

RESEARCH ARTICLE

Deteksi Kanker Kulit Melanoma Menggunakan Derivative of Gaussian dan Convolutional Neural Network

Rahadian Yusuf Abdul Rahman, Febryanti Sthevanie* and d Gamma Kosala

Fakultas Informatika, Universitas Telkom, Bandung, 40257, Jawa Barat, Indonesia

*Corresponding author: febryantisthevanie@telkomuniversity.ac.id

Received on 25 April 2024; accepted on 24 May 2024

Abstrak

Melanoma merupakan jenis kanker kulit paling mematikan di seluruh dunia. Namun, kanker kulit merupakan salah satu kanker yang mudah untuk disembuhkan karena terletak di bagian luar tubuh dan apabila dapat dideteksi sejak dini maka penyakit ini dapat ditangani lebih efektif. Ada beberapa metode klasifikasi yang digunakan untuk deteksi dini kanker kulit seperti SVM, KNN dan CNN. Dalam hal ini, jurnal ini menunjukkan metode untuk mendeteksi kanker kulit melanoma menggunakan derivative of gaussian(DOG) dan convolutional neural network(CNN). Akan dilakukan 5 skenario pengujian untuk mengetahui bagaimana performa dari apa yang diusulkan, salah satu dari skenario tersebut menggunakan gaussian filter.

Key words: CNN, DOG, Kanker Kulit, Skenario, Gaussian filter

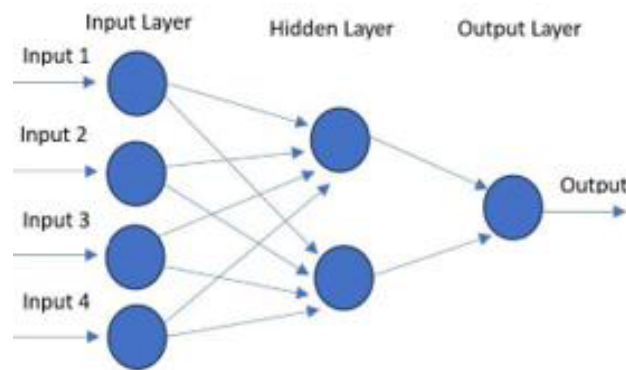
Pendahuluan

Penyakit kulit adalah kelainan kulit yang diakibatkan adanya jamur, kuman, parasit, virus maupun infeksi yang dapat menyerang siapa saja. Penyakit kulit dapat menyerang seluruh atau sebagian tubuh tertentu dan dapat membahayakan kondisi kesehatan penderita jika tidak ditangani dengan serius. Faktor-faktor yang dapat menyebabkan gangguan pada kulit yang sering ditemui misalnya faktor lingkungan, iklim, tempat tinggal, kebiasaan hidup yang kurang sehat, alergi dan lain-lain[8]. Human Development Index(HDI) sangat berhubungan dengan tingkat kematian akibat penyakit kulit. Semakin tinggi HDI, maka akses fasilitas Kesehatan, deteksi dan penanganan dini pada penyakit semakin baik, sehingga akan mengurangi angka kematian[2]. Dalam statistik, tingkat bertahan hidup dalam jangka waktu 5 tahun adalah 98% bagi penderita yang terdeteksi pada stadium awal, sedangkan hanya sekitar 2050% yang bertahan hidup dalam jangka waktu 5 tahun yang terdeteksi pada stadium lanjut[3]. Ada beberapa metode klasifikasi yang digunakan untuk mendeteksi kanker kulit yakni SVM, KNN dan CNN dengan akurasi CNN yang terbaik diantara ketiganya[4]. Karena deteksi dini penyakit kulit sangat penting dan akurasi CNN yang tinggi pada deteksi kanker kulit, maka diajukan deteksi penyakit kulit menggunakan derivative of gaussian(DOG) dan convolutional neural network(CNN).

Tinjauan Pustaka

Convolutional Neural Network

Convolutional neural network sejalan dengan ANN yakni tersusun atas neuron-neuron yang melakukan optimasi mandiri dengan belajar



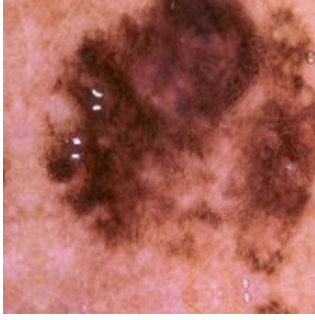
Gambar 1. Struktur dasar neural network

atau learning. Semua tip dan trik dalam pengembangan ANN dapat digunakan dalam pengembangan CNN [1].

