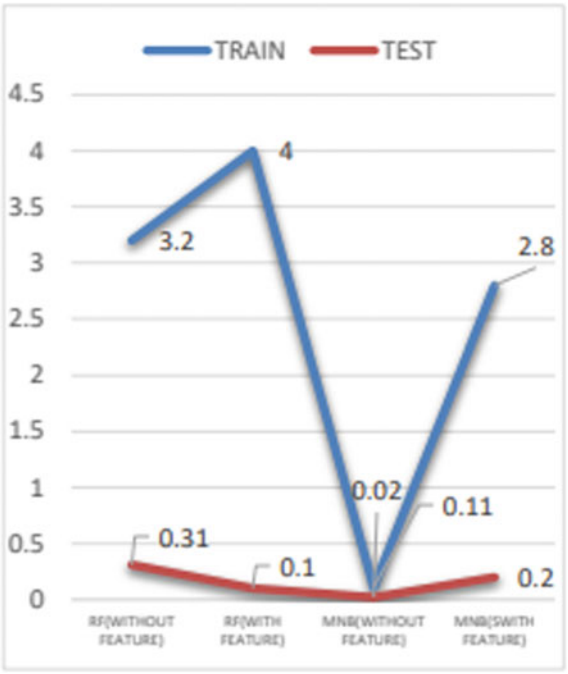

Fig. 3 Train/test time (in sec) for Email spam detection

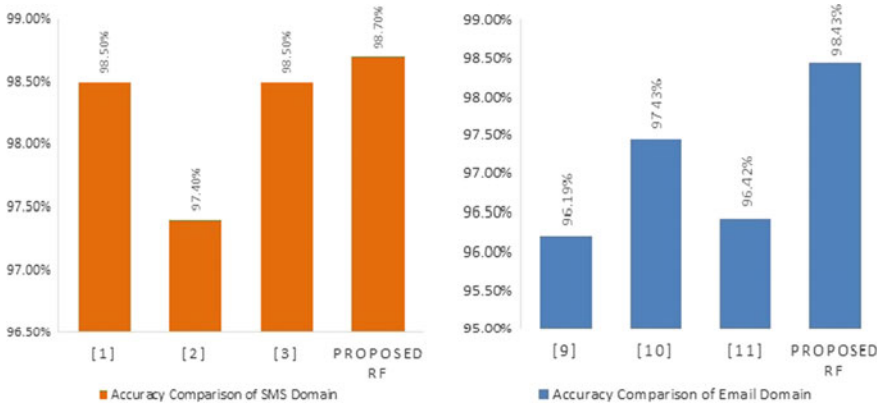

Fig. 4 Accuracy comparison for SMS and Email spam detection

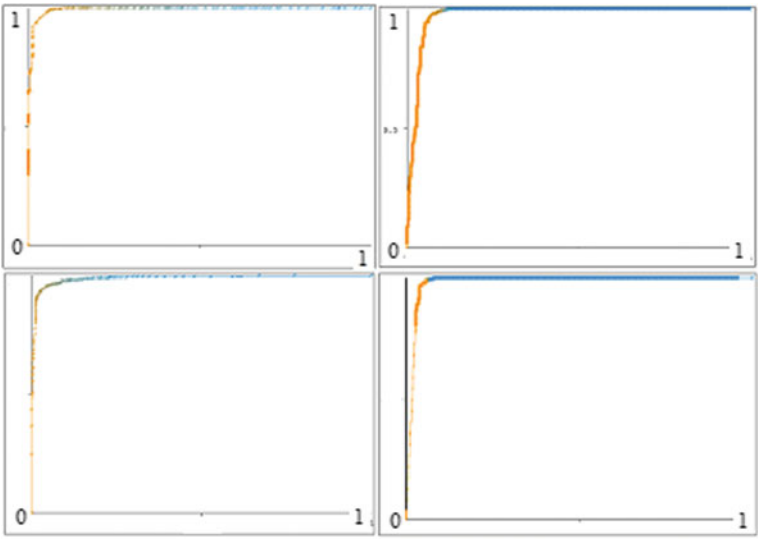

Fig. 5 ROC plots for SMS and Email spam detection

results even better than the hybrid models [9]. RF scores 0.99 ROC value in both model (with and without feature selection/extraction) for SMS spam detection. For MNB, improved 0.984 ROC value is achieved with feature selection/extraction than the 0.982 ROC value without feature selection/extraction for SMS spam detection. For Email spam detection, RF scores 0.98 ROC value for in both model (with and without feature selection/extraction). Also , MNB score 0.994 ROC value with feature selection/extraction as compared to 0.987 ROC value without feature selection/extraction. Further Fig. 5 represents the plot for ROC for spam detection in SMS and Email with feature selection/extraction.

