

**PENDEKATAN *PROBABILISTIC NEURAL NETWORK (PNN)*  
BERBASIS *EXPECTATION MAXIMUM (EM)*  
UNTUK PERMASALAHAN KLASIFIKASI GABUNGAN**

***Probabilistic Neural Network (Pnn) Approaches Based On Expectation  
Maximum (Em) For Combining Classification Problems***

**Wawan Setiawan<sup>1)</sup>, Wiweka<sup>2)</sup>**

<sup>1)</sup> *Ilmu Komputer, Universitas Pendidikan Indonesia, pik@upi.edu*

<sup>2)</sup> *Kedeputian Penginderaan Jauh, LAPAN, wiweka@lapan.go.id*

Diterima (received): 27-3-2012, disetujui untuk publikasi (accepted): 18-5- 2012

**ABSTRAK**

Makalah ini menyajikan desain pengklasifikasi dengan pendekatan jaringan neural (syaraf) yang didasarkan pada penggunaan metode Ekapentasi Maksimum (EM). Aturan keputusan pengklasifikasi Bayes menggunakan kesalahan minimum dalam klasifikasi citra gabungan multi waktu. Dalam khusus ini, model jaringan neural perceptron multi lapis Probabilistic Neural Network (PNN), digunakan untuk mengestimasi nonparametrik probabilitas posterior suatu kelas. Korelasi temporal citra dihitung dengan probabilitas gabungan masing-masing kelas secara otomatis dengan menerapkan formula khusus yaitu algoritma ekspektasi maksimum dari citra multi waktu. Eksperimen dilakukan pada dua citra multi waktu yaitu citra Saguling yang diambil pada dua waktu yang berbeda. Berdasarkan hasil eksperimen pada dua daerah uji tersebut, dapat ditunjukkan bahwa tingkat akurasi pengklasifikasi PNN rata-rata lebih baik dibandingkan dengan model Propagasi Balik (BP), dan Ekepektasi Maksimum (EM) dapat meningkatkan kemampuan suatu pengklasifikasi. Pengklasifikasi PNN dengan menerapkan ekpektasi maksimum memiliki kemampuan pengenalan yang konsisten untuk citra multi waktu, dan juga konsisten untuk setiap pengenalan kategori kelas obyek. Metodologi klasifikasi yang diusulkan dapat memecahkan permasalahan multi waktu secara efektif.

**KataKunci:** Probabilistik Jaringan Syaraf, Ekspektasi Maksimm, Citra Multitemporal, Kelas Objek, Pengenalan

**ABSTRACT**

This paper presents a classifiers design of neural network approachbased on Expectations Maximum (EM), a Bayes classifier decision rule using the Minimum Error to clasify combined multi-temporal imageries. In this particular, multilayer perceptron neural network model with Probabilistic Neural Network (PNN) is used for nonparametric estimation of posterior class probabilities. Temporal image correlation was calculated automatically usingprior joint probabilities of each class by applying a special formula that is algorithm expectation maximum of multi-temporal imagery. Experiments wasperformed on two multi-temporal Saguling imagestakenat two different epochs. Based on experimental results on two test areas, it can be shown that the average accuracy rate of PNN classifier is better than the Back Propagation (BP), and the Expectation Maximum (EM) can increase the classifiers ability. Multinomial PNN classifierapplying the maximum expected have a consistent recognition capability for multitemporal imagery, and also consistent for each object class category. The proposed classification methodology can effectively solve the problem when classifying multi-temporalimagery.

**Keywords:** Probabilistic Neural Networks, Expectation Maximum, Multitemporal Images, Class Object, Recognition

**PENDAHULUAN**

Penerapan jaringan neural untuk klasifikasi pola telah dikaji secara intensif

dalam beberapa tahun terakhir. Berbagai macam arsitektur jaringan neural seperti Multilayer Perceptron (MLP), radial basis function (RBF), dan self organizing map

(SOM) telah diusulkan. Dengan kemudahan dalam pelatihan dan dasar statistik yang baik dalam teori estimasi Bayesian, Probabilistic Neural Network (PNN) menjadi perangkat yang cukup efektif untuk memecahkan berbagai masalah klasifikasi. Namun, terdapat masalah terkait dengan struktur PNN, yaitu bagaimana menentukan ukuran jaringan, menempatkan neuron layer pola sebaik mungkin sesuai nilai parameter. Sesuai fakta, layer pola pada PNN sering kali memuat sampel pelatihan yang tidak efektif (berguna). Sampel yang berlebihan berpotensi mengakibatkan struktur jaringan yang membengkak, yang menyebabkan dua masalah. Pertama, akan menyebabkan cost komputasi yang tinggi hanya karena komputasi untuk sebuah pola yang tidak diketahui apakah proporsional dengan ukuran jaringan. Kedua, konsekuensinya cenderung terlalu sensitif dan kemungkinan menunjukkan kemampuan generalisasi yang rendah untuk data yang tidak pernah ada. Di sisi lain, parameter penghalus juga memainkan peran penting pada pengklasifikasi PNN, dan seringkali bergantung pada data.

Kedua masalah di atas telah direalisasikan oleh beberapa peneliti dan diantaranya melakukan reduksi sampel pelatihan. Pendekatan vektor kuantisasi telah digunakan untuk mengelompokkan sampel pelatihan dan menemukan pusat kluster untuk PNN. Fungsi probabilitas densitas dalam sebuah PNN diaproksimasi oleh sebuah angka kecil dari komponen densitas dan parameter komponen disestimasi dari pengaturan pelatihan dengan menggunakan algoritma *Self*

*Organizing Map* (SOM) Gaussian. Teknik pengelompokkan dengan paradigma *Energi Coulomb* terbatas, digunakan untuk menemukan pusat kluster dan disesuaikan dengan jumlah sampel perwakilan dari setiap kluster. Pada dasarnya, semua algoritma reduksi seperti PNN, didasarkan pada pendekatan pengelompokkan. Karena kesalahan klasifikasi tidak dimanfaatkan secara langsung dalam proses seleksi neuron, algoritma tersebut dapat dikategorikan ke dalam kategori pembelajaran tak terawasi.

