

Klasifikasi Fungsi Family Protein Transport Menggunakan Radial Basis Neural Network

Classification Function of Transport Protein Family Using Radial Basis Neural Network

Green Arther Sandag¹, Fergie Kaunang²

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Klabat
Jl. Arnold Mononutu, Airmadidi – Minahasa Utara, Manado, Sulawesi Utara
E-mail: ¹greensandag@unklab.ac.id, ²fergie@unklab.ac.id

Abstrak

Transporter adalah protein transmembran yang penting dalam proses masuk dan keluarnya ion atau molekul sel di seluruh protein membran dan memainkan peran penting dalam mengenali sistem kekebalan tubuh dan transduser energi. Dalam beberapa tahun terakhir, penelitian sebelumnya telah dilakukan untuk menganalisis protein transport, terutama diskriminasi kelas dan familynya dalam memainkan peran penting dalam system control sel, mengangkut air, sinyal kimia dan listrik. Protein transport membrane cenderung membentuk system pompa dan channel span, serta span cell membrane. Oleh karena itu, membedakan kelas dan family transport protein adalah tugas penting dalam ilmu komputasi biologi dan diperlukan bagi para ahli biologi untuk mendapatkan pemahaman yang lebih baik tentang fungsi protein transport. Oleh karena itu, dalam penelitian ini, telah dilakukan pengembangan metode untuk mengidentifikasi fungsi kelas utama dan family protein transport menggunakan radial basis neural network. Peneliti telah menganalisis karakteristik komposisi asam amino, komposisi residu pair pada protein transport. Metode dalam klasifikasi kelas protein transport untuk mengetahui fungsi protein transport peneliti menggunakan PSSM dengan metode quickRBF classifier memberikan hasil akurasi terbaik dibanding dengan metode yang lain. Hasil akurasi sebesar 84,84% untuk cross validation dan 80,71% untuk independent data, oleh karena itu maka metode yang peneliti usulkan dapat digunakan secara efektif untuk mengidentifikasi dan mendiskriminasi transporter ke dalam kelas protein transport dengan peningkatan 6-10 % dari penelitian yang sejenis.

Keywords—*transporter, membran, quickRBF, PSSM*

Abstract

Transporters are important transmembrane proteins that involve the cellular entry and exit of ions or molecules throughout the membrane proteins and thereby play important roles in recognizing the immune system and energy transducers. In recent years, several studies have been conducted to analyze the transport proteins; especially the discrimination class of transporters and their subfamilies play crucial roles in cell control system, transporting water, chemical and electric signals. Membrane transport proteins tend to form an intricate system of pumps and channel span, and span cell membranes. Hence, discriminating the specific class of transporters and their subfamilies is an essential task in computational biology and necessary for biologists to gain better understanding about the function of transport proteins. Therefore, in this study, an attempt has been made to develop a method that used radial basis neural network to identify the function of transport proteins in major class and family. We have

analyzed the characteristics of amino acid composition (AAC), dipeptide pair composition (DPC) in transport proteins, also we used PSSM with quickRBF classifier method give the best accuracy results compared to other methods. Accuracy results of 84.84% for cross validation and 80.71% for independent data, therefore the method we proposed can be used effectively to identify and discriminate transporters with increasing an accuracy of 6-10%.

Keywords—Transporters, AAC, DPC, PSSM, quickRBF, cross validation

1. PENDAHULUAN

Protein membrane adalah protein yang ditemukan dalam membrane biologis, protein ini dapat didefinisikan sebagai protein yang terkait dengan sel membrane [1]. Protein membrane mencakup berbagai fungsi dalam organisme [2], dan merupakan objek target lebih dari 50% dalam pembuatan obat di era modern [3], serta diperkirakan bahwa 20-30% semua gen dalam genom dikodekan berdasarkan protein membrane [4]. Salah satu kelas utama protein membrane adalah transporter atau protein transport. Protein transport membrane cenderung membentuk system pompa dan rentang saluran yang rumit, dan membentang di sekitar sel membrane. Umumnya, klasifikasi transporter untuk protein transport dikenal sebagai channels/pores, electrochemical transporters, active transporters, group trans locators, electron carriers, factor involved in transport, dan incompletely characterized transport system channels/pores [4]. Klasifikasi transporter berdasarkan kelas dan family yang berbeda merupakan isu penting untuk menentukan struktur dan fungsi genomic. Dalam beberapa tahun terakhir, beberapa metode telah diusulkan untuk membedakan protein membrane, terutama membedakan kelas transport protein dan subfamilynya. Berdasarkan metode yang diusulkan dari Gromiha dan Yabuki [5] dengan menggunakan neural network dalam mengklasifikasikan transport protein ke dalam tiga kelas utama yaitu Channels/pores, electrochemical, dan active transporters. Dalam penelitian sebelumnya juga, Ou telah menganalisis komposisi asam amino, komposisi pasangan residu dan sifat asam amino dalam tiga kelas dan enam family dengan menggunakan metode Position Specific Scoring Matrix (PSSM) dengan tingkat akurasi sebesar 78% [6-8]. Penelitian sebelumnya juga telah membuat tools untuk prediksi protein transport menggunakan PSSM [7]. Melihat hasil yang didapat pada penelitian sebelumnya maka peneliti mengharapkan dan mengajukan metode klasifikasi fungsi protein dalam transport protein berdasarkan kelas protein transport dan dapat meningkatkan tingkat akurasi tersebut dengan menggunakan algoritma pembelajaran mesin.

