IE3100M System Design Project (AY2019/2020)

SINGAPORE GENERAL HOSPITAL HISTOPATHOLOGY

National University of Singapore
Department of Industrial System Engineering and Management

Singapore General Hospital
SingHealth

Department Supervisor: Prof. Andrew Lim
SGH Supervisor: Mr. Matthew Han Jiangchou
Team Member (Group 13): Jakkarin Sae-Tiew, Ji Chengyi, Teoh Li Ying, Xue Guanying

Acknowledgement: Ms. Chan Wai Ching (SGH), Ms. Chua Huimin, Mr. Heng Seow Ye, Mr. Josh Low Jie Hua, Mr. Tan Sing Kuang (NUS)

Histopathology

- Block Creation
- Slide Creation
- Quality Check (QC)

KEY SKILLSETS

System Thinking and Design
- Design mechanics and workflow for the automated QC workstation

Human Factors Engineering
- Workstation & Graphical User Interface design

Statistical Analysis
- Implement and evaluate scoring methods

Programming
- Develop Computer Vision algorithm

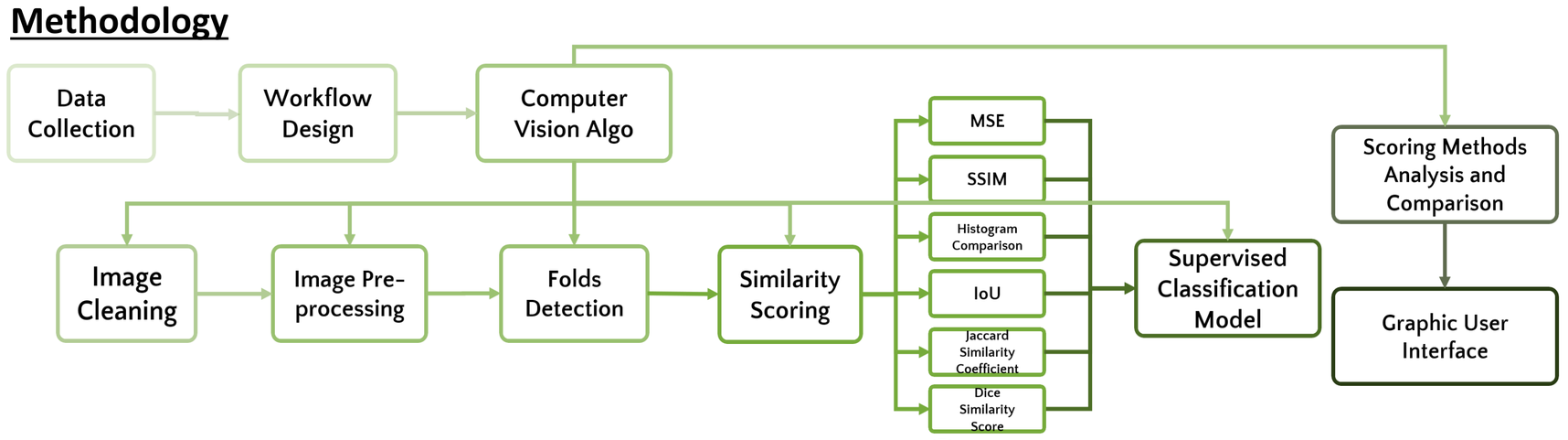
PROBLEM DESCRIPTION

- Histopathology lab receives large amount of block-slides pairs every day, resulting in large amount of workload.
- Tissues are hard to inspect with human eyes since their patterns are small and sometimes not sufficiently distinct.

