

JTIK ID 3399

by Jtik 3399

Submission date: 24-Mar-2020 10:29AM (UTC+0700)

Submission ID: 1280890889

File name: 3399-12667-1-RV.docx (1.02M)

Word count: 2420

Character count: 15042

Studi Awal Deteksi COVID-19 Menggunakan Citra CT Berbasis Deep Learning

Abstrak

Pada bulan Desember 2019, virus COVID-19 menyebar ke banyak negara, termasuk di Indonesia yang kemudian menjadi pandemi dan menimbulkan masalah serius karena masih belum adanya vaksin untuk mencegah penularan. Uji spesimen saluran nafas atas dan saluran nafas bawah saat ini merupakan salah satu metode yang efektif untuk mengetahui apakah seseorang terinfeksi COVID-19 atau tidak. Dengan semakin banyaknya *suspect*, maka permintaan uji spesimen di laboratorium juga semakin meningkat. Penelitian ini merupakan studi awal untuk melihat apakah citra CT dari organ thorax dapat digunakan sebagai alternatif untuk mendeteksi infeksi virus COVID-19. Metode deep learning digunakan untuk membuat sebuah model dengan citra CT sebagai masukan. Total 140 data citra CT yang terbagi menjadi 2 yaitu citra dari pasien terinfeksi dan citra dari subjek normal digunakan sebagai masukan pada deep learning. Proses pelatihan dilakukan menggunakan CNN dengan arsitektur VGG16 dan optimizer SGD dan Adam. Hasil yang didapatkan adalah akurasi sebesar 92,86% untuk mengklasifikasikan infeksi COVID-19 dan normal. Nilai spesifisitas dan sensitivitas sebesar 100% dan 85,71% untuk pelatihan dengan menggunakan optimizer SGD.

Kata kunci: *Deep Learning, COVID-19, VGGNet, CT-Scan*

Preliminary Study of COVID-19 Detection Using CT Image Based on Deep Learning

Abstract

In December 2019, the COVID-19 virus spread to many countries, including Indonesia which later became a pandemic and caused serious problems because there was still no vaccine to prevent transmission. Tests of upper and lower respiratory tract specimens are now an effective method of finding whether a person is infected with COVID-19 or not. With the increasing number of suspects, the demand for specimen testing in the laboratory is also increasing. This research is a preliminary study to see whether CT images of the thorax organs can be used as an alternative to detect COVID-19 virus. The deep learning method is used to create a model with CT images as input. A total of 140 CT image data which are divided into 2 images from infected patients and images from normal subjects are used as input for deep learning. The training process is carried out using CNN with VGG16 architecture and SGD and Adam optimizers. The results obtained are 92.86% accuracy for classifying COVID-19 infections and normal. Specificity and sensitivity values were 100% and 85.71% for training using the SGD optimizer.

Keywords: *Deep Learning, COVID-19, VGGNet, CT-Scan*

1. PENDAHULUAN

Coronavirus (COVID-19) adalah jenis baru virus yang diidentifikasi pertama kali di Wuhan, China pada bulan Desember 2019 (Li dkk, 2020). Infeksi ini kemudian menyebar ke seluruh China dan negara-negara lain di seluruh dunia yang ditetapkan sebagai pandemi oleh Organisasi Kesehatan Dunia (WHO).

Tanda-tanda umum infeksi dari virus ini adalah demam tinggi ($>38^{\circ}\text{C}$), batuk, sesak napas dan kesulitan bernapas. Pada kasus yang lebih parah, infeksi dapat menyebabkan pneumonia, sindrom

pernapasan akut, gagal ginjal, dan bahkan kematian (Wang dkk, 2020).

Di Indonesia, kasus pertama COVID-19 diidentifikasi pada tanggal 2 Maret 2020. Hingga 17 Maret 2020, ada 134 kasus yang telah terkonfirmasi yang tersebar di delapan provinsi, yaitu Bali, Banten, DKI Jakarta, Jawa Barat, Jawa Tengah, Kalimantan Barat, Sulawesi Utara dan Yogyakarta (KOMPAS, 2020).

