Thesis Plan

Submission target is March 2018, this will include a version of the full 2016 dataset analysis with emphasis on the vector boson fusion production mode and use of machine learning techniques in the analysis. This will include an improved VBF tag that uses deep learning. A plan of the thesis structure, section contents, and estimated page counts are given below.

1. **Introduction (2 Pages)**
Overview, context and motivation

2. **Theory (10 Pages)**
Gauge theory and the relevance of symmetry groups to bosons, the breaking of these gauge symmetries and the consequences, the standard model with emphasis on the electroweak sector and its breaking by the Higgs mechanism plus fundamental mass generation. Description of the resulting scalar boson.

3. **Apparatus (10 Pages)**
LHC and CMS. Detailed description of the relevant subsystems of CMS with particular emphasis on ECAL and its affect upon the measurement of EM objects.

4. **Object Reconstruction (20 Pages)**
Description of how physics objects are constructed from detector-level information, with detailed treatment of photons and jets. Detail on how the ECAL's measurement affects object reconstruction and the Higgs diphoton decay, plus how it is calibrated.

5. **Machine Learning (10 Pages)**
Concepts, models and techniques in machine learning required for subsequent chapters. This should include general concepts and terminology, boosting, regularisation, optimisation. Particularly concepts needed to understand dense convolutional deep networks. (This will most likely need to be pruned and have parts moved to appendices)

6. **Event Selection and Categorisation (30/40 Pages)**
    I. How events are selected up to tagging: trigger, vertexing, photon ID, diphoton BDT
    II. General description of tagging scheme: extra objects from production mode, description of the fall-through sequence, description of tags other than VBF
    III. VBF Tag
        a. The problem formulation: jet images and their analysis, engineered features, why this particular formulation of channels, problems particular to this image classification problem, etc.
        b. Dense convolutional network: description of architecture and its optimisation, training, outcomes, discussion of the architecture choice
        c. Analysis of the network: front filters, confusion matrix/images, recursive feature removal in the wide part of the architecture to infer what features the deep part is learning, interpretation
        d. (Maybe) Machine learning model survey: method and results

7. **Statistical Analysis and Results (25/30 Pages)**
Description of signal and background modelling (detail on envelope method or can this be handled more coarsely?). Description of the systematics with emphasis on VBF. Final results: mention significance? production rate and compatibility with the SM, compatibility with the SM per tag with VBF emphasis, mass measurement?

8. **Conclusion (2 Pages)**
Closing remarks and conclusions

Appendices (Possible)
Boosting and Tree Ensembles
Bayesian optimisation and application to hyperparameter tuning
Deep learning glossary

Approximate length will be 109 to 124 pages plus references and appendices