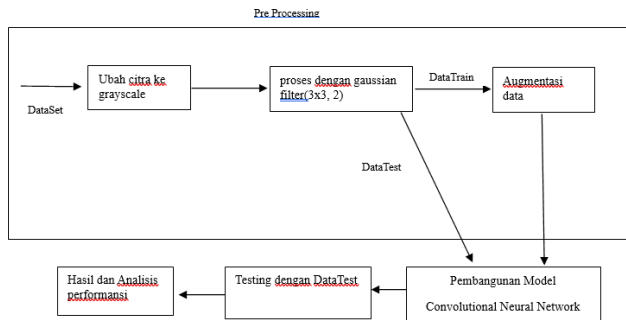
Perbedaan paling mencolok pada ANN dan CNN adalah CNN utamanya digunakan untuk pengenalan pola pada gambar [1].

Derivatives of Gaussian

Derivatives of Gaussian digunakan untuk pengurangan noise/denoise dan deteksi edge/batas dalam 1 langkah. Berikut adalah fungsi univariate Gauss:



Gambar 2. Contoh penyakit kulit (Melanoma)



Gambar 3. Metodologi

$$g_{\sigma\mu}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

σ adalah standar deviasi dan μ sebagai mean. Dalam filtering, mean selalu $\mu = 0$, standar deviasi σ sebagai parameter yang menentukan lebar dari filter [2].

Penyakit Kulit

Penyakit kulit adalah kelainan kulit yang diakibatkan adanya jamur, kuman, parasit, virus maupun infeksi yang dapat menyerang siapa saja. Penyakit kulit dapat menyerang seluruh atau sebagian tubuh tertentu dan dapat membahayakan kondisi kesehatan penderita jika tidak ditangani dengan serius. Faktor-faktor yang dapat menyebabkan gangguan pada kulit yang sering ditemui misalnya faktor lingkungan, iklim, tempat tinggal, kebiasaan hidup yang kurang sehat, alergi dan lain-lain.

Metodologi Penelitian

Metodologi

Berikut penjelasan dari masing-masing tahapan dalam membangun model:

1. **Pre-processing:** DataTrain dan DataTest diinput, lalu diubah menjadi citra grayscale. Citra kemudian diperhalus dengan fungsi Gaussian filter dengan ukuran kernel 3x3 dan $\sigma = 2$, kemudian disimpan dalam folder dataset. DataTrain yang disimpan di folder tersebut kemudian dilakukan augmentasi data dengan probabilitas rotasi 70% dengan maksimum rotasi ke kiri sebesar 10° dan maksimum rotasi ke kanan sebesar 10° .
2. **Processing:** DataTrain dan DataTest kemudian masuk ke CNN. CNN akan melakukan training menggunakan DataTrain dengan jumlah epoch=30. DataTrain juga digunakan sebagai data validasi untuk mencegah overfitting. Setelah itu dilakukan pengujian menggunakan DataTest.

Class	Class Name	DataTrain
0	squamous cell carcinoma	181
1	dermatofibroma	95
2	actinic keratosis	114
3	pigmented benign keratosis	462
4	nevus	357
5	vascular lesion	139
6	seborrheic keratosis	77
7	basal cell carcinoma	376
8	melanoma	438
Total		2239

Gambar 4. Tabel Data Train



Gambar 5. Sample Dataset

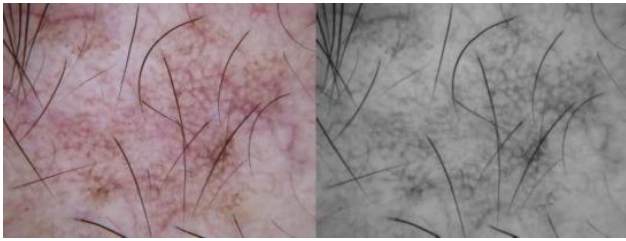
3. **Hasil dan analisis performansi:** Hasil proses berupa klasifikasi 9 kelas kelainan kulit yakni squamous cell carcinoma, dermatofibroma, actinic keratosis, pigmented benign keratosis, nevus, vascular lesion, seborrheic keratosis, basal cell carcinoma, melanoma. Setelah itu dianalisis performansi yang dihasilkan.