Protokol untuk melakukan pengujian apakah subjek terinfeksi COVID-19 harus berdasarkan pada faktor klinis, epidemiologis serta penilaian dari kemungkinan infeksi. Pemeriksaan spesimen di

laboratorium dengan menggunakan Reverse Transcriptase PCR (RT-PCR) dan dikonfirmasi dengan teknik sekuensing dilakukan untuk memastikan infeksi COVID-19 (WHO, 2020). Spesimen yang diambil adalah spesimen dari saluran 11 nafasan atas (nasofaring dan orofaring) dan saluran nafas bawah seperti dahak, aspirat trakea dan bilasan bronkoalveolar.

Seperti halnya di negara-negara lain, suspect Corona (orang yang menunjukkan gejala terinfeksi COVID-19 dan diduga pernah melakukan kontak dengan pasien positif COVID-19) semakin meningkat di Indonesia sehingga permintaan uji spesimen pada laboratorium yang telah ditunjuk oleh pemerintah juga meningkat (RRI, 2020). Untuk itu diperlukan suatu metode lain dalam melakukan pendeteksian COVID-19 yang lebih praktis dan cepat.

Pemeriksaan radiologis dengan menggunakan citra CT dapat digunakan sebagai deteksi awal infeksi COVID-19 untuk mengidentifikasi infeksi paru fase awal (Zu dkk, 2020). Temuan dari (Ng dkk, 2020) menunjukkan adanya ground-glass opacity (GGO) pada citra CT penderita COVID-19. Lebih lanjut, (Zhou dkk, 2020) mengamati citra CT dari 62 penderita COVID-19 dan sebanyak 40,3% terdapat GGO pada citra CT; 62,9% terdapat GGO dengan pola reticular; 45,2% terdapat vacuolar sign; 56,5% terdapat microvascular dilation sign; dan 33,9% terdapat fibrotic streak. Dari studi tersebut, dapat dilihat bahwa ada perubahan fisiologis yang nampak dari organ thorax penderita COVID-19.

Penelitian ini akan mengobservasi apakah pola-pola perubahan fisiologis dari organ thorax penderita COVID-19 yang diakuisisi dengan CT dapat dikenali menggunakan deep learning. Algoritma dari deep learning yang akan diimplementasikan adalah convolution neural network (CNN) yang telah banyak diimplementasikan untuk melakukan analisis dan mengenali objek visual dari citra. Bahasa pemrograman Python dan library Tensorflow digunakan untuk melakukan training dan membuat model dari CNN.

2. Metode Penelitian

Alur penelitian yang dilakukan dibagi menjadi beberapa tahapan, yaitu: analisis kondisi, pengumpulan data, desain sistem, proses training dan evaluasi.

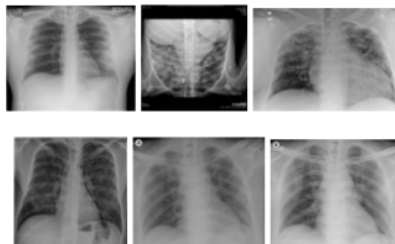
2.1. Analisis Kondisi

Langkah ini menganalisis kondisi-kondisi yang perlu diperhatikan saat penelitian dilakukan. Beberapa kondisi yang perlu diperhatikan antara lain: citra CT yang akan dianalisis merupakan citra dengan posisi posterior-anterior (PA) dari organ thorax. Ukuran citra perlu diseragamkan resolusinya dengan cara melakukan penskalaan (*scaling*) agar didapatkan citra dengan dengan resolusi yang sama.

Terakhir, citra dipisahkan menjadi 2 jenis data yaitu data pasien yang terinfeksi COVID-19 (data positif) dan data normal yang tidak terinfeksi COVID-19 (data negatif).

