

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/326893279>

Perbandingan Jaringan Learning Vector Quantization dan Backpropagation pada Klasifikasi Daun Berbasiskan Fitur Gabungan

Article · June 2017

CITATION

1

READS

307

2 authors:



Sigit Adinugroho

Brawijaya University

25 PUBLICATIONS 19 CITATIONS

[SEE PROFILE](#)



Yuita Arum Sari

Brawijaya University

78 PUBLICATIONS 68 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Text clustering [View project](#)



Melanoma Identification [View project](#)

Perbandingan *Jaringan Learning Vector Quantization* dan Backpropagation pada Klasifikasi Daun Berbasis Fitur Gabungan

Sigit Adinugroho¹, Yuita Arum Sari²

Fakultas Ilmu Komputer, Universitas Brawijaya^{1,2}

Jalan Veteran 8, Malang, INDONESIA^{1,2}

Email : sigit.adinu@gmail.com¹, yuita@ub.ac.id²

Abstract— *Accurate automated plant identification is required to solve various agricultural problems. In this research, a comparison between Backpropagation neural network and Learning Vector Quantization is conducted to classify plants based on leaves' characteristics. First, 31 features which represent shape, color, and texture are extracted from leaves images. Then, Backpropagation network and LVQ are utilized to classify the data. Evaluation on Backpropagation network yields 0.952 of accuracy while the maximum accuracy of LVQ reaches 0.420.*

Intisari—Identifikasi tanaman otomatis yang akurat diperlukan untuk menyelesaikan berbagai permasalahan di bidang pertanian. Pada penelitian ini, dilakukan perbandingan kinerja jaringan syaraf tiruan *Backpropagation* dan *Learning Vector Quantization* untuk mengelompokkan tanaman berdasarkan karakteristik daun. Langkah pertama, dilakukan ekstraksi 31 fitur dari daun yang merepresentasikan bentuk, warna, dan tekstur. Selanjutnya, jaringan *Backpropagation* dan *LVQ* digunakan untuk klasifikasi data daun. Proses evaluasi pada *Backpropagation* menghasilkan akurasi maksimal sebesar 0,952 sedangkan akurasi maksimal pada jaringan *LVQ* adalah 0,420.

Kata Kunci— identifikasi daun, klasifikasi daun, *LVQ*, *Backpropagation*.

I. PENDAHULUAN

Klasifikasi tanaman secara otomatis diperlukan sebagai dasar untuk menyelesaikan masalah-masalah kompleks seperti prediksi jumlah panen, estimasi laju pertumbuhan tanaman, serta memperkirakan kesehatan tanaman [1]. Selain itu, proses klasifikasi tanaman juga mampu menjaga kelestarian tanaman di muka bumi [2].

Proses identifikasi tanaman umumnya melibatkan ahli taksonomi dengan bantuan herbarium, yaitu sekumpulan contoh tanaman yang diawetkan. Seorang ahli taksonomi akan membandingkan suatu tanaman asing dengan contoh yang terdapat pada herbarium untuk mengetahui spesies dari tanaman asing tersebut [3].

Terdapat kelemahan dari proses identifikasi tumbuhan secara manual. Pertama, jumlah ahli taksonomi cenderung berkurang [4]. Kedua, sulitnya melakukan klasifikasi karena kategori tanaman yang sangat banyak [5]. Keempat, proses pengenalan secara manual memerlukan adanya referensi, yaitu herbarium [6]. Kelima, penggunaan herbarium dalam indentifikasi tanaman mempunyai kelemahan, yaitu persebaran herbarium yang cenderung lokal [7]. Artinya, suatu herbarium mungkin hanya tersedia di suatu wilayah tertentu saja.

Salah satu alternatif dalam pengenalan tumbuhan adalah menggunakan proses automasi dengan bantuan komputer.

Sistem ini biasa disebut dengan *Computer Assisted Taxonomy (CAT)* [8]. Tahap awal dari proses identifikasi tanaman secara otomatis adalah pengambilan gambar dari objek. Pengambilan gambar dapat dilakukan secara *online* menggunakan *scanner* ataupun *offline* menggunakan kamera digital. Selanjutnya, dilakukan pengolahan awal gambar untuk mengurangi derau. Setelah itu, fitur-fitur pada gambar diambil. Jika diperlukan, proses reduksi dimensi data dapat dilakukan. Terakhir, data yang telah diproses akan dikenali oleh suatu metode pengenalan pola dengan cara membandingkan terhadap pola-pola yang tersedia pada data latih.

Daun merupakan bagian tanaman yang dapat digunakan untuk mengetahui spesies dari sebuah tanaman. Penggunaan daun untuk melakukan identifikasi mempunyai keuntungan yaitu jumlahnya yang banyak dan hampir selalu tersedia sepanjang musim [1]. Selain itu daun juga memiliki struktur sederhana yang dapat disajikan dalam ruang dua dimensi. Lain halnya dengan bunga, biji, dan morfologi tanaman yang rumit dan memerlukan ruang tiga dimensi untuk menyajikannya [9]. Daun juga diketahui memiliki karakteristik unik yang mampu membedakan jenis tanaman. Beberapa karakteristik tersebut adalah bentuk, warna, tekstur, dan garis tepi [10].

