Klasifikasi Wisconsin Diagnostic Breast Cancer Data dengan Menggunakan Sequential Feature Selection dan Possibilistic C-Means

Aini Suri Talita
Jurusan Teknik Informatika, Fakultas Teknologi Industri, Universitas Gunadarma
Jl. Margonda 100, Depok
ainisuri@staff.gunadarma.ac.id

Abstrak

Pada masalah klasifikasi yang melibatkan data berdimensi tinggi, metode pengurangan dimensi data menjadi salah satu hal penting. Hal ini diakibatkan adanya masalah *curse of dimensionality* yang mengimplikasikan bahwa dimensi data secara signifikan berpengaruh terhadap kompleksitas waktu maupun ruang dari tahapan pemrosesan data. Salah satu metode pengurangan dimensi adalah metode seleksi fitur (*feature selection*). Secara umum, metode seleksi fitur dapat dikategorikan menjadi tiga jenis, yaitu *filter based method, wrapped based*, dan *embedded method*. Metode pemilihan fitur sekuensial merupakan salah satu algoritma seleksi fitur yang klasik berbasiskan metode filter. Selain daripada metode pengurangan dimensi yang digunakan, metode klasifikasi itu sendiri merupakan kunci penting dalam suksesnya pemecahan masalah klasifikasi. Salah satu metode klasifikasi berbasiskan metode *fuzzy* yang umum digunakan adalah metode Fuzzy C-Means. Metode ini memiliki kekurangan dalam hal adanya kemungkinan nilai keanggotaan yang dihasilkan tidak merepresentasikan secara akurat nilai kemungkinan suatu data berada di suatu kelas tertentu. Kekurangan ini dapat diatasi dengan metode Possibilistic C-Means. Dalam makalah ini, metode klasifikasi Possibilistic C-Means dan metode pengurangan dimensi melalui seleksi fitur dengan menggunakan metode pemilihan fitur sekuensial akan digunakan untuk memecahkan masalah klasifikasi data Wisconsin Diagnostic Breast Cancer.

Kata Kunci : fitur, klasifikasi, pemilihan fitur, pengurangan dimensi

Pendahuluan

Data berdimensi tinggi pada ruang masukan pada umumnya kurang efektif untuk masalah yang berkaitan dengan pengenalan pola, seperti masalah klasifikasi, dikarenakan adanya masalah "*curse of dimensionality*" yaitu dimensi data secara signifikan meningkatkan kompleksitas waktu dan ruang yang dibutuhkan untuk memproses data karena data yang dibutuhkan untuk "melatih" *classifier* bertambah secara eksponensial dengan bertambahnya dimensi data. Pada masalah praktis, masalah ini dapat diartikan dengan untuk suatu ukuran data tertentu, terdapat suatu nilai banyaknya fitur maksimum dimana setelah melewati nilai tersebut, pertambahan fitur dapat menyebabkan menurunnya performa *classifier* yang digunakan. Pengurangan fitur yang dapat mengakibatkan hilangnya sebagian informasi data pada banyak kasus dapat ditanggulangi dengan pemetaan yang tepat pada ruang berdimensi yang lebih rendah. Keberadaan fitur data yang irelevan maupun *redundant* dapat pula mengakibatkan metode-metode berbasis pembelajaran menjadi kurang efektif.

Salah satu cara untuk mengurangi dimensi data (*dimensionality reduction*) diantaranya adalah seleksi fitur maupun ekstraksi fitur yang bertujuan untuk menentukan himpunan fitur data yang akan dipilih. Ekstraksi fitur dan seleksi fitur memiliki perbedaan yang mendasar. Pada ekstraksi fitur, fitur data ditransformasi ke ruang yang berdimensi lebih rendah, sedangkan pada seleksi fitur, dipilih himpunan bagian dari fitur untuk menjadi fitur data yang akan diproses (tanpa adanya transformasi). Contoh dari metode ekstraksi fitur diantaranya adalah metode Principal Component Analysis (PCA) dan Linear Discriminant Analysis (LDA). Secara umum, masalah seleksi fitur dapat dibedakan menjadi tiga jenis [1] yaitu masalah seleksi fitur berbasiskan *filter*, *wrapper*, dan metode tertanam (*embedded*).

