

PENEMBERENGAN NUKLEUS SEL PAYUDARA PADA IMEJ PATOLOGI MENGGUNAKAN ALGORITMA PEMBELAJARAN MENDALAM

Wan Nur Aqila, Afzan Binti Adam

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Pada 2021, Kementerian Kesihatan Malaysia (KKM) telah mengeluarkan kenyataan bahawa seorang dari 19 wanita di Malaysia akan menghidap kanser payudara dan 43% daripada jumlah itu telah berada di tahap 3 atau 4. Di dalam bidang pembelajaran mesin, pembelajaran mendalam atau kecerdasan buatan boleh diaplikasikan untuk mengesan kawasan sel-sel kanser menggunakan model seperti U-Net kajian Andrew et. al (2021) yang mempunyai 0.3163 min purata ketepatan (mAP). Walau bagaimanapun, bilangan penyelidikan dan penggunaan imej patologi dalam mengesan kawasan sel-sel kanser adalah rendah berbanding imej mammogram dalam lingkungan lima tahun kebelakang dan masih belum meluas. Kajian ini akan mengkhusus dalam mencadangkan penggunaan pembelajaran mendalam atau kecerdasan buatan untuk menembereng kawasan kanser payudara di dalam imej patologi digital. Proses penilaian sebuah slaid merupakan kaedah terbaik dalam mengesan kehadiran kanser namun terdapat beberapa kekangan dalam menilai slaid tersebut, antaranya ialah masa. Cui (2019) menyatakan imej bersaiz 1000 x 1000 mengambil masa sekurang-kurangnya lima jam untuk menjalankan proses anotasi. Tiga algoritma dipilih untuk dibanding bezakan di dalam kajian ini ialah Resnet-50, DenseNet 201, ResNet 101. Sebelum eksperimen dijalankan, proses augmentasi dijalankan ke atas dataset latihan data bagi memastikan jumlah dataset yang akan digunakan untuk melatih model bertambah. Eksperimen dijalankan ke atas algoritma ini menggunakan dataset TNBC dan MoNuSeg dengan jumlah data sebanyak 85 imej. Eksperimen tersebut dijalankan dengan kaedah membahagikan dataset kepada 70% imej untuk latihan data dan 30% untuk ujian data. Hasil dari tiga eksperimen tersebut mendapati bahawa ResNet-10 mempunyai nilai purata Precision yang tertinggi iaitu 0.681 sementara ResNet-50 memperolehi nilai purata Precision yang terendah iaitu 0.583 dan DenseNet 201 memperolehi purata Precision 0.603.

1 PENGENALAN

Secara amnya, kanser payudara terdiri dari lima jenis kanser berdasarkan penyataan Dr. Arief (2016) di myhealth.gov dan ia adalah; kanser setempat ('Carcinoma In Situ'), kanser saluran (DCIS), kanser lobular (LCIS). Kanser saluran merebak (IDC) dan yang terakhir adalah kanser lobul merebak (ILC). Akhbar The

Star (Murugappan, 2022) melaporkan satu lagi jenis kanser yang sukar untuk dikesan mahupun dirawat iaitu Kanser Payudara Triple Negatif atau lebih dikenali sebagai TNBC. TNBC ini merangkumi 18-20% dari keseluruhan pecahan jenis- jenis kanser payudara dan pesakit TNBC ini terdiri dari wanita berumur di bawah 40 tahun (Murugappan, 2022). Dr. Mastura menyatakan TNBC merupakan sejenis payudara kanser yang agresif dan ia sering terjadi apabila kanser tersebut kebiasanannya dikesan lewat dan sel-sel kanser TNBC telah merebak apabila ditemui (Murugappan, 2022).

Kanser payudara bukanlah lagi istilah yang asing bagi rakyat Malaysia, lebih-lebih lagi apabila barah ini telah meragut sebanyak tiga ribu lima ratus dan lapan pesakit kanser payudara pada tahun 2020. Namun, kanser payudara boleh dicegah dengan menjalani pemeriksaan awal di hospital dan mendapatkan nasihat dari doktor. Sekiranya dilihat dari sudut klinikal, para doktor akan menjalankan pemeriksaan ke atas pesakit dan antara prosedur yang dijalankan untuk mengenal pasti kawasan sel-sel kanser ialah melalui pemerhatian ke atas patologi slid. Slaid tersebut diperolehi hasil dari biopsi dari pesakit dan tisu tersebut akan diperhatikan di bawah mikroskop.