Terdapat berbagai macam algoritme untuk melakukan pengenalan pola yang dapat diterapkan untuk melakukan identifikasi tanaman berdasarkan fitur daun. Namun selama ini sebagian besar metode pengenalan tanaman menggunakan metode berbasis *nearest neighbor* [11]. Metode berbasis ketetanggaan memiliki keunggulan, yaitu sederhana dan mudah diimplementasikan. Namun, terdapat pula kelemahan yaitu lambat jika data yang digunakan berjumlah banyak [12]. Metode alternatif yang dapat digunakan untuk mengenali pola daun adalah jaringan syaraf tiruan. Perbandingan kinerja antara metode berbasis ketetanggaan dan jaringan syaraf tiruan menunjukkan bahwa jaringan syaraf tiruan mampu mengidentifikasi daun lebih baik daripada *nearest neighbor* [13].

Jaringan *Backpropagation* dan *Learning Vector Quantization (LVQ)* merupakan algoritme jaringan syaraf tiruan yang paling sering digunakan untuk menyelesaikan permasalahan pengenalan pola. *Backpropagation* memiliki keunggulan yaitu menemukan hasil yang optimum, meskipun hanya *local optima*, berdasarkan fungsi *error*. Di sisi lain, *LVQ* melakukan klasifikasi berdasarkan vektor contoh dari data latih yang bersifat semi-optimal [14]. Dari segi kinerja, *LVQ* dan

Backpropagation memiliki kinerja yang tidak jauh berbeda [15].

Berdasarkan permasalahan yang ada, paper ini akan membandingkan kinerja jaringan syaraf tiruan *Backpropagation* dan LVQ dalam melakukan identifikasi tanaman. Fitur-fitur yang digunakan berasal dari karakteristik yang terdapat pada daun.














II. METODE DAN DATA



A. Dataset

Data yang digunakan pada penelitian ini berasal dari dataset *Swedish Leaf Dataset* [16]. Dataset tersebut berisi kumpulan foto daun dari 15 spesies tanaman. Setiap spesies memiliki 75 gambar daun, sehingga total gambar yang tersedia adalah 1125. Detail spesies dan contoh daun ditampilkan pada TABEL I. Dataset *Swedish Leaf Dataset* dapat diunduh di <http://www.cvl.isy.liu.se/en/research/datasets/swedish-leaf/>.

Gambar daun pada *Swedish Leaf Dataset* diperoleh melalui proses *scanning* sehingga mudah digunakan karena persebaran cahaya yang merata dan tidak adanya daun yang bertumpukan. Permasalahan yang terdapat pada dataset tersebut adalah tingginya kemiripan daun antara spesies yang berbeda [16]. Penelitian ini fokus pada jenis daun tunggal, sehingga spesies *Sorbus aucuparia* dan *Sorbus intermedia* tidak digunakan karena mengandung daun majemuk.

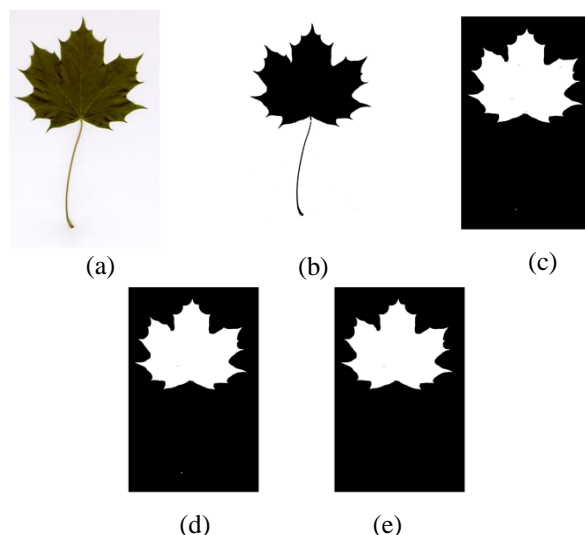
TABEL I
DETAIL SWEDISH LEAF DATASET

No	Spesies	Contoh daun
1	<i>Ulmus carpinifolia</i>	
2	<i>Acer</i>	
3	<i>Salix aurita</i>	
4	<i>Quercus</i>	
5	<i>Alnus incana</i>	
6	<i>Betula pubescens</i>	
7	<i>Salix alba 'Sericea'</i>	
8	<i>Populus tremula</i>	
9	<i>Ulmus glabra</i>	
10	<i>Sorbus aucuparia</i>	
11	<i>Salix sinerea</i>	
12	<i>Populus</i>	
13	<i>Tilia</i>	

14	<i>Sorbus intermedia</i>	
15	<i>Fagus silvatica</i>	

B. Segmentasi

Proses segmentasi dilakukan untuk membuang bagian yang tidak diperlukan dari gambar daun dan mengekstrak *mask* dari daun saja. Proses ini juga menghilangkan tangkai daun yang terdapat pada beberapa gambar daun. Segmentasi diawali dengan mengubah gambar RGB menjadi gambar biner. Selanjutnya, tangkai daun dihilangkan melalui proses erosi menggunakan *structuring element* berbentuk horizontal. Lubang-lubang kecil pada daun dihilangkan dengan filter median dengan ukuran ketetanggaan 3x3. Tahap terakhir dari proses segmentasi adalah pengambilan *mask* pada gambar yang memiliki luasan terbesar.