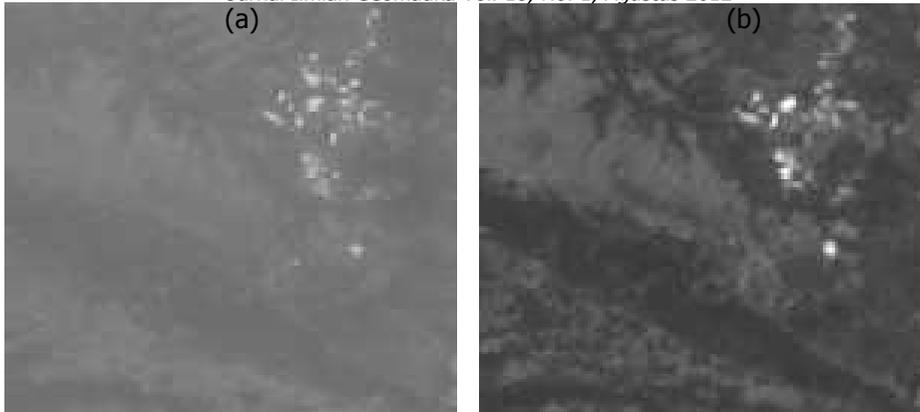
Untuk tujuan ini, dua pengklasifikasi PNN (untuk waktu  $t_1$  dan  $t_2$ ) perlu dilatih secara terpisah pada dua kumpulan data pelatihan, dengan target kebenaran yang telah tersedia. Dua kumpulan pelatihan harus independen satu sama lain.

Sebagai hasil dari fase pelatihan, diperoleh nilai-nilai optimal parameter internal dari jaringan neural (seperti, bobot dan bias). Dalam hal ini, setiap jaringan neural dapat digunakan untuk menghitung estimasi probabilitas posterior kelas obyek pada waktu yang bersesuaian (keluaran jaringan neural), bergantung vektor fitur  $X_i$  (masukan PNN).

## **METODOLOGI DAN EKSPERIMEN**

### **Dara Eksperimen**

Sepasang citra (Landsat TM) multi waktu daerah Saguling digunakan dalam penelitian ini. Citra Landsat TM Saguling yang diambil pada 4 Juli 1987 ( $t_1$ ) dan 9 Juli 1994 ( $t_2$ ) seperti ditunjukkan pada Gambar 1. Metodologi klasifikasi citra ditunjukkan seperti diagram blok pada



**Gambar 1.** Citra Lansat TM daerah Saguling pada (a) tahun 1987 dan (b) tahun 1994. (Sumber data: LAPAN)

Masukan terdiri atas data citra suatu daerah yang direkam pada dua waktu yang berbeda. Sampel target digunakan untuk pelatihan (untuk melatih jaringan neural dan estimasi probabilitas prior dari satu kelas pada suatu waktu) dan pengujian (untuk evaluasi kinerja). Dua vektor fitur  $X_1$ , dan  $X_2$  (masing-masing waktu  $t_1$  dan  $t_2$ ) terdiri atas 12 fitur tekstur yang diturunkan dari matriks kookuren (Murni, 1997). Dua jaringan neural (PNN) digunakan untuk mengestimasi probabilitas posterior  $P(w_i/X_1)$  dan  $P(v_j/X_2)$ . Arsitektur jaringan neural dengan layer hidden tunggal semuanya terhubung dan independen masing-masing untuk  $t_1$  dan  $t_2$  (Lee et al. 1987).

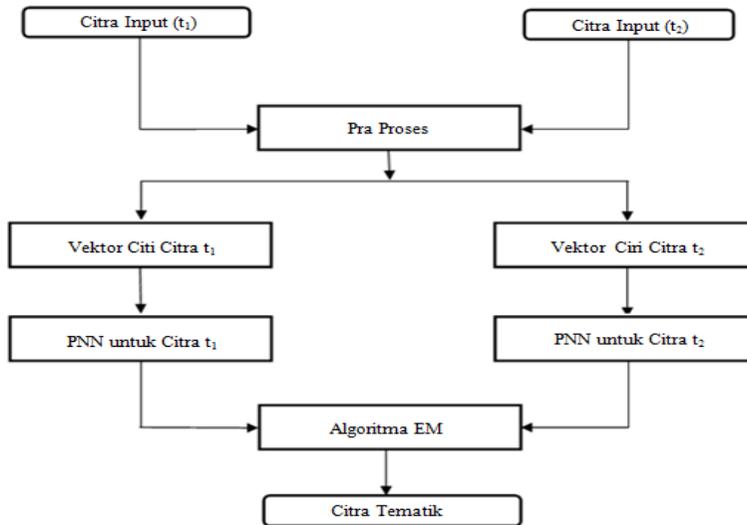
Dalam penerapannya, algoritma EM melakukan pemilihan parameter konvergensi yaitu  $\epsilon = 0,001$ . Probabilitas join untuk pengujian diperoleh dari iterasi terakhir pelatihan algoritma EM. Korelasi temporal diantara dua citra dinyatakan dengan estimasi EM dengan rata-rata probabilitas join yang mereduksi kesalahan klasifikasi secara signifikan

pada suatu kumpulan data (Dempster, 1997).

Tabel 1 menunjukkan kelas dan jumlah piksel terkait dengan data kumpulan pelatihan dan testing, tabel 2 menunjukkan atribut jaringan neural probabilistik (PNN), dan tabel 3 menunjukkan atribut algoritma ekspektasi maksimum (EM).

### Ekstraksi Fitur

Ekstraksi fitur menggunakan perhitungan statistik tingkat dua yang melibatkan hubungan karakteristik tertentu (tingkat keabuan) diantara piksel suatu citra pada suatu posisi (ketetanggaan) tertentu. Pada level ini meliputi perhitungan probabilitas okuren dari piksel dengan tingkat keabuan  $g_1$  berdekatan dengan piksel dengan tingkat abu-abu  $g_2$  (kookuren), penghitungan tingkat keabuan yang berbeda diantara dua piksel dengan jarak dan arah tertentu (semivariogram).



Gambar 2. Skema Klasifikasi

Keberanan sampel target digunakan untuk membuat data kumpulan pelatihan (digunakan untuk melatih jaringan neural dan untuk mengestimasi probabilitas prior kelas waktu tunggal) dan kumpulan data testing (digunakan untuk evaluasi dan komparasi kinerja).