Pada awal abad 20-an, sains fiksyen memperkenalkan sebuah idea di mana robot mahupun mesin mempunyai keupayaan untuk bercakap dan kebolehan berfikir seperti manusia. Idea dan fantasi ini diteruskan ke dalam media seperti filem, novel dan majalah seterusnya dipopularkan lagi oleh MGM dengna filem Wizard of Oz, menurut Rockwell Anyoha (2019), watak seperti Tin Man seolah-olah menggambarkan robot yang berkeupayaan seperti manusia. Idea ini dibawa ke hadapan dan kebarangkalian mesin untuk mempunyai kecerdasan seperti manusia ditemui oleh Alan Turing ada 1950. Beliau menyatakan sebagaimana manusia mampu menggunakan pengetahuan, data dan informasi yang sedia ada untuk membuat keputusan, mesin juga berupaya untuk membuat keputusan (Anyoha, 2019). Idea ini diteruskan lagi sehingga lahirnya cabang ilmu Kecerdasan Buatan dan Pembelajaran Dalam yang mampu membantu manusia untuk membuat keputusan yang lebih tepat berdasarkan data dan maklumat yang diberikan.

Membantu manusia untuk membuat keputusan merupakan bidang kecerdasan buatan yang boleh diaplikasikan untuk menembereng kawasan sel-sel kanser menggunakan imej slaid patologi yang diperolehi dari biopsi pesakit. Kecerdasan buatan merupakan cabang disiplin ilmu sains computer dan menurut Frankenfield (2022), kecerdasan buatan merujuk kepada stimulasi kecerdasan manusia yang diatur untuk berfikir seperti manusia. Tambahnya lagi; istilah kecerdasan buatan juga boleh diaplikasikan ke atas mesin yang mempamerkan sifat pemikiran manusia seperti keupayaan menyelesaikan masalah dan pembelajaran. Frankenfield (2022) menyatakan kecerdasan buatan boleh dibahagikan lagi kepada beberapa subset disiplin ilmu seperti pembelajaran mendalam dan pembelajaran mesin.

Pembelajaran mesin merujuk kepada konsep di mana computer program berupaya untuk mempelajari secara automatic malah berkemampuan untuk beradaptasi ke atas data baru tanpa memerlukan bantuan manusia (Frankenfield,2022). Kata Frankfield(2020) lagi,menyatakan pembelajaran mendalam pula merupakan teknik yang membolehkan pembelajaran automatik dijalankan walaupun menerima data yang tidak berstruktur seperti video, mesej dan gambar. Belyadi (2021) pula memberi contoh kegunaan pembelajaran mesin di dalam kehidupan seharian seperti pengecaman muka menggunakan imej, kereta pandu sendiri dan penterjemahan teks.

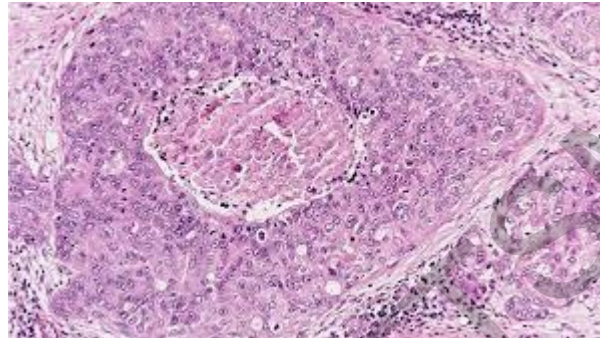
Jabeer (2022) melaporkan di National Cancer Institute (NIH) ujian seperti mammogram dan Pap dijalankan secara kerap untuk mengesan simptom-simtom awal sel-sel kanser bagi memastikan rawatan awal dapat dijalankan sebelum sel-sel tersebut merebak. Beliau juga melaporkan bahawa saintis telah membangunkan alat kecerdasan buatan untuk menyaring beberapa jenis kanser untuk membantu doktor mentafsir imej mammogram sejak 20 tahun yang lalu. Dakwaan ini disokong dengan kenyataan dari Dr.Turkbey yang menyatakan model kecerdasan buatan [Kumpulan National Cancer Institute, NCI] memudahkan lengkung pembelajaran ahli radiologi dan meminimumkan peratusan ralat (Jabeer, 2022).

Aplikasi disiplin ilmu kecerdasan buatan dalam menembereng kawasan sel-sel kanser payudara dapat membantu mengesan kehadiran sel-sel kanser pesakit lantas dapat memastikan pesakit mendapatkan rawatan awal. Ini boleh mencegah sel-sel kanser merebak di dalam tubuh pesakit sebelum sel-sel tersebut membiak dan membahayakan nyawa pesakit. Oleh itu, sistem penemberengan kawasan sel-sel kanser payudara ini dapat menyumbang dalam memastikan pesakit menerima rawatan awal dalam usaha mencegah penyakit kanser di kalangan rakyat Malaysia.

