JURNAL ILMIAH FLASH
Vol. 4 No. 2, Halaman: 78 - 83
Desember 2018

KLASIFIKASI KANKER LEUKIMIA MENGGUNAKAN MICROARRAY EKSPRESI GEN

Prisca Deviani Pakan

Departemen Mikrobiologi, Fakultas Kedokteran, Universitas Nusa cendana, Kupang
E-mail: priscapakan@staf.undana.ac.id

Abstrak

Kanker adalah sebuah penyakit yang disebabkan oleh pembelahan secara berlebihan dan tak terkendali darisel-sel dalam tubuh. Teknologi DNA microarray telah memungkinkan untuk mengamati beribu-ribu ekspresi gen dalam waktu bersamaan. Tingkat ekspresi gen dapatdigunakan untuk menentukan jenis sel kanker dari seorang penderita. Penelitian ini bertujuan untuk mengevaluasi kemampuan machine learning dalam mengklasifikasi kanker leukimia menggunakan data microarray ekspresi gen. Hasil percobaan menunjukan bahwa jaringan syaraf tiruan memiliki akurasi sebesar 98% lebih tinggi dibandingkan dengan algoritma lain.

Kata kunci: Klasifikasi, Kanker Leukimia, Microarray, Ekspresi Gen, Machine Learning

PENDAHULUAN

Klasifikasi kanker adalah salah satu tahapan penting dalam penanganan kanker. Hal ini karena dengan mengetahui kelas sebuah kanker, maka seorang dokter akan dengan mudah memberikan penanganan yang tepat sesuai dengan kondisi yang sebenarnya. Secara komputasional, masalah ini dikenal sebagai pengelompokan dan klasifikasi. Penelitian ini membahas klasifikasi leukemia myeloid akut (AML)dan leukemia limfoblastik akut (ALL). Leukemia adalah jenis kanker darah karena gangguan hematologi. Dalam banyak kasus, jika leukemia terjadi dalam limfosit di sumsum tulang maka disebut sebagai leukemia limfoblastik akut di sisi lain ketika gangguan akut terjadi pada sel sumsum tulang, sel darah merah atau trombosit, maka, itu disebut sebagai leukemia myeloid akut. Ribuan orang terkena leukemia dan terbukti menjadi salah satu kanker paling mematikan di antara semua jenis kanker. Identifikasi dan klasifikasi kanker leukemia sangat penting karena pengobatan bervariasi sesuai dengan subtipe leukemia. Pendekatan konvensional mengklasifikasikan kanker berdasarkan karakteristik morfologis telah dipastikan tidak memadai karena kerumitan yang mendasari dan ketidakjelasan dalam klasifikasi kanker. Dibutuhkan sumber daya yang sangat terampil untuk mendeteksi perbedaan di antara sel-sel tumor. Prosedur ini memerlukan waktu dan sangat mahal. Dengan demikian maka prosedur penanganan seperti ini bukanlah suatu solusi yang tepat. Sel bisa tampak sama secara morfologis tetapi bereaksi sangat berlawanan dengan obat dan perawatan sitotoksik (Haferlach. 2005, dkk). Batasan-batasan teknik konvensional ini menunjukan bahwa diperlukan penggunaan kriteria lain untuk klasifikasi tumor. Profil ekspresi gen sel memberikan informasi yang berguna untuk klasifikasi tersebut. Microarray DNA berperan penting karena dapat mengukur ekspresi gen dari ribuan gen secara bersamaan. Di sini, tujuan eksperimental adalah untuk menilai tingkat ekspresi sejumlah besar gen yang berbeda dalam sampel sel tertentu. Ini biasanya dilakukan untuk memutuskan gen mana yang terlibat dalam transformasi dalam fungsi sel seperti keadaan tidak sehat. Oleh karena itu, dimungkinkan untuk menentukan tingkat ekspresi gen sel dalam kondisi yang berbeda. Peralatan microarray DNA dapat digunakan untuk berbagai penggunaan yang luas di mana penting untuk mengukur jumlah relatif atau absolut dari sejumlah besar hasil DNA yang eksplisit. Ini terdiri dari aplikasi seperti sekuensing DNA, deteksi jumlah salinan Genome luas, hibridisasi genom komparatif (CGH), genotipe skala besar dan analisis ekspresi gen. Untuk selanjutnya, analisis profil ekspresi gen dapat memberikan wawasan untuk klasifikasi yang ditetapkan pada indikasi ekspresi gen sel yang sedang dipertimbangkan. Sejumlah penelitian telah