Gambar 1. Hasil dari setiap tahapan segmentasi. (a) Gambar RGB (b) Gambar biner (c) Hasil erosi (d) Hasil median filter (e) Hasil akhir berupa *mask* daun

C. Ekstraksi fitur

Berdasarkan gambar hasil segmentasi, sejumlah fitur diekstrak untuk keperluan pengenalan tanaman. Penelitian ini menggunakan 31 fitur yang menggambarkan karakteristik dasar dari daun, yaitu bentuk, warna, dan tekstur.

1) *Bentuk* : Terdapat 15 fitur yang menggambarkan karakteristik daun tanaman.

a) Roundness

Roundness pada sebuah daun menjelaskan rasio antara keliling dan luas daun [17]. Jika diketahui luas daun sebesar A dan keliling P , nilai *roundness* R dapat dihitung menggunakan Persamaan 1.

$$R = \frac{7\pi}{P^2} \quad (1)$$

b) Rasio keliling dengan diameter

Rasio keliling dengan diameter menjelaskan perbandingan antara keliling P dengan diameter D dari sebuah daun [18]. Rumus perhitungan rasio keliling dengan diameter disajikan pada Persamaan 2.

$$P_D = \frac{P}{D} \quad (2)$$

c) *Compactness*

Compactness mengukur rasio antara keliling daun dengan luas daun. *Compactness* mampu membedakan kompleksitas dan bentuk antar daun yang berbeda [19]. Perhitungan *compactness* C dilakukan menggunakan Persamaan 3.

$$C = \frac{P}{\sqrt{A}} \quad (3)$$

d) *Eccentricity*

Eccentricity menunjukkan seberapa besar perbedaan bentuk daun terhadap lingkaran sempurna [1]. Berdasarkan titik fokus f dan panjang sumbu mayor a dari sebuah daun, *eccentricity* E dihitung menggunakan persamaan sebagai berikut :

$$E = \frac{f}{a} \quad (4)$$

e) *Aspect ratio*

Aspect ratio AR menunjukkan rasio panjang sumbu mayor L dengan panjang sumbu minor W . Ukuran ini mengukur seberapa lebar atau panjang sebuah daun [1]. Perhitungan *aspect ratio* ditunjukkan dalam Persamaan 5.

$$AR = \frac{L}{W} \quad (5)$$

f) *Rectangularity*

Rectangularity R mendeskripsikan seberapa mirip bentuk daun dengan sebuah persegi panjang berdasarkan rasio luas A dengan batas persegi panjang minimum LW [9]. Perhitungan *rectangularity* dilakukan dengan Persamaan 6.

$$N = \frac{A}{LW} \quad (6)$$

g) *Narrow factor*

Narrow factor NF merupakan rasio dari diameter daun D dengan panjang sumbu mayor L [18]. *Narrow factor* dihitung dengan Persamaan 7.

$$NF = \frac{D}{L} \quad (7)$$

h) Rasio antara keliling dengan panjang sumbu mayor dan minor

Rasio antara keliling dengan panjang sumbu mayor dan minor P_{LW} mengukur rasio antara keliling daun P dengan jumlah panjang sumbu mayor L dan minor W . Persamaan untuk menghitung P_{LW} ditunjukkan oleh Persamaan 8.

$$P_{LW} = \frac{P}{(L+W)} \quad (8)$$

i) *Hu moment*

Hu moment merupakan 7 *moment* yang mendeskripsikan bentuk dari sebuah benda. *Hu moment* tidak dipengaruhi oleh transisi, rotasi, dan perubahan ukuran [20]. Pada bidang visi computer, *Hu moment* telah teruji mampu mengenali pola dalam sebuah gambar [21].

