2 Conclusions and Future Work

The objective of this thesis is to investigate the viability of machine learning methods applied to FTIR spectroscopy for use as prognostic tools in head and neck cancer. Despite phenomenal progress in the development of treatment options for OSCC over the past few decades, there is still work to be done. Additional treatment methods exist in the form of pre-operative neoadjuvant therapy, these treatment options show promise for improving clinical outcomes for patients with OSCC. However, research progress into neoadjuvant therapies is hindered by the difficulty of identifying patients eligible for clinical window trials. Due to the increased risk of adverse effects and additional health risks associated with hormone treatment and chemotherapy, it is unethical to accept patients who would not be responsive to treatment onto a trial. The difficulty lies in the fact that using pre-existing prognostic biomarkers it is only possible to determine a patient's prognosis after post-surgical nodal resection has taken place. The work undertaken in this thesis has shown that using sophisticated statistical methods to analyse FTIR spectroscopy data it is possible to predict the prognosis of a patient from primary tumour tissue sections.

A list of criteria was set out in **??** which summarises the requirements of any diagnostic method aiming to be widely adopted in a clinical setting. The performance of a diagnostic or prognostic test must be demonstrably superior to or at least be able to supplement existing methods; the method must also be sensitive to its intended range. Both of these criteria have been established

to be true of FTIR spectroscopy in **??** and chapter 1; a rigorous analysis procedure was performed which demonstrated that developed statistical models were able to stratify patients by risk status. Survival analysis was performed to ascertain the prognostic utility of an FTIR spectrum; results showed that risk groups had clearly distinguished survival curves. **??** contains a comparison to a known prognostic biomarker: ASMA, which was shown to be substantially less effective than FTIR in the data set used in the study. The two prognostic biomarkers were then combined to create a hybrid model which demonstrated superior performance in some cases over the two individual models. This suggests that combining prognostic biomarkers from a variety of assay methods may be an effective method of developing capable prognostic tools for use in a clinical setting.

Another important criterion is the identification of sources of error, whether or not these sources of error can be addressed is crucial for the prospects for any diagnostic test. **??** covers the development of a distributed preprocessing optimisation framework for use with FTIR spectra; the work was undertaken in collaboration with another PhD student Barney Ellis. The optimisation framework was designed to provide a means of determining an effective preprocessing and classification pipeline for a given prediction objective. It is crucial to explore the parameter space of a pipeline as the classification performance of a pipeline is highly dependent upon the choice of configuration. The framework proved to be very successful, the top performing pipeline configurations showed substnatially better scores with the difference between the best and worst pipelines being $>$70%. The configurations of the top ranked pipelines tended to be simpler but reenforced some conventional wisdom about preprocessing FTIR data. The framework was developed on a widely adopted parallel processing framework, and could be easily adapted for other multivariate tasks with similar constraints. The framework will be made available for public use.

An evaluation of deep learning based FTIR analysis in Chapter 1 included a comparison between two different network architectures which were determined through a network optimisation process]. Both models showed some promise as prognostic tools, with the CNN model performing particularly well in comparison to all models developed in this thesis. This is an encouraging result as CNNs are more robust to measurement inconsistencies which would likely be a hinderence to the wider adoption of FTIR as a clinical tool.

To take the work covered here forward for further development; larger collaborative efforts should be made to diversity and quality of data available for model training. CNNs should be explored more thoroughly as they have the potential to alleviate measurement issues, increasing the likelihood of wider adoption of the technique.