**Tabel 1.** Data Pelatihan dan Pengujian.

Jumlah Pixel Setiap Kelas	
Pelatihan	Pengujian
1.800	2.700

**Tabel 2.** Atribut Probabilistic Neural Network (PNN).

Parameter PNN	Data
Network Model	Basic
Kernel	Gauss
Allowable Error	0,001
Sigma Low	0,003
Sigma High	5,0
Sigma Tries	5,0

**Tabel 3.** Atribut Algoritma Ekspektasi Maksimum (EM).

Parameter of EM	Data
Error Criteria	0,001
Size of Join Probabilitas (JP)	12 x 12
Number of Epoch	288

Salah satu aspek dari tekstur terkait dengan distribusi spasial dan ketergantungan spasial tingkat keabuan dalam wilayah tertentu. Salah satu bentuk dapat dinyatakan bahwa ketergantungan dari probabilitas kemunculan bersama (kookuren) piksel-pikseltingkat keabuan  $g_1$  dengan tingkat keabuan  $g_2$ . Tingkat keabuan kookuren dapat direpresentasikan dalam sebuah matriks C di mana setiap unsur  $c_{ij}$  nya adalah nilai fungsi  $P(i, j, d, a)$ . Fungsi  $P(i, j, d, a)$  dapat dilihat sebagai probabilitas okuren pikseltingkat keabuan  $g_i$  dengan jarak  $d$  dengan piksel tingkat keabuan  $g_j$  pada arah  $a$   $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , dan  $135^\circ$ . Langkah-langkahnya adalah sebagai berikut:

- Tentukan banyaknya tingkat keabuan yang berbeda dalam citra tersebut, kemudian diurutkan dari kecil ke besar.
- Bentuklah matrik A berukuran  $k \times k$  dimana  $k$  adalah jumlah tingkat keabuan, sedangkan elemen  $a_{ij}$

menyatakan kemunculan piksel dengan tingkat keabuan  $g_i$  muncul berdampingan dengan piksel tingkat keabuan dari  $g_j$  pada arah  $0^\circ$  dimana  $\leq i, j \leq k$ .

- c. Matriks kookuren C dibentuk dengan membagi setiap elemen dari matriks A dengan n jumlah dari semua elemen matriks A.
- d. Langkah selanjutnya adalah melakukan perhitungan statistik pada matriks kookuren C. Jika  $C_{ij}$  merupakan elemen dari kolom dan baris ke-j dan ke-i dari matriks C berukuran  $k \times k$ , maka beberapa karakteristik tekstur dapat diperoleh melalui perhitungan di bawah.

1) **Angular Second Moment (ASM)**

$$\sum_{i=1}^k \sum_{j=1}^k c_{ij}^2 \dots\dots\dots(1)$$

ASM menunjukkan homogenitas atau keragaman tekstur. Tekstur yang lebih homogeny apabila semakin kecil ukuran matriks, tetapi nilai setiap elemannya semakin besar, maka tekstur yang lebih homogen.

2) **Entropy**

Entropi adalah derajat keacakan (randomness) dari tekstur. Entropi diambil dari nilai terbesar jika semua elemen dari  $c_{ij}$  sama.

$$-\sum_{i=1}^k \sum_{j=1}^k c_{ij}^2 \log c_{ij}^2 \dots\dots\dots(2)$$

3) **Elemen Difference Moment of  $m^{th}$  Order**

Momen tingkat kedua (perbedaan momen elemen-elemeorde dua) yang sering disebut sebagai teksturkontras.

$$\sum_{i=1}^k \sum_{j=1}^k (i-j)^m c_{ij} \dots\dots\dots(3)$$

4) **Inverse Elemen Difference Moment of  $m^{th}$  Order**

Momen tingkat pertama (elemen invers perbedaan momen tingkat pertama) yang sering disebut sebagai tekstur homogenitas.

$$\sum_{i=1}^k \sum_{j=1}^k \frac{c_{ij}}{(i-j)^m} \dots\dots\dots(4)$$

5) **Maximum Probability (max  $c_{ij}$ )**

Nilai terbesar dari probabilitas kemunculan tingkat keabuan  $g_i$ , berdekatan dengan tingkat keabuan  $g_j$  dalam citra.

6) **Minimum Probability (min  $c_{ij}$ )**

Nilai minimum probabilitas kemunculan yang kecil tingkat keabuan  $g_i$ , berdekatan dengan tingkat keabuan  $g_j$  dalam citra.

7) **Avarage Probability (avr  $c_{ij}$ )**

Nilai rerata probabilitas kemunculan yang besar dari tingkat keabuan  $g_i$ , berdekatan dengan tingkat keabuan  $g_j$  dalam citra.

8) **Mean Probability (mean  $c_{ij}$ )**

Nilai mean probabilitas kemunculanyang besar dari tingkat keabuan  $g_i$ , berdekatan dengan tingkat keabuan  $g_j$  dalam citra.

9) **Median Probability (med  $c_{ij}$ )**

Nilai median probabilitas kemunculan yang besar dari yang dominan dari tingkat keabuan  $g_i$ , berdekatan dengan tingkat keabuan  $g_j$  dalam citra.

10) **Modus Probability (mods  $c_{ij}$ )**

Nilai modus dari probabilitas kemunculan yang besar dari tingkat

keabuan  $g_i$ , berdekatan dengan tingkat keabuan  $g_j$  dalam citra.

**11) Correlation**

$$\sum_{i=1}^k \sum_{j=1}^k (i-m) (j-m) c_{ij} \dots\dots(5)$$

$m$  adalah probabilitas rata-rata okuren (nilai rata-rata dari semua elemen matriks C).

**12) Cluster Shade**

$$\sum_{i=1}^k \sum_{j=1}^k (i-j-2m)^2 c_{ij} \dots\dots\dots(6)$$

$m$  adalah probabilitas rata-rata okuren (nilai rata-rata dari semua elemen matriks C).

**Jaringan Neural Probabilistik (PNN)**

Jaringan neural yang sering digunakan untuk mengklasifikasikan pola didasarkan pada pembelajaran dari beberapa sampel. Perbedaan paradigma jaringan neural terdapat pada penggunaan aturan pembelajaran yang berbeda, tetapi kesemuanya memiliki cara yang sama dalam menentukan pola statistik yaitu dari suatu kumpulan sampel pelatihan kemudian mengklasifikasi pola-pola baru berdasarkan statistik tersebut.

Metode saat ini seperti backpropagasi menggunakan pendekatan heuristik untuk menemukan statistik yang mendasari pola kelas obyek. Pendekatan heuristik biasanya melibatkan banyak modifikasi terhadap parameter sistem yang secara bertahap meningkatkan

kinerja sistem. Selain membutuhkan waktu komputasi lama untuk pelatihan, pendekatan adaptif yang dapat meningkatkan backpropagasi terbukti rentan menuju nilai minimum yang salah (local minima). Untuk memperbaiki pendekatan ini, metode klasifikasi berdasarkan prinsip-prinsip statistik yang telah mantap berusaha ditemukan.