2 PERNYATAAN MASALAH

Ahli patologi dan pakar perubatan dapat mengesan kawasan sel-sel kanser dengan menilai sesebuah slaid patologi secara manual (Gupta,2019). Tugas menilai sesebuah slaid patologi juga memakan masa memandangkan ia terdiri dari proses yang rumit. Prosedur biopsi akan dijalankan di mana sampel tisu pesakit akan diambil untuk dicerap di bawah mikroskop. Ahli patologi akan mengekstrak kawasan tisu yang dikesan mempunyai tumor kemudiannya salid tersebut akan diwarnakan dengan noda hispatologi hematoxylin dan eosin (H&E). Noda hispatologi H&E ini digunakan untuk mengesan kawasan sel kanser dengan memerhatikan kawasan yang maglinan dan benigna (Dabeer,2019). Namun, cara tradisional ini mengundang risiko untuk melakukan kesalahan dalam mengdiagnosis sesebuah slaid patologi kerana ia merupakan sebuah tugas yang berat dan jerih (Abdoallahi 2022).Dabeer(2019) menjelaskan lagi bahawa aplikasi Pembelajaran Mesin mampu menembreg kawasan sel-sel kanser dengan tepat. Archarya (2022) mendefinisikan penemeberengan

imej sebagai kompute visi yang mebahagikan imej digital kepada beberapa bahagian atau lebih dikenali sebagai segmen imej dengan tujuan mengurangkan kesukaran imej tersebut apabila menjalankan proses menganalisa setiap segmen imej tersebut.



Rajah 1.0 : Sel kanser saluran

(Sumber: pathology.jhu.edu)

Slaid patologi yang mengandungi sampel tisu tersebut akan dinilai secara manual iaitu dengan dicerap di bawah mikroskop (Gupta, 2019). Dabeer(2019) menjelaskan bahawa proses menilai sebuah slaid patologi memakan masa yang lama untuk mengesan kawasan kanser payudara. Krishna (2022) berpendapat proses pencerap slaid patologi digital ini boleh menyebabkan peningkatan kadar ralat dan sukar untuk diuruskan dengan pertambahan bilangan slaid patologi digital untuk dicerap dan dinilai. Rahsmi (2021) menyenaraikan beberapa kekangan yang dihadapi oleh ahli patologi dalam menganalisis slaid patologi digital antaranya:

1. Terdapat jenis-jenis tumor ataupun kanser yang tidak boleh dikesan menggunakan analisis slaid patologi digital yang menggunakan node H & E.
2. Penilaian salid patologi digital tersebut bergantung kepada pengalaman dan kemahiran ahli patologi yang boleh membawa kepada variasi kepada keputusan pencerapan dan analisis slaid patologi digital.
3. Proses pencerapan dan analisis slaid patologi mengambil masa yang lama dan merupakan proses yang jerih. Misalnya, sifat sel yang kompleks, pewarnaan slaid yang tidak seragam dan variasi keterangan slaid.

3 OBJEKTIF

Mengaplikasikan beberapa algoritma Kecerdasan Buatan untuk menganalisis kesesuaian algoritma kecerdasan buatan terhadap penemberengan kawasan tumor dan stroma kanser payudara jenis TNBC.

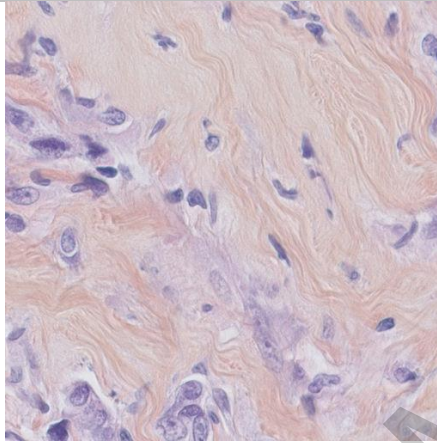
4 METODOLOGI

Bahagian ini menerangkan keperluan yang diperlukan sepanjang kajian ini dijalankan. Ia termasuk keperluan dari segi perisian, perkakasan, dataset dan lain – lain. Bahagian ini juga akan mencadangkan model atau algoritma yang akan digunakan untuk menembereng kawasan kanser payudara menggunakan imej slaid patologi digital. Pemilihan algoritma tersebut akan terus dipertimbangkan di dalam kajian dan eksperimen projek ini dan ia juga mengambil kira sorotan kesusasteraan yang telah dijalankan sebelum ini.

