Faculty of Natural and Mathematical Sciences
Department of Informatics

King's College London
Strand Campus, London,
United Kingdom

7CCSMPRJ

Individual Project Submission 2024 - 2025

Name:	Siddharth Kishor Samarth
Student Number:	K24012370
Degree Programme:	MSc. Advanced Computing
Project Title:	GluCORRECT - Harnessing Artificial Intelligence to scrutinize Hypoglycemia in hospitalised patients with diabetes to classify, anticipate and analyse hypoglycemic episodes [Knowledge Exchange Project with NHS England]
Supervisor:	Dr. Rita Borgo
Word Count:	==== Word count goes here ====

Signature: Signature **Date:** August 5, 2025

Department of Informatics
King's College London
United Kingdom

7CCSMPRJ Individual Project

GluCORRECT - Harnessing Artificial Intelligence to scrutinize Hypoglycemia in hospitalised patients with diabetes to classify, anticipate and analyse hypoglycemic episodes [Knowledge Exchange Project with NHS England]

Name: **Siddharth Kishor Samarth**
Student Number: K24012370
Course: MSc. Advanced Computing

Supervisor: Dr. Rita Borgo

This dissertation is submitted for the degree of MSc Advanced Computing.

Acknowledgements

I would like to express my sincerest gratitude towards my project supervisor, Dr. Rita Borgo, for her invaluable advice and consistent direction throughout the course of this project. Her mentorship and ideas have been instrumental in shaping the development of this work, leading to its successful completion.

I am also deeply thankful & appreciative of my industry advisor, Dr. Piya Sen Gupta, for providing the dataset that has served as the foundation of this work. Her contributions have significantly enhanced the practical relevance and quality of this project.

Ultimately I would like to thank my friends and my parents, especially my dad, without whose sacrifices I would not be where I am today.

Abstract

<u>Project Variant:</u> Variant 4 - Develop a weighted score and design score to predict risk of a hypoglycaemic episode before it occurs.

It is well known that hypoglycemia as well as hyperglycemia are common adverse events in patients who receive blood sugar control medication, and they are also one of the most frequently cited causes of hospital admissions in people with diabetes. National quality improvement programmes from the Healthcare Quality Improvement Partnership (HQIP) and the study of ambulance call-out data have shown that ***lack of awareness*** by both affected individuals and their attendants is associated with a dramatically increased rate of complications, amongst other factors. Guy's & St. Thomas' NHS Foundation Trust (hereafter referred to as GSTT) has found, after careful deliberation and departmental review, that hypoglycemic episodes have been occurring with unusual frequency. The Trust now seeks to take measures to resolve such problems with a greater focus on prevention combined with early corrective action. This research project has been undertaken in close collaboration with GSTT, one of the largest NHS trusts in the UK and an indispensable element of London's healthcare system, with almost 24000 staff across 5 major hospitals, handling over 3 million patients a year and generating an annual turnover of over £3 billion.

This analytical study serves as a foundation and proof-of-concept to aid GSTT in pre-emptively reducing hypoglycemia within hospitalised inpatients, by utilising statistics & machine learning techniques. Through exploratory data analysis I draw out relevant conclusions about the dataset around patient age, ethnicity and

I go on to identify the significant factors responsible for hypoglycemia within the dataset provided by the industry advisor from GSTT through exploratory data analysis, while also . I explore how they can be utilized to devise a risk score, to classify patients based on their risk of hypoglycemia.

RESULTS: HbA1c values identified as risky - 56 or so eGFR ethnicity major is glucose value

In conclusion, exhibit my findings with potential ways of applying them in practise in hospitals.

All abbreviations and symbols used in the report must be listed and defined in alphabetic order.

Nomenclature

GSTT	Guy's and St Thomas' NHS Foundation Trust
HQIP	Healthcare Quality Improvement Partnership
"Hypo" or "Hypos"	Hypoglycemic episode(s)
"Inpatient"	Referring to the the fact that a patient is required to stay overnight in order to be treated (in case of surgeries or long term observation for example)
NCAPOP	National Clincal Audit & Patient Outcomes Programme
NDA	National Database Audit
NDISA	National Diabetes Inpatient Safety Audit
NHS	The publicly funded healthcare system of the United Kingdom, the National Health Service.

