



PEMBINAAN SISTEM PINTAR UNTUK
PENENTUAN KUALITI AIR
BERDASARKAN RANGKAIAN NEURAL

DR NOR ASHIDI MAT ISA

UNIVERSITI SAINS MALAYSIA
KAMPUS KEJURUTERAAN
2008



Laporan Akhir Projek Penyelidikan Jangka Pendek

**Pembinaan Sistem Pintar untuk
Penentuan Kualiti Air Berdasarkan
Rangkaian Neural**

oleh

**Dr. Nor Ashidi Mat Isa
Dzati Athiar Ramli
Dr. Kamal Zuhairi Ramli
Dr. Wan Maznah Wan Omar
Fakroul Ridzuan Hashim**



UNIVERSITI SAINS MALAYSIA

LAPORAN TEKNIKAL GERAN JANGKA PENDEK

PEMBINAAN SISTEM PINTAR UNTUK PENENTUAN KUALITI AIR BERDASARKAN RANGKAIAN NEURAL (*DEVELOPMENT OF INTELLIGENT SYSTEM FOR CLASSIFYING QUALITY OF WATER BASED ON NEURAL NETWORK*)

KETUA PENYELIDIK:

DR NOR ASHIDI MAT ISA

PENYELIDIK-PENYELIDIK BERSAMA:

PN. DZATI ATHIAR BINTI RAMLI

DR. KAMAL ZUHAIRI ZAMLI

DR. WAN MAZNAH BINTI WAN OMAR

FAKROUL RIDZUAN BIN HASHIM

UNIVERSITI SAINS MALAYSIA

NOVEMBER 2007

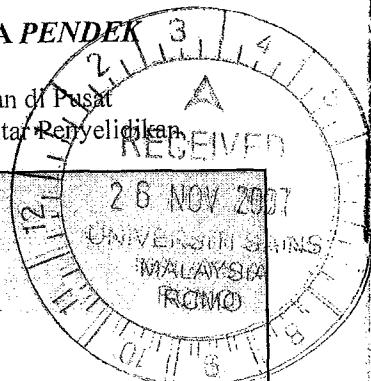
LAPORAN AKHIR PROJEK PENYELIDIKAN JANGKA PENDEK
FINAL REPORT OF SHORT TERM RESEARCH PROJECT

Sila kemukakan laporan akhir ini melalui Jawatankuasa Penyelidikan di Pusat Pengajian dan Dekan/Pengarah/Ketua Jabatan kepada Pejabat Pelantar Penyelidikan.

1. Nama Ketua Penyelidik:
Name of Research Leader

Profesor Madya/
 Dr.
 Assoc. Prof.

Encik/Puan/Cik
 Mr/Mrs/Ms



2. Pusat Tanggungjawab (PTJ):
School/Department

Pusat Pengajian Kejuruteraan Elektrik dan Eletronik

3. Nama Penyelidik Bersama:
Name of Co-Researcher

Pn Dzati Athiar Ramli, Dr Kamal Zuhairi Zamli, Dr Wan Maznah Wan Omar

4. Tajuk Projek:
Title of Project

Development of Intelligent System for Classifying Quality of Water Based on Neural Network

5. Ringkasan Penilaian/Summary of Assessment:

	Tidak Mencukupi Inadequate	Boleh Diterima Acceptable	Sangat Baik Very Good
i) Pencapaian objektif projek: <i>Achievement of project objectives</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ii) Kualiti output: <i>Quality of outputs</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
iii) Kualiti impak: <i>Quality of impacts</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
iv) Pemindahan teknologi/potensi pengkomersialan: <i>Technology transfer/commercialization potential</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
v) Kualiti dan usahasama : <i>Quality and intensity of collaboration</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
vi) Penilaian kepentingan secara keseluruhan: <i>Overall assessment of benefits</i>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

6. Abstrak Penyelidikan

(Perlu disediakan di antara 100 - 200 perkataan di dalam Bahasa Malaysia dan juga Bahasa Inggeris. Abstrak ini akan dimuatkan dalam Laporan Tahunan Bahagian Penyelidikan & Inovasi sebagai satu cara untuk menyampaikan dapatan projek tuan/puan kepada pihak Universiti & masyarakat luas).

Abstract of Research

(An abstract of between 100 and 200 words must be prepared in Bahasa Malaysia and in English). This abstract will be included in the Annual Report of the Research and Innovation Section at a later date as a means of presenting the project findings of the researcher/s to the University and the community at large)

Sila lihat Lampiran A

7. Sila sediakan laporan teknikal lengkap yang menerangkan keseluruhan projek ini.

[Sila gunakan kertas berasingan]

Applicant are required to prepare a Comprehensive Technical Report explaining the project.
(This report must be appended separately)

Sila lihat Lampiran B

Senaraikan kata kunci yang mencerminkan penyelidikan anda:

List the key words that reflects your research:

Bahasa Malaysia

Pentaksiran kualiti air

Rangkaian neural

Komposisi alga

Sistem pintar

Bahasa Inggeris

Water quality determination

Neural Network

Algae composition

Intelligent system

8. Output dan Faedah Projek
Output and Benefits of Project

(a) * **Penerbitan Jurnal**

Publication of Journals

(Sila nyatakan jenis, tajuk, pengarang/editor, tahun terbitan dan di mana telah diterbit/diseraikan)
(State type, title, author/editor, publication year and where it has been published/submitted)

Sila lihat Lampiran C

- (b) Faedah-faedah lain seperti perkembangan produk, pengkomersialan produk/pendaftaran paten atau impak kepada dasar dan masyarakat.
State other benefits such as product development, product commercialisation/patent registration or impact on source and society.

Satu sistem pengenalpastian jenis dan kualiti air sungai menggunakan rangkaian neuron telah berjaya dibina. Sistem ini mempunyai nilai komersil yang tinggi & dalam penurunan kadar pencemaran air sungai. Kaedah yang digunakan adalah mesin pengguna kerana pendekatan yang digunakan adalah secara 'biology indicator' iaitu penentuan pencemaran alga merupakan alga & dalam air.

* Sila berikan salinan/Kindly provide copies a. berdasarkan komposisi alga & dalam air.

(c) Latihan Sumber Manusia
Training in Human Resources

i) Pelajar Sarjana: Fathirou Ridzuan Hashim

Graduates Students

(Perincikan nama, ijazah dan status)

(Provide names, degrees and status)

Pentaksiran Jenis dan Kualiti Air Sungai Berdasarkan Komposisi:

Alga menggunakan Rangkaian Neuron

Status: Tunggu untuk viva

ii) Lain-lain:
Others

9. Peralatan yang Telah Dibeli:
Equipment that has been purchased

Sebuah notebook

Tandatangan Penyelidik
Signature of Researcher

15 / 11 / 2007

Tarikh
Date

Komen Jawatankuasa Penyelidikan Pusat Pengajian/Pusat
Comments by the Research Committees of Schools/Centres

Projek ini telah dijalankan dengan jayanya. Beberapa kertas kerja didalam jurnal dan proceeding telah dikenakan. 1 kertas kerja telah dihorta dan sedang diteliti didalam 'ISI' Indexed journal.

Dr. Kamal Zahari Zaini
Deputy Dean
Postgraduate And Research
School of Electrical and Electronic Engineering
Universiti Sains Malaysia, Engineering Campus

16/11/07

Tarikh
Date

TANDATANGAN PENGERUSI
JAWATANKUASA PENYELIDIKAN
PUSAT PENGAJIAN/PUSAT
*Signature of Chairman
[Research Committee of School/Centre]*

**PEMBINAAN SISTEM PINTAR UNTUK PENENTUAN KUALITI AIR
BERDASARKAN RANGKAIAN NEURAL**

LAMPIRAN A

ABSTRAK

ABSTRAK

Alga merupakan organisma mikro yang digunakan dalam pemerhatian secara biologi bagi penentuan kualiti air sungai. Pendekatan ini mampu meramal tahap kualiti sungai air lebih baik berbanding pemantauan secara fizikal dan kimia namun permasalahan timbul di kalangan ahli limnologi untuk menetapkan kelas kualiti air sungai yang diperolehi. Sebagai penyelesaian, satu sistem pintar direkabentuk dengan rangkaian neural digunakan sebagai pengkelas yang mengkelaskan jenis dan kualiti air sungai secara serentak dan automatik. Penyelidikan ini menganalisis keupayaan 78 jenis alga sungai sebagai data masukan kepada rangkaian neural untuk proses pengkelas. Keputusan analisis diskriminan menunjukkan hanya 21 jenis alga sahaja yang dominan sebagai data masukan. Bagi proses pengkelas, 2 rangkaian neural konvensional iaitu rangkaian perseptron berbilang lapisan (MLP) dan fungsi asas jejarian (RBF) digunakan. Selain itu, penyelidikan ini mencadangkan penggunaan rangkaian neural berbilang lapisan hibrid (HMLP), perseptron berbilang lapisan berhirarki (HiMLP) dan perseptron berbilang lapisan hibrid berhirarki (H^2 MLP) untuk meningkatkan keupayaan pengkelas. Keputusan yang diperolehi menunjukkan rangkaian H^2 MLP mampu mengkelaskan jenis air sungai kepada air tawar dan air payau serta kualitinya kepada bersih, pertengahan dan tercemar dengan peratus kejituhan yang tinggi iaitu 97.09% di samping ralat pengkelas yang sangat rendah iaitu 0.0096. Rangkaian H^2 MLP juga menunjukkan kadar pembelajaran yang pantas dan menumpu dengan baik serta mempunyai struktur rangkaian yang ringkas. Ini menunjukkan bahawa penyelidikan ini berjaya membuktikan bahawa komposisi alga sesuai dijadikan masukan kepada rangkaian neural untuk mengkelaskan jenis dan kualiti air sungai. Selain itu, rangkaian neural juga telah dibuktikan berkeupayaan tinggi sebagai alat pengkelasan pintar bagi tujuan tersebut.

**DEVELOPMENT OF INTELLIGENT SYSTEM FOR CLASSIFYING QUALITY
OF WATER BASED ON NEURAL NETWORK**

ABSTRACT

Algae are microorganisms which are being used in biological monitoring to determine the quality of river's water. This approach is used to predict the quality level of river's water which is better than physical and chemical monitoring, where problems arise among the limnologists to determine the quality of the river's water. Thus, an intelligent system is developed to solve this problem by using neural network as a classifier to classify river's type and water quality simultaneously and automatically. Breakdown analysis is determined with 78 types of river's algae as an input data to neural network for classification process. The result obtained shows that 21 species of algae are dominant to be used as input data. Two conventional neural networks namely the multilayered perceptron (MLP) and the radial basis function (RBF) were employed for the classification purpose. Nevertheless, the hybrid multilayered perceptron (HMLP), the hierarchical multilayered perceptron (HiMLP) and the hierarchical hybrid multilayered perceptron (H^2 MLP) network are suggested in this research to improve the classification performance. The result shows, the H^2 MLP network has high capability to classify river's water into fresh or brackish water and further classifies its quality into clean, moderate or polluted, with high accuracy at 97.09% and as low as 0.0096 classification error. Furthermore, the H^2 MLP network shows the fastest learning rate with a good convergence as compared to other networks. These results prove that algae composition is suitable as an input data of neural network for classifying the types and qualities of river's water. Besides, neural network has also proven as a good intelligent classifier with high classification performance.

LAMPIRAN B

LAPORAN TEKNIKAL



LAPORAN TEKNIKAL GERAN JANGKA PENDEK
PEMBINAAN SISTEM PINTAR UNTUK PENENTUAN KUALITI
AIR BERDASARKAN RANGKAIAN NEURAL
*(DEVELOPMENT OF INTELLIGENT SYSTEM FOR CLASSIFYING
QUALITY OF WATER BASED ON NEURAL NETWORK)*

KETUA PENYELIDIK:

DR NOR ASHIDI MAT ISA

PENYELIDIK-PENYELIDIK BERSAMA:

PN. DZATI ATHIAR BINTI RAMLI

DR. KAMAL ZUHAIRI ZAMLI

DR. WAN MAZNAH BINTI WAN OMAR

FAKROUL RIDZUAN BIN HASHIM

UNIVERSITI SAINS MALAYSIA

NOVEMBER 2007

SUSUNAN KANDUNGAN

Muka surat

SUSUNAN KANDUNGAN

i

ABSTRAK

iii

ABSTRACT

iv

BAB SATU: PENGENALAN

1

- 1.1 Pengenalan Kepada Klasifikasi Kualiti Air Sungai 1
1.2 Penggunaan Rangkaian Neural dalam Menentukan Kualiti Air 1
1.3 Objektif dan Skop Penyelidikan 3

BAB DUA: KAJIAN ILMIAH

5

- 2.1 Pengenalan 5
2.2 Penentuan Kualiti Air Sungai 5
 2.2.1 Klasifikasi Air Sungai 6
 2.2.2 Kaedah Pengklasifikasian Kualiti Air 6
2.3 Komposisi Alga 9
 2.3.1 Kesan Komposisi Alga Kepada Persekutaran 9
2.4 Rangkaian Neural 10
 2.4.1 Rangkaian Neuron Biologi 10
 2.4.2 Rangkaian Neural Buatan 11
 2.4.2.1 Pemodelan Neuron 12
 2.4.2.2 Seni Bina 13
 2.4.2.3 Proses Pembelajaran 14
 2.4.3 Aplikasi Rangkaian Neural Buatan 16
2.5 Ringkasan 17

**BAB TIGA: PENENTUAN JENIS DAN KUALITI AIR SUNGAI
MENGGUNAKAN RANGKAIAN NEURAL**

18

- 3.1 Pengenalan 18
3.2 Penentuan Jenis dan Kualiti Air Sungai 19
 3.2.1 Aplikasi Analisis Diskriminan 19
 3.2.2 Tahap Keertian Pembolehubah Diskriminan 22

	3.2.3 Kaedah Titik Pemisah	23
3.3	Rangkaian Fungsi Asas Jejarian	24
	3.3.1 Seni Bina dan Ciri-ciri Rangkaian	24
	3.3.2 Penempatan Pusat Rangkaian	25
3.4	Rangkaian Perseptron Berbilang Lapisan	26
	3.4.1 Seni Bina dan Ciri-ciri Rangkaian	27
	3.4.2 Pemberat Sambungan Rangkaian	28
	3.4.2.1 Algoritma Perambatan Balik	29
	3.4.2.2 Algoritma Lavenberg Marquardt	29
	3.4.2.3 Algoritma Bayesian Regularization	30
3.5	Rangkaian Perseptron Berbilang Lapisan Hibrid	31
	3.5.1 Seni Bina dan Ciri-ciri Rangkaian	31
	3.5.2 Pemberat Sambungan Rangkaian	33
3.6	Rangkaian Neural Berhirarki	34
3.7	Sampel Data	35
3.8	Metodologi	36
3.9	Kesimpulan	37
	BAB EMPAT: KEPUTUSAN DAN PERBINCANGAN	38
4.1	Pengenalan	38
4.2	Keputusan Analisis Diskriminan	38
	4.2.1 Ujian Univariat	38
	4.2.2 Ujian Multivariat	41
	4.2.3 Fungsi Diskriminan	44
	4.2.4 Penentuan Titik Pemisah	46
	4.2.5 Prestasi Pengkelasan Data	47
4.3	Keputusan Analisis Rangkaian Neural	48
4.6	Kesimpulan	55
	BAB LIMA: KESIMPULAN	57
5.1	Kesimpulan	57
5.2	Cadangan-cadangan	60
	RUJUKAN	62

