

IDENTIFIKASI PENYAKIT PENYERTA KANKER PARU-PARU MENGUNAKAN ALGORITMA DETEKSI KOMUNITAS DAN TEKNIK OPTIMASI

ESKAWATI KURNIA DWIMARDYASTUTI



PROGRAM MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Identifikasi Penyakit Penyerta Kanker Paru-Paru Menggunakan Algoritma Deteksi Komunitas dan Teknik Optimasi” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2026

Eskawati Kurnia D M
G6501231034

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



RINGKASAN

ESKAWATI KURNIA D M. Identifikasi Penyakit Penyerta Kanker Paru-Paru Menggunakan Algoritma Deteksi Komunitas dan Teknik Optimasi. Dibimbing oleh WISNU ANANTA KUSUMA, MUSHTHOFA dan IRANDI PUTRA PRATOMO.

Kanker paru-paru memiliki dua subtype utama, yaitu *Non-Small Cell Lung Cancer* (NSCLC) dan *Small Cell Lung Cancer* (SCLC), dengan karakteristik biologis dan klinis yang berbeda. Keberadaan penyakit penyerta dapat meningkatkan kompleksitas kondisi pasien, sedangkan hubungan antarpenyakit dapat membentuk komunitas tumpang tindih. Penelitian ini bertujuan menganalisa performa DPSO-Fuzzy, pengaruh topologi jaringan, serta mengidentifikasi kandidat penyakit signifikan dan tumpang tindih pada jaringan NSCLC dan SCLC.

DPSO-Fuzzy menggabungkan *Discrete Particle Swarm Optimization* untuk membentuk komunitas awal dengan *fuzzy refinement* untuk menghasilkan distribusi *membership* setiap node. Metode ini dibandingkan dengan sembilan algoritma pada lima jaringan *benchmark*, yaitu Karate, Dolphins, Football, Yeast-D2, dan Y2H. Evaluasi dilakukan menggunakan ONMI, Omega Index, *community precision*, *community recall*, dan *community F1*, kemudian dirangkum melalui *Principal Component Analysis*. Karakteristik topologi jaringan dianalisa menggunakan *degree skewness* dan Gini coefficient.

Jaringan penyakit dibangun berdasarkan semantik similaritas Disease Ontology dari hasil *text mining* artikel PubMed. *Threshold* similaritas 0,6 menghasilkan jaringan NSCLC dengan 410 node dan 1.053 edge serta jaringan SCLC dengan 456 node dan 1.246 edge. Analisa komunitas dilakukan pada komponen terbesar masing-masing jaringan.

Hasil evaluasi menunjukkan bahwa tidak ada algoritma yang selalu unggul pada seluruh jaringan. DPSO-Fuzzy memperoleh performa agregat tertinggi pada Yeast-D2 dan Y2H serta menunjukkan hasil kompetitif pada Dolphins dan Football. Berdasarkan *decision score*, DPSO-Fuzzy, Fuzzy LPA, dan PSO-LPA menjadi tiga kandidat utama pada jaringan NSCLC dan SCLC. Fuzzy LPA menghasilkan *fuzzy modularity* dan *stability* tertinggi, sedangkan PSO-LPA menghasilkan jumlah *overlapping nodes* terbanyak. DPSO-Fuzzy lebih selektif dengan rata-rata 8,70 *overlapping nodes* pada NSCLC dan 10 pada SCLC serta memperoleh *purpose score* tertinggi pada kedua jaringan. Analisa sensitivitas menghasilkan empat *stable overlapping nodes* pada NSCLC dan enam pada SCLC. *Pulmonary embolism* pada kedua jaringan dan *cerebellar ataxia* pada SCLC memiliki dukungan klinis paling kuat. Temuan lainnya lebih banyak mencerminkan hubungan biologis, anatomis, hierarkis, atau ontologis. Hasil penelitian ini merupakan kandidat hubungan penyakit yang masih memerlukan validasi menggunakan data klinis, *multi-omics*, atau rekam medis elektronik.

