term
risk_score
complexity_score
urgency_score
recall_initiated
recall_level
1.00 -0.02 0.02 0.00 0.01 -0.00 -0.01 -0.01 0.02
-0.02 1.00 -0.01 -0.01 0.00 0.01 -0.00 0.01 0.01
0.02 -0.01 1.00 0.00 -0.02 0.00 0.01 0.00 0.00
0.00 -0.01 0.00 1.00 0.01 -0.02 0.01 -0.00 -0.02
0.01 0.00 -0.02 0.01 1.00 -0.01 0.02 0.03 0.04
-0.00 0.01 0.00 -0.02 -0.01 1.00 -0.01 -0.00 -0.02
-0.01 -0.00 0.01 0.01 0.02 -0.01 1.00 0.00 -0.03
-0.01 0.01 0.00 -0.00 0.03 -0.00 0.00 1.00 -0.00
0.02 0.01 0.00 -0.02 0.04 -0.02 -0.03 -0.00 1.00
1.0
0.8
0.6
0.4
0.2
0.0

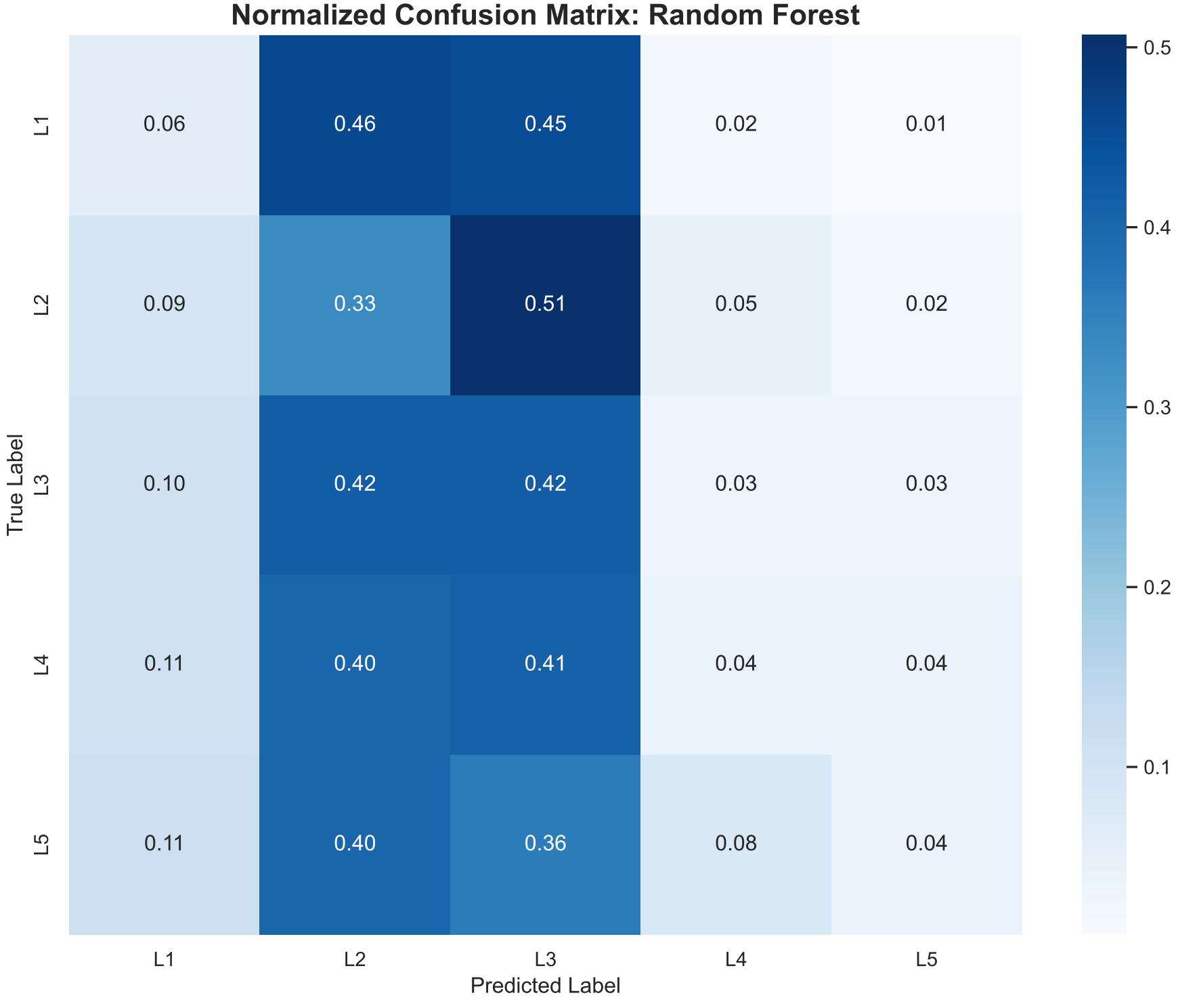

Model Performance Comparison
Training Accuracy
Validation Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
Accuracy
0.170
0.139
0.995
0.264
0.453
0.185
Logistic Regression
Random Forest
SVM
Models