**PEMBINAAN SISTEM PINTAR UNTUK PENENTUAN KUALITI AIR
BERDASARKAN RANGKAIAN NEURAL**

ABSTRAK

Alga merupakan organisma mikro yang digunakan dalam pemerhatian secara biologi bagi penentuan kualiti air sungai. Pendekatan ini mampu meramal tahap kualiti sungai air lebih baik berbanding pemantauan secara fizikal dan kimia namun permasalahan timbul di kalangan ahli limnologi untuk menetapkan kelas kualiti air sungai yang diperolehi. Sebagai penyelesaian, satu sistem pintar direkabentuk dengan rangkaian neural digunakan sebagai pengkelas yang mengelaskan jenis dan kualiti air sungai secara serentak dan automatik. Penyelidikan ini menganalisis keupayaan 78 jenis alga sungai sebagai data masukan kepada rangkaian neural untuk proses pengkelas. Keputusan analisis diskriminan menunjukkan hanya 21 jenis alga sahaja yang dominan sebagai data masukan. Bagi proses pengkelas, 2 rangkaian neural konvensional iaitu rangkaian perseptron berbilang lapisan (MLP) dan fungsi asas jejarian (RBF) digunakan. Selain itu, penyelidikan ini mencadangkan penggunaan rangkaian neural berbilang lapisan hibrid (HMLP), perseptron berbilang lapisan berhirarki (HiMLP) dan perseptron berbilang lapisan hibrid berhirarki (H^2 MLP) untuk meningkatkan keupayaan pengkelas. Keputusan yang diperolehi menunjukkan rangkaian H^2 MLP mampu mengelaskan jenis air sungai kepada air tawar dan air payau serta kualitinya kepada bersih, pertengahan dan tercemar dengan peratus kejituhan yang tinggi iaitu 97.09% di samping ralat pengkelas yang sangat rendah iaitu 0.0096. Rangkaian H^2 MLP juga menunjukkan kadar pembelajaran yang pantas dan menumpu dengan baik serta mempunyai struktur rangkaian yang ringkas. Ini menunjukkan bahawa penyelidikan ini berjaya membuktikan bahawa komposisi alga sesuai dijadikan masukan kepada rangkaian neural untuk mengelaskan jenis dan kualiti air sungai. Selain itu, rangkaian neural juga telah dibuktikan berkeupayaan tinggi sebagai alat pengkelasan pintar bagi tujuan tersebut.

**DEVELOPMENT OF INTELLIGENT SYSTEM FOR CLASSIFYING QUALITY
OF WATER BASED ON NEURAL NETWORK**

ABSTRACT

Algae are microorganisms which are being used in biological monitoring to determine the quality of river's water. This approach is used to predict the quality level of river's water which is better than physical and chemical monitoring, where problems arise among the limnologists to determine the quality of the river's water. Thus, an intelligent system is developed to solve this problem by using neural network as a classifier to classify river's type and water quality simultaneously and automatically. Breakdown analysis is determined with 78 types of river's algae as an input data to neural network for classification process. The result obtained shows that 21 species of algae are dominant to be used as input data. Two conventional neural networks namely the multilayered perceptron (MLP) and the radial basis function (RBF) were employed for the classification purpose. Nevertheless, the hybrid multilayered perceptron (HMLP), the hierarchical multilayered perceptron (HiMLP) and the hierarchical hybrid multilayered perceptron (H^2 MLP) network are suggested in this research to improve the classification performance. The result shows, the H^2 MLP network has high capability to classify river's water into fresh or brackish water and further classifies its quality into clean, moderate or polluted, with high accuracy at 97.09% and as low as 0.0096 classification error. Furthermore, the H^2 MLP network shows the fastest learning rate with a good convergence as compared to other networks. These results prove that algae composition is suitable as an input data of neural network for classifying the types and qualities of river's water. Besides, neural network has also proven as a good intelligent classifier with high classification performance.

BAB 1

PENGENALAN

1.1 Pengenalan Kepada Klasifikasi Kualiti Air Sungai

Pengelasan kualiti air yang dilakukan adalah bergantung kepada ciri-ciri fizikal, kimia dan biologi air itu. Perubahan suhu air serta suhu persekitaran di samping julat pH air juga memainkan peranan penting dalam mengelaskan kualiti air ini. Parameter-parameter seperti oksigen terlarut (*Dissolved oxygen, DO*), permintaan oksigen biokimia (*Bio-chemical oxygen demand, BOD*), permintaan oksigen kimia (*Chemical oxygen demand, COD*) dan pepejal terampai (*Suspended solids, SS*) telah digunakan sebagai parameter ukuran kepada tahap kualiti air (DOE, 1998). Maznah dan Mansor, (2002) telah membuktikan bahawa spesis-spesis alga yang terdapat di dalam air sungai mampu mengawal masalah pencemaran sungai daripada terus berlaku (Maznah & Mansor, 2002).

1.2 Penggunaan Rangkaian Neural dalam Menentukan Kualiti Air

Kecerdikan buatan menjadi satu kaedah permodelan yang sangat popular bagi menyelesaikan masalah-masalah yang kompleks. Salah satu cabang daripada kecerdikan buatan yang sering digunakan ialah rangkaian neural. Aplikasi-aplikasi permodelan rangkaian neural untuk meramal kehadiran alga biomas dalam ekosistem air tawar telah dilakukan di Tasik Tuusulunjarvi di Finland, Tasik Kasumigaura dan Tasik Biwa di Jepun dan Sungai Darling di Austria (Racknegal *et al.*, 1997). Penambahbaikan terhadap model rangkaian neural juga telah dilakukan dengan membangunkan satu rangkaian neural separa pengulangan (*recurrent neural network*) berdasarkan siri-siri masa (*time series*) (Jeong *et al.*, 2001). Model yang dibangunkan ini telah mengkaji kehidupan phytoplankton di Sungai Nakdong (Korea). Keputusan daripada kajian yang dilakukan di Sungai Nakdong telah membuktikan rangkaian neural sesuai untuk digunakan dalam meramal kehadiran spesis-spesis alga.

1.3 Objektif dan Skop Penyelidikan

Objektif utama penyelidikan ini ialah membangunkan satu sistem pintar bagi menentukan tahap kualiti air sungai. Sistem pintar ini akan melakukan ramalan berdasarkan sampel komposisi spesis-spesis alga dominan. Rangkaian neural akan melakukan proses penentuan tahap kualiti air sungai berdasarkan komposisi alga yang hadir. Kebanyakan penyelidik terdahulu hanya mengkelaskan kualiti air kepada dua klasifikasi sahaja iaitu bersih (*clean*) dan kotor (*dirty*) (Joergensen & Bendoricchio, 2001, Dzeroski & Todorovski, 2003 & Wilson & Recknegal, 2001).

Kajian ini pula mengkelaskan kualiti air kepada tiga klasifikasi iaitu bersih (*clean*), pertengahan (*moderate*) dan tercemar (*polluted*). Kajian yang dijalankan ini bukan hanya mengkelaskan kualiti air sahaja, malah kajian yang dijalankan juga melangkah satu langkah kehadapan dengan mengkelaskan air sungai kepada beberapa jenis air terlebih dahulu sebelum proses pengelasan kualiti air dijalankan. Hasil kajian ilmiah didapati terdapat tiga jenis air yang terdapat di mukabumi ini iaitu air tawar (*freshwater*), air payau (*brackish-water*) dan air masin (*salt-water*) (Choudhury et. al., 2001, Kromhout, 2005 & Pareek et. al., 2006). Kajian ini dijalankan terhadap air sungai sahaja. Oleh itu pengelasan terhadap jenis air sungai hanya dikelaskan kepada air tawar dan air payau sahaja kerana air masin hanya terdapat di lautan. Bagi jenis air tawar, kualiti air dikelaskan kepada tiga jenis iaitu bersih, pertengahan atau tercemar. Berbeza pula bagi jenis air payau, kualiti air hanya dapat dikelaskan kepada dua, sama ada pertengahan ataupun tercemar.

Penyelidikan ini mencadangkan penggunaan rangkaian perseptron berbilang lapisan (*Multilayered Perceptron, MLP*) dan fungsi asas jejarian (*Radial Basis Function, RBF*) dalam menentukan tahap kualiti air. Rangkaian MLP ini akan dilatih oleh algoritma perambatan balik (*backpropagation, BP*), algoritma Lavenberg Marquardt (*LM*) dan algoritma Bayesian Regularization (*BR*). Rangkaian MLP terubahsuai yang mempunyai beberapa sambungan lelurus tambahan yang dikenali sebagai rangkaian perseptron berbilang lapisan hibrid (*Hybrid Multilayered Perceptron, HMLP*)

dicadangkan dalam kajian ini mampu memberikan kejituhan yang lebih tinggi berbanding rangkaian MLP.

Di samping itu juga, rangkaian HMLP mampu mengurangkan saiz rangkaian yang dibentuk oleh rangkaian MLP (Mashor, 2000a). Bagi mendapatkan keputusan kejituhan yang lebih baik, pengelasan secara berhirarki dicadangkan dalam penyelidikan ini. Penyelidikan ini mencadangkan pengelasan berhirarki dilakukan kepada rangkaian MLP dan HMLP dengan masing-masing membentuk rangkaian perseptron berbilangan lapisan berhirarki (*Hierarchical Multilayered Perceptron, HiMLP*) dan perseptron berbilang lapisan hibrid berhirarki (*Hierarchical Hybrid Multilayered Perceptron, H²MLP*). Seperti rangkaian MLP, rangkaian HiMLP juga akan dilatih menggunakan algoritma BP, LM dan BR.

Secara umumnya, penyelidikan ini mempunyai beberapa objektif iaitu:

- (i) berbanding penyelidikan-penyalidikan terdahulu yang yang mengkelaskan kualiti air sungai, penyelidikan ini mencadangkan penggunaan rangkaian neural untuk mengkelaskan jenis air sungai dan kualiti air sungai secara serentak. Jenis air akan dikelaskan kepada air tawar dan air payau, manakala kualiti air akan dikelaskan kepada bersih, pertengahan dan tercemar. Penyelidikan ini akan menguji keupayaan beberapa jenis rangkaian neural konvensional seperti rangkaian MLP dan RBF untuk tujuan tersebut. Penyelidikan ini juga mencadangkan penggunaan rangkaian neural hibrid dan rangkaian neural berhirarki. Prestasi semua rangkaian neural yang dicadangkan akan dibandingkan di antara satu sama lain.
- (ii) menganalisis keupayaan komposisi alga sebagai parameter masukan kepada rangkaian neural dalam (i). 78 jenis alga akan dianalisis iaitu *Acnanthes Exigua*, *Acnanthes Exigua Var Heterovalva*, *Acnanthes Minutissima*, *Acnanthes Oblongela*, *Acnanthes Wolterickii*, *Amphora Libica*, *Anomoeoneis Brachysira*, *Coccconeis Pediculus*, *Coccconeis Placentula*, *Coccconeis Species*, *Coccconeis Thumensis*, *Coscinodiscus Antiquus*, *Coscinodiscus Argus*, *Coscinodiscus Decipiens*, *Coscinodiscus Excentricus*, *Coscinodiscus Granii*, *Cycbellia Species*, *Cyclotella*

Comta, *Cyclotella Striata*, *Cyclotella Striata Var Baltica*, *Cymbella Inaequalis*, *Diatoma Elongatum*, *Diatoma Species*, *Diploneis Bombus*, *Diploneis Decipiens*, *Diploneis Elliptica*, *Diploneis Interrupta*, *Eunotia Faba*, *Eunotia Grunowi*, *Eunotia Monodon Var Alpina*, *Eunotia Pectinalis*, *Eunotia Sadetica Var Inasa*, *Eutonia Pectinalis*, *Eutonia Pectinalis Var Minor*, *Flagilaria Capucina*, *Flagilaria Species 1*, *Flagilaria Species 2*, *Fragilaria Crotonensis*, *Frustulia Rhomboids*, *Frustulia Rhomboides Var Saxonica*, *Frustulia Saxonica*, *Gomphonema Acuminatum*, *Gomphonema Gracile*, *Gomphonema Longiceps*, *Gomphonema Parvalum*, *Gomphonema Subventricosum*, *Hantzschia Amphioxys*, *Navicula Cryptocephala*, *Navicula Cuspidata*, *Navicula Halophila*, *Navicula Hustedtii*, *Navicula Obtusa*, *Navicula Obtusa Var Scalpelliformis*, *Navicula Radiosa*, *Navicula Sigma*, *Navicula Species*, *Neidium Affine*, *Nitzschia Amphibia*, *Nitzschia Angustata*, *Nitzschia Fanticola*, *Nitzschia Littoralis*, *Nitzschia Palea*, *Pinnularia Biceps*, *Pinnularia Biceps F. Petersenii*, *Pinnularia Borealis*, *Pinnularia Braunii Var Amphicephala*, *Pinnularia Maior Var Transverve*, *Pinnularia Mesolepta*, *Pinnularia Microstauron*, *Pinnularia Ruttneri*, *Pinnularia Species 1*, *Pinnularia Splendida*, *Pinnularia Viridis*, *Psammothidium Bioretii*, *Stauroneis Obtusa*, *Surirella Linearis*, *Surirella Parma Soveriegn*, *Surirella Species 1 & Surirella Tenuissima*. Penggunaan jenis-jenis alga ini dicadangkan berdasarkan keupayaan mereka sebagai penunjuk biologi dalam penentuan kualiti air (Maznah & Mansor, 2002).

- (iii) melakukan analisis kepada 78 jenis alga dengan menggunakan analisis diskriminan bagi mengenalpasti komposisi alga yang mampu memberikan impak yang tinggi kepada pengelasan jenis dan kualiti air sungai. Hanya komposisi alga yang memberikan impak yang tinggi sahaja akan digunakan ke dalam rangkaian neural sebagai parameter masukan bagi mengelaskan jenis dan kualiti air sungai.
- (iv) membina satu sistem pintar yang menentukan jenis dan kualiti air sungai secara automatik. Sistem yang dibina mempunyai ciri-ciri dari (i) & (ii) dengan menggunakan alga-alga yang diperolehi dari (iii).

BAB 2

KAJIAN ILMIAH

2.1 Pengenalan

Pengurusan kualiti air merupakan salah satu cabang kepada kajian mengenai ekosistem akuatik. Ekosistem akuatik adalah sangat kompleks disebabkan oleh kepelbagaiannya jenis hidup di dalamnya serta hidupan yang terkandung di dalamnya membiak dengan pesat. Satu spesis yang kecil mampu membiak dengan pantas dalam kuantiti yang tinggi cenderung untuk menjadi penyebab kepada masalah pencemaran (Boyra et al., 2004). Spesis alga biru-kehijauan merupakan satu contoh terbaik yang menyebabkan persekitaran tercemar teruk. Spesis ini dapat menurunkan tahap kualiti air sungai, tasik dan di kawasan tadahan air (Boyra et al., 2004). Pendekatan yang digunakan oleh Wilson & Recknegel (2001), dalam meramal pertumbuhan spesis-spesies alga menggunakan rangkaian neural memungkinkan penentuan kualiti air menggunakan rangkaian neural dapat dilakukan dengan menggunakan komposisi alga sebagai parameter masukan.

