

ABSTRAK

Kanker ginjal, khususnya subtype *Kidney Renal Clear Cell Carcinoma* (KIRC), merupakan tantangan kesehatan global yang signifikan dengan mekanisme penyakit yang dikendalikan oleh interaksi protein (*Protein-Protein Interaction*/PPI). Protein dalam jaringan ini bersifat dinamis dan sering kali memiliki peran ganda dalam berbagai jalur fungsional secara bersamaan, sehingga membentuk struktur komunitas yang saling tumpang tindih (*overlapping*). Masalah utama dalam studi bioinformatika saat ini adalah ketergantungan pada metode deteksi komunitas konvensional yang bersifat eksklusif atau *disjoint*. Metode tersebut gagal menangkap sifat multi-fungsional protein, sehingga informasi mengenai modul biologis menjadi kurang akurat. Selain itu, hasil analisis dari model *machine learning* yang kompleks sering kali sulit diinterpretasikan oleh peneliti medis karena kurangnya sistem visualisasi integratif yang menghubungkan hasil teknis dengan analisis pengayaan fungsional yang mendalam.

Penelitian ini menggunakan pendekatan kuantitatif non-eksperimental berbasis *in silico* dengan mengadopsi kerangka kerja CRISP-DM. Data sekunder diperoleh dari platform IntOGen dan cBioPortal, mencakup dataset *cohort* penderita kanker ginjal. Konstruksi jaringan interaksi protein dilakukan melalui basis data STRING-DB dengan parameter *Medium Confidence* (0,400), menghasilkan jaringan dengan 585 simpul protein dan 11.812 interaksi. Algoritma yang diimplementasikan adalah *Neural Overlapping Community Detection* (NOCD), sebuah model *deep learning* yang menggabungkan *Graph Neural Network* (GNN) sebagai *encoder* dan model generatif Bernoulli-Poisson sebagai *decoder*. Pengembangan sistem visualisasi dilakukan melalui empat fase utama yaitu analisis, perancangan, integrasi, dan pengujian menggunakan bahasa Python dengan *framework* Django untuk *backend* dan React untuk *frontend*.

Hasil penelitian menunjukkan bahwa model NOCD berhasil mengidentifikasi 16 komunitas tumpang tindih dari 585 simpul protein. Keunggulan model dibuktikan melalui peningkatan metrik topologi yang signifikan, di mana nilai *density* meningkat dari *baseline* 0,0691 menjadi 0,3445 dan *clustering coefficient* naik dari 0,0011 menjadi 0,0324. Model juga menghasilkan nilai *coverage* 0,8486 dan *conductance* 0,5552, yang menunjukkan cakupan interaksi luas serta pemisahan komunitas yang baik. Validasi biologis melalui *enrichment analysis* menggunakan platform g:Profiler mengonfirmasi bahwa komunitas yang terbentuk memiliki relevansi kuat dengan patogenesis kanker ginjal. Temuan krusial menunjukkan Komunitas 2 sebagai klaster utama yang mengandung protein *master regulator* VHL dan HIF1A, sementara Komunitas 11 memiliki signifikansi statistik tertinggi terkait proses organisasi kromatin. Kontribusi penelitian ini terletak pada penyediaan alur kerja integratif yang mampu menggabungkan akurasi teknis GNN dengan interpretasi biologis yang mendalam, sehingga memberikan wawasan baru bagi upaya pemahaman mekanisme molekuler kanker ginjal secara aplikatif.

Kata Kunci: Kanker Ginjal, *Protein-Protein Interaction* (PPI), *Graph Neural Network* (GNN), Deteksi Komunitas Tumpang Tindih, NOCD, *Enrichment Analysis*.

ABSTRACT

Kidney cancer, particularly the Kidney Renal Clear Cell Carcinoma (KIRC) subtype, represents a significant global health challenge with disease mechanisms governed by Protein-Protein Interaction (PPI). Proteins within these networks are dynamic and frequently perform dual roles across multiple functional pathways simultaneously, resulting in overlapping community structures. A major challenge in current bioinformatics studies is the reliance on conventional community detection methods that are disjoint or exclusive. These methods fail to capture the multi-functional nature of proteins, leading to inaccurate identification of biological modules. Furthermore, results from complex machine learning models are often difficult for medical researchers to interpret due to the lack of integrative visualization systems that bridge technical outputs with in-depth functional enrichment analysis.

This study employs a non-experimental quantitative in-silico approach, adopting the CRISP-DM framework. Secondary data were acquired from the IntOGen and cBioPortal platforms, encompassing cohort datasets of kidney cancer patients. The protein interaction network was constructed using the STRING-DB database with a Medium Confidence parameter (0.400), resulting in a network of 585 protein nodes and 11,812 interactions. The implemented algorithm is Neural Overlapping Community Detection (NOCD), a deep learning model that integrates a Graph Neural Network (GNN) as the encoder and a Bernoulli-Poisson generative model as the decoder. The visualization system development followed four main phases, analysis, design, integration, and testing. Utilizing Python with the Django framework for the backend and React for the frontend.

The results demonstrate that the NOCD model successfully identified 16 overlapping communities from the 585 protein nodes. The superiority of the model is evidenced by significant improvements in topological metrics, where the density value increased from a baseline of 0.0691 to 0.3445, and the clustering coefficient rose from 0.0011 to 0.0324. The model also yielded a coverage value of 0.8486 and a conductance of 0.5552, indicating broad interaction coverage and effective community separation. Biological validation through enrichment analysis using the g:Profiler platform confirmed that the formed communities have a strong relevance to the pathogenesis of kidney cancer. Crucial findings highlight Community 2 as a primary cluster containing the master regulator proteins VHL and HIF1A, while Community 11 exhibited the highest statistical significance related to chromatin organization processes. The contribution of this research lies in providing an integrative workflow capable of combining the technical accuracy of GNNs with profound biological interpretation, thereby offering new insights into the molecular mechanisms of kidney cancer in an applicable manner.

Keywords: *Kidney Cancer, Protein-Protein Interaction (PPI), Graph Neural Network (GNN), Overlapping Community Detection, NOCD, Enrichment Analysis.*