

For each method, all the 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 15, linear kernel, and class imbalance correction is applied.
- For Random Forest, the number of estimators is 100.
- For Xgboost, class imbalance correction is applied.
- The maximum number of features used are 10000, and character ngram range is (1,5) and word ngram range is (1,2).

Fixed parameters across all deep learning (DL):

- The learning rate was set to 5e-5.
- Number of epochs set to 4.
- Batchsize set to 128.
- Maximum sequence length = 64
- Seed value is 21.
- 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	Features	LSTM dim.	Attention dim.	CNN Filters
C-biLSTM	BERT	150	N.A.	200
biLSTM-Attention	BERT	100	100	N.A.
biLSTM	BERT	150	N.A.	N.A.
CNN	BERT	N.A.	N.A.	150
C-biLSTM	ELMO	200	N.A.	200
biLSTM-Attention	ELMO	200	100	N.A.
biLSTM	ELMO	200	N.A.	N.A.
CNN	ELMO	N.A.	N.A.	100