

# PENGELASAN PENYAKIT CHRONIC MYELOID LEUKAEMIA DARI IMEJ SLAID DARAH MENGGUNAKAN TEKNIK PEMBELAJARAN MENDALAM

CHUNG YEN WANN (FTSM)

DR. AFZAN ADAM (FTSM)

Raja Zahratul Azma Raja Sabudin (HUKM)

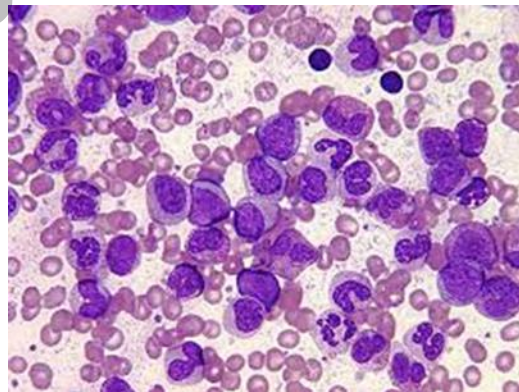
## ABSTRAK

Penyakit *chronic myeloid leukaemia* (CML) adalah sejenis kanser darah yang berpunca dari perubahan genetik pada sel myeloid yang belum matang di sumsum tulang. Ini mengakitbatkan pembentukan sel myeloid yang tidak normal hingga melimpah ke saluran darah dan adakalanya boleh mendap di limpa. Untuk menentukan seseorang itu ada menghadapi penyakit tersebut atau tidak, pakar hematologi akan menyediakan slaid darah dari sampel darah untuk membuat analisis. Selepas itu, pakar hematologi perlu menganalisis slaid darah tersebut secara manual dengan menggunakan mikroskop untuk mengelaskan dan mengira sel granulosit. Sel-sel darah menganalisis secara manual telah membeban kerja mereka jadi pembangunan model ini akan mengurangkan beban pakar hematologi di mana model akan mengelaskan imej darah dengan sel myeloid secara automatik dengan menggunakan teknik pembelajaran mendalam. Kajian ini mencadangkan dan penyiasatan seperti model seni bina rangkaian neural konvolusi iaitu Inception v3 untuk menentukan sama ada ia berfungsi paling baik dari segi ketetapan dan kecekapan dengan dataset imej slaid chronic myeloid leukaemia dengan melatih semula lapisan akhir seni bina Inception v3 yang dibangunkan oleh Google untuk pendekatan pengelasan ini. Kajian ini telah mencapai ketetapan 88% dengan kadar *misclassification* yang sangat rendah.

Sel darah putih atau juga dipanggil sebagai leukosit merupakan sejenis sel sistem imun yang membentuk komponen darah dan ia berasal dari sel multipoten di sumsum tulang. (Berry, n.d.) Sel darah putih terdiri dari dua jenis iaitu granulosit dan agranulosit. Granulosit terdiri daripada eosinophil, basophil dan neutrofil dan agranulosit terdiri daripada monosit dan limfosit, (Lynee Eldridge, 2018).

Sekiranya sumsum tulang seseorang menghasilkan terlalu banyak sel darah putih yang dipanggil granulosit(neutrofil, eosinofil, basofil). Penyakit yang berpunca daripada penghasilan granulosit yang banyak ialah *chronic myeloid leukaemia*(CML). (Laura J. Martin, 2018) Sel-sel putih tersebut biasanya membantu badan untuk melawan jangkitan dan penyakit. Dalam *chronic myeloid leukaemia*, sel-sel ini rosak dan tidak boleh berfungsi secara normal kerana sel granulosit ini tidak matang. Sel-sel tersebut tumpah keluar dari sumsum tulang, beredar di sekeliling badan dalam aliran darah dan terkumpul di dalam pelbagai organ. *Chronic myeloid leukaemia* biasanya mempamerkan secara relatif penyakit yang semakin perlahan(fasa kronik) tetapi dengan ketiadaan rawatan yang berkesan ia boleh berubah menjadi jenis penyakit yang lebih agresif(fasa pecutan atau pecahan) apabila sumsum tulang menghasilkan jumlah sel granulosit yang tidak matang yang dikenali sebagai sel *blast* atau *leukaemia blasts*.

