



Clustering Protein Diabetes Melitus menggunakan Algoritma MCL

Wa Ode Rahma Agus Udaya Manarfa *¹

¹ Teknik Informatika, Universitas Dayanu Ikhsanuddin, Baubau

e-mail: rahmamanarfa@unidayan.ac.id ¹

* Corresponding Author

Received: 22 Maret 2024

Revised: 1 April 2024

Accepted: 5 Mei 2024

Abstrak

Diabetes Melitus merupakan penyakit yang kompleks dengan berbagai tingkat komplikasi sistemik. *Clustering* PPI merupakan salah satu metode komputasi yang ditawarkan untuk mengidentifikasi protein-protein yang berpengaruh terhadap Diabetes Melitus tipe 2. Teknik *clustering* merupakan salah satu metode yang dapat digunakan dalam menemukan protein pelengkap sehingga menemukan obat menjadi lebih mudah. Penelitian sebelumnya mengidentifikasi protein-protein signifikan yang berasosiasi dengan penyakit Diabetes Melitus (DM) tipe 2 menggunakan analisis topologi jaringan interaksi protein yang hasilnya menunjukkan bahwa terdapat 21 protein signifikan yang berasosiasi dengan Diabetes Melitus (DM) tipe 2. Tujuan dari penelitian ini adalah mendapatkan *cluster* protein signifikan yang memiliki anggota Protein yang berinteraksi langsung dengan penyakit DM tipe 2. Penelitian ini telah dilakukan dengan beberapa tahapan yaitu data protein DM tipe 2, pencarian protein signifikan *clustering* MCL yang menghasilkan data protein kompleks DM tipe 2. Jaringan *backbone* yang dihasilkan memuat protein signifikan pada penyakit DM tipe 2, hal ini merepresentasikan bahwa hasil *output* penggunaan algoritma adalah jaringan inti pada penyakit DM tipe 2.

Kata kunci: DM tipe 2, markov clustering, interaksi antar protein, graf protein

Abstract

Diabetes Mellitus is a complex disease with varying degrees of systemic complications. PPI clustering is one of the computational methods offered to identify proteins that influence type 2 diabetes mellitus. The clustering technique is one method that can be used to find complementary proteins so that finding drugs becomes easier. Previous research identified significant proteins associated with type 2 Diabetes Mellitus (DM) using protein interaction network topology analysis, the results of which showed that there were 21 significant proteins associated with type 2 Diabetes Mellitus (DM). The aim of this research was to obtain clusters. significant proteins that have protein members that interact directly with type 2 DM disease. This research has been carried out in several stages, namely type 2 DM protein data, searching for significant proteins clustering MCL which produces type 2 DM complex protein data. The resulting backbone network contains significant proteins in type 2 DM, this represents that the output result of using the algorithm is the core network in type 2 DM.

Keywords: DM type 2, markov clustering, protein-protein interaction, protein graph

PENDAHULUAN

Diabetes Melitus (DM) merupakan penyakit yang kompleks dengan berbagai tingkat komplikasi sistemik (Leman 2012). Menurut *World Health Organization* (WHO 2017), DM merupakan gangguan metabolisme yang ditandai dengan naiknya kadar glukosa dalam darah (hiperglikemi), yang menimbulkan komplikasi kronik. Indonesia menduduki peringkat kelima se-Asia Tenggara penyandang Diabetes terbanyak setelah India. Diabetes adalah penyebab utama kebutaan, gagal ginjal, serangan jantung, stroke dan amputasi anggota badan bagian bawah.