2.2 Penentuan Kualiti Air Sungai

Di Malaysia, pembangunan ekonomi yang pesat menyebabkan hidrologi dan ekologi ekosistem terjejas teruk disebabkan oleh masalah pencemaran. Jabatan Alam sekitar (*Department of Environment, DOE*) dan beberapa firma perunding persendirian beserta pakar-pakar dari beberapa universiti tempatan telah memulakan program pemantauan terhadap sungai-sungai yang tercemar dengan mengumpul data mengenai pencemaran yang telah berlaku. Ciri-ciri fizikal air, kandungan bahan kimia di dalam air serta kandungan biologi air akan dikenalpasti (Cuffney, 2000). Ciri-ciri ini akan digunakan bagi meramal sifat-sifat dan tahap kualiti air bagi sistem-sistem sungai di Malaysia.

2.2.1 Klasifikasi Air Sungai

Kualiti air biasanya akan diklasifikasikan kepada empat kategori sama ada sangat bersih, bersih, kotor atau sangat kotor (Wilson & Recknegal, 2001). Kualiti air merupakan ungkapan yang ditujukan kepada tahap kesesuaian air untuk digunakan dalam pelbagai aplikasi. Setiap penggunaan air mestilah memenuhi spesifikasi dari segi sifat fizikal, kandungan biologi yang terdapat di dalam air tersebut dan kandungan bahan kimia yang terdapat di dalamnya. Oleh itu, keadaan fizikal air, kandungan biologi dalam air dan kandungan bahan kimia dalam air dapat menentukan tahap kualiti air tersebut. Peningkatan kajian mengenai ekosistem akuatik menjadi opsyen kepada pengurusan kualiti air. Maklumat mengenai fizikal air, kandungan bahan kimia dan kandungan biologi di dalam air di samping beberapa maklumat tambahan dapat dikumpul daripada perhatian terhadap sampel air.

2.2.2 Kaedah Pengklasifikasian Kualiti Air

Terdapat tiga tahap pemantauan yang sering dijalankan bagi pengklasifikasian kualiti air (Windelspecht, 2002). Pemantauan tahap pertama adalah pemantauan kepada fizikal air diikuti pemantauan tahap kedua iaitu pemantauan kepada kandungan kimia di dalam air manakala pemantauan tahap ketiga adalah pemantauan kepada komposisi biologi di dalam air.

Pemantauan tahap pertama adalah pemantauan kepada ciri-ciri fizikal air. Pemantauan kepada ciri-ciri fizikal air merupakan pemantauan yang paling mudah untuk dilakukan. Kaedah ini adalah paling efektif kerana tidak memerlukan sebarang latihan sebelum melakukan pemantauan, tidak memerlukan peralatan canggih yang mahal serta tidak memerlukan masa tertentu untuk pemantauan terhadap kualiti air dijalankan. Bagi menjalankan penentuan kualiti air, peralatan seperti sepasang kasut getah dan sebuah buku nota untuk mencatat sebarang maklumat mengenai keadaan air diperlukan. Pemantauan dapat dilakukan sekali seminggu atau sekurang-kurangnya

sekali bagi setiap bulan. Hanya beberapa jam sahaja diperlukan setiap kali pemantauan dijalankan.

Chang (2005) & Quevauviller (2006) telah mengklasifikasikan kualiti air menggunakan pemantauan tahap kedua iaitu pemantauan kepada kandungan bahan kimia di dalam air. Beberapa parameter telah digunakan bagi mewakilkan pemantauan terhadap keseimbangan jisim. Parameter kualiti air yang digunakan adalah pemintaan oksigen biokimia (*Bio-Chemical Oxygen Demand, BOD*), permintaan oksigen kimia (*Chemical Oxygen Demand, COD*), oksigen terlarut (*Dissolved Oxygen, DO*) dan pepejal terampai (*Suspended Solids, SS*).

Parameter BOD akan mengukur kuantiti oksigen yang diperlukan oleh mikroorganisma dan bakteria bagi menstabilkan kandungan bahan organik biosorot di dalam air (Sawyer et. al., 2003). Bahan organik biosorot ini mengandungi unsur karbon dan nitrogen. Proses pengoksidaan akan berlaku apabila bahan organik diurai oleh mikroorganisma. Hasil penguraian akan menghasilkan tenaga yang akan digunakan oleh mikroorganisma itu sendiri. Faktor suhu, masa dan kuantiti cahaya akan mempengaruhi kandungan BOD di dalam air. Parameter COD akan menentukan kandungan bahan kimia yang terdapat di dalam air (Kambe et. al., 2007). Parameter ini dapat memendekkan masa ujian kerana bahan organik akan dioksidakan secara kimia dan bukan terurai secara semulajadi seperti BOD yang memerlukan masa sehingga lima hari untuk diuraikan.

Parameter DO pula akan menentukan tahap pencemaran air (Kambe et. al., 2007). Kandungan oksigen terlarut adalah dipengaruhi oleh suhu. Petambahan suhu akan menyebabkan penurunan kadar kandungan oksigen terlarut. Oksigen terlarut ini amat penting kepada semua ekosistem akuatik untuk meneruskan kelangsungan hidup. Kandungan oksigen terlarut yang tidak mencukupi akan mengakibatkan ekosistem akuatik akan mati akibat kesukaran bernafas. Parameter SS biasanya terdiri daripada zarah-zarah organik dan zarah-zarah tidak organik yang mempunyai saiz yang lebih besar daripada 0.001mm dan tidak larut dalam air (Kambe et. al., 2007).

Zarah-zarah organik terdiri daripada alga, bakteria dan protozoa manakala zarah-zarah tidak organik terdiri daripada tanah liat dan kelodak. Kehadiran pepejal terampai akan menyebabkan air menjadi keruh, berkeladak, berbau serta beracun. Mikroorganisma aerobik yang mengurai bahan organik di bahagian atas air akan menghasilkan karbon dioksida, air dan sel-sel baru. Mikroorganisma anerobik pula akan mengurai bahan organik dan sel-sel yang terenap kepada asid organik dan seterusnya kepada karbon dioksida, air, hidrogen sulfida dan metana.

Pemantauan secara biologi terhadap tahap kualiti air telah berkembang sejak dua dekad yang lalu. Pemantauan biologi dilakukan terhadap ekosistem akuatik di sepanjang sungai (Maznah & Mansor, 2002). Kitar hidup spesis ekosistem akuatik yang terdiri daripada tumbuhan dan haiwan adalah singkat di samping pembiakan dan kepupusan yang silih berganti menarik perhatian ahli-ahli biologi. Pelbagai usaha telah dijalankan untuk menjadikan haiwan sebagai parameter pemantauan (Bradley & Ormerod, 2001). Satu sistem telah direkabentuk berdasarkan kehadiran jenis haiwan serta bilangannya yang berada di satu-satu kawasan pemantauan. Namun kehidupan spesis haiwan yang bermusim adalah tidak sesuai dijadikan sebagai penunjuk (Maznah & Mansor, 2002).

Rosenberg (1998) telah menjalankan satu penyelidikan dan hasilnya beliau mendapati pemantauan secara biologi adalah lebih baik berbanding pemantauan secara kimia dan merumuskan tumbuhan sesuai dijadikan penunjuk kepada kualiti air. Milner et. al. (2006) telah menjalankan ujian penentuan kualiti air dengan menggabungkan beberapa parameter kimia kepada ekosistem akuatik. Hasil kajian menunjukkan pemantauan secara kimia kurang berkesan berbanding pemantauan secara biologi. Pensampelan biologi telah digunakan dalam program pemantauan terhadap sungai-sungai yang tercemar. Hasil kajian mendapati spesis-spesis alga mempunyai perkaitan yang sangat hampir kepada pencemaran air (Maznah et. al., 2000).

2.3 Komposisi Alga

Spesis alga adalah populasi tumbuhan akuatik yang merupakan kumpulan organisma akuatik yang terbesar dan kebanyakannya daripadanya adalah terdiri daripada organisma berbilang sel (*multisel*) dan selebihnya adalah organisma sel tunggal (*unisel*) (Boyla et. al., 2004). Spesis alga juga merupakan tumbuhan hijau yang mempunyai klorofil di mana mampu menghasilkan makanannya sendiri melalui proses fotosentesis. Spesis alga juga akan mengeluarkan bahan organik seperti karbon terlarut dan asid hidropobik kepada persekitaran. Sebatian tak organik pula seperti karbon dioksida, ammonia, nitrat dan fosforus juga menjadi punca makanan kepada alga untuk membiakkan spesis-spesis alga yang baru seterusnya menghasilkan oksigen.

2.3.1 Kesan Komposisi Alga Serta Penggunaan Alga Kepada Persekutaran

Bahan organik yang dilepaskan oleh spesis alga kepada persekitaran sangat berbahaya disamping mengandungi toksin. Spesis alga *Cylindrospermopsis Raciborskii* contohnya boleh menyebabkan pengaratan air terjadi (Backer, 2002). Sistem rekreasi akuatik akan menjadi tidak seimbang disebabkan oleh pengaratan air. Populasi alga ini boleh membiak dalam air yang cetek dan menyebabkan kematian kepada hidupan akuatik akibat kekurangan oksigen terlarut yang terkandung di dalam air. Selain itu, toksin yang terhasil daripada spesis *Cylindrospermopsis Raciborskii* ini boleh mengancam nyawa manusia (Wilson & Recknagel, 2001). Toksin ini juga mungkin terkandung di dalam air minuman atau wujud di dalam sistem rantai makanan manusia.

Lu & Hodgkiss (2004) membuktikan toksin-toksin daripada alga mampu melepas kaedah rawatan air yang dilakukan secara normal dan tegar kepada suhu didih air. Ia akan mengakibatkan gastrik dan keracunan kepada hati. Bornet et. al (2005) pula melaporkan toksin daripada spesis alga seperti *Alexandrium*, *Dinophysis* dan *Psuedo-nitzschia Multiseries* boleh menyebabkan hidupan laut pupus akibat keracunan. Perkara ini akan menyebabkan pengurangan kepada permintaan eksport

makanan laut. Spesis alga seperti *Heterosigma* Akashiwo dan *Cryptosporidium Parvum* mampu membunuh spesis-spesis ikan dengan mengakibatkan insang ikan tersumbat teruk (Yoon et. al., 2005 & Coyne et. al., 2005). Selain itu, spesis alga seperti *Cyanophytes* menyebabkan bau, rasa dan warna kepada bekalan air minuman berubah sama sekali seterusnya meningkatkan rungutan di kalangan pelanggan (Komarek, 2005). Masalah-masalah ini mampu diatasi atau sekurang-kurangnya kadar kes dapat diturunkan sekiranya pembiakan populasi spesis alga ini dapat dikekang pada peringkat awal lagi.

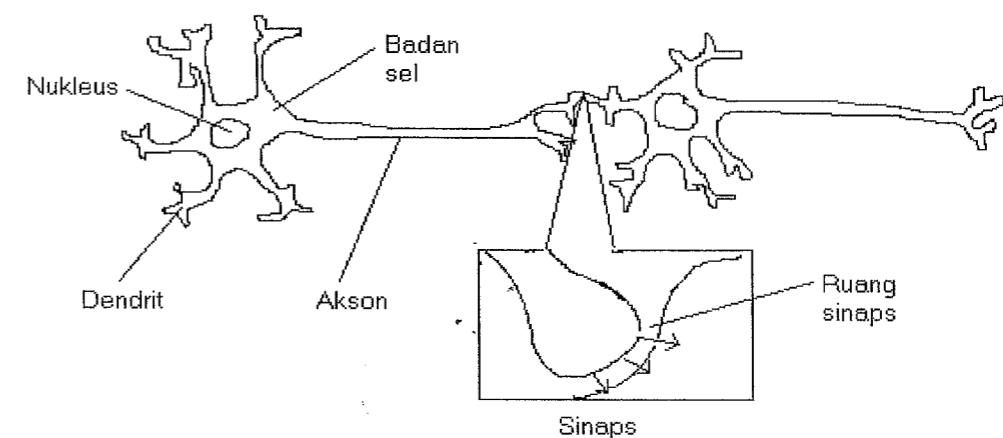
2.4 Rangkaian Neural

Penggunaan kecerdikan buatan kini adalah selari dengan arus pembangunan negara. Bidang ini memainkan peranan yang penting dalam membaiki taraf kehidupan rakyat. Rangkaian neural buatan adalah salah satu cabang dalam kecerdikan buatan yang digunakan bagi memodelkan fungsi otak. Berdasarkan prinsip otak beroperasi, rangkaian neural buatan direkabentuk menyerupai pengoperasian otak seperti pembinaan struktur senibina, teknik pembelajaran dan teknik pengoperasian.

2.4.1 Rangkaian Neuron Biologi

Neuron merupakan struktur asas bagi sistem saraf dalam tubuh badan manusia. Senibina neuron adalah terdiri daripada nukleus, sel badan, dendrit, akson dan sinaps seperti yang ditunjukkan di dalam Rajah 2.1 (Nicholls et. al., 2001). Setiap satu neuron terdiri daripada satu badan sel yang mempunyai banyak cabang-cabang dan cabang-cabang ini dinamakan dendrit yang berfungsi untuk menerima maklumat daripada neuron-neuron lain melalui akson. Akson pula akan menghantar maklumat daripada badan sel pada satu neuron kepada badan sel neuron yang lain. Setiap satu neuron disambungkan dengan satu neuron yang lain melalui satu ruang kecil yang terdapat di antara dendrit dan akson yang dikenali sebagai ruang sinaps. Proses penghantaran dan penerimaan segala maklumat berlaku pada ruang sinaps ini. Setiap

maklumat dikodkan berdasarkan perubahan denyut elektrik. Apabila denyut elektrik ini melepas had atau nilai ambang tertentu pada sinaps tertentu, sinaps tersebut akan terangsang. Penghantaran dan penerimaan informasi di antara dua neuron akan berlaku. Sebaliknya bagi denyut elektrik yang kurang daripada nilai ambang yang ditetapkan, sinaps akan terencat. Perencatan sinaps akan menghalang perhubungan di antara dua neuron.



Rajah 2.1: Gambarajah neuron biologi.