Hal ini akan ditunjukkan bahwa jaringan yang dihasilkan, walaupun struktur yang serupa dengan backpropagasi dan perbebedaan utama dalam fungsi aktivasi sigmoid diganti dengan statistik turunan kesatu, memiliki fitur unik yang mudah menemukan kondisi tertentu yang memenuhi keputusan yang dapat diimplementasi oleh jaringan neural probabilistik (PNN) secara asimtot mendekati keputusan Bayes yang optimal.

Memahami paradigma dasar PNN, penting untuk memulai pembahasan tentang startegy keputusan Bayes dan estimator nonparametrik dari fungsi probabilitas densitas. Kemudian ditunjukkan bagaimana memetakan teknik statistik pada tipe struktur jaringan neural umpan-maju dengan banyak pemroses sederhana ("neuron") yang bisa berfungsi secara paralel.

Akurasi keputusan tertentu tergantung pada keakuratan dengan estimasi menggunakan fungsi probabilitas densitas PDF. Membuat sebuah famili fungsi untuk estimasi fungsi  $f(X)$  seperti di bawah ini.

$$f_n ( X ) = \frac{1}{n \lambda} \sum_{i=1}^n \varpi \left( \frac{X - X_{Ai}}{\lambda} \right) \dots\dots\dots(7)$$

Misal  $X_{A1}, \dots, X_{Ai}, \dots, X_{-an}$  adalah variabel acak terikat identik terdistribusi sebagai suatu variabel acak  $X$  yang mana fungsi distribusi  $F(X) = P [x \leq X]$  pasti kontinyu. Kondisi Parzen pada fungsi bobot  $\varpi(y)$  sebagai berikut:

$$\sup_{-\infty < y < +\infty} |\varpi(y)| < \infty \dots\dots\dots (8)$$

di mana sup menunjukkan supremum.

$$\int_{-\infty}^{+\infty} |\varpi(y)| dy < \infty, \dots\dots\dots (9)$$

$$\lim |\varpi(y)| = 0, \dots\dots\dots (10)$$

$$E |f_n(X) - f(X)|^2 \rightarrow 0 \text{ as } n \rightarrow \infty \dots\dots\dots (13)$$

Konsistensi yang terdefinisi menyebutkan bahwa ekspektasi kesalahan semakin kecil dengan estimasi berdasarkan pada kumpulan data yang lebih besar, sangat penting karena hal itu berarti bahwa distribusi yang benar akan didekati dengan cara yang halus.

$$f_A(X) = \frac{1}{(2\pi)^{p/2} \sigma^p} \frac{1}{m} \sum_{i=1}^m \exp[ -(X - X_{Ai})^t (X - X_{Ai}) / 2 \sigma^2 ] \dots\dots\dots(14)$$

where

$i$  = jumlah pola

$m$  = jumlah pola pelatihan

$X_{Ai}$  = pola pelatihan ke  $i^{th}$  dari kategori  $\theta_A$

$\sigma$  = parameter perata

dan

$$\int_{-\infty}^{+\infty} \varpi(y) dy = 1, \dots\dots\dots (11)$$

pada persamaan (7),  $\lambda = \lambda(n)$  dipilih sebagai fungsi dari  $n$  sedemikiansehingga:

$$\lim_{y \rightarrow \infty} \lambda(n) = 0. \dots\dots\dots (12)$$

Terbukti bahwa fungsi estimasi  $f_n(X)$  konsisten dalam mean kuadrat yang berarti bahwa:

Hasil Parzen bisa dikembangkan untuk estimasi dalam kasus khusus dimana kernel multivariasi adalah hasil dari kernel univariasi. Dalam kasus tertentu kernel Gaussian, estimasi multivariasi dapat dinyatakan sebagai berikut.:

$p$  = dimasi ukuran ruang

Perhatikan bahwa  $f_A(X)$  adalah penjumlahan sederhana dari distribusi Gaussian multivariasi yang kecil yang terpusat pada setiap sampel pelatihan. Namun, jumlahnya tidak terbatas untuk menjadi Gaussian. Hal ini dapat terjadi,

pada kenyataannya, estimasi fungsi probabilitas densitas yang halus.

Ada kesamaan yang nyata antara jaringan analog paralel yang mengklasifikasi beberapa pola menggunakan estimator nonparametric sebuah fungsi probabilitas densitas (PDF) dan jaringan neural umpan maju yang digunakan dengan algoritma pelatihan yang lain (Swain, 1978).

**Klasifikasi data Gabungan**

Pertimbangkan masalah klasifikasi sebuah wilayah geografis yang diperoleh pada dua waktu  $t_1$  dan  $t_2$ , masing-masing. Setiap kumpulan data mungkin memuat gambar berasal dari waktu yang berbeda. Asumsikan bahwa semua gambar dari dua kumpulan data mengacu pada luas tanah yang sama, dan bahwa mereka coregistered dan tepat berubah menjadi resolusi spasial yang sama.

Secara umum, informasi kontekstual spasial dan temporal berperan penting dalam proses klasifikasi. Salah satu tujuan utama dari penelitian ini untuk menguji yang potensi teknik yang diusulkan untuk mengestimasi probabilitas prior join kelas, yang berhubungan dengan konteks temporal dari dua kumpulan data. Oleh karena itu, sederhananya, penelitian berfokus pada konteks waktu saja, dan tidak secara eksplisit mempertimbangkan konteks spasial. Selanjutnya, diasumsikan bahwa, untuk setiap piksel

dari suatu kumpulan data, semua informasi dalam konteks temporal disalurkan oleh piksel spasial yang bersesuaian dengan kumpulan data lainnya. Hal ini tampaknya asumsi yang masuk akal untuk prosedur saat ini dengan yang mempertimbangkan hanya dua waktu pengambilan dan mengabaikan konteks spasial.

Hal ini dapat menentukan dua jenis klasifikasi: 1) "Klasifikasi Gabungan" dari setiap pasangan piksel  $(x_1, x_2)$ , yang meliputi penemuan "terbaik" pasangan kelas  $(w_i, v_j)$  yang akan diberikan pada setiap pasangan piksel; atau 2) klasifikasi piksel salah satu dari dua kumpulan data citra dengan memanfaatkan informasi yang terkandung di kedua kumpulan data citra  $I_1$  dan  $I_2$ .

