

Analisis Pengaruh Metode Reduksi Dimensi *Minimum Redundancy Maximum Relevance* pada Klasifikasi Kanker Berdasarkan Data *Microarray* Menggunakan *Classifier Support Vector Machine*

Analysis of The Influence of Minimum Redundancy Maximum Relevance as Dimensionality Reduction Method on Cancer Based on Microarray Data Using Support Vector Machine Classifier

Firda A. Ma'ruf¹, Adiwijaya², Untari. N. Wisesty³

²Prodi Ilmu Komputasi, Fakultas Teknik Informatika, Universitas Telkom

^{1,4}Prodi Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

¹firdamaruf@student.telkomuniversity.ac.id,²adiwijaya@telkomuniversity.ac.id,

³untarinw@telkomuniversity.ac.id

Abstrak

Kanker adalah salah satu penyebab utama kematian di seluruh dunia dengan total 8,8 juta orang meninggal akibat kanker di tahun 2015 [1]. Teknologi DNA *microarray* yang dikembangkan oleh Patrick O. Brown, Joseph DeRisi, dan David Botstein memungkinkan untuk mengamati beribu-ribu ekspresi gen dalam waktu bersamaan. Namun, permasalahan yang dihadapi terhadap penggunaan teknologi *microarray* ini adalah karakteristik data *microarray* yang memiliki dimensi yang besar. Dimensi yang besar ini akan berpengaruh terhadap performansi klasifikasi. Untuk menyelesaikan masalah tersebut, penelitian ini mengusulkan penggunaan *Minimum Redundancy Maximum Relevance* (MRMR) sebagai metode reduksi dimensi dan *Support Vector Machine* (SVM) sebagai *classifier*. *Principal Component Analysis* (PCA) juga digunakan sebagai perbandingan. MRMR adalah metode reduksi dimensi yang menyeleksi fitur berdasarkan nilai relevansi yang tinggi dan nilai redundansi yang rendah. Sedangkan PCA adalah metode yang mengekstraksi fitur berdasarkan *principal component* yang diperoleh dari perhitungan nilai *eigen* dan vektor *eigen*. Data yang digunakan pada penelitian ini adalah data kanker usus besar, data kanker paru-paru, data kanker ovarium, dan data kanker leukemia. Hasil pengujian pada data kanker paru-paru membuktikan bahwa sistem yang diusulkan dapat memberikan *F1-score* sebesar 1 dari penggunaan metode reduksi dimensi MRMR dan *classifier* SVM dengan fungsi *kernel linear* dan *kernel polynomia*, dengan jumlah fitur yang digunakan untuk klasifikasi adalah sebesar 10% dari jumlah fitur asli. Hal ini berarti akurasi dari klasifikasi adalah 100% dan performansi sistem yang dibangun sangatlah bagus.

Kata kunci: data *microarray*, kanker, klasifikasi, SVM, PCA, MRMR

Abstract

ABSTRACT

Cancer is one of the main cause of death in the world with the total of 8,8 million cases of death cause by cancer in 2015 [1]. DNA *microarray* technology that was developed by Patrick O. Brown, Joseph DeRisi, and David Botstein provides a way to monitor thousands of gen expressions at the same time. But the problem with *microarray* technology is that *microarray* data has a high dimensionality. This high dimensionality will effect to the classification performance. In order to solve this issue, this research proposed the use of Minimum Redundancy Maximum Relevance (MRMR) as dimension reduction methods and Support Vector Machine (SVM) as *classifier*. Principal Component Analysis (PCA) method is also used as a comparison to MRMR. MRMR is a dimension reduction method that selects features based on its high relevance values and low redundancy values. While PCA is a method that extracts features based on the principal component obtained from the calculation of eigenvalues and eigenvectors. Data used in this research are colon cancer data, lung cancer data, ovarian cancer data, and leukemia cancer data. Test results on lung cancer data prove that the proposed system is able to provide an *F1-score* of 1 from the use of MRMR and SVM *classifier* with linear *kernel* function and polynomial *kernel* function, with the number of features used for the classification is 10% of the original features total number. This means the accuracy of classification is 100% and the performance of the system built is very good.

Keywords: microarray data, cancer, classification, SVM, PCA, MRMR.