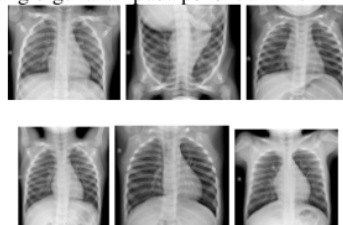
2.2. Pengumpulan Data

Pada penelitian ini data citra CT dari pasien yang terinfeksi COVID-19 berasal dari dataset GitHub yang dikumpulkan oleh Dr. Joseph Cohen (<https://github.com/ieee8023/covid-chestxray-dataset>). Data ini dirilis secara publik yang terdiri dari citra CT dari kasus COVID-19, MERS, SARS dan ARDS. Data citra CT dipilih hanya untuk kasus COVID-19 dengan posisi organ thorax posterior-anterior. Total citra CT yang didapatkan untuk data pasien sebanyak 70. Gambar 1 menunjukkan beberapa contoh citra CT dari pasien COVID-19 yang digunakan sebagai data latih pada penelitian ini.



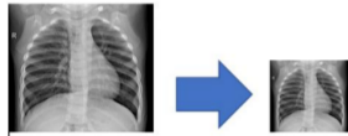
Gambar 1. Contoh data pasien sebagai data latih

Untuk data normal, didapatkan dari dataset Kaggle yang diambil secara acak sebanyak 70 data (<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>). Dataset ini semua terdiri dari CT subjek yang normal (tidak ada pneumonia) dan dalam posisi PA. Gambar 2 adalah contoh data normal yang digunakan pada penelitian ini.



Gambar 2. Contoh data normal sebagai data latih

Total data citra CT adalah 140 data. Preprocessing untuk semua data ini perlu dilakukan sebelum digunakan sebagai input pada CNN. Pada penelitian ini, preprocessing yang dilakukan adalah penskalaan menjadi 224x224 sebagai standard input untuk CNN dan arsitektur VGG16. Contoh proses *scaling* yang dilakukan pada penelitian ini dapat dilihat pada gambar 3.

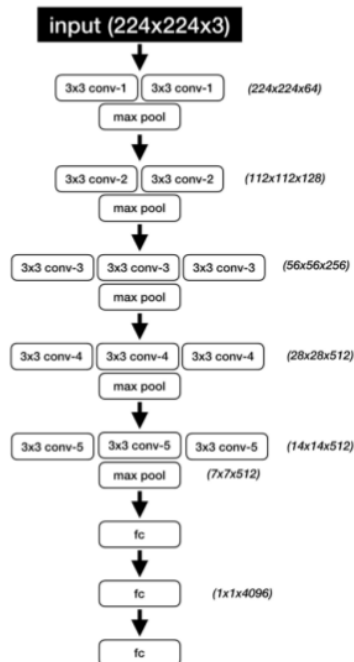


Gambar 3. Scaling pada praproses citra

2.3. Arsitektur CNN

Pada tahapan ini akan dibuat desain dari CNN yang akan digunakan dalam proses training serta deteksi infeksi COVID-19 pada citra CT. Arsitektur CNN yang digunakan pada penelitian ini adalah VGG16 (Simonyan dan Zisserman, 2014). Arsitektur VGG16 secara umum terdiri dari 16 layer. Di setiap layer-nya terdiri dari convolutional layer, max pooling layer, activation layer dan diakhiri dengan fully connected layer. Gambar 4 menunjukkan diagram arsitektur VGG16.

Citra input untuk arsitektur VGG16 adalah citra berukuran 224×224 dengan 3 channel (RGB). Pada conv-1, diberikan 64 filter dengan kernel 3×3 yang diikuti dengan fungsi aktivasi rectified linear unit (ReLU) dan max pooling untuk mengurangi dimensi spasial. Pada conv-2 terdapat dengan 128 filter, conv-3 terdapat 256 filter, conv-4 terdapat 512 filter, dan conv-5 terdapat 512 filter. Layer berikutnya adalah 3 fully connected (fc) layer dengan jumlah node sebesar 4096 dengan aktivasi ReLU dan diakhiri dengan output layer dengan aktivasi softmax. Total ada 13 layer convolutional, 5 layer max pooling, dan 3 fully connected layer.