2) *Warna*

Warna merupakan fitur penting pada daun, yang mampu membedakan jenis antar daun. Fitur-fitur yang berhubungan dengan warna umumnya didefinisikan pada *color space* tertentu. Dataset *Swedish Leaf Dataset* menggunakan *color space* RGB sehingga fitur-fitur yang akan dibentuk mengikuti *color space* tersebut. Fitur warna yang paling akurat untuk klasifikasi daun adalah *Color Moment (CM)* [22]. Keuntungan menggunakan *CM* adalah dimensinya yang kecil dan komputasinya yang sederhana [11]. *CM* terdiri empat

deskriptor sederhana yaitu rata-rata, standar deviasi, kurtosis, dan *skewness*. Secara total, terdapat 12 fitur warna karena masing-masing deskriptor pada *CM* dihitung pada masing-masing komponen warna *RGB*. Rata-rata, standar deviasi, kurtosis, dan *skewness* pada masing-masing komponen warna x dihitung menggunakan Persamaan 9 – 12

$$\bar{x} = \frac{\sum_{i=1}^N \sum_{j=1}^M x_{ij}}{NM} \quad (9)$$

$$\sigma_i = \sqrt{\frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (x_{ij} - \bar{x})^2} \quad (10)$$

$$S_i = \sqrt[3]{\frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (x_{ij} - \bar{x})^3} \quad (11)$$

$$K_i = \sqrt[4]{\frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (x_{ij} - \bar{x})^4} \quad (12)$$

3) *Tekstur*

Tekstur menggambarkan kehalusan atau kekasaran suatu benda dan merupakan fitur yang umum digunakan dalam pengenalan pola gambar [23]. Sifat tersebut dianalisis berdasarkan relasi sekumpulan piksel dalam sebuah gambar keabuan (*grayscale*).

Salah satu fitur tekstur yang banyak digunakan adalah *Gray Level Co-occurrence Matrix (GLCM)* [11]. GLCM merupakan probabilitas sebuah level keabuan i muncul di sekitar level keabuan j dengan jarak d , sudut θ , dan jumlah level warna N [24].

$$G = \Pr(i, j | d, \theta, N) \quad (13)$$

Bentuk simetris dari GLCM dihasilkan dengan menjumlahkan matriks GLCM dengan transposenya dan melakukan normalisasi dengan membagi setiap elemen GLCM dengan jumlah semua elemen matriks GLCM.

Matriks GLCM simetris ternormalisasi menjadi dasar pembentukan statistik yang mendeskripsikan tekstur dari sebuah objek. Empat statistic diantaranya adalah kontras, korelasi, energi, dan homogenitas.

a) *Kontras*

Kontras menggambarkan variasi dari nilai GLCM. Berdasarkan elemen (i, j) pada matriks GLCM simetris ternormalisasi $S(i, j)$ dengan jarak, sudut, dan jumlah level tertentu, kontras dapat dihitung menggunakan Persamaan 14.

$$\text{Kontras} = \sum_{i=1}^N \sum_{j=1}^N [(i - j)^2 S(i, j)] \quad (14)$$

b) *Korelasi*

Korelasi mengukur peluang kemunculan suatu pasangan nilai. Korelasi dihitung menggunakan Persamaan 15.

$$\text{Korelasi} = \frac{1}{\sigma^2} \{ (\sum_i \sum_j i \cdot j \cdot S(i, j)) - \mu^2 \} \quad (15)$$

c) *Energi*

Energi mengukur jumlah dari kuadrat elemen-elemen pada matriks GLCM simetris ternormalisasi. Energi dihitung menggunakan Persamaan 16.

$$\sum_{i=1}^N \sum_{j=1}^N \{S(i,j)\}^2 \quad (16)$$

d) Homogenitas

Homogenitas mengukur seberapa homogen tekstur dalam gambar, atau dengan kata lain, mengukur seberapa mirip i dan j . Homogenitas dihitung menggunakan Persamaan 17.

$$\sum_{i=1}^N \sum_{j=1}^N \frac{S(i,j)}{1+(i-j)^2} \quad (17)$$

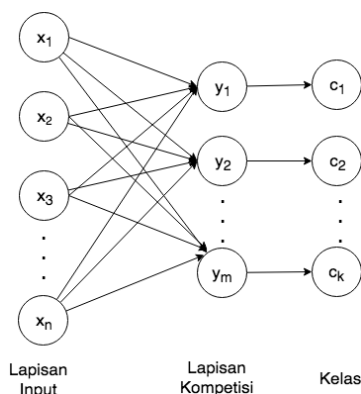
D. Klasifikasi Menggunakan Jaringan Syaraf Tiruan

1) Learning Vector Quantization

Learning Vector Quantization (LVQ) merupakan metode klasifikasi berbasis jaringan syaraf tiruan yang mengimplementasikan konsep kompetisi. Kompetisi pada LVQ berbasiskan *winner-take-all* yang berarti hanya ada satu neuron pemenang dalam proses kompetisi. Dibandingkan dengan model jaringan syaraf tiruan lainnya, LVQ memiliki keunggulan dari sisi kecepatan [25].

Jaringan LVQ merupakan jaringan *single layer* yang terdiri dari lapisan kompetisi, selain lapisan input. Arsitektur LVQ ditampilkan pada Gambar 2. Banyaknya neuron pada lapisan input sesuai dengan banyaknya fitur pada data. Setiap neuron pada lapisan kompetisi diberi sebuah label/kelas sesuai dengan kelas pada data, namun umumnya beberapa neuron pada lapisan kompetisi memiliki kelas yang sama. Oleh karena itu, banyaknya neuron pada lapisan kompetisi paling tidak sama dengan banyaknya kelas pada data [26].