Sebagai strategi klasifikasi, perlu mengadopsi aturan Bayes untuk meminimalkan kesalahan dan menerapkannya pada jenis klasifikasi, seperti, "Kklasifikasi Gabungan" dari  $x_1, x_2$ ,  $x_1 \in W_n$  dan  $x_2 \in V_n$ , yaitu :

$$P(w_i, v_j / X_1, X_2) = \max_{w_i, v_j} \{P(w_i, v_j / X_1, X_2)\} \dots \dots \dots (15)$$

Untuk klasifikasi tipe kedua, yaitu jika kumpulan data citra  $I_2$  diklasifikasikan, aturan Bayes menjadi  $x_2 \in V_n$  maka:

$$P(v_j / X_1, X_2) = \max_{v_j} \{P(v_j / X_1, X_2)\} \dots \dots \dots (16)$$

Masalah kedua klasifikasi, serta solusinya saling terkait erat. Dalam penelitian ini, akan terfokus hanya pada bentuk klasifikasi gabungan dua

kumpulan data citra. Dengan asumsi konvensional bahwa kondisi kelas bersifat independen dalam domain waktu, dapat ditulis sebagai berikut.

$$p(X_1, X_2 / w_i, v_j) = p(X_1 / w_i) p(X_2 / v_j) \dots\dots\dots (17)$$

Konsekuensinya, dari hasil penelitian sebelumnya dapat merealisasi nilai maksimum dengan :

$$\max_{w_i, v_j} \left\{ \frac{P(w_i / X_1) P(v_j / X_2) P(w_i, v_j)}{P(w_i) P(v_j)} \right\} \dots\dots\dots (18)$$

Secara umum, asumsi di atas dapat menjadi solusi suboptimal. Misalnya, apabila penutup lahan bersosiasi dengan piksel yang tidak berubah antara  $t_1$  dan  $t_2$  [misal,  $w_i \equiv v_j$  in (10)], seperti halnya bahwa beberapa sifat lahan, yang berkontribusi menentukan nilai-nilai ukuran sensorik, dapat disimpan diantara dua waktu akuisisi. Ini berarti korelasi antara  $X_1$  dan  $X_2$  tidak masuk perhitungan. Namun, mengadopsi asumsi di atas memungkinkan penyederhanaan masalah menjadi signifikan.

pemetaan many-to-one dari sebuah distribusi observasi. Sebagai sebuah algoritma populer berguna terutama dalam estimasi komponen distribusi gabungan.

Sebuah probabilitas prior  $P(w_i)$  dan  $P(v_j)$  diestimasi dari dua kumpulan pelatihan dengan menghitung frekuensi relatif masing-masing kelas. Estimasi dari term sisa (misal, probabilitas posterior kelas dan probabilitas join prior) perlu dipertimbangkan lebih seksama.

Formulasi umum dari algoritma EM terdiri atas dua langkah utama: langkah ekspektasi dan langkah maksimalisasi. Ekspektasi dihitung sehubungan dengan variabel yang tidak diketahui, menggunakan estimasi saat ini pada parameter dan kondisi pengamatan. Langkah memaksimalkan memberikan estimasi baru dari beberapa parameter. Kedua langkah ini diulang sampai konvergen.

**Estimasi Probabilitas Join dengan Algoritma Ekspektasi Maksimum (EM)**

Sebuah aspek penting dari algoritma EM adalah sifat konvergensi tersebut. Hal ini dimungkinkan untuk membuktikan bahwa, pada setiap iterasi, parameter estimasi memberikan peningkatan fungsi kemiripan sampai maksimum lokal tercapai. Terlepas dari kenyataan bahwa konvergensi dapat dipastikan, namun mustahil untuk memastikan bahwa algoritma akan menemukan kemungkinan maksimum global (hanya dalam kasus-kasus tertentu

memungkinkan untuk menjamin konvergensi ke maksimum global).

Penjelasan detail tentang algoritma EM dan aspek teoritis terkait ada di luar lingkup penelitian ini. Perlu memperhatikan literatur untuk analisis mendalam suatu algoritma.

**Estimasi Probabilitas Join Prior**

Dalam kasus ini estimasi hanya untuk

probabilitas join prior beberapa kelas, asumsinya tidak perlu memperbarui probabilitas posterior dan probabilitas prior beberapa kelas selama beberapa iterasi berurutan.

Probabilitas  $P(w_i, v_j)$  dianggap sebagai elemen matriks JP, (ukuran  $M_1 \times M_2$ ), yang dihitung dengan memaksimalkan persamaan seperti di bawah ini.

$$L(JP) = \prod_{q=1}^S \left( \sum_{w_n \in \Omega} \sum_{v_m \in N} P(w_n, v_m) P(X_1^q, X_2^q / w_n, v_m) \right) \dots\dots\dots (19)$$

dimana S adalah jumlah total piksel yang diklasifikasi dan  $X_k^q$  adalah piksel ke q dari citra  $I_k$ . Hal itu memungkinkan untuk

membuktikan bahwa persamaan rekursif digunakan untuk mengestimasi  $P(w_i, v_j)$  dengan memaksimalkan persamaan (20).

$$P_{k+1}(w_i, v_j) = \frac{1}{S} \frac{\sum_{q=1}^S P_k(w_i, v_j) P(X_1^q, X_2^q / w_i, v_j)}{\sum_{q=1}^S \sum_{w_n \in \Omega} \sum_{v_m \in N} P_k(w_n, v_m) P(X_1^q, X_2^q / w_n, v_m)} \dots\dots\dots(20)$$

dimana  $P_k(w_i, v_j)$  adalah estimasi iteratif probabilitas joinr pada iterasi ke k. Estimasi tersebut diinisialisasi dengan

memberikan probabilitas yang sama untuk setiap pasangan kelas.

$$P_0(w_i, v_j) = \frac{1}{M_1 M_2} \quad \forall w_i \in \Omega, v_j \in N \quad \dots\dots\dots (21)$$

Pada hipotesis yang dibuat dalam pendekatan ini, hal itu memungkinkan

untuk membuktikan bahwa persamaan (21) dapat ditulis sebagai:

$$P_{k+1}(w_i, v_j) = A_{ij} \sum_{q=1}^S \frac{P_k(w_i, v_j) P(w_i / X_1^q) P(v_j / X_2^q)}{\sum_{w_n \in \Omega} \sum_{v_m \in N} \frac{P_k(w_n, v_m)}{P_k(w_n) P_k(v_m)} P(w_n / X_1^q) P(v_m / X_2^q)} \dots\dots\dots (22)$$

Dimana  $A_{ij} = \frac{I}{S P(w_i)P(v_j)}$  ..... (23)