2. METODE PENELITIAN

2.1 Dataset

Dataset dalam penelitian ini dikumpulkan dari tiga kelas utama TCDB (Transport Classification Database) yaitu protein channels/pores, electrochemical transporters, dan active transporters yang memiliki jumlah data yang cukup besar pada TCDB. Kemudian, peneliti menghapus sequence protein yang tidak memiliki evidence pada protein level dengan menggunakan BLAST. Sequence protein yang memiliki similarity >20% identitas dihapus atau dikeluarkan dari dataset. Peneliti juga memeriksa dataset setelah dihapus similaritynya dari TCDB pada UniProt database dengan menggunakan protein ID dan hanya menggunakan reviewed section pada UniProt [9] sebagai dataset pada penelitian ini. Akhirnya, peneliti mendapatkan dataset yang berisi 870 protein channels/pores, 701 electrochemical transporters, 777 active transporters, dan 589 protein group translocators.

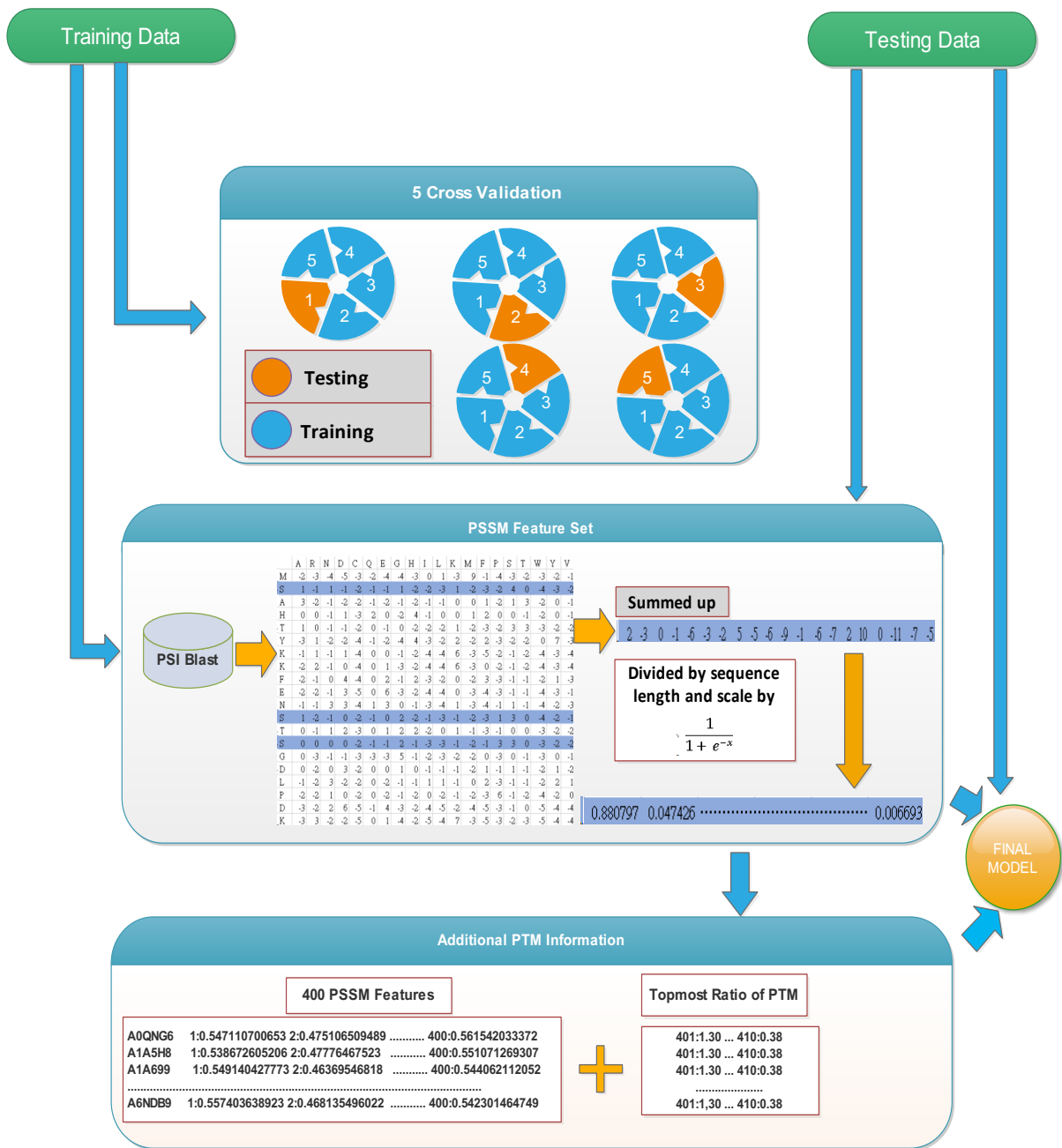
Tabel 1 Dataset Yang Digunakan Dalam Penelitian