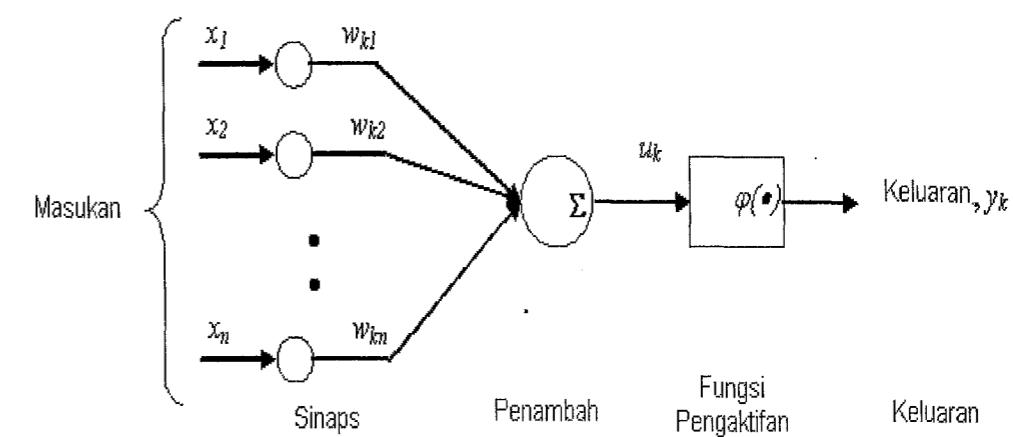
2.4.2 Rangkaian Neural Buatan

Rangkaian neural merupakan suatu sistem pemprosesan maklumat yang mengandungi satu set unit pemprosesan yang disambungkan melalui saluran lain secara selari mengikut struktur yang tertentu. Rangkaian ini direkabentuk bagi memodelkan sebahagian daripada fungsi otak manusia seperti pemprosesan isyarat, operasi matematik, perhubungan, pengenalpastian sistem dan banyak lagi. Selain itu, rangkaian neural juga mampu menerangkan hubungan linear dan hubungan tidak linear sesuatu sistem dan mampu mempelajari hubungan-hubungan ini secara terus daripada model data. Haykin (2001) menakrifkan rangkaian neural sebagai pemproses teragih selari besar yang mempunyai satu kecenderungan semulajadi untuk penyimpanan pengetahuan berpengalaman dan menggunakanannya. Ia menyerupai otak

manusia berdasarkan dua perkara berikut, pertama pengetahuan diperolehi oleh rangkaian melalui proses latihan dan pembelajaran dan kedua kekuatan sambungan di antara neuron yang dikenali sebagai pemberat sinaps digunakan untuk menyimpan pengetahuan.

2.4.2.1 Pemodelan Neuron

Pemodelan neuron bagi rangkaian neural buatan adalah diadaptasikan daripada pemodelan neuron McCulloch-Pitts (Haykin, 2001). Neuron merupakan medium pemrosesan dan penghantaran maklumat bagi satu-satu rangkaian neural buatan. Berdasarkan kepada Rajah 2.2, tiga komponen penting yang merupakan pembentukan neuron adalah satu set sinaps atau sambungan rangkaian, satu penambah dan satu fungsi pengaktifan. Setiap sinaps bagi setiap neuron diberikan satu nilai pemberat. Anggap neuron k yang dianalisa mempunyai n bilangan sinaps atau n data masukan. Data atau signal x_j pada masukan sinaps ke- j yang disambungkan pada neuron k akan didarabkan dengan nilai pemberat sinaps ke- j tersebut, w_{kj} . Pendaraban ini menunjukkan kepentingan dan pengaruh nilai pemberat sinaps kepada keluaran pemrosesan bagi sesuatu neuron. Penambah pula berfungsi untuk menambah semua signal atau data x_j yang telah diberatkan atau didarabkan dengan pemberat sinaps masing-masing. Hasil tambah nilai-nilai tersebut akan dihantar kepada fungsi pengaktifan.



Rajah 2.2: Pemodelan neuron tak lelurus.

Pemodelan neuron secara matematik berdasarkan Rajah 2.2 boleh ditakrifkan berdasarkan dua persamaan berikut:

$$u_k = \sum_{j=1}^n w_{kj} x_j \quad (2.1)$$

dan

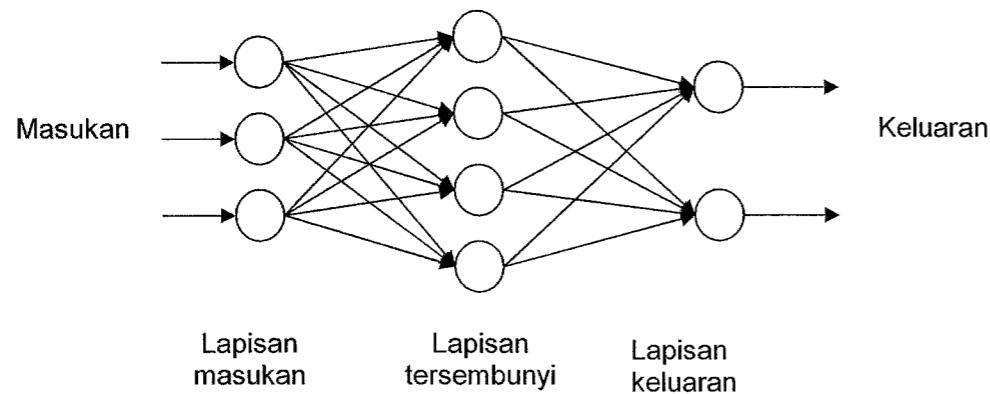
$$y_k = \varphi(u_k) \quad (2.2)$$

Berdasarkan persamaan (2.1) dan (2.2), u_k adalah keluaran penambah, x_j adalah data atau signal masukan pada sinaps ke- j , w_{kj} adalah pemberat sinaps ke- j untuk neuron k dan $\varphi(\cdot)$ adalah fungsi pengaktifan. Manakala, y_k dan n masing-masing adalah keluaran bagi neuron k dan bilangan masukan atau sinaps.

2.4.2.2 Seni Bina

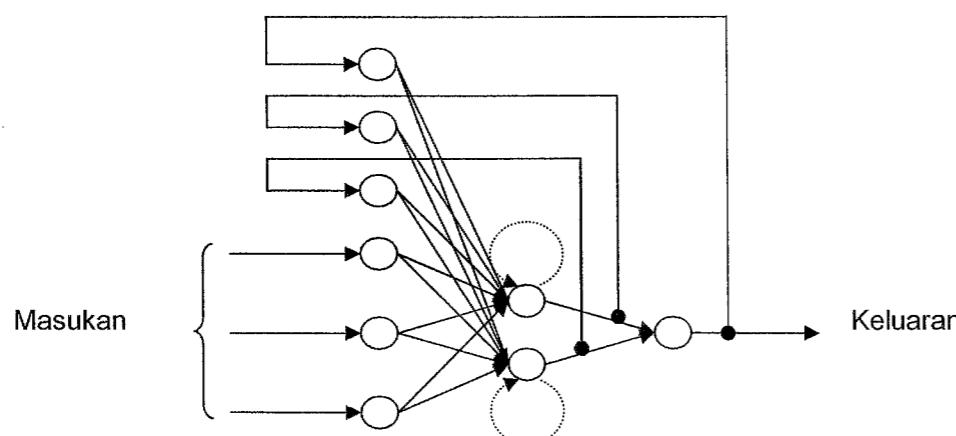
Rangkaian neural buatan ini dibezakan di antara satu sama lain oleh seni bina dan teknik atau algoritma pembelajarannya (Whelstead, 1994). Pemilihan seni bina rangkaian neural buatan yang sesuai adalah penting untuk menentukan kecekapan rangkaian neural tersebut. Seni bina rangkaian neural terbahagi kepada dua iaitu rangkaian neural suap depan (*feedforward neural network*) dan rangkaian neural suap

balik (feedback neural network) (Norgaard et al., 2000, Gallant, 1995, Haykin, 2001). Rajah 2.3 menunjukkan gambarajah rangkaian neural suap depan. Rangkaian ini terdiri daripada tiga lapisan utama iaitu lapisan masukan, tersembunyi dan keluaran. Lapisan tersembunyi ini bertindak sebagai perantaraan di antara lapisan masukan dan lapisan keluaran.



Rajah 2.3: Rangkaian neural suap depan.

Rajah 2.4 menunjukkan gambarajah rangkaian neural suap balik. Rangkaian ini membenarkan data atau isyarat bergerak pada pelbagai arah berdasarkan gelung pergerakan yang ditetapkan oleh rangkaian. Data atau isyarat boleh merambat ke hadapan dan boleh disuap balik untuk menjadi masukan kepada neuron sebelumnya.



Rajah 2.4: Rangkaian neural suap balik.

Berdasarkan kepada seni bina di antara kedua-dua jenis rangkaian neural ini, rangkaian neural suap depan mampu memberikan keputusan lebih pantas berbanding rangkaian neural suap balik disebabkan oleh keupayaan rangkaian neural suap depan untuk memodelkan sesuatu sistem dengan satu laluan ke hadapan sahaja tanpa perlu menyuap-balik sebarang keluaran (Nelson & Illingworth, 1992).

2.4.2.3 Proses Pembelajaran

Proses pembelajaran merupakan satu proses yang penting dalam rangkaian neural. Proses ini sangat penting bagi membolehkan rangkaian neural dapat berfungsi sama seperti rangkaian neural biologi. Terdapat dua jenis paradigma pembelajaran yang selalu diaplikasikan dalam rangkaian neural iaitu pembelajaran terselia (*supervised learning*) dan pembelajaran tak terselia (*unsupervised learning*) (Haykin, 2001). Set data pula akan dibahagikan kepada dua bahagian iaitu set data latihan dan set data ujian, masing-masing akan digunakan pada fasa latihan dan pada fasa ujian (Guo et. al., 2004).

Bagi pembelajaran terselia, pembelajaran ini mampu membentuk fungsi-fungsi berdasarkan set data-data latihan. Set data latihan yang digunakan perlu mengandungi sepasang parameter masukan dan parameter keluaran yang dikehendaki. Satu-satu keluaran bagi fungsi merupakan satu nilai yang berterusan yang dipanggil regresi atau nilai yang boleh meramal label-label kelas bagi data-data masukan yang dipanggil pengklasifikasi. Pembelajaran tak terselia pula merupakan kaedah pembelajaran yang mana model latihan telah ditetapkan bagi melakukan anggaran. Pembelajaran ini berlainan daripada pembelajaran terselia kerana pembelajaran ini tidak mempunyai sasaran keluaran. Dalam pembelajaran tak terselia, set data masukan akan dikumpulkan. Pembelajaran tak terselia akan menganggap data masukan adalah satu set pembolehubah rawak.

2.4.3 Aplikasi Rangkaian Neural Dalam Menentukan Komposisi Alga

Penggunaan rangkaian neural telah berjaya diaplikasikan dalam menentukan kualiti air sungai. Pelbagai parameter masukan telah digunakan dalam mencari parameter manakah yang mampu memberikan tahap kejituhan yang tertinggi. Kebanyakan pembolehubah yang digunakan sangat bergantung kepada musim. Oleh sebab itu, para penyelidik terpaksa mencari parameter yang kurang bergantung kepada perubahan musim. Kumpulan alga telah diketahui umum merupakan kumpulan ekosistem akuatik yang terbesar. Kajian yang dilakukan kepada komposisi alga didapati bahawa komposisi alga ini mampu menjadi satu parameter masukan yang baik disamping paling kurang bergantung kepada sebarang perubahan musim.

Setakat kajian ilmiah yang dijalankan, didapati masih belum pernah lagi komposisi alga dijadikan sebagai parameter masukan kepada rangkaian neural. Pembangunan rangkaian neural dengan menggunakan komposisi alga sebagai parameter masukan hanya diaplikasikan dalam meramal pertumbuhan alga pada masa akan datang. Reckhegel *et. al.*, (1997) telah menjalankan kajian untuk meramal pertumbuhan alga menggunakan rangkaian neural. Rangkaian neural telah digunakan sebagai alternatif kepada pendekatan semasa yang selalu digunakan. Penggunaan rangkaian neural dalam meramal pertumbuhan alga di dalam air telah mendapat pengesahan ahli-ahli limnologi di mana ramalan yang telah dilakukan adalah tepat dan keputusan yang diberikan boleh diterimapakai. Keputusan ramalan itu juga telah menunjukkan bahawa rangkaian neural mampu disesuaikan dengan kekompleksan dan ketidak lurusan komposisi alga.

Berikutkan kejayaan yang ditunjukkan oleh kajian Rechnagel, maka satu rangkaian neural telah dibangunkan dengan menggunakan komposisi alga sebagai parameter masukan. Komposisi alga mampu dijadikan parameter masukan dengan baik untuk meramal pertumbuhan alga pada masa akan datang. Oleh itu, pembangunan satu sistem penentuan kualiti air berdasarkan komposisi alga menggunakan rangkaian neural tidak mustahil dapat dilakukan.

2.5 Ringkasan

Penerangan yang diberikan dalam bahagian 2.2 menerangkan klasifikasi kualiti air sungai dan beberapa kaedah bagi menentukan kualiti air tersebut. Penentuan kualiti air adalah penting bagi mengukur tahap pencemaran air yang telah berlaku. Beberapa kaedah telah digunakan dalam memantau dan menentukan kualiti air iaitu dengan menggunakan kaedah pemantauan secara fizikal, pemantauan secara kimia dan pemantauan secara biologi.

Komposisi alga pula diterangkan dalam bahagian 2.3, di mana bahagian ini menerangkan penggunaan alga yang mampu digunakan sebagai penunjuk kepada tahap kualiti air. Komposisi alga adalah salah satu parameter yang telah digunakan di dalam kaedah pemantauan secara biologi. Daripada penyelidikan terdahulu, komposisi alga mampu mengakibatkan keracunan kepada sumber air seterusnya mendatangkan penyakit kepada manusia, haiwan, dan tumbuh-tumbuhan. Rangkaian neural pula diterangkan dalam bahagian 2.4. Analogi konsep pengoperasian kepada pemikiran otak manusia telah berjaya diaplikasikan menggunakan rangkaian neural. Kesesuaian sesuatu rangkaian neural untuk digunakan sebagai fungsi tertentu bergantung kepada pemilihan jenis seni bina, pemilihan teknik pembelajaran, penyediaan data masukan, pemilihan nilai-nilai parameter awalan dan saiz rangkaian, dan analisa akhir prestasinya.

Penyelidikan terdahulu telah berjaya menggunakan rangkaian neural seperti rangkaian perseptron berbilang lapisan (*Multilayered Perceptron, MLP*) dalam menentukan tahap kualiti air. Kajian penentuan kualiti air menggunakan rangkaian neural berdasarkan parameter-parameter kimia telah dijalankan dengan meluas namun kajian penentuan kualiti air menggunakan rangkaian neural berdasarkan parameter-parameter biologi buat masa kini masih belum diterokai. Oleh itu, kajian yang sedang dijalankan ini akan menentukan kualiti air sungai menggunakan rangkaian neural dengan menggunakan parameter-parameter biologi sebagai parameter masukan.

BAB 3

PENENTUAN KUALITI AIR SUNGAI MENGGUNAKAN RANGKAIAN NEURAL

3.1 Pengenalan

Penggunaan rangkaian neural sebagai teknik pengklasifikasian tahap kualiti air telah berjaya meningkatkan mutu penentuan kualiti air yang sebelum ini hanya menggunakan kaedah manual (Zaheer & Bai, 2003, Suen & Eheart, 2003). Komposisi alga telah dipilih sebagai parameter masukan kepada rangkaian neural berikutnya kejayaan yang ditunjukkan semasa pemantauan secara biologi yang dilakukan dengan menggunakan komposisi alga sebagai penunjuk (Maznah & Mansor, 2002).