1. Pendahuluan

Kanker adalah sebuah penyakit yang disebabkan oleh pembelahan secara berlebihan dan tak terkendali dari sel-sel dalam tubuh. Sebanyak 8,8 juta orang meninggal akibat kanker di tahun 2015.[1] Akan tetapi kematian akibat kanker bisa dihindari sebanyak 30-50%, dengan cara mengidentifikasi jenis kanker yang menyerang pasien sedari dini. Namun, merupakan tantangan yang sulit bagi para peneliti untuk mendiagnosa kanker berdasarkan struktur morfologisnya karena perbedaan morfologis yang sangat tipis untuk jenis kanker yang berbeda.

Kesulitan di atas mendorong penggunaan algoritma penambangan data, secara spesifik penggunaan ekspresi gen untuk menentukan jenis sel kanker. Tingkat ekspresi gen dapat menunjukkan keaktifan sebuah gen di dalam sel tubuh berdasarkan jumlah mRNA di dalam sel tubuh tersebut. Tingkat ekspresi gen dalam tubuh manusia dapat diukur melalui eksperimen *microarray* DNA. Teknologi *microarray* DNA memiliki kemampuan untuk menentukan level dari beribu-ribu gen secara bersamaan di dalam sebuah eksperimen [2].

Begitu banyak pendekatan yang sudah diimplementasikan dalam melakukan klasifikasi ekspresi gen *microarray*, salah satu yang sering digunakan selama 10 tahun terakhir adalah *Support Vector Machine* (SVM). SVM dipercaya merupakan salah satu algoritma *supervised learning* yang terbaik dalam hal performansi. Furey et. al. membuktikan bahwa SVM memberikan akurasi performansi sebesar 100% dalam mengklasifikasi data kanker usus besar [8].

Karakteristik dari data *microarray* adalah jumlah gen (p) yang sangat besar, dapat mencapai puluhan ribuan (dimensi yang besar), sedangkan jumlah sampelnya (n) yang sedikit. Maka dari itu diperlukan seleksi fitur atau ekstraksi fitur pada data *microarray* untuk menghindari *overfitting* pada *classifier*. Salah satu seleksi fitur yang mampu meningkatkan kinerja *classifier* adalah *Minimum Redundancy Maximum Relevance* (MRMR), yang sudah dibuktikan oleh Ding et al. [14] dapat memberikan akurasi performansi sebesar 100% untuk data Leukemia. Sedangkan ekstraksi fitur yang sering digunakan adalah *Principal Component Analysis* (PCA), yang dapat memberikan akurasi performansi rata-rata sebesar 80% [9][17].

2. Studi Literatur

Pengembangan DNA *microarray* oleh Patrick O. Brown, Joseph DeRisi, dan David Botstein di pertengahan tahun 1990an memungkinkan untuk mengamati beribu-ribu ekspresi gen dalam waktu bersamaan. Berbagai macam metode digunakan untuk mengklasifikasi data *microarray*, salah satu yang sering digunakan adalah *Support Vector Machine* (SVM) Furey et al. membuktikan bahwa SVM dengan *kernel* yang digunakannya, berhasil mengklasifikasikan semua kanker usus besar secara akurat [8]. *Kernel* sendiri adalah sebuah metode yang memiliki fungsi untuk menganalisa pola data. Metode *kernel* memungkinkan SVM untuk bekerja di dimensi yang tinggi.

Karakteristik dari data *microarray* adalah data ini memiliki dimensi yang tinggi dan jumlah gen yang jauh lebih banyak dibandingkan jumlah sampelnya. Dimensi yang tinggi ini bagai pisau dengan 2 sisi, dimana dimensi yang tinggi ini kemungkinan besar akan bisa dipisahkan dengan garis linear (*linearly separable*), namun di sisi yang lain dimensi yang tinggi akan membuat analisis data menjadi sebuah tantangan [13]. Permasalahan ini bisa dipecahkan dengan bantuan dari metode reduksi dimensi. Hal ini dilakukan untuk mengurangi kompleksitas dari data dan menurunkan error pada saat klasifikasi dengan metode *classifier* yang dipilih. MRMR mungkin tidak termasuk salah satu metode seleksi fitur yang populer untuk mereduksi dimensi data *microarray*, namun penelitian yang dilakukan oleh Chris Deng dan Hanchuan Peng [14] membuktikan bahwa model yang dibangun dengan *classifier* SVM dan MRMR memberikan hasil presentasi error yang kecil untuk klasifikasi data kanker kontinu dibandingkan dengan metode *baseline* tanpa metode reduksi dimensi.