Class	Description	Original Data	Reduced Data	Reviewed (Uniprot)	CV Data	Independent Data
1	Channels/Pores	3839	2696	870	713	157
2	Electrochemical Potential-driven Transport	3607	1459	701	562	139
3	Primary Active Transporters	3507	1630	777	629	148
4	Group Translocators	2481	1259	589	471	118
Total		13434	7044	2937	2375	562
CV, Cross Validation						

Seperti pada Table 1 menunjukkan adanya 4 class protein transport. Kemudian peneliti memilih secara acak 562 protein untuk independent test dan sisanya untuk cross validation sebanyak 2375 protein. Pada tahap ini terdapat training data untuk Class 1 adalah 713 protein, Class 2 sebanyak 562 protein, Class 3 sebanyak 629 protein, dan Class 4 sebanyak 471 protein. Selanjutnya untuk independent dataset terbagi menjadi Class 1 sebanyak 157 protein, Class 2 sebanyak 139 protein, Class 3 sebanyak 148 protein, dan Class 4 sebanyak 118 protein.

2.2 Arsitektur pembuatan model

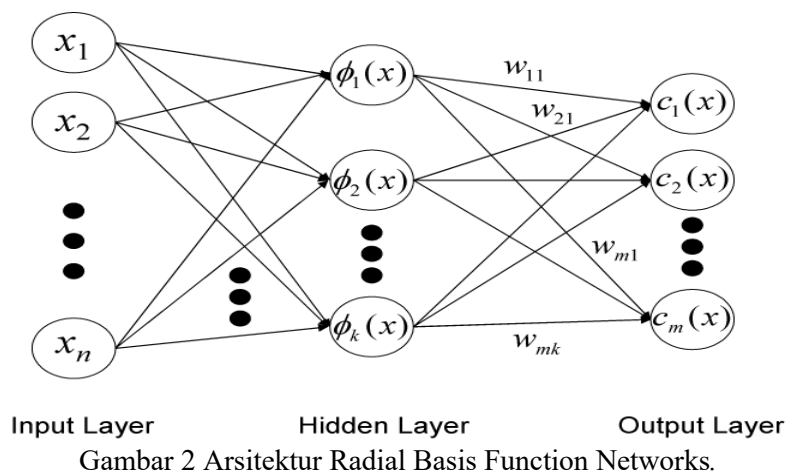
Arsitektur pembuatan model klasifikasi untuk mengidentifikasi fungsi protein transport dapat dilihat pada Gambar 1. Data training dilakukan uji 5-fold cross validation kemudian menggunakan PSSM, AAC, dan DPC untuk mendapatkan model prediksi. Peningkatan performance dari model prediksi dapat dilakukan dengan menambahkan beberapa informasi ke dalam dataset, informasi tersebut seperti F-Score.



Gambar 1 Arsitektur untuk Mengidentifikasi Fungsi Protein Transport

2.3 Algoritma Radial Basis Neural Network

Pada penelitian ini peneliti menggunakan package QuickRBF untuk membangun klasifikasi RBFN dengan pengaturan default [10]. Arsitektur dari RBFN ditunjukkan pada gambar 2. Seperti pada Gambar 2, RBFN pada umumnya terdiri dari tiga layer, yaitu input layer, hidden layer, dan output layer. Input layer akan melakukan broadcasts koordinat input vector ke masing-masing node di hidden layer. Setiap node di hidden layer menghasilkan radial basis kernel function. Setiap node di output layer menghasilkan kombinasi linear dari aktivasi hidden node.



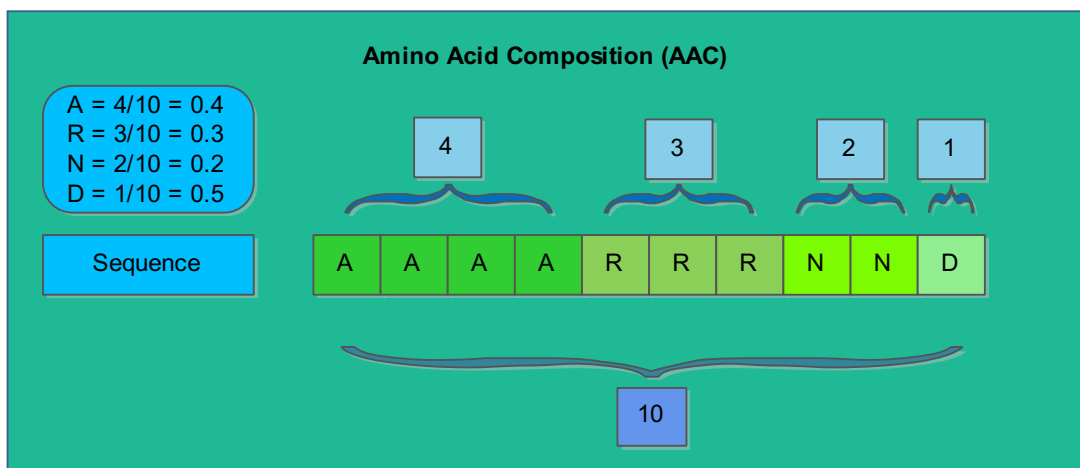
Gambar 2 Arsitektur Radial Basis Function Networks.