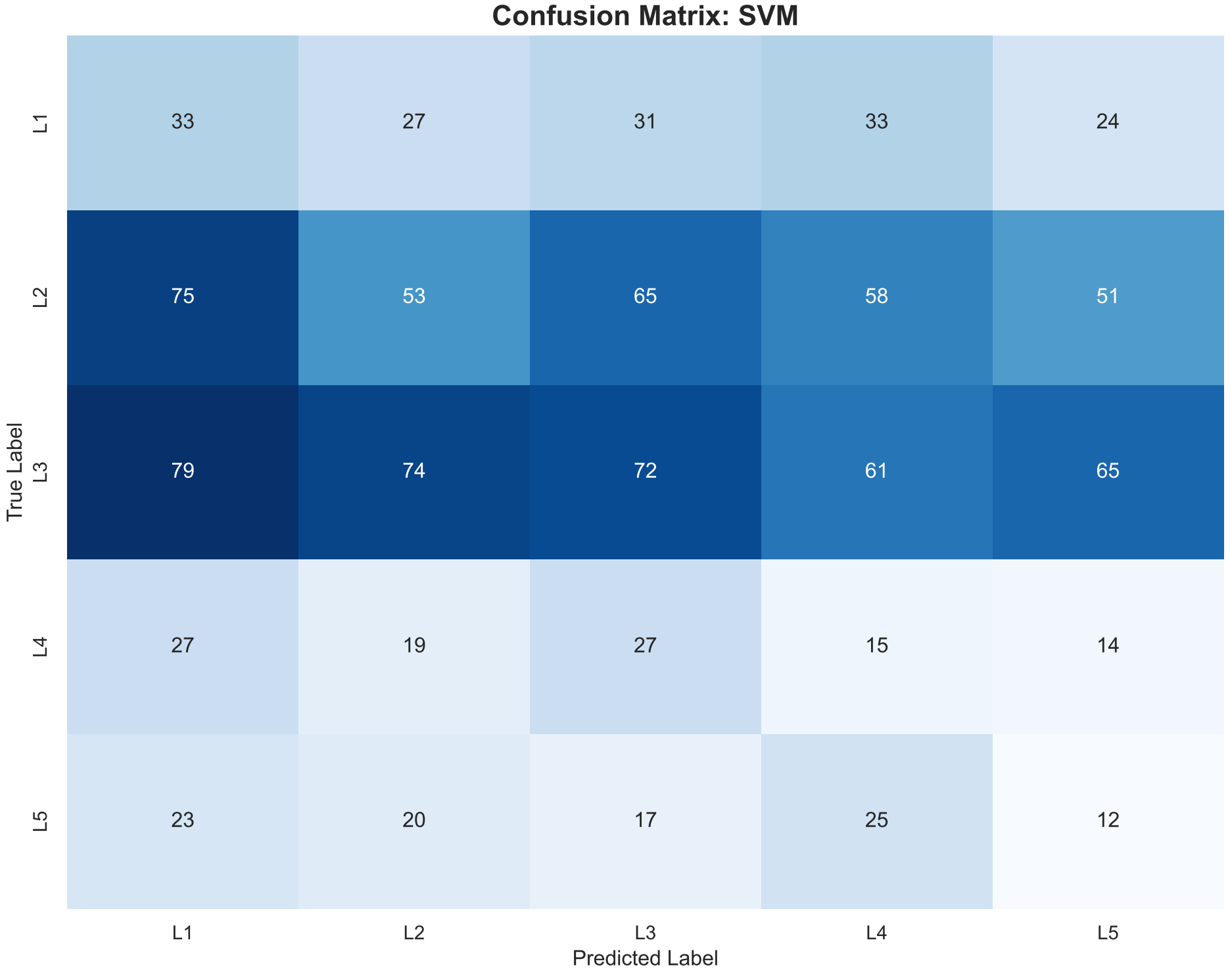

Model Training Time Comparison
Training Time (seconds)
0
2
4
6
8
10
0.02s
6.00s
10.85s
Logistic Regression
Random Forest
SVM

