

SKRIPSI

MESIN PEMBELAJAR
BERBASIS *CONVOLUTIONAL NEURAL NETWORK*
UNTUK MENDETEKSI DINI KANKER SERVIKS



JOANNA AYULIA

NPM: 6161901101

PROGRAM STUDI MATEMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2023

FINAL PROJECT

**MACHINE LEARNING
BASED ON CONVOLUTIONAL NEURAL NETWORK
FOR EARLY DETECTION OF CERVICAL CANCER**



JOANNA AYULIA

NPM: 6161901101

**DEPARTMENT OF MATHEMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2023**

LEMBAR PENGESAHAN

**MESIN PEMBELAJAR
BERBASIS *CONVOLUTIONAL NEURAL NETWORK*
UNTUK MENDETEKSI DINI KANKER SERVIKS**

Joanna Ayulia

NPM: 6161901101

Bandung, 5 Agustus 2023

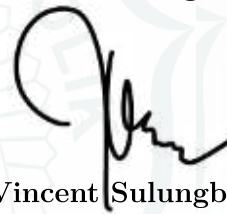
Menyetujui,

Pembimbing 1



Prof. M. Wono Setya Budhi

Pembimbing 2



Janto Vincent Sulungbudi, Drs.

Ketua Penguji



Dr. Livia Owen

Anggota Penguji



Liem Chin, M.Si.

Mengetahui,

Ketua Program Studi



Dr. Livia Owen

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

**MESIN PEMBELAJAR
BERBASIS *CONVOLUTIONAL NEURAL NETWORK*
UNTUK MENDETEKSI DINI KANKER SERVIKS**

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
5 Agustus 2023



A handwritten signature in black ink, appearing to read 'Joanna', is written over the stamp and QR code.

Joanna Ayulia
NPM: 6161901101

ABSTRAK

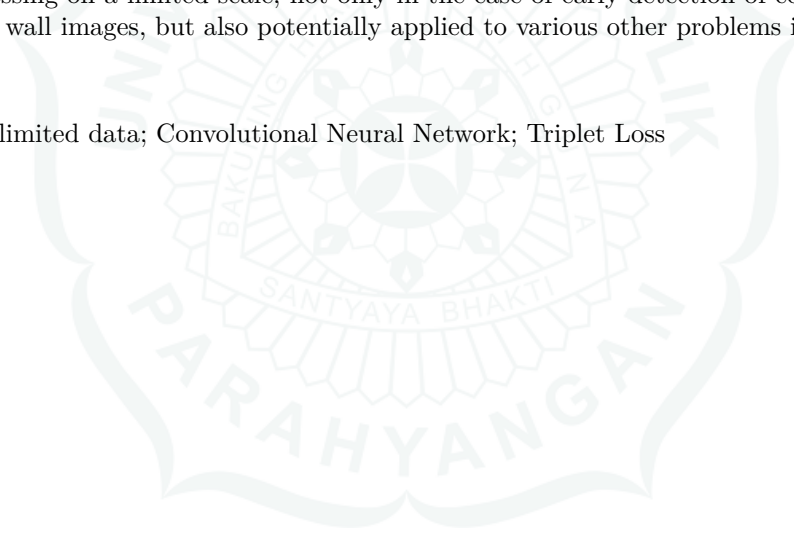
Penggunaan *Convolutional Neural Network* (CNN) telah membawa dampak positif yang signifikan dalam berbagai bidang, termasuk pengolahan citra medis. Penelitian sebelumnya telah menunjukkan kinerja yang baik dari model CNN dalam pengolahan data berukuran besar. Namun, tantangan baru muncul ketika dihadapkan pada keterbatasan data. Penelitian ini bertujuan untuk mengatasi hambatan tersebut dengan membangun sebuah model yang memungkinkan kinerja CNN yang optimal dengan data yang terbatas. *Pre-trained* model akan digunakan untuk melatih model CNN menggunakan data citra dinding rahim sebelum dan sesudah diusapkan larutan asam asetat. Kemudian *Triplet Loss* diimplementasikan untuk meningkatkan kemampuan model dalam mengenali fitur dan pola dari data citra dinding rahim yang terbatas dengan tingkat akurasi yang tinggi. Hasil penelitian menunjukkan bahwa model yang dikembangkan mampu mengatasi keterbatasan data dan menghasilkan prediksi yang akurat. Diharapkan hasil penelitian ini dapat memberikan solusi bagi tantangan pengolahan data dalam skala yang terbatas, tidak hanya pada kasus deteksi dini kanker serviks menggunakan citra dinding rahim, tetapi juga berpotensi diterapkan pada berbagai masalah lainnya di masa yang akan datang.

