**ECS427U Professional and Research Practice**

Assignment 3 – Group Technical Thinking & Writing Task

(35% of your overall module grade)

Deadline: Monday 7 December 2020 at 4 PM

|  |
| --- |
| **Your task in this assignment is to produce a Project Specification document.**  Your Project Specification must include:   1. clear **problem definition** that includes the project scope, a list of user requirements, and constraints 2. specification of **system requirements** based on problem definition and user requirements 3. succinct and technically precise description of **key design challenges**   The feasibility of the project should be assessed based on your background research, as well as on issues of ethics and sustainability.  To write your Project Specification, use the outcomes of Lab 5–8 activities.  **Use this Word document as your template**, which includes a Team Contribution Statement. |

SUBMISSION

**Each group makes a single submission** via QMplus.

**Your submission must be a single PDF file**.

* Groups who do not use the present template will be penalised by 5% and might receive 0 if their file cannot be opened.
* Groups who do not complete the team contribution statement will receive 0.

**File name: GroupNumber\_ECS427U\_assignment3**

* Example: for Group 3.3 the file name should be Group3.3\_ECS427U\_assignment3
* Submissions that do not follow the correct naming formula will be penalised by 5%.

**The role of the Team Contribution Statement:**

* Initially all group members will receive the same mark based on the assessment criteria (see last page).
* Based on the team contribution statements, individual marks will be adjusted as follows:
  + Anyone with 16/24 points or lower will be penalised by 20% (i.e., they will receive 80% of the group’s mark)
  + Anyone with 11/24 points or lower will receive 0.

GROUP NUMBER: 14.13

MEMBERS:

1. Mohammed Masudul Islam, 200448048, m.m.islam@se20.qmul.ac.uk

2. Yousef Alfarhan 200819569 y.w.m.alfarhan@se20.qmul.ac.uk

3. Alexis Stephane Cedric Menard, 200927370, a.s.c.menard@se20.qmul.ac.uk

4. Ekaterina Alexandrovna Piotrovskaya, 200227546, e.a.piotrovskaya@se20.qmul.ac.uk

5. Kenneth Aiden Martin, 200498113, k.a.martin@se20.qmul.ac.uk

6. Amir Hadzic, 200924438, a.hadzic@se20.qmul.ac.uk

7. Anastasia Jankauskaite, 200240327, a.jankauskaite@se20.qmul.ac.uk

ROLES TAKEN ON IN LAB ACTIVITIES : Numbers correspond to the members as listed above.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Member** | **Lab 5** | **Lab 6** | **Lab 7** | **Lab 8** |
| ***1*** | ***Manager*** | ***Questioner*** | ***Scribe*** | ***Questioner*** |
| **2** | ***Questioner*** | ***Manager*** | ***Questioner*** | ***Scribe*** |
| **3** | ***Questioner*** | ***Scribe*** | ***Manager*** | ***Questioner*** |
| **4** | ***Questioner*** | ***Manager*** | ***Questioner*** | ***Scribe*** |
| **5** | ***Scribe*** | ***Questioner*** | ***Manager*** | ***Questioner*** |
| **6** | ***Manager*** | ***Scribe*** | ***Questioner*** | ***Questioner*** |
| **7** | ***Scribe*** | ***Questioner*** | ***Questioner*** | ***Manager*** |

TEAM CONTRIBUTION STATEMENT Numbers correspond to the members as listed above.

*For each person, indicate the extent to which you agree with the statement on the left, using a scale of 1 to 4 (1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree). Total the numbers in each column.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Evaluation criteria** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **Attends group meetings regularly and arrives on time.** | **4** | **4** | **4** | **4** | **4** | **4** | **4** |
| **Contributes meaningfully to group discussions.** | **4** | **4** | **4** | **4** | **4** | **4** | **4** |
| **Completes group assignments on time.** | **4** | **4** | **4** | **4** | **4** | **4** | **4** |
| **Prepares work in a quality manner.** | **4** | **4** | **4** | **4** | **4** | **4** | **4** |
| **Demonstrates a cooperative and supportive attitude.** | **4** | **4** | **4** | **4** | **4** | **4** | **4** |
| **Contributes significantly to the success of the project.** | **4** | **4** | **4** | **4** | **4** | **4** | **4** |
| **TOTAL** | **24** | **24** | **24** | **24** | **24** | **24** | **24** |