Jaringan protein-protein interaksi membentuk peta jaringan protein skala besar dan sistematis atau disebut dengan jaringan *interactome*, dimana sebagian besar protein saling berinteraksi dengan beberapa pasangan protein lain, sedangkan sebagian kecil protein yang merupakan protein

signifikan berinteraksi dengan banyak pasangan protein (Han *et al.* 2004). *Clustering* PPI merupakan salah satu metode komputasi yang ditawarkan untuk mengidentifikasi protein-protein yang berpengaruh terhadap Diabetes Melitus tipe 2. Teknik *clustering* merupakan salah satu metode yang dapat digunakan dalam menemukan protein pelengkap sehingga menemukan obat menjadi lebih mudah (Singh dan Singh 2011). Penelitian sebelumnya oleh Singh dan Singh (2011), melakukan analisis *cluster* protein berdasarkan *bipartite graph* yang mencakup urutan protein dan *Sequence Derived Features* (SDF) untuk setiap urutan protein yang digunakan untuk memprediksi kelas protein pada manusia. Selain itu, penelitian lain yang dilakukan oleh Usman (2019) mengidentifikasi protein-protein signifikan yang berasosiasi dengan penyakit Diabetes Melitus tipe 2 menggunakan analisis topologi jaringan interaksi protein yang hasilnya menunjukkan bahwa terdapat

21 protein signifikan yang berasosiasi dengan Diabetes Melitus tipe 2.

Data protein DM tipe 2 direpresentasikan dalam graf. Data tersebut memiliki keterhubungan satu sama lain dengan jumlah interaksi yang banyak yang tidak banyak memberikan informasi dalam pencarian obat multi target dan peramalan kemungkinan munculnya penyakit baru. Oleh karenanya diperlukan pencarian *cluster* protein yang memuat hubungannya dalam skala yang lebih kompleks menggunakan algoritma MCL. Tujuan dari penelitian ini adalah mendapatkan *cluster* protein signifikan yang memiliki anggota Protein yang berinteraksi langsung dengan penyakit DM tipe 2.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini menggunakan pendekatan *unsupervised learning*. *Unsupervised learning* bekerja dengan cara menemukan struktur dari data tanpa perlu diberi label terlebih dahulu.

Waktu dan Tempat Penelitian

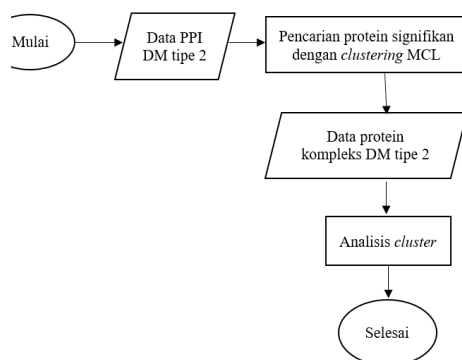
Penelitian dilakukan di Laboratorium Bioinformatika Institut Pertanian Bogor pada Februari 2018 sampai Agustus 2019.

Target Penelitian

Target penelitian ini adalah menemukan pola hubungan antara protein tertarget penyakit Diabetes Melitus tipe 2.

Prosedur

Penelitian ini dilakukan dengan beberapa tahapan yaitu data protein DM tipe 2 diperoleh dari penelitian Usman *et al.* (2019), pencarian protein signifikan *clustering* MCL yang menghasilkan data protein kompleks DM tipe 2, dan analisis *cluster*. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Data, Instrumen, dan Teknik Pengumpulan Data

Data yang digunakan pada penelitian ini adalah (1) data protein yang berasosiasi dengan DM tipe 2 yang didapat dari penelitian Usman *et al.* (2019) sebanyak 63 protein/gen, (2) data protein interaksi DM tipe 2 diambil dari basis data STRING

(*Search Tool for Retrieval of Interacting genes/protein*) dengan cara penambangan dari data protein yang berasosiasi dengan DM tipe 2 menghasilkan interaksi protein sebanyak 2341 interaksi protein.

Teknik Analisis Data

Pencarian Protein Signifikan dengan Markov Clustering (MCL)

Setelah data protein diperoleh, tahap selanjutnya adalah mencari protein signifikan dengan menggunakan algoritme Markov Clustering. Pada tahapan ini menggunakan data PPI DM tipe 2 yang kemudian dilakukan clustering dengan algoritme MCL sehingga menghasilkan data protein kompleks DM tipe 2.

