

## KLASIFIKASI TUMOR OTAK JINAK (*BENIGNA*) DAN GANAS (*MALIGNA*) MENGGUNAKAN EKSTRAKSI FITUR GLCM DAN SVM

<sup>1</sup>Rohmawati

<sup>1</sup>Program Studi Teknik Informatika, Fakultas Teknik Universitas Yudharta Pasuruan

<sup>1</sup>Email : rochma.rokmund@gmail.com

### ABSTRAK

*Brain tumors are a deadly disease that can attack anyone without knowing age. Technology that continues to grow makes the world of health inseparable from technology. One of the technologies used to identify a disease is CT Scan and MRI. Clinically it is difficult to distinguish between benign and malignant brain tumors because like normal brain tissue, doctors can diagnose the disease without having to do surgery. This study aims to detect benign and malignant brain tumors by using extraction of GLCM and SVM features. GLCM is one method for obtaining statistical characteristics by calculating the probability of the neighboring relationship between two pixels at a certain distance and angle orientation. While the SVM method is due to the best class and classification separation and is able to work on high-dimensional datasets. The ct-scan image that is entered will be segmented which will later be extracted using GLCM features, the features used include mean, contrast, correlation, homogeneity, IDM, variance and entropy. After testing, it can be concluded that the accuracy rate is 94.5%. While using the WEKA application is 91.6666% and an error of 8.3334%.*

*Keywords : Classification, Ct- Scan and MRI, brain tumor, GLCM, SVM*

### I PENDAHULUAN

Tumor otak merupakan suatu penyakit mematikan yang dapat menyerang siapa saja tanpa mengenal usia. Pertumbuhan sel yang tidak normal (abnormal) di jaringan otak atau pusat tulang belakang yang dapat mengganggu fungsi otak juga di sebut tumor otak. Menurut data dari International Agency for Research of Cancer, lebih dari 126.000 orang di dunia setiap tahunnya mengidap tumor otak dan lebih dari 97.000 jiwa meninggal dunia (Sabbih hamoud & Ghazali, 2014). Lebih dari 688.000 orang menderita tumor otak, dimana 63% tumor otak jinak dan 37% tumor otak ganas (National brain tumor, 2013)..

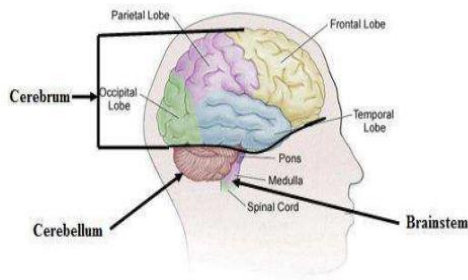
Oleh karena itu, untuk membedakan antara tumor otak jinak dan ganas dibutuhkan sebuah teknologi. Teknologi yang terus berkembang membuat dunia kesehatan tidak dapat di pisahkan, salah satunya dalam dunia medis, teknologi sangat di butuhkan dalam pengidentifikasian suatu penyakit, seperti mengidentifikasi tumor otak. Salah satu alat teknologi yang berguna untuk mengidentifikasi penyakit adalah Magnetic Resonance Imaging (MRI), Ultrasound dan Computed Tomography. Computed Tomography merupakan alat diagnosa dengan teknik radiografi yang memiliki kemampuan untuk mendeteksi penyakit dan menghasilkan gambar. Pada penelitian ini akan digunakan sebuah metode yaitu metode GLCM DAN SVM. GLCM digunakan karena metode ini merupakan ekstraksi fitur yang paling baik dan merupakan metode yang cukup efektif dalam melakukan klasifikasi dan identifikasi karena mampu memberikan informasi yang detail tentang suatu citra dalam hal tekstur (Maritaa, Nurhasanah, & Sanubarya, 2014). Sedangkan penggunaan metode SVM karena pemisahan kelas dan klasifikasi yang paling baik dan mampu bekerja pada dataset yang berdimensi tinggi. Di harapkan dengan adanya penelitian ini tumor otak jinak dan ganas dapat teridentifikasi dengan baik dan menghasilkan tingkat akurasi yang cukup tinggi dalam mengidentifikasi.

