

UNIVERSITAS INDONESIA

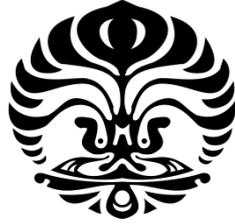
**KLASIFIKASI KANKER OTAK (ASTROCYTOMA)
MENGUNAKAN INDEPENDENT COMPONENT
ANALYSIS DAN POSSIBILISTIC C-MEANS**

SKRIPSI

ANGGI PANDYO WIBOWO

0305010076

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
DEPARTEMEN MATEMATIKA
DEPOK
JULI 2010**



UNIVERSITAS INDONESIA

**KLASIFIKASI KANKER OTAK (ASTROCYTOMA)
MENGUNAKAN INDEPENDENT COMPONENT ANALYSIS
DAN POSSIBILISTIC C-MEANS**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
sarjana sains**

ANGGI PANDYO WIBOWO

0305010076

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
DEPARTEMEN MATEMATIKA
DEPOK
JULI 2010**

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar

Nama : Anggi Pandyo Wibowo
NPM : 0305010076
Tanda Tangan :
Tanggal : 5 Juli 2010

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh

Nama : Anggi Pandyo Wibowo
NPM : 0305010076
Program Studi : Matematika
Judul Skripsi : Klasifikasi Kanker Otak (Astrocytoma)
Menggunakan Independent Component Analysis
dan Possibilistic C-Means

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjan Sains pada Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia

DEWAN PENGUJI

Pembimbing : Dr. Zuherman Rustam, DEA.

Pembimbing : dr. Jacob Pandelaki, SpRad.

Penguji : Prof. Dr. Belawati H Widjaja.

Penguji : Dr. Hengki Tasman, M.Si.



(.....)

(.....)

(.....)

(.....)

Ditetapkan di : Depok

Tanggal : 5 Juli 2010

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi ini. Penulisan skripsi dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Sains jurusan Matematika pada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Indonesia.

Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi penulis untuk menyelesaikan skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

- (1) Dr. Zuherman Rustam DEA, selaku pembimbing pertama yang telah menyediakan waktu, tenaga, pikiran, serta kesabaran dalam membimbing dan mengajarkan penulis dalam penyelesaian skripsi ini;
- (2) dr. Jacob Pandelaki SpRad, selaku pembimbing kedua yang telah mengajarkan dan membimbing penulis untuk memahami lebih dalam mengenai masalah kedokteran dan kanker otak, serta keramahan yang selalu diberikan kepada penulis;
- (3) Bapak Suryadi, M.T. selaku pembimbing akademik, yang telah memberikan bantuan dan perhatian, mulai dari awal masa perkuliahan hingga penulisan skripsi ini selesai;
- (4) Prof. Dr. Belawati H Widjaja dan Dr. Hengki Tasman, M.Si., yang telah bersedia menjadi penguji dalam sidang sarjana penulis;
- (5) Abah dan Ibu yang telah mendidik dan merawat penulis sejak kecil dan yang selalu mendoakan dan menemani penulis dalam pembuatan skripsi ini;
- (6) Eyang Kakung, Eyang Putri, dan Mas Jalu yang selalu mendoakan dan menanyakan kemajuan skripsi penulis;

- (7) Riesa Anandya Elfitra yang selalu memberikan semangat kepada penulis, menemani penulis disetiap kesempatan, serta selalu mendoakan penulis;
- (8) Sahabat GEMBOSS yang selalu menghibur dan juga menasehati penulis;
- (9) Yanuar yang telah membantu penulis dalam pembuatan program dalam skripsi ini;
- (10) Ardibian yang telah menjadi rekan skripsi penulis yang sangat baik dan rajin;
- (11) Sahabat Magic The Gathering, yang telah mengisi waktu luang penulis ketika berada di departemen selama penulisan skripsi ini;
- (12) Fika, Dimas, Hairu, Fia, Hamdan, Melati, Dian, Qq, Nisma, Karlina yang telah bersama-sama berjuang menyelesaikan skripsi masing-masing pada semester ini;
- (13) Akmal, Inul, dan May yang telah banyak membantu penulis baik dalam skripsi atau hal-hal diluar skripsi;
- (14) Semua Teman-teman angkatan 2005 yang tidak dapat disebutkan satu per satu, yang telah menjadi teman yang baik bagi penulis selama di matematika;
- (15) Seluruh Dosen Matematika, yang telah mendidik dan memberikan perhatian kepada penulis selama penulis menjalani masa perkuliahan hingga akhir penyusunan skripsi ini;
- (16) Seluruh Staff administrasi departemen matematika, khususnya Mba Santi, yang telah membantu penulis dalam berbagai hal yang berhubungan dengan masalah administrasi kuliah maupun skripsi;
- (17) Dan semua orang yang tidak dapat disebutkan satu per satu, yang telah membantu penulis dalam menyelesaikan skripsi;

Akhir kata, penulis berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu.

Penulis

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama : Anggi Pandyo Wibowo
NPM : 0305010076
Program Studi : Matematika
Departemen : Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul :

Klasifikasi Kanker Otak (Astrocytoma) Menggunakan Independent Component Analysis dan Possibilistic C-Means

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya,

Dibuat di : Depok
Pada Tanggal : 5 Juli 2010
Yang menyatakan

(Anggi Pandyo Wibowo)

ABSTRAK

Nama : Anggi Pandyo Wibowo
Program Studi : Matematika
Judul : Klasifikasi Kanker Otak (*Astrocytoma*) Menggunakan
Independent Component Analysis dan Possibilistic C-Means

Magnetic Resonance Spectroscopy (MRS) membantu ahli radiologi untuk mengetahui tingkat keganasan dari kanker otak (*astrocytoma*). Dalam tugas akhir ini akan dibahas proses klasifikasi terhadap data hasil MRS untuk mengetahui tingkat keganasan dari *astrocytoma* yang terdiri dari Tingkat rendah (*Low Grade*), Tingkat tinggi (*High Grade*), dan Normal. Data yang digunakan dalam tugas akhir ini berasal dari RSUD Pusat Nasional Dr. Cipto Mangunkusumo, Jakarta. Metode yang digunakan untuk klasifikasi adalah metode *Independent Component Analysis* dan metode *Possibilistic C-Means*. Hasil percobaan yang dilakukan menunjukkan bahwa metode *Independent Component Analysis* mempunyai nilai akurasi 96,67% sementara nilai akurasi dari metode *Possibilistic C-Means* mencapai 90,91%. Dalam tugas akhir ini, akan dibuat sebuah perangkat lunak untuk pendukung keputusan yang membantu memberikan informasi mengenai tingkat keganasan dari *astrocytoma*.

Kata Kunci : *Astrocytoma*, MRS, klasifikasi, ICA, PCM
xi+34 halaman ; 9 gambar; 6 tabel
Daftar Pustaka : 15 (1985-2010)

ABSTRACT

Name : Anggi Pandyo Wibowo
Program Study : Mathematic
Title : Brain Cancer (Astrocytoma) Classification Using Independent
Component Analysis and the Possibilistic C-Means

Magnetic Resonance Spectroscopy (MRS) can help radiologists to identify the level of malignancy of brain cancer (astrocytoma). This final project will discuss the process of classification of the data result of MRS to determine the level of malignancy of astrocytoma which consists of a low level (Low Grade), High Level (High Grade), and Normal. Data used in this final project comes from the National Center Hospital Dr. Cipto Mangunkusumo, Jakarta. The methods used for classification is the method of Independent Component Analysis and Possibilistic C-Means method. Experimental results show that the method of Independent Component Analysis has 96.67% accuracy value while the value of the accuracy of Possibilistic C-Means method reaches 90.91%. In this final project, a software for decision support that helps to provide information about the level of malignancy of the astrocytoma will be made.

Keywords : Astrocytoma, MRS, classification, ICA, PCM
xi+34 pages ; 9 pictures; 6 tables
Bibliography : 15 (1985-2010)

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR.....	iv
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	vi
ABSTRAK	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xi
1. PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penulisan.....	3
1.5 Sistematika Penulisan	3
2. KANKER OTAK DAN METODE POSSIBILISTIC C-MEANS.....	4
2.1 Kanker Otak	4
2.2 Magnetic Resonance Spectroscopy (MRS).....	6
2.2.1 Penjelasan MRS	6
2.2.2 Biokimia MRS	7
2.3 Possibilistic C-Means.....	8
3. INDEPENDENT COMPONENT ANALYSIS (ICA).....	14
3.1 Ruang Hasil Kali Dalam dan Norm	14
3.1.1 Hasil Kali Dalam.....	14

3.1.2 Norm Vektor	15
3.1.3 Norm Matriks	16
3.2 Nilai dan Vektor Eigen.....	17
3.3 Statistik Dasar	17
3.3.1 Definisi Ekspektasi.....	17
3.3.2 Estimasi dari Ekspektasi	18
3.3.3 Definisi Matriks Korelasi.....	18
3.3.4 Definisi Matriks Kovariansi.....	19
3.4 Independent Component Analysis	19
3.4.1 Blind Source Separation (BSS).....	19
3.4.2 Definisi dasar ICA.....	20
3.4.3 <i>Infomax Learning Algorithm</i>	22
4. HASIL DAN PEMBAHASAN	28
4.1 Tinjauan Data	28
4.2 Hasil Pengujian Data MRS dengan Metode ICA.....	29
5. KESIMPULAN DAN SARAN	32
DAFTAR PUSTAKA	33

DAFTAR GAMBAR

Gambar 2. 1 Kanker otak secara Histologis	5
Gambar 2. 2 Flowchart PCM untuk training data	11
Gambar 2. 3 Flowchart data testing	13
Gambar 3. 1 Ilustrasi masalah Blind Source Separation (BSS)	20
Gambar 3. 2 Kerangka Teoritis dari Algoritma untuk menyelesaikan model ICA	22
Gambar 3. 3 Flowchart tahap preprocessing ICA	25
Gambar 3. 4 Flowchart Algoritma Infomax dengan Natural Gradient Learning..	25
Gambar 3. 5 Flowchart data testing pada ICA	27
Gambar 4. 1 Grafik perbandingan akurasi metode ICA dengan PCM	31

DAFTAR TABEL

Tabel 2. 1 Ukuran ppm dari metabolit yang digunakan.....	7
Tabel 3. 1 Aturan Norm Vektor	15
Tabel 3. 2 Contoh aturan Norm Matriks	16
Tabel 4. 1 Ukuran ppm dari metabolit yang diatur pada mesin GE.....	29
Tabel 4. 2 Prosentase Hasil Akurasi Percobaan Metode Independent Component Analysis untuk Klasifikasi Data MRS	30
Tabel 4. 3 Perbandingan Hasil Terbaik dari Metode ICA dengan Metode PCM .	30

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Tumor otak merupakan salah satu bentuk kanker yang sangat berbahaya dan sulit untuk diidentifikasi. Tumor otak dapat dibagi menjadi beberapa jenis, salah satunya adalah *Astrocytoma*. *Astrocytoma* adalah kanker otak yang berasal dari sel pada otak yang disebut *astrocytes*. Berdasarkan WHO, *Astrocytoma* dibagi menjadi beberapa tingkat keganasan. Untuk mengetahui tingkat keganasan dari suatu *Astrocytoma*, diperlukan pengangkatan jaringan yang memiliki kelainan tersebut, yang kemudian akan diperiksa oleh bagian patologi. Selain itu, tingkat keganasan dari *Astrocytoma* dapat diketahui dengan menggunakan *Magnetic Resonance Spectroscopy* (MRS).