Dataset

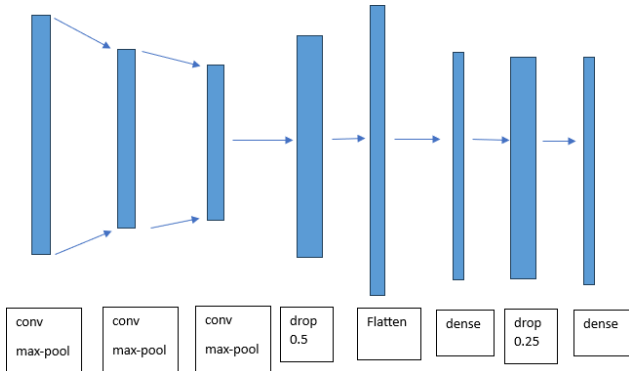
Dataset yang digunakan berasal dari ISIC. Dataset berupa sembilan kelas berbagai macam kelainan kulit. Dataset terdiri dari 2239 DataTrain dengan 118 DataTest. Berikut adalah persebaran DataTrain:

Augmentasi Data

Augmentasi data menggunakan library Augmentor dengan probabilitas rotasi 70% dengan maksimum rotasi ke kiri sebesar 10° dan maksimum rotasi ke kanan sebesar 10° .



Gambar 6. (a) Original (b) Diperhalus dengan Gaussian filter



Gambar 7. Arsitektur CNN

Gaussian Filter

Gaussian filter yang digunakan untuk membangun model ini memiliki ukuran kernel 3x3 dengan nilai $\sigma = 2$. Dataset diubah menjadi citra grayscale kemudian kernel ini diaplikasikan pada dataset grayscale tersebut.

Convolutional Neural Network

Model CNN yang digunakan dalam artikel ini memiliki tiga Convolution layers dengan ukuran filter 3x3 dengan fungsi ReLU activation, menggunakan tiga spatial max pool layers dengan ukuran 2x2, menggunakan dua dropout layers masing-masing dengan nilai 0.5 dan 0.25, menggunakan dua dense layers dan satu flatten layer.

Hasil dan Pembahasan

Dalam penelitian ini ada beberapa skenario yang dilakukan dalam upaya meningkatkan performansi dari sistem yang dibangun. Berikut penjelasan singkatnya untuk tiap skenario:

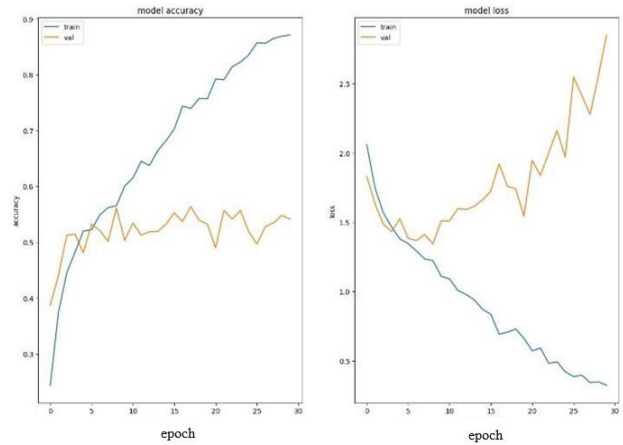
Skenario 1

Skenario 1 menggunakan model dasar CNN merujuk pada Bab 3, yakni menggunakan 3 convolutional layers, dengan dropout layer 1 bernilai 0,5 dan dropout layer 2 bernilai 0,25. Diperoleh hasil akurasi training 0,56376 dalam 30 epoch, ketika dilakukan pengujian menggunakan data test didapatkan hasil 0,3305.

