Appendix A1: MASTER PROJECT BRIEF FORM 2018-19-20

1. Your details:

Full Name: **Thomas Staite**
Student ID: **6422510**
E-mail @coventry.ac.uk: **staitet@uni.coventry.ac.uk**
Module Code: **M08CDE**
Course of Study: **MSc Data Science & Computational Intelligence**
Project Supervisor: **Dr Mauro Innocente**

2. Project title (provisional) [Meaningful, relevant and concise]

Development of Convolutional Neural Network for the Detection of Railway Track Maintenance

3. Outline (synopsis) of your project. [What are the aim and objectives of the project?]

To perform secondary research on Convolutional Neural Networks (CNNs) for the application of image classification – with emphasis on developing an understanding of their inner-workings, why they are preferred over other network architectures for this use case, which data pre-processing techniques are necessary and why, and how such networks can be optimized to yield greater network performance.

To critically review existing literature and attempts in image classification using Neural Networks - identifying and discussing the benefits, drawbacks, limitations, and obtained results in previous studies and relating this back to my own project for constant improvement and refinements. Reviewing previous work will help influence decision making in the project.

To apply the researched data pre-processing techniques to the provided dataset to prepare it for training and testing of the CNN.

To fully develop, train, and test the researched CNN for the application of classifying images of railway tracks into those that require maintenance, and those that do not. To analyse the performance of the network using a range of metrics based on the resulting Confusion Matrix, and to implement parameter tuning and network optimisation techniques in attempt to improve performance.

To design and develop a suitable Graphical User Interface (GUI) to act as the gateway between the user and the developed network – allowing the user to view the locations of the railway track that require maintenance.

4. Intended user or group of users and their requirements. [a) Who is the intended user or group of users? b) Why you think there is need for this project? c) What are the needs of the intended user that your product should satisfy?]

The initial intended user for this project is my supervisor; Dr Mauro Innocente, who works with the Research Institute for Future Travel and Cities (FTC) at Coventry University and whom listed the requirements for the project on the Moodle website.

A potential use case for this project could be to help railway engineers efficiently locate areas of a railway track that require maintenance without having to manually check the whole track themselves – saving both time and energy. The GUI will provide a helpful interface between the end-user (who

may be computer-illiterate) and the developed network and will provide the necessary information needed to allow these users to fulfil their role.

5. Systems requirements and project deliverables. [a) What are the characteristics/properties that the final product should possess? b) What are the process stages and the corresponding deliverables that will enable you to create the final product?]

The Project Report will be written using Microsoft Word and will contain evidence of the research and literature review that was carried out before the development of the CNN. In addition to this, evidence of data pre-processing techniques, final program code, training/testing results, network optimisation, evidence of the developed GUI, project discussions (benefits, drawbacks, and future work), and project management materials will also be included in this report.

Development of the network architecture will be completed in Python programming language and documented in Jupyter Notebook - which enables the combination of Python code and Plain text with various formatting capabilities.

The network should be incorporated as a Python program with a suitable GUI that can be clearly understood and easily navigated by users of all levels of computing ability. User Acceptance Testing (UAT) may be implemented to gain an understanding of end-user's experience, and adaptations may be implemented based on the feedback received.

When the Python code has been extracted from the Jupyter Notebook to its own file, a GitHub repository should be set up to store the program code – allowing for adjustments to be submitted and the ability to rollback to previous versions if errors arise.

6. Research [a) How will you investigate/identify in detail the needs of the specified user in (3) b) How will you investigate the background of the project?]

Weekly meetings with Mauro will provide opportunities for me to ask questions and find out information about areas in the project that have a degree of ambiguity. Feedback from Mauro will help influence the decisions I make and will ensure that the project will run as smoothly as possible and that the final product meets (and hopefully surpasses) his expectations.

Further research into the technologies that I will be using throughout the project will involve examining external resources such as content on the internet and from the University Library. Full care must be made to ensure the sources of this secondary research is reliable, accurate, and up-to-date. In addition to this, when it comes to reviewing literature of previous attempts, I should be vigilant about when and where the study was conducted, whether it's up-to-date, and whether the results obtained are accurate and trustworthy.

When it comes to developing the CNN, functions from a selection of external Python libraries will be used to aid with the development. Official documentation for these libraries can be accessed online and should be the main source of information regarding these libraries.

7. Evaluation. [a) What makes a product successful? b) How will you demonstrate that your product fulfils the needs of the user in (3)? c) How will you evaluate the product?]

A conclusive piece of research into CNNs and their application in image recognition, along with a critical literature review of previous studies will demonstrate competence in project research. The work completed in this section will be presented to Mauro who may provide feedback.

A successful data preparation stage will involve overcoming the limitations of the supplied dataset (namely, the lack of classification labels, the limited number of images, and further issues with the images themselves) to produce a new dataset which will be fit to train/test the model. This new dataset should be correctly labelled, should have a much-increased quantity of images, and should be resolved of any issues with the image themselves.

For the CNN itself, it will be trained and tested using the resulting dataset from the Data Preparation stage. Performance testing metrics will be employed to provide details on how well the model has been able to train and generalise, and optimisation will hopefully further improve the model. A successful model should yield high predictability power on both the training and the testing set.

As previously mentioned, a User Acceptance Test (UAT) may be implemented to provide Mauro and any other potential end-users with an opportunity to try the product. Any feedback from this test will be used to implement changes; whether necessary, or preferable. This stage will also act as a demonstration of the work that has been completed.