Data yang digunakan dalam skripsi ini berasal dari hasil MRS di RSU Pusat Nasional Dr. Cipto Mangunkusumo, Salemba, Jakarta. Dalam skripsi ini akan dilakukan proses klasifikasi terhadap data hasil MRS untuk mengetahui tingkat keganasan dari *Astrocytoma* yang terdiri dari Tingkat rendah (*Low Grade*), Tingkat tinggi (*High Grade*), dan Normal.

Penelitian yang terkait mengenai klasifikasi *Astrocytoma* ini sudah pernah dilakukan dengan menggunakan metode *Fuzzy C-means* (Akmal Fikri. 2009), *Possibilistic C-means* (Anggi P. Wibowo, et al. 2010), *Spherical K-means* dan *Online Spherical K-means* (Ardibian Krismanti, et al. 2010). Dan dalam tugas akhir ini, untuk klasifikasi *Astrocytoma*, akan digunakan metode *Independent Component Analysis*.

Independent Component Analysis (ICA) adalah metode untuk memunculkan faktor atau komponen yang tersembunyi dari sehimpunan variabel random, pengukuran, dan signal.

ICA merupakan perluasan dari *principal component analysis* (PCA) dan analisis faktor. ICA merupakan metode yang lebih kuat dalam menemukan faktor

atau komponen yang tersembunyi di mana metode lain gagal dalam menemukannya.

ICA erat hubungannya dengan masalah *blind source separation* (BSS) atau *blind signal separation*. *Source* disini artinya signal asli (original) atau bisa disebut *independent component*. *Blind* artinya kita hanya mengetahui sedikit sekali, dan asumsi yang lemah pada sumber signal.

Metode ICA masih tergolong penemuan baru. Diperkenalkan pada tahun 1980an, dan pada pertengahan 1990an diperkenalkan algoritma-algoritma baru seperti dalam masalah *cocktail-party effect*. Dalam kurun waktu 20 tahun terakhir, telah dipublikasikan banyak jurnal dan konferensi yang berhubungan dengan metode ini. Penerapan dari metode ini antara lain, Pengambilan Fitur (*Feature Extraction*), *Brain Imaging*, Telekomunikasi, Finansial, Pemisahan Audio, *Clustering*, dsb. (Hyvarinen, et al., 2001)

Dalam penulisan tugas akhir kali ini, metode ICA akan digunakan untuk melakukan *clustering* terhadap data hasil MRS berdasarkan *grade* dari *Astrocytoma*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, masalah yang dihadapi adalah bagaimana merancang suatu program berdasarkan algoritma ICA dan PCM untuk mengelompokkan *Astrocytoma* dari data hasil MRS. Aplikasi akan dibuat menggunakan pemrograman MATLAB 2009a. Perumusan Masalah dapat diberikan sebagai berikut:

1. Bagaimana menerapkan metode ICA pada klasifikasi *Astrocytoma* berdasarkan hasil MRS.
2. Bagaimana menerapkan metode PCM pada klasifikasi *Astrocytoma* berdasarkan hasil MRS.

1.3 Batasan Masalah

Berdasarkan rumusan di atas, maka masalah ini dibatasi pada klasifikasi *Astrocytoma* yang berdasarkan pada data hasil MRS

1.4 Tujuan Penulisan

Sesuai dengan permasalahan yang diangkat dalam penulisan ini, maka tujuan pada penulisan ini adalah sebagai berikut:

- Menerapkan metode ICA dan PCM dalam klasifikasi *Astrocytoma* berdasarkan data hasil MRS
- Melakukan pengujian terhadap keakuratan metode ICA dan PCM dalam melakukan klasifikasi *Astrocytoma*

1.5 Sistematika Penulisan

Penulisan tugas akhir ini tersusun dalam 5 (lima) bab. Bab 1 merupakan pendahuluan. Bab 2 berisi penjelasan mengenai kanker otak, *Magnetic Resonance Spectroscopy* (MRS), dan metode *Possibilitic C-Means* (PCM). Bab 3 membahas lebih lanjut mengenai metode ICA. Bab 4 berisi penerapan dari metode, hasil percobaan, dan tingkat keakuratan dari metode dalam melakukan klasifikasi. Bab 5 merupakan kesimpulan yang didapat dari pembahasan pada bab sebelumnya.

BAB 2

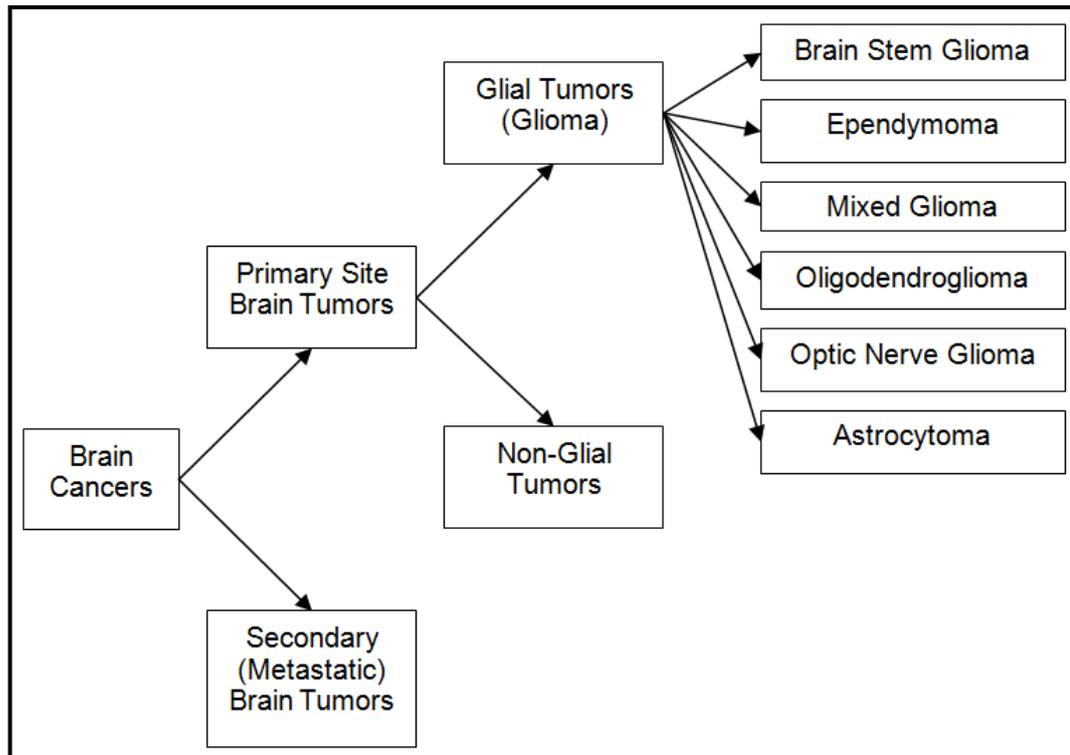
KANKER OTAK DAN METODE POSSIBILISTIC C-MEANS

Pada bab ini akan dibahas mengenai kanker otak dan *Magnetic Resonance Spectroscopy* yang digunakan dalam penerapan metode independent component analysis. Pembahasan dimulai dengan penjelasan kanker otak, dilanjutkan dengan penjelasan *Magnetic Resonance Spectroscopy*, dan pada bagian akhir, akan dibahas metode *Possibilistic C-Means* yang akan digunakan sebagai pembandingan terhadap metode ICA.

2.1 Kanker Otak

Dalam dunia kedokteran, kanker otak merupakan salah satu bentuk kanker yang sangat berbahaya dan sulit diidentifikasi. Secara Histologi kanker otak dibagi menjadi beberapa jenis, seperti pada Gambar 2.1. Tumor glial (selanjutnya disebut dengan *glioma*) merupakan tumor primer otak yang paling sering terjadi pada manusia. *Glioma* dapat dibedakan menjadi beberapa jenis. Pada tugas akhir ini, jenis *glioma* yang digunakan adalah jenis *astrocytoma*. *Astrocytoma* memiliki tingkat keganasan yang dibagi menjadi empat tingkat berdasarkan *World Health Organization* (WHO), yaitu *Grade I (pilocytic astrocytoma)*, *Grade II (fibrillary astrocytoma)*, *Grade III (anaplastic astrocytoma)*, serta *Grade IV (glioblastoma)* (Bernstein. 2000).

Untuk mendiagnosis pasien yang terkena Glioma atau tidak, tidak dapat dilakukan hanya dengan pemeriksaan klinis saja, hal ini dikarenakan gejala klinis yang ditemukan tergantung lokasi dari jenis kanker. Kecepatan pertumbuhan masa kanker dan kecepatan timbulnya gejala tekanan tinggi intrakranial, serta efek dari masa kanker ke jaringan otak dapat menyebabkan kompresi, infasi, dan destruksi dari jaringan otak. Pemeriksaan klinis biasanya dilakukan oleh dokter hanya untuk memastikan apakah ada gangguan pada otak pasien. Dari hasil pemeriksaan klinis, dokter baru melakukan rekomendasi apakah perlu dilakukan pemeriksaan lebih lanjut atau tidak.



[Sumber : Gallas, Sherrie. *Brain Cancer: Statistics and Cluster Investigation*]

Gambar 2. 1 Kanker otak secata Histologis

Untuk mendiagnosis pasien yang terkena *Glioma* atau tidak, tidak dapat dilakukan hanya dengan pemeriksaan klinis saja, hal ini dikarenakan gejala klinis yang ditemukan tergantung lokasi dari jenis kanker. Kecepatan pertumbuhan masa kanker dan kecepatan timbulnya gejala tekanan tinggi intrakranial, serta efek dari masa kanker ke jaringan otak dapat menyebabkan kompresi, infasi, dan destruksi dari jaringan otak. Pemeriksaan klinis biasanya dilakukan oleh dokter hanya untuk memastikan apakah ada gangguan pada otak pasien. Dari hasil pemeriksaan klinis, dokter baru melakukan rekomendasi apakah perlu dilakukan pemeriksaan lebih lanjut atau tidak.

Pemeriksaan lanjutan yang dilakukan oleh dokter sebelum memutuskan untuk melakukan operasi pengangkatan atau tidak ialah pemeriksaan radiologi. Pemeriksaan radiologi yang umum dilakukan adalah *CT-Scan* dan *Magnetic Resonance Imaging (MRI)*. Sarana diagnosis radiologi seperti *CT (Computed Tomography)* dan *MRI* sangat membantu untuk memberikan pertimbangan pada

dokter ahli bedah syaraf dalam melakukan diagnosis glioma, serta memutuskan tindakan lanjutan. Kedua pemeriksaan tersebut hanya menghasilkan gambar atau keadaan anatomi yang normal dan yang ada kelainan, tetapi tidak memperlihatkan metabolisme kimia dari jaringan.