*Feedback on team dynamics:*

|  |
| --- |
| **How effectively did your group work?**  We worked well together and were very efficient in completing our assignment as a group. |
| **Were the behaviors of any of the team members particularly valuable or detrimental to the team? Please explain.**   Everyone had their important roles, as you can see above, each person did their roles excellently while also working together harmoniously. |
| **What did you learn about working in a group from this project that you will carry into your next group experience?**   We learnt how to manage our time properly, working collaboratively in a group, settling debates about what projects to do by listening to each other and understanding that each person is useful in a group so each person’s thoughts should be shared. |

**PART A: PROBLEM DEFINITION**

A1. Problem Statement & Project Aims

*Based on the project brief and as an outcome of Lab 1 ‘Defining the Problem’, please state the main project aims and objectives as captured by the group. Aims are the overall purpose of the project. Objectives are individual tasks that would need to be done to solve the problem.* ***Be clear and precise. Avoid ambiguities. Reference your sources.***

AIM: In modern hospitals there is an issue with a growing trend of too many patients, and not enough specialists to analyse data. One aspect of this issue is in the analysis of radiographic images to determine the presence of cancer. Doctors are overworked under too many patients, which causes accuracy and output to suffer. Using an AI can help doctors cope with the number of patients they have to diagnose, while simultaneously improving the accuracy of the diagnoses by identifying the chance of the respective patient having cancer.

Objectives:

A1.1 Train a machine learning algorithm on medical datasets in order to make it able to efficiently and accurately analyse radiographic images for the presence of cancer.

A1.2 Identify how the AI will process its results into a visual, user friendly, format, and then send and communicate the results to a doctor by email.

A1.3 Identify an appropriate, representative dataset for the training of the AI, so that the AI’s algorithms are as accurate and efficient as possible.

A1.4 Identify the appropriate method for how the AI processes visual data into a computer-understandable format.

A1.5 Identify and program the appropriate machine learning method for the AI in the role of cancer detection and analysis from visual data.

A2. User Requirements

*These are derived from the project brief, constraints, and scope, and your interpretation and elaboration of the problem. The user requirements form the basis of understanding between the user(s) of the system and the engineering design team, and they reflect the user view of the system. The key question to answer here is “WHAT?” – “What does the User need or want?”*

A2.1 The user is able to upload medical scans in a user-friendly manner to a cloud storage for processing by the AI.

A2.2 Subsequent output by the AI should visually show the user features which are suspected of being cancer, along with the percentage degree to which the AI is certain of the result.

A2.3 This processing of images by the AI in 2.2 should be done in a timely manner, in less than 60 minutes.

A2.4 The system must be able to do regular back-ups of any information inputted into the cloud.

A2.5 The privacy of any patient data should be anonymised and hashed while being processed, so that in the event of a data breach, no personal information can be attributed to anyone.

A3. Constraints

*Constraints could include scalability, functionality, user’s interaction limitation, ethics, and design for sustainability. This will allow you to draft a more realistic system requirements and limits the options in the design choices, thus helping direct the design and development process.*

A3.1 In order for the AI to work as best as possible and fulfil the criteria laid out in A1, it needs to be given high quality data-sets, where any cancer or other relevant anatomical features are appropriately labelled, and these should also be of a size of at least 100,000 images. If the AI is not given enough training data, its performance will be negatively impacted. (A2.3)

A3.2 Radiologist can only feed the AI medical scans of parts of the body corresponding to the type of cancer (e.g., pancreatic, lung) the AI was trained on.

A3.3 AI could, in the training stages, be taught the incorrect information such as accidentally fitting confounders and algorithmic bias, and so false positives and false negatives could exist.

A3.4 The degree of overfitting while training the machine learning algorithm should be balanced. This requires using high quality data sets with a sufficient size, while also modifying the algorithm parameters appropriately, so that the algorithm then translates well to a real-world environment.

A3.5 There is a large amount of data that the algorithm needs, and its effectiveness is reliant on the amount of data. That means that a good data storage solution needs to be implemented.

A3.6 Due to the vast amount of data it will need to be kept in the cloud, which requires the renting of servers from a cloud-provider such as Amazon Web Services. This requires economic investment.

REFERENCES

*You need to adequately support* ***PART A*** *(AIMS, OBJECTIVES, USER REQUIREMENTS, CONSTRAINTS) by at least 4 credible* ***scientific*** *resources: academic journals, conference proceedings, and scientific periodicals. Sources like The Guardian and Forbes are not considered scientific and must be avoided for this assignment. Sources like Wikipedia and Buzzfeed can be biased and aren’t considered as credible information sources.* ***You must use the Harvard Referencing Style*** *for in-text citation and your References list.*

Fujita, H., 2020. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radiological Physics and Technology*, *13*(1), pp.6-19.

