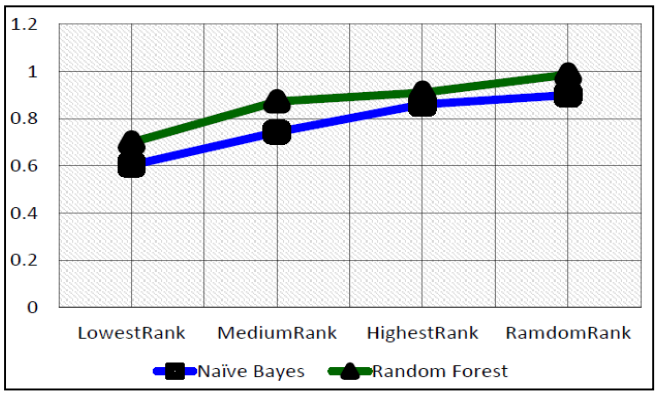

Fig. 9. Receiver Operation Curve (ROC) for dynamic features

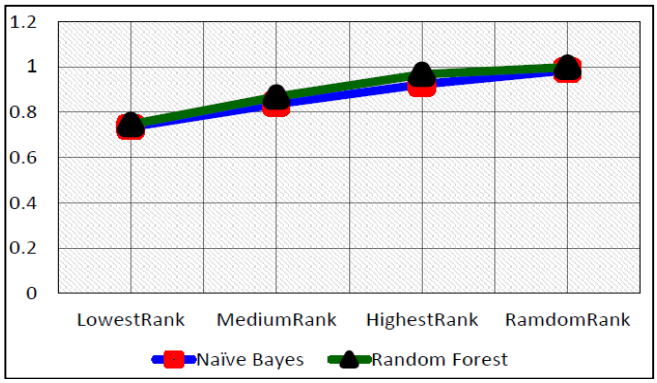

Fig.10. Receiver Operation Curve (ROC) for integrated features

IV. Conclusion

The experiment conducted between static features, dynamic features and integrated features on the various ranking group (Lowest_Rank, Medium_Rank, Highest_Rank, and Random_Rank). The Random Forest algorithm is the best classification algorithm because of it achieved high accuracy result, low means absolute error and high ROC value as compare to Naïve Bayes classifier for the most dataset. Our proposed integrated features show the promising result when tested on the various ranking group regarding accuracy; low means absolute error and high ROC value as compared to static and dynamic features. This shows that the selected integrated feature contribute well for virus classification and the given set of training feature vectors is most important for discriminating between the classes to be learned.

We plan to explore the integrated features to improve polymorphic computer virus classification. For instance, future work has to use an optimal number of features. This may reveal the information on processing time and also the effectiveness of the system since good features can affect the performance of the classifier. Thus, the integrated mechanism can update the classifier when notice new possible features in real time

Acknowledgment

The authors express appreciation to the Universiti Tun Hussein Onn Malaysia (UTHM). Postgraduate Research Grant vot number U610 supports this research, Short Term Grant vot number U653 and Gates IT Solution Sdn. Bhd. under its publication scheme.

References

[1] S. Chaumette, O. Ly, and R. Tabary, “Automated extraction of polymorphic virus signatures using abstract interpretation,” Proc. - 2011 5th Int. Conf. Netw. Syst. Secure. NSS 2011, pp. 41–48, 2011.

[2] A. A. E. Elhadi, “Malware Detection Based on Hybrid Signature Behaviour Application Programming Interface Call Graph,” Am. J. Appl. Sci., vol. 9, no. 3, pp. 283–288, 2012.

[3] H. Lim, Y. Yamaguchi, H. Shimada, and H. Takakura, “Malware classification method based on sequence of traffic flow BT - 1st International Conference on Information Systems Security and Privacy, ICISSP 2015, February 9, 2015 - February 11, 2015,” 2015, pp. 230–237.

[4] G. Nascimento and M. Correia, “Anomaly-based intrusion detection in software as a service,” Proc. Int. Conf. Dependable Syst. Networks, pp. 19–24, 2011.