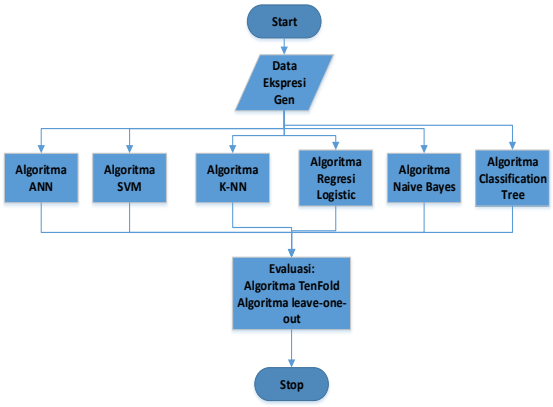
dilakukan dengan menggunakan data ekspresi gen microarray untuk klasifikasi kanker.

Penelitian tentang klasifikasi kanker yang dibangun di atas profil ekspresi gen microarray telah diusulkan oleh Golub et al. [1]. Dalam penelitian ini dikembangkan sistem untuk mengotomatisasi klasifikasi kanker leukemia dengan mengidentifikasi ekspresi gen berkorelasi dari sampel yang diketahui. Studi ini menunjukkan bahwa kanker leukemia dapat diklasifikasikan hanya berdasarkan profil ekspresi gen sel tumor. Metode lain yang digunakan untuk klasifikasi leukemia meliputi metode berdasarkan berbagai teknik Bayesian dan pendekatan pemrograman linier [3, 4]. Peneliti lain telah meneliti penggunaan mesin vektor dukungan untuk klasifikasi leukemia [5, 6] sementara yang lain telah menyelidiki penggunaan jaringan saraf untuk klasifikasi tersebut [7, 8]. Subclass dari ALL dan AML telah diklasifikasikan oleh Berrar et al. [7] menggunakan jaringan saraf probabilistik dengan akurasi 62%. Teori pembelajaran mesin menunjukkan bahwa hasil klasifikasi tergantung pada fitur set input, pada algoritma pelatihan dan pada kemampuan algoritma untuk menyesuaikan data yang mendasarinya. Oleh karena itu, wajib untuk menilai perilaku berbagai pengklasifikasi pada data yang diberikan. Oleh karena itu, perlu diperhatikan kerangka kerja pendekatan pembelajaran mesin yang dapat diterapkan untuk klasifikasi leukemia. Metode pembelajaran mesin telah menunjukkan tingkat perhatian yang tinggi dalam penelitian [7]. Seperti yang dilaporkan dalam berbagai penelitian terbaru, teknik pembelajaran mesin memiliki potensi dalam memberikan akurasi tinggi dalam klasifikasi sebagaimana disamakan dengan prosedur klasifikasi data lainnya [9, 10]. Mencapai akurasi yang terlihat dalam prediksi sangat penting karena dapat memberikan petunjuk untuk tindakan pencegahan yang sesuai.

Tujuan dari penelitian ini adalah untuk mengevaluasi penggunaan enam algoritma machine learning dalam melakukan klasifikasi terhadap kanker leukimia ALL dan AML berdasarkan data microarray ekspresi gen

METODE PENELITIAN

Pada penelitian ini, dataset yang digunakan adalah sebanyak 46 sampel leukimia, terdiri atasAcute Lymphoblastic Leukemia (ALL) sebanyak 32 sampel dan Acute Myeloid Leukemia (AML)sebanyak 14 sampel. Setiap sampel memiliki 7129 profil ekspresi gen.Data – data ini akan dilatih menggunakan algoritma jaringan syaraf tiruan, support vector machine, naive bayes, logistic regression, k-nearest negihbor, dan classification tree. Untuk mengukur kinerja dari metode – metode tersebut maka akan dievaluasi menggunakan ten fold validation dan metode leave-one-out