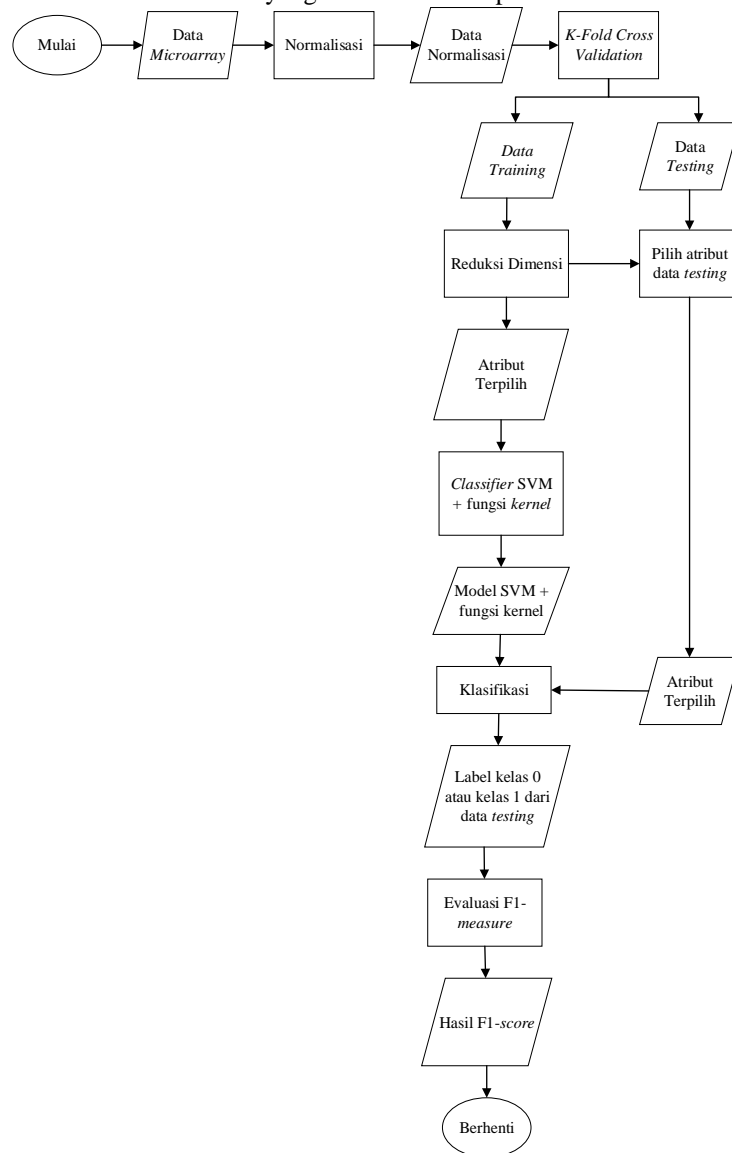
Mundra et al. [10] menggunakan metode mRMR dan SVM-RFE (Recursive Feature Elimination) sebagai *feature selection* dan *classifier*. Untuk akurasi performansi, SVM-RFE + mRMR memberikan akurasi sebesar 91,68% dalam mengklasifikasikan data kanker usus besar, 98,35% dalam mengklasifikasikan data leukemia, 88,12% dalam mengklasifikasikan kanker hepato (ginjal), dan 98.29% dalam mengklasifikasikan kanker prostat. Perbandingan yang dilakukan menunjukkan bahwa SVM-RFE + mRMR mengungguli metode-metode seperti Bayes + KNN, Bayes + SVM, dan LS-Bound + SVM.

3. Data dan Perancangan Sistem

Proses yang dilakukan dalam sistem ini terbagi menjadi 4 proses utama, yaitu:

1. *Preprocessing* data *microarray* dengan menormalisasikan data menjadi nilai *range* 0-1 menggunakan metode min-maks.
2. Reduksi dimensi dari data *microarray* yang dilakukan oleh metode seleksi fitur MRMR dan ekstraksi PCA.
3. Klasifikasi data *microarray* ke dalam dua kelas kanker.

