



**PENERAPAN EKSTRAKSI FITUR HARALICK KLASIFIKASI CITRA
COMPUTERIZED TOMOGRAPHY MENGGUNAKAN XGBOOST
(STUDI KASUS: CITRA *COMPUTERIZED TOMOGRAPHY* COVID-19)**

SKRIPSI

untuk memenuhi persyaratan

menyelesaikan program S1 PS Ilmu Komputer di FMIPA ULM

Oleh

MUHAMMAD ANDRE FIRDAUS

NIM 1611016210016

**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2023



**PENERAPAN EKSTRAKSI FITUR HARALICK KLASIFIKASI CITRA
COMPUTERIZED TOMOGRAPHY MENGGUNAKAN XGBOOST
(STUDI KASUS: CITRA *COMPUTERIZED TOMOGRAPHY* COVID-19)**

SKRIPSI

untuk memenuhi persyaratan

menyelesaikan program S1 PS Ilmu Komputer di FMIPA ULM

Oleh

MUHAMMAD ANDRE FIRDAUS

NIM 1611016210016

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2023**

SKRIPSI

PENERAPAN EKSTRAKSI FITUR HARALICK KLASIFIKASI CITRA COMPUTERIZED TOMOGRAPHY MENGGUNAKAN XGBOOST (STUDI KASUS: CITRA COMPUTERIZED TOMOGRAPHY COVID-19)

Oleh

MUHAMMAD ANDRE FIRDAUS
NIM 1611016210016

Telah dipertahankan didepan Dosen Penguji pada tanggal Juni 2023
Susunan Dosen Penguji :

Pembimbing I



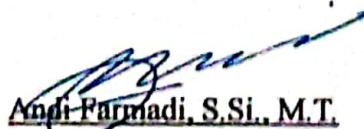
Muliadi, S.Kom., M.Cs.
NIP. 197804222010121002

Dosen Penguji I



Radityo Adi Nugroho, S.T., M.Kom.
NIP. 198212042008011006

Pembimbing II



Andi Firmadi, S.Si., M.T.
NIP. 197307252008011006

Dosen Penguji II



Friska Abdi, S.Kom., M.Kom.
NIP. 19880913201612110001



Banjarnegara, Juni 2023
Koordinator PS. Ilmu Komputer,

Andi Firmadi, S.T., M.Kom.
NIP. 197307252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, Juni 2023



Muhammad Andre Firdaus

NIM. 1611016210016

ABSTRAK

**PENERAPAN EKSTRAKSI FITUR HARALICK KLASIFIKASI CITRA
COMPUTERIZED TOMOGRAPHY MENGGUNAKAN XGBOOST (STUDI
KASUS: CITRA *COMPUTERIZED TOMOGRAPHY* COVID-19)** (Oleh:
Muhammad Andre Firdaus; Pembimbing: Muliadi, S.Kom., M.Cs. dan Andi Farmadi
S.Si., M.T.; 2023; 53 halaman)

COVID-19 merupakan penyakit jenis baru yang belum pernah diidentifikasi sebelumnya pada manusia. Virus penyebab COVID-19 ini dinamakan Sars-CoV-2. Virus corona adalah zoonosis (ditularkan antara hewan dan manusia). Adapun, hewan yang menjadi sumber penularan COVID-19 ini masih belum diketahui. Berdasarkan bukti ilmiah, COVID-19 dapat menular dari manusia ke manusia melalui percikan batuk/bersin (droplet), Orang yang paling berisiko tertular penyakit ini adalah orang yang kontak erat dengan pasien COVID-19 termasuk yang merawat pasien COVID-19. Pada saat kasus COVID-19 metode yang paling banyak digunakan untuk identifikasi COVID-19 adalah metode *Real-Time Reverse Transcription-Polymerase Chain Reaction* (RT-PCR). Tetapi, identifikasi menggunakan metode RT-PCR memakan waktu yang cukup lama. Alternatif yang bisa digunakan untuk identifikasi adalah dengan citra *Computerized Tomography* (CT) dan citra *X-Ray*. Pada penelitian ini dibuat suatu model untuk mengklasifikasi citra CT Scan COVID-19 menggunakan ekstraksi fitur Haralick dan XGBoost. Citra CT Scan yang digunakan berisi 397 citra CT pasien negatif COVID-19 dan 349 citra CT pasien positif COVID-19 dari citra CT 216 pasien. Ekstraksi fitur Haralick dengan klasifikasi XGBoost mampu memperoleh AUC yang paling tinggi yaitu 0.915 pada seleksi fitur $top\ k = 24$ dari fitur Haralick.

Kata Kunci: *Haralick*, *XGBoost*, Seleksi Fitur, Citra CT Scan COVID-19.