Gambar 4. Arsitektur VGG16

2.4. Proses Training

Proses *training* CNN untuk arsitektur VGG16 secara umum memiliki 2 bagian utama, yaitu *feedforward* dan *backpropagation* (Simonyan dkk, 2015). Proses *feedforward* adalah proses citra sebagai *input layer* memulai proses dari *convolution layer*, *pooling layer*, hingga sampai memasuki *fully connected layer* yang merupakan lapisan *output*. Hasil dari proses ini adalah bobot akhir yang merupakan klasifikasi dari citra masukan. Apabila perlu dilakukan perbaikan bobot, maka proses selanjutnya adalah proses *backpropagation*. Proses *backpropagation* ini menggunakan hasil dari proses *feedforward* sebelumnya untuk melakukan proses *feedforward* yang akan dilakukan. Proses *backpropagation* ini berfungsi untuk memperbaiki bobot yang dihasilkan oleh proses *feedforward*.

Dari total 140 data citra yang terbagi menjadi 70 data pasien dan 70 data normal. Sebanyak 110 data akan dijadikan data training dan 30 data dijadikan data testing yang dipilih secara acak. Teknik augmentasi dalam bentuk rotasi juga akan diimplementasikan guna meningkatkan jumlah data training.

2.5. Uji Coba dan Evaluasi

Uji statistik dilakukan untuk melihat akurasi model dalam menguji infeksi COVID-19 terhadap data pasien dan data normal. Pengujian dilakukan dengan melihat nilai dari sensitivitas (*true positive rate*) dan spesifisitas (*true negative rate*). Pengujian sensitivitas didapatkan melalui pembandingan jumlah identifikasi infeksi COVID-19 yang tepat terhadap seluruh kelas yang ada (pasien dan normal). Pengujian spesifisitas didapatkan melalui pembandingan jumlah identifikasi data normal terhadap seluruh kelas yang ada. Nilai sensitivitas yang tinggi menunjukkan akurasi dalam mengidentifikasi infeksi COVID-19 dengan tepat dari data yang ada, sedangkan nilai spesifisitas yang tinggi menunjukkan akurasi dalam mengidentifikasi data normal (negatif/tidak terinfeksi COVID-19) dengan tepat.

Untuk mendapatkan nilai sensitivitas dan spesifisitas, terlebih dulu dibuat sebuah confusion matrix yang melakukan perhitungan terhadap jumlah true positive (TP), true negative (TN), false positive (FP) dan false negative (FN). Bentuk dari *confusion matrix* ini dapat dilihat pada Tabel 1.

Tabel 1. Matrix Confusion untuk proses klasifikasi

	Klasifikasi	
	Terinfeksi	Normal
Terinfeksi	TP	FN
Normal	FP	TN

Pada penelitian ini, hanya ada 2 kelas yang digunakan, yaitu terinfeksi dan normal. Nilai TP

didapatkan apabila suatu data positif terinfeksi COVID-19 dan dikategorikan ke dalam kelas terinfeksi. Nilai FP didapatkan apabila suatu data normal dan dikategorikan ke dalam kelas terinfeksi. Nilai FN didapatkan apabila suatu data positif terinfeksi dan dikategorikan ke dalam kelas normal. Terakhir, nilai TN didapatkan apabila suatu data normal (tidak terinfeksi COVID-19) dan dikategorikan ke dalam kelas normal.

Rumus untuk menghitung akurasi dengan menggunakan *matrix confusion* dapat dilihat pada rumus (1).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

TP dan TN adalah klasifikasi yang tepat (terinfeksi diklasifikasikan sebagai terinfeksi, dan normal diklasifikasikan sebagai normal). Total dari TP dan TN akan dibandingkan dengan keseluruhan jumlah data yang ada (TP, TN, FP dan FN). Jika tidak didapatkan nilai FP dan FN, maka akurasi akan bernilai 1 yang berarti dapat mengklasifikasikan semua data dengan tepat.