Algoritma ini diiterasi sampai konvergen dan akan tercapai apabila perbedaan

$$\max_{w_i, v_j} \{ |P_{k+1}(w_i, v_j) - P_k(w_i, v_j)| \} < \epsilon, \quad w_i \in \Omega, \quad v_j \in N \quad \dots\dots\dots (24)$$

Dimana  $\epsilon \in [0,1]$ , estimasi  $P(w_i, v_j)$  diperoleh dalam konvergensi yang kemudian diterapkan pada aturan klasifikasi gabungan

**Klasifikasi Join**

Dua citra penginderaan jauh multi waktu yang diperoleh pada  $t_1$  dan  $t_2$  pada wilayah yang sama untuk diuji. Perhatikan pasangan piksel terdiri atas piksel dari citra multi waktu yang diperoleh pada  $t_1$  dan piksel spasial yang sesuai dari citra multi waktu yang diperoleh pada waktu  $t_2$ : misal piksel tersebut ditandai dengan dimensi-d vektor fitur  $X_1$  dan  $X_2$ , masing-masing. Misal  $\Omega = \{w_1, w_2, \dots, w_n\}$  adalah himpunan kemungkinan kelas penutup lahan pada waktu  $t_1$ , dan  $N = \{v_1, v_2, \dots, v_m\}$  adalah himpunan kemungkinan kelas penutup lahan pada  $t_2$ . Perubahan penutup lahan dalam pasangan piksel

maksimum diantara estimasi pada dua iterasi berturutan di bawah suatu ambang. Lebih tepatnya, kriteria henti didefinisikan dengan persamaan berikut:

akan dideteksi jika dua kelas  $w_i$  dan  $v_j$ , untuk piksel tersebut dinyatakan berbeda.

Jika kita mengabaikan informasi kontekstual dalam domain spasial, misal kita klasifikasi setiap pasangan piksel secara independen satu dengan yang lainnya atas dasar hanya vektor fitur  $X_1$  dan  $X_2$ , klasifikasi yang optimal, dalam arti probabilitas kesalahan minimum, diberikan oleh aturan Bayes untuk kasus masalah klasifikasi gabungan. Aturan seperti itu perlu pasangan kelas  $(w_i, v_j)$ , memberikan vektor fitur yang teramati sebagai  $X_1$  dan  $X_2$ :

$$\max_{w_i, v_j} \left\{ P \left( \frac{w_i, v_j}{X_1 X_2} \right) \right\} \quad \dots\dots (25)$$

Pasangan kelas  $(w_i, v_j)$  yang menyediakan nilai maksimum dinyatakan sebagai berikut:

$$\max_{w_i, v_j} \left\{ \frac{p \left( \frac{X_1 X_2}{w_i, v_j} \right) P \left( \frac{v_j}{w_i} \right) p(w_i)}{p(X_1 X_2)} \right\} \Leftrightarrow \max_{w_i, v_j} \left\{ p \left( \frac{X_1 X_2}{w_i, v_j} \right) P \left( \frac{v_j}{w_i} \right) p(w_i) \right\} \quad \dots\dots\dots (26)$$

dimana  $p(X_1, X_2)$  dapat diabaikan, karena tidak tergantung pada  $w_i$  dan  $v_j$ .

Kedua persamaan di atas melibatkan estimasi fungsi  $n \times m$  yang didefinisikan

dalam dimensi ruang  $(2 \times d)$ . Estimasi ini dapat dilakukan dengan menggunakan satu kumpulan piksel pelatihan ("training set"). Sayangnya, dalam situasi nyata, sulit untuk memiliki

kumpulan pelatihan yang sesuai, seperti sejumlah besar piksel pelatihan untuk setiap kombinasi kemungkinan dari kelas  $w_i$  dan  $v_j$  diperlukan. Untuk menyederhanakan fungsi estimasi tersebut, dikenalkan hipotesis. Pertimbangkan vektor fitur  $X_i$  ( $i = 1,2$ ), terkait dengan waktu  $t_i$ , sebagai gabungan dari komponen sinyal  $S_i$  dan komponen gangguan  $N_i$ .

$$X_1 = S_1 + N_1 \dots\dots\dots (27)$$

dan

$$X_2 = S_2 + N_2 \dots\dots\dots (28)$$

Dengan asumsi bahwa sinyal  $S_i$  hanya tergantung pada kelas penutup lahan pada  $t_i$ , dan gangguan  $N_i$  hanya bergantung pada kelas penutup lahan pada  $t_i$  dan kemungkinan pada  $S_i$  (misalnya, untuk gangguan multiplikatif pada citra SAR ). Berdasarkan hipotesis ini, ketergantungan probabilitik kelas pada dua waktu, dan kita dapat menulis kan:

$$P\left(\frac{X_1, X_2}{w_i, v_j}\right) = P\left(\frac{X_1}{w_i}\right)P\left(\frac{X_2}{v_j}\right) \dots\dots\dots (29)$$

Dengan mensubstitusi, dan menerapkan beberapa transformasi, diperoleh bahwa nilai maksimum dapat digunakan pengambilan keputusan berikut:

$$\max_{w_i, v_j} \left\{ \frac{P\left(\frac{w_i}{X_1}\right)P\left(\frac{v_j}{X_2}\right)P\left(\frac{v_j}{w_i}\right)}{P(v_j)} \right\} \dots(30)$$

Dengan hipotesis yang didefinisikan di atas, untuk membentuk klasifikasi gabungan dua citra penginderaan jauh multi waktu perlu untuk mengestimasi probabilitas prior  $P(v_j)$  dari kelas obyek

pada  $t_1$  dan  $t_2$ , waktu tunggal, multivariasi, probabilitas kondisional  $P(w_i.X_1)$  dan  $P(v_j.X_2)$  pada dua waktu, dan probabilitas transisi  $P(v_j/w_i)$ .

**HASIL DAN DISKUSI**

Ukuran citra uji yang digunakan 350 x 350 piksel. Ciri turunan yang dihasilkan berdasarkan model kookuren. Jumlah kelas obyek citra Saguling adalah empat kelas meliputi air, lahan terbuka, vegetasi, dan perkampungan. Jumlah sampel masing-masing kelas adalah 4,500 piksel, 40% nya (1.800 piksel) digunakan untuk pelatihan, dan 60% nya (2.700 piksel) digunakan untuk pengujian.

Kesalahan klasifikasi obyek air dan perkampungan relatif jauh lebih kecil dari kelas lahan terbuka dan vegetasi. Klasifikasi obyek air dan perkampungan, konflik dengan dua kelas lainnya, sementara lahan terbuka dan vegetasi konflik dengan semua kelas yang ada. Pengklasifikasi ini lebih cocok untuk obyek air dan perkampungan.

