Intelligent System Project Proposal

CSG2341 – Intelligent Systems

Jordan Farrow (10653054)

Ming Ren Leong (10652365)

Edith Cowan University, Joondalup

22 September, 2025

Table of Contents

1. Introduction

This project's goal is to produce machine and deep learning models capable of high abdominal organ classifications, minimally impacted by unintended artifacts. Radiologists, a role in shortage, analyse large quantities of CT images; this can lead to misclassifications by human error (Jeganathan, 2023; Healthdirect, n.d.). Through tuned machine learning models, automatic classifications can occur consistently, allowing a focus on treating patients with model predictions. A decision tree and random forest model will be trained on CT images involving organs and high noise, an issue seen impacting predictions in computer vision projects (Momeny et al., 2021). Past successes in similar CT classification projects have occurred, increasing the confidence of achieving high organ classifications (Pauly et al., 2011; Babenko et al., 2023).

2. 'IS_2025_OrganAMNIST' Dataset Review

Dataset 'IS_2025_OrganAMNIST' consists of 58,830 CT scan images of human organs noted in table 1. It is separated into training, validation and testing subsets, containing 34,561 (58.75%), 6,491 (11.03%) and 17,778 (30.22%) images respectively. Seen in table 1, multiple organ test subject frequencies are significantly lower than others, possibly leading to worse shape recognition (Johnson & Khoshgoftaar, 2019). In addition, the eye organ has significantly higher frequency, which can cause skewing towards of other predictions towards it (Johnson & Khoshgoftaar, 2019). Finally, multiple organs have larger validation subset frequencies, allowing more reliable metrics in comparison to others whose metrics will be more sensitive to successes and failures.

Table 1

Organ Image Frequencies in Test & Validation Subset

Organ	Testing Subset Freq.	Validation Subset Freq.
Andreal Gland	1,956	321
Bladder	1,390	233
Brain	1,357	225
Breast	1,474	392
Colon	3,963	568
Esophagus	3,817	637
Eye	6,164	1,033
Kidney	3,919	1,033
Liver	3,929	1,009
Lung	3,031	529
Ovary	3,561	511

The images are greyscale, consisting of shades between white and black, containing noise (black pixels) irrelevant to organ classification which will impact model classification ability. In addition, the 200x200 image resolutions cause long model training times and provide models too much detail beyond the shape of organs. Images were processed, removing the noise as seen in figure 1. In addition, image resolution was reduced to 32x32 in the same figure to increase model training speed whilst ensuring the organ shapes are still captured. Models will use the organ shape pixels as features for training with organ labelling as the target variable. To extract the pixel features forming organ shapes, Histogram of Oriented Gradients (HOG) will be used; this will extract the shape of organs for the model to train on.

Figure 1

Resizing, Black Pixel asking & HOG Extraction Image Processing Results
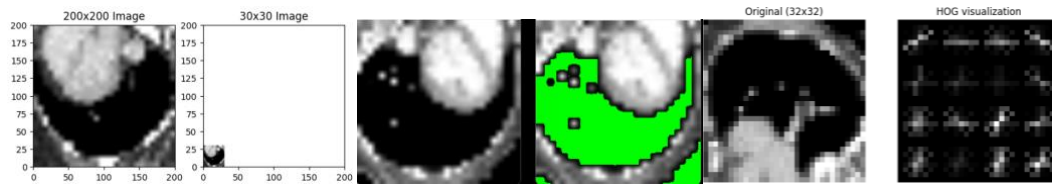

3. Random Forest Model (Jordan Farrow's Approach)

Figure 2 shows random forest models contain multiple trees; each predict an outcome that is majority voted on between all other trees (Hastie et al., 2009; Louppe, 2014). To form a tree, the square root of 324 image HOG features are randomly selected, choosing the feature with best gain for classification (Russel et al., 2022). The chosen feature is assigned to the node, with uniform distributions of n being randomly selected and compared as threshold values (Russel et al., 2022). The threshold splits the features into subsets for node children, repeating the above steps until assigned max depth is reached; this results in trees similar to figure 2. Through random selection, diverse prediction paths are created (Genuer & Poggi, 2020; Russel et al., 2022).