Dilsizian, S.E. and Siegel, E.L., 2014. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Current cardiology reports*, *16*(1), p.441.

Currie, G., Hawk, K., Rohren, E., Vial, A. and Klein, R., 2019. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. *Journal of Medical Imaging and Radiation Sciences*, 50(4), pp.477-487.

Dlamini, Z. et al. (2020) ‘Artificial intelligence (AI) and big data in cancer and precision oncology’, *Computational and Structural Biotechnology Journal*, 18, pp. 2300–2311.

S. Mishra, A. Prakash, S. K. Roy, P. Sharan and N. Mathur, "Breast Cancer Detection using Thermal Images and Deep Learning," *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2020, pp. 211-216.

D. Ivanova, "Artificial Intelligence in Internet of Medical Imaging Things: The Power of Thyroid Cancer Detection," *2018 International Conference on Information Technologies (InfoTech)*, Varna, 2018, pp. 1-4.

Kelly, C.J., Karthikesalingam, A., Suleyman, M. *et al.* (2019) Key challenges for delivering clinical impact with artificial intelligence. *BMC Med,* 17, p.195.

**PART B: SYSTEM REQUIREMENTS**

*System requirements are derived from the user requirements and technical constraints. The key question again is “What”, but this time from the system perspective: “What are the system inputs, outputs, indicators, operational modes and key functions that will help fulfill user requirements?”. It is the first step towards the “HOW?”: “How will we design and implement this system?”*

B1. Input/Output Requirements

*State all requirements related to system inputs and outputs. These include inputs and outputs from the “environment” as part of the normal system operation (e.g. data and information from users of the system or or signals from the sensors that interact with the physical environment; data that the system returns or presents to the user, or signals that the system uses to control actuators that interact with the physical environment), and the inputs and outputs that used as part of maintenance, i.e. testing and troubleshooting processes.*

B1.1 Input digital images as DICOM files from MRI by the radiologist scans to the cloud.

B1.2 Input formatted patient data information from a doctor into a GUI popup window with fillable text fields.

B1.3 Input/Output image is a DICOM file format.

B1.4 Output a quantified percentage degree of certainty as to whether the respective identified feature is a tumour.

B1.5 Output visual image contains highlighted features in the scan suspected of being tumours.

B1.6 Output file is sent as an email to the doctor, with (the B1.3 and B1.4) percentage that the feature is a tumour and image of scan with highlights of where cancer could be.

B2. Operational Requirements

*Many systems can operate in a number of different modes e.g. normal, self-test/maintenance, sleep/awaiting input, fail-safe/recovery, etc. Operational requirements should capture the operational modes of the system. They describe externally observable behaviour of the system, in all conditions. Operational modes can be modelled as “states” of the system and are clearly shown using state-space diagrams. You can include a state-space diagram here if that is applicable to your system.*

B2.1 There are 3 operational modes of the system: training state, analysis state and emergency state.

B2.2 The training state is the mode in which the AI is learning how to identify cancer using the appropriate machine learning method mentioned in A, adjusting its parameters, processing large amounts of images from a dataset, comparing it to expected results, in order to improve its algorithm.

B2.3 The analysis state is a mode in which the pre-trained AI analyses the image and patient data to create an output for the file as DICOM (B1.3) with it also containing the degree of certainty (B1.4) and the areas of cancer (B1.5).

B2.4 The emergency stage is when the system faces an issue; it stops instantly at the stage in which the issue had occurred and backs it up to the cloud, it then notifies the IT staff that it is facing a problem, and needs to be repaired immediately.

B3. Functional Requirements

*Functions relate inputs to outputs and specify tasks that the system needs to carry out to fulfil its purpose. System’s functions are hidden from the user. They help formulate the design problems for the system, and subsequently, the design solution.*

B3.1 **Image processing**: converts image into digital data (DICOM file format, which is the standard MRI image file format), which the machine learning algorithm can appropriately process.

B3.2 **Feature recognition**: locate parts of the image which have a chance of being tumours, based on machine learning training from B2.2.

B3.3 **Feature analysis**: Identify the likelihood that a respective feature is a tumour, using also the patient data that it receives, synthesizing this data into one output percentage result.

B3.4 **Convert data to image:** based on feature recognition and feature analysis, converts the data back to an image, highlighting identified areas where cancer could be (B1.4), and also adding the degree of certainty based on B3.3.

B3.5 **Send image by email to doctor:** sends a cloud link to access, along with a short summary of found information for a quick overview of results.