Dalam diagnosa glioma terkadang sulit membedakan jaringan yang terkena kanker glioma dengan kelainan lain, sebagai contoh, glioma derajat keganasan rendah, secara pemeriksaan MRI konvensional sulit dibedakan dengan infark ataupun infeksi, sedangkan penanganan kelainan tersebut sangat berbeda. Oleh karena itu diagnosis secara tepat akan membantu dokter ahli bedah syaraf untuk mengobati penyakit dengan benar.

2.2 Magnetic Resonance Spectroscopy (MRS)

2.2.1 Penjelasan MRS

Proton *Magnetic Resonance Spectroscopy* (MRS) didasari oleh distribusi dan sifat fisika-kimia dari air sehingga dari perbedaan medan magnet dapat dinilai konsentrasi metabolit di dalam jaringan otak. Sehingga dengan MRS, dapat lebih mengetahui kelainan di tingkat metabolisme seluler sebelum terjadinya kelainan yang dapat dilihat dengan MRI konvensional.

MRS dihasilkan sama seperti melakukan pemeriksaan MRI yaitu, tanpa sepengetahuan pasien. Gambar yang didapat dari area yang dipilih (*localizer*), lalu dibuat *volume of interest* (VOI) dan hasil akhirnya akan ditampilkan sebagai grafik atau pengkodean (*coding*) warna dari konsentrasi metabolit setelah alat MRI menekan air (*water suppression*) pada VOI.

Setelah ditentukan lokasi VOI pada gambar MRI, lalu diproduksi spektrum metabolit dari jaringan pada lokasi tersebut. Puncak dari spektrum menggambarkan jumlah proton didalam grup molekul tersebut.

Terdapat tujuh metabolit utama yaitu, *N-Acetyl Aspartate* (NAA), *Creatine* (Cr), *Choline* (Cho), *Lipid* (Lip), *Lactate* (Lac), *myoinositol* (mI), dan *Glutamine-glutamate* (Glx). Metabolit memiliki lokasi atau area tertentu dalam spektrum

dengan satuan *parts per million* (ppm). Lokasi untuk ke-tujuh metabolit tersebut adalah sebagai berikut:

Tabel 2. 1 Ukuran ppm dari metabolit yang digunakan

Metabolit	Ppm
Choline	3.19 – 3.34
Creatine	2.97 – 3.12
NAA	1.96 – 2.13
Lipid	0.81 – 1.3
mI	3.49 – 3.65
Glutamine	2.1 – 2.54
Lactate	1.31 – 1.35

[Sumber : Pandelaki, Jacob. dr, SP.Rad. 2006. *Makalah Seminar S-3*]

2.2.2 Biokimia MRS

Sebelum menganalisa hasil dari spektroskopi otak, terlebih dahulu akan dijelaskan pengertian dari struktur metabolit yang perlu diketahui, yaitu :

(Brandao, 1999; Danielsen, 2004; dan Mukherji, 1998)

- *N-Acetyl Aspartate* (NAA) berada eksklusif di badan sel neuron dan akson sehingga dinyatakan sebagai penanda neuron dan akson. Kerusakan jaringan otak pada tumor atau penyebab lain akan mengakibatkan penurunan puncak NAA.
- *Choline* (Cho) adalah komponen fosfolipid yang merupakan komponen utama membran sel. Perubahan puncak *choline* menggambarkan “*turn over*” membrane sel. Pada proses suatu penyakit akan berhubungan dengan kerusakan membrane sel kemudian terjadi peningkatan sintesa membrane sel sehingga puncak *choline* akan meningkat. Metabolisme *choline* menggambarkan kerusakan mielin atau lesi demielinisasi aktif dan meningkat pada tumor.
- *Creatine* (Cr) adalah indikator untuk pertukaran energi yang berhubungan dengan fosfat energi tinggi. Puncak *Creatine* cenderung meningkat pada

metabolisme jaringan yang rendah dan sebaliknya tetapi umumnya mempunyai puncak yang stabil.

- *Lipid* (Lip) diindikasikan sebagai kerusakan selubung saraf dan adanya nekrosis. Peningkatan lipid dapat terjadi pada inflamasi, abses dan tumor.
- *Lactate* (Lac) dihasilkan dari metabolisme anaerobik (proses glikolisis), yang akan meningkat pada glioma, inflamasi dan infark iskemik
- *Myo-inositol* (mI) adalah khas untuk penanda glial pada astrosit dan hasil dari degradasi mielin (mI merupakan regulator penting dari volume sel). Rasio mI/Cr berhubungan dengan penderajatan tumor, akan meninggi pada tumor derajat rendah. Peningkatan rasio mI/Cr dapat dipikirkan suatu lesi non-neoplasma.
- *Glutamine* merupakan penanda astrosit dan *glutamate* adalah neurotoksin. Jika puncak Glx lebih dari 1/3 puncak NAA menandakan peningkatan Glx. Glx akan meninggi pada lesi non-neoplasma seperti infeksi atau ensefalitis, infark dan ensefalopati hepatis. Pengecualian pada meningioma karena akan terlihat peningkatan Glx.

2.3 Possibilistic C-Means

Clustering telah menjadi metode yang populer untuk *unsupervised pattern recognition*. Salah satu metode yang sangat baik dalam melakukan *clustering* adalah *Fuzzy C-means* (FCM). Sebagian besar *clustering* berbasis *fuzzy* yang diturunkan dari algoritma *Bezdek's Fuzzy C-Means* (Bezdek, 1981). Algoritma FCM tersebut telah berhasil dalam banyak aplikasi. Pada beberapa kasus, metode FCM ini akan memberikan hasil kurang maksimal, untuk mengatasinya dilakukan beberapa modifikasi. Modifikasi ini dikemukakan oleh Raghu Krishnapuram dan M. Keller (1993) dan metode tersebut dinamakan metode *Possibilistic C-means* (PCM).

Secara umum metode PCM menggunakan fungsi objektif dalam bentuk:
(Krishnapuram, 1993)

$$J_w(\mathbf{V}, \mathbf{U}, \mathbf{X}) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^w d(\mathbf{x}_j, \mathbf{v}_i) + \sum_{i=1}^c \eta_i \sum_{j=1}^n (1 - u_{ij})^w \quad (2.3.1)$$

di mana

- n = Banyak data
- c = Banyak *cluster*
- \mathbf{U} = Matriks berukuran $c \times n$ yang disebut matriks *fuzzy C-partition* atau matriks *membership*
- u_{ij} = Elemen dari matriks \mathbf{U}
- \mathbf{X} = Matriks data berukuran $n \times m$, di mana m adalah besarnya dimensi
- \mathbf{x}_j = baris ke- j dari matriks \mathbf{X} , di mana $j = 1, 2, \dots, n$
- \mathbf{v}_i = vektor *centroid* dari cluster ke- i , di mana $i = 1, 2, \dots, c$
- w = Derajat kefuzzy-an, di mana $w \geq 1$
- η_i = Suatu konstanta positif, di mana $i = 1, 2, \dots, c$

Pengalokasian data ke dalam *cluster* pada metode PCM didasarkan pada perbandingan jarak antara *centroid* dengan data yang ada. Data dialokasikan ke dalam cluster yang memiliki *centroid* terdekat dengan data tersebut.

Clustering dalam PCM dibagi menjadi tiga tahap, tahap pertama disebut tahap *preprocessing*, tahap kedua adalah tahap *training*, kemudian tahap ketiga adalah tahap *testing*.

Dalam tahap *preprocessing*, tentukan banyaknya data training yang akan digunakan, yang kemudian akan digunakan sebagai data input untuk tahap *training*.

Dalam tahap kedua akan digunakan algoritma PCM untuk *training* yang secara umum dapat dilihat sebagai berikut :

Input : $\mathbf{X}, w, c, K, \text{maxiter}, \varepsilon$

Output : \mathbf{v}_i

1. Hitung *centroid* \mathbf{v}_i untuk masing-masing *cluster* dengan menggunakan persamaan

$$\mathbf{v}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbf{x}_{ki} \quad (2.3.3)$$

2. Inisialisasi (acak) matriks *membership* $\mathbf{U} = [u_{ij}]$, $i = 1, \dots, c, j = 1, \dots, n_i$
3. Hitung nilai η_i dengan menggunakan persamaan

$$\eta_i = K \frac{\sum_{j=1}^n u_{ij}^w d_{ij}^2}{\sum_{j=1}^n u_{ij}^w} \quad (2.3.4)$$

di mana K adalah *learning state constant*, dan

$$d_{ij}^2 = (\mathbf{x}_j - \mathbf{v}_i)^T (\mathbf{x}_j - \mathbf{v}_i) \quad (2.3.5)$$