### II TINJAUAN PUSTAKA

#### 2.2.1 Otak

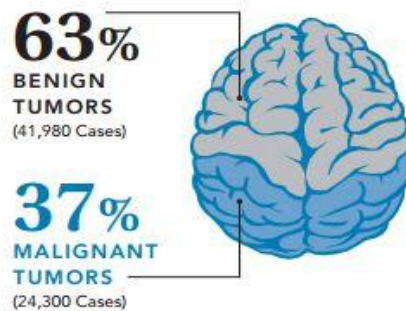
Otak (*Brain*) adalah pusat sistem saraf yang memiliki volume sekitar 1.350cc dan terdiri atas 100 juta sel saraf dan *neuron*. Otak manusia berfungsi sebagai pusat untuk mengendalikan semua bagian tubuh manusia yang memungkinkan manusia untuk mengartikulasi kata – kata, gerakan perilaku dan mengatur fungsi tubuh seperti detak jantung, tekanan darah dan keseimbangan suhu tubuh.

Gambar 2.1 Otak manusia

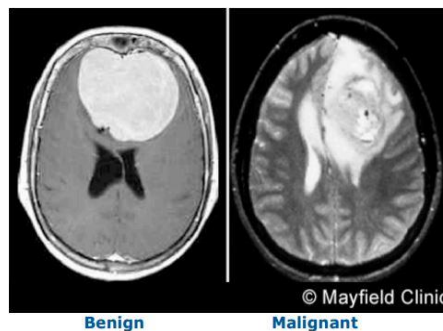


### 2.2.2 Tumor Otak

Tumor juga disebut *neoplasma* atau *lesi*, yaitu jaringan abnormal yang tumbuh oleh sel yang tidak terkontrol dan sel tumor bereproduksi tak terkontrol. Tumor otak adalah pertumbuhan sel yang tidak normal di jaringan otak atau pusat tulang belakang yang dapat mengganggu fungsi otak. Tumor otak menurut sel mulai tumbuh bersifat kanker (ganas) atau tidak (jinak).



gambar 2.2 Tumor otak



gambar 2.3 Tumor otak jinak dan ganas

### 2.2.3 Computed Tomography– Scan

*Computed tomography* memiliki dua kata yaitu *computed* dan *tomography*. *Computed* adalah kata kerja dari bahasa Inggris yang berarti dihitung atau terkomputasi (menggunakan komputer). Sedangkan *Tomography* mempunyai arti gambaran dari sebuah objek melalui refleksi data dengan memperjelas objek dari arah yang berbeda.

### 2.2.5 Segmentasi Citra

Segmentasi merupakan teknik untuk membagi suatu citra menjadi beberapa daerah (*region*), dimana setiap daerah memiliki kemiripan atribut. Atribut yang paling dasar adalah amplitudo pencahayaan (*luminance*) citra untuk gambar monokrom dan komponen warna gambar berwarna.

### 2.2.6 Data Mining

*Data mining* adalah teknik untuk mengekstraksi atau mengetahui dari kumpulan banyak data dari data yang berukuran atau berjumlah besar atau serangkaian proses untuk menggali nilai dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.

### 2.2.7 Klasifikasi (Classification)

Klasifikasi adalah proses pembelajaran suatu fungsi aturan (target)  $f$  yang memetakan tiap himpunan atribut  $x$  ke satu dari label kelas  $y$  yang didefinisikan sebelumnya. Fungsi target di sebut juga model klasifikasi

### 2.2.9 Gray Level Co - occurrence Matrik (GLCM)

GLCM (*Gray Level Co - occurrence Matrik*) pertama kali diusulkan oleh Haralick pada tahun 1973 untuk menjelaskan pola spasial (kulkarni, 1994). (Abdul & Adhi, 2013). GLCM merupakan salah satu metode untuk memperoleh ciri statistic dengan cara menghitung probabilitas hubungan ketetanggaan antara dua piksel pada jarak dan orientasi sudut tertentu.

Ciri atau fitur statistik GLCM yang akan digunakan antara lain:

1. Mean

$$\mu = \frac{1}{MN} \sum M_i = i \sum N_j = 1 P_{ij}$$

$\mu$  = nilai rata – rata piksel ( $P_{ij}$ )

$M$  dan  $N$  = nilai panjang dan lebar piksel

$P_{ij}$  = nilai pada baris  $i$  dan kolom  $j$  pada matriks *co-occurrence*.

