

# Experiments

## Performance Comparison

Category	Method	Accuracy	AUC-ROC	Fake News			Real News		
				Precision	Recall	F <sub>1</sub>	Precision	Recall	F <sub>1</sub>
Supervised	LIWC-LR	0.528	0.558	0.604	0.160	0.253	0.517	0.896	0.655
	LIWC-SVM	0.568	0.598	0.574	0.521	0.546	0.563	0.614	0.587
	LIWC-RF	0.590	0.616	0.613	0.483	0.541	0.574	0.696	0.629
	LSTM	0.733	0.799	0.876	0.543	0.670	0.669	<b>0.923</b>	0.775
	CNN	0.747	0.834	0.869	0.580	0.696	0.685	0.913	0.783
	EANN	0.767	0.803	0.863	0.634	0.731	0.711	0.899	0.794
Semi-supervised	LSTM <sub>semi</sub>	0.753	0.841	0.854	0.611	0.713	0.697	0.895	0.784
	CNN <sub>semi</sub>	0.759	0.848	0.850	0.630	0.723	0.706	0.889	0.787
Automatically annotated	WeFEND— WeFEND	0.807 <b>0.824</b>	0.858 <b>0.873</b>	0.846 <b>0.880</b>	<b>0.751</b> <b>0.751</b>	0.795 <b>0.810</b>	0.776 <b>0.783</b>	0.863 0.898	0.817 <b>0.836</b>

- Observe that WeFEND achieves the best results in terms of Accuracy, AUC-ROC, precision, recall and F1 measurement

# Experiments

## Performance Comparison

Category	Method	Accuracy	AUC-ROC	Fake News			Real News		
				Precision	Recall	F <sub>1</sub>	Precision	Recall	F <sub>1</sub>
Supervised	LIWC-LR	0.528	0.558	0.604	0.160	0.253	0.517	0.896	0.655
	LIWC-SVM	0.568	0.598	0.574	0.521	0.546	0.563	0.614	0.587
	LIWC-RF	0.590	0.616	0.613	0.483	0.541	0.574	0.696	0.629
	LSTM	0.733	0.799	0.876	0.543	0.670	0.669	<b>0.923</b>	0.775
	CNN	0.747	0.834	0.869	0.580	0.696	0.685	0.913	0.783
	EANN	0.767	0.803	0.863	0.634	0.731	0.711	0.899	0.794
Semi-supervised	LSTM <sub>semi</sub>	0.753	0.841	0.854	0.611	0.713	0.697	0.895	0.784
	CNN <sub>semi</sub>	0.759	0.848	0.850	0.630	0.723	0.706	0.889	0.787
Automatically annotated	WeFEND— WeFEND	0.807 <b>0.824</b>	0.858 <b>0.873</b>	0.846 <b>0.880</b>	<b>0.751</b> <b>0.751</b>	0.795 <b>0.810</b>	0.776 <b>0.783</b>	0.863 0.898	0.817 <b>0.836</b>

- LIWC-LR achieves the worst performance. The reason is that LIWC-LR is a linear model and hard to discriminate the complicated distributions of fake and real news content