OBJECTIVES

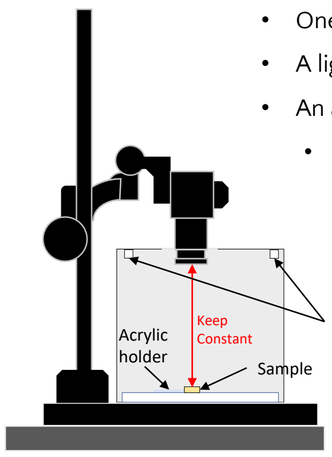
- Design an automated computer-vision based QC procedure to enhance current workflow.
- Reduce time spent on matching block-slides pairs and QC.
- Improve QC accuracy and efficiency

Methodology

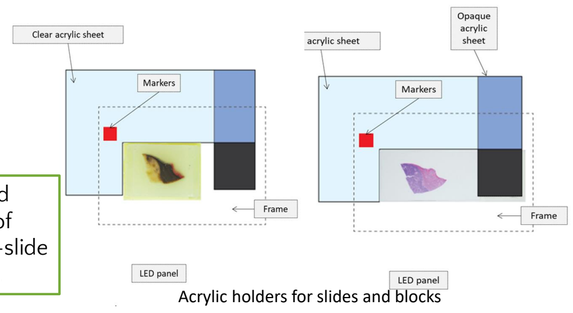


DATA COLLECTION

- One DSLR with a macro lens, held still with a tripod
- A lightbox with 2 LED strips and 1 LED panel attached;
- An acrylic sample holder is secured on the bottom of the lightbox
  - ensure the positions of samples are fixed at each shot taken

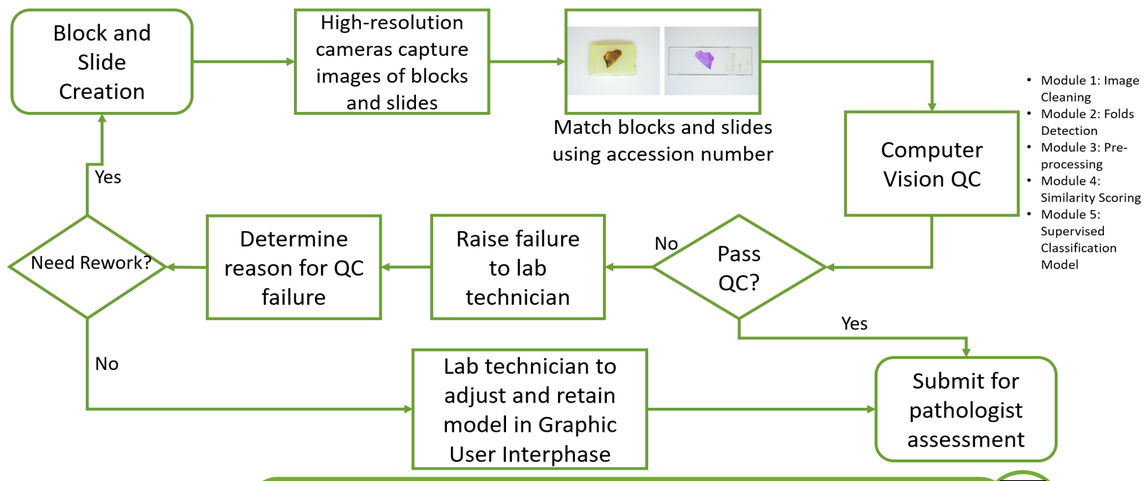


our group collected pictures of a total of 1,219 unique block-slide pairs as input data.



Acrylic holders for slides and blocks

WORKFLOW



COMPUTER VISION ALGORITHM

1. Image Cleaning

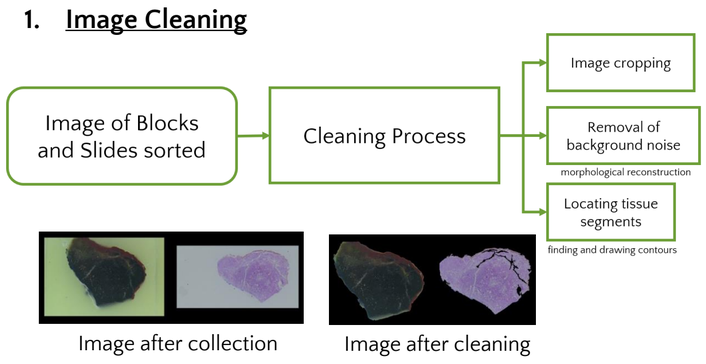
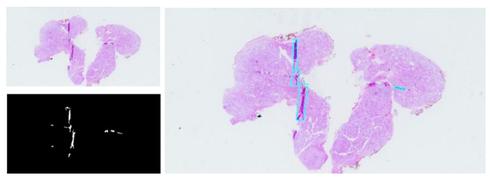