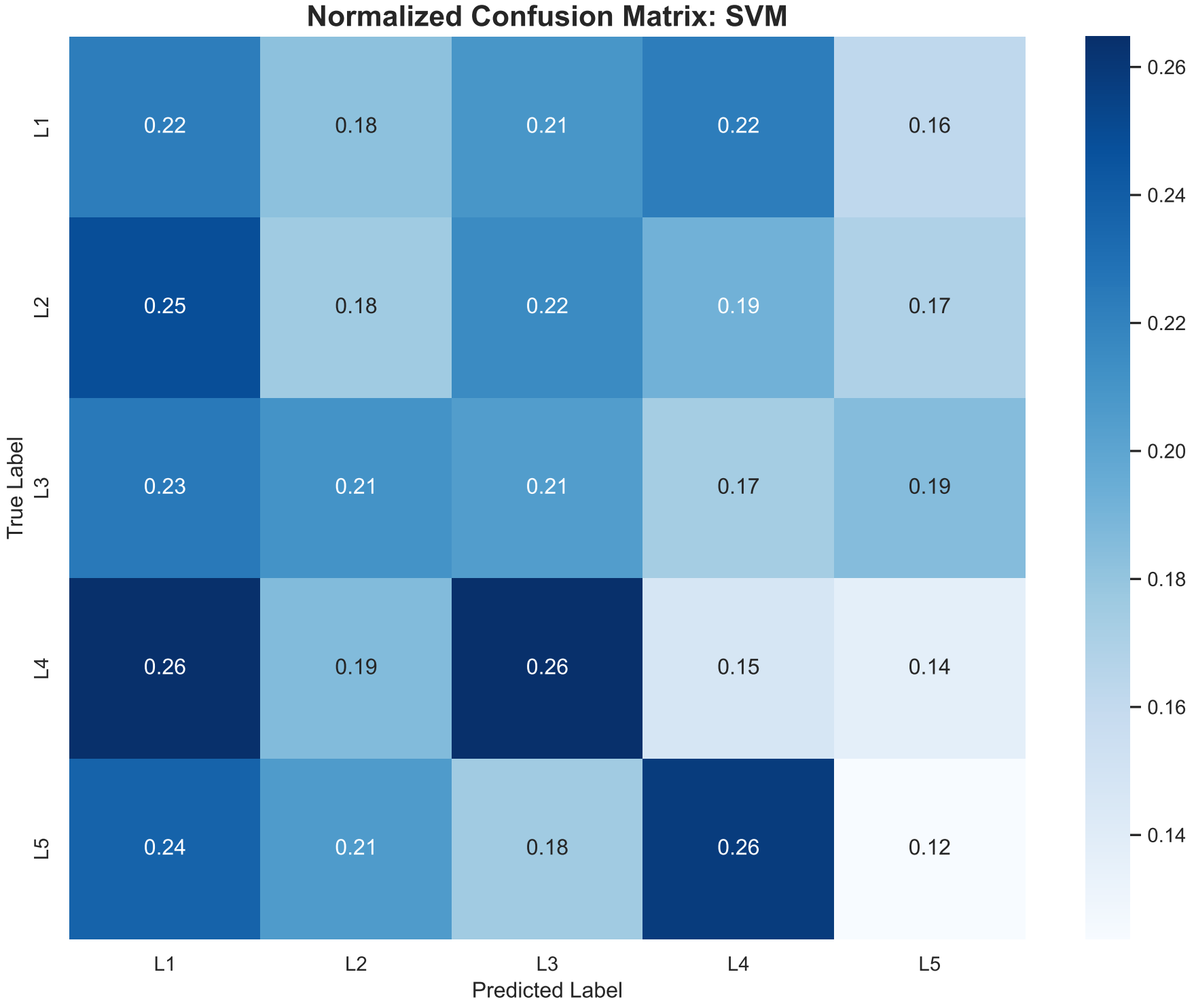

Confusion Matrix: Logistic Regression
True Label
L1
L2
L3
L4
L5
38
29
3
41
37
73
46
9
93
81
82
68
8
89
104
28
14
1
21
38
16
20
7
28
26
L1
L2
L3
L4
L5
Predicted Label