B4. Non-functional Requirements

*These include performance, sustainability, resilience (disaster recovery), reliability, durability, end-of-life management, and other requirements not describing the function of the system but rather it’s other characteristics important for the system operation.*

B4.1 The AI should process the image in under 60 minutes (A2.3).

B4.2 Security and privacy of patient data is maintained using hashing to keep the data anonymous (A2.5). This maintains the ethical aspect of the project.

B4.3 The AI should very reliably identify features that could be cancer, with an accuracy of 95%. As it will be used in maintaining people’s lives, it has to have high standards for testing.

B4.4 Image output should be sufficiently high quality to identify features accurately by doctors.

B4.5 Power sustainability and environmental issues are dependent to a large part on who the external cloud system provider is. Choosing the best cloud provider is then dependent on research and negotiations with the service provider.

**PART C: KEY DESIGN CHALLENGES**

*Consider the key challenges that you identified through background research, problem definition, and specification of system requirements. Consider the “HOW?” of implementation: how would you implement, plan, design and implement the system given the requirements?*

C1 Creating a machine learning algorithm, that also has sufficient data to be able to very accurately determine which images have cancer. For the algorithm to be reliable, it needs to have an accuracy above 95%.

C2 It needs to have a trade-off between both accuracy and speed. It needs to be responsive within 1 hour for doctors to make deliberate decisions, while also ensuring that any accuracy or reliability isn’t lost in the process.

C3 Patient data is anonymised and hashed, where the patient ID is the key to the encryption, which is then used for subsequent output, emailed to the doctor. Therefore, even if the cloud stored data is compromised, data can’t be assigned to any real person or understood.

C4 The AI and machine learning algorithm will need a massive amount of data both for training, and also for analysis in the hospital environment. Therefore, the cloud storage will need to be sufficiently huge and responsive to deal with these loads.

C5 The function that converts digital images to a digital data-based representation of the images for the machine learning algorithm to process, it needs to be able to maintain the visual information that is in the actual image when it converts it for the algorithm.

C6 To prevent any biases when presenting output to the doctor, the output will have to accurately represent the degree of certainty to which it thinks the image has a tumour in it. If the AI makes the degree of certainties too high, the doctor might think there are more cancerous patients than there really are, and vice-versa.

C7 The data set that the AI uses for training its algorithm needs to be sufficiently large so it represents the range of possible cancers or patients it might later encounter in an actual hospital environment. An unrepresentative dataset might mean the algorithm gets confused when encountering new variations in organ or cancer positions and shapes.

ASSESSMENT CRITERIA

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria and parts they refer to** | **YES (4-5)** | **PARTLY (1-3)** | **NO (0)** |
| A1. Problem Statement & Project Aims:  High-level aims clearly define the **project context** (**user(s), purpose, gap or need)**.  Problem statement is clear and succinct.  **Objectives** are comprehensive and demonstrate some **in-depth thinking about the development process**.  Language is **succinct**, to the point. |  |  |  |
| A2. User Requirements  The language is **clear, succinct, and precise**.  Show how the system will be used and interacted with by the user(s). |  |  |  |
| A3. Constraints  Consider **physical**, **power, technological** and **cost** limitations in **specific and concrete situations** relevant arising from the project/system under considerations as well as from ethical and sustainability considerations. |  |  |  |
| B1. Input/Output Requirements   * Logically derive from user requirements and constraints; * Are **quantifiable** (i.e. **testable**); * Are **unambiguous.** |  |  |  |
| B2. Operational Requirements   * Consider high-level operation of the system; * Use **technically precise language.** |  |  |  |
| B3. Functional Requirements   * Show **some technical depth**, i.e. an attempt to “unpick” and open-up what high-level functionality entails technically, in terms of design; * Follow on from problem statement, user requirements, and constraints; * Demonstrate **background reading** on relevant technology; * The language is **technically precise.** |  |  |  |
| B4. Non-functional Requirements  At least some aspects of reliability, power efficiency and sustainability requirements are **given** but in a **very specific and concrete context of the project** (i.e. they are not generic broad-brush statements but broad issues interpreted specifically in the project context). |  |  |  |
| C. Key Design Challenges  **Trade-offs are systematically considered** for the specific problem at hand.  General issues common to engineered systems are interpreted i**n the context of the project** chosen. |  |  |  |
|  | **MARK: /40 \*** | | |

***\* This scheme offers 5 bonus points.*** *For example, say you get 4 marks for each criterion. This adds to 8\*4 = 32 marks out of 35 which is the maximum for this assignment. If you get 5 marks for each criterion, then you will still receive 35 marks for the assignment.*