# CONCLUSION

This work focused on two major concept, firstly, the proposal of a real-time water monitoring network for gathering data on water parameters from water bodies. Secondly, the application of machine learning (ML) models as means of assessing water quality. The developed water monitoring network is based on Lo Ra, a low power long range protocol for data transmission, and was developed using the City of Cape Town as case study. Results of the simulation done in Radio Mobile, revealed a partial mesh network topology as the most adequate network to cover the city. Data gathered from this monitoring network would ideally be aggregated on a Cloud server, where ML models can then be applied to assess the water's fitness of use for drinking or irrigation purposes. Due to the absence of relevant datasets, two suitable datasets were built in this work and used to training and testing three ML models considered, which are Random Forest (RF), Logistic Regression (LR) and Support Vector Machine (SVM). Results of the test showed that LR performed best for drinking water, as it gave the highest classification accuracy and lowest false positive and negative values, while SVM was better suited for irrigation water. Finally, a model for identifying the most influential water parameter(s) w.r.t classification accuracies of the ML models was then explored using recursive feature elimination (RFE). Obtained results showed that pH, and total hardness were the least influential parameters in drinking water, while SSP was the least for irrigation water.

Though the authors acknowledge the possible application of deep learning models, these were not used in this work. In future works, deep learning models such as the various variants of neural networks could be considered as expansion to this work. Furthermore, water quality indices were manually calculated and used to assess the ``fitness for use'' of water, future works could explore the application of unsupervised ML models as alternatives to manually calculated water quality indices. In the same vein, rather than using RFE, other approaches such as multi criteria decision making could also be considered to identify influential parameters. Finally, incorporating usage prediction models and microbial monitoring into the water network as well as tracking sources of water contaminates could also be avenues to further this work.