

## KLASIFIKASI DATA KANKER BERDASARKAN MIKRO ARRAY GEN MENGGUNAKAN JARINGAN SYARAF TIRUAN

**Prisca Deviani Pakan**

Departemen Mikrobiologi, Fakultas Kedokteran, Universitas Nusa Cendana

Jl. Adi Sucipto, Penfui, Kupang

\* E-mail: [priscapakan@staf.undana.ac.id](mailto:priscapakan@staf.undana.ac.id)

### Abstrak

Klasifikasi kanker merupakan suatu topik penting dalam dunia kedokteran. Data mikro array merupakan suatu kumpulan data lengkap dengan ekspresi gen yang diambil dari beberapa pasien kanker. Penelitian ini bertujuan untuk melakukan klasifikasi kanker menggunakan ekspresi gen sebagai fitur yang dilatihkan pada jaringan syaraf tiruan. Penelitian ini menggunakan Elman Recurrent Neural Network sebagai mesin klasifikasi

**Kata kunci:** Klasifikasi Kanker, Jaringan Syaraf Tiruan, Ekspresi Gen, Elman Recurrent

### PENDAHULUAN

Kanker adalah salah satu penyakit yang menjadi penyebab utama kematian manusia diberbagai negara. Menurut WHO pada tahun 2012 terjadi 14 juta kasus baru dan 8,2 juta terkait kematian yang disebabkan oleh kanker, dan jumlah tersebut diperkirakan akan meningkan 70% selama dua dekade mendatang [1]. Karena itu, sangat diperlukan suatu teknologi yang mampu mendeteksi penyakit kanker lebih awal dengan analisis yang akurat, sehingga penyakit kanker dapat ditangani sejak dini

Pada beberapa tahun terakhir, suatu analisis dan diagnosis kanker dilakukan dengan menggunakan microarray data. DNA microarray, merupakan suatu teknologi yang digunakan untuk memonitor suatu ekspresi data (gen) dalam jumlah yang sangat besar dan berbeda secara bersamaan. Teknologi microarray telah digunakan untuk diagnosis medis dan analisis gen, terutama untuk memeriksa bagaimana pola ekspresi gen berubah dalam kondisi yang berbeda. Dengan analisis data ekspresi gen berupa microarray tersebut, memungkinkan suatu medis untuk dapat mengetahui apakah seseorang terkena suatu kanker atau tidak. Maka dari itu microarray sangat memiliki peran dalam diagnosis kanker.

Aplikasi microarray banyak digunakan dalam deteksi kanker, dimana sel kanker mengalami abnormalitas dalam meng-ekspresikan gennya. Pada dasarnya, DNA microarray akan mengamati ekspresi genetik dari berbagai gen dengan jumlah yang

mencapai ratusan bahkan ribuan secara bersamaan. Dalam analisis kanker, digunakan DNA normal dan DNA kanker yang kemudian akan di hibridisasi lalu akan dibaca dengan menggunakan laser (scanner). Klasifikasi data microarray merupakan salah satu kajian dalam bidang bioinformatika yang digunakan, salah satunya yaitu untuk analisis suatu kanker. Klasifikasi pada deteksi kanker, dapat dilakukan dengan mengelompokkan data ke dalam kelas yang sudah ditetapkan.

Microarray mampu menentukan ekspresi ribuan gen dan secara simultan memantau proses biologis yang sedang berlangsung [3]. Microarray merupakan bagian dari high dimensional data karena memiliki ratusan sampai dengan ribuan fitur [4]. Namun permasalahan yang terjadi adalah data microarray mempunyai jumlah variabel yang lebih besar dibandingkan dengan jumlah observasinya. Oleh karena itu, perlu dilakukan seleksi variabel dengan menggunakan metode fast correlation based filter untuk menyelesaikan permasalahan klasifikasi.