Pakar hematologi yang perlu menganalisis slaid darah dari sampel darah seseorang yang menghadapi CML dengan menggunakan mikroskop untuk mengelaskan dan megira bilangan sel granulosit akan mengambil masa lebih panjang. Jadi, kajian ini akan membantu pakar hematologi menyenangkan beban mereka dengan pembanguanan perisian yang menggunakan teknik pembelajaran mendalam.



Rajah 1.1 *Chronic Myeloid Leukaemia*(CML)

## 2 PENYATAAN MASALAH

Pengelasan sel myeloid untuk menganalisis penyakit *Chronic Myeloid Leukemia*(CML) memerlukan pengetahuan pakar hematologi. Tetapi bidang hematologi menghadapi masalah kekurangan bilangan anggota baru. Permintaan untuk perkhidmatan hematologi telah berkembang dengan pesat dalam beberapa tahun kebelakangan ini, tetapi kakitangan tidak meningkat pada kadar yang sama. Tambahan pula, pengelasan sel myeloid memerlukan masa yang panjang untuk analisis kerana CML mempunyai tiga fasa iaitu fasa kronik, fasa pacutan dan fasa pecahan. Pakar hematologi juga mengemukakan masalah kekangan kewangan kerana kos untuk membeli mesin berkaitan adalah tinggi. Oleh itu, kajian ini dibangunkan bertujuan untuk mengurangkan beban pakar hematologi.

## 3 OBJEKTIF KAJIAN

Objektif kajian ini adalah membangunkan sebuah model pengelasan bagi imej *periferel blood film* bagi penyakit CML menggunakan teknik pembelajaran mendalam.

## 4 METOD KAJIAN

Dalam pembangunan model ini, modal proses yang akan digunakan ialah modal *Agile*. Modal *Agile* adalah gabungan modal proses *Incremental* yang memberi tumpuan kepada kesesuaian proses dan kepuasan pengguna dengan penghantaran pesat produk perisian kerja, (Gabry, 2017). Modal *Agile* memecah produk dengan membina tambahan kecil. Binaan ini disediakan dalam lelaran. Setiap lelaran biasanya berlangsung dari kira-kira satu hingga tiga minggu.



Rajah 4.1 Modal Agile

#### 4.1 Fasa Perancangan

Dalam fasa ini, perancangan projek akan dibuat dengan mengenalpasti masalah kajian, objektif kajian dan menentukan skop kajian. Selepas itu, pengumpulan dan pembacaan jurnal atau kajian lepas akan dijalankan untuk mencari idea yang boleh membantu dalam kajian ini. Topik yang dikaji utama dalam kajian ini ialah topik tentang pengelasan imej dengan menggunakan teknik pembelajaran mendalam seperti rangkaian neural konvolusi(CNN). Maklumat berkaitan telah dicari dengan penggunaan internet dan temu bual dengan pakar iaitu Assoc. Prof. Dr. Raja Zahratul Azma Raja Sabudin. Langkah ini akan memastikan kelancaran projek.

## 4.2 Fasa Analisis

Perbincangan dengan pengguna akan dijalankan untuk menganalisis keperluan yang perlu dalam sistem perisian ini. Peringkat ini penting dalam kajian ini kerana ia melibatkan komunikasi yang kerap dengan pengguna untuk menentukan fungsi tertentu yang pengguna harap yang ada dalam model ini. Dalam fasa reka bentuk, prototaip akan dibuat dengan keperluan pengguna yang membincang dalam fasa menganalisis keperluan.