Gambar 1.Metodologi penelitian

HASIL DAN PEMBAHASAN

Pengujian terhadap kinerja algoritma yang digunakan adalah seperti pada tabel di bawah

Tabel 1. Confussion Matrix untuk ANN

		Predicted class			
		Tenfold		Leave-one-out	
		ALL	AML	ALL	AML
Actual class	ALL 32	32 (97.0%)	0 (0.0%)	32 (97.0%)	0 (0.0%)
	AML 14	1 (3.0%)	13 (100.0%)	1 (0.0%)	13 (100.0%)
Total	46	33	13	33	13

Tabel 2. Confussion Matrix untuk SVM

		Predicted class			
		Tenfold		Leave-one-out	
		ALL	AML	ALL	AML
Actual class	ALL 32	29 (96.7%)	3 (18.8%)	29 (96.7%)	3 (18.8%)
	AML 14	1 (3.3%)	13 (81.2%)	1 (3.3%)	13 (81.2%)
Total	46	30	16	30	16

Tabel 3. Confussion Matrix untuk Regresi

		Predicted class			
		Tenfold		Leave-one-out	
		ALL	AML	ALL	AML
Actual class	ALL 32	28 (96.6%)	4 (23.5%)	29 (96.7%)	3 (18.8%)
	AML 14	1 (3.4%)	13 (76.5%)	1 (3.3%)	13 (81.2%)
Total	46	29	17	30	16

Tabel 4. Confussion Matrix untuk Naive Bayes

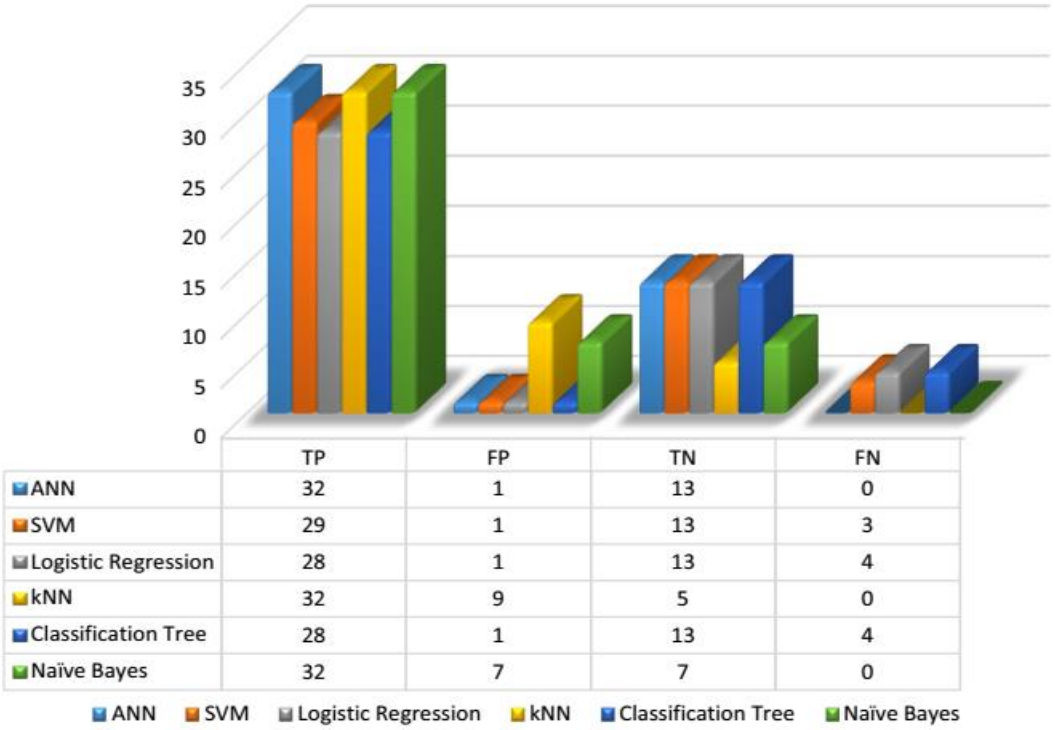
		Predicted class			
		Tenfold		Leave-one-out	
		ALL	AML	ALL	AML
Actual class	ALL 32	32 (82.1%)	0 (0.0%)	31 (91.2%)	1 (12.5%)
	AML 14	7 (17.9%)	7 (100.0%)	7 (8.8%)	7 (87.5%)
Total	46	39	7	34	8