Formula dalam RBFN adalah

$$g_j(x) = \sum_{i=1}^k w_{ij} \phi(\|x - \mu_i\|; \sigma_i);$$

$g_j(x)$ adalah fungsi yang sesuai dengan node output j dan merupakan kombinasi linear k dari fungsi radial basis $\phi(\mu_i)$ dengan pusatnya μ_i and layer σ_i dan w_{ij} menunjukkan bobot yang terkait dengan korelasi antara node output j . Metode RBFN memiliki banyak aplikasi dalam memecahkan masalah bioinformatika dan telah banyak digunakan untuk meprediksi cleavage sites dalam protein[11], fungsi protein transport [12], protein disorder[13], diskriminasi protein membrane [14] dan sebagainya.

2.4 Komposisi Asam Amino



Gambar 3 Metode membuat AAC

Protein terdiri dari 20 jenis asam amino, masing-masing asam amino memiliki kode yaitu A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V. Pada penelitian ini menggunakan 20 asam amino dalam data training. AAC adalah jumlah asam amino dari masing-masing jenis dinormalisasi dengan jumlah total residu. Ini dedefinisikan sebagai

Gambar 3 menunjukkan metode AAC untuk menghitung proporsi dari 20 jenis asam amino pada masing-masing komponen dalam protein sequence [15]. Formula menghitung AAC adalah:

$$AAC_{aa,i} = \sum ni * 100/N$$

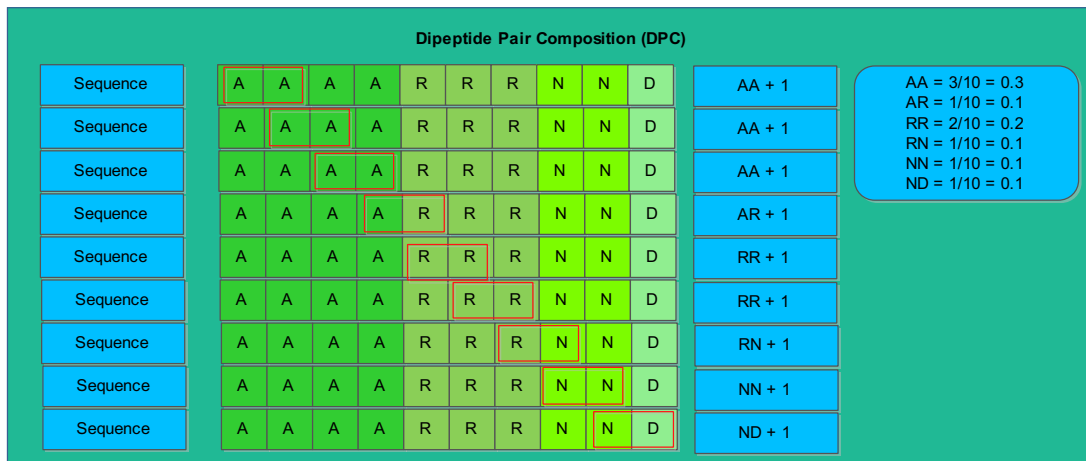
Dimana i adalah 20 residu asam amino (aa) yang terdapat pada protein, n_i adalah total jumlah residu setiap amino acid, dan N adalah panjang sequence .

2.5 Komposisi Asam Amino Pair

Dipeptides adalah jumlah pasangan residu asam amino dalam satu sequence protein. Metode ini dapat menghasilkan 400 kombinasi dipeptide yang berbeda dari 20 asam amino. 400 elemen menunjukkan kemunculan dari semua pasangan residu asam amino kemudian dinormalisasi dengan jumlah total residu dalam protein. Jumlah pasangan residu umumnya dinormalisasi dengan jumlah total residu dalam protein. Pada penelitian ini, normalisasi telah dilakukan dengan jumlah total residu dalam protein sehingga dapat memberikan informasi pasangan residu [16]. Gambar 4 menunjukkan metode dalam menghitung DPC. Formula perhitungan DPC adalah sebagai berikut:

$$DPC_{aa,i,j} = \frac{\sum n_{aa,i,j}}{N}$$

Dimana $\sum n_{aa,i,j}$ adalah jumlah residu (i) pada asam amino (aa) yang berdekatan dengan jenis residu j dan N adalah total jumlah residu. i and j adalah variasi dari 1 to 20 asam amino dan total kombinasinya mencapai 400.