Metode berbasiskan *filter* mengurutkan fitur sebelum algoritma pembelajaran diterapkan, dan memilih fitur dengan nilai urutan tertinggi,

misalnya uji F-ANOVA, Fisher Score, uji-t dan uji chi-square (χ^2). Salah satu penelitian yang menggunakan metode seleksi fitur uji F-ANOVA adalah penelitian [2]. Penelitian ini menggunakan metode *enhanced* SVM untuk menyelesaikan masalah klasifikasi *e-mail spam*.

Contoh metode seleksi fitur berbasis *filter method* lainnya adalah metode Fisher Score (FS). Metode ini memilih masing-masing fitur secara terpisah berdasarkan nilai mereka sesuai dengan kriteria Fisher yang akan menuju pada terbentuknya suatu suboptimal himpunan bagian fitur. Penelitian [1] mengajukan suatu pengembangan dari metode FS yang disebut Generalized Fisher Score (GFS) yang bertujuan untuk secara bersamaan memilih fitur-fitur yang akan dipilih. Himpunan bagian fitur yang dibentuk memaksimumkan batas bawah dari metode FS. Masalah seleksi fitur yang baru ini merupakan masalah pemograman bilangan bulat campuran (*mixed integer programming*) yang dapat diformulasikan sebagai masalah pemrograman linier dengan batas kuadratik (Quadratically Constrained Linear Programming, QCLP). Berdasarkan percobaan dengan menggunakan *benchmark data* GFS memberikan hasil yang lebih baik dari metode FS.

Penelitian [3] mengembangkan modifikasi dari pengukuran peringkat uji-t dan menerapkannya pada data genotype HapMap dengan cara mengurutkan semua SNP (Single Nucleotide Polymorphisms), yaitu variasi genetik yang menentukan perbedaan antara dua individu yang tidak berelasi. Data HapMap memiliki empat grup populasi dengan sekitar 10 juta SNP, sehingga pemilihan maupun ekstraksi fitur sangat diperlukan dalam masalah ini. HapMap *classifier* yang digunakan pada penelitian ini adalah Support Vector Machines (SVM).

Metode *wrapped-based* mengurutkan nilai fitur dengan terlebih dahulu melatih model prediktif pada himpunan bagian dari fitur, misalnya metode Sequential Feature Selection atau Genetic Algorithm. Sedangkan metode ketiga yaitu *embedded based method* mengkombinasikan masalah seleksi fitur dengan algoritma pembelajaran, yaitu menentukan himpunan bagian dari fitur yang optimal secara langsung dengan bobot hasil *training* pada metode klasifikasi.

Pada masalah klasifikasi, hal penting lain selain masalah seleksi fitur data adalah masalah penentuan metode klasifikasi itu sendiri. Terdapat beberapa metode klasifikasi yang sering digunakan, seperti metode Fuzzy C-Means yang digunakan pada [4] untuk mengklasifikasikan data Iris, metode Fuzzy Kernel K-Medoids yang merupakan kombinasi dari metode Fuzzy K-Medoids dan metode Kernel [5], maupun penelitian [6] yang memanfaatkan metode Possibilistic C-Means (PCM) dan Fuzzy PCM untuk menyelesaikan masalah model *clustering blind speaker*. Terkait pentingnya pemilihan fitur data pada masalah klasifikasi, pada makalah ini akan dibahas penggunaan metode seleksi fitur Sequential Feature Selection dan metode klasifikasi Possibilistic C-Means untuk menyelesaikan masalah pengurangan dimensi dan klasifikasi data Winsconsin Diagnostic Breast Cancer (WDBC).