## 4.3 Fasa Reka Bentuk

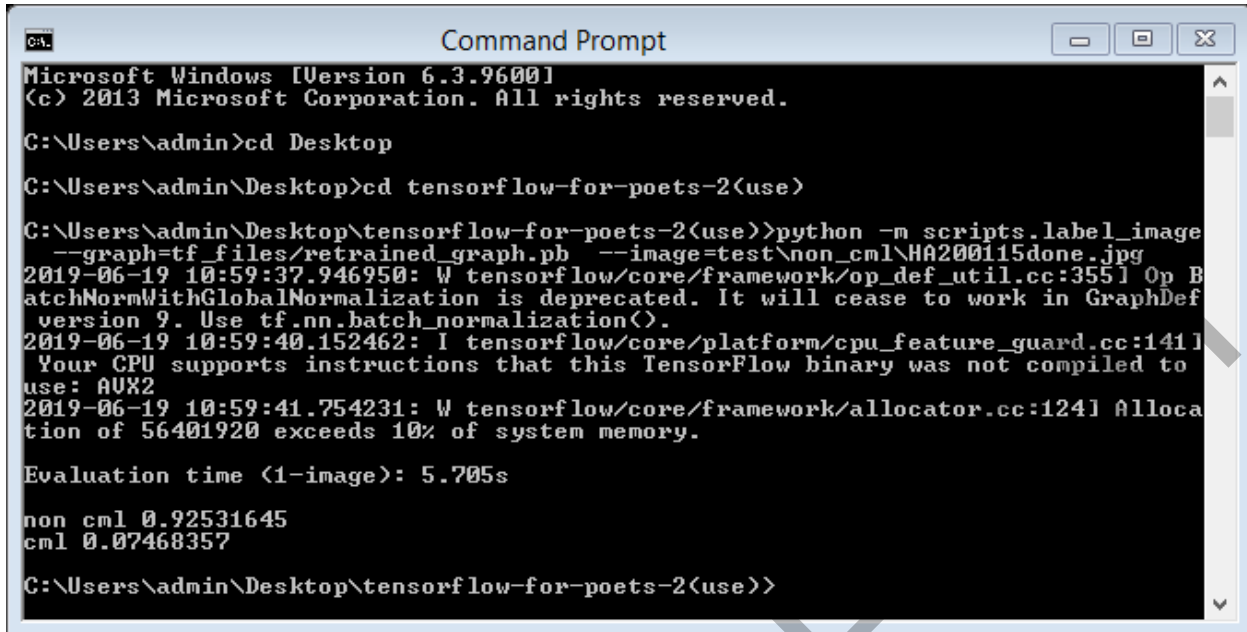
Fasa ini akan membincangkan spesifikasi reka bentuk sistem perisian dengan terperinci. Prototaip akan dibuat dengan keperluan pengguna yang membincang dalam fasa menganalisis keperluan. Selepas itu, segala maklumat dan data diperolehi akan digunakan untuk pembangunan model dengan menggunakan teknik pembelajaran mendalam.

Model ini melatih dengan menggunakan rangka kerja iaitu Tensorflow dan menjalankan komen melatih untuk memanggil komputer melatih imej slaid darah chronic myeloid leukaemia. Komen yang digunakan untuk melatih imej slaid darah telah ditunjukkan di bawah:

```
python -m scripts.retrain --bottleneck_dir=tf_files/bottlenecks --model_dir=tf_files/models
--summaries_dir=tf_files/training_summaries --output_graph=tf_files/retrained_graph.pb
--how_many_training_steps=200 --output_labels=tf_files/retrained_labels.txt --
architecture=inception_v3 --image_dir=tf_files/bloodcancer_photos
```

Selepas berjaya melatih imej slaid darah, komen yang ditunjukkan di bawah akan membantu pengguna menjalankan pengelasan iaitu:

```
python -m scripts.label_image --graph=tf_files/retrained_graph.pb --
image=test\non_cml\HA200115done.jpg
```



```

Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\Users\admin>cd Desktop
C:\Users\admin\Desktop>cd tensorflow-for-poets-2(use)
C:\Users\admin\Desktop\tensorflow-for-poets-2(use)>python -m scripts.label_image
--graph=tf_files/retrained_graph.pb --image=test\non_cml\HA200115done.jpg
2019-06-19 10:59:37.946950: W tensorflow/core/framework/op_def_util.cc:355] Op BatchNormWithGlobalNormalization is deprecated. It will cease to work in GraphDef
version 9. Use tf.nn.batch_normalization().
2019-06-19 10:59:40.152462: I tensorflow/core/platform/cpu_feature_guard.cc:141]
Your CPU supports instructions that this TensorFlow binary was not compiled to
use: AVX2
2019-06-19 10:59:41.754231: W tensorflow/core/framework/allocator.cc:124] Allocation
of 56401920 exceeds 10% of system memory.

Evaluation time (1-image): 5.705s

non_cml 0.92531645
cml 0.07468357

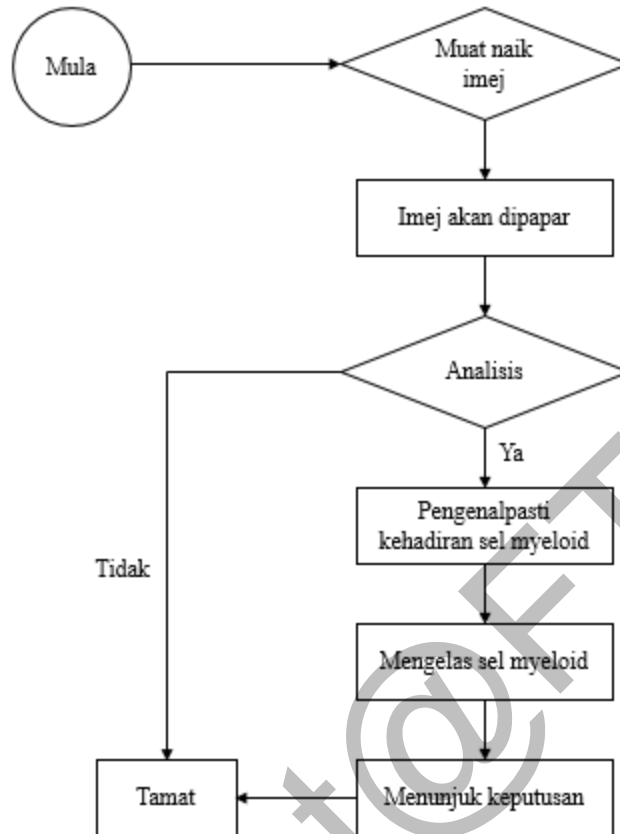
C:\Users\admin\Desktop\tensorflow-for-poets-2(use)>