Adapun beberapa peneliti yang telah bekerja untuk melakukan klasifikasi microarray data yaitu, Moorthy [14] mengusulkan skema klasifikasi menggunakan algoritma. Random Forest pada microarray data. Data yang digunakan adalah data Adenocarcinoma, Brain, Breast2, Breast3, Colon, Leukemia, Lymphoma, NCI60, Prostate, SRBCT. Untuk klasifikasi menggunakan Random Forest mereka melakukan uji coba pada nilai OOB error rate dengan kondisi, yaitu (1) nilai OOB error rate terendah dan jumlah subset paling sedikit (2) nilai OOB error rate terendah dan jumlah subset paling banyak (3) proses standar

atau default algoritma Random Forest. Dari penelitiannya diperoleh akurasi klasifikasi standar atau default algoritma Random Forest untuk seleksi gen selalu lebih rendah dari OOB error rate terendah dan jumlah subset paling sedikit dan OOB error rate terendah dan jumlah subset paling banyak. Vanitha [15] melakukan klasifikasi menggunakan algoritma Support Vector Machine (SVM) dengan menggunakan 2 microarray data yaitu Colon cancer dan Lymphoma. Mutual information (MI) antara gen dan label kelas digunakan untuk mengidentifikasi gen yang informatif. Gen yang terpilih digunakan untuk learning menggunakan SVM dan evaluasi pada penelitian ini menggunakan metode Leave-one-Out Cross Validation (LOOCV). Metode yang diusulkan pada penelitian ini mampu mengurangi dimensi fitur dengan mengidentifikasi subset gen yang paling informatif serta memperbaiki akurasi klasifikasi. Tsun-Chen Lin, Ru-Sheng Liu, Ya-Ting Chao, and Shu-Yuan Chen [20] tujuan dari penelitian yang dilakukan adalah untuk mengeksplorasi penggunaan data ekspresi gen dalam membedakan jenis kanker heterogen. Mereka menggunakan hybrid learning methodology algoritma genetika (GA) dan jaringan syaraf tiruan (JST) untuk menemukan subkumpulan gen yang optimal untuk klasi\_kasi jaringan / kanker. Metode yang diusulkan diuji pada dua dataset microarray yaitu NCI60 dan GCM. Hasil dari penelitian yang dilakukan menunjukkan metode GA dan ANN tidak hanya unggul dalam pendekatan prediksi tetapi juga dapat mengurangi jumlah gen prediktif yang diperlukan dalam klasifikasi.

Campos [18] pada penelitiannya menerapkan algoritma bayesian network untuk proses klasifikasi menggunakan microarray data, dimana dataset yang digunakan adalah Breast-cancer, CNS, Colon-cancer, DLBCL-MIT, DLBCL-Stanford, Leukemia ALL-AML, Leukemia- MLL, Lung-cancer, Prostate-tumor. Pada penelitiannya, ekspresi gen dengan 3 cara, yaitu (1) klasifikasi Bayesian yang diinduksi dari data microarray. (2) skema preprocessing untuk data ekspresi gen digunakan menginduksi pengklasifikasi Bayesian, dan (3) evaluasi Bayesian yang berbeda untuk jenis data. Lavanya [18] membangun kerangka komprehensif yang menggabungkan fitur seleksi dan teknik klasifikasi. Teknik klasifikasi yang digunakan adalah algoritma Naive Bayes. Sedangkan untuk fitur seleksi menggunakan pengukuran T-Test, Chi-Square Test yang digunakan untuk memilih kemungkinan gen yang berhubungan dengan kanker dari data ekspresi gen microarray yang besar. Dari kerangka klasifikasi yang dibangun berhasil melakukan

klasifikasi sampel baru seperti yang terinfeksi atau normal. Deegalla [19] melakukan penelitian klasifikasi pada microarray data menggunakan algoritma KNN.