Metode Penelitian

Pada makalah ini, metode pengurangan dimensi data yang akan digunakan adalah salah satu metode klasik dari metode seleksi fitur yaitu Metode Seleksi Fitur Sekuensial (Sequential Feature Selection). Metode ini dapat dibagi menjadi dua jenis yaitu metode Sequential Forward Feature Selection (SFFS) dan Sequential Backward Feature Selection (SBFS). Secara umum, masalah seleksi fitur dapat didefinisikan dengan: diberikan suatu himpunan fitur data $\{x_i,\ i = 1, 2, \dots, M\}$ akan dipilih himpunan bagiannya yaitu $\{x_{i1}, x_{i2}, \dots, x_{iN}, i \in \{1, 2, \dots, M\}, N < M\}$. Pada SFFS, himpunan fitur terpilih dimulai dari himpunan kosong, secara sekuensial tambahkan fitur x^+ yang menyebabkan nilai fungsi objektif tertentu mencapai nilai optimum pada himpunan fitur terpilih. Hal ini terus dilakukan sampai tidak ada lagi penambahan fitur yang dapat mengoptimumkan nilai fungsi objektif. Sebaliknya pada SBFS, himpunan fitur terpilih dimulai dari himpunan fitur awal data, kemudian secara sekuensial mengurangi fitur pada himpunan tersebut yang dapat mengoptimumkan nilai fungsi objektif. Metode klasifikasi (*classifier*) yang digunakan pada makalah ini adalah metode Possibilistic C-Means (PCM) [7] yang akan dibahas lebih lanjut pada bagian Hasil dan Pembahasan. Berikut ini diberikan penjelasan mengenai data BWDC yang akan diujicoba pada makalah ini.

Wisconsin Diagnostic Breast Cancer Data (WDBC)

Data yang digunakan dalam penelitian ini adalah data Wisconsin Diagnostic Breast Cancer (WDBC). Data ini didapatkan dari Rumah Sakit University of Wisconsin, Madison yang merupakan milik dari Dr. William H. Wolberg. Data ini pertama kali digunakan pada penelitian [8] mengenai ekstraksi fitur inti sel berkaitan dengan pendiagnosaan tumor payudara. Sebagian dari fitur pada data WDBC yaitu fitur nomor 2 sampai dengan 10 digunakan pada penelitian [9] yang menerapkan metode *multisurface pattern separation* pada sitologi payudara. Kesepuluh fitur pada data ini dihitung dari citra terdigitalisasi payudara dengan metode *fine needle aspiration* (FNA) yang mendeskripsikan karakteristik dari inti sel pada citra. Data ini terbagi menjadi dua kelas (*cluster*) yaitu kanker payudara diagnosis benign (B) dan malignant (M). Pada himpunan data label ini disimbolkan dengan "2" untuk benign dan "4" untuk malignant. Sepuluh fitur bernilai riil yang dihitung pada setiap inti sel menyatakan: radius (rata-rata jarak dari titik pusat ke titik perimeter), tekstur (standar deviasi dari nilai skala keabuan), perimeter, luas, kehalusan (variasi lokal di sepanjang radius), kepadatan, titik kecekungan, kesimetrian, dan dimensi fraktal [10]. Fitur kehalusan dihitung dengan mengukur perbedaan antara panjang jari-jari dan rata-rata garis di sekitarnya, kecekungan dihitung dengan mengukur kecekungan kurva pada batas (*boundary*) inti sel. Untuk mengukur kesimetrian, sumbu utama atau segmen terpanjang yang melalui titik pusat harus dihitung. Kemudian dihitung selisih antara garis yang tegak lurus sumbu utama menuju batas inti sel di kedua arah. Sedangkan fitur tekstur didapatkan dengan menentukan variansi dari intensitas skala keabuan pada komponan – komponen piksel citra [10].

Tabel 1: Contoh Data WDBC

Data	Kelas
1000025,5,1,1,1,2,1,3,1,1	2
1002945,5,4,4,5,7,10,3,2,1	2
1017122,8,10,10,8,7,10,9,7,1	4
1018561,2,1,2,1,2,1,3,1,1	2
1033078,2,1,1,1,2,1,1,1,5	2
1036172,2,1,1,1,2,1,2,1,1	2
1041801,5,3,3,3,2,3,4,4,1	4
1044572,8,7,5,10,7,9,5,5,4	4
1015425,3,1,1,1,2,2,3,1,1	2
1054590,7,3,2,10,5,10,5,4,4	4
1016277,6,8,8,1,3,4,3,7,1,	2
1017023,4,1,1,3,2,1,3,1,1	2
1050670,10,7,7,6,4,10,4,1,2	4
1018099,1,1,1,1,2,10,3,1,1	2
1054593,10,5,5,3,6,7,7,10,1	4

Metode PCM dan metode SFS yang digunakan pada penelitian ini diapliasikan dengan menggunakan *software* Matlab 2012 yang selanjutnya diujicobakan pada data WDBC. Uji coba dilakukan pada komputer pribadi dengan *processor* i-7, 1TB HD, dan 16 GB RAM.