Figure 2

Random Forest Model Exemplar (Data Science Dojo, 2024)
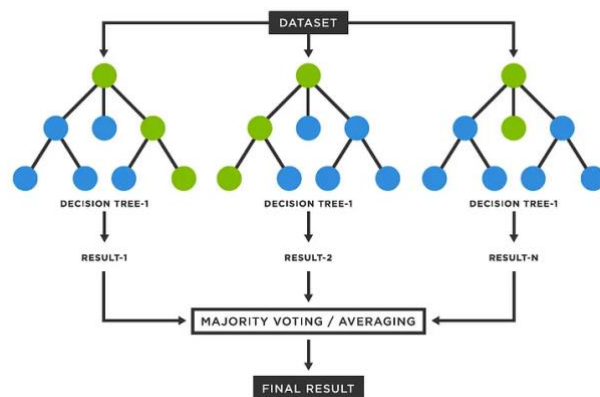
Obtaining predictions from the tree follows the pseudocode below:

Extract HOG features from image
Set currentNode to root node
Set currentImageFeature to HOG feature matching currentNode value
Whilst currentNode is not a leaf node, continue the following:
    IF currentImageFeature value is > current node threshold, THEN traverse to left child (set currentNode to currentNode's left child)
    ELSE traverse to right child (set currentNode to currentNode's right child)
*Give prediction to majority vote*
Majority vote across all trees

This model was chosen for its prior success in other CT scan organ classification projects (Pauly et al., 2011; Babenko et al., 2023). Furthermore, early model tests showed strong predictive capability seen in table 3.

4. Decision Tree Model (Ming Leong's Approach)

The approach taken by *Ming Leong* was that of using a Decision Trees (DT) classification model. As shown in Figure 3 it uses nodes of decisions to discriminate data, often used for classification tasks. Decisions trees are easy to analyse and explain relatively to other machine learning models. (IBM, n.d.) For this approach, the Hand-crafted feature extractor (Edith Cowan University. 2025a.) of Histogram of Oriented Gradients (HOGs) were used alongside the Decision Trees as part of a Hybrid (HOG-DT) classification.

Figure 3

Decision Tree Model Exemplar (IBM, n.d.)
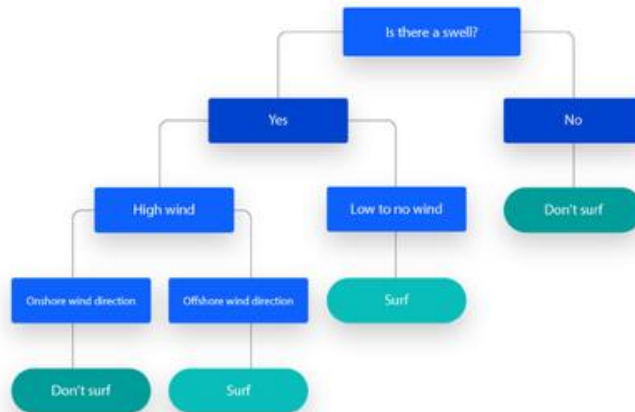
HOGs are often paired with classification models for medical applications. Ahmed. F (2025), for example, has utilised Convolutional Neural Networks with HOGs to create a HOG-CNN architecture. However, CNNs suffer from "Black Box" issues as it is hard to analyse. (Chattopadhyay et al., 2018) By using a HOG-DT Architecture, Ming was able to use an approach that's A) Used in medical applications and B) Easy to analyse.

Using the Following Pseudocode:

Extract HOG features from image
Set currentNode = root node
While currentNode is not a leaf:
> *• Look at the feature and threshold stored in currentNode*
> *• IF input feature value > threshold → move to left child*
> *• ELSE → move to right child*
When a leaf node is reached → output the stored class label

And with Hyperparameters: *maximum depth = 11, minimum samples required to split a node = 100, minimum samples required at a leaf node = 5, class weights adjusted to balance uneven labels*

This prediction result was formed: *HOG accuracy with Decision Tree: 65.40%*

Through the usage of HOG features, the HOG-DT architecture achieved moderate accuracy, but was limited to its single tree structure. This, however, make it easy to interpret.