4. Perbaharui *centroid* \mathbf{v}_i dengan menggunakan

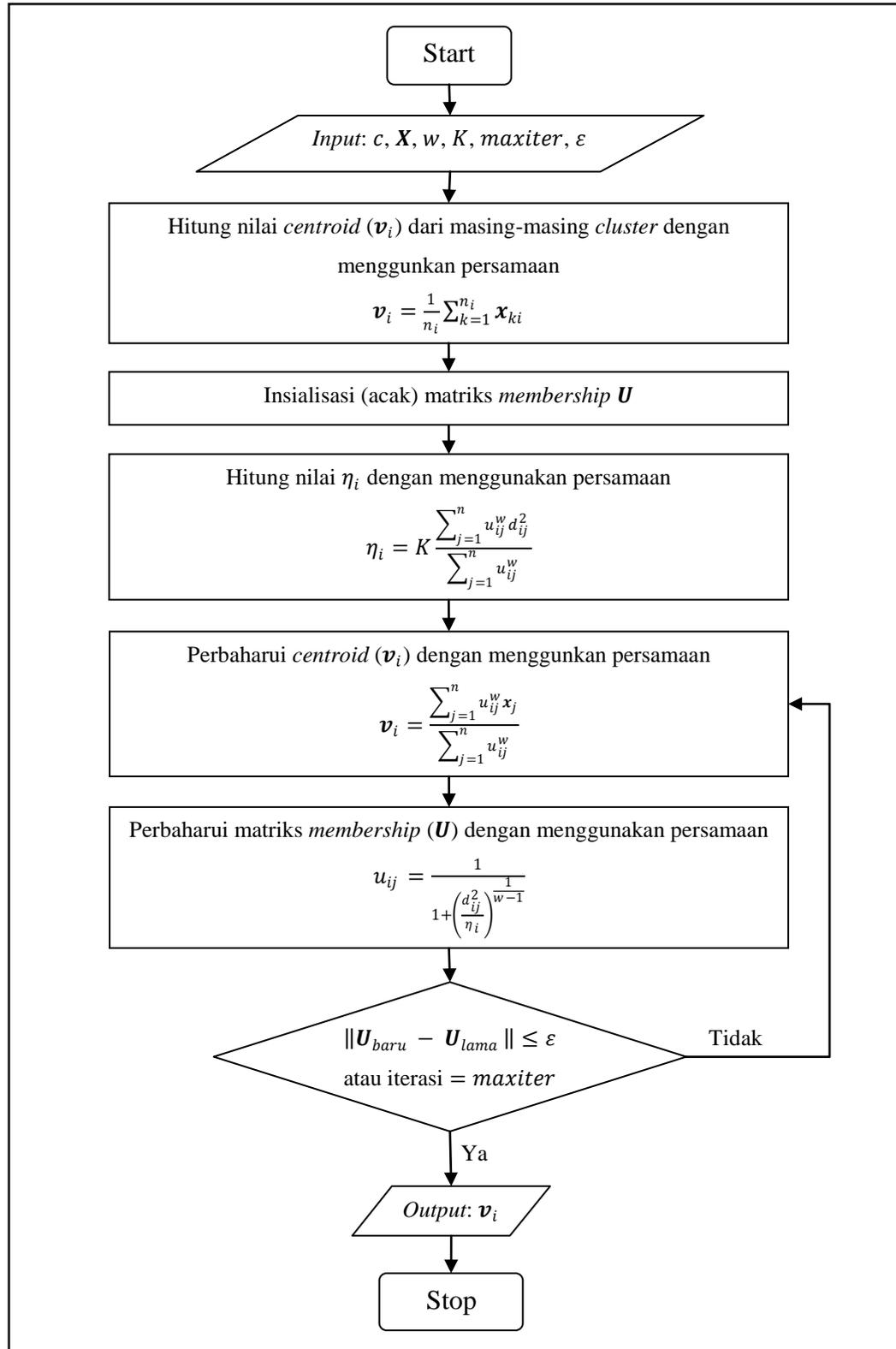
$$\mathbf{v}_i = \frac{\sum_{j=1}^n u_{ij}^w \mathbf{x}_j}{\sum_{j=1}^n u_{ij}^w} \quad (2.3.6)$$

5. Perbaharui matriks *membership* \mathbf{U} dengan menggunakan

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_i}\right)^{\frac{1}{w-1}}} \quad (2.3.7)$$

6. Periksa apakah $\|\mathbf{U}_{baru} - \mathbf{U}_{lama}\| \leq \varepsilon$ atau iterasi mencapai maksimum (*maxiter*). Jika tidak kembali ke langkah 4

Diagram alurnya (*flowchart*) dapat dilihat pada Gambar 2.2.



Gambar 2. 2 Flowchart PCM untuk training

Dalam tahap ketiga dilakukan *testing*. Dalam tahap ini sisa data yang tidak digunakan dalam *training*, akan digunakan sebagai data uji. Secara umum algoritma untuk *testing* dapat dilihat sebagai berikut :

Input : $\mathbf{v}_i, \mathbf{X}_{test}$

Output : akurasi

1. Hitung jarak data uji ke *centroid* dari tiap *cluster* dengan menggunakan persamaan

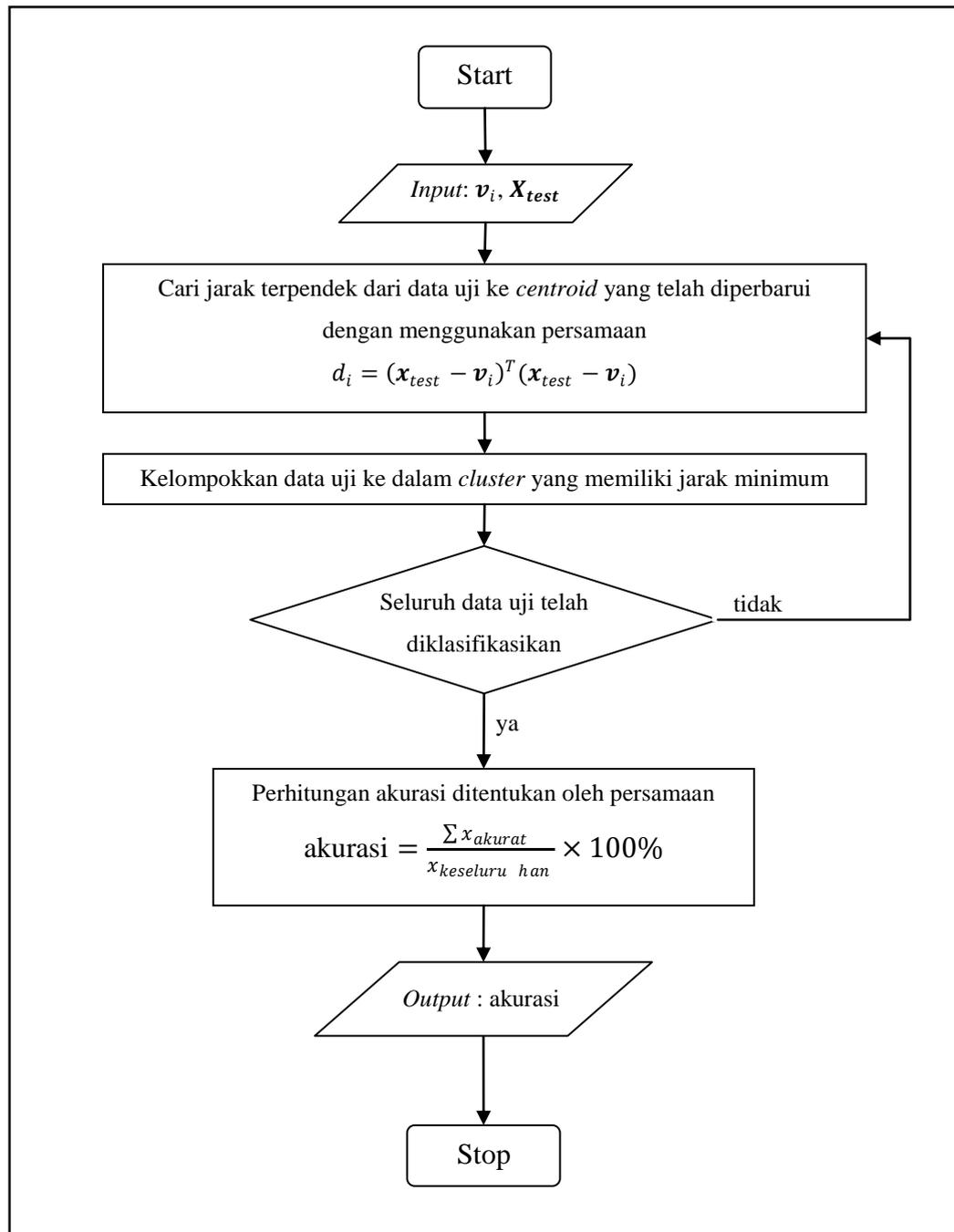
$$d_i = (\mathbf{x}_{test} - \mathbf{v}_i)^T (\mathbf{x}_{test} - \mathbf{v}_i) \quad (2.3.8)$$

2. Kemudian kelompokkan data uji ke dalam *cluster* yang memiliki jarak paling minimum
3. Ulangi langkah 1 hingga seluruh data uji sudah diklasifikasikan
4. Hitung akurasi dengan menggunakan persamaan

$$\text{akurasi} = \frac{\sum x_{akurat}}{x_{keseluruhan}} \times 100\% \quad (2.3.9)$$

di mana x_{akurat} menyatakan data uji yang sesuai dengan kelasnya dan $x_{keseluruhan}$ menyatakan total keseluruhan data uji.

Diagram alurnya dapat dilihat pada gambar 2.3.



Gambar 2. 3 Flowchart testing pada PCM

BAB 3

INDEPENDENT COMPONENT ANALYSIS (ICA)

Pada bab ini akan dibahas teori-teori yang berhubungan dengan metode *Independent Component Analysis* serta algoritma yang digunakan. Pembahasan akan dimulai dengan penjelasan ruang hasil kali dalam dan norm, serta penjelasan nilai eigen dan vektor eigen. Kemudian akan dibahas mengenai beberapa definisi statistik, yaitu ekspektasi, matriks korelasi, dan matriks kovariansi. Sedangkan dalam bagian akhir akan dibahas lebih lanjut mengenai *independent component analysis*.

3.1 Ruang Hasil Kali Dalam dan Norm

3.1.1 Hasil Kali Dalam

Definisi 3.1.1 (Anton, 2000)

Suatu hasil kali dalam pada suatu ruang vektor real V adalah suatu fungsi yang menghubungkan setiap pasangan vektor \mathbf{u} dan \mathbf{v} dalam V dengan suatu bilangan real $\langle \mathbf{u}, \mathbf{v} \rangle$, sedemikian sehingga untuk semua vektor \mathbf{u}, \mathbf{v} , dan \mathbf{w} dalam V dan semua skalar k , berlaku sifat berikut :

- $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$; Kesimetrian
- $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$; Penjumlahan
- $\langle k\mathbf{u}, \mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle$; Kehomogenan
- $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, dan $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ jika dan hanya jika $\mathbf{v} = \mathbf{0}$; Kepositifan

Suatu ruang vektor real dengan suatu hasil kali dalam disebut suatu ruang hasil kali dalam real.

3.1.2 Norm Vektor

Definisi 3.1.2 (Anton, 2000)

Jika V adalah suatu ruang hasil kali dalam, maka norm suatu vektor \mathbf{u} di V dinyatakan dengan $\|\mathbf{u}\|$ dan didefinisikan sebagai.

$$\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{\frac{1}{2}} \quad (3.1.1)$$

Sedangkan, jarak antara dua vektor, \mathbf{u} dan \mathbf{v} yang dinyatakan oleh $d(\mathbf{u}, \mathbf{v})$, didefinisikan sebagai

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| \quad (3.1.1)$$

Berdasarkan definisi, dapat dilihat jarak antara dua buah vektor (titik) akan sama dengan norm dari selisih dua vektor (titik) tersebut.

Beberapa aturan norm vektor dapat dilihat pada tabel 3.1.

Tabel 3. 1 Aturan Norm Vektor

No.	Fungsi Norm	Formula
1	1-norm	$\ \mathbf{x}\ _1 = \sum_{i=1}^n x_i $
2	2-norm	$\ \mathbf{x}\ _2 = \sqrt{\sum_{i=1}^n x_i ^2}$
3	∞ -norm	$\ \mathbf{x}\ _\infty = \max_k x_k $
4	Frobenius	$\ \mathbf{x}\ _{fro} = \sqrt{\sum diag(\mathbf{x}^T \mathbf{x})}$

Dalam tugas akhir ini norm vektor yang akan digunakan adalah norm vektor *Euclidean* (2-norm).

