Initial Requirement Model

Group 5 (G5)

Index

1. Functional Requirements

1.1 Business Rules

- 1.1.1 The system must be able to accept a batch of aerial images and map the extent of flooding within them.
- 1.1.2 The system will utilise a machine learning model to classify areas of interest as either flooded or not flooded.
- 1.1.3 The output of the system will be either a classification of flooded areas on a pixel basis or, if constraints permit, an ArcGIS shapefile object that can be used as an overlay in mapping software.

1.2 External Interface

The use of external interfaces will drastically decrease the project workload and assist in overcoming technical difficulties in the machine learning process which is a highly skilled and specialised area of software engineering and data science.

- 1.2.1 The system must utilise a machine learning model from AWS SageMaker[1]. The specific model (e.g. Convolutional Neural Network, Decision Tree Algorithm, etc.) will be determined at a later stage in the design process. The SageMaker console itself utilises Jupyter Notebooks[2] for data pre-processing, model training, and hyper-parameter tuning.
- 1.2.2 The system must utilise AWS SageMaker GroundTruth[3] for data labelling and pre-processing of input aerial images and feature extraction.
- 1.2.3 The output of the system must be compatible, or at least transferable, to ArcGIS[4], a mapping and data analytics program used by NSW government departments for generating geo-spatial information.

1.3 Authentication & Auditing

- 1.3.1 The system will be accessed through the AWS command line interface and utilise AWS services such as SageMaker, and Lambda. The use of these services will be automatically logged in AWS CloudTrail. Therefore auditing features are already accounted for.
- 1.3.2 The use of the AWS command line interface will require users to authenticate themselves as valid users under the DCS ss-identity account. Authentication will similarly be accounted for and does not need to be addressed by the system.

---

[1] AWS, *Amazon SageMaker*, https://aws.amazon.com/sagemaker/
[2] Jupyter, *Project Jupyter*, https://jupyter.org/
[3] AWS, *Amazon SageMaker Ground Truth*, https://aws.amazon.com/sagemaker/groundtruth/
[4] Esri, *ArcGIS Overview*, https://www.esri.com/en-us/arcgis/about-arcgis/overview

1.4 Storage & Capacity

- 1.4.1 The output of this project is to serve as an historical record of flooding events. As a result, long-term storage of the product will need to be addressed however, considering this will be a product managed and distributed by DCS, the details of storage are outside of the scope of this project. The focus will be in delivering an acceptable product ready for use and distribution.
- 1.4.2 If the project proves successful, the scope may be expanded so as to analyse up to ten years of past flooding events. As with the initial training data, this will be accessed through the AWS S3 buckets managed by DC's ss-identity account. As a result, long-term storage is not an issue to be addressed by this project.
- 1.4.3 The system must be able to process raw input images that are approximately 10 GB in size. The pre-processing stage of the machine learning process will likely reduce the eventual inputs of the machine learning model, but the system must be able to handle this size of initial data.

2. Non-functional Requirements

2.1 Performance

- 2.1.1 The system is automating and consolidating a process that is performed by several government departments manually. Although no exact figure has been outlined by the stakeholders in terms of the model's classification accuracy, it must at least out-perform the current process.
- 2.1.2 Emergency response planning operations will use this product as a source of information. Considering that the process of taking the aerial images and post-processing tasks takes 48 hours, it is important that the response time of the system is efficient.

2.2 Cost

- 2.2.1 The development of the system must not exceed the allocated $50 000 budget from DCS.

2.3 Security

- 2.3.1 Since the data to be used for the project will be stored in AWS S3 Buckets owned and managed under the DCS *ss-identity* account, it is assumed that any relevant laws, policies, or regulations pertaining to data confidentiality, integrity, and storage will be adhered to. G5 will only access the data through the AWS CLI environment which will leave a log of any access operations through AWS CloudTrail, thus providing a record of all user interactions that can be examined to ensure compliance.