The project will be written professionally using a Microsoft Word document and an accompanying template. Every stage of the project should be fully documented so that when it is submitted, my supervisor, and 2nd assessor has evidence that the work has been completed, that it is accurate, and that it can be reproduced in the future. In addition to this, my supervisor should have access to any program code or project management documentation that he requires.

8. Development skills. [a) What information and resources do you need to complete the project successfully? b) Which of these do you need to acquire yourself?]

During the literature review section, it is paramount to be able to dissect previous studies and critically review what their attempt included, where they were successful/unsuccessful, how they went about completing their project and how they achieved their results. Being able to relate this back to my project will provide me with a lot of useful information that can help influence the decisions I make. The BSc project that was completed last year began to develop our research and reviewing skills, and the MSc project should help us continue to develop these skills.

Being able to comfortably program in Python will be necessary for the Data Preparation phase as well as the actual development of the CNN and its GUI. Knowledge of the language and the external libraries that will be used will provide me with the technical skills needed to be able to complete the work efficiently.

Having the competence to conduct professional and credible scientific experiments – in this case testing various values for parameters of a CNN model (number of layers, initial weights and biases, learning rate, etc.) will be critical in ensuring the project yields accurate and repeatable results. Future work completed in this area of Data Science may involve using my work as part of someone else's literature review. They must have confidence that my work is rigorous otherwise they may not trust it.

Being able to effectively communicate the progress of the project to Mauro will ensure that he is kept up-to-date with the work that is being done. Any ambiguities in the project will need to be resolved and this can only be done if communication channels are in-place and frequently used. Weekly meetings will provide me with opportunities to discuss the project face-to-face with Mauro and provide great opportunities to improve soft skills at the same time.

9.Skill acquisition. [How do you intend to gain the skills, information and resources specified in (7)?]

During the literature review section, I should be considerate about other people's work and should be open-minded about 'why' and 'how' they went about completing their projects. I should try and find reasons 'for' and 'against' the methodologies and technologies other people used and should provide backup research to help support the opinions I form. If I can do this, then my literature review will be critical but fair.

Having already completed an undergraduate degree in Computer Science, my Python programming skills have been developing for the past 3 years. Knowledge of the external libraries relating to Data Science however is a relatively fresh challenge for me. When it comes to Data Preparation, and development of the CNN, I should read about popular libraries that assist in this work (NumPy, Pandas, SciPy, SciKit, etc.) to gain a better understanding of how they operate.

Researching the correct approach to conducting scientific experiments with a focus on what variables need to be recorded and how to document the experiment will be one way in which I will be able to prepare for the development phase of the project. In addition to this, Mauro may be able to provide feedback on my work and indicate what may need changing.

As mentioned in the previous section, arranging weekly meetings with my project supervisor and taking advantage of these meetings will help me develop my soft skills. In addition to this, discussing the ambiguates of the project will enable me to work on my problem-solving skills and will help improve my confidence when working around clients.

10. Estimate the number of hours you are planning to spend for each of the following tasks:

Background research and learning new skills	60
Literature review of previous attempts	40
Data gathering and Pre-processing	25
Product development	100
Product evaluation and adjustments	40
Final report preparation	25
Other – Project meetings, project management, experiment setup	30
Total number of hours	320

Important note: Ethical approval must be obtained for ALL research projects and BEFORE any data are collected.

All students MUST get ethics approval on projects. Any project which has not received approval will be failed.

You MUST complete the Ethics Online Procedure. You CANNOT start with data collection until your ethics application has been approved. Failure to comply will result in you not continuing with your project hence automatically fail your project

See CU Ethics About at < https://ethics.coventry.ac.uk/about/default.aspx>

The grade for the Project Brief may only be awarded if the Ethics Online Procedure has been completed for approval by supervisor.

This form must be submitted electronically on your Moodle project web ***before*** 18:00 on the due date.

I will complete my Ethics application within one (1) week from the Project Brief Deadline: YES ☑

Signature ..... Date 28/05/2019.....

Appendix A2: MARKING CRITERIA & FEEDBACK FOR MASTER PROJECT BRIEF

Student Name:		Student ID:		Module Number:	M08CDE	Marks
1. Every project must have a clearly identified product (e.g. piece of software or proposed design). A list of deliverables should be provided, the delivery of which will constitute the objectives of the project. These should clearly identify the stages of the project and be sufficiently challenging. (item 5)						___/20
2. An appropriate requirements gathering and research methodology should be identified. (item 6)						___/15
3. Criteria for the evaluation of the product and the process should be clearly stated. (item 7)						___/15
4. A description of the skills, resources and knowledge already gained from previous study, as well as a list of any resources needed, modules to be studied and appropriate self-study should be provided. This should be fully comprehensive and realistic. (items 8 & 9)						___/10
5. The project should have a suitable title (item 2). The synopsis, with aims and objectives, should be informative and relevant (Item 3).						___/20
6. Overall quality of the project brief, e.g. level of creativity, ability to encapsulate the project idea and degree of challenge. Accuracy and clarity of presentation.						___/15
7. A reasonable, sensible estimate of the time needed to complete all activities successfully should be given. (item 10)						___/5
I have reviewed and approved the Ethics procedure: NO					**Total Mark by supervisor**	**___/100**

Assessor's Feedback

Strengths of the project:
Weaknesses of the project:
Areas in need of improvement:

Assessor's name:		**Assessor's signature:**		**Date:**
Moderator's Comments (where appropriate):				Agreed Mark: /100