4. Hitung performansi dari klasifikasi menggunakan model yang dibangun dengan *F1-score*. Berikut ini adalah skema umum dari sistem yang dibuat di dalam penelitian ini:



Gambar 1 Skenario Model Klasifikasi Dengan Metode Reduksi Dimensi PCA atau MRMR dan Classifier SVM + kernel

Data *microarray* merupakan data yang berukuran sangat besar, maka dari itu diperlukan *preprocessing* terhadap data *input*. Data perlu dinormalisasi ke dalam bentuk *range* 0 – 1 agar dapat mudah diproses, karena data yang ada sekarang ini memiliki perbedaan nilai yang cukup jauh. Digunakan metode normalisasi min-maks untuk mengubah data, yang formulanya adalah sebagai berikut:

$$Xn_i = \left(\frac{X_i - \min(X)}{\max(X) - \min(X)} \right) \quad (1)$$

Dimana:

Xn_i = data aktual normalisasi ke- i

X_i = data aktual dengan *range* data asli ke- i

X = data actual dengan *range* data asli

3.1. Data

Data *microarray* yang digunakan untuk klasifikasi kanker pada penelitian ini adalah data kanker usus besar, kanker paru-paru, kanker ovarium, dan leukemia yang didapatkan dari Kent Ridge Biomedical Data Repository [15]. Spesifikasi dari data tersebut bisa dilihat pada Tabel.1.

Tabel 1 Spesifikasi Beberapa Data

Data	Jumlah Kelas	Sample	Feature
Kanker Usus Besar	2	62 (22 Positif, 40 Negatif)	2.000
Kanker Paru-Paru	2	181 (31 Mesothelioma, 150 ADCA)	12.533
Kanker Ovarium	2	253 (91 Normal, 162 Cancer)	15.154
Leukemia	2	38(27 ALL, 28 AML)	7.129

Kelas dari setiap data kanker sendiri diubah ke dalam bilangan biner untuk memudahkan proses klasifikasi. Berikut adalah keterangan kelas dari tiap data kanker:

Tabel 2 Tabel Perubahan Jenis Kelas Kanker ke Kelas Biner

Jenis Kanker	Jenis Kelas	Kelas Biner
Kanker Usus Besar	Negatif	1
	Positif	0
Kanker Paru-Paru	<i>Mesothelioma</i>	1
	ADCA	0
Kanker Ovarium	Normal	1
	Kanker	0
Kanker Leukemia	ALL	1
	AML	0

3.2. Preprocessing

Preprocessing merupakan proses yang dilakukan untuk membuat suatu data menjadi lebih mudah digunakan. Yang dilakukan dalam tahap *preprocessing* ini yaitu mengubah skala nilai atau *range* nilai data (normalisasi data) dengan *range* 0 sampai 1.

3.3. Reduksi Dimensi

A. Minimum Redundancy Maximum Relevance (MRMR)

Untuk metode MRMR yang digunakan di dalam penelitian ini adalah MRMR untuk variabel kontinu karena data yang dimiliki adalah data *microarray* kontinu. Berikut adalah fungsi seleksi fitur yang digunakan:

1. FCD (*F-test Correlation Difference*): mengkombinasikan *F-test* dengan korelasi kemudian dikurangkan.

$$\max_{i \in \Omega_S} \left[F(i, h) - \frac{1}{|S|} \sum_{j \in S} |c(i, j)| \right] \quad (1)$$

2. FCQ (*F-test Correlation Quotient*): mengkombinasikan *F-test* dengan korelasi menggunakan hasil bagi.

$$\max_{i \in \Omega_S} \left[F(i, h) / \frac{1}{|S|} \sum_{j \in S} |c(i, j)| \right] \quad (2)$$

B. Principal Component Analysis (PCA)

Sedangkan langkah-langkah pengerjaan PCA dijelaskan dalam 6 langkah berikut ini:

1. Kumpulkan sampel n dari m -dimensi data $\vec{x}_1 \dots \vec{x}_n$ vektor dalam \mathbb{R}^m .
2. Lakukan pengurangan *mean* (rata-rata) dari setiap dimensi data dengan rumus berikut:

$$\vec{\mu} = \frac{1}{n} (\vec{x}_1 + \dots + \vec{x}_n) \quad (3)$$

Dimana: n = jumlah sampel atau jumlah data observasi.

3. Bentuk matriks data (B) dan matriks *covariance* (S)

B adalah matriks berukuran $m \times n$, dengan kolom i -nya $\vec{x}_i - \vec{\mu}$. Untuk membentuk matriks data, gunakan formula:

$$B = |\vec{x}_1 - \vec{\mu}| \dots |\vec{x}_n - \vec{\mu}| \quad (4)$$

Untuk membentuk matriks *covariance* (yang akan berukuran $m \times m$), gunakan formula:

$$S = \frac{1}{n-1} B B^T \quad (5)$$

Dimana: n = jumlah sampel atau jumlah data observasi.