3.1.3 Norm Matriks

Definisi 3.1.2 (Roger, 2005)

Misal $M_n(\mathbb{R})$ adalah himpunan matriks $n \times n$ atas \mathbb{R} . Fungsi $\|\cdot\|: M_n \rightarrow \mathbb{R}$ disebut norm matriks jika untuk setiap $\mathbf{A}, \mathbf{B} \in M_n$ memenuhi sifat-sifat berikut :

- $\|\mathbf{A}\| \geq 0$; Nonnegatif
- $\|\mathbf{A}\| = 0 \leftrightarrow \mathbf{A} = \mathbf{0}$; Positif
- $\|c\mathbf{A}\| = |c|\|\mathbf{A}\|$ untuk setiap skalar kompleks c ; *Homogeneous*
- $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$; Ketaksamaan segi tiga
- $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$; *Submultiplicative*

Beberapa contoh aturan norm berdasarkan Roger (2005) dapat dilihat pada tabel 3.2.

Tabel 3. 2 Contoh aturan Norm Matriks

No.	Fungsi Norm	Formula
1	L_1 atau <i>Hamming norm</i>	$\ \mathbf{A}\ _1 = \sum_{i,j=1}^n a_{ij} $
2	L_2 atau <i>Euclidean norm</i>	$\ \mathbf{A}\ _2 = \left(\sum_{i,j=1}^n a_{ij} ^2 \right)^{\frac{1}{2}}$

[Sumber : Roger, 2005]

Dalam tugas akhir ini, norm matriks yang digunakan adalah norm matriks *Euclidean* (L_2).

3.2 Nilai dan Vektor Eigen

Definisi 3.2 (Anton, 2000)

Jika \mathbf{A} adalah suatu matriks $n \times n$, maka vektor tak-nol \mathbf{x} pada \mathbf{R}^n disebut suatu vektor-eigen dari \mathbf{A} jika \mathbf{Ax} adalah suatu penggandaan skalar dari \mathbf{x} ; yaitu,

$$\mathbf{Ax} = \lambda \mathbf{x} \quad (3.2.1)$$

Di mana skalar λ disebut nilai-eigen dari \mathbf{A} , dan \mathbf{x} disebut suatu vektor-eigen dari \mathbf{a} yang berpadanan dengan λ .

3.3 Statistik Dasar

Pada bagian ini akan dibahas beberapa definisi statistik dasar yang digunakan dalam metode ICA.

3.3.1 Definisi Ekspektasi

Definisi 3.3.1 (Robert 2005)

Misal x adalah suatu variabel random. Jika x merupakan variabel random yang kontinu dengan pdf $f(x)$ dan

$$\int_{-\infty}^{\infty} |x|f(x) dx < \infty \quad (3.3.1)$$

Maka ekspektasi dari X adalah

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx \quad (3.3.2)$$

Definisi 3.3.2 (Hyvarinen 2001)

Misal $\mathbf{g}(\mathbf{X})$ menyatakan suatu kuantitas yang berasal dari suatu vektor random \mathbf{X} . Kuantitas dari $\mathbf{g}(\mathbf{X})$ dapat merupakan suatu skalar, vektor, atau sebuah matriks. Ekspektasi dari $\mathbf{g}(\mathbf{X})$ dinyatakan oleh $E\{\mathbf{g}(\mathbf{X})\}$, dan didefinisikan dengan

$$E\{\mathbf{g}(\mathbf{X})\} = \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{X})p_{\mathbf{X}}(\mathbf{X})d\mathbf{X} \quad (3.3.3)$$

3.3.2 Estimasi dari Ekspektasi

Biasanya, *probability density* dari suatu vektor random \mathbf{X} tidak diketahui, tetapi seringkali terdapat himpunan sampel $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K$ dari \mathbf{X} . Dengan menggunakan sampel tersebut, ekspektasi dapat diestimasi dengan (Hyvarinen, 2001)

$$E\{\mathbf{g}(\mathbf{X})\} \approx \frac{1}{K} \sum_{j=1}^K \mathbf{g}(\mathbf{x}_j) \quad (3.3.4)$$

3.3.3 Definisi Matriks Korelasi

Korelasi r_{ij} antara komponen ke- i dan ke- j dari vektor random \mathbf{x} , di mana $\mathbf{x} = (x_1, \dots, x_k)$ adalah

$$r_{ij} = E\{x_i x_j\} = \int_{-\infty}^{\infty} x_i x_j p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j p_{x_i, x_j}(x_i, x_j) dx_j dx_i \quad (3.3.5)$$

Misal diberikan $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n]$, matriks korelasi $n \times n$ dinyatakan sebagai

$$\mathbf{R}_{\mathbf{X}} = E\{\mathbf{X}\mathbf{X}^T\} \quad (3.3.6)$$

di mana seluruh korelasi r_{ij} merupakan elemen dari baris ke- i dan kolom ke- j dari $\mathbf{R}_{\mathbf{X}}$. Matriks korelasi $\mathbf{R}_{\mathbf{X}}$ memiliki beberapa sifat penting, yaitu (Hyvarinen, 2001):

- Matriks $\mathbf{R}_{\mathbf{X}}$ merupakan matriks simetris: $\mathbf{R}_{\mathbf{X}} = \mathbf{R}_{\mathbf{X}}^T$.
- Matriks $\mathbf{R}_{\mathbf{X}}$ merupakan matriks *semidefinite* positif:

$$\mathbf{y}^T \mathbf{R}_{\mathbf{X}} \mathbf{y} \geq 0 \quad (3.3.7)$$

untuk sembarang vektor \mathbf{y} . Dalam praktek biasanya $\mathbf{R}_{\mathbf{X}}$ adalah definit positif, artinya untuk sembarang vektor $\mathbf{y} \neq 0$.

- Setiap nilai eigen dari $\mathbf{R}_{\mathbf{X}}$ real dan non-negatif (jika $\mathbf{R}_{\mathbf{X}}$ definit positif, maka nilai eigennya bernilai positif). Setiap vektor eigen dari $\mathbf{R}_{\mathbf{X}}$ real dan selalu dapat dipilih sedemikian sehingga vektor eigen tersebut saling orthonormal.

3.3.4 Definisi Matriks Kovariansi

Matriks kovariansi didefinisikan dengan

$$\mathbf{C}_X = E\{(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T\} \quad (3.3.8)$$

di mana elemen dari matriks \mathbf{C}_X adalah

$$c_{ij} = E\{(x_i - \bar{X}_i)(x_j - \bar{X}_j)\} \quad (3.3.9)$$

yang disebut juga kovariansi. Matriks kovariansi \mathbf{C}_X memenuhi sifat-sifat yang sama dengan matriks korelasi \mathbf{R}_X . Dengan menggunakan sifat-sifat dari ekspektasi, dapat diperoleh

$$\mathbf{R}_X = \mathbf{C}_X + \bar{\mathbf{X}}\bar{\mathbf{X}}^T \quad (3.3.10)$$

Jika mean vector $\bar{\mathbf{X}} = \mathbf{0}$, matriks korelasi dan matriks kovariansi akan bernilai sama. Dalam penerapan *independent component analysis*, hal ini sering terjadi karena dalam tahap pra-pengolahan, data yang digunakan dapat dijadikan *zero mean*, dengan mengurangkan *mean* vektor dari data. (Hyvarinen, et al. 2001)

3.4 Independent Component Analysis

Pada bagian ini akan dibahas definisi dari *independent component analysis* dan algoritma yang digunakan dalam tugas akhir ini. Sebelumnya akan dibahas mengenai *Blind Source Separation* (BSS) yang merupakan konsep dasar dari metode *Independent Component Analysis*.

3.4.1 Blind Source Separation (BSS)

Misalkan dalam suatu ruangan terdapat empat orang berbicara secara bersamaan. Misalkan di dalam ruangan tersebut juga terdapat empat buah mikropon yang terletak diempat tempat yang berbeda seperti pada Gambar 3.1. Mikropon tersebut memberikan empat hasil rekaman suara (*recorded time signal*), yang dapat dinyatakan dengan $x_1(t)$, $x_2(t)$, $x_3(t)$, dan $x_4(t)$, di mana x_1, x_2, x_3 , dan x_4 merupakan amplitudo (*amplitude*), dan t adalah indeks waktu (*time index*). Tiap hasil rekaman merupakan *weighted sum* dari suara yang dihasilkan empat pembicara, di mana suara dari empat pembicara tersebut dinyatakan dengan $s_1(t)$, $s_2(t)$, $s_3(t)$, dan $s_4(t)$.

Yang akan dilakukan disini adalah bagaimana mendapatkan sumber sinyal asli (suara dari pembicara) dengan hanya menggunakan sinyal dari hasil observasi (mikropon). Masalah ini dinamakan *Blind Source Separation (BSS)*. *Blind* disini artinya hanya sedikit informasi mengenai sumber sinyal asli. Masalah tersebut dapat dinyatakan sebagai persamaan linear berikut:

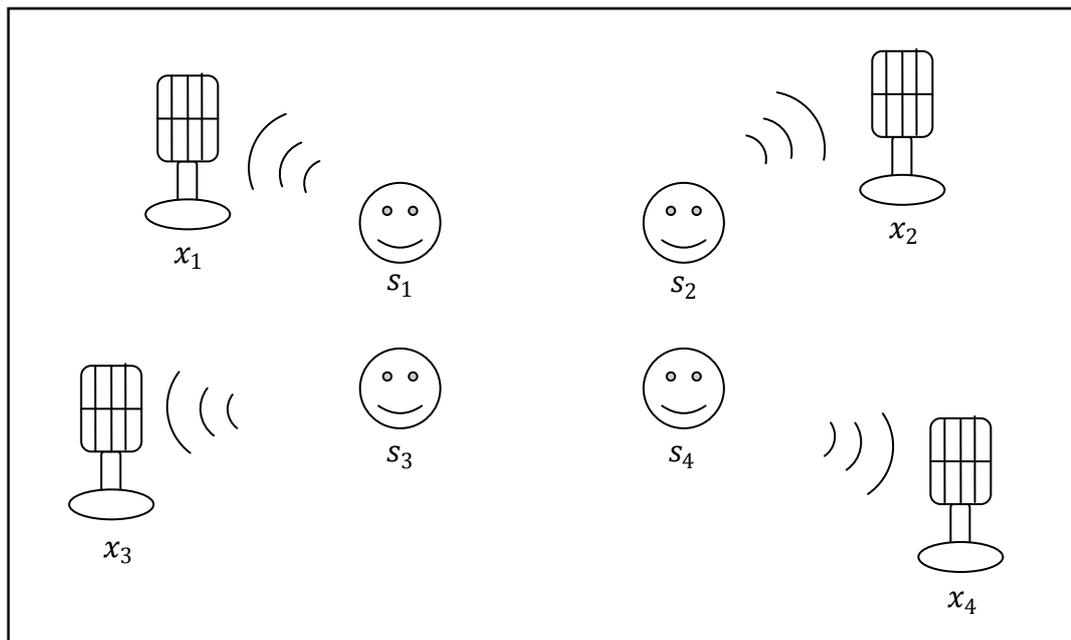
$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) + a_{14}s_4(t) \quad (3.4.1)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) + a_{24}s_4(t) \quad (3.4.2)$$

$$x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) + a_{34}s_4(t) \quad (3.4.3)$$

$$x_4(t) = a_{41}s_1(t) + a_{42}s_2(t) + a_{43}s_3(t) + a_{44}s_4(t) \quad (3.4.4)$$

di mana a_{ij} , $i, j = 1, \dots, 4$ adalah suatu parameter yang bergantung pada jarak mikropon dari pembicara. Nilai a_{ij} dan $s_i(t)$ tidak diketahui.