Kata-kata kunci: data terbatas; *Convolutional Neural Network*; *Triplet Loss*

ABSTRACT

The use of Convolutional Neural Network (CNN) has brought significant positive impacts in various fields, including medical image processing. Previous research has shown the good performance of CNN models in processing large data. However, new challenges arise when faced with data limitations. This research aims to overcome these obstacles by building a model that enables optimal CNN performance with limited data. Pre-trained model will be used to train the CNN model using image data of the uterine wall before and after applying acetic acid solution. Then Triplet Loss is implemented to improve the model's ability to recognize features and patterns from limited uterine wall image data with a high level of accuracy. The results show that the developed model is able to overcome data limitations and produce accurate predictions. It is expected that the results of this study can provide solutions to the challenges of data processing on a limited scale, not only in the case of early detection of cervical cancer using uterine wall images, but also potentially applied to various other problems in the future.

Keywords: limited data; Convolutional Neural Network; Triplet Loss



KATA PENGANTAR

Terpujilah Tuhan atas kasih dan pertolongan-Nya yang tak terbatas. Dengan berkat-Nya, saya berhasil memulai dan menyelesaikan skripsi ini. Skripsi berjudul “Mesin Pembelajaran Berbasis *Convolutional Neural Network* untuk Mendeteksi Dini Kanker Serviks” telah disusun selama dua semester dan selesai diujikan pada tanggal 14 Juli 2023. Skripsi ini merupakan salah satu syarat untuk menyelesaikan studi Strata-1 di Program Studi Matematika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan.

Tidak dapat disangkal bahwa penyelesaian skripsi ini memerlukan usaha dan kerja keras. Namun, skripsi ini tidak akan bisa selesai tanpa kehadiran orang-orang di sekeliling saya yang memberikan dukungan dan bantuan. Oleh karena itu, pada kesempatan ini, saya ingin mengucapkan terima kasih.

- Kepada Papi, Markus Benyamin, dan Mami, Intan Erawati. Terimakasih atas kasih, dukungan dan doa dalam setiap langkah perjalanan skripsi ini.
- Kepada Koko, Joshua Satriana. Terimakasih juga atas kasih, dukungan dan doanya.
- Kepada Bapak Prof. M. Wono Setya Budhi dan Bapak Janto Vincent Sulungbudi, Drs. yang telah membimbing saya. Terima kasih telah memberikan arahan dan wawasan berharga selama proses penyusunan skripsi.
- Kepada civitas akademika Program Studi Matematika UNPAR. Terima kasih atas ilmu dan pengalaman berharga yang telah saya peroleh selama kuliah di sini.
- Kepada Femilia dan Angela, teman-teman dekat selama masa perkuliahan. Terima kasih telah memberikan warna dan keceriaan dalam perjalanan kuliah.
- Kepada Amira, Enrico, Syawqi, Joycelynn, Daniel, teman-teman “Bootcamp” dan semua teman-teman lain yang tak dapat disebut satu per satu. Terima kasih telah hadir dan memberikan momen-momen berharga yang tak terlupakan.
- Kepada diri saya sendiri, terima kasih atas kegigihan, kesabaran, dan upaya tanpa henti dalam menyelesaikan skripsi ini. Perjuangan ini menjadi bukti bahwa ketekunan membawa hasil yang luar biasa.

Saya menyadari bahwa skripsi ini bukanlah akhir dari perjuangan, melainkan awal dari kesuksesan yang lebih besar. Saya juga menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, saya menerima setiap kritik dan saran untuk perbaikan dan pengembangan lebih lanjut. Semoga skripsi ini dapat menjadi langkah awal yang baik dalam mencapai tujuan-tujuan masa depan. Selamat membaca!