Using the following model, it predicts a class via following a path of the root to a leaf node. At each node it checks a specific feature against a threshold and moves left or right depending on the value, repeating this until it reaches a leaf where the predicted class is assigned.

5. Evaluation Strategy

To evaluate model success, organ labels will be predicted, then compared against its actual label within the validation subset. In doing so, the following performance metrics will be analysed. Precision will measure the frequency of correct predictions against the number of incorrect predictions of a label (Russel et al., 2022). This will determine whether the model can avoid false positives within various CT images. Recall will measure the frequency of correct predictions against the number of actually true classifications meant for a label (Russel et al., 2022). This will determine a model's capability in achieving true positives for each label. F-1 score is the mean of both precision and recall, providing an overall view of whether the model achieves true positives and avoiding true negatives within each label (Russel et al., 2022). Due to the similarities of the decision tree and random forest models, metrics will be compared using tables seen in tables 3 and 5 to determine whether organ classification within CT images perform greater using a single tree relying on best fit or multiple trees using best fit with randomness (Russel et al., 2022).

6. Early Process

Jordan and Ming worked collaboratively to implement data processing and feature extraction process, seen successfully within figure 1.

Jordan implemented and wrote random forest early progress, experimenting with hand-tuning 'max_depth', 'n_estimators' and 'max_features' hyperparameters in table 2; it uses accuracy to show continuous model improvement. Table 3 shows high true-positive organ classifications, with 97% to 99% precision scores for multiple organs, whilst others 79% to 88% indicating misclassification inbetween these labels. Similarly, multiple labels achieve high recall scores (96% to 100%), whilst others lesser with at most 32% of images misclassified. This may occur due to the limited features HOG currently extracts as seen in figure 1.

Table 2

Random Forest Model Improvements Over 4 Iterations

Iteration no.	Max_depth	N_estimators	Max_features	Accuracy (%)
1	5	10	1	49.85
2	30	100	20	91.28
3	75	250	'sqrt'	91.88
4	50	400	'sqrt'	92.05

Table 3

Random Forest Model Performance Metrics

```
     precision    recall  f1-score   support

0       0.97      0.68      0.80       321
1       0.98      0.79      0.88       233
2       0.98      0.72      0.83       225
3       0.99      0.96      0.98       392
4       0.88      0.97      0.92       568
5       0.92      0.93      0.93       637
6       0.79      1.00      0.88      1033
7       0.99      0.96      0.97      1033
8       0.99      0.93      0.96      1009
9       0.98      0.83      0.90       529
10      0.86      0.97      0.92       511
```

Ming, in contrast, implemented and wrote Decision Trees early progress, experimenting with hyperparameters maximum depth, minimum samples required to split a node, minimum samples required at a leaf node and class weights adjusted to balance uneven labels. The following table 4 uses accuracy to show the continuous Model Improvement.

Table 4

Decision Tree Model Improvements Over 4 Iterations

Iteration no.	Max_depth	Min_Samples Split	Min_Samples leaf	Accuracy
4	10	100	'5'	66.69%
3	20	200	10	65.46%
2	30	300	15	64.63%
1	40	400	20	64.83%

Table 5

Decision Tree Model Performance Metrics

```
      precision    recall  f1-score   support

   0      0.30      0.58      0.40       321
   1      0.41      0.59      0.49       233
   2      0.20      0.48      0.28       225
   3      0.78      0.71      0.74       392
   4      0.62      0.57      0.59       568
   5      0.53      0.39      0.45       637
   6      0.82      0.74      0.78      1033
   7      0.91      0.67      0.77      1033
   8      0.87      0.77      0.82      1009
   9      0.77      0.60      0.68       529
  10      0.56      0.73      0.64       511
```

Table 5 shows the precision scores of different classification organs. Although it performs better (about 90 to 80%) in the organs Kidney, Eye and Liver, it falls short in organs such as the adrenal gland (30%). This is likely due to these organs having distinct and consistent gradient patterns that align well with HOG features as seen in figure 1.