Gambar 3. 1 Ilustrasi masalah Blind Source Separation (BSS)

3.4.2 Definisi dasar ICA

Dari masalah BSS, didapatkan definisi dasar dari ICA. Misalkan terdapat n variabel *random* x_1, \dots, x_n , yang dimodelkan sebagai kombinasi linear dari n variabel *random* s_1, \dots, s_n :

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n, \text{ untuk setiap } i = 1, \dots, n \quad (3.4.5)$$

di mana $a_{ij}, i, j = 1, \dots, n$ adalah koefisien bilangan real dan s_i saling independent.

Model ICA di atas dapat ditulis dengan menggunakan notasi vektor sebagai berikut. Misalkan vektor *random* \mathbf{x} di mana elemen dari vektor tersebut adalah x_1, \dots, x_n . Demikian juga dengan vektor *random* \mathbf{s} di mana elemennya adalah s_1, \dots, s_n . Dan matriks \mathbf{A} di mana elemennya adalah a_{ij} . Dengan menggunakan notasi vektor tersebut, model ICA dapat ditulis :

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \text{noise} \quad (3.4.6)$$

atau

$$\mathbf{x} = \sum_{i=1}^n \mathbf{a}_i s_i + \text{noise} \quad (3.4.7)$$

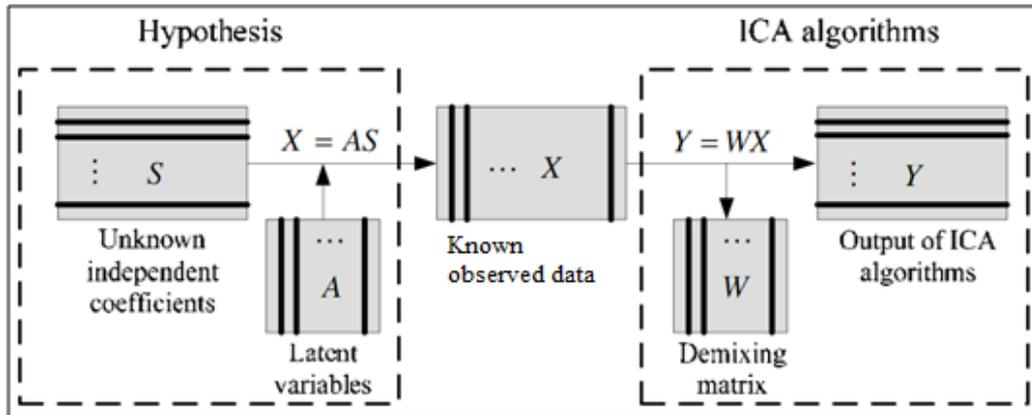
di mana vektor \mathbf{s} dan \mathbf{A} tidak diketahui.

Tanpa mengetahui sumber sinyal \mathbf{s} dan mixing matrix \mathbf{A} , kita ingin mendapatkan sumber sinyal dari sinyal yang diobservasi \mathbf{x} dengan transformasi berikut

$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad (3.4.8)$$

di mana $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_n]^T$ merupakan estimasi dari sumber sinyal \mathbf{s} , dan \mathbf{W} adalah *unmixing matrix*. Gambar 3.2 menunjukkan kerangka teoritis dari algoritma untuk menyelesaikan model ICA.

Terdapat beberapa algoritma untuk menyelesaikan model ICA, seperti algoritma *FastICA* (Hyvarinen, 2001) , *Infomax* (Sejnowski, 1995).



[Sumber : Kong, et al. *A review of independent component analysis application to microarray gene expression data*. 2008]

Gambar 3. 2 Kerangka Teoritis dari Algoritma untuk menyelesaikan model ICA

3.4.3 Infomax Learning Algorithm

Dalam tugas akhir ini, akan digunakan *Infomax Learning Algorithm* yang diperkenalkan oleh Bell dan Sejnowski (1995), dengan *natural gradient learning rule*

$$\mathbf{W}(t + 1) = \mathbf{W}(t) + \mu(\mathbf{I} - \phi(\mathbf{y})\mathbf{y}^T)\mathbf{W}(t) \quad (3.4.8)$$

di mana

\mathbf{W} ; *Unmixing matrix* yang diestimasi

\mathbf{y} ; *Source signal* yang diestimasi

μ ; *Learning state constant*

\mathbf{I} ; Matriks Identitas

dan $\phi(\cdot)$ dinamakan *scorefunction* yang dinyatakan dengan :

$$\phi(\mathbf{y}) = \left[\left(-\frac{f''(y_1)}{f'(y_1)} \right), \dots, \left(-\frac{f''(y_M)}{f'(y_M)} \right) \right]^T \quad (3.4.9)$$

di mana $f(y)$ adalah fungsi *Sigmoid*, yang memiliki bentuk

$$f(y) = \tanh(y) \quad (3.4.10)$$

Clustering dalam ICA dibagi menjadi tiga tahap, tahap pertama dinamakan tahap *preprocessing*. Tahap *preprocessing* secara umum diberikan sebagai berikut:

Input : $c; \mathbf{X}_i, i = 1, 2, \dots, c$

Output : $\tilde{\mathbf{X}}$

1. Hitung *mean* dari data untuk setiap *cluster* dengan menggunakan

$$\bar{\mathbf{X}}_i = E\{\mathbf{X}_i\} = \frac{1}{n_i} \sum_{k=1}^{n_i} \mathbf{x}_{ki} \quad (3.4.13)$$

2. Kurangkan tiap baris dari data dengan *mean* dari data tersebut. Langkah ini dinamakan langkah *mean-subtracted*
3. Hitung kovariansi dari matriks \mathbf{X}_i dengan menggunakan

$$\text{Cov}(\mathbf{X}_i) = E\{\mathbf{X}_i \mathbf{X}_i^T\} \quad (3.4.14)$$

4. Lakukan data *whitening* dengan menggunakan

$$\tilde{\mathbf{X}}_i = \mathbf{E}_i \mathbf{D}_i^{-\left(\frac{1}{2}\right)} \mathbf{E}_i^T \mathbf{X}_i \quad (3.4.15)$$

di mana \mathbf{E} adalah matriks orthogonal dari vektor eigen dari $\text{Cov}(\mathbf{X}_i)$, \mathbf{D} adalah matriks diagonal dari nilai eigen dari $\text{Cov}(\mathbf{X}_i)$, dan

$$\mathbf{D}^{-\left(\frac{1}{2}\right)} = \text{diag}(e_1^{-\left(\frac{1}{2}\right)}, e_2^{-\left(\frac{1}{2}\right)}, \dots, e_n^{-\left(\frac{1}{2}\right)}) \quad (3.4.16)$$

Diagram alurnya pada dapat dilihat pada gambar 3.3.

Setelah tahap *preprocessing*, selanjutnya akan diestimasi *unmixing matrix* \mathbf{W} dengan menggunakan algoritma *natural gradient*. Sebelumnya data yang sudah di-*whitening*, dibagi menjadi dua, pertama sebagai data *training*, dan yang kedua sebagai data uji. Dalam tahap kedua ini akan digunakan data *training*. Algoritma *natural gradient* secara lengkap diberikan sebagai berikut :

Input : $\tilde{\mathbf{X}}; \text{maxiter}; \mu; \varepsilon$

Output : $\mathbf{W}; \mathbf{y}$

1. Inisialisasi awal *unmixing matrix* (\mathbf{W})
2. Hitung *source signal* (\mathbf{y}) yang diestimasi, dengan menggunakan

$$\mathbf{y} = \mathbf{W}\tilde{\mathbf{X}} \quad (3.4.17)$$

3. Perbarui *unmixing matrix* dengan menggunakan

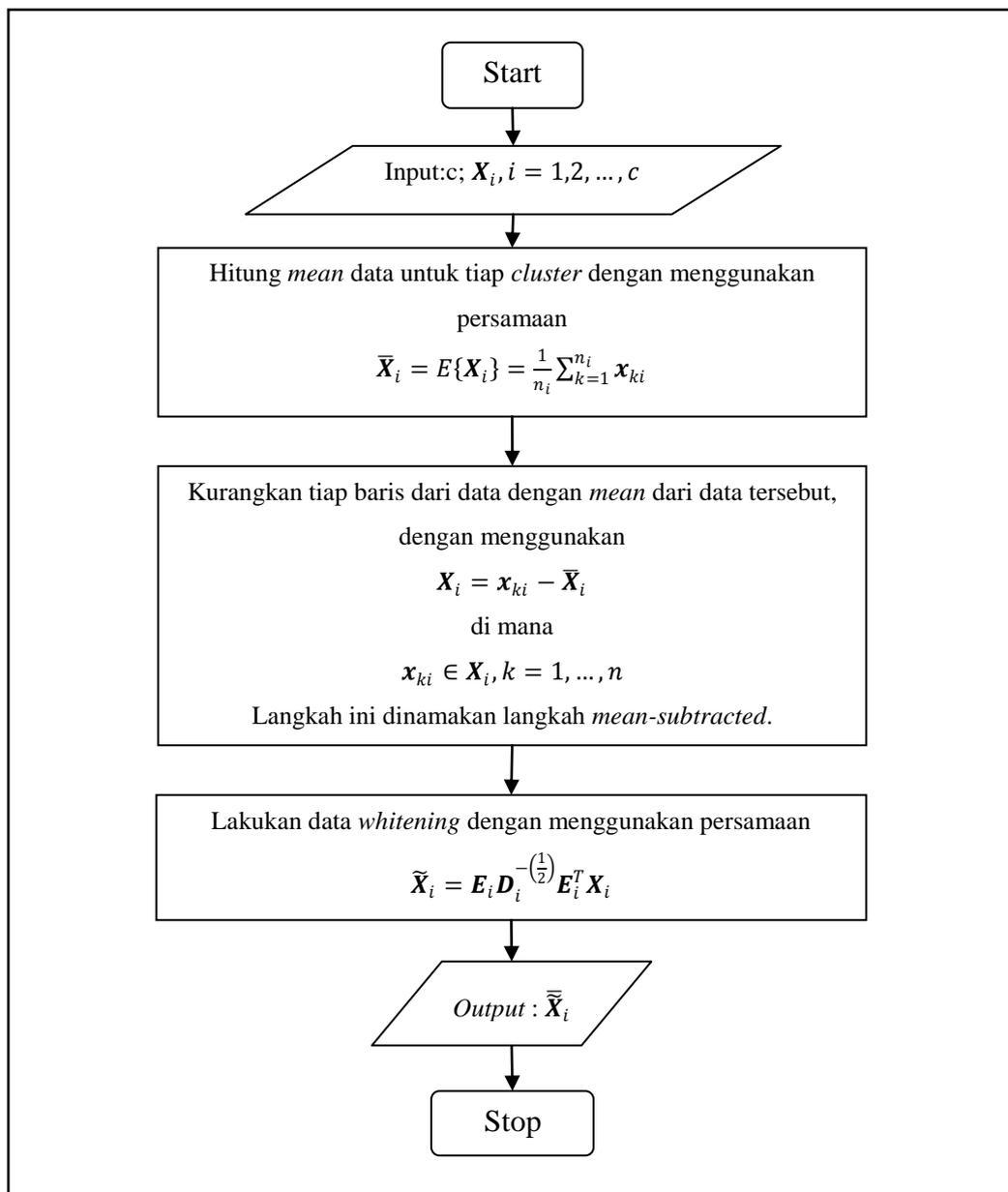
$$\mathbf{W}_{baru} = \mathbf{W}_{lama} + \mu(\mathbf{I} - \phi(\mathbf{y})\mathbf{y}^T)\mathbf{W}_{lama} \quad (3.4.18)$$