Bandung, 5 Agustus 2023

Penulis

DAFTAR ISI

KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	2
1.4 <i>State of the Art</i>	2
2 LANDASAN TEORI	4
2.1 Pembelajaran Mendalam	4
2.2 Jaringan Saraf Tiruan	5
2.2.1 Lapisan dan Neuron	5
2.2.2 Bobot, dan Bias	6
2.2.3 Fungsi Aktivasi	7
2.2.4 Fungsi <i>Loss</i>	9
2.2.5 Metode Optimasi	10
2.2.6 <i>Backpropagation</i>	11
2.2.7 <i>Overfitting</i> dan <i>Underfitting</i>	14
3 Convolutional Neural Network	16
3.1 Arsitektur Model	16
3.1.1 Lapisan Konvolusi	16
3.1.2 Lapisan <i>Pooling</i>	18
3.1.3 Lapisan <i>Flatten</i>	19
3.2 <i>Transfer Learning</i>	19
3.3 <i>Triplet Loss</i>	21
3.4 <i>Siamese Neural Network</i>	23
4 ANALISIS PREDIKSI	24
4.1 Data Pelatihan	24
4.2 Eksplorasi Penggunaan <i>Triplet Loss</i>	25
4.2.1 Pelatihan Model	26
4.2.2 Evaluasi Model	28
4.3 Hasil Pelatihan Model	29
4.3.1 Implementasi Klasifikasi Citra Dinding Rahim dengan <i>Triplet Loss</i>	30
4.3.2 Implementasi Klasifikasi Citra Dinding Rahim tanpa <i>Triplet Loss</i>	31
4.3.3 Perbandingan Hasil Model Klasifikasi	32