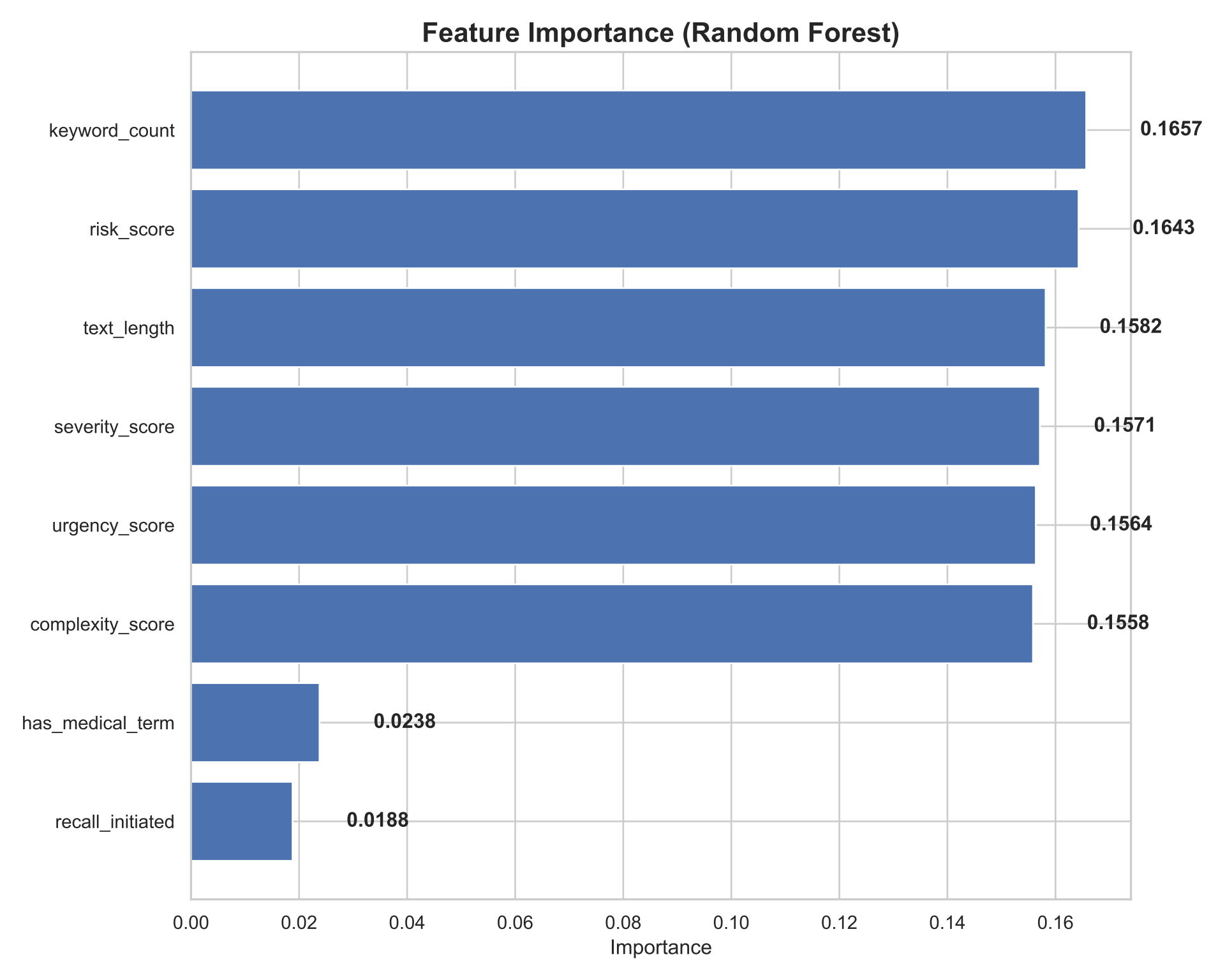

Normalized Confusion Matrix: Logistic Regression
True Label
L1
L2
L3
L4
L5
0.26
0.20
0.02
0.28
0.25
0.24
0.15
0.03
0.31
0.27
0.23
0.19
0.02
0.25
0.30
0.27
0.14
0.01
0.21
0.37
0.16
0.21
0.07
0.29
0.27
L1
L2
L3
L4
L5
Predicted Label
0.35
0.30
0.25
0.20
0.15
0.10
0.05