Arsitektur jaringan LVQ yang digunakan pada penelitian ini terdiri dari 31 *input neuron* yang sesuai dengan banyaknya fitur pada data. Sementara itu, banyaknya neuron pada lapisan kompetisi akan diuji pada Bab III untuk menentukan banyaknya neuron dengan akurasi tertinggi. Banyaknya neuron pada lapisan kompetisi paling sedikit 13 neuron, menyesuaikan banyaknya kelas pada data.



Gambar 2. Arsitektur LVQ

Proses pembelajaran (*training*) pada LVQ menggabungkan konsep kompetisi dan *supervised learning*. Sama seperti proses *supervised learning* lainnya, LVQ memerlukan data latih yang

telah diberi label/kelas. Proses pembelajaran pada LVQ secara lengkap adalah sebagai berikut :

1. Inisialisasi bobot pada lapisan kompetisi (w_j) dan tentukan nilai α (*learning rate*).
2. Selama kondisi berhenti belum terpenuhi, lakukan langkah 3-7.
3. Untuk setiap pasangan data latih beserta targetnya ($x_i : t_i$) lakukan langkah 4-5.
4. Tentukan neuron pemenang (J) pada lapisan kompetisi yang memiliki jarak Euclidean terkecil dengan x_i .
5. Ubah bobot neuron pemenang (w_j) sebagai berikut:
 - Jika $c_j \neq t_j$ maka $w_j = w_j - \alpha[x_i - w_j]$
 - Jika $c_j = t_j$ maka $w_j = w_j + \alpha[x_i - w_j]$
6. Kurangi nilai α
7. Uji kondisi berhenti. Iterasi dihentikan setelah mencapai *epoch* tertentu atau jika nilai α mencapai nilai yang kecil.

Inisialisasi bobot awal biasanya menggunakan bilangan acak atau menggunakan sebagian data latih yang mewakili setiap kelas yang ada.

2) Backpropagation

Jaringan *Backpropagation* merupakan sebuah jaringan *multilayer* yang terdiri dari sebuah *input layer*, sebuah *output layer*, dan satu atau lebih *hidden layer*. Setiap *layer* dalam jaringan *Backpropagation* terhubung ke *layer* di depannya tanpa ada hubungan balik sehingga jaringan ini masuk ke dalam kategori jaringan *Feedforward*. Banyaknya neuron input pada jaringan *Backpropagation* sama dengan banyaknya fitur pada data dan banyaknya neuron output sesuai dengan banyaknya banyaknya digit yang digunakan untuk mengkodekan kelas. Di sisi lain, jumlah neuron pada *hidden layer* menentukan kompleksitas dan akurasi dari jaringan *Backpropagation*. Penentuan jumlah neuron pada *hidden layer* normalnya dilakukan melalui *trial and error* [26].

Arsitektur *Backpropagation* yang digunakan dalam penelitian ini terdiri dari satu *input layer*, satu *hidden layer*, dan satu *output layer*. Banyaknya neuron pada *input layer* adalah 31, sesuai dengan banyaknya fitur pada data sedangkan banyaknya neuron pada *output layer* adalah 13 yang sesuai dengan banyaknya kelas pada data. Banyaknya *hidden neuron* akan diujikan pada Bab III untuk menentukan jumlah *hidden neuron* dengan akurasi tertinggi.

Fungsi aktivasi yang digunakan dalam penelitian ini adalah fungsi *hyperbolic tangent sigmoid*, baik pada *hidden layer* maupun pada *output layer*. Fungsi *hyperbolic tangent sigmoid* dihitung menggunakan Persamaan 18.

$$tansig(x) = \frac{2}{1+e^{-2x}} - 1 \quad (18)$$

Terdapat beberapa algoritme pelatihan pada jaringan *Backpropagation*. Pada penelitian ini, jaringan dilatih menggunakan aturan *Backpropagation* dan bobot serta bias diubah berdasarkan metode optimasi Levenberg-Marquardt. Selama proses pelatihan, usaha yang dilakukan untuk

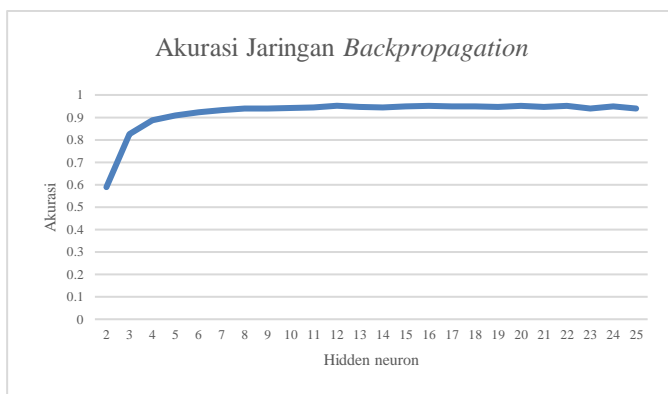
mengurangi *overfitting* adalah dengan membagi data latih menjadi 2 bagian : 70% data latih dan 30% data validasi. Jika *error* yang dihasilkan dalam proses training ketika dilakukan pengujian menggunakan data validasi meningkat selama 6 iterasi secara berturut-turut, maka proses pelatihan dihentikan.