Kata kunci: deteksi komunitas tumpang tindih, dpso-fuzzy, jaringan komorbiditas, kanker paru-paru, penyakit penyerta



SUMMARY

ESKAWATI KURNIA D M. Identification of Lung Cancer Comorbidities Using Community Detection Algorithms and Optimization Techniques. Supervised by WISNU ANANTA KUSUMA, MUSHTHOFA, and IRANDI PUTRA PRATOMO.

Lung cancer has two main subtypes, namely non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), with different biological and clinical characteristics. The presence of comorbidities can increase the complexity of patients' conditions, while relationships among diseases can form overlapping communities. This study aimed to analyze the performance of DPSO-Fuzzy, the influence of network topology, and identify significant and overlapping disease candidates in NSCLC and SCLC networks.

DPSO-Fuzzy combines Discrete Particle Swarm Optimization to form initial communities with fuzzy refinement to generate the membership distribution of each node. This method was compared with nine algorithms on five benchmark networks, namely Karate, Dolphins, Football, Yeast-D2, and Y2H. The evaluation was performed using ONMI, Omega Index, community precision, community recall, and community F1, which were then summarized using Principal Component Analysis. Network topological characteristics were analyzed using degree skewness and the Gini coefficient.

The disease networks were constructed based on Disease Ontology semantic similarity from the results of text mining PubMed articles. A similarity threshold of 0.6 produced an NSCLC network with 410 nodes and 1.053 edges and an SCLC network with 456 nodes and 1.246 edges. Community analysis was performed on the largest component of each network.

The evaluation results showed that no algorithm consistently outperformed the others across all networks. DPSO-Fuzzy achieved the highest aggregate performance on Yeast-D2 and Y2H and showed competitive results on Dolphins and Football. Based on the decision score, DPSO-Fuzzy, Fuzzy LPA, and PSO-LPA were the three main candidates for the NSCLC and SCLC networks. Fuzzy LPA produced the highest fuzzy modularity and stability, whereas PSO-LPA produced the largest number of overlapping nodes. DPSO-Fuzzy was more selective, with an average of 8,70 overlapping nodes in NSCLC and 10 in SCLC, and obtained the highest purpose score in both networks. The sensitivity analysis produced four stable overlapping nodes in NSCLC and six in SCLC. Pulmonary embolism in both networks and cerebellar ataxia in SCLC had the strongest clinical support. The other findings mostly reflected biological, anatomical, hierarchical, or ontological relationships. The results of this study represent candidate disease relationships that still require validation using clinical data, multi-omics data, or electronic medical records.

Keywords: DPSO-Fuzzy, comorbidities, comorbidity network, lung cancer, overlapping community detection



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

IDENTIFIKASI PENYAKIT PENYERTA KANKER PARU-PARU MENGUNAKAN ALGORITMA DETEKSI KOMUNITAS DAN TEKNIK OPTIMASI

ESKAWATI KURNIA DWIMARDYASTUTI

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Magister Ilmu Komputer

**PROGRAM MAGISTER ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Tesis:

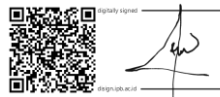
1. Rina Trisminingsih, S.Kom, M.T., Ph.D.



Judul Tesis : Penyerta Kanker Paru-Paru Menggunakan Algoritma Deteksi
Komunitas Dan Teknik Optimasi
Nama : Eskawati Kurnia Dwimardiyastuti
NIM : G6501231034

Disetujui oleh

Pembimbing 1:
Prof. Dr. Eng Wisnu Ananta Kusuma, S.T., M.T.



Pembimbing 2:
Dr. Mushthofa, S.Kom., M.Sc.



Signed with  privy

Pembimbing 2:
dr. Irandi Putra Pratomo, Ph.D., Sp.P(K)



Diketahui oleh

Ketua Program Studi:
Dr. Toto Haryanto, S.Kom., M.Si.
NIP 198211172014041001

Dekan Sekolah Sains Data, Matematika dan Informatika:
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom
NIP. 196607021993021001

Tanggal Ujian:
(12 Agustus 2026)

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanahu wa ta'ala atas segala rahmat dan karunia-Nya sehingga karya ilmiah ini dapat diselesaikan. Penelitian yang dilaksanakan sejak Agustus 2024 hingga Agustus 2026 ini mengangkat bidang bioinformatika dengan judul “Identifikasi Penyakit Penyerta Kanker Paru-Paru Menggunakan Algoritma Deteksi Komunitas dan Teknik Optimasi”.