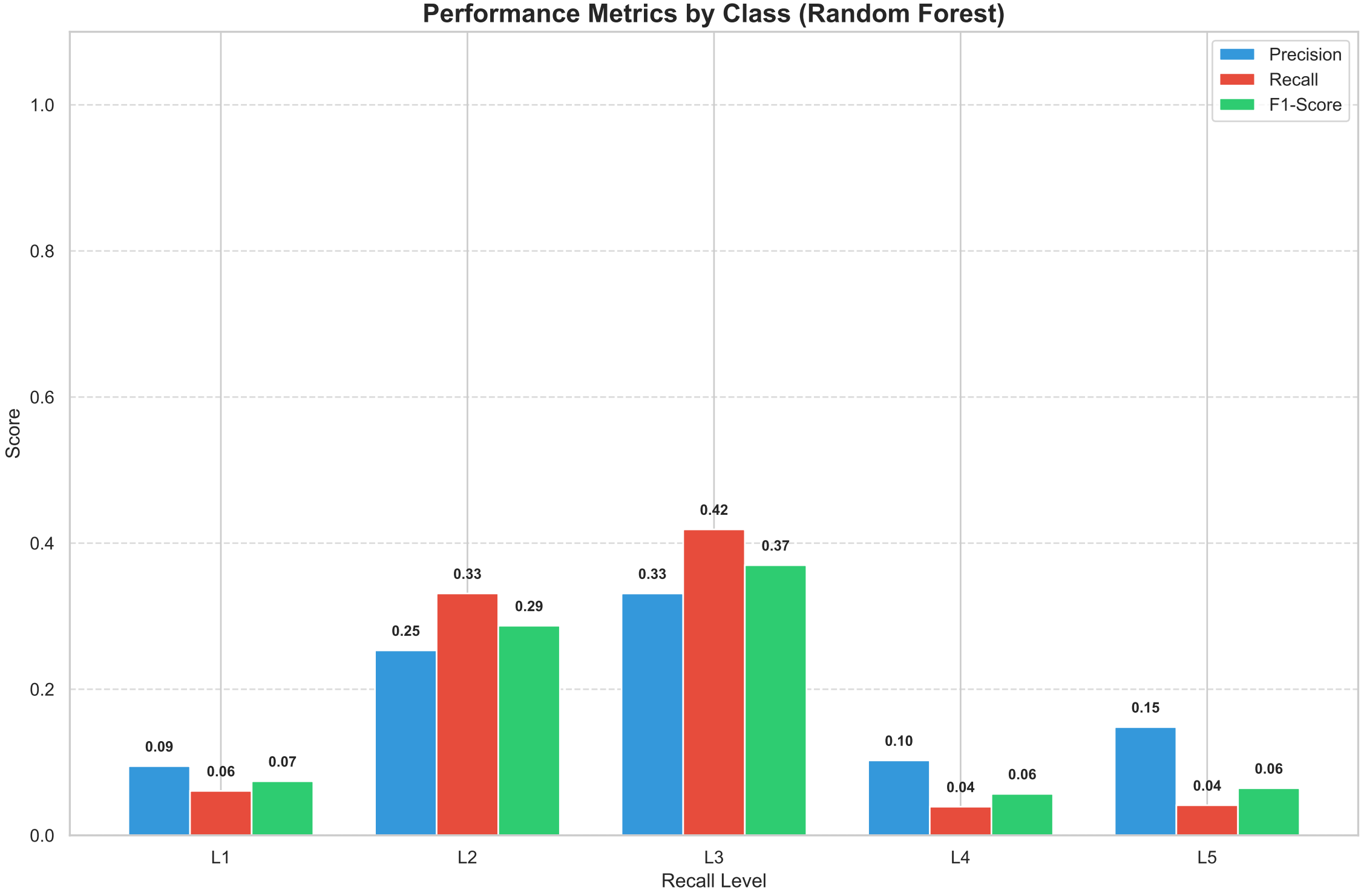

Confusion Matrix: Random Forest
True Label
L1
L2
L3
L4
L5
9
68
67
3
1
28
100
153
14
7
36
147
147
10
11
11
41
42
4
4
11
39
35
8
4
L1
L2
L3
L4
L5
Predicted Label


Normalized Confusion Matrix: Random Forest
True Label
L1
L2
L3
L4
L5
0.06
0.46
0.45
0.02
0.01
0.09
0.33
0.51
0.05
0.02
0.10
0.42
0.42
0.03
0.03
0.11
0.40
0.41
0.04
0.04
0.11
0.40
0.36
0.08
0.04
L1
L2
L3
L4
L5
Predicted Label
0.5
0.4
0.3
0.2
0.1


Confusion Matrix: SVM
True Label
L1
L2
L3
L4
L5
33
27
31
33
24
75
53
65
58
51
79
74
72
61
65
27
19
27
15
14
23
20
17
25
12
L1
L2
L3
L4
L5
Predicted Label


Normalized Confusion Matrix: SVM
True Label
L1
L2
L3
L4
L5
0.22 0.18 0.21 0.22 0.16
0.25 0.18 0.22 0.19 0.17
0.23 0.21 0.21 0.17 0.19
0.26 0.19 0.26 0.15 0.14
0.24 0.21 0.18 0.26 0.12
L1
L2
L3
L4
L5
Predicted Label
0.26
0.24
0.22
0.20
0.18
0.16
0.14


Feature Importance (Random Forest)
keyword_count
risk_score
text_length
severity_score
urgency_score
complexity_score
has_medical_term
recall_initiated
0.1657
0.1643
0.1582
0.1571
0.1564
0.1558
0.0238
0.0188
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
Importance


Performance Metrics by Class (Random Forest)
Precision
Recall
F1-Score
Score
1.0
0.8
0.6
0.4
0.2
0.0
0.09
0.06
0.07
0.25
0.33
0.29
0.33
0.42
0.37
0.10
0.04
0.06
0.15
0.04
0.06
L1
L2
L3
L4
L5
Recall Level

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