2. Contrast

$$\sum_{i,j} |i - j|^2 p(i,j)$$

$i$  dan  $j$  adalah nilai pada kolom

$\sum$  = sigma

3. Correlation

$$\sum_{i,j} \frac{(i - \bar{i})(j - \bar{j})p(i,j)}{\sigma_i \sigma_j}$$

$\bar{i}$   $\bar{j}$  nilai rata elemen pada kolom  $j$  pada matriks,

$\sigma_i \sigma_j$  = standart deviasi dari baris dalam matriks

4. Energy

$$\sum_{i,j} p(i,j)^2$$

Nilai energy di tunjukkan dalam baris dan kolom  $i,j$

5. Homogeneity

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|}$$

$i$  dan  $j$  adalah nilai pada baris dan kolom

6. Variance

$$- \sum_i \sum_j (i - \bar{i})(j - \bar{j})P(i,j)$$

$P(i,j)$  = nilai pada baris  $i$  dan kolom  $j$  pada matriks kookurensi.

$\bar{i}$  = nilai rata – rata elemen pada kolom  $i$  pada baris matriks.

7. Inverse Different Moment

$$- \sum_i \sum_j \frac{1}{1+(i,j)^2} P(i,j)$$

$P_{i,j}$  adalah nilai pada baris  $i$  dan kolom  $j$  pada matriks.

8. Entropy

$$- \sum_i \sum_j P(i,j). 2 \log P(i,j)$$

$P_{i,j}$  adalah nilai pada baris  $i$  dan kolom  $j$  pada matriks

### 2.2.13 Support Vector Machine

Support Vektor Machine (SVM) adalah algoritma pembelajaran yang diawasi dengan menggunakan pemetaan non – linier untuk menganalisis data dan digunakan untuk klasifikasi. *Support vector machine* pertama kali di kembangkan oleh Vladimir Vapnik, Bernhard Boser dan Isabelle Guyon pada tahun 1992 di Annual Workshop on Computational Learning Theory. Konsep dasar SVM merupakan usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah class pada input space.

## III METODE PENELITIAN

### 3.1 Metode Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh suatu informasi yang dibutuhkan untuk mencapai tujuan dari peneliti. Berikut merupakan metode pengumpulan data :

1. Studi literature  
Pengumpulan referensi dalam penelitian dilakukan dengan mencari jurnal, buku dan artikel.
2. Data

Data pada penelitian ini berupa data citra CT- Scan dan MRI tumor otak jinak dan ganas yang diambil dari <https://public.cancerimagingarchive.net>, <https://radiopaedia.org/>, <http://www.medinfo.cs.ucy.ac.cy/index.php/downloads/datasets>



Gambar 3.2 alur penelitian

Pengujian dilakukan menggunakan data yang di peroleh dari hasil segmentasi. Yang telah di ekstrasi fitur GLCM yang nantinya akan menjadi masukan untuk melakukan pengujian atau testing menggunakan metod SVM.

Hasil

Hasil adalah keluaran dari yang dikerjakan

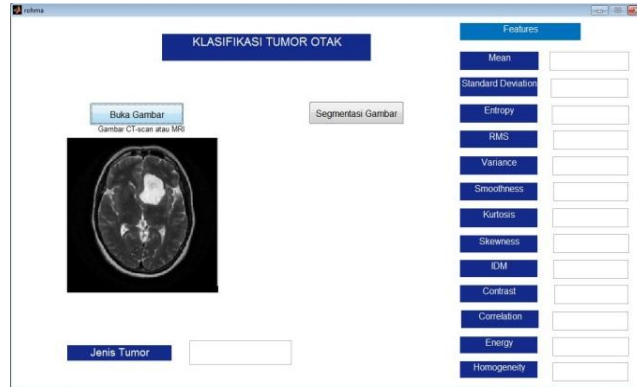
$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