Hasil dan Pembahasan

Metode Possibilistic C-Means

Salah satu pendekatan dalam metode *clustering* maupun klasifikasi adalah metode berbasiskan metode Fuzzy. Salah satunya yang sering digunakan adalah metode Fuzzy C-Means (FCM). FCM menggunakan syarat batas probabilistik bahwa total nilai keanggotaan (*membership*) untuk tiap-tiap data pada semua kelas adalah 1. Syarat batas tersebut digunakan untuk membangun fungsi keanggotaan (*membership*) dengan menggunakan suatu algoritma iteratif. Nilai-nilai keanggotaan hasil dari metode FCM tidak selalu berkaitan dengan konsep intuitif "kecocokan" atau "kemungkinan" dari suatu data untuk berada di suatu *cluster* atau kelas tertentu. Algoritma FCM juga memiliki kekurangan apabila digunakan pada data dengan tingkat *noise* yang tinggi. Penelitian [7] mengajukan metode Possibilistic C-Means yang menggunakan pendekatan posibilistik dalam artian bahwa nilai keanggotaan yang dihasilkan oleh metode ini dapat mencerminkan "kecocokan" atau "kemungkinan" dari suatu data untuk terletak pada suatu kelas tertentu. Fungsi objektif pada FCM dimodifikasi untuk menurunkan fungsi keanggotaan baru dari metode PCM.

Misalkan U menotasikan matriks partisi *fuzzy* yang dibangun pada algoritma FCM. Maka untuk setiap i, j, elemen u_{ij} dari U memenuhi:

$$u_{ij} \in [0,1] \quad (1)$$

$$0 < \sum_{j=1}^{N} u_{ij} < N, \forall i \text{ dan}$$

$$\sum_{j=1}^{c} u_{ij} = 1, \forall j \quad (2)$$

Dengan u_{ij} menyatakan nilai keanggotaan dari data $\boldsymbol{x}_j$ pada kelas β_i, c adalah banyaknya kelas, dan N adalah banyaknya fitur data. Selanjutnya notasi β_i akan digunakan untuk merepresentasikan kelas ke-i dan pusat *cluster*nya.

Fungsi objektif dari FCM yang ingin diminimumkan adalah:

$$J(L,U) = \sum_{i=1}^{c}\sum_{j=1}^{N}\left(u_{ij}\right)^m d_{ij}^2 \quad (3)$$

Dengan kendala

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \quad (4)$$

Dimana $L = (\beta_1, \beta_2, \dots, \beta_c)$ merupakan c-*tuple* dari *prototype* (pusat *cluster*), d_{ij}^2 adalah jarak dari data ke pusat *cluster*. Dalam hal ini u_{ij} merupakan nilai keanggotaan dari $\boldsymbol{x}_j$ pada *cluster* β_i dan m adalah nilai derajat ke-fuzzy-an, $m \in [1, \infty]$.

Dengan menyederhanakan syarat kendala pada (4) menghasilkan solusi trivial yaitu fungsi kriteria diminumkan dengan memberikan nilai 0 pada semua keanggotaan. Nilai keanggotaan untuk data perwakilan kelas diharapkan sebesar-besarnya dan nilai keanggotaan untuk data yang tidak mewakili kelas diharapkan bernilai sekecil mungkin. Fungsi objektif yang baru diberikan oleh:

$$J_m(L,U)\sum_{i=1}^{c}\sum_{j=1}^{N}\left(u_{ij}\right)^m d_{ij}^2 + \sum_{i=1}^{c}\eta_i\sum_{j=1}^{N}\left(1-u_{ij}\right)^m \quad (5)$$

Dengan nilai η_i diberikan oleh:

$$\eta_i = K\frac{\sum_{j=1}^{N} u_{ij}^m d_{ij}^2}{\sum_{j=1}^{N} u_{ij}^m} \quad (6)$$

Untuk mendapatkan nilai global minimum dari $J_m(L,U)$ fungsi untuk memperbaharui nilai keanggotaan haruslah bernilai:

$$u_{ij} = \left(1 + \left(\frac{d_{ij}^2}{\eta_i}\right)^{1/(m-1)}\right)^{-1} \quad (7)$$