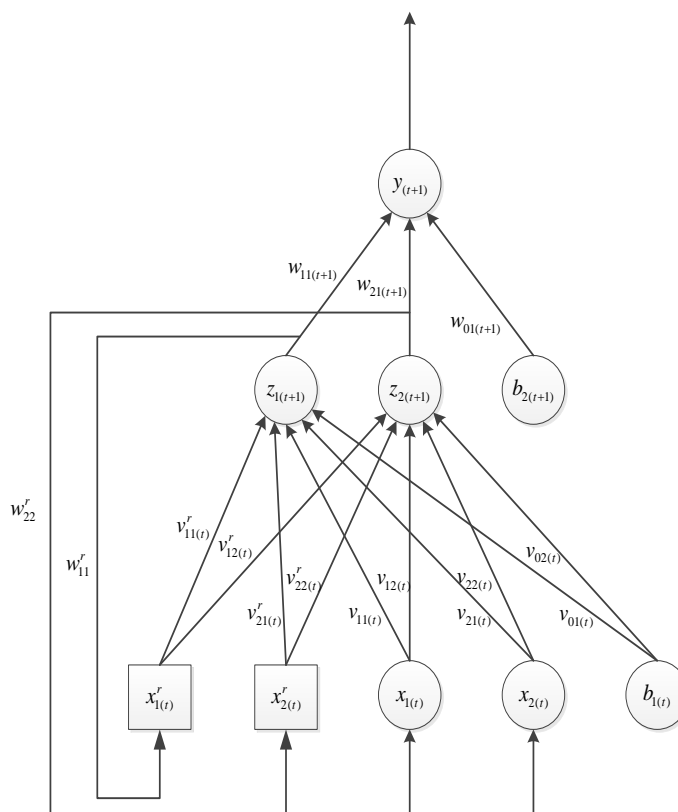
Pada penelitian ini dilakukan klasifikasi terhadap data ekspresi gen menggunakan Algoritma Jaringan Syaraf Tiruan Elman Recurrent

## METODE PENELITIAN

### Data

Jumlah data yang digunakan dalam penelitian ini sebanyak 174 sampel dengan 12533 gen. Setiap kelas dalam sampel menyatakan : 0 Ovary, 1 Bladder/Ureter, 2 kanker payudara, 3 Colorectal, 4 Gastroesophagus, 5 kanker ginjal, 6 Liver, 7 Prostate, 8 Pankreas, 9 Lung Adeno, 10 Lung Squamous

Arsitektur JST penelitian ini seperti pada gambar di bawah



Gambar 1. Arsitektur JST ERNN

**HASIL DAN PEMBAHASAN**

Tabel 1. Parameter Pengujian JST

| Variables     | Parameters         |
|---------------|--------------------|
| input         | 174                |
| hidden        | (50,50,50)         |
| output        | 11                 |
| Loss function | Mean Squared Error |
| Epochs        | 1000000            |
| 11            | 0.00001            |
| 12            | 0.00001            |
| nfolds        | 10                 |

Dengan menggunakan parameter – parameter tersebut di atas maka hasil pelatihan JST menunjukkan bahwa sistem mampu mencapai MSE: 1.38526e-11 sementara pada saat validasi, sistem mampu mencapai MSE: 0.15.

Hasil pengujian tersebut menunjukkan bahwa arsitektur JST ERNN yang digunakan cukup handal dalam melakukan klasifikasi data kanker berdasarkan mikro array

**PENUTUP**

**Kesimpulan**

Merujuk pada tujuan penelitian, hasil yang diperoleh selama pengujian maka dapat disimpulkan bahwa sistem dapat melakukan klasifikasi terhadap 10 kelas kanker dengan MSE pelatihan 1.38526e-11 dan MSE pengujian/validasi sebesar 0.15. Hal ini menunjukkan bahwa algoritma yang diusulkan dapat digunakan sebagai mesin klasifikasi

**Saran**

Beberapa hal yang dapat dilakukan untuk perbaikan penelitian ini antara lain:

- 1) Penelitian ini dapat diujicobakan pada dataset microarray lainnya, sehingga dapat diamati hasil yang didapat pada berbagai dataset microarray yang lain.
- 2) Perbaikan dapat dilakukan pada proses ekstraksi fitur dengan mencoba wavelet filter lainnya atau melakukan pemrosesan lebih lanjut pada nilai koefisien hasil dekomposisi sebelum dijadikan fitur masukan classifier, agar didapat hasil klasifikasi yang lebih baik.

**DAFTAR PUSTAKA**

[1] World Health Organization, Cancer fact sheet, <http://www.who.int/mediacentre/factsheets/fs297/en/index.html>, ccessed at March 2nd 2016.

[2] Vanitha, C.D.A., Devaraj, D. and Venkatesulu, M., 2015. Gene Expression Data Classification Using Support vector machines and Mutual Information-based Gene Selection. *Procedia Computer Science*, 47, pp.13-21.

[3] Siang, T.C., Soon, T.W., Kasim, S., Mohamad, M.S., Howe, C.W., Deris, S., Zakaria, Z., Shah, Z.A. and Ibrahim, Z., 2015. A Review of Cancer Classification Software for Gene Expression Data. *International Journal of Bio-Science and Bio-Technology*, 7(4), pp.89-108.