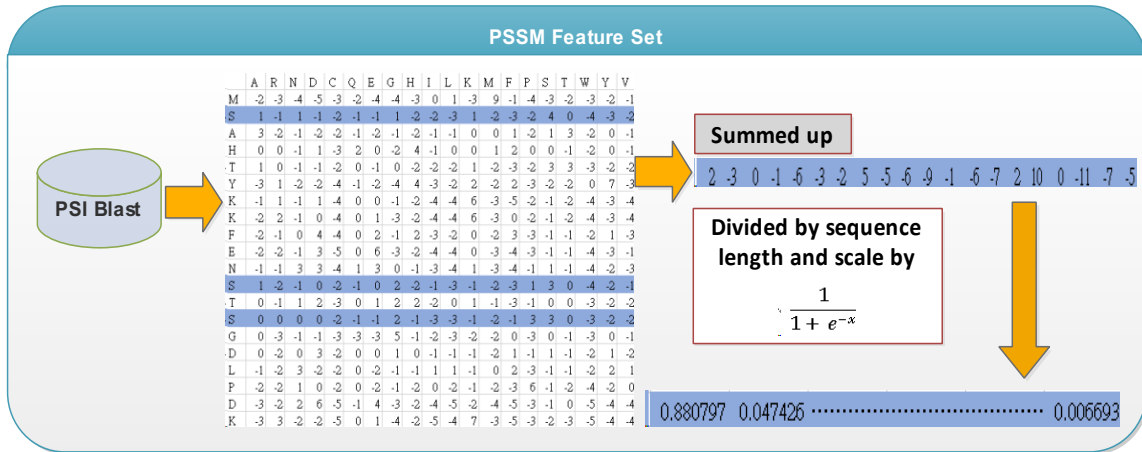


Gambar 4 Metode Membuat DPC

2.6 Position Specific Scoring Matrix

Dalam struktur protein, residu asam amino dapat mengalami mutasi, sehingga memungkinkan dua protein dapat berbagi struktur serupa dengan komposisi asam amino yang berbeda. Oleh karena itu, kami menggunakan PSSM untuk melakukan diskriminasi, metode ini juga telah banyak digunakan dalam memprediksi struktur protein, fungsi protein transport, prediksi target transporter serta masalah dalam bidang bionformatika lainnya [16]. Pada penelitian ini kami menggunakan PSSM profil untuk membuat feature dalam prediksi fungsi protein transport yang menghasilkan input vector 400D sebagai feature yang baru. PSSM diperoleh dengan kemungkinan mutase 20 residu asam amino. Setiap residu akan memiliki 19

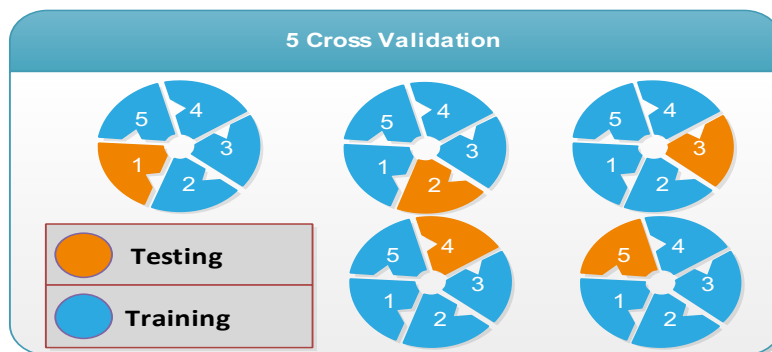
kemungkinan perubahan atau tidak sama sekali. Profil PSSM diperoleh dengan menggunakan database protein PSI-BLAST dan non-redundant (NR) database [16]. Pada Gambar 5 menunjukkan rincian data yang menghasilkan 400 fitur PSSM dari PSSM profil sebelumnya. Setiap elemen vector input sebanyak 400D itu dibagi dengan panjang baris dan kemudian dilakukan scale data dengan rumus $\frac{1}{1+e^{-x}}$.



Gambar 5 Metode Membuat PSSM

2.7 Performance Evaluation

Kinerja prediksi diperiksa dengan 5 cross-validasi dimana teknik ini digunakan untuk mengevaluasi model dengan mempartisi sampel asli menjadi training data untuk melatih model, dan sebuah tes yang ditetapkan untuk mengevaluasinya [17]. Dalam 5 cross-validasi seperti pada Gambar 6, sample asli dipartisi secara acak menjadi lima subsample dengan ukuran yang sama, satu subsample dipertahankan sebagai data validasi atau testing untuk pengujian model, dan sisanya empat subsample digunakan sebagai data pelatihan. Proses cross-validasi kemudian diulang sebanyak 5 kali dengan masing-masing subsample digunakan tepat satu kali sebagai data validasi. Kemudian kelima hasil tersebut dirata-ratakan untuk menghasilkan estimasi tunggal.



Gambar 6 five-fold cross validation

Untuk mengevaluasi hasilnya menggunakan sensitivity, specificity, accuracy, and MCC (Matthew's Correlation Coefficient)[15].

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

[TP – True Positive; FN – False Negative; TN – True Negative; FP – False Positive]

3. HASIL DAN PEMBAHASAN

Percobaan dari penelitian ini dilakukan untuk mengevaluasi kinerja prediksi dari *classifier* RBF *network* dan metode yang diusulkan. Peneliti menggunakan dua dataset, dataset untuk *cross validation* dan *independent*. Dalam percobaan ini, dataset *cross validation* berisi lima subsamples dimana empat subsamples digunakan sebagai data training dan sisanya satu subsamples digunakan sebagai data testing. Kelima hasil dari *cross validation* kemudian dihitung rata-ratanya untuk menghasilkan estimasi tunggal.