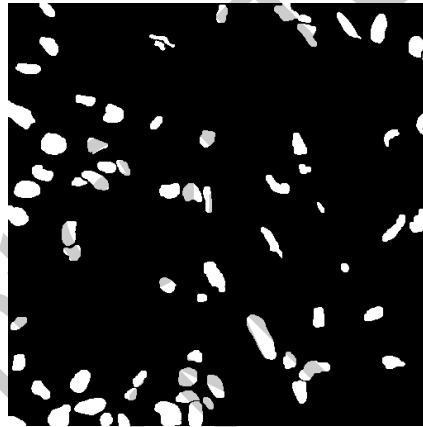
Data kajian ini mengandungi imej slaid patologi sel kanser yang disampel melalui proses biopsi pesakit dari portal Genomic Data Common portal dan telah ditandakan dengan noda hematoxylin dan eosin (H&E) untuk menandakan sel nukleus dengan warna ungu dan sitoplasma dengan warna biru. Cahaya warna yang lain pula merupakan gabungan warna ini di sekitar sel tersebut. Dataset ini terdiri dari dataset TNBC dan MoNuSeg.

Jadual 4.0 – Spesifikasi Keperluan Sistem

Keperluan Dataset	Dataset yang diperolehi ialah dataset yang mengandungi TNBC dan MoNuSeg. Imej ujian Latihan mengandungi 55 imej. Setiap imej dipasangkan bersama dengan mask dan imej ini berada di dalam format .png
Dataset TNBC	Dataset MoNuSeg ini diambil dari laman CDC portal Imej di bawah menunjukkan contoh imej Latihan dari TNBC, terdapat 30 imej pewarnaan H&E di dalam dataset ini.



Rajah 4.1 – Imej TNBC 01_1



Rajah 4.2 – Mask TNBC 01_1

Di bawah merupakan senarai ID imej dataset TNBC yang digunakan di dalam projek ini:

A. Imej Latihan (Train):

01_1	03_4
01_2	03_5
01_3	04_1
01_4	04_2
01_5	04_3
01_6	04_4
01_7	04_5
02_1	04_6
02_2	04_7
02_3	04_8
03_1	05_1
03_2	05_2
03_3	

	<p>B. Imej Ujian (Testing):</p> <table border="0"> <tr><td>05_3</td><td>09_2</td></tr> <tr><td>05_4</td><td>09_3</td></tr> <tr><td>06_1</td><td>09_4</td></tr> <tr><td>06_2</td><td>09_5</td></tr> <tr><td>06_3</td><td>09_6</td></tr> <tr><td>07_1</td><td>10_1</td></tr> <tr><td>07_2</td><td>10_2</td></tr> <tr><td>07_3</td><td>10_3</td></tr> <tr><td>08_1</td><td>10_4</td></tr> <tr><td>08_2</td><td>11_1</td></tr> <tr><td>08_3</td><td>11_2</td></tr> <tr><td>08_4</td><td>11_3</td></tr> <tr><td>09_1</td><td></td></tr> </table>	05_3	09_2	05_4	09_3	06_1	09_4	06_2	09_5	06_3	09_6	07_1	10_1	07_2	10_2	07_3	10_3	08_1	10_4	08_2	11_1	08_3	11_2	08_4	11_3	09_1	
05_3	09_2																										
05_4	09_3																										
06_1	09_4																										
06_2	09_5																										
06_3	09_6																										
07_1	10_1																										
07_2	10_2																										
07_3	10_3																										
08_1	10_4																										
08_2	11_1																										
08_3	11_2																										
08_4	11_3																										
09_1																											
<p>Dataset MoNuSeg</p>	<p>Dataset MoNuSeg ini diambil dari laman web Genomic Data Commons (GDC) yang didirikan oleh National Cancer Institute. Imej di bawah menunjukkan contoh imej Latihan dari MoNuSeg, terdapat 30 imej pewarnaan H&E di dalam dataset ini.</p> <div data-bbox="740 1003 1102 1361" data-label="Image"> </div> <p>Rajah 4.3 - TCGA-18-5592-01Z-00-DX1 (Sumber: Portal GDC)</p> <div data-bbox="732 1471 1112 1850" data-label="Image"> </div> <p>Rajah 3.4 - TCGA-18-5592-01Z-00-DX1_bin_mask</p> <p>GDC. (n.d.). https://portal.gdc.cancer.gov/</p>																										

Di bawah merupakan senarai ID imej dataset MoNuSeg yang digunakan di dalam projek ini.