4. Hitung *eigenvectors* $\vec{\mu}_1 \dots \vec{\mu}_m$ dan *eigenvalues* $\vec{\lambda}_1 \dots \vec{\lambda}_m$ dari matriks *covariance* S .
5. Urutkan *eigenvectors* berdasarkan nilai *eigenvalues* dari yang terbesar hingga yang terendah dan pilih k *eigenvectors* dengan *eigenvalues* yang terbesar untuk membentuk matriks W dengan dimensi $m \times k$ (dimana setiap kolom merepresentasikan *eigenvector*).
6. Bentuk data set baru
Data set yang baru ini didapatkan dari:

$$Y = V_{row} * B \quad (6)$$

Dimana V_{row} didapat dari *Feature Vector*^T dan Y adalah set data final.

Untuk menentukan berapa banyak vektor *eigen* yang dipilih, penelitian ini menggunakan proporsi kumulatif varians (nilai *eigen*) terhadap total varians (nilai *eigen*). Proporsi varians menunjukkan besarnya presentase informasi variable-variabel asal yang terdapat di dalam setiap vektor ciri (vektor *eigen*) berdasarkan nilai *eigen* dan memberikan interpretasi mengenai seberapa besar data dapat terwakili dalam dimensi yang telah direduksi [20]. Proporsi varians untuk setiap komponen utama (vektor *eigen*) dapat dihitung menggunakan rumus berikut ini:

$$\left(\frac{\lambda_i}{\sum \lambda_i} \times 100\% \right) + \left(\frac{\lambda_{i-1}}{\sum \lambda_i} \times 100\% \right) \quad (7)$$

Dimana λ_i = nilai *eigen* dan λ_{i-1} = nilai *eigen* sebelumnya.

Vektor *eigen* yang terpilih berdasarkan *threshold* (kriteria pemilihan vektor *eigen* atau komponen utama) terhadap hasil PPV yang sudah ditentukan terlebih dahulu.

$$\text{Pemilihan vektor } eigen = \frac{\lambda_i}{\sum \lambda_i} \times 100\% > threshold \quad (8)$$

3.4. Klasifikasi

Setelah mengurangi kompleksitas dimensi, selanjutnya melakukan proses klasifikasi. Klasifikasi merupakan tujuan dari penelitian ini, dimana pada tahap ini akan dilakukan diagnosis (pengklasifikasian) apakah seseorang terkena suatu kanker atau tidak kanker atau jenis kanker yang kemungkinan diidap oleh orang tersebut. Metode yang digunakan sebagai *classifier* adalah *Support Vector Machine* (SVM).

SVM adalah salah satu *supervised learning* yang bertujuan untuk mencari *hyperplane* paling optimal dalam memisahkan kelas-kelas di ruang *input*. Sejatinnya SVM digunakan untuk mengklasifikasikan data dengan 2 kelas dan *linearly separable*. Namun nyatanya masalah yang ada kebanyakan berupa *non-linear*. Untuk mengatasi hal ini, digunakan *kernel trick* pada data *training* yang berfungsi untuk memetakan setiap data pada ruang *input* menjadi ruang vektor baru dengan dimensi yang lebih tinggi. *Kernel* akan merubah perhitungan *dot product* \vec{x} pada setiap persamaan dengan menambahkan fungsi-fungsi di bawah ini. Jadi *kernel* akan mentransformasi dari *input space* ke dalam *feature space*.

Berikut adalah jenis-jenis *kernel* yang digunakan dalam model ini:

$K(\vec{x}_i, \vec{x}_d)$ = kernel, dimana jenis kernel yang digunakan, diantaranya:

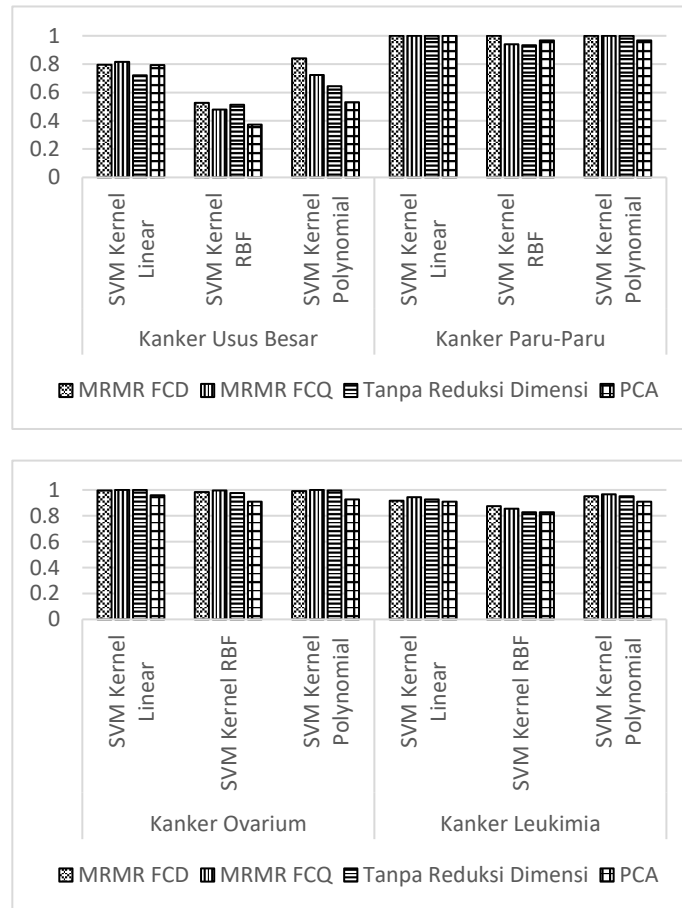
$$\text{Linear} \quad : K(\vec{x}_i, \vec{x}_d) = (\vec{x}_i, \vec{x}_d) \quad (9)$$

$$\text{Radial Basis Function (RBF)} \quad : K(\vec{x}_i, \vec{x}_d) = \exp\left(\frac{-|\vec{x}_i - \vec{x}_d|^2}{2\sigma^2}\right) \quad (10)$$

$$\text{Polinomial} \quad : K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \vec{x}_j + 1)^P$$

4. Pengujian dan Analisis

Implementasi dilakukan pada 4 jenis data kanker dengan menerapkan metode MRMR FCD + SVM (*kernel linear*, RBF, dan *polynomial*), MRMR FCQ + SVM (*kernel linear*, RBF, dan *polynomial*), PCA+ SVM (*kernel linear*, RBF, dan *polynomial*), dan metode SVM SVM (*kernel linear*, RBF, dan *polynomial*) tanpa reduksi dimensi pada aplikasi Matlab. Hasil akhir terbaik dari semua klasifikasi data kanker menggunakan *classifier* SVM dengan 3 fungsi *kernel* yang berbeda dapat dilihat pada grafik di bawah ini:



Gambar 2 Grafik Hasil Akurasi Terbaik Dari Semua Skenario Pengujian Dengan Classifier SVM + fungsi kernel (*linear*, RBF, *polynomial*).

(atas: hasil akurasi terbaik dari semua skenario pengujian dengan data kanker usus besar dan kanker paru-paru. bawah: hasil akurasi terbaik dari semua skenario pengujian dengan data kanker ovarium dan kanker leukemia)

Grafik Gambar 2 menyajikan F1-score terbaik dari skenario pengujian klasifikasi data kanker tanpa metode reduksi dimensi dan dengan metode reduksi dimensi (MRMR FCD, MRMR FCQ, dan PCA), menggunakan *classifier* SVM dengan 3 fungsi *kernel* yang berbeda, yaitu *linear*, RBF, dan *polynomial*. Klasifikasi data kanker usus besar, data kanker paru-paru, dan data kanker ovarium menggunakan *classifier* SVM *kernel linear* tanpa reduksi dimensi atau dengan metode reduksi dimensi (MRMR FCD, MRMR FCQ, dan PCA) rata-rata menghasilkan F1-score yang lebih besar dibandingkan hasil F1-score klasifikasi data tersebut dengan *classifier* SVM *kernel RBF* dan SVM *kernel polynomial*. Hal tersebut dikarenakan data awal dari ketiga data kanker tersebut masih dapat dipisahkan dengan mudah oleh *hyperplane linear* meskipun dimensinya yang besar.