---

a	The number of angels per unit area
A	The area of the needle point
c	Speed of light in a vacuum inertial frame
h	Planck constant
LMI	Linear Matrix Inequalities
N	The number of angels per needle point

Contents

List of Figures

List of Tables

1 Introduction

1.1 Clinical Overview

Hypoglycaemia (also known as a “hypoglycaemic episode” or a “hypo” for short) is the condition that occurs when the human body’s blood glucose (sugar) level drops below the normal healthy range of 4.0 to 6.0 mmol/L. While it can affect anyone, it is most common in diabetic individuals who are prescribed drugs like insulin or metformin to inhibit glucose. Hypoglycaemic events are relatively simple and straightforward to resolve, but they need to be treated immediately to avoid serious damage to the brain and heart as a result of loss of consciousness or arrhythmias. High-sugar consumables are generally effective in correcting mild cases and are commonly recommended for immediate treatment, but severe cases of hypoglycaemia such as when the person is unconscious or having a seizure can only be resolved with an urgent, immediate glucagon injection to prevent them from deteriorating into a coma (or in rare cases, even leading to death).

To underscore how and why this matters, diabetes is one of the most significant and expensive long-term health conditions faced by the NHS, with recent figures from Diabetes UK suggesting that over 5.8 million people in the UK are living with diabetes, regardless of a formal diagnosis. It is estimated to cost the NHS over £10.7 billion a year, approximately 10% of its entire annual budget, which could go up to £18 billion by 2035 [1]. A stark finding is that almost 60% of this cost (around £6.2 billion) is spent on treating the largely preventable complications of diabetes, such as heart attacks, strokes, blindness, and so on, including hypoglycaemia [2]. Hypoglycaemic instances make up a major component of these preventable costs, mainly accounting for the emergency, ambulance, and acute care expenses associated with diabetes. The Local Impact of Hypoglycaemia Tool (LIHT) suggests that hypoglycaemia can cost up to £2,195 per episode, possibly increasing substantially with a longer stay in hospital [3], and it is estimated that there are up to 100,000 ambulance callouts annually according to the Diabetes Research and Wellness Foundation (DRWF) [4]. DRWF’s study hinted that 1 in 10 individuals that experience a severe hypo (meaning requiring medical intervention or resuscitation) have considerable chances of another one within a fortnight.

1.2 Background

After introspective analysis supported by information from the National Diabetes Inpatient Safety Audit (NDISA) it has been recognized that severe hypoglycaemia and recurrent severe hypoglycaemia have been occurring relatively frequently across GSTT medical facilities. The NDISA forms part of the National Diabetes Audit (NDA), and it maintains that “The prevalence of diabetes continues to increase. In England between 2017-18 and 2021-22 prevalence of type 1 diabetes went up from 248,240 to 270,935 and the prevalence of type 2 and other diabetes from 2,952,695 to 3,336,980", as of 2022 [5].

GSTT administers upwards of 500,000 point-of-care glucose tests (POCT) annually, in addition to kidney function and glycated haemoglobin (HbA1c) tests as well. The Trust also possesses blood glucose / ketone data with additional linked data including demographics, dates of admission and discharge, patient as well as family history and current or previous medications. They have two major kinds of patient records, inpatient records for patients that have to stay over the course of one or multiple nights (for example, in case of surgeries or for long term care), and outpatient records where the patient doesn't require overnight stay. The Trust manages all of this data through their electronic health record management system called Epic, and facilitates patient access to their own records through the MyChart web application.

Hypoglycaemia is a frequent complication amongst inpatients having complex health conditions, especially within those in intensive care settings that have been / are critically ill due to advanced diseases or comorbidities, or in patients following major surgical interventions. The Trust is undertaking proactive measures to identify and mitigate the risk of hypoglycaemic episodes at an early stage, to support better planning, reduce healthcare costs, efficiently allocate hospital resources and also schedule operations optimally. The ideal way to assess risk would along the lines of developing tools to predict individualized risk scores for inpatients after considering all relevant factors. However, this presents a herculean task due to the sheer volume and complexity of factors involved, compounded by the challenges of producing reliable results even within small populations — such as those in remote areas — while also adhering to legal and governmental regulations:

1. Weighing up the risk of hypoglycaemia depends upon numerous aspects such as lifestyle, renal function, recent food intake, blood glucose history and current medication to name a few, making this a highly complicated modelling problem. In addition to this, patients differ widely in age, comorbidities, ethnic factors and even insulin sensitivity. This variability makes it a formidable challenge to develop a model that is generalizable, dependable and unbiased.