Komposisi alga telah disahkan mampu menjadi penunjuk atau parameter dalam menentukan tahap kualiti air (Maznah & Mansor, 2002). Tahap kualiti air dikelaskan kepada tiga kategori iaitu bersih (*clean*), pertengahan (*moderate*) dan tercemar (*polluted*). Penyelidikan ini juga telah mengambil satu langkah ke hadapan dengan mencadangkan penggunaan rangkaian neural untuk menentukan jenis air terlebih dahulu sebelum tahap kualiti air dikesan. Oleh kerana penyelidikan ini menentukan jenis dan kualiti air, maka sampel-sampel air sungai dijadikan parameter masukan. Setelah jenis air dikelaskan, barulah kualiti air itu dapat ditentukan. Ini bermakna setiap sampel air yang diuji akan ditentukan jenisnya sama ada dari jenis air tawar atau air payau kemudian, air ini akan ditetapkan kualitinya sama ada bersih, pertengahan atau tercemar.

Bab ini akan membincangkan tentang analisis diskriminan dan rangkaian neural yang akan digunakan dalam menentukan jenis dan tahap kualiti air sungai. Antara rangkaian-rangkaian neural yang digunakan ialah rangkaian RBF, MLP, HMLP, HiMLP dan H²MLP. Algoritma latihan yang digunakan untuk menentukan parameter-parameter rangkaian neural tersebut juga akan diterangkan. Selepas itu, metodologi penggunaan rangkaian-rangkaian neural tersebut akan dibincangkan.

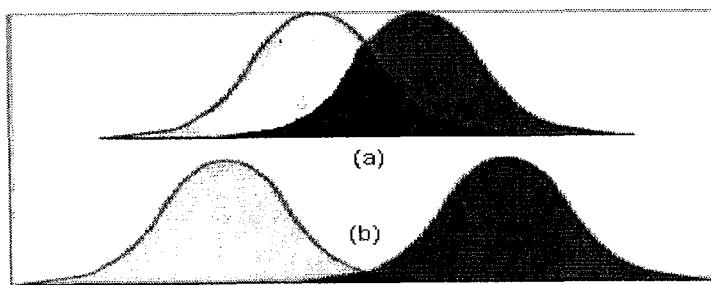
3.2 Analisis Diskriminan

Pada peringkat awal, analisis diskriminan (*Discriminant Analysis, DA*) hanya diaplikasikan dalam bidang perubatan (Golland *et. al.*, 1999) dan biologi sahaja (Zhang, 2000). Berdasarkan kejayaan yang ditunjukkan oleh kajian terdahulu, analisis diskriminan digunakan dengan meluas dalam bidang perniagaan (Tang & Chi, 2005), pembelajaran (Sheel, 2001) dan kejuruteraan (Mitra & Pal, 2002).

DA merupakan satu teknik statistik yang diaplikasikan bagi mengkelaskan sesuatu cerapan kepada beberapa kumpulan pengkelasan. Kumpulan pengkelasan perlu ditetapkan terlebih dahulu berdasarkan sifat-sifat cerapan. DA akan menganalisa data-data apabila pembolehubah bersandar yang berbentuk kategori kumpulan dan pembolehubah tak bersandar berada dalam satu selang yang sama. Objektif pertama DA adalah menghasilkan satu fungsi diskriminan atau kombinasi linear di antara pembolehubah tak bersandar yang terbaik yang akan menjadi penentu diskriminan yang akan mengasingkan data-data kepada beberapa kumpulan pengkelasan berdasarkan sifat-sifat penentu diskriminan tersebut dan kedua pula adalah menentukan pembolehubah tak bersandar yang memberikan impak yang tinggi terhadap kumpulan-kumpulan.

3.2.1 Aplikasi Analisis Diskriminan

DA dapat diaplikasikan kepada set-set data sekiranya pembolehubah bersandar dapat dikelaskan kepada beberapa kumpulan manakala bagi pembolehubah tak bersandar pula adalah bersifat selanjar. Keluaran yang diberikan oleh analisis diskriminan akan diukur berdasarkan taburan data dan darjah tindanan yang berlaku. Skor-skor diskriminan mestilah bertabur dalam taburan normal dan bertindan dalam darjah tindanan yang kecil. Rajah 3.1 menunjukkan taburan normal yang baik dan taburan normal yang kurang baik bagi analisis diskriminan.



Rajah 3.1: Taburan normal yang bagi analisis diskriminan bagi (a) taburan normal yang kurang baik dan (b) taburan normal yang baik.

Skor diskriminan adalah hasil tambah pemberat lelurus bagi pemberat diskriminan. Kesemua pemberat dipiawaikan bagi memastikan perbezaan skala di antara pembolehubah-pembolehubah dihapuskan. Skor diskriminan adalah dikira berdasarkan fungsi diskriminan yang diperolehi dan dinyatakan dalam persamaan (3.1):

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n + \beta \quad (3.1)$$

dengan

Z adalah skor bagi fungsi diskriminan,

$\alpha_1, \alpha_2, \dots, \alpha_n$ adalah pekali diskriminan,

X_1, X_2, \dots, X_n adalah pembolehubah tak bersandar,

dan β adalah pemalar yang ditetapkan.

Rajah 3.2 pula menunjukkan serakan data yang berlaku semasa analisis diskriminasi dijalankan. Analisis ini dijalankan kepada dua kumpulan pembolehubah bersandar. Oleh itu, data akan diserakkan kepada dua kumpulan dan setiap pembolehubah akan ditetapkan satu pusat serakan dengan data akan tertumpu di pusat itu.

3.2.2 Tahap Keertian Pembolehubah Diskriminan

Setiap pembolehubah dalam analisis diskriminasi mempunyai tahap keertian sama ada sama atau berbeza di antara satu sama lain. Oleh sebab itu, ujian kepada tahap keertian pembolehubah diskriminasi perlu dijalankan. Satu ujian univariat dijalankan kepada pembolehubah tak bersandar bagi mendapatkan nilai purata bagi setiap kumpulan pengelasan. Pembolehubah tak bersandar akan diuji secara satu persatu bermula daripada pembolehubah tak bersandar pertama sehingga pembolehubah tak bersandar terakhir. Satu hipotesis dibentuk seperti berikut:

$$H_o : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2 \quad (3.2)$$

dengan μ_1 dan μ_2 masing-masing merupakan purata bagi kumpulan satu dan kumpulan dua. Hipotesis nol, H_o menyatakan bahawa nilai purata bagi pembolehubah tak bersandar bagi kedua-dua kumpulan adalah sama manakala bagi hipotesis alternatif, H_a pula menyatakan nilai purata bagi pembolehubah tak bersandar adalah tidak sama. Hipotesis nol akan ditolak sekiranya paras keertian melebihi 5% manakala hipotesis nol diterima sekiranya paras keertian kurang daripada 5%.

Ujian multivariat juga dijalankan kepada pembolehubah tak bersandar bagi mendapatkan nilai purata bagi setiap kumpulan pengelasan. Pembolehubah tak bersandar bagi ujian multivariat akan diuji secara keseluruhan, berbanding ujian univariat yang melakukan ujian kepada pembolehubah tak bersandar secara satu persatu. Satu hipotesis dibentuk seperti berikut:

$$H_o : \begin{pmatrix} \mu_{1,x1} \\ \vdots \\ \mu_{1,xn} \end{pmatrix} = \begin{pmatrix} \mu_{2,x1} \\ \vdots \\ \mu_{2,xn} \end{pmatrix}$$

$$H_a : \begin{pmatrix} \mu_{1,x_1} \\ \vdots \\ \vdots \\ \mu_{1,x_n} \end{pmatrix} \neq \begin{pmatrix} \mu_{2,x_1} \\ \vdots \\ \vdots \\ \mu_{2,x_n} \end{pmatrix} \quad (3.3)$$

Sama seperti ujian univariat yang dilakukan, hipotesis nol akan ditolak apabila paras keertian ujian mutivariat melebihi 5% manakala hipotesis nol diterima apabila paras keertian kurang daripada 5%. Penolakan hipotesis nol memberikan keputusan purata bagi pembolehubah tak bersandar bagi kedua-dua kumpulan adalah tidak sama. Oleh kerana nilai paras keertian melebihi 5%, maka analisa seterusnya iaitu analisis fungsi diskriminan tidak dapat dibentuk. Sekiranya hipotesis nol diterima, purata keseluruhan penbolehubah tak bersandar adalah sama dan pembolehubah tak bersandar adalah bererti. Oleh itu analisa untuk menerbitkan fungsi diskriminan dapat dilakukan dan satu model diskrimian dapat dibentuk seperti Persamaan 3.1.

3.2.3 Kaedah Titik Pemisahan

Skor diskriminan yang diperolehi dapat digunakan bagi menentukan nilai titik pemisah di antara kelas-kelas. Bagi pengkelasan kepada dua kumpulan, secara amnya pada satu kumpulan pengkelasan akan mempunyai nilai skor diskriminan yang melebihi nilai titik pemisah manakala satu kumpulan pengkelasan lagi mempunyai nilai skor diskriminan kurang daripada nilai titik pemisah. Kaedah berikut selalu digunakan bagi mencari nilai titik pemisah seperti berikut:

$$\text{Titik pemisah, } Z_p = \left(\frac{\bar{Z}_1 + \bar{Z}_2}{2} \right); \quad (3.4)$$

di mana Z_1 dan Z_2 masing-masing adalah purata nilai skor diskriminan bagi kumpulan satu dan kumpulan dua.

3.3 Rangkaian Fungsi Asas Jejarian

Powell (1985), telah mempelopori penggunaan fungsi asas jejarian (*Radial Basis Function, RBF*) bagi menyelesaikan masalah penentu-dalam berbilang pembolehubah (*multivariable interpolation problem*). Berdasarkan fungsi yang telah dirintis oleh Powell (1985), Broomhead & Lowe (1988) telah menggunakan penemuan Powell (1985) di dalam pembinaan satu rangkaian neural yang dikenali sebagai rangkaian RBF. Beberapa penyelidikan telah dijalankan dan telah memberikan sumbangan besar terhadap teori, pembinaan struktur dan aplikasi rangkaian RBF (Mashor, 2000a & Poggio & Girosi, 1990).

3.3.1 Seni Bina dan Ciri-ciri Rangkaian

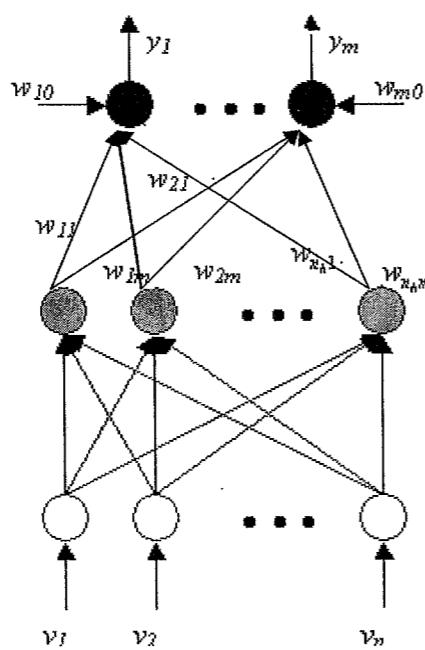
Ciri-ciri rangkaian RBF adalah berdasarkan konfigurasi rangkaian neural suap depan seperti yang ditunjukkan di dalam Rajah 3.5. Rangkaian ini terdiri daripada lapisan masukan, lapisan tersembunyi dan lapisan keluaran. Nod-nod berwarna putih pada Rajah 3.5 mewakili lapisan masukan kepada rangkaian RBF yang berfungsi untuk menyediakan data masukan. Nod-nod yang berwarna kelabu pada Rajah 3.5 pula mewakili lapisan tersembunyi dan disertakan bersama satu set fungsi asas jejarian yang telah dipilih. Setiap nod tersembunyi terdiri daripada satu susunan nod-nod yang mengandungi vektor parameter yang dinamakan pusat. Nod-nod berwarna hitam pada Rajah 3.5 mewakili lapisan keluaran di mana tindakbalas keluaran akan disediakan terhadap corak pengaktifan data masukan. Keluaran satu rangkaian RBF diwakili oleh persamaan:

$$\hat{y}_i(t) = w_{i0} + \sum_{j=1}^{n_h} w_{ij} \phi\left(\|v(t) - c_j(t)\|\right), \quad i = 1, 2, \dots, m \quad (3.5)$$

di mana m adalah bilangan nod keluaran, n adalah bilangan nod masukan dan n_h adalah bilangan nod tersembunyi. Berdasarkan Persamaan (3.3), w_{ij} ialah pemberat sambungan dari nod tersembunyi kepada nod keluaran, w_{i0} ialah pemberat

sambungan pincang, $c_j(t)$, ialah pusat rangkaian RBF, $v(i)$ ialah data masukan, $\phi(\bullet)$ ialah fungsi asas tak lelurus dan $\|\bullet\|$ ialah ukuran jarak yang kebiasaannya di dalam bentuk Euclidean.

Fungsi asas jejarian diaplikasikan pada setiap nod tersembunyi bagi memproses data masukan yang telah diberikan. Penggunaan fungsi asas jejarian terhadap setiap nod tersembunyi akan menyebabkan rangkaian RBF bersifat tak lelurus. Ini bermakna setiap sistem yang lelurus akan dimodelkan sebagai sistem tak lelurus apabila pemodelan dilakukan menggunakan rangkaian RBF.



Rajah 3.5: Seni bina rangkaian RBF.

3.3.2 Penempatan Pusat Rangkaian dan Pemberat Sambungan Rangkaian

Poggio & Girosi (1990) telah memperkenalkan konsep awal penempatan pusat rangkaian. Penyelidikan itu berjaya mengumpukkan semua data kepada beberapa bilangan pusat tertentu. Penyelidikan itu juga telah membuktikan bahawa aturan pengemaskinian pusat rangkaian RBF berdasarkan pendekatan penurunan cerun (*gradient descent*) mampu menempatkan pusat rangkaian ke dalam taburan data

majoriti. Penyelidikan oleh Hush & Horne (1993) mendapat algoritma pengelompokan purata- k sering digunakan disebabkan oleh teknik ini mudah diimplementasi dan keupayaannya untuk memberikan keputusan yang baik. Teknik pengelompokan ini juga telah digunakan dengan meluas bagi menempatkan pusat rangkaian RBF. Teknik pengelompokan ini diimplementasi dengan anggapan pusat awalan bagi rangkaian RBF telah diberikan.

Pertimbangkan satu masalah yang mempunyai N data yang akan dikelompokkan kepada n_c pusat. Katakan v_i adalah data ke- i dan c_j adalah pusat ke- j di mana $i = 1, 2, \dots, N$, dan $j = 1, 2, \dots, n_c$. Nilai awalan pusat pengelompokan c_1, c_2, \dots, c_k diberikan. Semua data v_i diumpukkan kepada pusat terdekat dan lokasi pusat pengelompokan menggunakan persamaan :

$$c_j = \frac{1}{n_j} \sum_{i \in c_j} v_i \quad (3.7)$$

Pusat-pusat pengelompokan c_1, c_2, \dots, c_k dikemaskinikan dengan mencari purata bagi data-data v_i . Pusat-pusat pengelompokan c_1, c_2, \dots, c_k dikemaskinikan sehingga data-data menumpu pada pusat-pusat pengelompokan dengan baik, di mana nilai-nilai pusat tidak lagi berubah.