[4] KR, S., 2011. Microarray Data Classification Using Support Vector Machine. *International Journal of Biometrics and Bioinformatics (IJBB)*, 5(1), p.10.

[5] Nurfalah A.,Adiwijaya, Suryani A.A ,“Cancer Detection Based On Microarray data Classification Using PCA And Modified Backpropogation”, *Far East Journal of Electronics and Communications*.2015

[6] Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M. and Haussler, D., 2000. Support vector machines Classification And Validation Of Cancer Tissue Samples Using Microarray Expression Data. *Bioinformatics*,16(10), pp.906-914.

[7] Purbolaksono, M.D., Widiastuti, K.C., Mubarak, M.S. and Ma’ruf, F.A., 2018, March. Implementation of mutual information and bayes theorem for classification microarray data. In *Journal of Physics: Conference Series* (Vol. 971, No. 1, p. 012011). IOP Publishing.

[8] Aydadenta, H., 2018, March. On the classification techniques in data mining for microarray data classification. In *Journal of Physics: Conference Series* (Vol. 971, No. 1, p. 012004). IOP Publishing.

[9] Nhita, F. and Adiwijaya, U.N., 2014. Forecasting Indonesian Weather through Evolving Neural Network (ENN) based on Genetic Algorithm. In *The Second International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE2014)* (pp. 78-82). The Society of Digital Information and Wireless Communication.

[10] Jinyan Li, Kent-ridge bio-medical data set repository, School of Computer Engineering Nanyang Technological University, Singapore, Downloaded at January 2013 from URL:<http://levis.tongji.edu.cn/gzli/data/mirror-kentridge.html>

[11] Adiwijaya, Wisesty, U.N. and Nhita, F., 2014. Studi of line search techniques on the modified backpropagation for forecasting of weather data in Indonesia. *Far East Journal*

- of Mathematical Sciences, 86(2), p.139.
- [12] Bai Anita and S. K. Rath. Classification and Clustering using Intelligence Techniques: Application to microarray cancer data. Master Thesis of Department of Computer Science and Engineering National Institute of Technology Rourkela, India, 2013.
- [13] Hieu Trung Huynh, Jung-Ja Kim, Yonggwan Won., 2009. Classification Study on DNA Microarray with Feed Forward Neural Network Trained by Singular Value Decomposition. *International Journal of Bio-Science and Bio Technology*, 1(1).
- [14] Moorthy, K. and Mohamad, M.S., 2012. Random forest for gene selection and microarray data classification. In *Knowledge Technology* (pp. 174-183). Springer, Berlin, Heidelberg
- [15] Vanitha, C.D.A., Devaraj, D. and Venkatesulu, M., 2015. Gene expression data classification using support vector machine and mutual information-based gene selection. *procedia computer science*, 47, pp.13-21.
- [16] Li, T., Zhang, C. and Ogihara, M., 2004. A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, 20(15), pp.2429-2437.
- [17] De Campos, L.M., Cano, A., Castellano, J.G. and Moral, S., 2011, November. Bayesian networks classifiers for gene-expression data. In *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*(pp. 1200-1206). IEEE.
- [18] Lavanya, C., Nandihini, M., Niranjana, R. and Gunavathi, C., 2014. Classification of microarray data based on feature selection method. *International Journal of Innovative Research in Science, Engineering and Technology*, 3(1).
- [19] Deegalla, S. and Boström, H., 2007, December. Classification of microarrays with knn: Comparison of dimensionality reduction methods. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 800-809). Springer, Berlin, Heidelberg
- [20] Lin, T.C., Liu, R.S., Chao, Y.T. and Chen, S.Y., 2006, August. Multiclass microarray data classification using GA/ANN method. In *Pacific Rim International Conference on Artificial Intelligence* (pp. 1037-1041). Springer, Berlin, Heidelberg.
- [21] Husna Aydadenta, Adiwijaya, (2018), A Clustering Approach for Feature Selection in Microarray Data Classification using Random Forest, *Journal of Information Processing System* 14(5).
- [22] Adiwijaya, U. N. Wisesty, E. Lisnawati, A. Aditsania, D. S. Kusumo, (2018). Dimensionality Reduction using Principal Component Analysis for Cancer Detection based on Microarray Data Classification, *Journal of Computer Science* 14(10).