Modularity merupakan salah satu ukuran tingkat konektivitas dalam pengelompokan graf. Graf dengan nilai *modularity* yang tinggi memiliki konektivitas yang padat antara *node* dalam *cluster*, tetapi memiliki konektivitas yang jarang antara *node* dalam *cluster* yang berbeda. *Modularity* sering digunakan dalam metode optimasi untuk mendeteksi struktur komunitas dalam graf. Rumus 1 merupakan Rumus untuk digunakan untuk mendapatkan nilai *modularity* pada suatu graf (Ochieng *et al.*, 2017).

$$C_{mod} = \frac{E_{in}}{E_{out}} \quad (1)$$

dimana:

C_{mod} : nilai *modularity*.

E_{in} : jumlah *edge* diantara *node* yang menjadi anggota dalam sebuah *cluster*.

E_{out} : jumlah *edge* diantara *node* yang bukan anggota dalam sebuah *cluster*.

Analisis Cluster

Analisis *cluster* pada penelitian ini didasarkan pada data protein signifikan DM tipe 2 pada basis data DAVID (*database for anotation, visualization and integrated discovery*). Tahap kedua adalah melakukan analisis berdasarkan hasil penelitian yang dilakukan oleh Usman *et al.* (2019) yang melakukan identifikasi protein-protein signifikan dengan menganalisis topologi interaksi protein-protein pada penyakit DM. Protein-protein signifikan yang diperoleh pada penelitian Usman *et al.* (2019) akan dilihat keanggotaannya dalam *cluster* protein yang memiliki nilai *modularity* tertinggi.

HASIL PENELITIAN DAN PEMBAHASAN

Persiapan Data

Data protein yang berasosiasi dengan DM tipe 2 yang diambil dari penelitian Usman *et al.* (2019) sebanyak 63 protein/gen yang diperoleh dari basis data OMIM (*online mendelian inheritance in man*). Selanjutnya data protein tersebut digunakan sebagai masukan pada basis data STRING (*search tool for retrieval of interacting genes/protein*) untuk dilakukan penambangan data PPI DM tipe 2. Kemudian dari data protein yang berasosiasi dengan DM tipe 2 diperoleh data interaksi protein

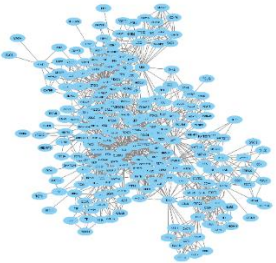




sebanyak 2341 interaksi protein yang melibatkan 481 protein yang berinteraksi.






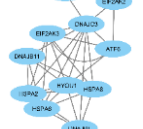





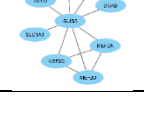
Pencarian Protein Signifikan dengan Markov Clustering (MCL)




Clustering jejaring interaksi protein dengan algoritme MCL ini dilakukan dengan menggunakan aplikasi Cytoscape yang telah menyediakan package untuk melakukan clustering dengan menggunakan algoritme MCL. Pada penelitian ini, nilai parameter-parameter yang digunakan adalah sebagai berikut: 1) Granularity parameter (inflation value) = 2; 2) Weak edge pruning threshold = 1×10^{-10} ; 3) Number of iterations = 15; dan 4) Maximum residual value = 1×10^{-3} .

Clustering jejaring interaksi protein dengan menggunakan algoritme MCL ini menghasilkan 21 cluster protein. Jejaring pada cluster 1 merupakan jejaring interaksi protein yang memiliki modularity score tertinggi yang akan digunakan pada tahap selanjutnya. Jejaring tersebut tersusun atas 238 protein yang terdiri dari 38 protein yang berasosiasi langsung dengan penyakit DM tipe 2 dan 200 protein yang tidak berasosiasi langsung dengan penyakit DM tipe 2. Protein-protein yang berasosiasi dengan penyakit DM tipe 2 tersebut adalah ABCC8, AKT2, CCR5, CDKAL1, CTLA4, EPO, FOXP3, GCGR, GCK, HMGA1, HNF1A, HNF1B, HNF4A, IAPP, IGF2BP2, IL2RA, INS, INSR, IRS1, IRS2, KCNJ11, KLF11, MAPK8IP1, NEUROD1, PAX4, PDX1, PON1, PPARG, PTPN22, RETN, SLC2A2, SLC30A8, SOD2, SUMO4, TBC1D4, TCF7L2, UCP3 dan WFS1.