Parameter rangkaian RBF perlu ditentukan nilai pemberat sambungan tak lelurus di antara nod tersembunyi dengan nod keluaran. Secara tipikal, nilai pemberat sambungan rangkaian pada lapisan keluaran dapat ditentukan melalui penggunaan algoritma kuasa dua terkecil Givens (*Givens least square, GLS*) (Mashor, 2000a). Algoritma GLS ini diperkenalkan oleh Gentleman (1973).

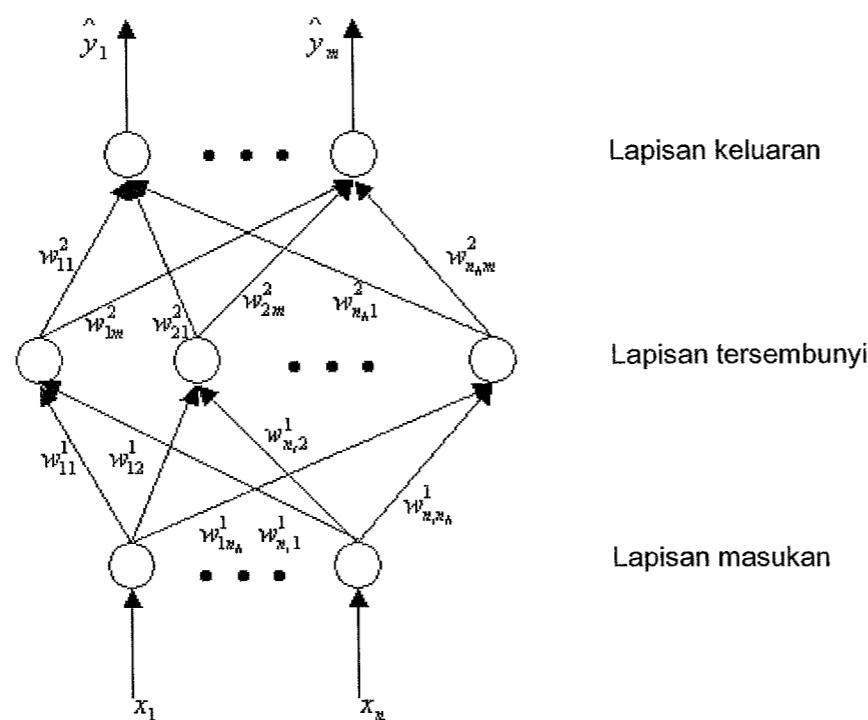
3.4 Rangkaian Perseptron Berbilang Lapisan

Rangkaian MLP berkeupayaan untuk digunakan sebagai pengimplementasi fungsi logik Boolean, mampu untuk mengklasifikasikan data serta berkeupayaan untuk mengimplementasikan transformasi tak lelurus bagi satu-satu proses penganggaran

(Crone et al., 2004). Penyelidikan ini mencadangkan penggunaan rangkaian MLP dalam menentukan tahap kualiti air sungai. Rangkaian ini akan dilatih menggunakan algoritma perambatan balik (*backpropagation*, BP), algoritma Lavenberg-Marquardt (LM) dan algoritma Bayesian Regularization (BR).

3.4.1 Seni Bina dan Ciri-ciri Rangkaian

Rosenblatt (1958) merupakan perintis kepada permodelan perseptron (Hush & Horne, 1993). Beberapa perseptron yang dikaskadkan di dalam satu lapisan akan membentuk satu rangkaian yang dikenali sebagai rangkaian MLP. Perseptron di dalam rangkaian MLP ini dirujuk sebagai neuron atau nod. Rajah 3.6 menunjukkan seni bina rangkaian MLP dengan n_i nod masukan, n_h nod tersembunyi dan m nod keluaran.



Rajah 3.6: Seni bina rangkaian MLP.

Berdasarkan kepada Rajah 3.6, data masukan, x_i akan merambat ke lapisan tersembunyi. Data masukan, x_i akan dihubungkan ke nod tersembunyi melalui

sambungan berpemberat, w_{ij}^1 . Setiap nod pada lapisan tersembunyi akan beroperasi seperti model perseptron dengan fungsi sigmoid sebagai fungsi pengaktifannya. Keluaran bagi nod tersembunyi ke- j , u_j diberi oleh persamaan:

$$u_j = \varphi\left(\sum_{i=1}^{n_i} w_{ij}^1 x_i + b_j^1\right); \quad \text{untuk } 1 \leq j \leq n_h \quad (3.28)$$

Berdasarkan Persamaan (3.28), $\varphi(\bullet)$ ialah fungsi pengaktifan, b_j^1 ialah nilai ambang pada nod tersembunyi ke- j .

Keluaran bagi setiap nod tersembunyi pula akan bergerak ke lapisan keluaran. Data keluaran nod tersembunyi akan dihubungkan ke nod keluaran melalui sambungan berpemberat, w_{jk}^2 . Keluaran bagi neuron keluaran ke- k , \hat{y}_k diberikan oleh persamaan:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 \varphi\left(\sum_{i=1}^{n_i} w_{ij}^1 x_i + b_j^1\right); \quad \text{untuk } 1 \leq k \leq m \quad (3.29)$$

3.4.2 Pemberat Sambungan Rangkaian

Kaedah pembelajaran merupakan satu prosedur untuk mengubahsuai pemberat dan nilai ambang. Kaedah ini juga dapat mengurangkan ketidaksepadanan antara keluaran sebenar dengan keluaran yang diberikan oleh rangkaian. Aggarwal (2005) menyatakan bahawa algoritma pembelajaran digunakan untuk mengubah pemberat-pemberat dan ambang bagi menggerakkan nilai-nilai keluaran rangkaian menghampiri nilai-nilai sasaran. Kajian ini mencadangkan penggunaan tiga jenis algoritma pembelajaran yang berlainan untuk melatih struktur rangkaian MLP. Algoritma latihan yang digunakan adalah perambatan balik (*back propagation*, BP), Levenberg Marquardt (LM) dan Bayesian Regularization (BR).

3.4.2.1 Algoritma Perambatan Balik

Algoritma BP telah diperkenalkan oleh Werbos (1974) melalui kajian yang dilakukannya. Rumelhart & McClelland (1986) telah membangunkan semula kajian dan penggunaan oleh Werbos dan kajian itu telah digunakan sehingga kini. Penggunaan algoritma BP mampu mengecilkan ralat purata kuasa dua (*Mean Square Error, MSE*) bagi pengiraan keluaran rangkaian.

Berikan nilai awalan kepada semua pemberat rangkaian dan nilai ambang. Masukan data dimasukkan kepada rangkaian neural setelah kesemua pemberat dan nilai ambang ditetapkan nilai awalannya. Keluaran bagi nod tersembunyi diperolehi menggunakan persamaan:

$$u_j = \varphi \left(\sum_{i=1}^{n_h} w_{ij}^1 x_i + b_j^1 \right) ; \text{ untuk } 1 \leq j \leq n_h \quad (3.30)$$

di mana w_{ij} adalah pemberat, u_j adalah keluaran pada lapisan tersembunyi, b_j^1 adalah nilai ambang dan φ adalah fungsi pengaktifan sigmoid. Keluaran pada nod tersembunyi pula diperolehi berdasarkan Persamaan (3.29). Ralat keluaran pada lapisan keluaran diberikan oleh persamaan:

$$e_j(t) = y(t)_j - \hat{y}_j(t) \quad (3.31)$$

dengan $\hat{y}_j(t)$ adalah keluaran ramalan neuron j pada lelaran t dan $\hat{y}_j(t)$ adalah keluaran sebenar neuron j pada lelaran t .

3.4.2.2 Algoritma Lavenberg Marquardt

Penggunaan algoritma LM sangat efisyen dalam memberikan tahap kejutan yang tinggi. Algoritma ini jelas bervariasi kepada kepentasan yang diberikan oleh kaedah Newton serta memberikan kestabilan yang tinggi seperti yang ditunjukkan oleh kaedah penurunan cerun. Algoritma ini digunakan untuk meminimumkan hasil tambah bagi fungsi-fungsi tak-lelurus. Langkah-langkah di bawah telah diaplikasikan oleh

Wilamowski *et. al.*, (2001) dalam kajian mereka. Dalam kajian tersebut, indeks prestasi rangkaian neural yang optimum diberikan oleh:

$$F(w) = \sum_{i=1}^i \left[\sum_{j=1}^j (\sigma_{ij} - y_{ij})^2 \right]^2 \quad (3.36)$$

dengan σ_{ij} dan y_{ij} masing-masing adalah keluaran ramalan dan keluaran sebenar rangkaian pada keluaran ke i pada lapisan ke j . $w = [w_1, w_2, \dots, w_n]^T$ adalah pemberat kepada rangkaian. Persamaan (3.37) boleh diringkaskan kepada:

$$F(w) = \hat{E}^T \hat{E} \quad (3.37)$$

dengan vektor ralat $\hat{E} = [e_1, e_2, \dots, e_k]^T$ dan ralat rangkaian adalah:

$$e_k = \sum_{p=1}^p (\sigma_{pj} - y_{pj})^2 \quad (3.38)$$

3.4.2.3 Algoritma Bayesian Regularization

Selain daripada menggunakan algoritma BP dan algoritma LM dalam melatih rangkaian MLP, rangkaian MLP ini dapat dilatih menggunakan algoritma latihan Penetapan Bayesian (*Bayesian Regularization, BR*). Algoritma latihan ini menganggap pemberat-pemberat dan pengambangan kepada rangkaian adalah dalam bentuk pembolehubah rawak berdasarkan taburan Gaussian dan berkait rapat dengan varians-variанс yang tidak diketahui. Duda *et. al.*, (2001) telah mencadangkan penggunaan hukum jiran terdekat- k (k -Nearest-Neighbor Rule) dalam menyelesaikan permasalah pengitlakan data bagi rangkaian ini berdasarkan algoritma BR.

Dengan menganggap $\omega_m(x)$ adalah persekutuan maksimum, kebarangkalian maksimum adalah:

$$P(\omega_m | x) = \max_i P(\omega_i | x) \quad (3.50)$$

Analisis dilakukan terhadap sifat-sifat hukum jiran terdekat akan menghasilkan keadaan tak terhingga bagi purata kebarangkalian ralat $P(e|x)$. Purata kebarangkalian ralat dilakukan terhadap kesemua data x :

$$P(e) = \int P(e|x)p(x)dx \quad (3.51)$$

$P(e)$ dapat diminimumkan dengan meminimumkan $P(e|x)$ bagi setiap data x .

3.5 Rangkaian Perseptron Berbilang Lapisan Hibrid

Penyelidikan ini juga menggunakan rangkaian HMLP bagi menentukan tahap kualiti air sungai. Rangkaian HMLP ini dipilih berdasarkan keupayaan rangkaian HMLP ini memberikan prestasi yang lebih baik berbanding rangkaian MLP konvensional (Mashor 2000b, 1999). Rangkaian ini akan dilatih menggunakan algoritma ralat ramalan jadi semula ubahsuai (*modified recursive prediction error, MRPE*) yang diperkenalkan oleh Mashor (2000b, 1999). Kajian tersebut membuktikan bahawa rangkaian HMLP yang dilatih menggunakan algoritma tersebut berupaya menumpu dengan lebih baik.

3.5.1 Seni Bina dan Ciri-ciri Rangkaian

Pengkaskadan beberapa perseptron di dalam satu lapisan akan membentuk satu rangkaian yang dinamakan rangkaian MLP. Rangkaian MLP mempunyai sifat tak lelurus yang tinggi yang menyebabkan sistem lelurus terpaksa dimodelkan dengan model tak lelurus. Maka sambungan lelurus ditambahkan kepada rangkaian MLP konvensional untuk mewujudkan satu rangkaian HMLP yang mampu memodelkan sistem lelurus tersebut dengan lebih baik seperti yang ditunjukkan di dalam Rajah 3.7 (Mashor, 2000b).

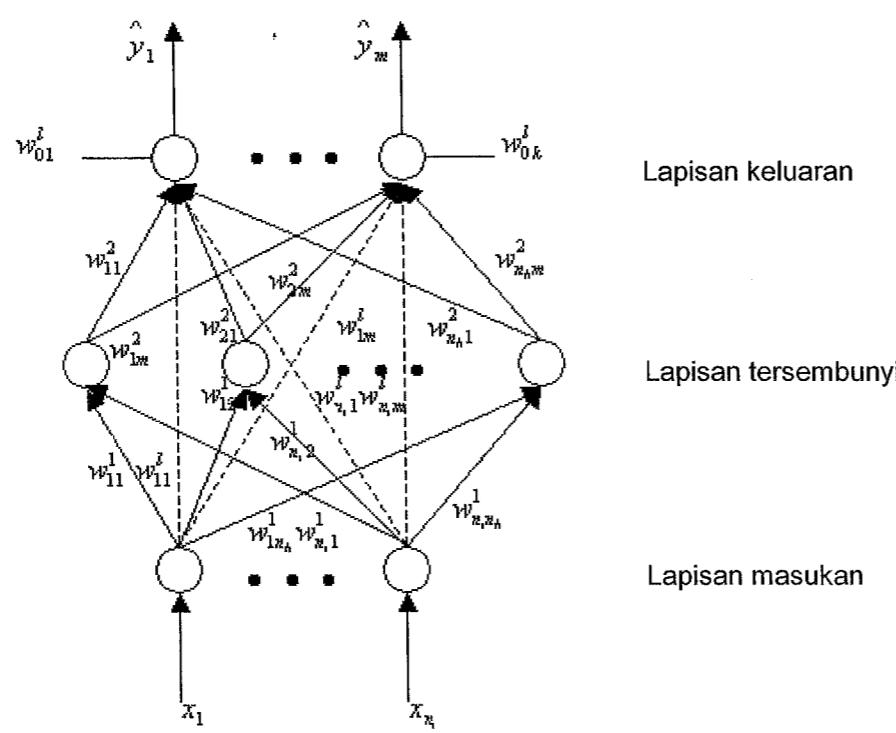
Berdasarkan Rajah 3.7, sambungan lelurus terbentuk daripada sambungan yang menghubungkan nod masukan terus kepada nod keluaran. Keluaran bagi rangkaian HMLP, \hat{y}_k diberikan oleh persamaan:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 \varphi \left(\sum_{i=1}^{n_h} w_{ij}^1 x_i + b_j^1 \right) + \sum_{i=0}^{n_h} w_{ik}^1 x_i \quad \text{untuk } 1 \leq k \leq m \quad (3.57)$$

Berdasarkan Persamaan (3.57), keluaran nod tersembunyi akan dihubungkan ke nod keluaran melalui sambungan berpemberat, w_{jk}^2 . Sebutan kedua pada bahagian kanan persamaan pula mewakili sambungan lelurus di antara nod masukan dengan nod keluaran dan w_{ik}^1 ialah pemberat sambungan lelurus tersebut. Berdasarkan Persamaan (3.57), apabila nilai $i = 0$, w_{ik}^1 dan x_i di dalam sebutan kedua masing-masing akan mewakili pemberat sambungan dan masukan untuk masukan pincang rangkaian. φ adalah satu fungsi pengaktifan dan penyelidikan ini memilih fungsi sigmoid sebagai fungsi pengaktifan. w_{jk}^2 , w_{ik}^1 , w_{ij}^1 dan nilai ambang b_j^1 tidak diketahui dan dikemaskini untuk meminimumkan ramalan ralat yang diberikan oleh :

$$\varepsilon_k(t) = y_k(t) - \hat{y}_k(t) \quad (3.58)$$

di mana $y_k(t)$ dan $\hat{y}_k(t)$ masing-masing adalah keluaran sebenar dan juga keluaran rangkaian.



