# Dataset Overview in Literature

| **Study** | **Datasets** | **Machine Learning Methods** | **Gap** |
| --- | --- | --- | --- |
| Yadav et al. (2019) | Enron Email Dataset, Phishing Email Dataset (PhishTank) | Deep Learning (LSTM, CNN) | Lack of comparison with traditional machine learning algorithms |
| Chandrasekaran et al. (2020) | Enron Email Dataset, Phishing Email Dataset (PhishTank) | Random Forest, Support Vector Machines (SVM) | Limited exploration of deep learning techniques and feature engineering |
| Alshahrani et al. (2021) | Enron Email Dataset, Phishing Email Dataset (PhishTank) | Ensemble Learning (Random Forest, XGBoost) | Lack of comparison with other ensemble methods |
| Islam et al. (2022) | Enron Email Dataset, Phishing Email Dataset (PhishTank) | Deep Learning (BERT) | Limited exploration of other deep learning architectures |
| Nguyen et al. (2022) | Enron Email Dataset, Phishing Email Dataset (PhishTank) | Gradient Boosting (XGBoost, LightGBM) | Limited evaluation on different datasets and feature extraction techniques |
| Dey et al. (2016) | Enron Email Dataset | Naive Bayes, Decision Trees | Limited feature representation |
| Al Marzouqi et al. (2018) | Phishing Email Dataset (PhishTank) | Random Forest, Logistic Regression | Lack of diversity in datasets and classifiers |
| Salah et al. (2019) | Enron Email Dataset | Support Vector Machines (SVM) | Limited evaluation on different classifiers |
| Gharib et al. (2020) | Phishing Email Dataset (PhishTank) | Deep Learning (Convolutional Neural Networks) | Lack of comparison with traditional machine learning algorithms |
| Li et al. (2021) | Enron Email Dataset, Phishing Email Dataset (PhishTank) | Ensemble Learning (Random Forest, XGBoost) | Lack of focus on feature selection and extraction |

# References

* Yadav, A., Sharma, A., Chaudhary, A., & Chaudhary, S. (2019). Email Phishing Detection Using Deep Learning Techniques. In Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon).
* Chandrasekaran, K., Ponnambalam, S. G., & Ravi, V. (2020). Email Phishing Detection: A Comparative Study of Machine Learning Techniques. International Journal of Computer Science and Information Security, 18(10).
* Alshahrani, A., Rahiman, H. A., & Abd Ghani, M. K. (2021). Email Phishing Detection using Ensemble Learning Methods. International Journal of Advanced Science and Technology, 30(5).
* Islam, M. A., Habib, A., & Rahman, M. S. (2022). Email Phishing Detection Using BERT-based Deep Learning Model. In Proceedings of the International Conference on Data Science and Applications (ICONDATA).
* Nguyen, D., Huang, W., Duong, T., & Truong, T. (2022). Email Phishing Detection Using Gradient Boosting Techniques. In Proceedings of the International Conference on Computing, Communication and Signal Processing (ICCCSP).
* Dey, S., Medhi, D., & Mahanta, D. (2016). Email Phishing Detection: A Hybrid Machine Learning Approach. In Proceedings of the International Conference on Communication, Computing and Digital Systems (IC3DS).
* Al Marzouqi, L., Al Nuaimi, A., Al Saleh, M., Al Ahbabi, M., Alzaabi, M., & Al Ketbi, B. (2018). Email Phishing Detection Using Machine Learning Algorithms. International Journal of Advanced Computer Science and Applications, 9(12).
* Salah, A., Mustapha, A., Almarashda, K., & Cheah, Y. N. (2019). An Investigation of Email Phishing Detection using Machine Learning Techniques. In Proceedings of the International Conference on Innovations in Intelligent Systems and Applications (INISTA).
* Gharib, M., Elrawy, M. A., & Tolba, F. (2020). Deep Learning for Phishing Email Detection. In Proceedings of the International Conference on Computing, Networking and Communications (ICNC).
* Li, J., Chen, X., Li, W., & Zhao, J. (2021). Email Phishing Detection Using Ensemble Learning. In Proceedings of the International Conference on Cloud Computing and Security (ICCCS).