**Tabel 4.** Klasifikasi menggunakan Jaringan Neural Backpropagasi (BP)

Object Class	Water body	Open Area	Vegetation	Villages	Producer's Accuracy - PA (%)
Water Body	2.562	67	62	-	95,21
Open Area	71	2.529	34	85	93,01
Vegetation	67	67	2.533	97	91,64
Villages	-	33	71	2.518	96,03
User's Accuracy - UA (%)	94,89	93,67	93,81	93,26	Overall Accuracy - OA=93.91
					PA=93,97 UA=93,94

**Tabel 5.** Klasifikasi menggunakan Jaringan Neural Probabilistik (PNN)

Object Class	Water body	Open Area	Vegetation	Villages	Producer's Accuracy - PA (%)
Water Body	2.609	57	42	-	96,34
Open Area	51	2.579	-	55	96,05
Vegetation	40	64	2.573	57	94,11
Villages	-	-	85	2.588	96,82
User's Accuracy - UA (%)	96,63	95,52	95,30	95,85	Overall Accuracy - OA=95.82
					PA=95,83 UA=95,82

Kelas air diinterpretasi oleh BP sebagai lahan terbuka dan vegetasi hampir sama. Demikian pula perkampungan diinterpretasi keliru sebagai lahan terbuka dan vegetasi. Dua kesalahan yang ada dikarenakan ukuran jendela kookuren terlalu kecil sehingga unsur substansi yang sama dapat dimiliki oleh kelas konflik (Tabel 4).

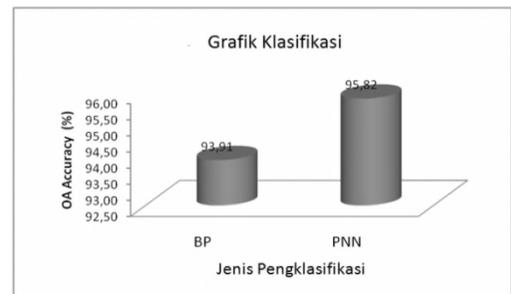
Pengklasifikasi PNN menunjukkan hasil yang lebih baik dari BP (Tabel 5), semua kelas konflik maksimum dengan satu kelas lainnya. PNN dapat mengurangi konflik antar kelas. Kasus konflik obyek air dan perkampungan masih sama tetapi jumlahnya dapat diturunkan. Kesalahan klasifikasi karena substansi obyek merupakan kelas yang berbeda, sebagai contoh substansi air dan ruang hijau di vegetasi dapat menyebabkan piksel dikategorikan sebagai obyek air dan desa. Kemudian substansi vegetasi dan ruang terbuka di perkampungan tidak diinterpretasi sebagai perkampungan.

Ekspektasi maksimum (EM) dapat meningkatkan kemampuan pengklasifikasi BP, mengurangi jumlah kelas konflik, semua kelas hanya konflik dengan maksimal satu kelas lainnya

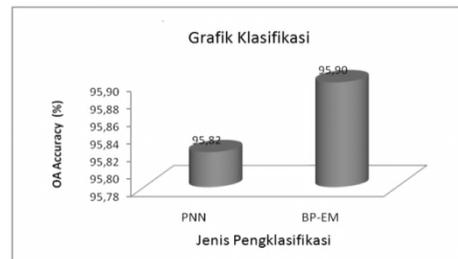
Kombinasi dari BP-EM juga lebih baik dari PNN sekitar 0,1%. Ekspektasi maksimum dapat berfungsi untuk memberikan kepastian bahwa substansi keputusan klasifikasi obyek yang tidak logis ditetapkan dalam kelas obyek terdekat (Tabel 6, Gambar 3).

**Tabel 6.** Klasifikasi menggunakan Jaringan Neural Backpropagasi dengan Ekspektasi Maksimum

Object Class	Water body	Open Area	Vegetation	Villages	Producer's Accuracy - PA (%)
Water Body	2.596	51	38	-	96,15
Open Area	55	2.587	-	58	96,15
Vegetation	49	62	2.583	51	96,15
Villages	-	-	79	2.591	96,15
User's Accuracy - UA (%)	96,15	95,81	95,67	95,96	Overall Accuracy - OA=95,90



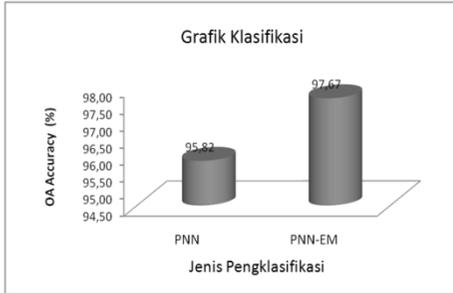
**Gambar 3.** Akurasi BP dan PNN.



**Gambar 4 .** Akurasi Pengklasifikasi PNN dan BP-EM.

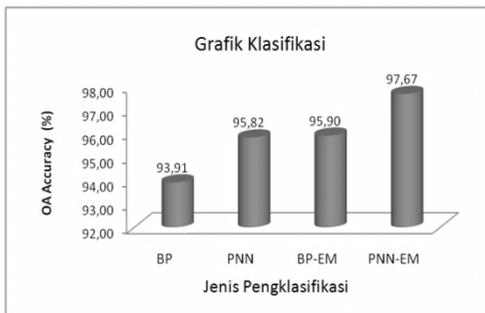
**Tabel 7.** Klasifikasi menggunakan Jaringan Neural Probabilistik dengan Ekspektasi Maksimum

Object Class	Water body	Open Area	Vegetation	Villages	Producer's Accuracy - PA (%)
Water Body	2.648	24	25	-	98,18
Open Area	8	2.625	-	42	98,13
Vegetation	44	51	2.637	20	95,82
Villages	-	-	34	2.638	98,73
User's Accuracy - UA (%)	98,07	97,22	97,81	97,70	Overall Accuracy - OA=97,67



**Gambar5.** Akurasi Klasifikasi PNN dan PNN-EM

Penggabungan PNN dengan EM dapat meningkatkan pengenalan obyek citra multi waktu sekitar 2% lebih dari skema yang lain. Kesalahan klasifikasi obyek air sebagai vegetasi masih sulit direduksi dengan EM (Tabel 7, Gambar 5).