```

Rajah 4.2 Keputusan Pengelasan Imej Slaid Darah

Rajah 4.2 menunjukkan salah satu imej slaid darah dari kategori NON-CML telah berjaya dikelaskan.

Selepas model berjaya melatih dan mengelas, model ini akan dirancang untuk import ke dalam sesuatu antaramuka pengguna. Rajah 4.3 menunjukkan carta alir sistem perisian yang dirancang dalam projek ini. Dengan carta alir ini, pengguna akan lebih mudah untuk memahami aktiviti dalam sistem perisian ini. Dari rajah 4.3, langkah pertama proses ini ialah pengguna perlu memuat naik imej slaid darah dan imej akan dipaparkan dalam antaramuka sistem jika imej dimuat naik dengan berjaya. Selepas itu, langkah lain akan diteruskan. Jika imej tidak berjaya dianalisis, sistem tidak akan meneruskan langkah yang seterusnya. Setelah imej berjaya dianalisis keputusan akan dikeluarkan oleh sistem.



Rajah 2.3 Carta Alir Sistem

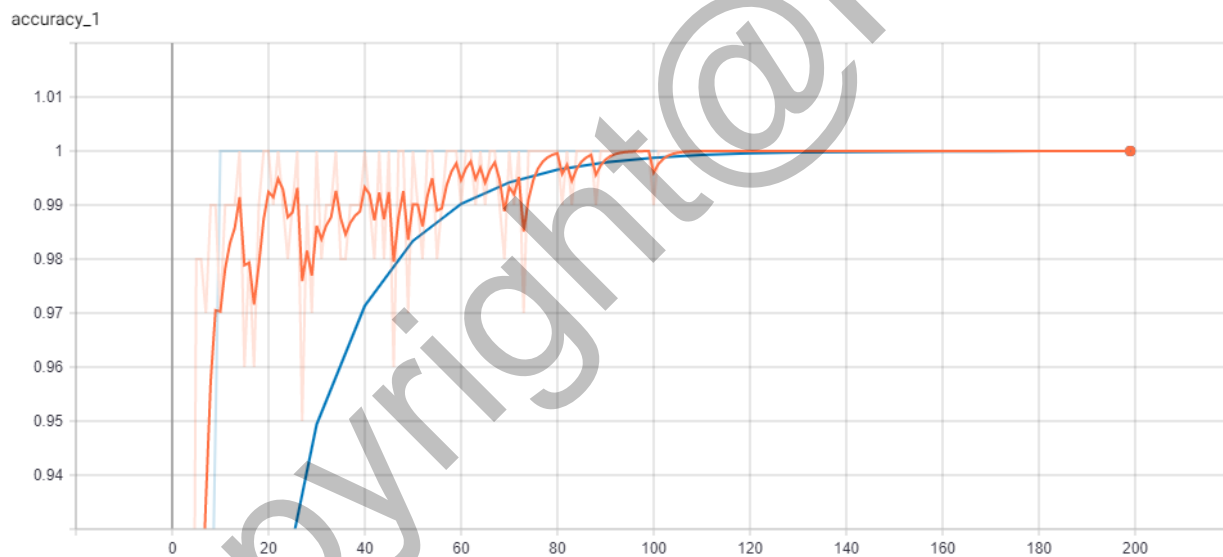
#### 4.4 Fasa Pengujian

Fasa ini merupakan fasa untuk menguji model yang telah dihasilkan. Dalam pengujian ini sebanyak 186 keping imej slaid darah akan digunakan iaitu 93 keping ialah imej slaid chronic myeloid leukaemia dan 93 keping ialah bukan chronic myeloid leukaemia. Data set yang digunakan untuk pengujian