5 Conclusions and Future Work

After experimentation on the selected state-of-the-art, various conclusions are drawn. It is concluded that by selecting appropriate feature selection/extraction techniques, the overall performance of the spam detection model can be improved. RF achieved 98.7% accuracy for spam detection in SMS domain and 98.43 for spam detection in Email domain. Also MNB with feature selection and extraction has accuracy of 97.8% and 97.96% for spam detection in SMS and Email domain, respectively. For spam detection in SMS domain, based on all evaluation measures it can be concluded that RF outperformed over other techniques and can be followed for spam detection. MNB perform better in SMS domain with less time to train and test the model. Thus, it can be used as spam detection technique in Email domain. Also, it can be concluded

from Figs. 2 and 3 that time taken for training the model with and without feature selection/extraction varies for SMS and Email spam detection and can be seen in result of different evaluation measures used. As a future work, hybrid methods and evolutionary computing-based algorithms can be followed for spam detection for SMS and Email domains.

Acknowledgements Author's are thankful to Shri G. S. Institute of Technology and Science, Indore for providing all facilities and necessary support to carry out the research work.

References

1. Sethi, P., Bhandari, V., & Kohli, B. (2017). SMS spam detection and comparison of various machine learning algorithms. In *2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)* (pp. 28–31). https://doi.org/10.1109/IC3TSN.2017.8284445
2. Gadde, S., Lakshmanarao, A., & Satyanarayana, S. (2021). SMS spam detection using machine learning and deep learning techniques. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* (p. 358362).
3. Navaney, P., Dubey, G., & Rana, A. (2018). SMS spam filtering using supervised machine learning algorithms. In *2018 8th international conference on cloud computing, data science & engineering (confluence)* (pp. 43–48). https://doi.org/10.1109/CONFLUENCE.2018.8442564
4. Tang, S., Mi, X., Li, Y., Wang, X., & Chen, K. (2022). *Clues in tweets: Twitter-guided discovery and analysis of SMS spam*. arXiv preprint arXiv:2204.01233. 2022 Apr 4.
5. ElBakrawy, L. M. (2019). *Hybrid particle swarm optimization and Pegasos algorithm for spam Email detection*. https://doi.org/10.25728/assa.2019.19.3.751
6. Jain, T., Garg, P., Chalil, N., Sinha, A., Verma, V. K., & Gupta, R. (2022). SMS spam classification using machine learning techniques. In *2022 12th international conference on cloud computing, data science & engineering (confluence)* (pp. 273–279). https://doi.org/10.1109/Confluence52989.2022.9734128
7. Annareddy, S., & Tammina, S. (2019). A comparative study of deep learning methods for spam detection. In *2019 third international conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 66–72).
8. Agarwal, K., & Kumar, T. (2018). Email spam detection using integrated approach of Naïve Bayes and particle swarm optimization. *Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018,* 685–690. https://doi.org/10.1109/ICCONS.2018.8662957.
9. Iyengar, A., Kalpana, G., Kalyankumar, S., & GunaNandhini, S. (2017). Integrated SPAM detection for multilingual emails. *International Conference on Information Communication and Embedded Systems (ICICES), 2017,* 1–4. https://doi.org/10.1109/ICICES.2017.8070784.
10. Ponmalar, A., Rajkumar, K., Hariharan, U., Kalaiselvi, V. K. G., & Deeba, S. (2021). Analysis of spam detection using integration of logistic regression and PSO algorithm. In *2021 4th International Conference on Computing and Communications Technologies (ICCCT)* (pp. 396–402). https://doi.org/10.1109/ICCCT53315.2021.9711903
11. Raza, M., Jayasinghe, N. D., & Muslam, M. M. A. (2021). A comprehensive review on email spam classification using machine learning algorithms. *International Conference on Information Networking (ICOIN), 2021,* 327–332.
12. GuangJun, L., Nazir, S., Khan, H. U., & Ul Haq, A. (2020). Spam detection approach for secure mobile messgae communication using machine learning algorithms. In *Hindawi, security and communication netwroks* (Vol. 2020, Article id: 8873639)