**Future work**

As discussed in the previous sections of this report, our approach focused on simplicity and speed of computation. The chosen model was also the one that minimized the number of false negatives and therefore minimized the risks of an IoT system being infected.

There are several possible avenues for future development of this work.

**Expand number of features / try different feature selection methods**

Other methods in the literature on this subject built models using more features than in our best model (which used 10 features). Rezvy et. al built a highly accurate model for a similar purpose using 36 features derived from the same dataset [1]. In light of this, a future development may be to gradually increase the number of features used to train the model and evaluate the performance iteratively to find the optimum number of features. For this project, there was not enough time to carry out the necessary iterations to achieve this. Increasing the number of features may improve the accuracy of the model but at the cost of model speed and simplicity.

Another feature selection method worth exploring would be the use of an auto-encoder (a neural network that compresses a dataset) as this is also very popular in the literature [2]. The trade-off here would be the loss of interpretability from combining / compressing features.

**More complex learning algorithms**

One potential future development would be to use more complex learning algorithms. The project mostly focused on simpler algorithms such as random forests and KNN. It would be interesting to explore artificial neural networks as a learning approach using this dataset and evaluate the results to compare against the current model. This approach is popular in the literature.

In general, most avenues for future development would be to increase the complexity and assess the impact this has on model accuracy versus speed.

**Input from domain experts**

The view taken in this project was that it is desirable to look to minimise the number of false negatives and therefore the risk of the system being infected. A next step would be to look for input from domain experts on the trade-off between sensitivity and specificity in the context of this issue. This could allow us to refine our choice of algorithm and give us direction for which avenues to explore further. Further work and deeper reading of the literature could also be informative on this subject.

**References**

[1] Rezvy et. al, “An efficient deep learning model for intrusion classification and prediction in 5G and IoT networks” in 53rd Annual Conference on Information Sciences and Systems (CISS), 2019.

[2] Kolias et. al, “Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset”, IEEE Communication Surveys & Tutorials, Vol. 18, No. 1, First Quarter 2016