Penulis menyampaikan terima kasih kepada para pembimbing, Prof. Dr. Eng. Wisnu Ananta Kusuma, S.T., M.T., Dr. Mushthofa, S.Kom., M.Sc., dan dr. Irandi Putra Pratomo, Ph.D., Sp.P(K), yang telah memberikan bimbingan, arahan, ilmu, dan berbagai saran selama pelaksanaan penelitian dan penyusunan tesis ini. Ucapan terima kasih juga disampaikan kepada moderator seminar dan penguji luar komisi pembimbing atas masukan yang diberikan untuk penyempurnaan karya ilmiah ini.

Penulis menyampaikan terima kasih kepada Dr. Toto Haryanto, S.Kom., M.Si. yang telah memberikan izin menggunakan akun server G6 untuk mendukung proses komputasi dalam penelitian ini. Terima kasih juga disampaikan kepada Bapak Muhammad Asyhar Agmalaro, S.Si., M.Kom., Ibu Dr. Ana Ratnasari, M.Pd., dan Ibu Ajeng Eka Sagitta, S.E., M.E. atas dukungan, motivasi, dan surat rekomendasi yang diberikan sebagai salah satu persyaratan untuk melanjutkan pendidikan di Program Pascasarjana IPB.

Ungkapan terima kasih dan penghargaan yang tulus penulis sampaikan kepada suami tercinta M. Nur Rahmadi, A.Md., kedua orang tua Bapak Supratman, S.Pd. dan Ibu Sri Sudarwati, S.Pd., serta mertua Ibu Pudjiwati, atas doa restu, kasih sayang, kesabaran, pengertian, dan dukungan moral maupun material yang senantiasa diberikan selama penulis menempuh pendidikan dan menyelesaikan penelitian ini.

Terima kasih juga penulis sampaikan kepada teman-teman Pascasarjana Ilmu Komputer lintas angkatan tahun 2022–2024 yang tidak dapat disebutkan satu per satu atas kebersamaan, bantuan, semangat, dan dukungan selama proses pendidikan dan penyusunan tesis.

Semoga karya ilmiah ini dapat memberikan manfaat bagi pihak yang membutuhkan serta berkontribusi bagi perkembangan ilmu pengetahuan.

Bogor, Agustus 2026

Eskawati Kurnia Dwimardiyastuti



DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
II TINJAUAN PUSTAKA	4
2.1 Kanker Paru-Paru	4
2.2 Penyakit Penyerta	4
2.3 Teknik Optimasi Menggunakan DPSO	4
2.4 Algoritma Deteksi Komunitas	7
III METODE	11
3.1 Tahapan Penelitian	11
3.2 Kebutuhan Pengembangan	26
IV HASIL DAN PEMBAHASAN	27
4.1 Konstruksi Jaringan	27
4.2 Analisa Topologi Jaringan	29
4.3 Evaluasi Algoritma pada Data <i>Benchmark</i>	31
4.4 Performa algoritma pada data <i>benchmark</i>	35
4.5 Analisa topologi dan pemilihan algoritma	39
4.6 Hasil deteksi komunitas pada jaringan NSCLC dan SCLC	44
4.7 Identifikasi node signifikan pada jaringan NSCLC dan SCLC	45
4.8 Identifikasi node <i>overlapping</i> pada jaringan NSCLC dan SCLC	50
4.9 Interpretasi klinis komorbiditas signifikan	54
4.10 Interpretasi molekuler <i>overlapping nodes</i> yang stabil	55
V SIMPULAN DAN SARAN	60
5.1 Simpulan	60
5.2 Saran	60
DAFTAR PUSTAKA	62
LAMPIRAN	69
RIWAYAT HIDUP	81



