For each method, all the hyper-parameter tuning is done using the validation data. The hyper-parameters values for the baselines and proposed semi-supervised approaches are as follows. P_{min} and P_{max} are set to 4 and 7 respectively. The minimum labeled data coverage S_m is set to 1300 by observing the class distribution in the original labeled data. For each method, we pick the optimal confidence threshold T and Top_p based on F_{macro} on the validation set. Hyper-parameter tuning is done using the validation data. The hyper-parameters values for the baselines and proposed methods are as follows.

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 is 10000, and the character word-ngram range is (1,2).

Fixed parameters across all deep learning (DL), semi-supervised baselines and for proposed methods:

- 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

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.
USE-biLSTM-Attention	300	600	N.A.
BERT-biLSTM-Attention	300	600	N.A.
C-biLSTM	300	0	N.A.
CNN-biLSTM-Attention	300	500	100
CNN-Kim	0	0	150

Tuned hyper-parameter values for semi supervised baselines

Approach	Threshold (T)	LSTM dim.	Attention dim.
tBERT-biLSTM-Attention	N.A.	300	600
Opti-DL	N.A.	200	400

Random Sampling + Opti-DL	0.8	200	300
Mean-based self-training +Opti-DL	0.8	300	500

Tuned hyper-parameter values for Opti-DL on data augmented by proposed methods using Opti-DL as the base classifier

Approach	Threshold (T)	Top_p	Thu	Thl	LSTM dim	Attention dim
Basic	0.75	N.A.	N.A.	N.A.	200	500
IPSPC	0.75	N.A.	N.A.	N.A.	300	500
Diversity.label	0.75	0.8	N.A.	N.A.	300	500
Diversity.uniform	0.75	0.9	N.A.	N.A.	300	600
Support.weakest	0.75	0.9	N.A.	N.A.	300	400
Support.uniform	0.75	0.8	N.A.	N.A.	300	400
Hard.label	0.75	0.9	N.A.	0.8	200	500
Hard.uniform	0.8	0.9	0.85	N.A.	300	400
S(Diversity.label, Support.uniform)	0.75	0.8	N.A.	N.A.	300	500
S(Diversity.label, Support.weakest)	0.75	0.9	N.A.	N.A.	300	400
S(Diversity.unifirm, Support.uniform)	0.8	0.9	N.A.	N.A.	200	400
S(Diversity.uniform, Support.weakest)	0.75	0.8	N.A.	N.A.	200	500
Diversity.uniform ∩ Support.uniform	0.75	0.8	N.A.	N.A.	300	400
Diversity.uniform ∩ Support.weakest	0.75	0.8	N.A.	N.A.	100	100
Diversity.label ∩ Support.uniform	0.75	0.9	N.A.	N.A.	300	500
Diversity.label ∩ Support.weakest	0.75	0.9	N.A.	N.A.	200	500

Tuned hyper-parameter values for proposed classification methods

Approach	LSTM dim.	Attention dim.
Proposed Architecture (PA)	200	300

PA without pretraining	300	400
Multi-level PA	200	500

Tuned hyper-parameter values for proposed combined methods

Approach	LSTM dim	Attention dim
PA on data augmented by Diversity.label∩Support.weakest using Opti-DL as the base classifier	200	400
Opti-DL on data augmented by Diversity.label∩Sup-port.weakest using PA as the base classifier	200	300
PA on data augmented by Diversity.label∩Support.weakest using PA as the base classifier (PA-aug)	300	600
PA-aug with CBCE Loss	300	400
Multi-level PA-aug	200	300
Multi-level PA-aug with CBCE Loss	300	400