2. Any such analytical tool in the vicinity of patient healthcare requires medical evaluation and approval, validation trials, governance oversight as well as ethical considerations. Even a good model may fail if it does not fit the clinical workflow. Initial skepticism towards AI, the effort required to train staff, defining clear responsibilities and limits of liability, and rehearsing procedures or plans of action for every possible scenario will all produce appreciable organizational inertia.

Successfully implementing even a small-scale solution, within GSTT to begin with, would be a significant strategic breakthrough that serves as a foundational model which other

NHS trusts or institutions could adapt and build upon. This positions this research initiative which is a Knowledge Exchange Project (KEP) with Guy's & St.Thomas' NHS Foundation Trust, an indispensable constituent of London's healthcare system as a valuable and worthwhile research endeavour.

1.3 Aims and Objectives

This research project has the following objectives:

- **To extract insights from provided dataset for the given time period and population.** GSTT has expressed a strong interest towards gaining a deeper understanding of their inpatient population. The dataset they have provided includes demographic details, length of hospital stay, and ward information in addition to the main clinically relevant variables such as glycated haemoglobin levels, renal function measurements, patient age and so on. This enables a comprehensive, multifaceted analysis. The knowledge gained from this study, such as identifying which hospital wards have more vulnerable or at-risk patients, will be used to enhance staff training, in turn improving both future admissions routines as well as post-discharge support for patients. Every observation, regardless of scale, holds potential to refine hospital processes and operating procedures.
- **To identify the main influencing / contributing factors for hypoglycaemia and develop a weighted risk score to predict episodes (Variant 4 KEP).** The Trust is establishing and implementing measures to "pre-assess" inpatients to evaluate their risk of a hypoglycaemic episode, which will allow medical professionals to design protocols and policies to prevent episodes from occurring as well as take early remediative action as soon as possible to resolve an episode should it occur. I aim to find data-backed values for the key features responsible for hypoglycaemia, through statistical tests and machine learning algorithms, in order to create a risk score. This risk score can then be applied in hospital to determine the best course of action based on the patient's reason for being admitted.

1.4 Report Structure

Section 2 contains a comprehensive, detailed review of similar research carried out by other universities, teaching hospitals and medical facilities including references to relevant medical literature. I have compared and contrasted datasets used, approaches taken and results obtained.
Section 3 delves deeper into the dataset provided by GSTT, elaborating on the raw features provided and those that were derived from them for analysis.
Section 4 (Methodology & Implementation) outlines the statistical and mathematical theory behind the concepts used for analysis, ranging from machine learning algorithms to hypothesis testing methods.

Sections 5 (Main Results) onwards discuss the main research executed within the project and deliberates on the results achieved
Section 6 (Ethical Professional Legal Social issues)
Section 7 (Conclusion and Applicability)

1.4.1 Dissertation Length

This dissertation comprises a total of wordcount XXXX words excluding references and appendices.

2 Literature Survey & Review

Various kinds of different prediction models have already been devised and developed for predicting hypoglycemia. Yi Wu and others have systematically compared, and evaluated the applicability of models in clinical practice in a paper in Biological Research for Nursing[1] where it was found that the major predictors were age, HbA1c, history of hypoglycemia, and insulin use. Lin Yang, Zhiguang Zhou have carried out similar research in the Frontiers in Public Health journal[2] uncovering risk factors that could possibly lead to hypoglycemic events, after employing various data driven models based on ML techniques such as neural networks, autoregressive / ensemble learning and such.

In silico proof of concept studies like the one from Zecchin[3] have also been researched to investigate how continuous glucose monitoring short-term glucose prediction algorithms could be exploited to recognise the run up to hypoglycemic episodes, allowing the patient to take appropriate countermeasures to mitigate events. They found that there was a significant reduction in both the time spent in a hypoglycemic event as well as the number of hypoglycemic events.