Analisis Cluster

No.	Hasil clustering	Modularity score
1.		0,98
2.		0,06
3.		0,05
4.		0,04
5.		0,03

No.	Hasil clustering	Modularity score
6.		0,02
7.		0,02
8.		0,02
9.		0,02
10.		0,02
11.		0,02
12.		0,02
13.		0,02
14.		0,02
15.		0,02
16.		0,02
17.		0,02
18.		0,02

No.	Hasil clustering	Modularity score
19.		0,02
20.		0,02
21		0,01

Protein signifikan menurut penelitian Usman (2019) yang tersebar di masing-masing cluster terdapat pada Tabel 1.

Tabel 1. Protein signifikan menurut penelitian Usman yang terdapat pada cluster terpilih (2019)

No.	Protein Signifikan (Usman (2019))	Cluster 1	
		Ada	Tidak ada
1	INS	√	
2	AKT1	√	
3	TCF7L2	√	
4	KCNJ11	√	
5	PPARG		√
6	GCG	√	
7	NSR	√	
8	IAPP	√	
9	SOCS3		√
10	UBC	√	
11	STAT3		√
12	PTH		√
13	APOE	√	
14	PRKACA		√
15	FOXO1	√	
16	EP300	√	
17	SOD3	√	
18	WFS1	√	
19	MTNR1B		√
20	CTLA4	√	
21	PPARA		√

Protein signifikan menurut penelitian Usman (2019) yang terdapat pada cluster 1 hanya terdapat 14 protein.

KESIMPULAN DAN SARAN

Kesimpulan

Penerapan algoritma MCL pada clustering data PPI penyakit DM tipe 2 dapat digunakan untuk mendapatkan protein signifikan yang berinteraksi langsung dengan penyakit DM tipe 2. Kesimpulan yang didapat dari penerapan algoritme ini pada data PPI penyakit DM tipe 2 adalah cluster terpilih yang dihasilkan memuat protein signifikan pada penyakit DM tipe 2, hal ini merepresentasikan bahwa hasil output penggunaan algoritme adalah protein signifikan pada penyakit DM tipe 2.

Saran

Penerapan algoritme MCL pada penelitian penyakit DM tipe 2 dapat dilengkapi dengan integrative gen ontologi semisal biological process, cellular component dan molecular function sehingga didapat output yang lebih kompleks.

DAFTAR PUSTAKA

Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J. M., Cusick, M. E., Roth, F. P. & Vidal, M. 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, vol. 430, pp. 88-93.

Leman, M. A. 2012. Periodontal Disease In Diabetic Patients. *Jurnal Biodemik*, vol. 4, no. 1, pp. 1-4.

Ochieng, P. J., Kusuma, W. A. & Haryanto, T. 2017. Detection of protein complex from protein-protein interaction network using Markov clustering. *Journal of Physics*, vol. 835, no. 1, p. 012001.

Organization, W. H. 2018. World Health Organization. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/diabetes>.

Singh, M. & Singh, G. 2011. Cluster Analysis Technique based on Bipartite Graph. *International Journal of Computer Applications*, vol. 20, no. 3, pp. 22-27.

Usman, M. S., Kusuma, W. A., Afendi, F. M. & Heryanto, R. 2019. Identification of Significant Proteins Associated with Diabetes Mellitus Using Network Analysis of Protein Interactions. *Computer Engineering and Applications*, vol. 8, no. 1, pp. 41-52.