Sehingga, pada setiap iterasi, nilai terbarukan u_{ij} hanya bergantung pada jarak antara $\boldsymbol{x}_j$ dan β_i yang secara intuitif merupakan hasil yang baik. Dari sudut pandang "kecocokan dengan pusat *cluster*", nilai keanggotaan dari suatu titik data di suatu kelas seharusnya hanya ditentukan oleh sejauh mana jaraknya dengan pusat *cluster* dan tidak lagi dikalikan dengan lokasinya terhadap kelas lainnya. Akibat dari formulasi ini adalah dimungkinkan solusi keanggotaan optimum untuk seluruhnya terletak pada unit *hypercube*, tidak lagi terbatas pada suatu *hyperplane* yang direpresentasikan oleh syarat $\sum_{i=1}^{c} u_{ij} = 1$. Selanjutnya pusat *cluster* diperbaharui dengan $\beta_i = \frac{\sum_{j=1}^{N} u_{ij}^m \boldsymbol{x}_j}{\sum_{j=1}^{N} u_{ij}^m}$, untuk $i = 1, 2, \dots, c$.

Tabel 2. Algoritma Possibilistic C-Means

Input	:	Data masukan X, banyak kelas c, $\boldsymbol{\beta}$ sebagai inisialisasi awal dari himpunan pusat *cluster*, m sebagai parameter *fuzziness degree* dari PCM, K sebagai konstanta penghitungan η_i dan T sebagai batas iterasi
Output	:	Matriks membership U dan himpunan pusat *cluster* $\boldsymbol{\beta}$
	1.	Inisialisasi awal: $\boldsymbol{\beta}$ dipilih berisikan *mean* dari vektor masukan yang dijadikan data *training*
	2.	For $t = 1$ to T
	3.	$b = \frac{1}{m-1}$, $\eta_i = K\frac{\sum_{j=1}^{N} u_{ij}^m d_{ij}^2}{\sum_{j=1}^{N} u_{ij}^m}$
	4.	Perbaharui matriks keanggotaan dengan formula: $u_{ij} = \left(1 + \left(\frac{d_{ij}^2}{\eta_i}\right)^b\right)^{-1}$
	5.	Perbaharui pusat *cluster* dengan formula: $\beta_i = \frac{\sum_{j=1}^{N} u_{ij}^m \boldsymbol{x}_j}{\sum_{j=1}^{N} u_{ij}^m}$ $i = 1, 2, \dots, c$
	6.	$t = t + 1$

Tabel 2 memberikan ringkasan Algoritma Possibilistic C-Means yang menggunakan formula-formula yang telah dibahas sebelumnya.

Pada penelitian ini, kriteria penghentian adalah dengan membatasi jumlah iterasi. Terdapat alternatif pilihan kriteria lainnya yaitu apabila nilai-nilai pada matriks keanggotaan U telah konvergen ke suatu nilai maka algoritma telah konvergen, yaitu dengan menentukan apakah nilai $\|U_{old} - U_{new}\|$ telah bernilai cukup kecil kurang dari suatu toleransi ε.

Tabel 3. Hasil Klasifikasi Algoritma PCM

Data *Training*	Waktu Berjalannya Program (detik)		Persentase Keberhasilan	
	Menggunakan Seluruh *Feature*	Menggunakan *Feature* Pilihan *Sequential Feature Selection*	Menggunakan Seluruh *Feature*	Menggunakan *Feature* Pilihan *Sequential Feature Selection*
10%	2.280	0.156	56.9841	79.2063
20%	4.248	0.280	55.1786	78.9286
30%	6.751	0.401	53.0612	78.1633
40%	8.182	0.511	61.6667	78.3333
50%	7.147	0.622	69.4286	77.7143
60%	10.344	0.749	65.8363	77.2242
70%	9.213	0.854	56.3981	75.3555
80%	13.146	0.985	59.5745	76.5957

Metode pemilihan fitur SFS yang digunakan pada makalah ini adalah SBFS, yang diaplikasikan pada *software* Matlab dengan menggunakan perintah "*sequentialfs*". Hasil penerapan data WDBC adalah terpilihnya fitur ke-10 sebagai fitur tunggal, dengan nilai kriteria pengujian (*criterion value*) 0.018958. Pada Tabel 3 diberikan perbandingan hasil klasifikasi data WDBC dengan menggunakan metode PCM, baik dengan memilih fitur dengan SBFS maupun non seleksi fitur. Parameter yang dipilih berkaitan dengan metode PCM adalah nilai m = 2, K = 6, dan banyaknya iterasi maksimum adalah 500. Fungsi jarak yang digunakan adalah Euclidean Distance.