adalah imej slaid yang berbeza dengan data set untuk menjalankan latihan. Dalam pengujian ini, matriks kekeliruan akan digunakan. Matriks kekeliruan adalah jadual yang sering digunakan untuk menggambarkan prestasi model klasifikasi pada satu set data ujian yang mana nilai-nilai sebenar diketahui. Berikut merupakan matriks kekeliruan model ini.

## 5 HASIL KAJIAN

Dalam kajian ini, ketetapan latihan, ketetapan pengesahan, persilangan entropi dan matriks kekeliruan akan dibincangkan.

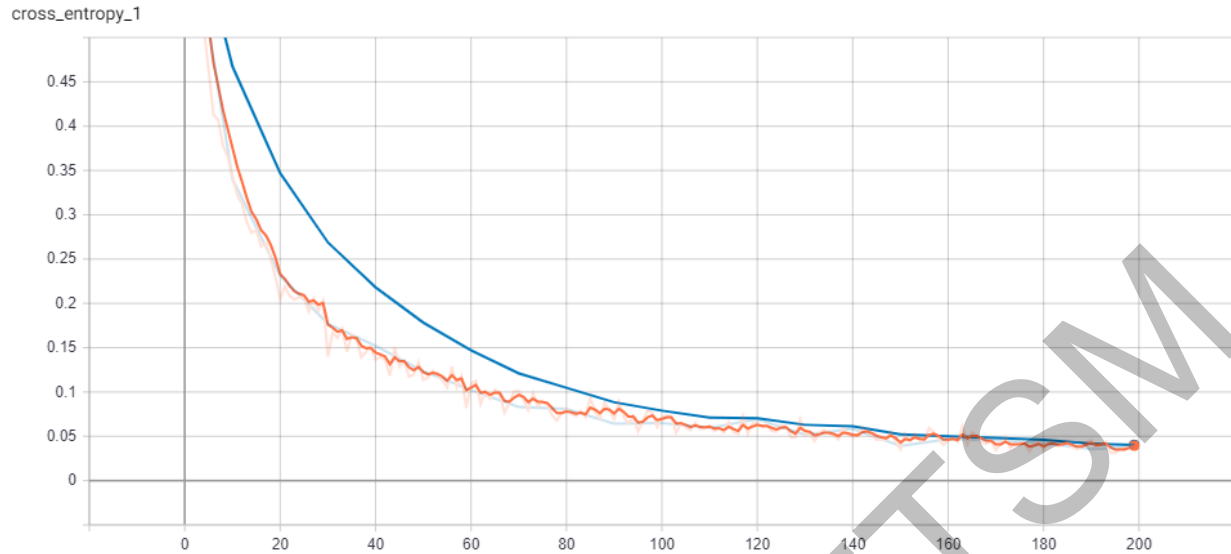
Dalam Rajah 5.1, paksi x merupakan bilangan langkah untuk melatih model pengelasan. Bilangan langkah yang digunakan dalam model ini ialah 200 dan paksi y merupakan ketetapan. Garis merah menunjukkan ketetapan latihan dan garis biru merupakan ketetapan pengesahan. Dari graph di atas, sebelum langkah 100, ketetapan latihan merupakan tidak stabil dan selepas langkah 100 ia menunjukkan ketetapan latihan yang stabil. Untuk ketetapan pengesahan, ia berjalan dengan lancar seperti yang ditunjukkan dalam graf. Ia menunjukkan bahawa model itu tidak ada yang kurang memuaskan atau mengatasi data kerana ketetapan pengesahan tidak menunjukkan lebih kecil atau lebih besar daripada ketetapan latihan.