III. HASIL PERCOBAAN

Pengujian kinerja metode *Backpropagation* dan LVQ dilakukan menggunakan 975 gambar yang terbagi menjadi 13 kelas sebagai data latih dan data uji. Pembagian data latih dan data uji dilakukan menggunakan mekanisme *10-fold* dimana semua data dibagi menjadi 10 bagian (*fold*). Proses *training* dilakukan menggunakan 9 *fold* sedangkan proses pengujian dilakukan menggunakan 1 *fold* sisanya. Khusus pada metode *Backpropagation*, data latih dibagi menjadi data latih dan data validasi dengan rasio 70% : 30%.

Metode evaluasi yang digunakan untuk menentukan kinerja metode jaringan syaraf tiruan adalah akurasi. Setelah melalui proses *training* menggunakan *training fold*, jaringan tersebut melakukan *testing* menggunakan *testing fold* yang menghasilkan kelas uji. Selanjutnya, kelas uji dibandingkan dengan kelas sebenarnya dari *testing fold* untuk menentukan nilai akurasi.

Pengujian pertama dilakukan untuk mengetahui pengaruh jumlah *hidden neuron* terhadap akurasi jaringan *Backpropagation*. Akurasi jaringan *Backpropagation* dengan jumlah *hidden neuron* mulai 2 sampai 25 dievaluasi nilai akurasinya dan hasilnya ditampilkan pada Gambar 3. Pada Gambar 3 terlihat bahwa nilai akurasi pada jaringan dengan 2 *hidden neuron* mencapai nilai 0,589. Selanjutnya, nilai akurasi meningkat drastis pada 3 *hidden neuron* menjadi 0,826 dan nilai akurasi terus meningkat seiring penambahan jumlah *hidden neuron*. Nilai akurasi mulai mencapai kondisi stagnan pada kisaran 0,932 ketika *hidden neuron* berjumlah 7. Nilai akurasi maksimum adalah 0,952 ketika jaringan *Backpropagation* memiliki *hidden neuron* sebanyak 12. Nilai akurasi ketika jumlah *hidden neuron* lebih dari 12 menunjukkan adanya penurunan yang tidak signifikan.

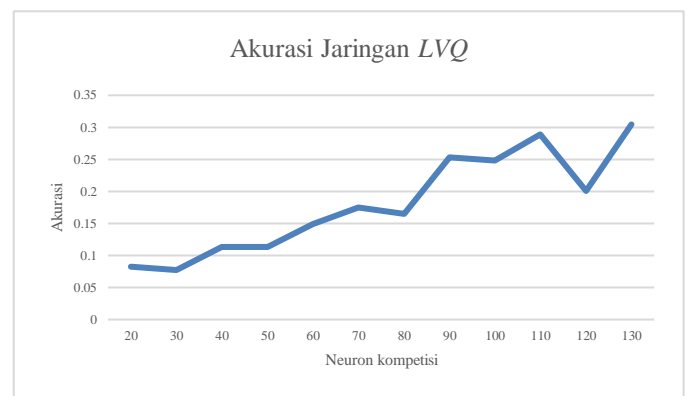


Gambar 3. Akurasi jaringan *Backpropagation*

Nilai akurasi yang rendah ketika *hidden neuron* berjumlah sedikit terjadi karena *hyperplane* yang terbentuk terlalu sederhana sehingga tidak mampu memisahkan atau mengklasifikasikan data yang memiliki kompleksitas tinggi.

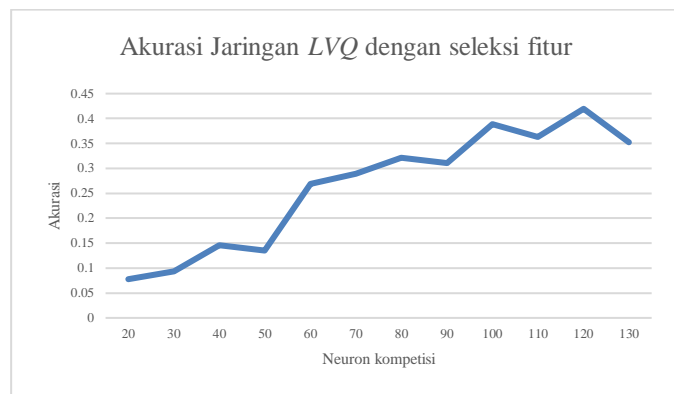
Dengan menambah jumlah *hidden neuron*, maka *hyperplane* yang terbentuk semakin kompleks sehingga mampu mengikuti kompleksitas data yang digunakan. Namun, jumlah *hidden neuron* yang terlalu banyak akan menyebabkan terbentuknya *hyperplane* yang terlalu kompleks dan menyebabkan *overfitting*. Efek dari *overfitting* adalah jaringan yang terbentuk tidak mampu mengenali pola baru. Dengan kata lain, tingkat generalisasi yang dimiliki rendah yang diindikasikan oleh nilai akurasi terhadap data uji yang rendah pula. Dilihat dari waktu komputasi, penambahan jumlah *hidden neuron* akan berefek negatif terhadap waktu komputasi. Semakin banyak *hidden neuron*, maka waktu komputasi semakin tinggi karena banyaknya bobot dan bias yang harus dilatih juga semakin banyak.