A. Imej Latihan (train) :

1. TCGA-18-5592-01Z-00-DX1
2. TCGA-21-5784-01Z-00-DX1
3. TCGA-21-5786-01Z-00-DX1
4. TCGA-38-6178-01Z-00-DX1
5. TCGA-49-4488-01Z-00-DX1
6. TCGA-50-5931-01Z-00-DX1
7. TCGA-A7-A13E-01Z-00-DX1
8. TCGA-A7-A13F-01Z-00-DX1
9. TCGA-AR-A1AK-01Z-00-DX1
10. TCGA-AR-A1AS-01Z-00-DX1
11. TCGA-AY-A8YK-01A-01-TS1
12. TCGA-B0-5698-01Z-00-DX1
13. TCGA-B0-5710-01Z-00-DX1
14. TCGA-B0-5711-01Z-00-DX1
15. TCGA-CH-5767-01Z-00-DX1
16. TCGA-DK-A2I6-01A-01-TS1
17. TCGA-E2-A1B5-01Z-00-DX1
18. TCGA-E2-A14V-01Z-00-DX1
19. TCGA-G2-A2EK-01A-02-TSB
20. TCGA-G9-6336-01Z-00-DX1
21. TCGA-G9-6348-01Z-00-DX1
22. TCGA-G9-6356-01Z-00-DX1
23. TCGA-G9-6362-01Z-00-DX1
24. TCGA-G9-6363-01Z-00-DX1
25. TCGA-HE-7128-01Z-00-DX1
26. TCGA-HE-7128-01Z-00-DX1
27. TCGA-HE-7130-01Z-00-DX1
28. TCGA-KB-A93J-01A-01-TS1
29. TCGA-NH-A8F7-01A-01-TS1
30. TCGA-RD-A8N9-01A-01-TS1

B. Imej Ujian (Testing):

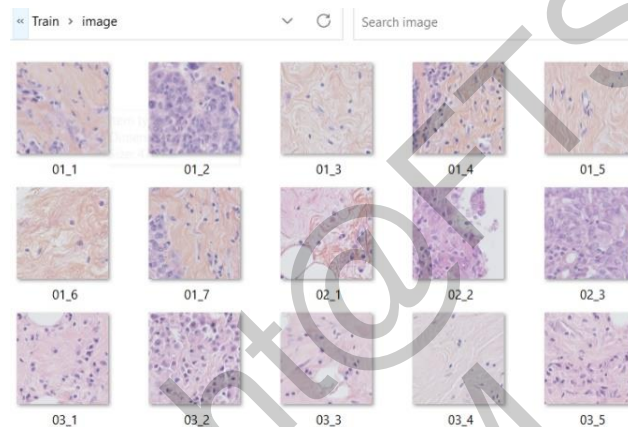
1. TCGA-2Z-A9J9-01A-01-TS1
2. TCGA-44-2665-01B-06-BS6
3. TCGA-69-7764-01A-01-TS1
4. TCGA-A6-6782-01A-01-BS1
5. TCGA-AC-A2FO-01A-01-TS1
6. TCGA-AO-A0J2-01A-01-BSA
7. TCGA-CU-A0YN-01A-02-BSB
8. TCGA-EJ-A46H-01A-03-TSC
9. TCGA-FG-A4MU-01B-01-TS1
10. TCGA-GL-6846-01A-01-BS1

- | | |
|--|--|
| | 11. TCGA-HC-7209-01A-01-TS1
12. TCGA-HT-8564-01Z-00-DX1
13. TCGA-IZ-8196-01A-01-BS1
14. TCGA-ZF-A9R5-01A-01-TS1 |
|--|--|

4.1 Pra - pemrosesan data

Jumlah dataset Latihan Data hanyalah 55, untuk memastikan dataset mempunyai variasi dan mempunyai kuantiti yang banyak, proses data augmentasi di jalankan.

Rajah di bawah menunjukkan dataset sebelum dijalankan proses augmentasi.