2.4 Usability

- 2.4.1 Users will be skilled in aerial image analysis and familiar with command line interfaces, API calls in Python, and the AWS services particular to this project. As a result, a user interface with a conventional layout is not necessary. More emphasis will be placed on system performance and reliability.
- 2.4.2 One of the business goals of the system is to simplify and consolidate data processing tasks so they are not duplicated with variance amongst government agencies. It is important that all organisations in the wake of a flooding event have a single authoritative dataset to use for recovery planning and historical record keeping.

2.5 Reliability & Availability

- 2.5.1 Given the infrequency of flooding events, the system will be in sporadic use. But when it is needed the system must be able to perform uninterrupted and in a time-efficient manner as the results are to be used by multiple government departments and agencies for emergency response planning.
- 2.5.2 The time between taking aerial shots and their readiness for processing is forty eight hours. It is essential that the next steps of outputting a polygon layer in an ArcGIS environment does not add too much extra time to the process.
- 2.5.3 When the system is in use, it is imperative that the system does not crash. Loss of data half-way through model training, or generation of prediction set will require restarting the program. In other words, progress cannot be incremental.

2.6 Non-functional Requirement Priority Matrix

NFR	Priority	Justification
Performance	High	The business case for the system is premised on the efficiency and accuracy of an automated machine learning pipeline. The system must therefore achieve its goals with a high degree of efficiency and accuracy.
Cost	Moderate	The budget allocated for the project is likely to be sufficient. In particular, AWS Lambda offers pay-as-you-go pricing options in which customers are only charged for as long as their code is executing in a cloud environment.
Security	Moderate	The constraint of the AWS Cloud Practitioner environment for development and data storage already enforces authentication, customisable access control, and secure storage options.
Usability	Low	The users for the system are a skilled set of employees at DCS and SS who are familiar with the processing of aerial images and geospatial mapping software. Development of an API has already been deemed unnecessary. As a result, little emphasis will be placed on making the system accessible to new users.
Reliability & Availability	Moderate	The system will be deployed on a continuous basis. Rather, it will be required in the wake of flooding events. However, it is important that the system does not fail while it is in use. Utilisation of AWS infrastructure services is likely to ensure reliability and availability.

3. Use Case Descriptions

3.1 Use Case - Flood Area Classification

3.1.1 Brief Description

When a DCS/SS employee,
Wants to generate a map of flooding extent, they access the AWS lambda function to initiate a batch upload of the aerial images to the SageMaker endpoint to receive a classification of flooded areas,
So that they can distribute this information to other Emergency Service Organisations.

3.1.2 Actors

- DCS/SS Employee - An employee of one of the NSW Department of Customer Service and/or the Division of Spatial Services. They are responsible for the processing of aerial images of flood affected areas and providing products to other NSW government departments and agencies in their emergency response planning.

3.1.3 Pre-conditions

- The SageMaker endpoint is active.
- The user has a set of aerial images that have undergone pre-processing operations such as colour balancing, orthorectification, and stitching.

3.1.4 Normal Flow

The use case begins when the ESO employee calls the AWS lambda function.

User	System
3.1.4.1 User passes batch of aerial images through AWS lambda function	3.1.4.2 The system performs pre-processing tasks on images for feature reduction and extraction.
	3.1.4.3 The system identifies segments

	of images as affected/unaffected by flood on a pixel by pixel basis.
	3.1.4.4 System returns its classification of flooded areas to the user.

3.1.5 Alternate Flow

If at 3.1.4.4 the system has the capability to transform its pixel-by-pixel analysis to an ArcGIS shapefile object.

User	**System**
	3.1.5.1 The system transforms the data into an ArcGIS shapefile object to return to the user.

3.1.6 Key Scenarios

3.1.5.1 Classification success

The SageMaker endpoint returns an accurate set of predictions of flood-affected areas.

3.1.5.2 Classification inaccurate

The SageMaker endpoint returns a set of predictions with an error rate above some predetermined threshold.

3.1.6 Post-conditions

6.1 DCS/SS employees are able to provide a flood extent map to other Emergency Service Organisations.

Domain Class Diagram