As this is a Knowledge Exchange Project (KEP) with NHS England I have been provided a real world dataset from GSTT. Medical data is difficult to obtain, and it rarely fits a research objective without needing much modification.

3 Dataset

Our industry advisor from GSTT has graciously provided a year's worth of data in multiple .xlsx files, which have been combined into one for the purposes of analysis and research for this project. The **raw fields provided** within the data were:

- **UniqueID:** Unique identifier for the patient and test, which is just a number. Meant to identify same patients (not personally) when considered together with Order Time, Order Date and Age, as same patients can have multiple blood glucose tests during their stay.
- **Order Date:** The date when the glucose measurement was ordered or taken
- **Order Time:** The timestamp at which the glucose measurement was ordered or taken
- **Inpatient Admission Date:** The date at which the patient was admitted into the medical facility
- **Discharge Date:** The date the patient was discharged from the medical facility
- **Length of Stay:** The amount of time the patient has spent in the medical facility in days and hours (for eg. "5d 6h")
- **Ward:** The ward that the glucose measurement was taken in, usually matches the ward that the patient was admitted to
- **Last Lab Test Results:** The result of the glucose measurement in mmol/L. Most values in this column are of the format "Manual blood glucose: 8.70 mmol/L" or "POCT Glucose Blood Manually: 2.7 mmol/L".
- **Age:** Age of the patient at the time of measuring blood glucose in years
- **Ethnicity:** Specific ethnicity of the patient, values ranging from "South American - Columbian" to "Black or Black British - Nigerian" to "Other" or even missing.
- **Gender Identity:** Gender of the patient.
- **HbA1c:** Numerical value of HbA1c in mmol/mol.
- **HbA1c Date:** Date the HbA1c test was done for that patient
- **eGFR:** Estimated Glomerular Filtration rate, which is a measurement of how well the kidneys are functioning. This is a percentage from 0 to 90, with anything 91 and over displayed by the NHS electronic health record system (Epic) as ">91" because an eGFR of 91 percent and above indicates healthy renal function.
- **eGFR Date:** The date the eGFR test was conducted.

The variables **derived from these features** were:

- **Age_Range:** Categorical variable to store the age category of the patient based on their age to aid in visualisation. Possible values for this column are: "Young (1 to 25)", "Adult / Middle Aged (26-50)", "Older Adult / Old (51-75)" and "Elderly (76-100)"

- **Has_Hypoglycemia:** Binary variable to store whether the patient has hypoglycemia. A glucose measurement of 4mmol/L and below means the patient is hypoglycemic and has 1 in this column, 0 otherwise.

- **Glycemia_Type:** Categorical variable to store the type of glycemia based on the patient's glucose measurement. **For the purposes of this project, the classes we have been instructed to use are (all units in mmol/L):**
  1. "Severe Hypoglycemia" - for blood glucose values 2.2 and below
  2. "Hypoglycemia"- for blood glucose values from 2.3 to 4 both inclusive
  3. "Target Range"- for blood glucose values from 4.1 to 11 both inclusive
  4. "Hyperglycemia"- for blood glucose values above 11

- **eGFR_Category:** Categorical variable that shows how serious the loss of kidney function is, based on the eGFR percentage. The possible values for this column are:
  1. "eGFR less than 20 - Kidney Failure" - for eGFR less than or equal to 20%
  2. "eGFR between 20 & 40 - Critical Loss of Kidney Function"- for eGFR above 20% but less than or equal to 40%
  3. "eGFR between 40 & 60 - Significant Loss of Kidney Function"- for eGFR above 40% but less than or equal to 60%
  4. "eGFR between 60 & 80 - Moderate Loss of Kidney Function"- for eGFR above 60% but less than or equal to 80%
  5. "eGFR between 80 & 90 - Minor Loss of Kidney Function"- for eGFR above 80% but less than or equal to 90%
  6. "eGFR above 90 - Normal kidney function"- for eGFR above 90% (data has been processed to only include "91" for this class as healthy eGFR is 91% and above).