Rajah 3.7: Seni bina rangkaian HMLP

Rangkaian HMLP ini telah dibuktikan mempunyai prestasi yang lebih baik berbanding rangkaian MLP konvensional (Mashor, 2000b, 1999). Penyelidikan yang terperinci telah dilakukan di dalam kedua-dua penyelidikan oleh Mashor tersebut untuk membuktikan keupayaan rangkaian HMLP tersebut. Rangkaian HMLP yang digunakan ini mampu menampung dengan lebih baik dan cepat berbanding rangkaian MLP dan RBF.

3.5.2 Pemberat Sambungan Rangkaian

Algoritma ralat ramalan jadi semula (*recursive prediction error, RPE*) telah digunakan bagi mengatasi beberapa kelemahan yang ditunjukkan oleh algoritma perambatan balik (*backpropagation, BP*). Antaranya kadar penumpuan yang perlakan ke atas rangkaian MLP. Algoritma ini mempunyai prestasi yang lebih baik berbanding algoritma perambatan balik. Mashor (2000b, 1999) pula telah melakukan sedikit mengubahsuai terhadap algoritma RPE. Pengubahsuai yang dilakukan ialah

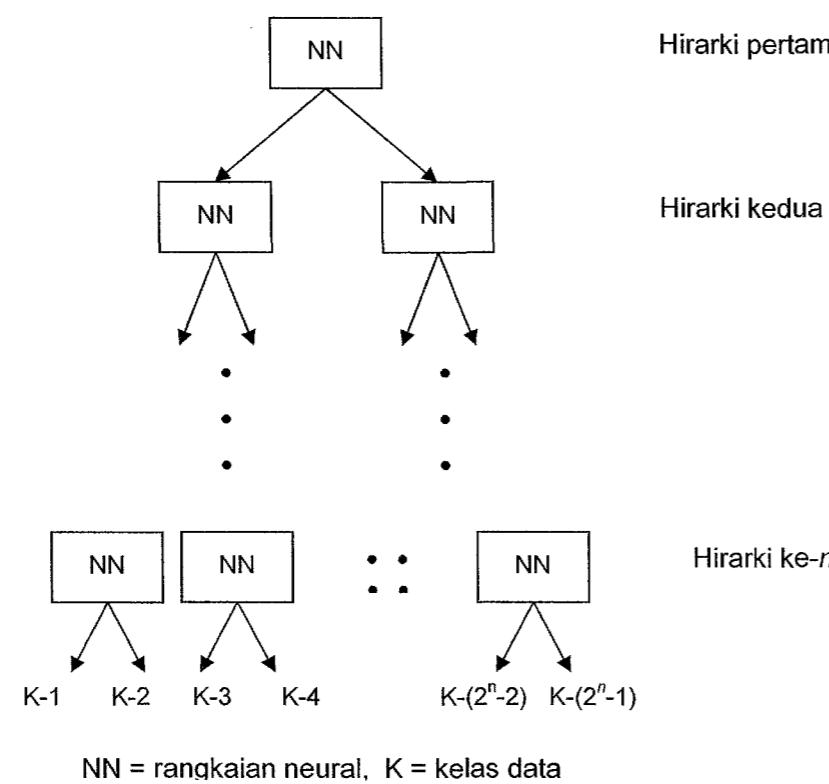
mengoptimum cara momentum dan kadar pembelajaran yang diumpukkan kepada algoritma RPE. Algoritma yang diubahsuai ini dikenali sebagai ralat ramalan jadi semula ubahsuai (*modified recursive prediction error, MRPE*). Rangkaian HMLP dengan algoritma MRPE mampu menumpu dengan kadar yang lebih cepat berbanding menggunakan algoritma RPE.

3.6 Rangkaian Neural Berhirark

Penyelidikan ini juga menggunakan rangkaian neural berhirarki bagi menentukan tahap kualiti air sungai. Rangkaian neural berhirarki ini digunakan kerana keupayaannya dalam membuat ramalan seterusnya mengklasifikasikan data-data ke dalam kumpulan-kumpulan tertentu. Mat-Isa (2002) telah menggunakan rangkaian berbilang lapisan hibrid berhirarki (*Heirarchical Hybrid Multilayered Perceptron, H²MLP*) dalam mengklasifikasikan sel barah pangkal rahim kepada sel normal, LSIL dan HSIL. Rajah 3.8 menunjukkan rangkaian neural berhirarki yang dicadangkan. bagi menyelesaikan masalah yang melibatkan lebih daripada dua klasifikasi. Pembangunan rangkaian neural berhirarki ini dapat menggantikan penggunaan satu rangkaian neural dengan nod keluaran yang berbilang. Rangkaian neural berhirarki ini hanya membenarkan satu nod keluaran sahaja pada satu-satu masa. Rangkaian ini membenarkan setiap masukan data diklasifikasikan kepada dua kelas yang berlainan sahaja.

Daripada Rajah 3.8, dapat diperhatikan rangkaian neural berhirarki ini mengandungi n peringkat hirarki di mana rangkaian akan berfungsi berperingkat-peringkat. Bermula pada hirarki peringkat pertama, rangkaian neural ini akan berfungsi sebagai data masukan kepada rangkaian neural pada hirarki kedua. Data keluaran rangkaian neural pada hirarki peringkat pertama akan menjadi data masukan kepada rangkaian neural pada hirarki peringkat kedua. Pada hirarki peringkat kedua, kelas pertama bagi hirarki peringkat pertama akan menjadi data masukan kepada rangkaian neural di sebelah kiri manakala kelas kedua bagi hirarki peringkat pertama akan

menjadi data masukan kepada rangkaian neural di sebelah kanan. Proses memasukkan data masukan kepada rangkaian neural pada setiap peringkat hirarki akan berterusan sehingga bilangan hirarki ke- n .



Rajah 3.8: Gambarajah blok rangkaian neural berhirarki.

3.7 Sampel Data

78 jenis spesis alga yang telah berjaya dicerap dan akan digunakan sebagai data masukan kepada rangkaian-rangkaian neural. Kehadiran alga semasa sampel dicerap dijadikan penunjuk kepada data masukan. Kehadiran alga di dalam sampel air akan ditandakan dengan '1' manakala '0' bagi alga yang tidak terdapat di dalam air semasa sampel dicerap. Sampel-sampel data telah dicerap oleh seorang ahli limnologi yang pakar dalam bidang ekologi alga, Dr Wan Maznah Wan Omar dari Pusat Pengajian Sains Kajihayat, USM. Bilangan data yang dicerap adalah sebanyak 320 data. Sebanyak 180 data (32 data air tawar berkualiti bersih, 32 data air tawar berkualiti pertengahan, 52 data air tawar berkualiti tercemar, 35 data air payau

berkualiti pertengahan dan 32 data air payau berkualiti tercemar) digunakan sebagai data latihan manakala sebanyak 140 data (23 data air tawar berkualiti bersih, 31 data air tawar berkualiti pertengahan, 39 data air tawar berkualiti tercemar, 16 data air payau berkualiti pertengahan dan 31 data air payau berkualiti tercemar) digunakan sebagai data ujian. Sampel data telah dicerap di 12 buat stesen cerapan di sepanjang lembangan Sungai Pinang, Pulau Pinang.

3.8 Metodologi

Penggunaan 78 jenis alga sebagai parameter masukan kepada rangkaian neural akan menyebabkan rangkaian neural menjadi sangat kompleks. Oleh itu, analisis diskriminan (*Discriminant Analysis, DA*) digunakan untuk memilih dan menentukan spesis-spesis alga yang bererti atau memberi impak yang tinggi dalam menentukan jenis dan tahap kualiti air sungai. Di samping itu juga, DA ini juga mampu melakukan pengelaskan jenis dan tahap kualiti air sungai. Penyelidikan ini menggunakan dua jenis rangkaian neural yang sering digunakan iaitu rangkaian MLP dan RBF konvensional sebagai sistem penentuan tahap kualiti air. Selain itu, penyelidikan ini juga mencadangkan penggunaan satu rangkaian hibrid iaitu rangkaian perseptron berbilang lapisan hibrid (*Hybrid Multilayered Perceptron, HMLP*). Penyelidikan ini juga mencadangkan penggunaan dua rangkaian berhirarki iaitu rangkaian perseptron berbilang lapisan berhirarki (*Heirarchical Multilayered Perceptron, HiMLP*) dan rangkaian perseptron berbilang lapisan hibrid berhirarki (*Heirarchical Hybrid Multilayered Perceptron, H²MLP*) bagi menentukan jenis dan kualiti air dengan lebih baik lagi. Spesis-spesis alga yang telah dikenalpasti oleh DA memberikan impak yang tinggi dalam mengelaskan jenis dan tahap kualiti air sungai akan digunakan sebagai data masukan kepada rangkaian-rangkaian neural tersebut.

3.9 Kesimpulan

Secara amnya, bab ini menerangkan tentang keseluruhan proses yang telah dijalankan di dalam kajian ini. Proses kajian dimulakan dengan memilih spesis-spesis alga yang dominan sehingga kepada proses pembangunan sistem penentuan jenis dan kualiti air sungai. Antara proses-proses yang perlu dilalui bagi melengkapkan kajian ini adalah penentuan spesis-spesis alga dominan, analisis kejituhan pengelasan jenis dan tahap kualiti air sungai serta pembangunan satu sistem penentuan jenis dan kualiti air sungai. Analisis diskriminan telah diaplikasikan bagi menentukan spesis-spesis alga yang memberikan impak yang tinggi dalam mengelaskan jenis dan kualiti air sungai. Sebanyak 78 spesis alga terdapat di dalam lembangan Sungai Pinang dan penentuan spesis alga yang dominan dapat mengurangkan kekompleksan rangkaian neural yang akan dibangunkan kerana hanya spesis-spesis yang dominan sahaja akan dijadikan masukan data kepada rangkaian neural. Antara rangkaian neural yang dibangunkan adalah rangkaian RBF, MLP, HMLP, HiMLP dan H²MLP. Rangkaian yang memberikan kejituhan ramalan yang tinggi akan dipilih untuk diimplementan ke dalam sistem penentuan jenis dan kualiti air sungai. Pembangunan sistem akan dilakukan menggunakan rangkaian neural terbaik bagi memperolehi satu pengelasan yang mempunyai kadar ketepatan yang tinggi.

BAB 4

KEPUTUSAN DAN PERBINCANGAN

4.1 Pengenalan

Di dalam bab ini, beberapa ujian akan dilakukan menggunakan analisis diskriminan bagi menentukan spesis-spesis alga yang dominan. Alga dominan ini akan digunakan sebagai parameter masukan kepada rangkaian neural. Ujian yang dijalankan merangkumi ujian univariat dan ujian multivariat. Keputusan ujian-ujian yang dilakukan akan dibincangkan dengan terperinci di dalam bab ini.

4.2 Keputusan Analisis Diskriminan

Berdasarkan kepada data-data spesis alga yang diperolehi, terdapat 78 spesis yang akan digunakan sebagai parameter masukan kepada rangkaian neural. Penggunaan kesemua spesis alga sebagai parameter masukan akan menyebabkan rangkaian menjadi sangat kompleks. Oleh itu, analisis diskriminan dilaksanakan bagi mengenalpasti spesis-spesis alga dominan yang memberikan impak yang tinggi sungai. Sampel-sampel data akan dikelaskan kepada lima kumpulan di mana setiap kumpulan akan mewakili jenis dan kualiti air sungai yang dicerap.

4.2.1 Ujian Univariat

Jadual 4.1 menunjukkan keputusan bagi ujian univariat yang dilakukan terhadap parameter-parameter diskriminan. Berdasarkan Jadual 4.1, didapati bahawa parameter diskriminan *Frustulia Rhomboides* (X_{38}), *Nitzschia Angustata* (X_{57}), *Pinnularia Mesolepta* (X_{67}) dan *Pinnularia Splendida* (X_{71}) memberikan taburan nilai-p yang lebih besar daripada 5%. Keputusan yang diperolehi juga menunjukkan bahawa keempat-empat parameter ini memberikan impak yang rendah kepada proses pengelasan manakala parameter-parameter selainnya pula memberikan impak yang

tinggi kepada proses pengelasan kerana taburan nilai-p adalah kurang daripada 5%.

Oleh itu hipotesis nol ditolak pada keempat-empat parameter di atas manakala hipotesis nol diterima bagi pembolehubah-pembolehubah tak bersandar yang lain.

Jadual 4.1: Keputusan ujian univariat bagi parameter-parameter diskriminan.

Komposisi Alga	Pengenalan	Taburan Nilai-p
<i>Acnanthes Exigua</i>	X_1	0.000
<i>Acnanthes Exigua</i> Var <i>Heterovalva</i>	X_2	0.000
<i>Acnanthes Minutissima</i>	X_3	0.000
<i>Acnanthes Oblongela</i>	X_4	0.000
<i>Amphora Libica</i>	X_5	0.000
<i>Anomoeneis Brachysira</i>	X_6	0.000
<i>Coccineis Pediculus</i>	X_7	0.000
<i>Coccineis Placetula</i>	X_8	0.000
<i>Coccineis Species</i>	X_9	0.000
<i>Coccineis Thumensis</i>	X_{10}	0.000
<i>Coscinodiscus Antiquus</i>	X_{11}	0.000
<i>Coscinodiscus Argus</i>	X_{12}	0.000
<i>Coscinodiscus Decipiens</i>	X_{13}	0.000
<i>Coscinodiscus Excentricus</i>	X_{14}	0.000
<i>Coscinodiscus Granii</i>	X_{15}	0.000
<i>Cymbella Species</i>	X_{16}	0.000
<i>Cymbella Inaequalis</i>	X_{17}	0.000
<i>Cyclotella Comta</i>	X_{18}	0.000
<i>Cyclotella Striata</i>	X_{19}	0.000
<i>Cyclotella Striata</i> Var <i>Baltica</i>	X_{20}	0.000
<i>Diatoma Elongatum</i>	X_{21}	0.000
<i>Diatoma Species</i>	X_{22}	0.000
<i>Diploneis Bombus</i>	X_{23}	0.000
<i>Diploneis Decipiens</i>	X_{24}	0.000
<i>Diploneis Elliptica</i>	X_{25}	0.002
<i>Diploneis Interrupta</i>	X_{26}	0.000
<i>Eutonia Faba</i>	X_{27}	0.000
<i>Eutonia Grunowi</i>	X_{28}	0.000
<i>Eutonia Monodon</i> Var <i>Alpina</i>	X_{29}	0.000
<i>Eutonia Pectinalis</i>	X_{30}	0.000