Rajah 5. 1 Graf Ketetapan Latihan dan Ketetapan Pengesahan

Dalam Rajah 5.2 telah menunjukkan persilangan entropi. Persilangan entropi adalah fungsi kerugian yang memberikan gambaran tentang bagaimana proses pembelajaran berjalan dengan lancar (bilangan yang lebih rendah adalah lebih baik). Rajah ini telah menunjukkan lengkung pengesahan adalah sama dengan lengkung latihan, ia boleh membuat kesimpulan bahawa rangkaian telah dilatih dengan betul.





Rajah 5. 2 Graf Persilangan Entropi

Seterusnya, pengujian model akan dibincangkan. Dalam pengujian ini, matriks kekeliruan akan digunakan. Matriks kekeliruan adalah jadual yang sering digunakan untuk menggambarkan prestasi model klasifikasi pada satu set data ujian yang mana nilai-nilai sebenar diketahui. Berikut merupakan matriks kekeliruan model ini.

		Jangkaan	
		Positif	Negatif
Sebenar	Positif	85	8
	Negatif	14	79

Jadual 5. 1 Matriks Kekeliruan

Ketetapan pengelasan model imej slaid chronic myeloid leukemia adalah 88%.

## 6 KESIMPULAN

Pengelasan imej chronic myeloid leukaemia memerlukan masa yang panjang untuk analisis kerana CML mempunyai tiga fasa iaitu fasa kronik, fasa pacutan dan fasa pecahan. Dengan masalah ini, projek ini akan membangunkan sebuah model yang membolehkan pakar hematologic

untuk menganalisis sel darah CML. Walaupun model ini akan membantu pakar hematologi untuk mengurangkan masa dalam pengelasan imej slaid chronic myeloid leukaemia tetapi model ini masih mempunyai kekurangan.

Kekurangan dalam model ini ialah ia tidak mempunyai sesuatu antaramuka pengguna yang senang untuk digunakan oleh pengguna. Selain itu, model 'pra-terlatih' yang digunakan untuk melatih imej slaid chronic myeloid leukaemia hanya menggunakan Inception v3 sahaja. Model ini juga hanya boleh mengelaskan imej ke kategori CML atau NON-CML sahaja. Cadangan penambahbaik yang boleh dibuat bagi model ini ialah menambah fitur lagi dalam model ini seperti mengelaskan fasa CML. Selain itu, model pra-terlatih untuk melatih imej dalam kajian ini hanya menggunakan model Inception v3. Jadi, untuk meningkatkan ketetapan, pengelasan ini boleh melatih dengan model yang lain seperti AlexNet untuk memerhati model mana yang sesuai untuk pengelasan ini. Selain itu, sesuatu antaramuka pengguna yang mesra kepada pengguna boleh dibangunkan.

Kesimpulannya, model ini dapat dibangunkan mengikut objektif dan skop kajian yang telah ditetapkan. Halangan yang terdapat pada model ini akan dijadikan sebagai panduan dalam cadangan untuk menambahbaik model ini supaya ia akan menjadi lebih efisien.

## 7 RUJUKAN

Berry, J. L. (n.d.). *What Are White Blood Cells*. Retrieved from University of Rochester Medical Center:

<https://www.urmc.rochester.edu/encyclopedia/content.aspx?ContentID=35&ContentTypeID=160>

Gabry, O. E. (2017, March 18). *Software Engineering—Software Process and Software Process Models (Part 2)*. Retrieved from A Medium Corporation[US]:

<https://medium.com/omarelgabrys-blog/software-engineering-software-process-and-software-process-models-part-2-4a9d06213fdc>

Lynne Eldridge, M. (2018, November 24). *Types and Function of White Blood Cell*. Retrieved from verywellhealth: <https://www.verywellhealth.com/understanding-white-blood-cells-and-counts-2249217>

DexterLei. (2017, October 13). *简书*. Retrieved from TensorFlow 学习笔记: Retrain Inception\_v3 (二) : <https://www.jianshu.com/p/55e5847f7d36>

Feedback from people affected by CML. (2017). *Chronic Myeloid Leukaemia A guide for patients and families*. Australia: Leukaemia Foundation.

Clopton, J. (2016, December 18). *Advanced Phases of Chronic Myelogenous Leukemia*. Retrieved from WebMD: <https://www.webmd.com/cancer/lymphoma/features/cml-chronic-phase-tips#1>

Copyright@FTSM