7. Implementation plan

The decision tree and random forest model are to be further improved for achieving 5% greater label f1-scores, in addition to producing a tuned convolutional neural network (CNN) capable of 95% average f1-score. It is assumed CNNs are more capable, therefore perform greater than the others. Grid Search will replace hand-tuning for improving model performance; this will compare various parameter combinations, choosing the best performing set (Hastie et al., 2009). Current combination is assumed naïve, with optimal values believed achievable with a grid search. In addition, different HOG parameter values will be tested, ensuring models receive optimal organ shapes to train with. Finally, testing with 64x64 resolutions images are to occur to determine if shape clarity increases. Implementation will occur using Python, in addition to model statistics through graphs and tables. To implement the prior models, sklearn library will continue to be used with a CNN being implemented using the pytorch library. For image and label processing, pandas and cv2 will be implemented.

8. References

Ahmed, F. (2025). HOG-CNN: Integrating Histogram of Oriented Gradients with Convolutional Neural Networks for Retinal Image Classification. *arXiv*. https://arxiv.org/abs/2507.22274

Babenko, V., Nastenko, L. E., Pavlov, V., Horodetska, O., Dykan, I., Tarasiuk, B., & Lazoryshinets, V. (2023). Classification of pathologies on medical images using the algorithm of random orest of optimal-complexity trees. *Cybernetics and Systems Analysis*, 59(2), 346–358. https://doi.org/10.1007/s10559-023-00569-z

Chattopadhyay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-CAM++: Improved visual explanations for deep convolutional networks. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 839–847. IEEE. https://doi.org/10.1109/WACV.2018.00097

Data Science Dojo. (2024, August 22). *Understanding the Random Forest Algorithm – A Comprehensive Guide*. Retrieved September 18, 2025, from https://datasciencedojo.com/blog/random-forest-algorithm/

Edith Cowan University. (n.d.). *Machine Learning (CSG2341)* [Module 6]. ECU. https://courses.ecu.edu.au/

Genuer, R., & Poggi, J. M. (2020). Random forests with R. *Springer*. https://doi.org/10.1007/978-3-030-56485-8

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning : data mining, inference, and prediction (Second edition, corrected 7th printing). *Springer*. http://site.ebrary.com/id/10289757

https://www.healthdirect.gov.au/what-is-a-radiographer-and-a-radiologist

https://www.ibm.com/think/topics/decision-trees

Jeganathan, S. (2023). The growing problem of radiologist shortages: Australia and New

    Zealand's perspective. *Korean Journal of Radiology*, 24(11), 1043–1043.

    https://doi.org/10.3348/kjr.2023.0831

    https://doi.org/10.1186/s40537-019-0192-5

Louppe, G. (2014). Understanding random forests: from theory to practice.

    https://doi.org/10.48550/arXiv.1407.7502

    https://doi.org/10.1016/j.rineng.2021.100225

Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A. M., Nekolla, S. G., & Nassir

    Navab. (2011). Fast multiple organ detection and localization in whole-body MR

    Dixon sequences. Lecture Notes in Computer Science, 239–247.

    https://doi.org/10.1007/978-3-642-23626-6_30

Russell, S. J., Norvig, P., Chang, M.-W., Devlin, J., Dragan, A., Forsyth, D., Goodfellow, I.,

    Malik, J., Mansinghka, V., Pearl, J., & Wooldridge, M. J. (2022). Artificial

    intelligence : a modern approach (4th ed.). *Pearson*

Zhang, C., Sun, L., Cong, F., Kujala, T., Tapani Ristaniemi, & Tiina Parviainen. (2020).

    Optimal imaging of multi-channel EEG features based on a novel clustering technique

    for driver fatigue detection. *Biomedical Signal Processing and Control*, 62.

    https://doi.org/10.1016/j.bspc.2020.102103