[5] R. Islam, R. Tian, L. Batten, and S. Versteeg, “Classification of Malware Based on String and Function Feature Selection,” 2010 Second Cybercrime Trust. Comput. Work., pp. 9–17, 2010.

[6] A. Tang, S. Sethumadhavan, and S. Stolfo, “Unsupervised Anomaly-based Malware Detection using Hardware Features,” Proc. Int. Symp. Res. Attacks, Intrusion Detect., p. 1, 2014.

[7] R. Sekar, a Gupta, J. Frullo, T. Shanbhag, a Tiwari, H. Yang, and S. Zhou, “Specification-based anomaly detection: a new approach for detecting network intrusions,” CCS ’02 Proc. 9th ACM Conf. Comput. Commun. Secur., pp. 265–274, 2002.

[8] E. Al Daoud, I. Jebril, and B. Zaqaibeh, “Computer virus strategies and detection methods,” Int. J. Open Probl. Comput. Math., vol. 1, no. 2, pp. 122–129, 2008.

[9] A. Techniques, “MALWARE: Threats and Attacks Part 1-D: How to protect from Malware attacks, Antivirus Techniques Malware threats and attacks,” 2012.

[10] Idika, N. (2007). A Survey of Malware Detection Techniques.

[11] R. Islam, R. Tian, L. Batten, and S. Versteeg, “Classification of Malware Based on String and Function Feature Selection,” 2010.

[12] F. Leder, B. Steinbock, and P. Martini, "Classification and detection of metamorphic malware using value set analysis," *2009 4th International Conference on Malicious and Unwanted Software (MALWARE)*, Montreal, QC, 2009, pp. 39-46.

[13] Y. Ye, T. Li, Q. Jiang, and Y. Wang, "CIMDS: Adapting Postprocessing Techniques of Associative Classification for Malware Detection," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 3, pp. 298-307, May 2010.

[14] K. Huang, Y. Ye and Q. Jiang, "ISMCS: An intelligent instruction sequence based malware categorization system," *2009 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication*, Hong Kong, 2009, pp. 509-512.

[15] N. Bayes, “Naive Bayes classifier,” pp. 1–9, 2006.

[16] “VXHeaven_Dataset,” 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC), 2014.

[17] A. R. Kakad, S. G. Kamble, S. S. Bhuvad, and V. N. Malavade, “Study and Comparison of Virus Detection Techniques,” Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 4, no. 3, pp. 251–253, 2014.

[18] R. Tian, R. Islam, L. Batten, and S. Versteeg, “Differentiating Malware from Cleanware Using Behavioural Analysis,” pp. 23–30, 2010.

[19] R. Islam, R. Tian, L. M. Batten, and S. Versteeg, “Journal of Network and Computer Applications Classification of malware based on integrated static and dynamic features,” vol. 36, pp. 646–656, 2013.

[20] H. Zhao, M. Xu, N. Zheng, J. Yao and Q. Ho, "Malicious Executables Classification Based on Behavioral Factor Analysis," *2010 International Conference on e-Education, e-Business, e-Management and e-Learning*, Sanya, 2010, pp. 502-506.

[21] Grégoire Jacob, Hervé Debar, Eric Filiol, "Malware detection using attribute-automata to parse abstract behavioral descriptions," *CoRR abs/0902.0322*, 2009.

[22] I.R.A Hamid, N.S Khalid, N.A. Abdullah, N. H. Ab Rahman, C.C. Wen, “Android Malware Classification Using K-Means Clustering Algorithm,” 2017 IOP: Conference Series: Materials Science and Engineering, Melaka, 2017, vol. 226.

[23] A. Zulkifli, I.R.A Hamid, W.M Shah, and Z. Abdullah, “Android Malware Detection Based on Network Traffic Using Decision Tree Algorithm,” 2018 *International Conference on Soft Computing and Data Mining*, pp. 485-494.