Dataset	Naive Bayes Test Accuracy	Naive Bayes F1 Score	Support Vector Machine Test Accuracy	Support Vector Machine F1 Score	Gradient Boosting Test Accuracy	Gradient Boosting F1 Score	Multi-layer Perceptron Test Accuracy	Multi-layer Perceptron F1 Score	CNN Test Accuracy	CNN F1 Score
George McIntire	86.19	0.8608	92.03	0.9173	88.16	0.8786	91.87	0.9137	93.29	90.9284
Liar	62.67	0.6972	61.68	0.6728	60.01	0.703	60.99	0.6446	59.17	0.6324
Twitter	98.11	0.9811	98.96	0.9896	95.09	0.9486	98.99	0.9898	99.45	0.9945
BanFakeNews	87.69	0.8735	90.19	0.8990	90.48	0.9011	91.54	0.9080	96.35	0.9598

TABLE I: Data from the Naive Bayes, SVM, Boosting, MLP and CNN models in our experiments

Dataset	LSTM Test Accuracy	LSTM F1 Score	dEFEND Test Accuracy	dEFEND F1 Score	NURG Test Accuracy	NURG F1 Score	Supervised Test Accuracy	Supervised F1 Score	Active Learning Test Accuracy	Active Learning F1 Score	Active Learning Fraction of Dataset
George McIntire	90.92	0.9053	93.37	0.9320	93.69	0.9320	95.42	0.9541	**94.00**	**0.9407**	**25%**
Liar	58.19	0.6033	60.60	0.6285	58.48	0.6310	62.17	0.7132	**61.04**	**0.6792**	**18%**
Twitter	99.48	0.9948	99.48	0.9948	99.51	0.9951	99.41	0.9941	**99.0**	**0.99**	**4%**
BanFakeNews	97.60	0.9739	97.31	0.9703	97.31	0.9698	97.12	0.9676	**94.13**	**0.9413**	**28%**

TABLE II: Data from LSTM, dEFEND, Neural User Response Generator and our proposed system in our experiments

In this section, we look at how our ensemble model fairs against the two other state-of-the-art models, Neural User Response Generator (NURG) and dEFEND, when all of them are trained using the proposed entropy-based active learning method. Figure 3 shows the results of this analysis through the accuracy against the fraction of the dataset used to train the model for the benchmark datasets. The blue, green, and purple curves correspond to the ensemble, NURG, and dEFEND respectively while the red line is the accuracy of the ensemble model during supervised training.

Observing the graphs in Figure 3, we wish to evaluate the performance of our ensemble model compared to the others under the same active learning settings. This can be done by observing whether, using the same active learning method, the comparison models achieve the threshold accuracy at a much lower fraction than our model. From Figure 3, we can see that our model performs better or the same for each of the datasets. For the McIntire and Liar dataset, the ensemble model achieves the threshold accuracy at a lower fraction before the other two models. For the Twitter and BanFakeNews datasets, all three models show similar performance as they converge to the threshold accuracy at the same fraction of the dataset.

V. Conclusion

In this research, we implement a novel active learning-based multi-model neural ensemble architecture for automated fake news detection using low amounts of data to pre-emptively prevent the detrimental spread of fake news. We tested our methods on several real-life datasets and against benchmarks to prove the efficacy of our model. The simulation results show that our proposed method significantly reduces the burden of labeling large news datasets for training a classifier. As part of future work, we wish to develop similar architecture based on active learning to other problems where data categorization and annotation are difficult and cumbersome such as political NLP research, hate speech, etc.

Acknowledgment

This work was supported by the ICT Division, Government of the People's Republic of Bangladesh.

References

[1] T.-S. Chang and W.-H. Hsiao, "Time spent on social networking sites: Understanding user behavior and social capital," *Systems Research and Behavioral Science*, vol. 31, no. 1, pp. 102–114, 2014.

[2] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

[3] W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648*, 2017.

[4] S. D. Bhattacharjee, A. Talukder, and B. V. Balantrapu, "Active learning based news veracity detection with feature weighting and deep-shallow fusion," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 556–565.

[5] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "defend: Explainable fake news detection," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 395–405.

[6] F. Qian, C. Gong, K. Sharma, and Y. Liu, "Neural user response generator: Fake news detection with collective user intelligence." in *IJCAI*, 2018, pp. 3834–3840.

[7] B. Settles, "Active learning literature survey," University of Wisconsin-Madison Department of Computer Sciences, Tech. Rep., 2009.

[8] K. Kowsari, M. Heidarysafa, D. E. Brown, K. J. Meimandi, and L. E. Barnes, "Rmdl: Random multimodel deep learning for classification," in *Proceedings of the 2nd International Conference on Information System and Data Mining*, 2018, pp. 19–28.

[9] S. Kwon, M. Cha, and K. Jung, "Rumor detection over varying time windows," *PloS one*, vol. 12, no. 1, 2017.

[10] M. Z. Hossain, M. A. Rahman, M. S. Islam, and S. Kar, "Banfakenews: A dataset for detecting fake news in bangla," *arXiv preprint arXiv:2004.08789*, 2020.

[11] M. Granik and V. Mesyura, "Fake news detection using naive bayes classifier," in *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*. IEEE, 2017, pp. 900–903.

[12] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks." in *Ijcai*, 2016, pp. 3818–3824.