Tabel 2 Diskriminasi kelas protein transport: *channels/pores (T1)*, *electrochemical (T2)*, *active transporters (T3)*, *group translocators (T4)*

Method	SENSITIVITY				SPECIFICITY				MCC				ACCURACY
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	TOTAL
Cross-validation													
AAC	73,5	79,2	76,9	76,6	79,03	84,93	74,90	78,9	0,52	0,62	0,54	0,64	79,56
DPC	65,5	80	71,8	76,7	71,34	86,04	76,70	71	0,36	0,64	0,47	0,58	77,80
PSSM	81,9	82,3	81,8	78,1	84,62	90,15	84,70	82,9	0,65	0,71	0,57	0,55	84,84
Independent dataset													
AAC	71,34	76,26	66,22	70,5	76,66	81,97	72,64	71,3	0,47	0,56	0,37	0,54	75,90
DPC	70,06	76,98	65,54	72,4	75,26	82,95	71,62	70,8	0,44	0,58	0,36	0,57	75,82
PSSM	75,16	81,29	70,27	74,5	82,93	85,25	75,34	73,7	0,57	0,64	0,44	0,55	80,7

3.1 Diskriminasi transporter berdasarkan empat kelas

Hasil eksperimen diskriminasi transporter berdasarkan tiga kelas dengan *cross validation* dataset dan *independent dataset* ditunjukkan pada Tabel 2 dengan menggunakan QuickRBF classifier. Hasil terbaik untuk *cross validation* dataset sebanyak 2375 protein diperoleh dari classifier QuickRBF dengan fitur set dengan metode PSSM dengan tingkat akurasi rata-rata sebesar 84,8%, sensitivity sebesar 81%, specificity sebesar 85,5%, dan MCC sebesar 0,62. Akurasi rata-ratanya meningkat sebesar 7% dari metode DPC. Sementara untuk *independent dataset* sebanyak 562 protein mendapatkan akurasi terbaik dengan fitur set PSSM mencapai rata-rata sebesar 80,7%, sedangkan rata-rata nilai sensitivity sebesar 75,3%, specificity sebesar 79,3%, dan MCC sebesar 0,55. Metode PSSM mendapatkan peningkatan akurasi sebesar 4-5% dari metode DPC. Dengan demikian, fitur yang memiliki nilai akurasi terbaik untuk diskriminasi protein transport terhadap empat kelas adalah metode PSSM.

3.2 Diskriminasi transporter berdasarkan empat kelas terhadap classifier yang berbeda

Tabel 3 Diskriminasi *channels/pores (T1)*, *electrochemical (T2)*, *active transporters (T3)*, dan *group translocators (T4)* berdasarkan classifier yang berbeda

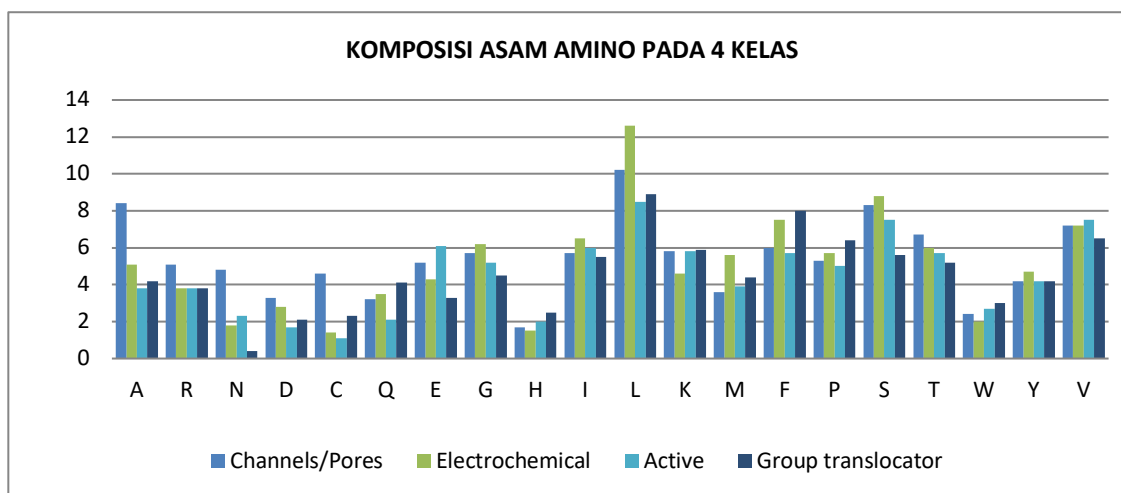
Method	SENSITIVITY				SPECIFICITY				MCC				ACCURACY
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	TOTAL
Cross-validation													
J48	67,6	77,3	59,8	76,6	72,9	80,3	82	79,9	0,4	0,55	0,42	0,54	75
Naïve Bayes	73,9	78	66,2	76,7	74,5	76,7	80	73,3	0,47	0,52	0,46	0,59	75,6
IBK	76,9	82,3	66,4	78,1	82,8	85,6	88,9	81,6	0,59	0,65	0,57	0,57	82,2
Random Forest	78,5	80,8	67,6	78	83,2	88,8	89,8	83,3	0,61	0,68	0,60	0,62	83,5
LIBSVM	79	82,8	76,3	77,6	83,2	86,7	86,8	83,3	0,62	0,67	0,63	0,6	83,5
QuickRBF	81	82,3	82,4	80,5	86,7	90,3	86,5	85,5	0,67	0,72	0,68	0,67	84,84
Independent dataset													
J48	64,9	64	55,4	72,5	78	79,6	74,3	70,3	0,43	0,43	0,29	0,54	72,7
Naïve Bayes	60,5	84,1	70,2	73,4	79	77,7	63,5	73,8	0,40	0,58	0,32	0,58	72,6
IBK	67,5	79,8	64,8	75,5	86	84,9	76,3	73,7	0,55	0,63	0,4	0,55	78,4
Random Forest	66,8	81,2	68,9	75	86,4	88,2	76,3	72,8	0,54	0,68	0,44	0,55	79,8
LIBSVM	69,4	84,8	68,9	77,3	85,7	86,8	76	73	0,56	0,7	0,43	0,58	79,9
QuickRBF	73,2	82,0	69,5	80,3	88,2	87,5	79,3	75,3	0,56	0,69	0,48	0,59	80,7