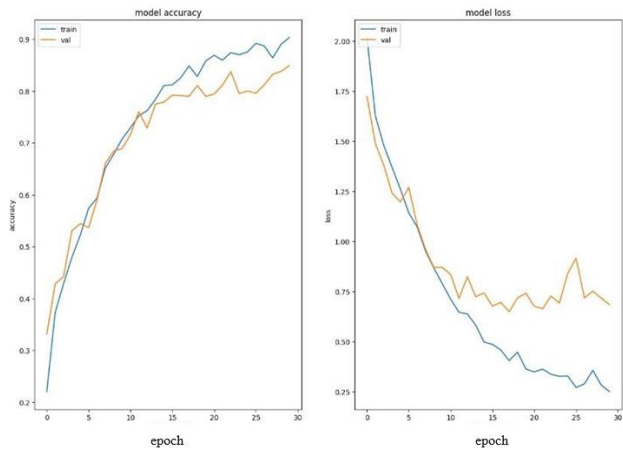
Berdasarkan dari hasil pengujian, hasil tersebut disebabkan adanya ketidak seimbangan data train untuk tiap-tiap kelas sehingga proses training CNN menjadi kurang maksimal.

Skenario 2

Skenario 2 menggunakan model dasar CNN merujuk pada Bab 3, yakni menggunakan 3 convolutional layers, dengan dropout layer 1 bernilai 0,5 dan dropout layer 2 bernilai 0,25, dilakukan modifikasi berupa



Gambar 8. Model Accuracy dan Model Loss Skenario 1



Gambar 9. Model Accuracy dan Model Loss Skenario 2

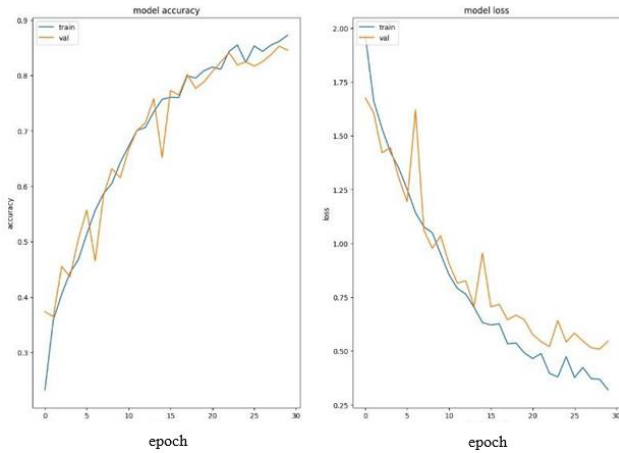
penambahan augmentasi data sehingga data train menjadi seimbang dan data yang sedikit menjadi lebih banyak. Diperoleh hasil akurasi training 0,84929 dalam 30 epoch. Ketika dilakukan pengujian dengan data test didapatkan hasil 0,4237.

Berdasarkan dari hasil pengujian di atas, peningkatan akurasi pengujian sebesar 0,0932 disebabkan oleh augmentasi data. Fungsi augmentasi data adalah untuk meng-generate data train baru berdasarkan data train yang sudah ada, sehingga hasil training pada tiap epoch CNN menjadi lebih optimal.

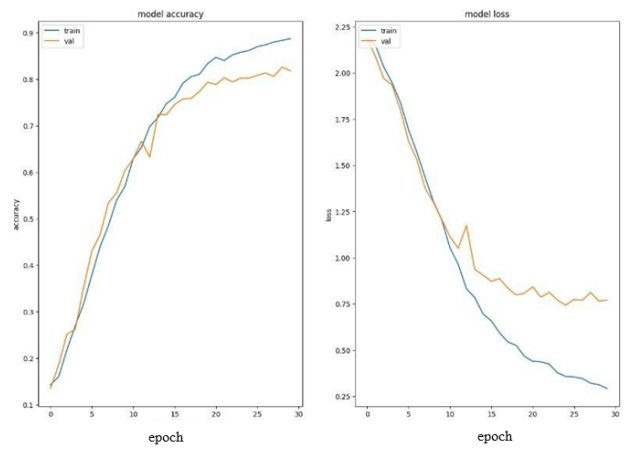
Skenario 3

Skenario 3 adalah model dari Skenario 2 dengan penambahan 1 convolutional layer baru sehingga jumlah convolutional layer menjadi 4. Diperoleh hasil akurasi training 0,85301 dalam 30 epoch. Ketika dilakukan pengujian dengan data test didapatkan hasil 0,4407.