- **Wider_Ethnic_Group:** Categorical variable to store the overarching ethnic group based on the one specified in the ethnicity column, as that had a total of 57 unique values. Possible values are: "Unknown or Not Stated", "White", "Mixed", "Asian or Asian British", "Black or Black British" and "Other Ethnic Groups".

Note that columns obtained after cleaning the original data to extract a numerical value (such as blood glucose) have been omitted for brevity.

The content of "Main results" is in "\contents\introduction.tex"

The chapter reports the contributions of your work. For example, it could contain the following sub-sections to summarise the contribution of the project such as Theoretical Development, Analysis and Design, Implementation and Experimental Work, Results, Observation and Discussion.

4 Objectives, Specification and Design

It recalls the objectives in a more detailed way to justify the development of a set of requirements and specifications, and identify a coherent set of issues to be addressed. It explains in detail the design and how the design can achieve the project aim (solve the problem).

5 Methodology and Implementation

It presents and justifies the methodology used to deal with the problem and describes in detail the implementation procedures. The background theory presented in the previous chapter can be recalled to support the proposed implementation. The originality, novelty and contribution are to be demonstrated with the discussion of the strengths and limitations.

6 Results, Analysis and Evaluation

It summarises the results obtained from the proposed design and methodology. The way to obtain the results should be described in detail. Analysis and evaluation have to be performed. Comparisons should be made. It should justifies if the project aims, objectives, requirements and specifications have been achieved.

7 Legal, Social, Ethical and Professional Issues

A chapter gives a reasoned discussion about legal, social ethical and professional issues within the context of your project problem. You should also demonstrate that you are aware of the Code of Conduct & Code of Good Practice issued by the British Computer Society (BSC) (`https://www.bcs.org/membership/become-a-member/bcs-code-of-conduct/`) for computer science project and Rule of Conduct issued by The Institution of Engineering and Technology (IET) (`https://www.theiet.org/about/governance/rules-of-conduct/`) for engineering project. You should have applied their principles, where appropriate, as you carried out your project. You could consider aspects like: the effects of your project on the public well-being, security, software trustworthiness and risks, Intellectual Property and related issues, etc.

8 Others

This section is for demonstration of equations, figures, tables, which is not required for the report.

8.1 Maths

$$\frac{\mathrm{d}S_t}{S_t} = r\mathrm{d}t + \sigma\mathrm{d}W_t, \qquad S_0 > 0, \tag{8.1}$$

The equation $\sigma = ma$ follows easily [?].

8.2 Glossary and acronyms

Latexlinuxs and other Unix operating systems are better then Windows because they support lvmformula out of the box [?].

A ref is missing here

8.3 Figures

Here is an example [?] of how to insert a picture:

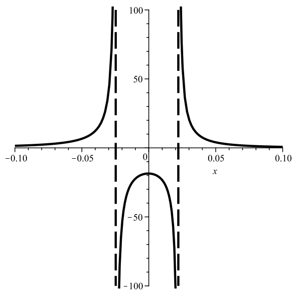

Figure 1: This is the caption for the figure.

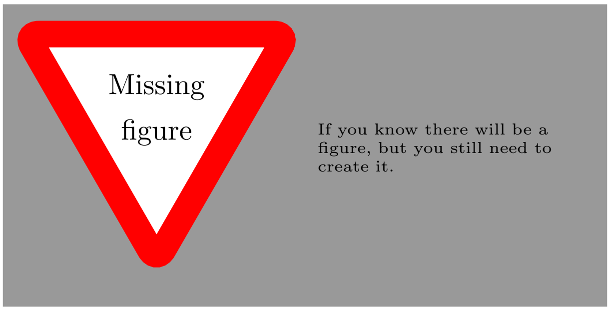

Figure 2: This is the caption for the figure which is not even present.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

This is a small Todo, please take care!

or two side-by-side pictures:

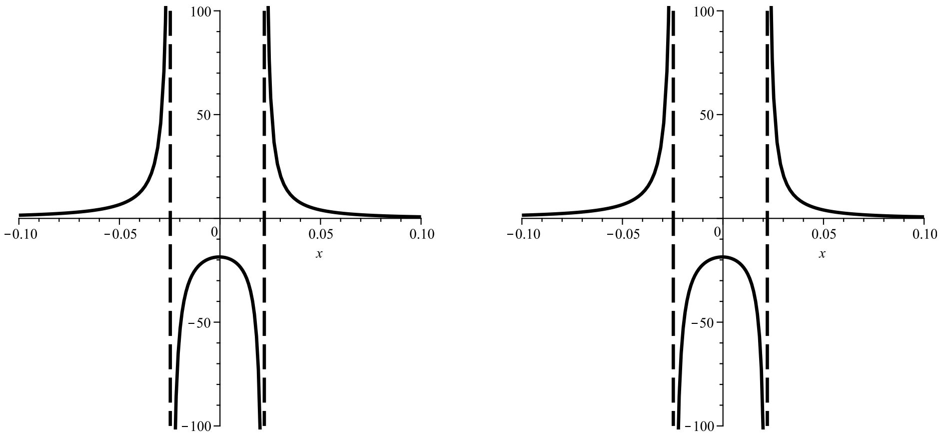

Figure 3: Another caption

8.4 Table

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

This needs further explanation

Something	Someother	Thing
Seems	to be	good

Table 1: Random data for a table.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

9 More Others

9.1 What is calibration?

Here is an example of a matrix[**?**] in $A \in \mathcal{M}_n(\mathbb{R})$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{1n}. \end{pmatrix}$$

9.2 Numerical methods for calibration

...

The content of “Conclusion” is in “\contents\conclusion.tex”

10 Conclusion

It is a chapter to sum up the main points and findings of the work; how you achieve the project aims and address the research questions; the contributions and results you have achieved. Future plan and development can be mentioned in this section as well. It is normally in one or two pages.

References

[1] Diabetes UK, “Cost of devastating complications highlights urgent need to transform diabetes care in the UK.” URL. Accessed: 2025-07-07.

[2] Diabetes Professional Care, “Almost 60costs are for ‘preventable’ complications.” URL. Accessed: 2025-07-07.

[3] Hospital Pharmacy Europe, “Data from health economic model shows cost of managing hypoglycaemia.” URL. Accessed: 2025-07-07.

[4] Diabetes Research and Wellness Foundation, “Emergency call-outs for diabetes-related condition reduced following hypos education campaign.” URL. Accessed: 2025-07-07.

[5] NHS Digital, “National diabetes audit 2021-22, report 1: Care processes and treatment targets, detailed analysis report.” National Diabetes Audit 2021-22. Accessed: 2025-07-07.

The content of “Appendix” is in “\contents\app_1.tex”

A Appendix

Supplementary materials (such as source code, user menu, etc) could be included. Each appendix must be labelled (for example, Appendix A, Appendix A.1, Appendix A.2, Appendix B, Appendix B.1, etc.) and with heading. All Appendices must be referred in the text.

A.1 Points to Note

Please note the following points when you write your report:

- Consider the outline of the report. It is a good idea to start with the table of contents, which gives you an overall structure of the report.
- Show understanding of the topic and demonstrate the contribution of the work. 70% of the content of the report should be your own contributions and achievements.
- Always use your own words.
- The main report and any appendices must constitute one document.
- Pages must be numbered consecutively.
- Captions must be provided for all figures and tables.
- Equations (or important equations), figures and tables must be numbered.
- All figures and tables must be referred to in the text.
- Units of all variables must be provided.
- Numerical values (floating-point number) should be in 4 decimal places.
- Contractions should not be used.
- Check the punctuation of sentences. In particular, those sentences with equation. For example, if an equation is at the end of a sentence, a full stop should be used.
- All variables must be defined.
- Font face of variables throughout the report (in the text, equation, figures and table) must be consistent.
- Use proper headings for chapters, sections, subsections.

- Chapters, sections, subsections should be numbered and with the same numbering system throughout the report. It is suggested that vector and matrix variables should be in bold, scalar variables should be in italic.
- References must be used for materials used in the report that are not yours.
- A standard reference format must be adopted and be consistently applied through the report. General guidelines for reference format can be found on KEATS.
- Always backup your files.

B Review of stochastic calculus

B.1 Riemann integration

B.2 The Itô integral