Rumus (2) dan (3) digunakan untuk mendapatkan nilai sensitivitas dan spesifisitas.

$$\text{Sensitivitas} = \frac{TP}{TP+FN} \quad (2)$$

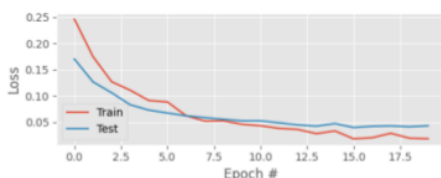
$$\text{Spesifisitas} = \frac{TN}{TN+FP} \quad (3)$$

3. HASIL DAN PEMBAHASAN

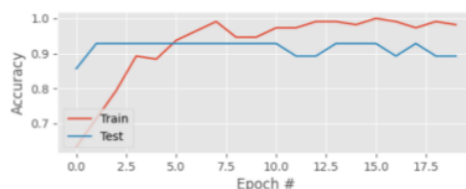
3.1. Loss Function dan Akurasi

Dalam proses training, digunakan stochastic gradient descent (SGD) dan Adam Optimizer yang merupakan pengembangan dari SGD (King dan Ba, 2014). Jumlah epoch yang dilakukan dalam proses training adalah sebanyak 20 epoch. Loss function yang digunakan adalah mean squared error (MSE).

Gambar 5 dan 6 menunjukkan grafik loss function dan akurasi pada saat training menggunakan Adam optimizer. Hasil akhir dari loss function adalah 0,0210 sedangkan nilai dari validation loss sebesar 0,0409. Ini menunjukkan adanya overfitting pada model yang mengindikasikan tingginya variasi pada data training. Untuk hasil akhir akurasi adalah 89,29%

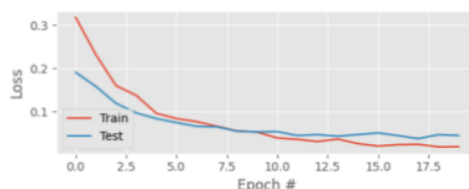


Gambar 5. Loss function untuk pelatihan menggunakan Adam optimizer

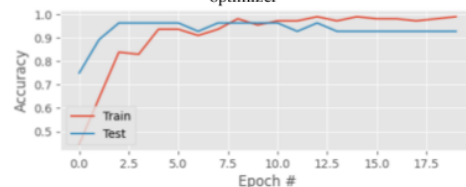


Gambar 6. Akurasi untuk pelatihan menggunakan Adam optimizer

Untuk loss function dan akurasi menggunakan SGD dapat dilihat pada gambar 7 dan 8.



Gambar 7. Loss function untuk pelatihan menggunakan SGD optimizer



Gambar 8. Akurasi untuk pelatihan menggunakan SGD optimizer

Hasil akhir loss function dari training yang menggunakan SGD optimizer adalah 0,0186, sedangkan validation loss sebesar 0,0431. Hasil ini mirip dengan Adam optimizer dan menunjukkan adanya tendensi overfitting pada model yang dihasilkan.

Pada kedua optimizer, nampak bahwa penurunan nilai loss dan peningkatan akurasi hanya sampai pada epoch ke-10. Penambahan jumlah epoch tidak banyak mengurangi nilai loss ataupun peningkatan akurasi.

3.2. Evaluation Metrics

Model yang dihasilkan dari proses training dievaluasi berdasarkan confusion matrix. Hasil dari confusion matrix untuk SGD optimizer ditunjukkan pada tabel 2 sedangkan hasil dari Adam Optimizer ditunjukkan pada tabel 3.

Tabel 2. Hasil confusion matrix untuk Adam Optimizer

	Klasifikasi	
	Terinfeksi	Normal
Terinfeksi	13	1
Normal	2	12

Tabel 3. Hasil confusion matrix untuk optimizer SGD

	Klasifikasi	
	Terinfeksi	Normal
Terinfeksi	14	0
Normal	2	12

Dari tabel 2 dan 3, didapatkan nilai akurasi, sensitivitas dan spesifisitas untuk masing-masing optimizer seperti yang ditunjukkan tabel 4.

Tabel 4. Perbandingan nilai akurasi, spesifisitas dan sensitivitas untuk optimizer Adam dan SGD

	Optimizer	
	Adam	SGD
Akurasi	89,29%	92,86%
Spesifisitas	92,86%	100%
Sensitivitas	85,71%	85,71%

Seperti yang nampak di tabel 4, nilai akurasi dan spesifisitas terbaik didapatkan menggunakan optimizer SGD. Optimizer SGD dapat melakukan klasifikasi pasien yang terinfeksi COVID-19 dengan tepat, sementara untuk data normal (negatif), ada 2 data yang terklasifikasikan sebagai pasien terinfeksi (positif).