Dapat dilihat pada Tabel 3 bahwa dengan menggunakan metode pemilihan fitur SBFS, waktu berjalannya program semakin singkat yang diakibatkan berkurangnya dimensi data *training* maupun data uji coba. Berkaitan dengan presentase keberhasilan metode klasifikasi, untuk setiap kemungkinan data training yang diujicoba, persentase keberhasilan dengan hanya menggunakan satu fitur hasil SBFS lebih tinggi daripada dengan menggunakan seluruh fitur data. Hal ini dimungkinkan terjadi karena pada beberapa kasus, terdapat fitur data yang tidak relevan ataupun *redundant*.

Kesimpulan dan Saran

Pada makalah ini telah diaplikasikan gabungan antara metode pemilihan fitur Sequential Backward Feature Classification (SBFS) dan metode klasifikasi Possibilistic C-Means pada masalah pengurangan dimensi dan klasifikasi data Wisconsin Diagnostic Breasr Cancer (WDBC). Berdasarkan hasil uji coba dengan menggunakan 10% , 20%, .., dan 80% bagian data sebagai data *training*, didapat bahwa dengan hanya menggunakan satu buah fitur hasil pemilihan dari metode SBFS, waktu berjalannya program semakin berkurang dan persentase keberhasilan klasifikasi semakin meningkat apabila dibandingkan dengan hanya menggunakan metode klasifikasi PCM tanpa adanya pemilihan fitur. Persentase keberhasilan tertinggi dicapai pada penggunaan 10% data *training* yaitu sebesar 79.2063%. Penelitian ini masih bisa ditingkatkan dengan pemilihan metode seleksi maupun metode klasifikasi yang lebih akurat untuk meningkatkan persentase keberhasilan klasifikasi.

Daftar Pustaka

[1] Q. Gu, Z. Li, and J. Han. "Generalized Fisher Score for Feature Selection" In *Proc. Of the 27th Conference on Uncertainty in Artificial Intelligence (UAI)*, 14-17 July, Barcelona, Spain, 2011.

[2] O. F. Elssied, O. Ibrahim, and A. H. Osman. "A Novel Feature Selection Based on One-Way ANOVA F-Test for E-mail Spam Clasiffication". *Research Journal of Applied Sciences, Engineering and Technology*, 7, 3, pp. 625-638, 2014.

[3] N. Zhou and L. Wang. "A Modified T-test Feature Selection Method and its Application on the HapMap Genotype Data" *Geno. Prot. Bioinfo* 5, 3-4, pp. 242-249, 2007.

[4] A. S. Talita. "Klasifikasi Data Iris Menggunakan Metode Fuzzy C-Means dan Modifikasi dari Canberra Distance" *UG Jurnal* 09, 11, pp. 12-19, 2015.

[5] Z. Rustam and A.S. Talita. "Fuzzy Kernel K-Medoids Algorithm for Multiclass Multidimensional Data Classification" *Journal of Theoretical and Applied Information Technology*, 80, 01, pp. 147-151, 2015.

[6] G. Gosztolya and L. Szilagyi. "Application of Fuzzy and Possibilistic C-Means Clustering Models in Blind Speaker Clustering". *Acta Polytechnica Hungarica* 12, 7, pp. 41-56, 2015.

[7] R. Krishnapuram and J. M. Keller. "A Possibilistic Approach to Clustering" *IEEE Transactions on Fuzzy Systems*, 1, 2, pp. 98-110, 1993.

[8] W. N. Street, W. H. Wolberg, and O.L. Mangasarian. "Nuclear Feature Extraction for Breast Tumor Diagnosis IS&T/SPIE" *International Symposium on Electronic Imaging: Science and Technology*, 1905, pp. 861-870, San Jose, California, 1993.

[9] W. H. Wolberg and O. L. Mangasarian. "Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology" *Proceedings of the National Academy of Sciences, U.S.A*, 87, pp. 9193-9196, 1990.

[10] G.I. Salama, M. B. Abdelhalim, and M. A. Zeid. "Breast Cancer Diagnosis on Three Different Datasets using Multi-classifiers" *International Journal of Computer and Information Technology*, 01, 01, pp. 36-43, 2012.