Pengujian kedua dilakukan untuk mengetahui pengaruh banyaknya neuron kompetisi terhadap akurasi jaringan LVQ. Neuron kompetisi pada jaringan LVQ diberi nilai 20 – 130 dengan kenaikan 10. Nilai akurasi ditampilkan pada Gambar 4. Nilai akurasi yang diperoleh rendah, dimana nilai akurasi maksimum adalah 0,304 ketika neuron kompetisi berjumlah 130.



Gambar 4. Akurasi jaringan LVQ

Pengujian ketiga dilakukan untuk meningkatkan kinerja jaringan LVQ dengan cara mengurangi kompleksitas jaringan melalui pengurangan jumlah *input neuron*. Pengurangan *input neuron* dilakukan dengan mengurangi fitur pada data. Seleksi fitur dapat dimanfaatkan untuk mengurangi fitur, dengan cara membuang fitur-fitur yang tidak relevan. Penelitian ini menggunakan *Pearson's coefficient* untuk menentukan tingkat kepentingan setiap fitur terhadap kelas yang terbentuk dan hanya 15 fitur dengan koefisien tertinggi saja yang digunakan. Akurasi jaringan LVQ dengan seleksi fitur ditampilkan pada Gambar 5. Berdasarkan Gambar 5, terlihat adanya kenaikan akurasi meskipun tidak terlalu signifikan. Rata-rata akurasi setelah menggunakan seleksi fitur adalah 0,264 sedangkan akurasi tanpa seleksi fitur adalah 0,181. Terdapat kenaikan akurasi sebesar 45,86%. Nilai akurasi maksimum yang berhasil diperoleh adalah 0,420 dengan neuron kompetisi berjumlah 120.



Gambar 5. Akurasi jaringan LVQ dengan seleksi fitur

Berdasarkan perbandingan antara jaringan *Backpropagation* dan LVQ, terlihat bahwa jumlah neuron pemroses (neuron kompetisi pada LVQ dan *hidden neuron* pada *Backpropagation*) memiliki pengaruh terhadap nilai akurasi. Peningkatan jumlah neuron pemroses diperlukan untuk mengatasi kompleksitas data. Namun, jaringan LVQ memerlukan neuron pemroses (neuron kompetisi) yang jauh lebih banyak dibandingkan dengan neuron pemroses (*hidden neuron*) pada jaringan *Backpropagation*. Hal ini selaras dengan kesimpulan Hawickhorst et al [27] yang menyatakan bahwa LVQ membutuhkan jumlah neuron yang lebih besar daripada RBF dan *Backpropagation*.

Perbandingan akurasi pada LVQ dan *Backpropagation* menunjukkan bahwa LVQ mengalami kesulitan dalam mengenali pola daun dalam dataset. Menurut Grbovic dan Vucetic [28], LVQ tidak memiliki kinerja yang baik pada data yang memiliki *noise* yang tinggi. Seperti diketahui, *Swedish Leaf Dataset* memiliki beberapa kelas yang bersinggungan karena bentuk daun pada kelas tersebut mirip. Selain itu, kinerja LVQ juga rendah jika beberapa kelas dalam dataset memiliki kerumitan yang berbeda-beda. Pada kelas-kelas dengan bentuk daun mirip, kompleksitasnya akan jauh lebih tinggi dibandingkan kelas-kelas dengan bentuk daun yang jauh berbeda.

IV. KESIMPULAN

Penelitian ini membandingkan kinerja jaringan syaraf tiruan *Backpropagation* dengan LVQ untuk melakukan klasifikasi tanaman berdasarkan karakteristik daun. Hasil evaluasi menunjukkan bahwa kinerja *Backpropagation* mengungguli LVQ dengan akurasi maksimal sebesar 0,952 sedangkan akurasi maksimal dari jaringan LVQ adalah 0,420. Kegagalan LVQ dalam mendapatkan hasil yang memuaskan disebabkan oleh tingginya kompleksitas data. Penelitian selanjutnya sebaiknya dilakukan untuk menurunkan kompleksitas data menggunakan teknik seperti seleksi fitur atau transformasi fitur.

