Traditional machine learning hyper-parameters:

- For SVM, soft margin (C) is set to 1.0 and class imbalance correction is applied.
- For Random Forest, the number of estimators is 100.
- For logistic regression, class imbalance correction is applied.
- The maximum number of features used are 10000, and character word-ngram range is (1,2).

Fixed parameters across all deep learning baselines and for proposed semi-supervised approaches:

- The learning rate was set to 0.001
- Number of epochs set to 10.
- Batchsize set to 64
- We employ 0.25 dropouts after each input as well as before the final fully connected layer.

Tuned hyper-parameter values for DL baselines for 23 categories

Approach	LSTM dim	Attention dim	CNN Filters
biLSTM	300	N.A.	N.A.
biLSTM-Attention	300	600	N.A.
Hierarchical-biLSTM-Attention	200	400	N.A.
BERT-biLSTM-Attention	300	600	N.A.
USE-biLSTM-Attention	300	600	N.A.
CNN-biLSTM-Attention	300	500	100
CNN-Kim	0	0	150
C-biLSTM	300	0	N.A.
tBERT-biLSTM-Attention	300	600	N.A.
Opti-DL	200	400	N.A.
Random Sampling+ Opti-DL	100	200	N.A.

Tuned hyper-parameter values for proposed semi-supervised methods for 23 ctegories

Approach	Threshold (T)	Top_p	LSTM dim	Attention dim
Basic + Opti-DL	0.75	N.A.	200	500
IPSPC + Opti-DL	0.75	N.A.	300	500
Diversity.label + Opti-DL	0.75	0.8	300	500
Diversity.uniform + Opti-DL	0.75	0.9	300	600
Support.weakest + Opti-DL	0.75	0.9	300	400
Support.uniform + Opti-DL	0.75	0.8	300	400
S(Diversity.label, Support.uniform) + Opti-DL	0.75	0.8	300	500
Diversity-label ‘intersection’ Support.weakest + Opti-DL	0.75	0.9	300	500