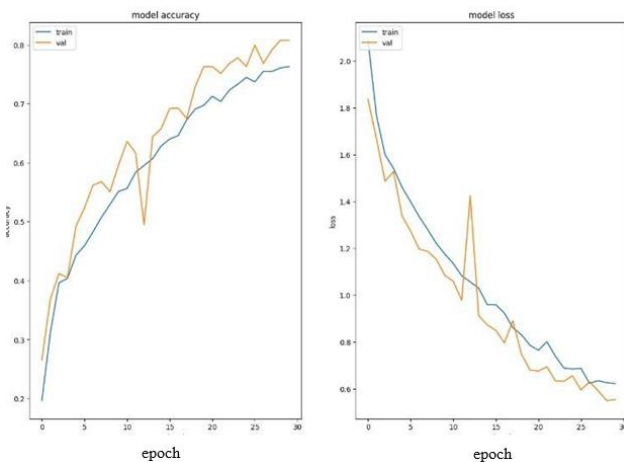
Berdasarkan dari hasil pengujian di atas, terdapat peningkatan yang sangat kecil yakni 0,017. Penambahan convolutional layer menjadi 4 meningkatkan sedikit akurasi model namun tidak memberikan peningkatan yang signifikan. Penambahan 1 convolutional layer pada CNN akan meningkatkan jumlah neuron yang digunakan untuk menganalisis citra.



Gambar 10. Model Accuracy dan Model Loss Skenario 3



Gambar 12. Model Accuracy dan Model Loss Skenario 5



Gambar 11. Model Accuracy dan Model Loss Skenario 4

Skenario 4

Skenario 4 menggunakan 4 convolutional layer dengan tambahan modifikasi berupa dropout layer 1 diubah menjadi 0,6 dan dropout layer 2 menjadi 0,4. Diperoleh hasil akurasi training 0,80772 dalam 30 epoch. Ketika dilakukan pengujian dengan data test didapatkan hasil 0,3983.

Berdasarkan dari hasil di atas, dropout layer 1 sebesar 0,6 dan dropout layer 2 sebesar 0,4 memiliki akurasi yang lebih rendah daripada Skenario 3 dengan selisih 0,0424. Penurunan akurasi kemungkinan besar disebabkan oleh ukuran dropout layer yang terlalu besar sehingga terlalu banyak neuron yang di-drop.

Skenario 5

Skenario 5 adalah model dari Skenario 3 dengan perubahan pada dataset yang dimanipulasi menggunakan Gaussian filter dengan kernel 3x3 dan sigma 2. Diperoleh hasil akurasi training 0,8258 dalam 30 epoch. Ketika dilakukan pengujian dengan data test didapatkan hasil 0,2712.

Hasil pengujian ini adalah hasil yang terburuk meskipun akurasi training cukup tinggi di angka 0,8258. Namun pada bagian loss juga cukup besar pada 0,7702, ini merupakan loss kedua terbesar setelah Skenario 1 yang menggunakan dataset default tanpa augmentasi. Dataset yang dimanipulasi menggunakan Gaussian filter pada kasus ini justru menjadi lebih buruk daripada sebelum dimanipulasi atau default, sehingga CNN kesulitan untuk melakukan klasifikasi.

Skenario	Class								Result	
	0	1	2	3	4	5	6	7		8
1	0	0,4325	0	0,1875	0,75	0,625	0	0,3125	0,667	0,3305
2	0,375	0,5	0,1875	0	0,875	0,75	0	0,3125	0,667	0,4237
3	0,3125	0,625	0,25	0,0625	0,875	0,6875	0	0,3125	0,667	0,4407
4	0,25	0,5	0,3125	0	0,9375	0,4375	0	0,375	0,667	0,398
5	0	0,5	0,125	0,1875	0,25	0,6875	0	0,1875	0,333	0,2712

Gambar 13. Tabel hasil akurasi per kelas dan keseluruhan pengujian

Analisis Hasil Pengujian

Cara hitung akurasi:

$$\text{Akurasi per kelas} = \frac{\text{Jumlah klasifikasi benar}}{\text{Jumlah data test per kelas}} \tag{2}$$

$$\text{Akurasi skenario} = \frac{\text{Jumlah klasifikasi benar}}{\text{Total semua data test}} \tag{3}$$

Dalam Tabel hasil pengujian (13) terdapat kelas yang konsisten pada akurasi 0%. Hal ini kemungkinan disebabkan oleh jumlah data yang paling sedikit di antara kelas lain, atau data tersebut memiliki kemiripan dengan kelas lain yang membuat CNN kebingungan dalam melakukan klasifikasi. Usaha untuk mengatasi hal ini menggunakan Gaussian filter untuk memanipulasi dataset, namun efek yang dihasilkan adalah sebaliknya.