REFERENSI

- [1] A. Aakif and M. F. Khan, "Automatic classification of plants based on their leaves," *Biosyst. Eng.*, vol. 139, pp. 66–75, Nov. 2015.
- [2] E. Elhariri, N. El-Bendary, and A. E. Hassanien, "Plant classification system based on leaf features," in *2014 9th International Conference on Computer Engineering Systems (ICCES)*, 2014, pp. 271–276.
- [3] P. Smith, "Wild-seed banks and taxonomy," *Taxon. Plant Conserv. Camb. Univ. Press Camb.*, pp. 294–304, 2006.
- [4] K. J. Gaston, "Taxonomy of taxonomists," *Nature*, vol. 356, pp. 281–282, 1992.
- [5] P. F. Culverhouse et al., "Automatic classification of field-collected dinoflagellates by artificial neural network," *Mar. Ecol. Prog. Ser.*, pp. 281–287, 1996.
- [6] P. J. Weeks and K. J. Gaston, "Image analysis, neural networks, and the taxonomic impediment to biodiversity studies," *Biodivers. Conserv.*, vol. 6, no. 2, pp. 263–274, 1997.
- [7] M. Nesbitt, R. P. H. McBurney, M. Broin, and H. J. Beentje, "Linking biodiversity, food and nutrition: The importance of plant identification and nomenclature," *J. Food Compos. Anal.*, vol. 23, no. 6, pp. 486–498, Sep. 2010.
- [8] K. J. Gaston and M. A. O'Neill, "Automated species identification: why not?," *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, vol. 359, no. 1444, pp. 655–667, Apr. 2004.
- [9] J.-X. Du, X.-F. Wang, and G.-J. Zhang, "Leaf shape based plant species recognition," *Appl. Math. Comput.*, vol. 185, no. 2, pp. 883–893, Feb. 2007.
- [10] T. Beghin, J. S. Cope, P. Remagnino, and S. Barman, "Shape and texture based plant leaf classification," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, 2010, pp. 345–353.
- [11] J. Wäldchen and P. Mäder, "Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review," *Arch. Comput. Methods Eng.*, pp. 1–37, Jan. 2017.
- [12] N. Bhatia and Vandana, "Survey of Nearest Neighbor Techniques," *Int. J. Comput. Sci. Inf. Secur.*, vol. 8, no. 2, pp. 302–305, May 2010.
- [13] J. Du, D. Huang, X. Wang, and X. Gu, "Shape Recognition Based on Radial Basis Probabilistic Neural Network and Application to Plant Species Identification," in *Advances in Neural Networks – ISNN 2005*, 2005, pp. 281–285.
- [14] P. Frasconi, M. Gori, and G. Soda, "Links between LVQ and Backpropagation," *Pattern Recognit. Lett.*, vol. 18, no. 4, pp. 303–310, Apr. 1997.
- [15] T.-S. Lim, W.-Y. Loh, and Y.-S. Shih, "A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms," *Mach. Learn.*, vol. 40, no. 3, pp. 203–228, Sep. 2000.
- [16] O. Söderkvist, "Computer Vision Classification of Leaves from Swedish Trees," PhD Thesis, Linköping University, Sweden, 2001.
- [17] W.-N. Tan, R. Sem, and Y.-F. Tan, "Blooming flower recognition by using eigenvalues of shape features," 2014, p. 91591R.
- [18] S. G. Wu, F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang, and Q. L. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network," in *2007 IEEE International Symposium on Signal Processing and Information Technology*, 2007, pp. 11–16.
- [19] C. Kalyoncu and Ö. Toygar, "Geometric leaf classification," *Comput. Vis. Image Underst.*, vol. 133, pp. 102–109, Apr. 2015.
- [20] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962.
- [21] Z. Huang and J. Leng, "Analysis of Hu's moment invariants on image scaling and rotation," in *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on*, 2010, vol. 7, pp. V7–476.
- [22] B. Yanikoglu, E. Aptoula, and C. Tirkaz, "Automatic plant identification from photographs," *Mach. Vis. Appl.*, vol. 25, no. 6, pp. 1369–1383, Aug. 2014.
- [23] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognit.*, vol. 45, no. 1, pp. 346–362, 2012.
- [24] J. Chaki, R. Parekh, and S. Bhattacharya, "Plant leaf recognition using texture and shape features with neural classifiers," *Pattern Recognit. Lett.*, vol. 58, pp. 61–68, Jun. 2015.
- [25] T. Kohonen, "Improved versions of learning vector quantization," in *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*, 1990, pp. 545–550.
- [26] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesus, *Neural network design*, 2nd edition. Wroclaw: Amazon Fulfillment Poland Sp. z o.o.

- [27] B. A. Hawickhorst, S. A. Zahorian, and R. Rajagopal, "A comparison of three neural network architectures for automatic speech recognition," *Intell. Eng. Syst. Artif. Neural Netw.*, vol. 5, p. 221, 1995.
- [28] M. Grbovic and S. Vucetic, "Learning Vector Quantization with adaptive prototype addition and removal," in *2009 International Joint Conference on Neural Networks*, 2009, pp. 994–1001.