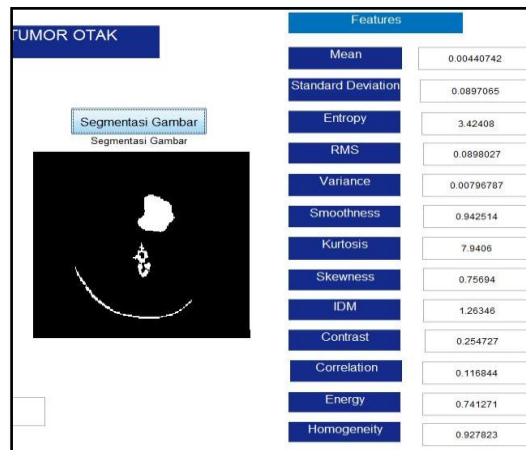
$$Recall = \frac{TP}{TP + FN}$$

$$F - score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

#### IV HASIL DAN PEMBAHASAN



Gambar 4.7 citra muncul



Gambar 4.8 hasil klasifikasi

gambar 4.13 hasil segmentasi dan ekstraksi fitur.

Setelah pengujian selesai, langkah selanjutnya Adalah mengetahui kinerja system, dalam hal ini ada 4 tingkat, yaitu

- a. *True Positive (TP)*
- b. *False Negative (FN)*
- c. *False Positive (FP)*
- d. *True Negative (TN)*

**Tabel 4.2 confusion matrix**

prediksi	Tumor	Normal
Tumor	TP	FN
Normal	FP	TN

- TP yaitu data tumor yang diklasifikasi sebagai tumor sesuai jenis.
- FN yaitu data tumor otak yang diklasifikasi sebagai data normal atau data tumor tidak sesuai dengan jenisnya.
- FP yaitu data normal yang diklasifikasikan sebagai data tumor.
- TN yaitu data normal yang diklasifikasi sebagai data normal. Adapun nilai - nilai TP, FN, FP dan TN berdasarkan hasil pengujian pada dapat dilihat pada table 4.2

**Tabel 4.3 hasil**

TP	FN	FP	TN
40	5	0	15

Berdasarkan nilai TP, FN, FP dan TN pada tabel 4.2, diketahui nilai *False Negative* sebanyak 5 dari total 60 data pengujian yang disebabkan oleh kesalahan dalam hasil pengklasifikasian. Rata-rata akurasi yang diperoleh yaitu:

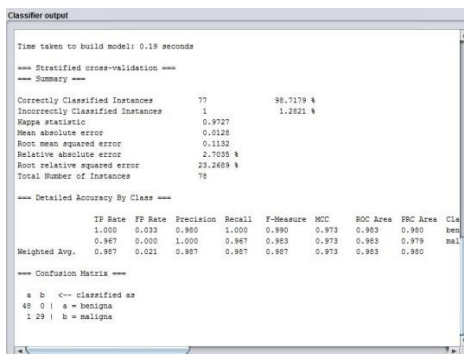
$$\frac{\text{True Positive} + \text{True Negative}}{\text{total data uji}} \times 100\% = 91.666\% = 92\%$$

- Nilai *precision*, *recall*, *specificity* dan *F-Score* seperti pada table 4.16

**Table 4.4 Pattern recognition**

Precision	$\frac{TP}{TP+FP} \times 100\% = 100\%$
Recall	$\frac{TP}{TP+FN} \times 100\% = 89\%$
Specificity	$\frac{TN}{TN+FP} \times 100\% = 100\%$
F - Score	$\frac{\text{precision} + \text{recall}}{2} \times 100\% = 94.5\%$

## 4.2 Hasil Klasifikasi



```

Classifier output
Time taken to build model: 0.19 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      77      98.119 %
Incorrectly Classified Instances     1      1.881 %
Kappa statistic                     0.9727
Mean absolute error                  0.0128
Root mean squared error              0.1132
Relative absolute error              2.7035 %
Root relative squared error          23.2689 %
Total Number of Instances           78

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
-----
1.000  0.033  0.980    1.000  0.990  0.973  0.983  0.980  ben
0.987  0.000  1.000    0.987  0.983  0.973  0.983  0.979  mal
Weighted Avg.  0.987  0.021  0.987  0.987  0.987  0.973  0.983  0.980

=== Confusion Matrix ===
  a b <- classified as
 40 0 | a = benign
 12 1 | b = malignant
  
```

Gambar 4.17 Hasil akurasi

Hasil akurasi  
Hasil = 91.6666%  
Error = 8.3334%.