4. Periksa

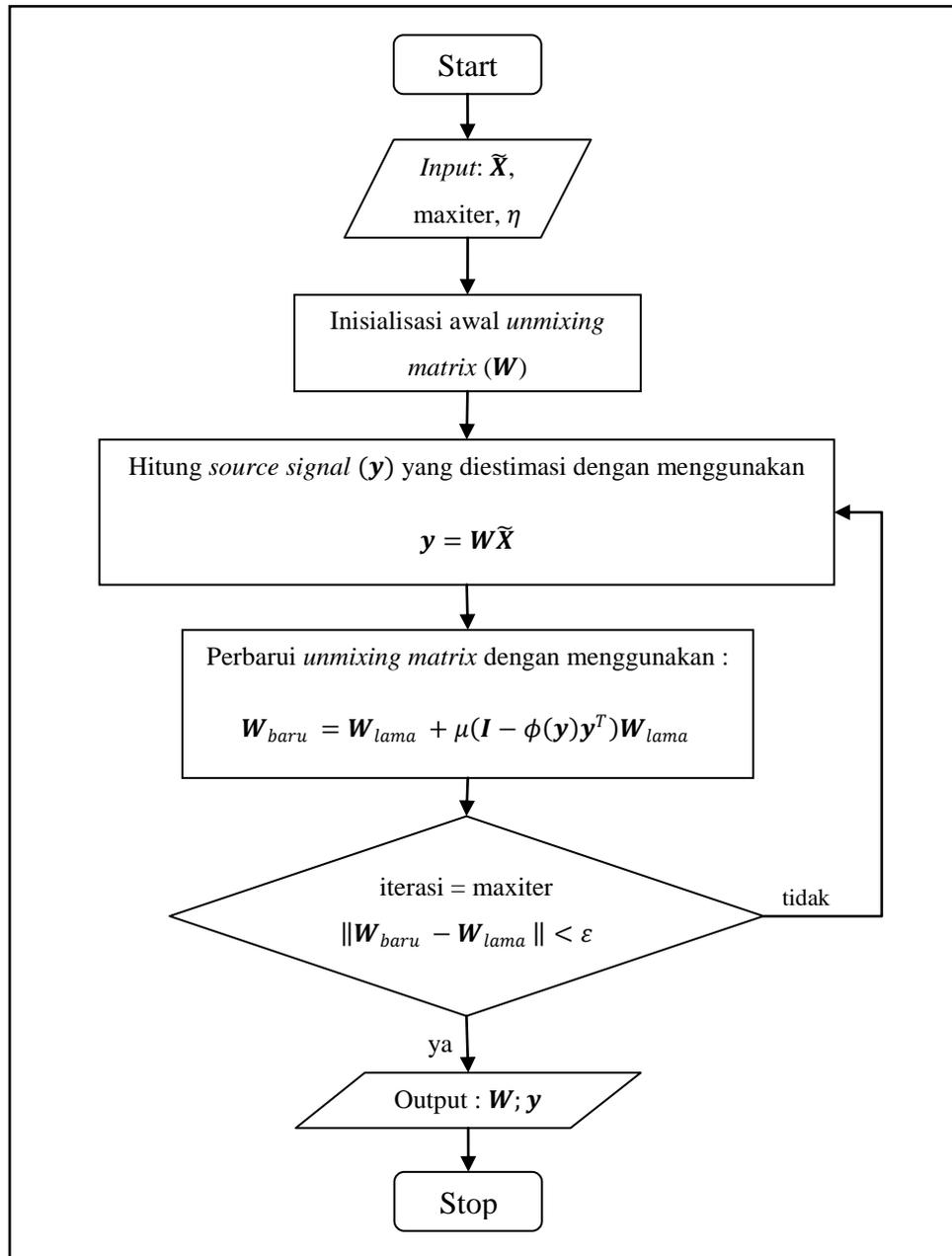
$$\|\mathbf{W}_{baru} - \mathbf{W}_{lama}\| < \varepsilon \quad (3.4.19)$$

dan apakah iterasi telah mencapai maksimum (*maxiter*). Jika tidak, ulangi langkah 2

Diagram alurnya pada dapat dilihat pada gambar 3.4.



Gambar 3. 3 Flowchart tahap preprocessing ICA



Keterangan :

- \tilde{X} menyatakan notasi dari data yang telah di-*whitening*
- Data yang digunakan adalah data *training*

Gambar 3. 4 Flowchart Algoritma Infomax dengan Natural Gradient Learning

Dalam tahap ketiga dilakukan *testing*. Dalam tahap ini sisa data yang tidak digunakan dalam tahap kedua, akan digunakan sebagai data untuk pengujian (*testing*). Secara umum algoritma untuk *testing* diberikan sebagai berikut :

Input : $\tilde{\mathbf{x}}_{test}; \mathbf{W}_i; \mathbf{y}_i$

Output : akurasi

1. Hitung \mathbf{y}_i^{test} , dengan menggunakan

$$\mathbf{y}_i^{test} = \mathbf{W}_i \tilde{\mathbf{x}}_{test} \quad (3.4.20)$$

2. Hitung *mean* dari \mathbf{y}_i dengan menggunakan

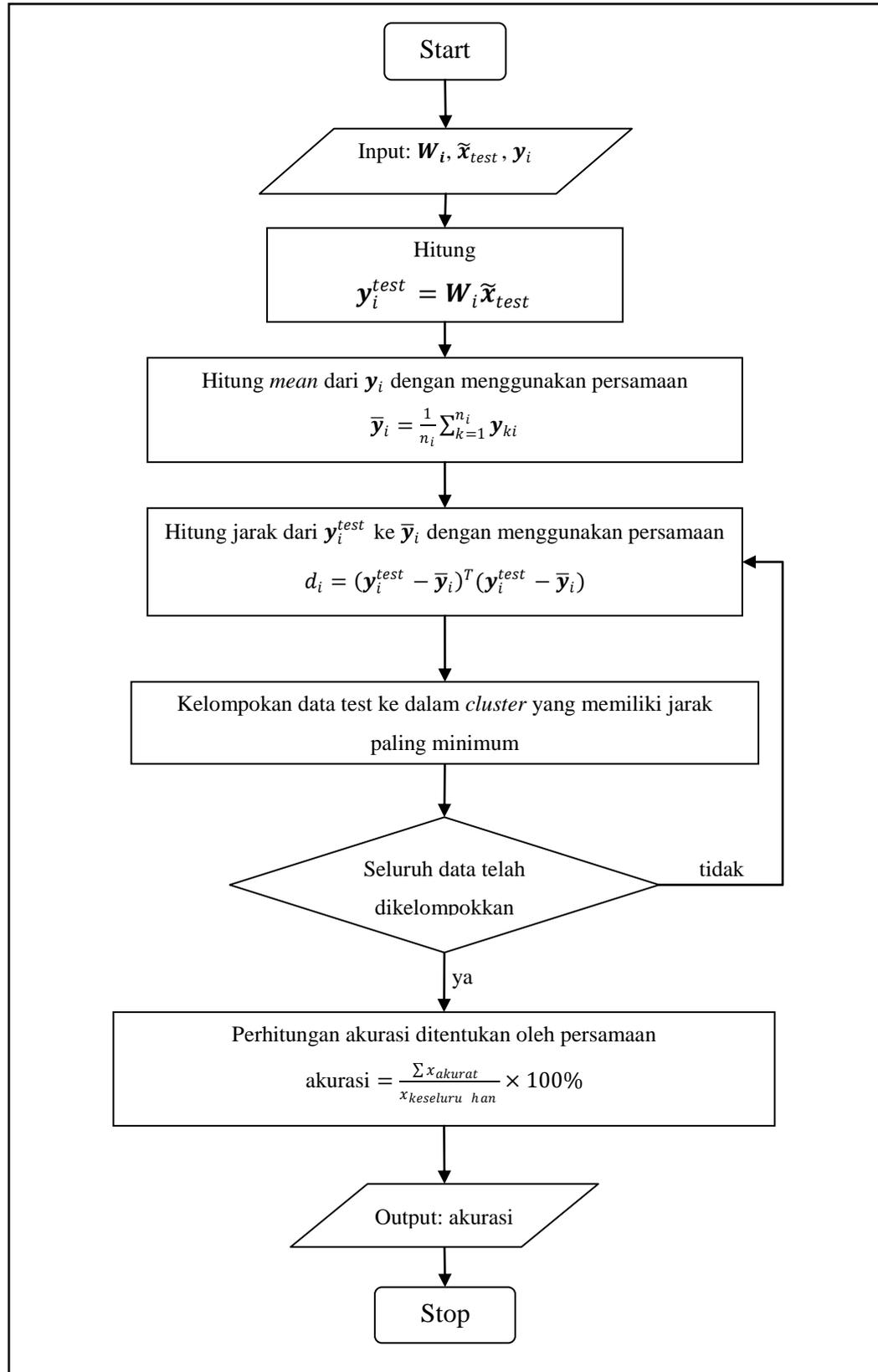
$$\bar{\mathbf{y}}_i = \frac{1}{n_{ki}} \sum_{k=1}^{n_{ki}} \mathbf{y}_{ki} \quad (3.4.21)$$

3. Hitung jarak dari \mathbf{y}_i^{test} ke $\bar{\mathbf{y}}_i$ dengan menggunakan

$$d_i = (\mathbf{y}_i^{test} - \bar{\mathbf{y}}_i)^T (\mathbf{y}_i^{test} - \bar{\mathbf{y}}_i) \quad (3.4.22)$$

4. Kelompokkan data test ke dalam *cluster* yang memiliki jarak paling minimum
5. Ulangi langkah 1 untuk data test berikutnya, hingga seluruh data test sudah diklasifikasikan
6. Hitung akurasi dengan menggunakan persamaan (2.3.9)

Diagram alurnya dapat dilihat pada pada gambar 3.5.



