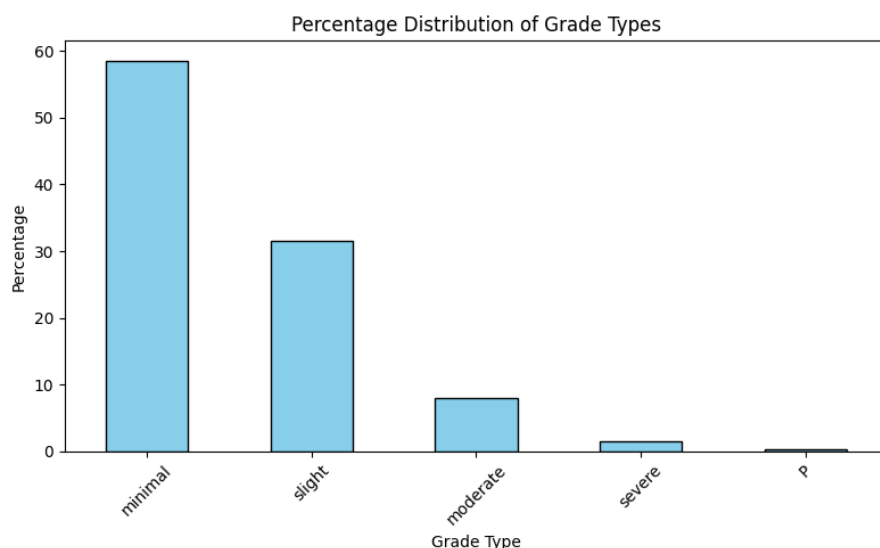# Results and Discussion

Original Data link : "TG-GATE"s (Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System)

1. Toxicity Level Distribution

2. Feature influencing on grade type

To prepare the dataset, extensive preprocessing was performed, including organ labeling, ID generation, and dataset merging. Feature selection was carried out by separating categorical and numerical variables, using Cramér's V for categorical correlation with toxicity grade, and Pearson correlation for numerical features. A total of 16 strongly correlated numerical variables and the most relevant categorical variable (COMPOUND_NAME) were retained.

I. Cramer's V scores for categorical variables
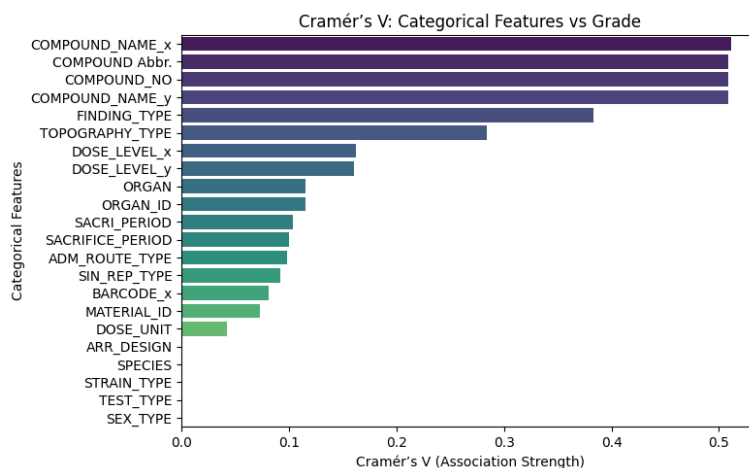
II. F score for numerical variables

```
        Feature        F_Score
28          PT(s)   149.980081
29        APTT(s)    95.644800
30     Fbg(mg/dL)    83.686835
39     CRE(mg/dL)    83.362465
12       LIVER(g)    72.404476
2      GROUP_ID_x    60.870791
0              ID    59.520018
7      GROUP_ID_y    57.260731
17        MCV(fL)    47.552675
31       ALP(IU/L)   43.677774
46     RALB(g/dL)    42.434339
47            A/G    36.965053
18        MCH(pg)    35.736505
38     BUN(mg/dL)    30.708144
13 KIDNEY_TOTAL(g)   30.061837
```

```
Selected features: ['PT(s)', 'APTT(s)', 'Fbg(mg/dL)', 'CRE(mg/dL)',
'LIVER(g)']
```