Tabel 5. Confussion Matrix untuk Classification Tree

		Predicted class			
		Tenfold		Leave-one-out	
		ALL	AML	ALL	AML
Actual class	ALL 32	28 (96.6%)	4 (23.5%)	31 (91.2%)	1 (8.3%)
	AML 14	1 (3.4%)	13 (76.5%)	3 (8.8%)	11 (91.7%)
Total	46	29	17	34	12

Tabel 6. Confussion Matrix untuk K-NN

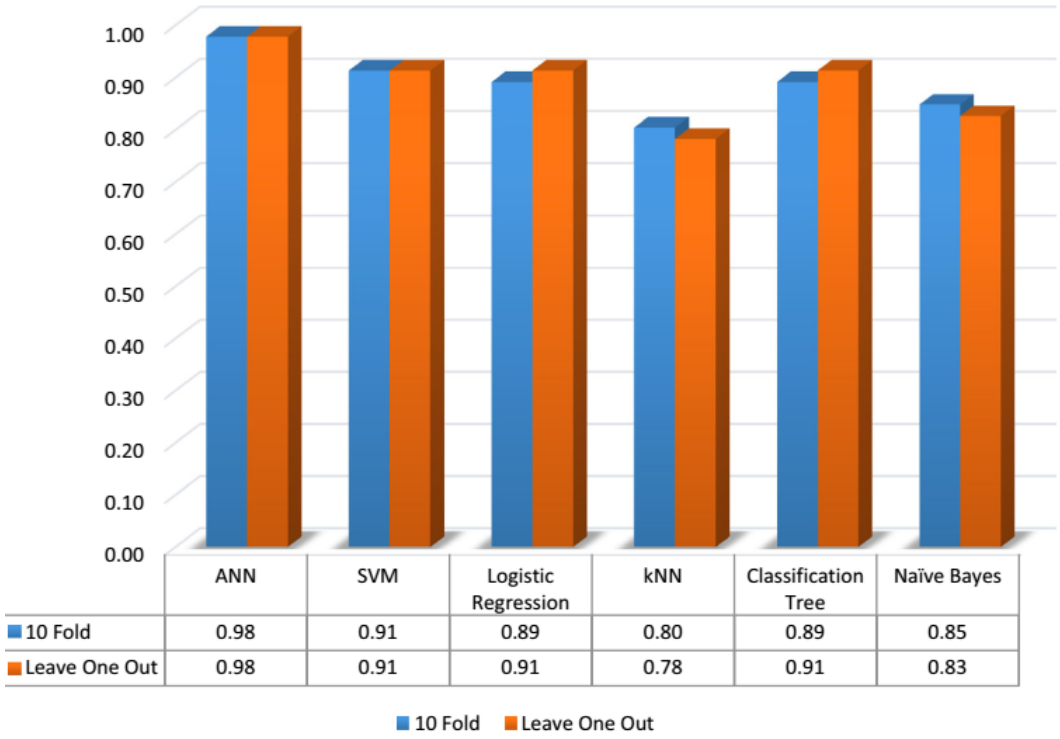
		Predicted class			
		Tenfold		Leave-one-out	
		ALL	AML	ALL	AML
Actual class	ALL 32	32 (78.0%)	0 (0.0%)	32 (76.2%)	0 (0.0%)
	AML 14	9 (22.0%)	5 (100.0%)	10 (23.8%)	4 (100.0%)
Total	46	31	5	42	4



	TP	FP	TN	FN
ANN	32	1	13	0
SVM	29	1	13	3
Logistic Regression	28	1	13	4
kNN	32	9	5	0
Classification Tree	28	1	13	4
Naïve Bayes	32	7	7	0