Hasil percobaan diskriminasi transporter berdasarkan classifier yang berbeda di empat kelas transport protein pada cross validation dataset dan independent dataset ditunjukkan pada Tabel 3. Peneliti mengukur kinerja beberapa classifier terhadap diskriminasi protein transport seperti: decision tree (J48), naïve bayes, KNN, random forest, libsvm dan quickRBF. Berdasarkan hasil yang didapat, classifier dari usulan peneliti bekerja lebih baik daripada classifier yang lain dalam hal akurasi, sensitivity, specificity, dan MCC. Hasil klasifikasi dalam Tabel 4 terhadap cross validation dataset menunjukkan bahwa quikRBF memiliki performance yang baik dengan tingkat akurasi sebesar 84.84%, sensitivity untuk kelas 1 sebesar 73,3%, kelas 2 sebesar 82%, kelas 3 sebesar 69,5%, dan kelas 4 sebesar 80,5, specificity untuk kelas 1 sebesar 88,2%, kelas 2 sebesar 87,5 %, kelas 3 sebesar 79,3%, dan kelas 4 sebesar 75,3% serta MCC untuk kelas 1 sebesar 0,56, kelas 2 sebesar 0,69, kelas 3 sebesar 0,48, dan kelas 4 sebesar 0,59. Sedangkan untuk independent dataset didapat quikRBF classifier mendapatkan hasil yang terbaik untuk akurasi, specificity, sensitifity, dan MCC. Tingkat akurasi pada quikRBF sebesar 80,7% untuk independent dataset. Terjadi peningkatan sebesar 8% dibanding naïve bayes classifier, hal yang sama juga dapat dilihat pada cross validation dataset, quickRBF mendapat peningkatan akurasi sebesar 10% dibanding decision tree (j48) classifier.

3.3 Analisis komposisi asam amino pada setiap kelas dalam protein transport

Analisis ini berfokus pada analisis komposisi asam amino pada setiap kelas protein transport. Peneliti mengamati perbedaan karakteristik urutan protein pada setiap kelas protein transport. Pada Gambar 7 menunjukkan perbedaan komposisi asam amino dan nilai varians dalam empat kelas protein transport. Peneliti memperoleh bahwa residu Ala, Leu, Ser, Var, dan Gly dalam electrochemical transporter memiliki komposisi lebih dari 7 %. Juga residu Cys, Leu,

dan Glu memiliki varians lebih dari 0.5 pada tiga kelas protein transport. Informasi tersebut penting bahwa residu tersebut terdapat dalam hydrophobic dan charged group yang paling dominan pada electrochemical transporter



Gambar 7 Komposisi asam amino pada tiga kelas protein tranport

3.4 Analisis komposisi residu pair berdasarkan tiga kelas protein transport

Analisis ini berfokus pada dipeptide pair variance dan komposisi residu pair pada tiga kelas protein transport. Peneliti menemukan bahwa terdapat perbedaan karakteristik dari 400 residu pair dari asam amino pada tiga kelas protein transport. Seperti yang ditunjukkan pada Tabel 4 terdapat perbedaan varians antara channel/pores, electrochemical, active transporters, dan group translocator. Polar residu pair antar Tyr dan Trp memiliki varians yang tinggi sebesar 1.0 diantara ketiga kelas tersebut.

Tabel 4 Residue Pairs dengan nilai varians tertinggi pada kelas: Channels/Pores, Electrochemical, Active Transporters, dan Group Translocators

Pairs	Channel/Pores	Electrochemical	Active	Group	Variance
YW	3,12	0,68	1,55	1,3	1,02
HY	2,93	0,75	2,38	1,5	0,86
KW	2,03	0,53	2,69	2,6	0,82
YM	2,83	0,63	1,73	1,8	0,8
HT	2,5	0,49	2,1	2,1	0,76
WT	2,89	0,85	1,35	1,4	0,75
WL	2,52	0,42	1,56	1,5	0,74
QW	2,7	0,61	1,5	1,5	0,74
WA	2,84	0,8	1,52	2,1	0,72
YP	2,91	0,88	1,7	2,4	0,7