**4.4 Proses Klasifikasi SVM**

Proses SVM diawali dengan menemukan *Support Vector*, yaitu 2 data dengan jarak terpendek yang berasal dari *class* yang berbeda menggunakan rumus jarak antara dua titik. Maka data bernomor urut 20 dan 44 dengan jarak terpendek sebesar 0.177473 dipilih sebagai *Support Vector* seperti tertera pada Tabel 4.17 berikut:

**Tabel 4.5 Data Latih SVM**

Data	mean	ent	var	.....
A	0.00486631	2.82343	0.00804304	.....
B	0.00421278	2.991	0.00805846	.....

Nilai *class* benigna pada data A dianggap sebagai 1 untuk mewakili *class* positif dan nilai *class* maligna pada data B dianggap sebagai -1 untuk mewakili *class* negative. Langkah-langkah penerapan perhitungan algoritma klasifikasi SVM selanjutnya adalah sebagai berikut:

- Menghitung *Kernel* tiap data:  
 $K(A, A) = 10.40472904$   
 $K(A, B) = 10.84022474$   
 $K(B, A) = 10.84022474$   
 $K(B, B) = 11.30721707$
- Memasukkan ke dalam persamaan *Lagrangian* berikut:  

$$\sum_i \lambda_i y_i = 1\lambda A - 1\lambda B = 0$$
- Memasukkan ke dalam persamaan *Decision Boundary* untuk *Positive Gutter* berikut:  

$$\sum_i \lambda_i y_i K(x_i, x) + b = 10.405\lambda A - 10.84\lambda B + 1b = 1$$
- Memasukkan ke dalam persamaan *Decision Boundary* untuk *Negative Gutter* berikut:  

$$\sum_i \lambda_i y_i K(x_i, x) + b = 10.84\lambda A - 11.307\lambda B + 1b = -1$$
- Menggunakan ketiga persamaan pada langkah 2, 3 dan 4 dengan persamaan sistem linier yang menghasilkan nilai  $\lambda$  (*Lagrangian*) dan  $b$  (*bias*). Hasilnya adalah  $\lambda A = 62.5$ ,  $\lambda B = 62.5$  dan  $b = 28.1875$ .
- Dilakukan pengujian menggunakan sampel data nomor 22 sebagai data uji seperti yang tertera pada Tabel 4.18 berikut:

**Tabel 4.6 Data Uji SVM**

Data	mean	ent	var	.....
X	0.00471326	2.56203	0.0080699	.....

- Menghitung *Kernel* data uji terhadap data latih:  
 $K(X, A) = 9.074523239$   
 $K(X, B) = 9.500725902$
- Mengklasifikasikan data X ke dalam suatu *class* menggunakan:  

$$f(X) = \sum_{i=1}^N \lambda_i y_i (x_i)(X) + b = (62.5 * 9.075 * 1) + (62.5 * 9.501 * (-1)) + 28.1875$$

$$= 567.1875 - 593.8125 + 28.1875 = 1.5625$$

Jadi, data uji X merupakan *class* positif karna  $1.5625 > 1$  dan memiliki nilai benigna pada atribut *class*-nya

3. Langkah 6 - 8 diulang hingga seluruh data uji selesai diklasifikasikan dan kemudian dilakukan *evaluation* berdasarkan nilai-nilai pada tabel *confusion matrix*.
4. Melakukan proses *evaluation* sebagaimana berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} = \frac{40 + 15}{40 + 0 + 5 + 15} = \frac{55}{60} = 0.91666 * 100\% = 91.666\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{40 + 0}{40} = \frac{40}{40} = 1$$

$$Recall = \frac{TP}{TP + FN} = \frac{40 + 5}{40} = \frac{40}{45}$$

$$F - measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2}{\frac{1}{1} + \frac{1}{0.126}} = 8.936 = 0.888$$

## V PENUTUP

### 5.1 Kesimpulan

Berdasarkan hasil penelitian, maka dapat disimpulkan :

1. Pada aplikasi matlab dengan hasil klasifikasi menggunakan metode GLCM dan SVM berhasil mengidentifikasi antara tumor otak jinak dan ganas dengan baik dengan tingkat akurasi sebesar 94.5 %.
2. Sedangkan menggunakan aplikasi WEKA menghasilkan tingkat akurasi sebesar 91.6666 % dan Error sebesar 8.3334 %.
3. Tingkat akurasi menggunakan Matlab lebih baik dari pada menggunakan aplikasi WEKA.