4. KESIMPULAN

Berdasarkan model yang telah dihasilkan dari proses training CNN dengan data input citra CT, didapatkan kesimpulan bahwa model bisa mendeteksi infeksi COVID-19 dengan akurasi 100% dengan nilai spesifisitas dan sensitivitas sebesar 92,86% dan 85,71% dengan menggunakan optimizer SGD.

Model ini merupakan studi awal dengan data training yang terbatas jumlahnya. Model ini juga tidak dapat digunakan sebagai pengganti diagnosis atau uji spesimen laboratorium. Dengan semakin bertambahnya data, maka model juga akan semakin baik dalam melakukan prediksi infeksi COVID-19.

DAFTAR PUSTAKA

1. Coronavirus Disease 2019 (COVID-19): A Perspective from China. *Radiology*, p.200490. 10
- KOMPAS, 2020. UPDATE: Tambah 17 Pasien, Total Ada 134 Kasus Positif Virus Korona. [5]ine Tersedia pada: KOMPAS.COM <<https://gaya.tempo.co/read/877228/9-juta-orang-di-indonesia-mengalami-depresi/full&view=ok>>. Diakses pada tanggal 17 Maret 2020
- 6 Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- 2 Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y. and Xing, X., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*.

Ng, M.Y., Lee, E.Y., Yang, J., Yang, F., Li, X., Wang, H., Lui, M.M.S., Lo, C.S.Y., Leung, 1, Khong, P.L. and Hui, C.K.M., 2020. Imaging profile of the COVID-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*, 2(1), p.e200034.

RRI, 2020. Sepekan, RS Unair Telah Uji Tes Corona 500 Pasien. [online] Tersedia pada: http://rri.co.id/post/berita/802394/daerah/sepekan_rs_unair_telah_uji_tes_corona_500_pasien.html. Diakses pada tanggal 17 Maret 2020.

1 Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y. and Zhao, Y., 2020. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *Jama*.

World 8 ealth Organization, 2020. *Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance, 2 March 2020* (No. WHO/COVID-19/laboratory/2020.4). World Health Organization.

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

14 Zhou, S., Wang, Y., Zhu, T. and Xia, L., 2020. CT Features of Coronavirus Disease 2019 (COVID-19) Pneumonia in 62 Patients in Wuhan, China. *American Journal of Roentgenology*, pp.1-8.

1 Zu, Z.Y., Jiang, M.D., Xu, P.P., Chen, W., Ni, Q.Q., Lu, G.M. and Zhang, L.J., 2020.

ORIGINALITY REPORT

11%

SIMILARITY INDEX

11%

INTERNET SOURCES

5%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1

pubs.rsna.org

Internet Source

2%

2

www.biorxiv.org

Internet Source

1%

3

pt.scribd.com

Internet Source

1%

4

text-id.123dok.com

Internet Source

1%

5

eprints.umm.ac.id

Internet Source

1%

6

www.groundai.com

Internet Source

1%

7

www.coursehero.com

Internet Source

1%

8

Giuseppe Lippi, Ana-Maria Simundic, Mario Plebani. "Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19)", Clinical

1%

Chemistry and Laboratory Medicine (CCLM), 2020

Publication

9	e-journal.uajy.ac.id Internet Source	<1 %
10	nasional.kompas.com Internet Source	<1 %
11	akbid-ku.blogspot.com Internet Source	<1 %
12	id.scribd.com Internet Source	<1 %
13	sentrin.filkom.ub.ac.id Internet Source	<1 %
14	drtonywuorlhns.pixnet.net Internet Source	<1 %
15	Bankat M. Patil. "Impact of K-Means on the Performance of Classifiers for Labeled Data", Communications in Computer and Information Science, 2010 Publication	<1 %
16	complexity.cecs.ucf.edu Internet Source	<1 %
17	www.koalahero.com Internet Source	<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off