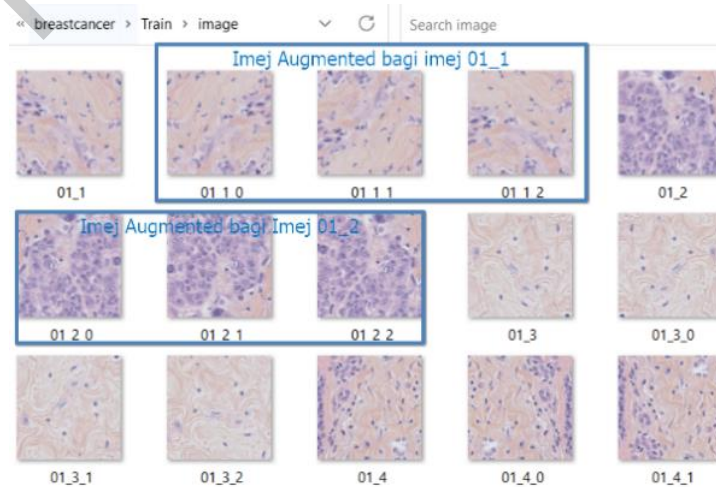


Rajah 4.1.0 – Imej latihan sebelum proses augmentasi

Selepas mengalami proses augmentasi, dataset Latihan Data bertambah menjadi 110.

Rajah 2.0 menunjukkan imej 01_1 mempunyai tiga variasi gambar yang berbeza.

Proses augmentasi yang dijalankan akan menghasilkan tiga imej yang flip horizontally, flip vertically dan rotate pada sudut 45 darjah.



Rajah 4.1.1 – Imej latihan selepas proses augmentasi

4.2 Latihan Data dan Ujian Data

Jumlah keseluruhan imej di dalam dataset yang digunakan ialah 94, dataset ini dibahagikan kepada dua bahagian iaitu Latihan Data (Train) dan Ujian Data (Testing). Latihan Data mengandungi imej berjumlah 55 manakala Ujian Data mengandungi imej berjumlah 39. Namun, selepas imej menjalani proses augmentasi, dataset untuk Latihan Data akan bertambah menjadi 110, hal ini kerana setiap imej akan memperoleh tambahan variasi 3 imej. Dataset ini dibahagikan dengan 73% dikhususkan untuk Latihan Data dan 27% untuk Ujian Data.

5 HASIL KAJIAN

Antara kaedah evaluasi yang digunakan ialah Indeks Jaccard, Presicion, Acc dan skor F1. Rajah 26.0 menunjukkan kod kritikal dalam pelaksanaan Latihan Data dan Ujian Data. Oleh kerana projek ini menggunakan tiga model yang berbeza, ketiga – tiga model ini menunjukkan prestasi yang berbeza. Parameter seperti epoch dan batch yang diletakkan adalah sama untuk setiap model ini. Jadual di bawah menunjukkan prestasi ketiga tiga model ini. Data yang ditunjukkan adalah purata untuk keseluruhan dataset Latihan Data dan perbezaan di antara tiga model ini boleh dilihat di dalam evaluasi setiap imej. Evaluasi imej yang digunakan di dalam jadual di bawah adalah imej berindeks indeks 32 dengan ID Imej ; TCGA-EJ-A46H-01A-03-TSC. Namun di dalam antara muka, model yang digunakan ialah ResNext 101.

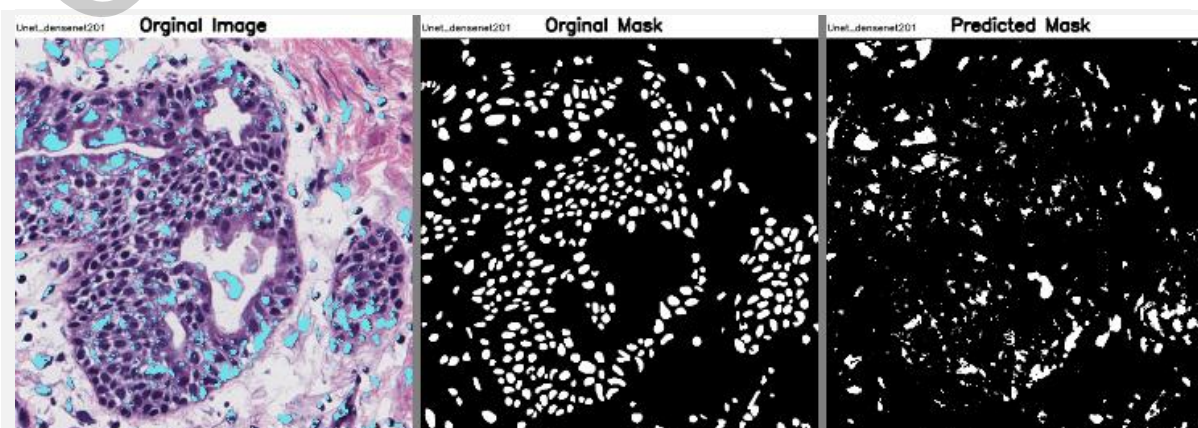
Objektif utama pengujian ini diadakan adalah untuk menguji model yang dapat memberikan hasil penemberengan yang: (1) Mempunyai hasil penembangan yang konsisten dan seragam untuk kesemua 39 imej Ujian Data, (2) Mendapat indeks Jaccard yang tinggi. Indeks Jaccard berfungsi sebagai kaedah penilaian untuk menilai ketepatan pertindihan di antara predicted mask dan ground truth masks. Ketepatan pertindihan tersebut dinilai menggunakan *Intersection Union* (IU) dan ia mengambil kira bilangan piksel yang tepat.