Image after collection

Image after cleaning

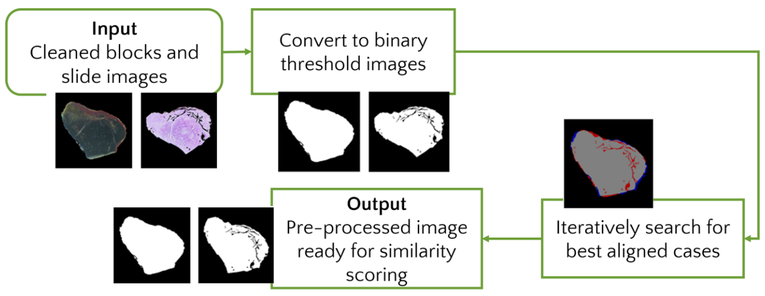
2. Folds Detection

- HSV-colour-based method
  - segment out areas of the tissue that are darker in colour and have elongated shapes.
- A bounding box is drawn on the slide image to visualise the detected tissue-folds.



Detected segments with long and darker colour (bottom left) and highlighted tissue-folds (right)

3. Image Pre-processing



4. Similarity Scoring

Pre-processed block image is compared with corresponding slide using similarity scoring methods. Numerical scores returned for each block-slide pair and be appended to the excel file.

Mean Squared Error (MSE)
- The smaller the value, the more similar.
- Large differences in pixel intensity may not suggest the image varies greatly.

Structural Similarity Index (SSIM)
- The range of this index is [-1,1].
- SSIM= -1, completely different
- SSIM = 1, very similar.

Jaccard Similarity Coefficient
- Calculates the similarity value between two vectors.
- Unable to provide useful insights on areas of fault without manual inspection.

Dice Similarity Score
- Calculated by dividing two times the area of overlap between two images by the total number of pixels in both images.

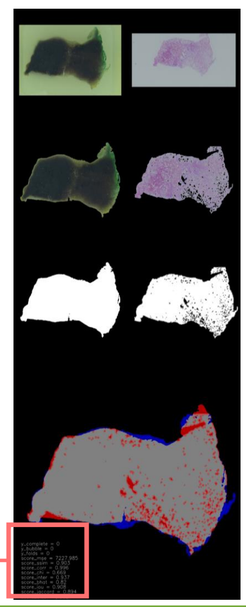
Histogram Comparison
- Shows a graphical representation of the colour tonal distribution of an image.

Intersection Over Union (IoU)
- Calculated by dividing the area of overlap between the two superimposed images by the area of union.
- The most intuitive method

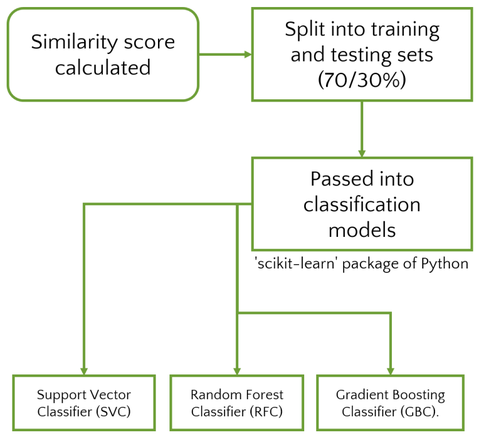
Unmatched areas are highlighted in red, matched areas in blue

```
y_complete = 0
y_bubble = 0
y_folds = 0
score_mse = 7227.985
score_ssim = 0.903
score_corr = 0.996
score_chi = 0.669
score_inter = 0.937
score_bhat = 0.82
score_iou = 0.908
score_jaccard = 0.894
score_dice = 0.952
```