ABSTRACT

IMPLEMENTATION OF HARALICK FEATURE EXTRACTION FOR CLASSIFICATION OF COMPUTERIZED TOMOGRAPHY IMAGES USING XGBOOST (CASE STUDY: COVID-19 COMPUTERIZED TOMOGRAPHY IMAGES) (By: Muhammad Andre Firdaus; Preceptor: Muliadi, S.Kom., M.Cs. dan Andi Farmadi S.Si., M.T.; 2023; 53 Pages)

COVID-19 is a new type of disease that has not been previously identified in humans. The virus causing COVID-19 is called SARS-CoV-2. Coronaviruses are zoonotic, meaning they are transmitted between animals and humans. The exact animal source of COVID-19 is still unknown. Based on scientific evidence, COVID-19 can spread from human to human through respiratory droplets when an infected person coughs or sneezes. The people at highest risk of contracting the disease are those who have close contact with COVID-19 patients, including those who care for them. During the COVID-19 outbreak, the most commonly used method for identification is the Real-Time Reverse Transcription-Polymerase Chain Reaction (RT-PCR) method. However, RT-PCR identification can be time-consuming. An alternative approach for identification is using Computerized Tomography (CT) and X-Ray images. In this study, a model was developed to classify COVID-19 CT scan images using Haralick feature extraction and XGBoost. The CT scan images used consisted of 397 CT images from COVID-19 negative patients and 349 CT images from COVID-19 positive patients, totaling 746 patients. Haralick feature extraction combined with XGBoost classification achieved the highest which is AUC = 0.915 from top k = 24 feature selection from Haralick feature.

Keywords: *Haralick, XGBoost, Feature Selection, COVID-19 CT Scan Image.*

PRAKATA

Puji dan syukur kepada Allah SWT karena atas berkat rahmat-Nya penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Ekstraksi Fitur Haralick Klasifikasi Citra *Computerized Tomography* Menggunakan XGBoost (Studi Kasus: Citra *Computerized Tomography* COVID-19)” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Tidak lupa penulis menyampaikan ucapan terimakasih kepada pihak-pihak terkait yang sangat mendukung dan membantu penulis dalam penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga yang senantiasa memberikan doa, semangat, dukungan, hingga kepercayaan yang membuat penulis selalu bekerja keras menyelesaikan skripsi.
2. Bapak Muliadi, S.Kom., M.Cs. selaku dosen pembimbing utama yang senantiasa membimbing, membantu, dan meluangkan waktu dalam proses penyelesaian skripsi ini.
3. Bapak Andi Farmadi, S.Si., M.T., selaku dosen pembimbing pendamping yang turut serta memberi arahan, membantu dan meluangkan waktu dalam proses penyelesaian skripsi ini.
4. Bapak Irwan Budiman, S.T., M.Kom. selaku ketua program studi Ilmu Komputer beserta seluruh dosen dan karyawan/staff pegawai Fakultas MIPA Universitas Lambung Mangkurat atas bantuan yang diberikan selama penulis mengikuti studi.
5. Keluarga besar Ilmu Komputer angkatan 2016, terimakasih sudah membantu memecahkan kebingungan-kebingungan saat masa pengerjaan skripsi ini dan mendukung dalam berbagai hal.
6. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam menyelesaikan skripsi ini.

Akhir kata penulis menyadari bahwa skripsi ini jauh dari sempurna, kritik dan saran yang membangun sangat diperlukan. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

DAFTAR ISI

| | |
|---|-----|
| Halaman judul | i |
| PERNYATAAN | iii |
| ABSTRAK | iv |
| ABSTRACT | v |
| PRAKATA | vi |
| DAFTAR ISI | vii |
| DAFTAR GAMBAR | ix |
| DAFTAR TABEL | x |
| DAFTAR LAMPIRAN | xi |
| BAB I | 1 |
| PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 2 |
| 1.3 Tujuan Penelitian | 2 |
| 1.4 Manfaat Penelitian | 2 |
| 1.5 Batasan Masalah | 3 |
| 1.6 Sistematika Penulisan | 3 |
| BAB II | 5 |
| TINJAUAN PUSTAKA | 5 |
| 2.1 Kajian Terdahulu | 5 |
| 2.2 Landasan Teori | 10 |
| 2.2.1 Dataset COVID-CT..... | 10 |
| 2.2.2 Ekstraksi Fitur Haralick | 11 |
| 2.2.3 <i>Stratified K Fold Cross Validation</i> | 13 |
| 2.2.4 Klasifikasi XGBoost | 14 |
| 2.2.5 Feature Selection | 15 |
| 2.2.5 <i>Confusion Matrix</i> | 16 |
| 2.2.6 Kurva ROC dan AUC | 17 |
| BAB III | 18 |
| METODE PENELITIAN | 18 |

| | | |
|-----------------------------|---|-----------|
| 3.1 | Alat penelitian | 18 |
| 3.2 | Bahan Penelitian..... | 18 |
| 3.3 | Prosedur Penelitian..... | 19 |
| BAB IV | | 21 |
| HASIL DAN PEMBAHASAN | | 21 |
| 4.1 | Hasil | 21 |
| 4.1.1 | Pengumpulan Data | 21 |
| 4.1.2 | Pembuatan Model..... | 22 |
| a. | Mengubah Citra Digital menjadi Matriks yang berisi Nilai Piksel | 22 |
| b. | Menghitung Nilai Fitur Haralick..... | 22 |
| c. | Hasil Ekstraksi Fitur Haralick | 26 |
| d. | <i>Framing Data</i> | 27 |
| d. | <i>Feature Importance</i> | 30 |
| e. | <i>Feature Selection</i> | 32 |
| f. | <i>XGBoost Classifier</i> | 32 |
| 4.1.3 | Evaluasi | 35 |
| 4.2 | Pembahasan..... | 50 |
| BAB V | | 53 |
| PENUTUP | | 53 |
| 5.1 | Kesimpulan | 53 |
| 5.2 | Saran | 53 |
| DAFTAR PUSTAKA | | 54 |
| LAMPIRAN | | 57 |