4.4	Sensitivitas dan Spesifisitas Model	33
5	KESIMPULAN DAN SARAN	35
5.1	Kesimpulan	35
5.2	Saran	35
	DAFTAR REFERENSI	37

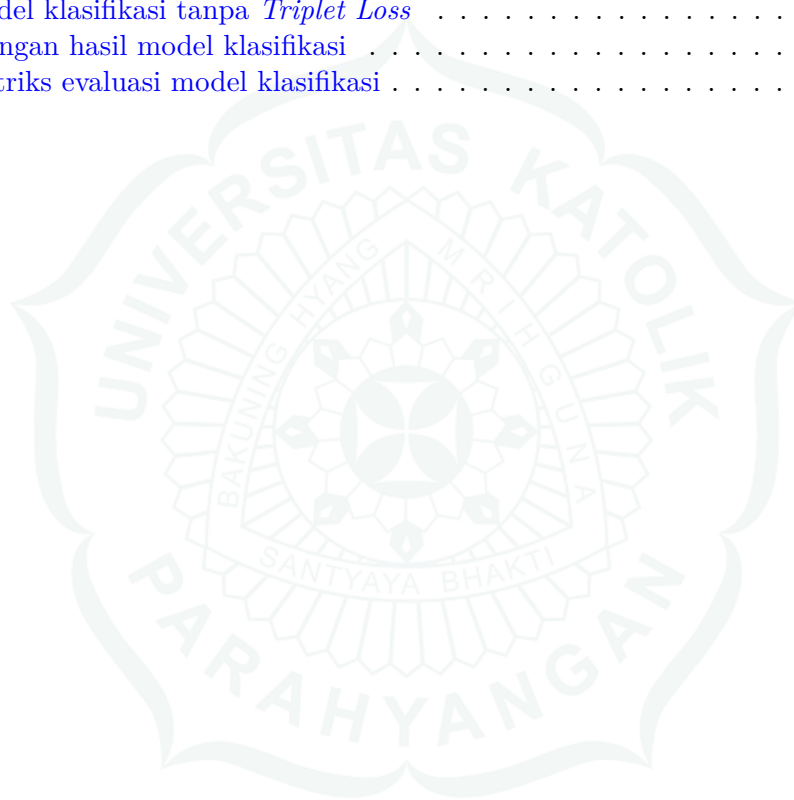


DAFTAR GAMBAR

2.1	Diagram Venn: hubungan antara kecerdasan buatan, pembelajaran mesin, dan pembelajaran mendalam	4
2.2	Jaringan Saraf Tiruan	5
2.3	Jaringan Saraf Tiruan dengan bobot dan bias	6
2.4	Ilustrasi fungsi aktivasi	7
2.5	Fungsi aktivasi ReLU	8
2.6	Fungsi aktivasi Sigmoid	9
2.7	Ilustrasi <i>backpropagation</i>	12
2.8	Plot model <i>overfitting</i>	14
2.9	Plot model <i>underfitting</i>	14
2.10	Plot model “ <i>good fit</i> ”	14
3.1	Ilustrasi proses konvolusi	16
3.2	Contoh masukan citra	17
3.3	Ilustrasi <i>max pooling</i> dan <i>average pooling</i> dengan submatriks berdimensi 2×2	18
3.4	Ilustrasi proses <i>flattening</i>	19
3.5	Arsitektur <i>pre-trained</i> model ResNet-50	20
3.6	Arsitektur blok konvolusi dan blok identitas pada <i>pre-trained</i> model ResNet-50	21
3.7	Skema proses pembelajaran dari <i>Triplet Loss</i>	22
3.8	<i>Siamese Neural Network</i>	23
4.1	Data dengan kelas positif, negatif dan mencurigakan	24
4.2	Ilustrasi rancangan model dengan 6 citra masukan dan <i>pre-trained</i> model ResNet-50	25
4.3	Contoh masukan citra	26
4.4	Contoh matriks <i>confusion</i>	28
4.5	Grafik <i>loss</i> terhadap <i>epochs</i> pada pelatihan data dengan <i>Triplet Loss</i>	30
4.6	Hasil model klasifikasi menggunakan <i>Triplet Loss</i> dengan 500 <i>epochs</i>	31
4.7	Hasil model klasifikasi tanpa menggunakan <i>Triplet Loss</i> dengan 500 <i>epochs</i>	32
4.8	Perbandingan hasil model klasifikasi	33
4.9	Matriks <i>confusion</i> model klasifikasi dengan <i>Triplet Loss</i>	33
4.10	Matriks <i>confusion</i> model klasifikasi tanpa <i>Triplet Loss</i>	33

DAFTAR TABEL

3.1 Hasil konvolusi dengan berbagai <i>kernel</i>	17
4.1 Contoh penambahan data	26
4.1 Contoh penambahan data	27
4.2 Hasil model klasifikasi dengan <i>Triplet Loss</i>	30
4.3 Hasil model klasifikasi tanpa <i>Triplet Loss</i>	31
4.4 Perbandingan hasil model klasifikasi	32
4.5 Hasil metriks evaluasi model klasifikasi	34



BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kemajuan yang pesat dalam pengembangan pembelajaran mesin, khususnya dengan penggunaan CNN, telah memberikan dampak besar dalam berbagai bidang, termasuk pengolahan citra medis. CNN merupakan model jaringan saraf atau *class* yang sangat efektif dalam mengenali fitur dari citra dan melakukan klasifikasi [1].

Pada tahun 2020, Yuan, C., dkk. melakukan penelitian yang berhasil mengenali abnormalitas pada citra dinding rahim dengan menggunakan CNN [2]. Penelitian tersebut menggunakan 22.330 data citra dinding rahim yang telah diusapkan larutan asam asetat, larutan yodium, dan juga informasi klinis pasien. Hasil dari penelitian tersebut menunjukkan bahwa CNN mencapai tingkat akurasi sebesar 84,10%. Penggunaan data dengan jumlah yang besar dalam penelitian tersebut memberikan beragam variasi data yang dapat meningkatkan kemampuan model CNN untuk melakukan generalisasi dan mendeteksi fitur-fitur pada citra dinding rahim. Data dengan jumlah yang besar memungkinkan model untuk memperoleh representasi fitur yang lebih baik dan meningkatkan akurasi model.

Mengacu pada penelitian tersebut, penelitian ini bertujuan untuk membangun sebuah model dengan data terbatas. Meskipun CNN telah terbukti efektif dalam penelitian sebelumnya, penelitian ini menghadapi tantangan baru, yaitu keterbatasan data. Keterbatasan data dapat menyebabkan masalah *overfitting*, di mana model tidak dapat digeneralisasi dengan baik dan menghasilkan kinerja yang buruk pada data baru.

1.2 Rumusan Masalah

Berikut masalah-masalah yang akan dikaji pada penelitian ini.