**Gambar 6.** Akurasi Pengklasifikasi

Kesalahan klasifikasi objek air sebagai lahan terbuka berkurang dengan EM. Lahan terbuka dan perkampungan masih menjadi obyek yang sulit untuk klasifikasi dengan EM namun kesalahan klasifikasi telah dapat dikurangi.

**KESIMPULAN**

Terfokus pada aspek multi waktu, asumsi kami bahwa, untuk kesederhanaan, korelasi temporal diantara dua kumpulan data dapat diperhitungkan dengan probabilitas prior join dari beberapa kelas pada dua tanggal. Data multi waktu kemudian

dihasilkan dengan menggunakan beberapa mean jaringan neural probabilistik, yang memberikan estimasi nonparametrik dari probabilitas posterior kelas yang berdasarkan waktu tunggal.

Selain fleksibilitas yang diberikan oleh teknik nonparametrik, penggunaan jaringan neural menawarkan keuntungan umum dari setiap pendekatan jaringan neural : paralelisme intrinsic, kemampuan beradaptasi terhadap data, dan kehandalan terhadap noise dan kesalahan data pelatihan. Khususnya untuk model neural yang diadopsi, ada aturan umum untuk menentukan topologi jaringan neural dan menetapkan prosedur dari proses pelatihan, namun sulit untuk menafsirkan perilaku jaringan.

Peningkatan akurasi dan waktu pelatihan, disini menggunakan PNN. Salah satu prinsip keuntungan paradigma PNN adalah bahwa lebih jauh lebih cepat dari paradigma sebelumnya backpropagasi untuk masalah di mana peningkatan waktu adaptasi backpropagasi yang signifikan dari waktu komputasi total. Akurasi klasifikasi kira-kira sebanding, backpropagasi 93,91% sedangkan PNN menghasilkan 95,62% untuk akurasi umum.

Inovasi utama adalah penerapan algoritma EM untuk estimasi probabilitas prior join beberapa kelas. Probabilitas prior join biasanya dipilih secara manual oleh ahli atas dasar pengetahuan dari karakteristik dari wilayah geografisi yang diamati dari interval waktu diantara beberapa pengambilan. Keuntungan dari

algoritma EM terdiri atas kemungkinan menghitung beberapa estimasi probabilitas join secara langsung dari kumpulan data yang diklasifikasikan secara otomatis. Hal ini mengatasi kelemahan akan kebutuhan intervensi orang dan ketergantungan pada keakuratan pengetahuan permulaan.

### **UCAPAN TERIMA KASIH**

Penelitian ini diperoleh dari Direktorat Pendidikan Tinggi (DIKTI) melalui program penelitian strategis nasional tahun 2011. Penelitian ini dilakukan dengan bantuan dan kerjasama yang baik dari semua pihak. Untuk itu, kami ucapkan terima kasih kepada Ketua Laboratorium Kecerdasan Komputasional dan Pengolahan Citra, Program Ilmu Komputer Universitas Pendidikan Indonesia. Secara khusus terima kasih kepada Prof Aniati Murni dari Universitas Indonesia, dan Dr. Wiweka dari LAPAN. Semoga hasil penelitian ini bisa berkontribusi dalam upaya mengoptimalkan pengembangan penggunaan lahan dan penutup lahan di Indonesia.

### **DAFTAR PUSTAKA**

- Aniati Murni, 1997. Metodologi interpretasi citra inderaja multitemporal dan multisensor berdasarkan klasifikasi uniform, *Disertasi Pascasarjana UI*, Jakarta.
- Bin Tian, et al., 2000. Temporal Updating Scheme for Probabilistic Neural Network with Application to Satellite Cloud Classification, *IEEE Trans. On Neural Network*, Vol. 11 No. 4, July 2000.
- Byeungwoo J. et. al, 1999. *Decision Fusion Approach Multitemporal Classifications*.
- Fausett, L., 1994. *Fundamentals of Neural Networks*, Prentice Hall.
- Goldberg, D. E., 1989. *Genetic Algorithm in Search, Optimization, and Machine Learning*. Addison Wesley.
- H. M. Kalayeh and D. A. Landgrebe, 1986. Utilizing multitemporal data by stochastic model. *IEEE Trans. Geosci. Remote Sensing*, vol. GE-24, pp. 792-795, Sept. 1986.
- H. Solberg, A. K. Jain, and T. Taxt, 1997. Multisource classification of remotely sensed data: Fusion of landsat TM and SAR images. *IEEE Trans. Geosci. Remote Sensing*. vol. 32, pp. 100-113, Jan. 1997.
- J. A. Benediktsson, and P. H. Swain. 1989. A method of statistical multisource classification with a mechanism to weight the influence of the data sources. *In Proc. IEEE Trans. Geosci. Remote Sensing*, pp. 517-520, July 1989.
- J. A. Benediktsson, P. H. Swain, and O. K. Ersoy. 1990. Neural network approach versus statistical method in classification of multisource remote sensing data. *IEEE Trans. Geosci. Remote Sensing*. vol. 28, pp. 540-552, July 1990.
- Jose R. G. A., et al, 1995. Applications of the EM Algorithm to the Analysis of Life Length Data", *Applied Statist.* Vo. 44, No. 3.
- Josef K. et al, 1998. On Combining Classifiers. *IEEE Trans. On Pattern*

- Analysis and Machine Intelligence*, Vol. 20, No. 3, March.
- L. Bruzzone and S. B. Serpico, 1997. An iterative technique for detection of land-cover transitions in multitemporal remote-sensing images. *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 858-867.
- L. Bruzzone, C. Conese, F. Maselli, and F. Roli, 1997. Multisource classification of complex rural areas by statistical and neural network approaches. *Photogramm. Eng. Remote Sens.*, vol. 63, no. 5, pp. 523-533, May 1997.
- L. Bruzzone, D. F. Prieto, and S. B. Serpico, 1999. A Neural – Statistical Approach to Multitemporal and Multisource Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 1350-1358, May 1999.
- Miguel A. et al, 2000. Mode-Finding for Mixtures of Gaussian Distributions. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, Nov. 2000.
- P. Dempster, N. M. Laird, and D. B. Rubin, 1997. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc.*, vo. 39, no. 1, pp. 1-38.
- P. H. Swain, 1978. Bayesian classification in a time-varying environment. *IEEE Trans. Syst. Man. Cyber.*, vol. SMC-8, pp. 880-883, Dec. 1978.
- T. Lee, J. A. Richard and P. H. Swain, 1987. Probabilistic approach for multisource data analysis. *IEEE Trans. Geosci. Remote Sensing*, vol. GE. 25, pp. 283-293.