Jadual 5.1 – Eksperimen yang dijalankan

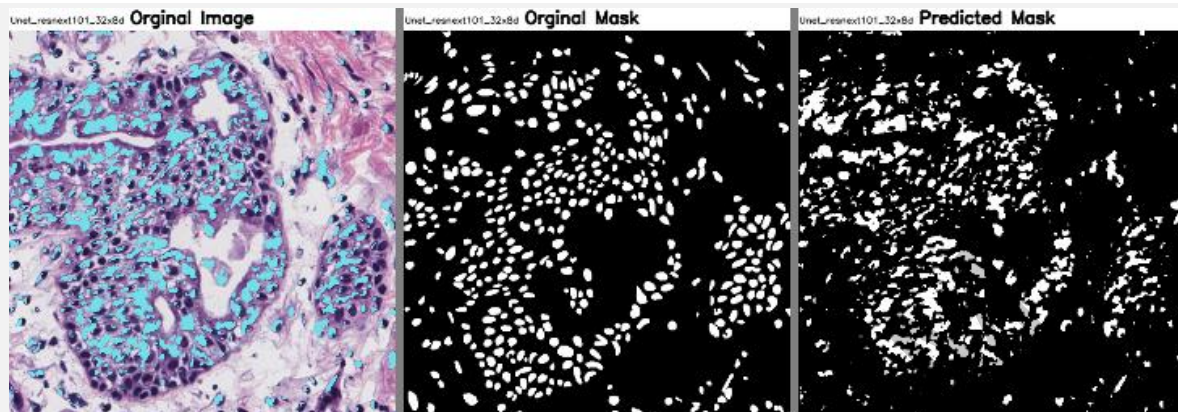
Eksperimen I :ResNet 50	Eksperimen II :DenseNet 201	Eksperimen III:ResNext 101
Jaccard avg: 0.3	Jaccard avg: 0.3	Jaccard avg: 0.3
F1 avg: 0.483	F1 avg: 0.479	F1 avg: 0.493
Recall avg: 0.61	Recall avg: 0.481	Recall avg: 0.478
Precision avg: 0.583	Precision avg: 0.603	Precision avg: 0.681
Acc avg: 0.807	Acc avg: 0.874	Acc avg: 0.877
Evaluasi imej di indeks 38		
Jaccard: 14.3	Jaccard: 13.3	Jaccard: 14.3
F1: 18.681	F1: 18.824	F1: 19.209
Recall: 18.74	Recall: 23.782	Recall: 18.641
Precision: 23.519	Precision: 22.721	Precision: 26.543
Acc: 34.081	Acc: 31.49	Acc: 34.186



Rajah 5.0 - Cat_Images model ; ResNet 50



Rajah 5.1 : Cat_Images model; DenseNet 201



Rajah 5.2 : Cat_Images model ResNext 101

Berdasarkan metrik penilaian (Indeks Jaccard, Skor F1, Recall, Presisi, dan Ketepatan) untuk tiga model (ResNet 50, DenseNet 201, dan ResNext 101) daripada Jadual 9.0, berikut adalah beberapa rumusan yang dibuat:

1. Indeks Jaccard mengukur pertindihan piksel dan menunjukkan sejauh mana *predicted masks* segmentasi sejajar dengan *ground truth*. Daripada ketiga-tiga model tersebut, ResNext-101 mencapai nilai Indeks Jaccard tertinggi iaitu 0.681, ini *predicted masks* model ResNext-101 mempunyai pertindihan yang paling tinggi dengan *ground truth* secara purata untuk keseluruhan 39 dataset Ujian Data. Sementara itu, DenseNet-201 mencapai Indeks Jaccard terendah iaitu 0.478.
2. Skor F1 adalah purata harmoni di antara *Presicion* dan *Recall*, dan ia menyeimbangkan kedua-dua metrik tersebut. ResNext 101 memiliki Skor F1 tertinggi iaitu 0.493, manakala DenseNet 201 memiliki Skor F1 terendah iaitu 0.479.
3. *Recall* mengukur keupayaan model untuk mengesan semua kejadian positif benar (*True Positive*) daripada jumlah keseluruhan kejadian positif (*True Positif + False Negatif*). ResNet -50 pula mencapai nilai *Recall* tertinggi iaitu 0.61, menunjukkan keupayaannya untuk mengesan positif benar dengan lebih baik, manakala DenseNet 201 mempunyai nilai recall terendah iaitu 0.481.
4. Presisi mengukur keupayaan model untuk mengenal pasti kejadian positif (positif benar) dengan betul daripada keseluruhan kejadian yang dikenal pasti sebagai positif (positif benar + positif palsu). ResNext-101 mencapai *Precision* tertinggi iaitu 0.681,

menunjukkan bahwa jangkauan positifnya memiliki lebih tinggi positif palsu. Sementara itu, DenseNet-201 mencapai presisi terendah yaitu 0.603.

5. Ketepatan (*accuracy*) mengukur keseluruhan kebetulan jangkauan model. ResNext 101 memiliki ketepatan tertinggi yaitu 0.877, manakala ResNet 50 memiliki ketepatan terendah yaitu 0.807.

Secara keseluruhannya, berdasarkan Indeks Jaccard, Skor F1, Recall, Presisi, dan Ketepatan, ResNext 101 menunjukkan prestasi yang lebih baik berbanding dua model lain (ResNet 50 dan DenseNet 201) dalam segmentasi. Ia mencapai nilai Indeks Jaccard dan Skor F1 yang lebih tinggi, yang menunjukkan ketepatan segmentasi yang lebih baik dan prestasi model secara purata berbanding dengan model ResNet-50 dan DenseNet-201. Selain itu, ResNext 101 memiliki presisi tertinggi, yang bermakna jangkauan positifnya mengeluarkan lebih positif palsu, dan ia juga memiliki *Recall* yang tinggi, menunjukkan keupayaannya untuk mengesan lebih banyak positif benar.

6 KESIMPULAN

Projek ini bertujuan untuk mengendalikan tiga komponen utama; (1) Segmentasi Nukleus Kanser Payudara Jenis TNBC; melatih model pembelajaran mendalam, seperti U-Net atau variasinya, untuk melakukan segmentasi nukleus pada imej kanser payudara. Matlamatnya adalah untuk mengenal pasti dan mengasingkan kawasan nukleus dalam imej, yang boleh menjadi penting untuk analisis dan diagnosis lanjut. Seterusnya, membandingkan model di dalam Eksperimen I, II dan III. Prestasi model tersebut menggunakan pelbagai metrik penilaian, seperti indeks Jaccard, skor F1, ketepatan (*precision*), kepekaan (*recall*), dan ketepatan umum (*accuracy*). Model tersebut dinilai dan dipilih salah satu untuk diaplikasikan di dalam antara muka. Yang terakhir ialah menyediakan *dashboard* untuk menjalankan proses penemberengan. *Dashboard* ini akan berfungsi sebagai antara muka pengguna berasaskan web di mana pengguna dapat berinteraksi dengan model yang dilatih.

7 RUJUKAN

Dabeer, S., Khan, M., & Islam, S. (2019, August 21). *Cancer diagnosis in histopathological image: CNN based approach*. Sciencedirect.

<https://www.sciencedirect.com/science/article/pii/S2352914819301133>

Gupta, K., & Chawla, N. (n.d.). *Analysis of Histopathological Images for Prediction of Breast Cancer Using Traditional Classifiers with Pre-Trained CNN*. Sciencedirect. Retrieved January 1, 2022, from

<https://www.sciencedirect.com/science/article/pii/S1877050920308930>

Rockwell Anyoha, R. (2020, April 23). The History of Artificial Intelligence. Science in the News. <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>

Frankfield, J. (2022, July 6). *Artificial Intelligence: What It Is and How It Is Used*.

Investopedia. <https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp>

Belyadi, H., & Haghghat, A. (2021). *Machine Learning Guide for Oil and Gas Using Python: A Step-by-Step Breakdown with Data, Algorithms, Codes, and Applications* (1st ed.). Gulf Professional Publishing.

Wan Nur Aqila (A184403)

Dr. Afzan binti Adam

Fakulti Teknologi & Sains Maklumat,

Universiti Kebangsaan Malaysia