### 5.2 Saran

Berdasarkan hasil yang diperoleh dari penelitian ini, terdapat beberapa saran yang berguna untuk penelitian-penelitian mendatang yakni:

1. Menggunakan metode lain agar hasil klasifikasi semakin baik
2. Menambahkan detail diagnosa pada hasil keluaran di matlab
3. Menggunakan lebih banyak gambar atau citra untuk klasifikasi agar tingkat akurasi semakin baik

## DAFTAR PUSTAKA

Analisis Diagnosis Pasien Kanker Payudara menggunakan Regresi dan Support Vector Machine (SVM) berdasarkan Hasil Mamografi 2012 *Jurnal Sains dan Seni ITS* 147-153

Brain tumour detection using mean shift clustering and GLCM features with edge adaptive total variation denoising technique 2017 *Alexandria Engineering Journal* 1-6

*data mining* 2013 Yogyakarta cv andi offset

DETEKSI KANKER KOLEREKTAL (KANKER USUS BESAR) MENGGUNAKAN METODE GRAY LEVEL COOCCURENCE MATRIX DAN K-NEAREST NEIGHBOR BERBASIS PENGOLAHAN CITRA 2017 *e-Proceeding of Engineering* 1-14

Dr.S. Palani 2012 BRAIN TUMOR MRI IMAGE CLASSIFICATION WITH FEATURE SELECTION AND EXTRACTION USING LINEAR DISCRIMINANT ANALYSIS *International Journal of Information Sciences and Techniques* 1-16

EKSTRAKSI FITUR DAN KLASIFIKASI MENGGUNAKAN METODE GLCM DAN SVM PADA CITRA MAMMOGRAM UNTUK IDENTIFIKASI KANKER PAYUDARA 2017 *Jurnal Teknologi Rekayasa* 18-26

GLCM Textural Features for Brain Tumor Classification 2012 *International Journal of Computer Science Issues* 1-6

ISSN : 2355-9365 e-Proceeding of Engineering 2015 *Deteksi Tumor Otak dengan Ekstraksi Ciri & Feature Selection menggunakan Linear Discriminant Analysis (LDA) dan Support Vector Machine (SVM)* 8

Journal of Theoretical and Applied Information Technology 2014 *TUMOR BRAIN DETECTION THROUGH MR IMAGES: A REVIEW OF LITERATURE* 387-403

Jurnal Ilmiah Komputer dan Informatika (KOMPUTA) 12013 *ANALISIS METODE JARINGAN SYARAF TIRUAN BACKPROPAGATION UNTUK PENGENALAN SEL KANKER OTAK* 8

Klasifikasi Citra Batik Menggunakan Jaringan Syaraf Tiruan Berdasarkan Gray Level Co-Occurrence Matrices (GLCM) 2014 *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)* 13

Maritaa Nurhasanah Sanubarya 2014 *PRISMA FISIKA*, ISSN : 2337 8204 *Identifikasi Tumor Otak Menggunakan Jaringan Syaraf Tiruan Propagasi Balik pada Citra CT-Scan Otak* 117-121

mayfield clinic 2016 *mayfield brain and spine* [mayfieldclinic.com](http://mayfieldclinic.com) 1-7

National brain tumor 2013 *national brain tumor society Frankly speaking about cancer brain tumor* 1-96

Seminar Nasional Aplikasi Teknologi Informasi 2010 *PENGENALAN POLA BERBASIS JARINGAN SYARAF TIRUAN DALAM ANALISA CT SCAN TUMOR OTAK BELIGNA* G26-G31 *teori dan aplikasi pengolahan citra* 2013 *yogyakarta cv andi* offs