Gambar 2. Perbandingan hasil klasifikasi menggunakan TP,TN, FN dan FP

Merujuk pada Tabel 1 sampai Tabel 6 bahwa evaluasi yang digunakan yakni tenfold dan leave-one-out terlihat bahwa kehandalan algoritma jaringan syaraf tiruan lebih baik dibandingkan dengan algoritma yang lain dalam mengklasifikasi data ekspresi gen. Perbandingan hasil akurasi dari setiap algoritma yang digunakan seperti pada Gambar 3 di bawah



	ANN	SVM	Logistic Regression	kNN	Classification Tree	Naïve Bayes
10 Fold	0.98	0.91	0.89	0.80	0.89	0.85
Leave One Out	0.98	0.91	0.91	0.78	0.91	0.83

Gambar 3. Perbandingan Akurasi Hasil Klasifikasi

KESIMPULAN

Penelitian bertujuan untuk melakukan klasifikasi terhadap data kanker ALL dan AMLmenggunakan microarray ekspresi gen. Hasil penelitian ini menunjukkan bahwa Jaringan Syaraf Tiruan memiliki kinerja yang lebih baik dibandingkan dengan algoritma KNN, SVM, Regresi Logistik, Classification Tree dan Naive Bayes.

DAFTAR PUSTAKA

[1] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA (1999) Molecular classification of cancer: class discovery andclass prediction by gene expression monitoring. Science 286(5439):531–537

[2] Haferlach T, Kohlmann A, Schnittger S, Dugas M, Hiddemann W, Kern W, Schoch C (2005) Global approach to the diagnosis of leukemia using gene expression profiling. Blood 106(4):1189–1198

[3] Mallick BK, Ghosh D, Ghosh M (2005) Bayesian classification of tumours by using gene expression data. J R Stat SocSer B StatMethodol 67(2):219–234

[4] Antonov AV, Tetko IV, Mader MT, Budczies J, Mewes HW (2004) Optimization models for cancer classification: extracting gene interaction information from microarray expression data. Bioinformatics 20(5):644–652

[5] Lee Y, Lee C-K (2003) Classification of multiple cancer types by multicategory support vector machines using gene expression data. Bioinformatics 19(9):1132–1139

[6] Peng S, Xu Q, Ling XB, Peng X, Du W, Chen L (2003) Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. FEBS Lett 555(2):358–362

[7] Berrar DP, Downes CS, Dubitzky W (2003) Multiclass cancer classification using gene expression profiling and probabilistic neural networks. In: Proceedings of the Pacific symposium on biocomputing, pp 5–16

[8] Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C (2001) Classification and diagnostic prediction of cancers using gene

expression profiling and artificial neural networks. Nat Med 7(6):673–679

[9] Dwivedi AK, Chouhan U (2016) Comparative study of artificial neural network for classification of hot and cold recombination regions in Saccharomyces cerevisiae. Neural Comput Appl. doi:10.1007/s00521-016-2466-6

[10] Dwivedi AK, Chouhan U (2016) Comparative study of machine learning techniques for genome scale discrimination of recombinant HIV-1 strains. J Med Imaging Health Inform 6(2):425–430

[11] Dwivedi AK (2016) Performance evaluation of different machine learning techniques for prediction of heart disease. Neural ComputAppl 27(7):1–9

[12] Garcı ´a-Pedrajas N, Herva´s-Martı ´nez C, Ortiz-Boyer D (2005) Cooperative coevolution of artificial neural network ensembles for pattern classification. IEEE Trans EvolComput 9(3):271–302

[13] Yao X, Liu Y (1998) Making use of population information in evolutionary artificial neural networks. IEEE Trans Syst Man Cybern Part B Cybern 28(3):417–425

[14] Haykin S (2010) Neural networks: a comprehensive foundation, 1994. McMillan, New Jersey

[15] Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, New York

[16] Bahrammirzaee A (2010) A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. Neural ComputAppl 19(8):1165–1195

[17] Hoptroff RG (1993) The principles and practice of time series forecasting and business modelling using neural nets. Neural ComputAppl 1(1):59–66

[18] Azar AT (2013) Fast neural network learning algorithms for medical applications. Neural ComputAppl 23(3–4): 1019–1034

[19] Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, Haussler D (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl AcadSci 97(1):262–267

[20] Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16(10):906–914