DAFTAR TABEL

1	Data <i>Benchmark</i>	12
2	Kombinasi parameter	14
3	Cuplikan matriks similaritas dari DOID NSCLC	27
4	Karakteristik topologi jaringan benchmark dan komorbiditas.	30
5	Ketepatan prediksi algoritma terbaik	43
6	Kesesuaian peringkat hasil validasi	43
7	Perbandingan kinerja algoritma pada jaringan NSCLC	44
8	Perbandingan kinerja algoritma pada jaringan SCLC	45
9	Nilai karakteristik hasil 20 kali run pada jaringan NSCLC dan SCLC	46
10	Normalisasi dan interpretasi klinis sepuluh <i>stable significant nodes</i> teratas pada jaringan NSCLC dan SCLC	46
11	Uji sensitivitas <i>threshold</i> pada jaringan NSCLC	50
12	Uji sensitivitas <i>threshold</i> pada jaringan SCLC	52
13	Interpretasi molekuler overlapping nodes yang stabil pada jaringan NSCLC	56
14	Interpretasi molekuler <i>overlapping nodes</i> yang stabil pada jaringan SCLC	57

DAFTAR GAMBAR

1	Ilustrasi representasi partikel dan mekanisme pembaruan label komunitas pada DPSO	6
2	Ilustrasi hasil deteksi komunitas pada struktur jaringan yang sama: (a) hasil <i>non-overlapping</i> yang menempatkan setiap node hanya pada satu komunitas dan (b) hasil <i>overlapping</i> yang menempatkan node <i>v5</i> sebagai anggota komunitas <i>C1</i> dan <i>C2</i> .	7
3	Ilustrasi Keanggotaan Fuzzy	9
4	Tahapan penelitian	11
5	Proses pembentukan jaringan komorbiditas	13
6	Flowchart DPSO-Fuzzy	16
7	Proses pemilihan algoritma berbasis topologi	22
8	Proses untuk identifikasi komorbiditas <i>overlapping</i>	26
9	Analisa sensitivitas topologi jaringan NSCLC	28
10	Analisa sensitivitas topologi jaringan SCLC	28
11	Visualisasi jaringan SCLC dengan <i>threshold</i> 0,6.	29
12	Evaluasi model pada dataset Karate	31
13	Evaluasi model pada dataset Dolphins	32
14	Evaluasi model pada dataset Football	32
15	Evaluasi model pada dataset Yeast-D2	33
16	Evaluasi model pada dataset Y2H	33
17	Visualisasi hasil deteksi komunitas pada data Karate. (a) Struktur komunitas <i>ground truth</i> . (b) DPSO <i>community assignment</i> sebelum proses fuzzy. (c) DPSO-Fuzzy dengan nilai $\alpha=0,4$.	35
18	Nilai PC1_norm pada Karate	36

19	Nilai PC1_norm pada Dolphins	36
20	Nilai PC1_norm pada Football	37
21	Nilai PC1_norm pada Yeast-D2	37
22	Nilai PC1_norm pada Y2H	38
23	Heatmap $\bar{Z}$ -Score	38
24	Heatmap korelasi Spearman	40
25	Perbandingan koefisien Gini dan nilai Omega pada DPSO-Fuzzy di lima dataset benchmark	40
26	Jarak topologi antara target dengan <i>benchmark</i>	41
27	Decision score pada jaringan NSCLC	42
28	Decision score pada jaringan SCLC	42
29	Diagram venn <i>overlapping nodes</i> pada jaringan NSCLC	51
30	Stable <i>overlapping nodes</i> pada jaringan NSCLC; <i>threshold</i> 0,5	51
31	Diagram venn <i>overlapping nodes</i> pada jaringan SCLC	52
32	Stable <i>overlapping nodes</i> pada jaringan SCLC dengan <i>threshold</i> 0,5	53

DAFTAR LAMPIRAN

1	Lampiran 1 DOID untuk jaringan NSCLC	70
2	Lampiran 2 DOID untuk jaringan SCLC	72
3	Lampiran 3 Parameter final yang digunakan	74
4	Lampiran 4 Hasil evaluasi pada data <i>benchmark</i>	76
5	Lampiran 5 Node signifikan yang stabil pada NSCLC	78
6	Lampiran 6 Node signifikan yang stabil pada SCLC	79

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.