DAFTAR GAMBAR

| Gambar | Halaman |
|---|----------------|
| Gambar 1. Skema 10-fold cross validation | 14 |
| Gambar 2. Alur penelitian..... | 19 |
| Gambar 3. Pengecekan Fitur Gabungan digabung dengan benar | 28 |
| Gambar 4. Pengecekan Nilai Kosong Fitur Gabungan | 29 |
| Gambar 5. Pengecekan Label Encoder | 30 |
| Gambar 6. Feature Importance..... | 31 |
| Gambar 7. Pohon ke-1..... | 33 |
| Gambar 8. Rule Pohon ke-1 | 33 |
| Gambar 9. Pohon ke – 2..... | 34 |
| Gambar 10. Rule Pohon ke – 2 | 34 |
| Gambar 11. Pohon ke – 100..... | 35 |
| Gambar 12. Rule Pohon ke - 100..... | 35 |
| Gambar 13. Grafik ROC dan AUC top k = 1..... | 36 |
| Gambar 14. Grafik ROC dan AUC top k = 2..... | 37 |
| Gambar 15. Grafik ROC dan AUC top k = 3..... | 37 |
| Gambar 16. Grafik ROC dan AUC top k = 4..... | 38 |
| Gambar 17. Grafik ROC dan AUC top k = 5..... | 38 |
| Gambar 18. Grafik ROC dan AUC top k = 6..... | 39 |
| Gambar 19. Grafik ROC dan AUC top k = 7..... | 39 |
| Gambar 20. Grafik ROC dan AUC top k = 8..... | 40 |
| Gambar 21. Grafik ROC dan AUC top k = 9..... | 40 |
| Gambar 22. Grafik ROC dan AUC top k = 10..... | 41 |
| Gambar 23. Grafik ROC dan AUC top k = 11..... | 41 |
| Gambar 24. Grafik ROC dan AUC top k = 12..... | 42 |
| Gambar 25. Grafik ROC dan AUC top k = 13..... | 42 |
| Gambar 26. Grafik ROC dan AUC top k = 14..... | 43 |
| Gambar 27. Grafik ROC dan AUC top k = 15..... | 43 |
| Gambar 28. Grafik ROC dan AUC top k = 16..... | 44 |
| Gambar 29. Grafik ROC dan AUC top k = 17..... | 44 |
| Gambar 30. Grafik ROC dan AUC top k = 18..... | 45 |
| Gambar 31. Grafik ROC dan AUC top k = 19..... | 45 |
| Gambar 32. Grafik ROC dan AUC top k = 20..... | 46 |
| Gambar 33. Grafik ROC dan AUC top k = 21..... | 46 |
| Gambar 34. Grafik ROC dan AUC top k = 22..... | 47 |
| Gambar 35. Grafik ROC dan AUC top k = 23..... | 47 |
| Gambar 36. Grafik ROC dan AUC top k = 24..... | 48 |
| Gambar 37. Grafik ROC dan AUC top k = 25..... | 48 |
| Gambar 38. Grafik ROC dan AUC top k = 26..... | 49 |
| Gambar 39. Grafik Nilai AUC berdasarkan banyak fitur yang dipilih | 51 |

DAFTAR TABEL

| Tabel | Halaman |
|--|----------------|
| Tabel 1. Keaslian Penelitian..... | 8 |
| Tabel 2. Fitur Haralick | 13 |
| Tabel 3. Confusion Matrix | 16 |
| Tabel 4. Data Citra CT-Scan yang digunakan | 21 |
| Tabel 5. Nilai Piksel dari Citra CT-Scan COVID-19 | 22 |
| Tabel 6. Nilai derajat keabuan suatu piksel | 23 |
| Tabel 7 Konversi Nilai Piksel menjadi Nilai Derajat Keabuan | 23 |
| Tabel 8 Matriks Kemunculan dengan derajat 0° | 24 |
| Tabel 9. Matriks Transpose dari Matriks Kemunculan..... | 24 |
| Tabel 10. Matriks Simetris..... | 24 |
| Tabel 11. Matriks Normalisasi..... | 25 |
| Tabel 12. Tabel Fitur Mean dan Range dari Haralick..... | 27 |
| Tabel 13. Feature Importance | 30 |
| Tabel 14. Hasil Pengujian Model Fitur Haralick dengan XGBoost Classifier | 49 |
| Tabel 15. Kriteria Nilai AUC..... | 52 |

DAFTAR LAMPIRAN

Lampiran 1. Source Code

Lampiran 2. Tabel Hasil Ekstraksi Fitur Haralick

Lampiran 3. Riwayat Hidup