Ada beberapa faktor yang mungkin menyebabkan hasil pengujian memiliki akurasi yang rendah. Tidak adanya data validasi, validasi ketika training

Kesimpulan

Dari beberapa skenario pengujian diatas terdapat beberapa kesimpulan yang dapat diambil sebagai berikut:

1. Augmentasi data digunakan untuk meng-generate data train sehingga apabila antar kelas dalam data train tidak seimbang, maupun terdapat keterbatasan jumlah dalam data train dapat diatasi, sehingga CNN dapat di train secara optimal.
2. Jumlah convolutional layer pada CNN dapat mempengaruhi performa CNN, terdapat peningkatan performansi CNN karena dengan menambah jumlah convolutional layer maka jumlah neuron pada CNN juga bertambah.
3. Dropout layers juga berpengaruh pada performa CNN. Dropout layers dengan ratio yang tidak tepat justru akan menurunkan performa dari CNN. Karena akan terlalu banyak neuron yang terbuang.

4. Skenario 3 merupakan skenario terbaik. Baik dalam hasil training model maupun dalam hasil pengujian menggunakan data test dengan akurasi training 0.85301 dan akurasi test 0.4407.
5. Skenario ke 5 merupakan skenario terburuk dari semua skenario pengujian yang dilakukan. Gaussian filter memang dapat digunakan untuk memanipulasi dataset namun dataset yang dihasilkan belum tentu lebih baik daripada dataset default, kemungkinan hasil akan berbeda jika filter yang digunakan diganti dengan 2d DoG, karena manipulasi citra bisa menjadi lebih baik.

Daftar Pustaka

1. O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. 2015 November.
2. Maucher PDJ. Object Recognition Lecture; 2021. Hochschule Der Medien.
3. Dimitriou F, Krattinger R, Ramelyte E, Barysch MJ, Micaletto S, Dummer R, et al. The world of melanoma: epidemiologic, genetic, and anatomic differences of melanoma across the globe. *Current oncology reports*. 2018;20(11):87.
4. Khazaei Z, Ghorat F, Jarrahi A, Adineh H, Sohrabivafa M, Goodarzi E. Global incidence and mortality of skin cancer by histological subtype and its relationship with the human development index (hdi); an ecology study in 2018. *World Cancer Research Journal*. 2019;6:13.
5. Schadendorf D, van Akkooi AC, Berking C, Griewank KG, Gutzmer R, Hauschild A, et al. Melanoma. *The Lancet*. 2018;392(10151):971-84.
6. Daghrir J, Tlig L, Bouchouicha M, Sayadi M. Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. In: *International Conference on Advanced Technologies for Signal and Image Processing*. IEEE; 2020. p. 1-6.
7. Wu X, Wang K, Zhang D. Palmprint Texture Analysis Using Derivative of Gaussian Filters. In: *International Conference on Control Automation Robotics & Vision (ICARCV)*. IEEE; 2006. p. 1-4.
8. Hay R, Bendeck SE, Chen S, Estrada R, Haddix A, McLeod T, et al. Disease Control Priorities in Developing Countries, Chapter 37: Skin Diseases. *The International Bank for Reconstruction and Development / The World Bank*; New York: Oxford University Press; 2006.
9. Li Q, Cai W, Wang X, Zhou Y, Feng DD, Chen M. Medical image classification with convolutional neural network. In: *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*. IEEE; 2014. p. 844-8.
10. Soliman ALKolifi ALEnezi N. A Method Of Skin Disease Detection Using Image Processing And Machine Learning. *Procedia Computer Science*. 2019;163:85-92.
11. Zhang B, Zhang L, Zhang L, Karray F. Retinal vessel extraction by matched filter with first-order derivative of Gaussian. *Computers in Biology and Medicine*. 2010;40(4):438-45.