Klasifikasi dengan SVM *kernel polynomial* menghasilkan F1-score yang tinggi pada klasifikasi data kanker usus besar tanpa metode reduksi dimensi yaitu 0,9947, pada klasifikasi data kanker paru-paru dengan metode MRMR FCD dan MRMR FCQ yang menghasilkan F1-score sebesar 1. SVM *kernel polynomial* juga mengungguli SVM *kernel linear* dan SVM *kernel RBF* pada hasil klasifikasi data kanker leukemia, dimana F1-score tertinggi adalah 0,9657, yang diperoleh dari penggunaan metode reduksi dimensi MRMR FCQ dengan *classifier* SVM *kernel polynomial*. Hasil ini membuktikan bahwa data kanker usus besar dan data kanker leukemia setelah di reduksi dimensi, variansi datanya akan berubah dan membutuhkan penggunaan fungsi *kernel non-linear* seperti

kernel polynomial dalam menentukan *hyperplane non-linear* optimal yang dapat mengklasifikasi data dengan dimensi yang cukup tinggi.

F1-score tertinggi untuk klasifikasi hanya dengan *classifier SVM* yaitu 1, yang diperoleh dari klasifikasi data kanker usus besar, data kanker paru-paru, dan data kanker ovarium, dengan *SVM kernel linear*. F1-score tertinggi untuk klasifikasi dengan metode reduksi dimensi MRMR FCD dan *classifier SVM* adalah 1, nilai ini didapatkan dari klasifikasi data kanker paru-paru dengan *classifier SVM kernel linear*, *classifier SVM kernel RBF*, dan *classifier SVM kernel polynomial*. F1-score tertinggi untuk klasifikasi dengan metode reduksi dimensi MRMR FCQ dan *classifier SVM* adalah 1, hasil ini didapatkan dari klasifikasi data kanker paru-paru dan data kanker ovarium dengan *classifier SVM kernel linear* dan *classifier SVM kernel polynomial*. Sedangkan F1-score tertinggi untuk klasifikasi dengan metode reduksi dimensi PCA dan *classifier SVM* adalah 1, yang diperoleh dari klasifikasi data kanker paru-paru menggunakan *classifier SVM kernel linear*.

5. Kesimpulan

Kesimpulan yang diperoleh dari proses dan hasil penelitian adalah:

1. Pemilihan fitur oleh metode MRMR, jumlah vektor *eigen* pada proses reduksi dimensi oleh metode PCA, serta pemilihan fungsi *kernel* pada SVM, amat sangat berpengaruh terhadap F1-score yang dihasilkan dan lamanya waktu komputasi pada proses klasifikasi data kanker pada setiap metode (SVM, MRMR + SVM, PCA + SVM).
2. Dari hasil akhir klasifikasi (F1-score) dengan metode reduksi dimensi MRMR + SVM, nilainya sama bagusnya atau bahkan lebih bagus dari hasil akhir klasifikasi data kanker tanpa metode reduksi dimensi. Salah satu buktinya adalah pada klasifikasi kanker usus besar, dimana MRMR FCQ + SVM *kernel linear* menghasilkan F1-score sebesar 0.81667 dengan pemilihan fitur oleh MRMR FCQ sebanyak 50% dari total 2000 fitur, dimana hasil klasifikasi tanpa metode reduksi dimensi dengan *classifier SVM kernel linear* yang perolehan F1-scorenya adalah 0.72. Hal ini membuktikan bahwa fitur-fitur yang terpilih ini meskipun sedikit tapi memiliki sifat generalisasi yang lebih baik dan secara efektif mencakup karakteristik dari kelas.
3. Semakin bertambah banyak fitur yang dipilih oleh metode MRMR untuk digunakan pada proses klasifikasi dengan SVM, tidak akan selalu meningkatkan performansi dan F1-score yang diperoleh dari klasifikasi data kanker.