1. Bagaimana implementasi CNN untuk mengatasi keterbatasan data, khususnya pada data citra dinding rahim?
2. Bagaimana mengatasi masalah *overfitting* yang disebabkan oleh keterbatasan jumlah data pada model CNN?
3. Bagaimana tingkat akurasi dalam penerapan *pre-trained* model ResNet-50 pada data citra dinding rahim yang terbatas?

1.3 Tujuan

Tujuan dari penyusunan penelitian ini adalah

1. mengimplementasikan CNN untuk mengatasi keterbatasan data pada data citra dinding rahim;
2. mengatasi masalah *overfitting* yang disebabkan oleh keterbatasan jumlah data dengan menggunakan tahap pembelajaran *Triplet Loss*;
3. mengukur tingkat akurasi dari penerapan *pre-trained* model ResNet-50 pada data citra dinding rahim yang terbatas.

1.4 *State of the Art*

Penelitian yang dilakukan oleh Yuan, C., dkk. [2] telah berhasil memanfaatkan CNN dalam mengenali abnormalitas pada citra dinding rahim untuk mendeteksi dini kanker serviks dengan tingkat akurasi yang tinggi. Namun, penelitian tersebut berfokus pada penggunaan data citra dinding rahim dengan jumlah data yang besar dan belum mempertimbangkan keterbatasan data dalam penggunaan model. Selanjutnya, penelitian oleh Bircanoğlu, Cenk [3], bertujuan untuk membandingkan kinerja berbagai fungsi *loss*. Hasil penelitian tersebut menunjukkan bahwa *Triplet Loss* menjadi salah satu fungsi *loss* yang paling efektif dalam tugas-tugas tertentu.

Dengan mengkombinasikan kedua penelitian tersebut, penelitian ini akan memanfaatkan model CNN dari penelitian oleh Yuan, C., dkk. dan mengimplementasikan *Triplet Loss* yang terbukti efektif dari penelitian oleh Bircanoğlu, Cenk, untuk meningkatkan kinerja model dalam mengenali abnormalitas pada citra dinding rahim sebelum dan sesudah diusapkan larutan asam asetat. Penggunaan *Triplet Loss* dan penambahan data dengan augmentasi diharapkan dapat membantu mengatasi masalah *overfitting* yang disebabkan oleh keterbatasan data dan meningkatkan akurasi dalam mendeteksi dini kanker serviks.

Penelitian ini memiliki perbedaan signifikan dalam jumlah masukan data citra yang digunakan. Pada umumnya, *Triplet Loss* menggunakan tiga kelompok masukan data, tetapi dalam penelitian ini, enam kelompok masukan data citra digunakan untuk membandingkan citra dinding rahim sebelum dan sesudah diusapkan larutan asam asetat. Ini memungkinkan model untuk mempelajari perbedaan dan fitur yang lebih spesifik antara kedua jenis citra tersebut.

Sejauh pengetahuan penulis, belum ada penelitian sebelumnya yang secara khusus memanfaatkan *Triplet Loss* dalam meningkatkan kinerja model dengan data yang terbatas untuk mendeteksi dini kanker serviks pada citra dinding rahim. Perlu diperhatikan bahwa deteksi citra dinding rahim ini tidak semudah mendeteksi citra yang sudah dikenal, seperti deteksi objek pada gambar pemandangan, deteksi wajah pada foto orang, klasifikasi gambar hewan, dan lainnya. Citra dinding rahim memiliki kompleksitas tinggi yang tidak dapat diprediksi dengan jelas, sehingga diperlukan model yang kompleks untuk dapat mempelajari struktur dan fitur citra yang khusus. Selain itu, data citra dinding rahim yang digunakan dalam penelitian ini sangat terbatas, yang menambah kompleksitas dalam proses pembelajaran model.

Oleh karena itu, penelitian ini memiliki kontribusi dalam mengoptimalkan kinerja model dengan memanfaatkan *Triplet Loss* dan menambahkan data dengan augmentasi pada data citra dinding rahim yang terbatas. Diharapkan model yang dibangun pada penelitian ini tidak hanya dapat memberikan solusi untuk mendeteksi dini kanker serviks pada citra dinding rahim dengan akurasi yang lebih tinggi, tetapi juga dapat menjadi acuan untuk mengatasi keterbatasan data pada kasus lainnya.