Output Scores



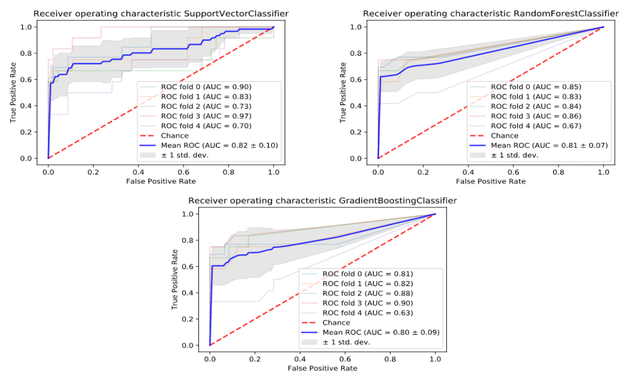
5. Supervised Classification Model



- Hyperparameters of the models are tuned using the grid search method with a 5-fold cross validation of the training set.

Results

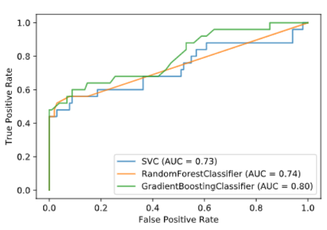
Section Completeness Performance



ROC curves of classifiers with a 5-fold cross validation

Table 1: Performances of classifiers

n=421	SVC (testing set)	RFC (testing set)	GBC (testing set)
Accuracy	85.83%	86.61%	89.76%
Sensitivity	28.00%	32.00%	48.00
Specificity (true negative rate)	100.00%	100.00%	100.00%
Precision (positive predictive value)	100.00%	100.00%	100.00%
Negative predictive value	85.00%	85.71%	88.70%



ROC curves of classifiers on testing set

The Gradient Boosting Classifier outperforms the other two models.

All three models had relatively low true positive rate.
- small missing tissue segments not picked by camera

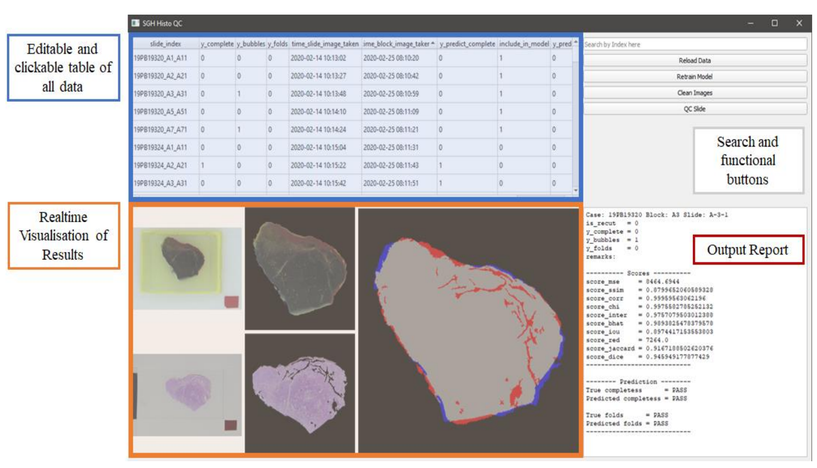
Tissue-folds Detection Performance

Correctly segment out tissue-folds from the rest of the slide's tissue section for 3 out of the 5 samples provided.
- HSV-colour-based segmentation method

Fold-like features are also misclassified
- Image resolution limitation

GRAPHICAL USER INTERFACE

- Model easily modified and retrained
- Continuous improvement
- Realtime visualisation
- Compiled into an executable file



CONCLUSION

Our group developed a preliminary setup with proper specifications for the QC workstation using a simple camera system.
- More advanced imaging techniques, such as hyperspectral camera can be evaluated in future.

This project focused on evaluating the quality of the tissue section completeness.
The QC tool is able to accurately evaluate the quality of a subset (40.0%, n=487) of the dataset.

Future plan:
- A pre-classification can also be done at the start of the QC procedure to classify tissues into various categories..
  - Use magnified virtual slide images for the detection of very small artefacts such as folds and bubble.