References

- [1] World Health Organization, Cancer Fact Sheet, <http://www.who.int/mediacentre/factsheets/fs297/en/>, diakses pada tanggal 11 Februari 2017.
- [2] L.-J. Zhang, Z.-J. Li and X. Hu.2007. *A Hybrid Gene Selection Method for Cancer Classification*.
- [3] Li, L., Weinberg, C., Darden, T., and Pedersen, L. (2001). *Gene Selection for Sample Classification Based on Gene Expression Data: Study of Sensitivity to Choice of Parameters of the GA/KNN Method*. Bioinformatics.
- [4] Fort G., Lambert-Lacroix, S. (2004). *Classification using Partial Least Squares with Penalized Logistic Regression*.
- [5] 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*.
- [6] Bhuvanewari V., Vanitha K. (2012). *Classification of Microarray Gene Expression Data by Gene Combinations using Fuzzy Logic*. International Journal of Computer Science, Engineering and Applications.
- [7] Tian, David & Keith Burley. *Classification of Micro-array Gene Expression Data using Neural Networks*.
- [8] Furey, T., Cristianini, T., Duffy, N., Bednarski, D., Schummer, M., and Haussler, D. (2000). *Support Vector Machine Classification and Validation of Cancer Tissue Samples using Micoarray Expression Data*.
- [9] Lisnawati, E., Adiwijaya, Murdiansyah, D. T., Wisesty, N., U. (2016). *Deteksi Kanker Berdasarkan Klasifikasi Microarray Data Menggunakan Principal Component Analysis dan Support Vector Machine*.
- [10] Mundra, A., P., and Rajapakse, C., J. (2009). *SVM-RFE with MRMR Filter for Gene Selection*. IEEE.
- [11] Singh, K., R., and Dr. M. Sivabalakrishnan. (2015). *Feature Selection of Gene Expression Data for Cancer*

Classification: A Review. ISBCC 2015.

- [12] Schimdt, Uli & C. Glenn Begley. (2003). *Cancer Diagnosis and Microarrays*. Geoffrey Laurent. *The International Journal of Biochemistry & Cell Biology*. Pp 119-124. Volume 35. Edisi 2. Elsevier.
- [13] NCBI, Principal components analysis to summarize *microarray* experiments: application to sporulation time series, <https://www.ncbi.nlm.nih.gov/pubmed/10902193>, diakses pada tanggal 13 April 2017.
- [14] Ding, Chris & Hanchuan Peng. *Minimum Redundancy Feature Selection from Microarray Gene Expression Data*.
- [15] Kent Ridge Biomedical Data Set Repository, <http://leo.ugr.es/elvira/DBCRepository/>, diakses pada tanggal 10 April 2017.
- [16] Elizondo, A., D., Benjamin N. P., Ralph B., and Andreas Huemer. *Dimensionality Reduction and Microarray Data*. Centre for Computational Intelligence, School of Computing, Faculty of Computing Sciences and Engineering, De Montfort University, Leicester, UK.
- [17] Genetics Home Reference, What Is DNA?, <https://ghr.nlm.nih.gov/primer/basics/dna>, diakses pada tanggal 10 April 2017.
- [18] Cristian Dima, Basics of support vector machine, <http://www.cristiandima.com/basics-of-support-vector-machines/>, diakses pada tanggal 11 Februari 2017.
- [19] Anindiya Apriliyanti, P., 2009. *Penentuan Banyak Kelompok dalam Fuzzy C-Means Cluster Berdasarkan Proporsi Eigen Value Dari Matriks Similarity dan Indeks XB (Xie dan Beni)*. Seminar Nasional Matematika dan Pendidikan Matematika 2009. Jurusan Pendidikan Matematika FMIPA. UNY.
- [20] Mariana, I.A.I.N., 2016. *Analisis Komponen Utama*. INTEGRAL, 2(2).
- [21] Nelson Spencer, Machine Learning: Supervised Learning pt. 2, <http://www.nelsonspencer.com/blog/2015/2/15/machine-learning-supervised-learning-pt-2>, diakses pada tanggal 11 Februari 2017.
- [22] Faria, Fabio & dos Santos, Jefersson & Rocha, Anderson & Torres, Ricardo. (2012). *Automatic Classifier Fusion for Produce Recognition*. Brazilian Symposium of Computer Graphic and Image Processing. 252-259. 10.1109/SIBGRAPI.2012.42.
- [23] Adiwijaya. 2014. *Aplikasi Matriks dan Ruang Vektor*. Yogyakarta: Graha Ilmu
- [24] Adiwijaya, 2016, *Matematika Diskrit dan Aplikasinya*, Bandung: Alfabeta
- [25] Nurfalah, A., Adiwijaya, and Suryani, A.A, 2016. *Cancer Detection Based On Microarray Data Classification Using PCA And Modified Bank Propagation*. Far East Journal of Electronics and Communications, 16 (2), p. 269
- [26] L Marianah, F Nhita, A Adiwijya. 2014. *Klasterisasi Data Microarray Menggunakan Metode Clique Partitioning*. Indonesian Symposium on Computing 2014