Gambar 3. 5 Flowchart data testing pada ICA

BAB 4

HASIL DAN PEMBAHASAN

4.1 Tinjauan Data

Data yang digunakan dalam tugas akhir ini adalah data hasil MRS yang diperoleh dari beberapa rumah sakit di Jakarta, diantaranya adalah RSUD Pusat Nasional Dr. Cipto Mangunkusumo dan Rumah Sakit Pantai Indah Kapuk. Pada data hasil MRS tersebut, ada tujuh variabel yang akan diukur dalam menentukan keganasan dari *astrocytoma*. Data yang didapat, dibentuk menjadi sebuah vektor $\mathbf{x} = [x_1 \ \cdots \ x_7]$, di mana

- x_1 = Konsentrasi Choline
- x_2 = Konsentrasi Cr atau Creatine
- x_3 = Konsentrasi Naa
- x_4 = Konsentrasi Lipid
- x_5 = Konsentrasi MI
- x_6 = Konsentrasi Glutamin
- x_7 = Konsentrasi Lactat

data tersebut akan diklasifikasi ke dalam tiga kelas, yaitu normal, *low grade*, atau *high grade*. Data yang ada berjumlah 32 data, yaitu 16 data normal, 8 data *low grade*, dan 8 data *high grade*.

Pengambilan data MRS dilakukan dengan mengolah data MRS menggunakan aplikasi yang digunakan di rumah sakit. Aplikasi yang digunakan berasal dari mesin *General Electric* (GE). Penentuan besaran ppm dari metabolit yang diatur pada mesin GE adalah sebagai berikut :

Tabel 4. 1 Ukuran ppm dari metabolit yang diatur pada mesin GE

Metabolit	Ppm
Choline	3.19 – 3.34
Creatine	2.97 – 3.12
NAA	1.96 – 2.13
Lipid	0.81 – 1.3
mI	3.49 – 3.65
Glutamine	2.1 – 2.54
Lactate	1.31 – 1.35

[Sumber : Pandelaki, Jacub. *Seminar S-3*, 2006]

Pengambilan data dilakukan dengan mengambil data pada voxel yang dicurigai mengandung kelainan, sedangkan untuk data normal, didapatkan dengan mengambil data pada voxel kontralateralnya. Pemilihan voxel dilakukan berdasarkan petunjuk dokter ahli radiologi. Data yang diambil sudah terlebih dahulu diketahui tingkat keganasannya (kelas) berdasarkan hasil pemeriksaan patologi anatomi. Data tersebut bertujuan untuk menghitung tingkat akurasi dari metode *Independent Component Analysis* dan *Possibilistic C-Means* dalam mengklasifikasi *astrocytoma*.

4.2 Hasil Pengujian Data MRS dengan Metode ICA

Pengujian dilakukan dengan melakukan seratus kali pengulangan, hingga didapatkan hasil yang bisa dilihat pada tabel 4.2.

Dari tabel 4.2 kita dapatkan tingkat akurasi yang diperoleh metode ICA setelah dilakukan seratus kali pengulangan. Akurasi maksimum yang diperoleh metode ICA mencapai 96,67% yang didapat dengan menggunakan 40% data training dan rata-rata akurasi maksimum yang diperoleh setelah seratus kali percobaan mencapai 92,52% yang didapat dengan menggunakan 60% data *training*. Dari hasil ini sudah terlihat bahwa metode ICA memberikan hasil yang sangat baik dalam klasifikasi *astrocytoma*.

Tabel 4. 2 Prosentase Hasil Akurasi Percobaan Metode Independent Component Analysis untuk Klasifikasi Data MRS

Prosentase data training (%)	Prosentase Akurasi Metode Independent Component Analysis (%)	
	Akurasi Maksimum	Akurasi Rata-Rata*
20	66,67	61,90
30	52,78	47,78
40	96,67	88,50
50	95,83	92,12
60	95,24	92,52
70	93,33	91,91
80	93,17	91,67

Keterangan : *) Akurasi rata-rata didapatkan setelah dilakukannya seratus kali pengulangan

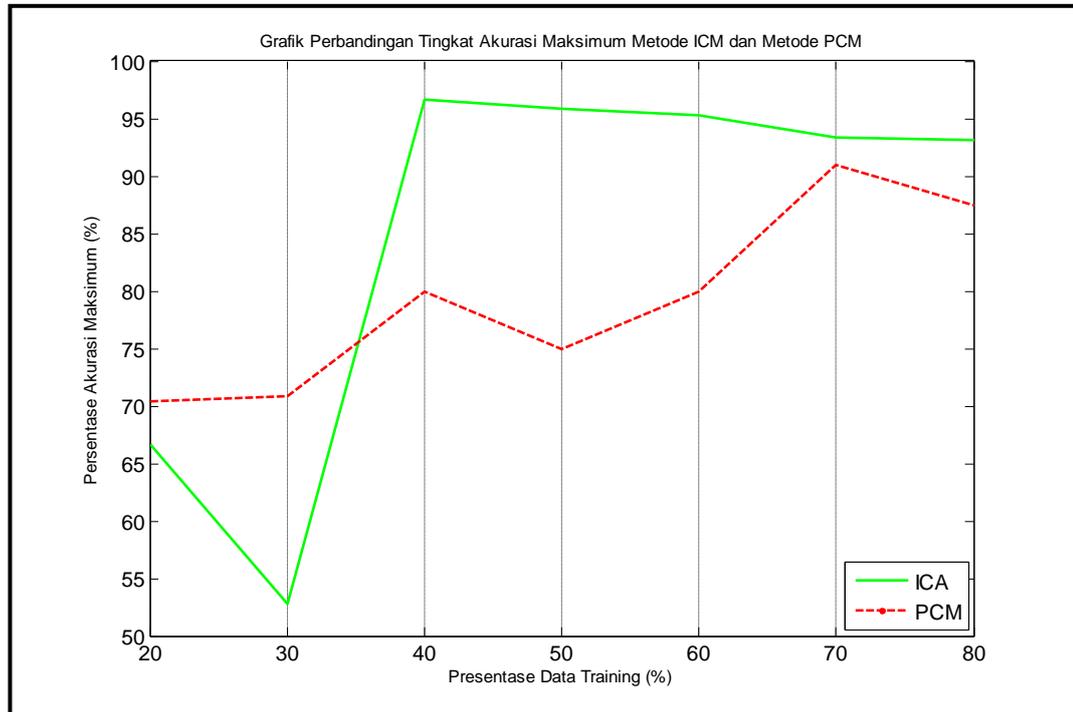
Selanjutnya akan dilihat perbandingan dari hasil yang diperoleh metode ICA dengan hasil terbaik yang diperoleh metode *Possibilitic C-Means* (Wibowo, Anggi P., et al., 2010). Hasil dapat dilihat pada tabel 4.3 dan gambar 4.1.

Dari tabel 4.3 terlihat bahwa metode ICA mendapatkan hasil akurasi yang lebih baik daripada metode PCM. Metode ICA mendapatkan tingkat akurasi 96,67% sementara metode PCM hanya 90,91%.

Tabel 4. 3 Perbandingan Hasil Terbaik dari Metode ICA dengan Metode PCM

Prosentase Data Training	Prosentase Keakuratan Klasifikasi (%)	
	PCM*	ICA
20	70,37	66,67
30	70,83	52,78
40	80,00	96,67
50	75,00	95,83
60	80,00	95,24
70	90,91	93,33
80	87,50	91,67

Keterangan: *) Hasil akurasi dari PCM didapatkan dari hasil percobaan yang dilakukan Anggi P. Wibowo, et al (2010)



Gambar 4. 1 Grafik perbandingan akurasi metode ICA dengan PCM

BAB 5

KESIMPULAN DAN SARAN

Berdasarkan percobaan yang dilakukan pada bab sebelumnya, metode *Independent Component Analysis* (ICA) dan metode *Possibilistic C-Means* (PCM) dapat digunakan untuk klasifikasi *astrocytoma*. Metode *Independent Component Analysis* merupakan metode yang lebih baik dibandingkan dengan metode *Possibilistic C-Means* dalam melakukan proses klasifikasi *astrocytoma*. Hal ini dapat terlihat pada tabel perbandingan kedua metode, di mana akurasi dari metode ICA mencapai 96,67% sementara metode PCM mencapai 90,91%.

Untuk selanjutnya dari hasil tersebut, akan dibuat perangkat lunak untuk penunjang keputusan, yang dapat langsung digunakan oleh dokter ahli untuk menentukan tingkat keganasan dari *astrocytoma*.

Untuk penelitian lebih lanjut, *astrocytoma* dapat digolongkan kedalam empat kelas yang berbeda, tidak hanya tiga kelas.

DAFTAR PUSTAKA

- Anton, Howard. 2000. *Elementary Linear Algebra*. Jakarta: Interaksara.
- Bauer, Christoph, et al. *Local Features in Biomedical Image Clusters extracted with Independent Component Analysis*. Germany: University of Rogensburg.
- Bell, Anthony J., dan Sejnowski, Terrence J. 1995. *An Information-Maximization Approach to Blind Separation and Blind Deconvolution*. *Neural Computation*, 7, 1129-1159.
- Bernstein M, Berger MS. 2000. *Neuro-Oncology The Essentials*. New York: Thieme.
- Bezdek, J., 1981. *Pattern Recognition with Fuzzy Objective Function Algorithm*. New York: Plenum Press.
- Brandao LA. 2004. *MR Spectroscopy of the Brain*. Philadelphia: Lippincott Williams & Wilkins.
- Fikri, Akmal. 2009. *Brain Cancer (Astrocytoma) Clustering Menggunakan Metode Fuzzy C-Means*. Depok: Universitas Indonesia.
- Gallas, Sherrie. M, MPH. *Brain Cancer: Statistics and Cluster Investigation*. MCDH Epidemiologist.
- Hyvärinen, Aapo., Karhunen, Juha., dan Oja, Erkki. 2001. *Independent Component Analysis*. United States: John Wiley & Sons.
- Krishnapuram, Raghu., Keller, James M., 1993. *Possibilistic Approach to Clustering*. *IEEE Transactions on Fuzzy Systems*, Vol. 1, No. 2, pp.98-110.

- Kong, Wei., et al. 2008. *A Review of Independent Component Analysis Application to Microarray Gene Expression Data*. BioTechniques, 45, 501-520.
- Pandelaki, Jacob. dr, SP.Rad. 2006. *Makalah Seminar S-3*.
- Roger, A. Horn dan Johnson, Charles R. 1985. *Matrix Analysis*. Cambridge University. Press: Cambridge.
- Stone, James V. *Independent Component Analysis: A Tutorial Introduction*. Cambridge: The MIT Press.
- Wibowo, Anggi P., Rustam, Zuherman., dan Pandelaki, Jacob. 2010. *Clustering Brain Cancer Menggunakan Possibilistic C-Means*. Depok: Seminar Nasional Matematika UI-UNPAD.