4. KESIMPULAN

Sebagai protein yang terkait dengan membrane, protein membrane meliputi berbagai fungsi yang luas dalam organisme hidup seperti, sebagai target pembuatan obat modern, sehingga banyak genom dalam protein membrane dapat dikodekan. Protein pada membrane plasma bertindak sebagai reseptor, saluran dan transporter. Protein transport membrane cenderung membentuk system yang rumit dari saluran sampai ke membrane sel. Transporter

adalah protein transmembrane yang penting dalam proses masuk dan keluarnya ion atau molekul sel diseluruh protein membrane dan dengan demikian memainkan peran penting untuk mengenali system kekebalan tubuh dan transduser energy. Peneliti telah menganalisis karakteristik komposisi asam amino, komposisi residu pair pada protein transport. Metode dalam klasifikasi kelas protein transport untuk mengetahui fungsi protein transport peneliti menggunakan PSSM dengan metode quickRBF classifier memberikan hasil akurasi terbaik dibanding dengan metode yang lain. Hasil akurasi sebesar 84,84% untuk cross validation dan 80,71% untuk independent data, oleh karena itu maka metode yang peneliti usulkan dapat digunakan secara efektif untuk mengidentifikasi dan mendiskriminasi transporter ke dalam kelas protein transport dengan peningkatan 6-10 % dari penelitian yang sejenis. Metode PSSM yang digabungkan dengan AAC juga merupakan metode yang terbaik untuk meningkatkan kinerja klasifikasi dan dapat mengidentifikasi fungsi protein dalam protein membrane.

5. SARAN

Untuk penelitian selanjutnya peneliti menyarankan agar dapat mengidentifikasi dan mempelajari lebih lanjut tentang fungsi protein pada kelas transport protein yang lain atau dengan meningkatkan hasil klasifikasi kelas transporter lain dengan metode pembelajaran mesin yang lain seperti deep learning menggunakan Convolutional Neural Network (CNN).

DAFTAR PUSTAKA

- [1] M. Pop, and S.L. Salzberg, *Bioinformatics challenges of new sequencing technology*. Trends in Genetics, 2008. **24**(3): p. 142-149.
- [2] Martí-Renom, M.A., et al., "Comparative protein structure modeling of genes and genomes," *Annual review of biophysics and biomolecular structure*, 2000. **29**(1): p. 291-325.
- [3] M. B. Eisen , et al., "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences*, 1998. **95**(25): p. 14863-14868.
- [4] R. Wernersson, and A.G. Pedersen, "RevTrans: multiple alignment of coding DNA from aligned amino acid sequences," *Nucleic acids research*, 2003. **31**(13): p. 3537-3539.
- [5] L. Holm, and C. Sander, "Protein structure comparison by alignment of distance matrices" *Journal of molecular biology*, 1993. **233**(1): p. 123-138.
- [6] D. T. Jones, "Protein secondary structure prediction based on position-specific scoring matrices," *Journal of molecular biology*, 1999. **292**(2): p. 195-202.
- [7] G. A. Sandag, and S. W. Taju. "Bioinformatics Tools for Data Processing and Prediction of Protein Function". *Cogito Smart Journal*, 4(2), 2019. 305-315.
- [8] D. T. -H. Chang, et al., "Prediction of protein secondary structures with a novel kernel density estimation based classifier," *BMC research notes*, 2008. **1**(1): p. 51.
- [9] U. Consortium, *Reorganizing the protein space at the Universal Protein Resource (UniProt)*. Nucleic acids research, 2011: p. gkr981.
- [10] Y.-Y. Ou, *QuickRBF: an efficient RBFN package*. software available at : <http://csie.org/~yien/quickrbf/quickstart.php>, 2005.

-
- [11] Z. R. Yang, and R. Thomson, "Bio-basis function neural network for prediction of protease cleavage sites in proteins," *IEEE Transactions on Neural Networks*, 2005. **16**(1): p. 263-274.
- [12] G.-Z. Zhang, and D.-S. Huang, "Prediction of inter-residue contacts map based on genetic algorithm optimized radial basis function neural network and binary input encoding scheme," *Journal of computer-aided molecular design*, 2004. **18**(12): p. 797-810.
- [13] C.-T. Su, C.-Y. Chen, and Y.-Y. Ou, "Protein disorder prediction by condensed PSSM considering propensity for order or disorder," *Bmc Bioinformatics*, 2006. **7**(1): p. 319.
- [14] Y.-Y. Ou, et al., "TMBETADISC-RBF: discrimination of-barrel membrane proteins using RBF networks and PSSM profiles," *Computational biology and chemistry*, 2008. **32**(3): p. 227-231.
- [15] N.Q.K. Le, G. A. Sandag, and Y.-Y. Ou. "Incorporating post translational modification information for enhancing the predictive performance of membrane transport proteins," *Computational biology and chemistry* 77 (2018): 251-260.
- [16] S.-A. Chen, et al., "Prediction of transporter targets using efficient RBF networks with PSSM profiles and biochemical properties," *Bioinformatics*, 2011. **27**(15): p. 2062-2067.
- [17] G. Zhang, et al., "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis," *European journal of operational research*, 1999. **116**(1): p. 16-32.