3. Feature Selection with model based ranking and correlation heat map

Correlation Heatmap of Numerical Features
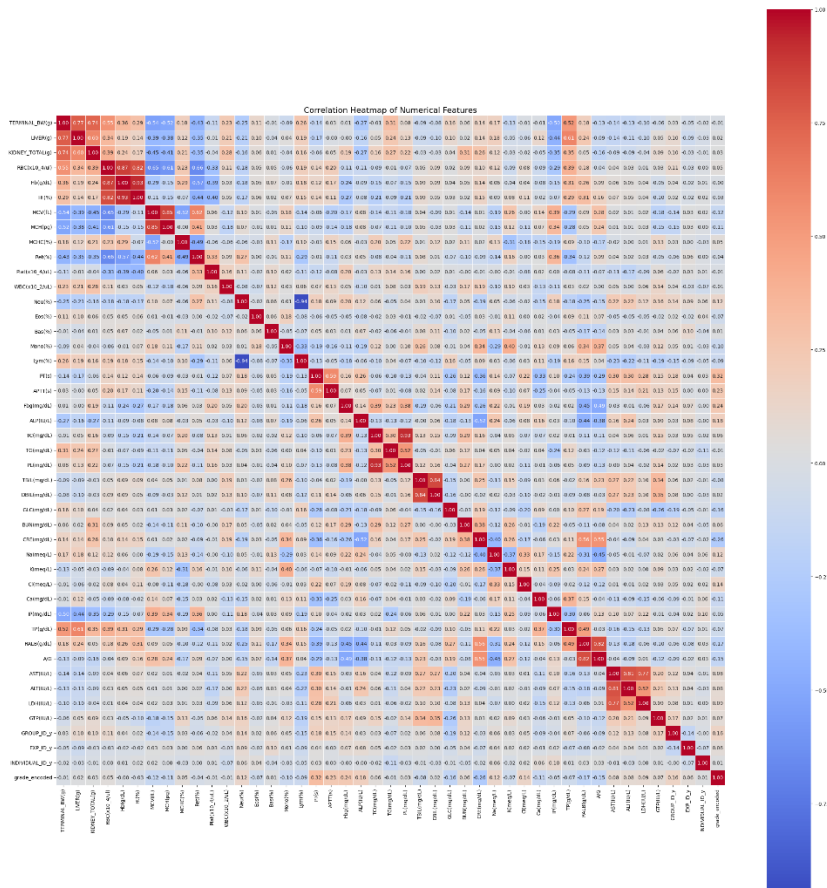
4. Model Performance Reports

A. Random Forest

```
 Random Forest Report:
Accuracy: 0.7615499254843517
              precision    recall  f1-score   support

           0       0.80      0.93      0.86       394
           1       0.53      0.19      0.27        54
           2       0.67      0.40      0.50        10
           3       0.69      0.62      0.65       213

    accuracy                           0.76       671
   macro avg       0.67      0.53      0.57       671
weighted avg       0.74      0.76      0.74       671
```

B. Support Vector Machine (SVM)

```
 SVM
Accuracy: 0.5886736214605067
              precision    recall  f1-score   support

           0       0.59      1.00      0.74       394
           1       0.00      0.00      0.00        54
           2       0.00      0.00      0.00        10
           3       1.00      0.00      0.01       213

    accuracy                           0.59       671
   macro avg       0.40      0.25      0.19       671
weighted avg       0.66      0.59      0.44       671
```

C. XGBoost Classifier

```
 XGBoost Report:
              precision    recall  f1-score   support

           0       0.82      0.90      0.86       394
           1       0.50      0.24      0.33        54
           2       0.62      0.50      0.56        10
           3       0.67      0.65      0.66       213

    accuracy                           0.76       671
   macro avg       0.65      0.57      0.60       671
weighted avg       0.74      0.76      0.75       671

Accuracy: 0.7600596125186289
```

Three machine learning models Random Forest, Support Vector Machine (SVM), and XGBoost were trained to classify compound-induced toxicity into four severity levels such as minimal, slight, moderate, and severe. Among these, the Random Forest model outperformed the others, achieving an accuracy of 97.32% and a weighted-average F1-score of 0.97. It showed strong performance in detecting low-toxicity cases, though struggled with moderate and severe cases due to class imbalance.