<i>Eutonia Sadetica</i> Var <i>Inasa</i>	X_{31}	0.000
<i>Eutonia Pectinalis</i>	X_{32}	0.000
<i>Eutonia Pectinalis</i> Var <i>Minor</i>	X_{33}	0.000
<i>Flagilaria Capucina</i>	X_{34}	0.000
<i>Flagilaria Species 1</i>	X_{35}	0.000
<i>Flagilaria Species 2</i>	X_{36}	0.000
<i>Flagilaria Crotonensis</i>	X_{37}	0.000
<i>Frustulia Rhomboides</i>	X_{38}	0.639
<i>Frustulia Rhomboides</i> Var <i>Saxonica</i>	X_{39}	0.000
<i>Frustulia Saxonica</i>	X_{40}	0.001
<i>Gomphonema Acuminatum</i>	X_{41}	0.001
<i>Gomphonema Gracile</i>	X_{42}	0.000
<i>Gomphonema Longiceps</i>	X_{43}	0.002
<i>Gomphonema Parvalum</i>	X_{44}	0.000
<i>Gomphonema Subventricosum</i>	X_{45}	0.000
<i>Hantzschia Ampioxys</i>	X_{46}	0.006
<i>Navicula Crytocephala</i>	X_{47}	0.000
<i>Navicula Cuspidata</i>	X_{48}	0.000
<i>Navicula Halophila</i>	X_{49}	0.000
<i>Navicula Hustedtii</i>	X_{50}	0.000
<i>Navicula Obtusa</i>	X_{51}	0.000
<i>Navicula Obtusa</i> Var <i>Scalpelliformis</i>	X_{52}	0.000
<i>Navicula Radiosa</i>	X_{53}	0.000
<i>Navicula Sigma</i>	X_{54}	0.000
<i>Navicula Species</i>	X_{55}	0.000
<i>Neidium Affine</i>	X_{56}	0.007
<i>Nitzschia Angustata</i>	X_{57}	0.072
<i>Nitzschia Amphibia</i>	X_{58}	0.000
<i>Nitzschia Fanticola</i>	X_{59}	0.000
<i>Nitzschia Littoralis</i>	X_{60}	0.000
<i>Nitzschia Palea</i>	X_{61}	0.000
<i>Pinnularia Biceps</i>	X_{62}	0.000
<i>Pinnularia Biceps</i> F. <i>Petersenii</i>	X_{63}	0.000
<i>Pinnularia Borealis</i>	X_{64}	0.000
<i>Pinnularia Braunii</i> Var <i>Amphicephala</i>	X_{65}	0.000
<i>Pinnularia Maior</i> Var <i>Transerve</i>	X_{66}	0.000
<i>Pinnularia Mesolepta</i>	X_{67}	0.452
<i>Pinnularia Microstauron</i>	X_{68}	0.000

Pinnularia Ruttneri	X_{69}	0.000
Pinnularia Species 1	X_{70}	0.000
Pinnularia Splendida	X_{71}	0.074
Pinnularia Viridis	X_{72}	0.001
Psammothidium Bioretii	X_{73}	0.000
Stauroneis Obtusa	X_{74}	0.000
Surirella Linearis	X_{75}	0.000
Surirella Parma Soveriegne	X_{76}	0.000
Surirella Species 1	X_{77}	0.000
Surirella Tenuissima	X_{78}	0.000

4.2.2 Ujian Multivariat

Parameter-parameter yang memberikan impak yang tinggi daripada keputusan ujian multivariat akan diuji menggunakan ujian multivariat bagi menentukan sama ada terdapat hubungan terbentuk di antara satu pembolehubah bersandar dengan pembolehubah bersandar yang lain. Terdapat hanya 74 pembolehubah tak bersandar yang memberikan impak yang tinggi berdasarkan keputusan ujian univariat. Jadual 4.2 menunjukkan keputusan yang diberikan apabila ujian multivariat dijalankan dan keputusan ujian menunjukkan terdapat hubungan di antara kumpulan-kumpulan pembolehubah bersandar berdasarkan taburan nilai-p yang diperolehi oleh setiap parameter diskriminan. Di dapati hanya 21 parameter sahaja mampu membentuk hubungan di antara kumpulan-kumpulan pembolehubah bersandar serta memberikan impak yang tinggi kepada proses pengelasan manakala parameter yang selebihnya tidak mampu membentuk hubungan di antara kumpulan pembolehubah bersandar dan memberikan impak yang rendah kepada proses pengelasan. Jadual 4.2 menunjukkan bahawa parameter yang memberikan impak tinggi dengan taburan nilai-p adalah kurang daripada 5% manakala parameter selain daripada parameter berkenaan memberikan impak yang rendah dengan taburan nilai-p yang lebih besar daripada 5%.

Daripada keputusan yang diperolehi, hipotesis nol diterima pada parameter yang dinyatakan kerana mampu membentuk hubungan di antara kumpulan-kumpulan pembolehubah bersandar serta memberikan impak yang tinggi kepada proses pengelasan. Hipotesis nol ditolak untuk parameter-parameter selain yang dinyatakan sebelum ini kerana tidak mampu membentuk hubungan di antara setiap kumpulan pembolehubah bersandar serta memberikan impak yang rendah kepada proses pengelasan, dengan taburan nilai-p melebihi 5%..

Jadual 4.2: Keputusan ujian multivariat bagi keseluruhan parameter diskriminan.

Komposisi Alga	Pengenalan	Taburan Nilai-p
Acnanthes Exigua	X_1	0.071
Acnanthes Exigua Var Heterovalva	X_2	0.631
Acnanthes Minutissima	X_3	0.002
Acnanthes Oblongela	X_4	0.076
Amphora Libica	X_5	0.836
Anomoeneis Brachysira	X_6	0.000
Coccconeis Pediculus	X_7	0.110
Coccconeis Placetula	X_8	0.000
Coccconeis Species	X_9	0.002
Coccconeis Thumensis	X_{10}	0.132
Coscinodiscus Antiquus	X_{11}	0.348
Coscinodiscus Argus	X_{12}	0.101
Coscinodiscus Decipiens	X_{13}	0.912
Coscinodiscus Excentricus	X_{14}	0.307
Coscinodiscus Granii	X_{15}	0.000
Cymbella Species	X_{16}	0.721
Cymbella Inaequalis	X_{17}	0.000
Cyclotella Comta	X_{18}	0.001
Cyclotella Striata	X_{19}	0.059
Cyclotella Striata Var Baltica	X_{20}	0.063
Diatoma Elongatum	X_{21}	0.052
Diatoma Species	X_{22}	0.000
Diploneis Bombus	X_{23}	0.396
Diploneis Decipiens	X_{24}	0.018
Diploneis Elliptica	X_{25}	0.070

Diploneis Interrupa	X_{26}	0.117
Eutonia Faba	X_{27}	0.732
Eutonia Grunowi	X_{28}	0.193
Eutonia Monodon Var Alpina	X_{29}	0.814
Eutonia Pectinalis	X_{30}	0.020
Eutonia Sadetica Var Inasa	X_{31}	0.067
Eutonia Pectinalis	X_{32}	0.318
Eutonia Pectinalis Var Minor	X_{33}	0.901
Flagilaria Capucina	X_{34}	0.467
Flagilaria Species 1	X_{35}	0.819
Flagilaria Species 2	X_{36}	0.119
Flagilaria Crotonensis	X_{37}	0.010
Frustulia Rhomboides Var Saxonica	X_{39}	0.057
Frustulia Saxonica	X_{40}	0.047
Gomphonema Acuminatum	X_{41}	0.001
Gomphonema Gracile	X_{42}	0.613
Gomphonema Longiceps	X_{43}	0.402
Gomphonema Parvalum	X_{44}	0.287
Gomphonema Subventricosum	X_{45}	0.665
Hantzschia Ampioxys	X_{46}	0.006
Navicula Cryptocephala	X_{47}	0.389
Navicula Cuspidata	X_{48}	0.192
Navicula Halophila	X_{49}	0.557
Navicula Hustedtii	X_{50}	0.919
Navicula Obtusa	X_{51}	0.787
Navicula Obtusa Var Scalpelliformis	X_{52}	0.361
Navicula Radiosa	X_{53}	0.000
Navicula Sigma	X_{54}	0.627
Navicula Species	X_{55}	0.656
Neidium Affine	X_{56}	0.005
Nitzschia Amphibia	X_{58}	0.015
Nitzschia Fanticola	X_{59}	0.115
Nitzschia Littoralis	X_{60}	0.150
Nitzschia Palea	X_{61}	0.080
Pinnularia Biceps	X_{62}	0.280
Pinnularia Biceps F. Petersenii	X_{63}	0.075
Pinnularia Borealis	X_{64}	0.637
Pinnularia Braunii Var Amphicephala	X_{65}	0.114

Pinnularia Maior Var Transerve	X_{66}	0.070
Pinnularia Microstauron	X_{68}	0.452
Pinnularia Ruttneri	X_{69}	0.648
Pinnularia Species 1	X_{70}	0.167
Pinnularia Viridis	X_{72}	0.001
Psammothidium Bioretii	X_{73}	0.000
Stauroneis Obtusa	X_{74}	0.000
Surirella Linearis	X_{75}	0.052
Surirella Parma Soveriegn	X_{76}	0.000
Surirella Species 1	X_{77}	0.516
Surirella Tenuissima	X_{78}	0.097

4.2.3 Fungsi Diskriminan

Berdasarkan ujian multivariat, didapati hanya 21 parameter mampu membentuk hubungan di antara setiap pembolehubah bersandar serta memberikan impak yang tinggi kepada proses pengelasan. Pembentukan fungsi diskriminan dilakukan dengan memberikan pekali diskriminan kepada setiap parameter yang memberikan impak yang tinggi. Jadual 4.3 menunjukkan pembolehubah-pembolehubah tak bersandar yang berimpak tinggi bersama pekali diskriminan yang diberikan semasa fungsi diskriminan terbentuk. Berdasarkan keputusan ujian multivariat dan pembentukan fungsi diskriminan, diketahui bahawa penggunaan 21 parameter atau spesis alga yang dominan ini sudah memadai untuk menentukan jenis dan tahap kualiti air sungai dengan baik.

Jadual 4.3: Pembolehubah tak bersandar yang dominan dan pekali diskriminan.

Komposisi Alga	Pengenalan	Pekali
Acnanthes Minutissima	X_3	-0.232
Anomeoneis Brachysira	X_6	0.970
Coccconeis Placentula	X_8	-0.203
Coccconeis Species	X_9	-0.356
Coscinodiscus Granii	X_{15}	-0.379

Cyclotella Comta	X_{17}	-1.832
Cymbella Inaequalis	X_{18}	1.686
Diatoma Species	X_{22}	2.140
Diploneis Decipiens	X_{24}	3.128
Eutonia Pectinalis	X_{30}	1.569
Fragilaria Crotonensis	X_{37}	-0.286
Frustulia Saxonica	X_{40}	10.025
Ghomponema Acuminatum	X_{41}	1.624
Hantzschia Amphioxys	X_{46}	-0.346
Navicula Radiosa	X_{53}	0.716
Neidium Affine	X_{56}	11.928
Hitzschia Amphibia	X_{58}	-2.660
Pinnularia Viridis	X_{72}	-0.207
Psammothidium Bioretii	X_{73}	-0.218
Stauroneis Obtusa	X_{74}	-0.105
Surirella Parma Sovereign	X_{76}	-0.100
Pemalar		-1.447

Daripada Jadual 4.3, fungsi diskriminan disusun menjadi seperti berikut:

$$\begin{aligned}
 Z_p = & -0.232 X_3 + 0.970 X_6 - 0.203 X_8 - 0.356 X_9 - 0.379 X_{15} - 1.832 X_{17} \\
 & + 1.686 X_{18} + 2.140 X_{22} + 3.128 X_{24} + 1.569 X_{30} - 0.286 X_{37} + 10.025 X_{40} \\
 & + 1.624 X_{41} - 0.346 X_{46} + 0.716 X_{53} + 11.928 X_{56} - 2.660 X_{57} - 0.207 X_{72} \\
 & - 0.218 X_{73} - 0.105 X_{74} - 0.100 X_{76} - 1.447
 \end{aligned} \quad (4.3)$$

Dalam kajian ini, masukan data yang diberikan adalah dalam bentuk '0' atau '1' sahaja. Oleh itu, nilai fungsi diskriminan akan diketahui dengan menggantikan pembolehubah tak bersandar dengan '0' sekiranya parameter atau alga berkenaan tidak wujud di dalam sampel air sungai dan '1' jika sebaliknya. Nilai fungsi diskriminan akan diperolehi dengan menjumlahkan hasil darab pekali diskriminan dengan pembolehubah tak bersandar iaitu spesis-spesis alga yang dominan. Jenis dan tahap kualiti air sungai dapat ditentukan dengan memeriksa julat penentuan titik pemisah.

4.2.4 Penentuan Titik Pemisah

Titik pemisah perlu ditentukan supaya nilai tersebut dapat digunakan untuk menentukan julat kelas bagi jenis dan tahap kualiti air sungai. Jadual 4.4 menunjukkan titik-titik tengah bagi setiap pembolehubah bersandar:

Jadual 4.4: Titik tengah bagi kumpulan-kumpulan kelas.

Kumpulan	Titik Tengah Kumpulan
1	-1.868
2	-1.274
3	14.453
4	0.777
5	-1.589

Z_1 adalah mewakili titik pusat bagi kumpulan satu, Z_2 adalah mewakili titik pusat bagi kumpulan dua, Z_3 adalah mewakili titik pusat bagi kumpulan tiga, Z_4 adalah mewakili titik pusat bagi kumpulan empat dan Z_5 adalah mewakili titik pusat bagi kumpulan lima. Dengan susunan menaik, didapati bahawa $Z_1 < Z_5 < Z_2 < Z_4 < Z_3$. Titik pemisah di antara satu kumpulan dengan kumpulan yang lain ditentukan seperti berikut.

Bagi titik pemisah kumpulan satu dan kumpulan lima adalah:

$$Z_1 | Z_5 = \left(\frac{(-1.868) + (-1.589)}{2} \right) = -1.729$$

Titik pemisahan bagi kumpulan lima dengan kumpulan dua pula adalah:

$$Z_5 | Z_2 = \left(\frac{(-1.589) + (-1.274)}{2} \right) = -1.432$$

Kumpulan dua dan kumpulan empat dipisahkan pada titik:

$$Z_2 | Z_4 = \left(\frac{(-1.274) + 0.777}{2} \right) = -0.249$$

Titik pemisahan bagi kumpulan empat dan kumpulan tiga pula adalah:

$$Z_4 | Z_3 = \left(\frac{0.777 + 14.453}{2} \right) = 7.615$$

Daripada titik-titik pemisahan yang diperolehi, julat bagi setiap kumpulan akan ditetapkan. Kumpulan satu terletak pada titik pemisah kurang daripada -1.729 manakala kumpulan lima terletak pada titik pemisah di antara -1.729 dan -1.432. Kumpulan dua terletak di antara -1.432 dan -0.249. Julat antara -0.249 dan 7.615 pula mewakili kumpulan empat sementara kumpulan tiga terletak pada titik pemisah yang lebih besar daripada 7.615.

4.2.5 Prestasi Pegkelasan Data

Selain daripada menentukan pembolehubah yang memberikan impak yang tinggi, ujian univariat dan ujian multivariat juga mampu mengkelaskan data-data kepada kumpulan-kumpulan yang ditetapkan. Analisis diskriminan yang dijalankan kepada set-set data spesis alga dan telah memberikan keputusan pengkelasan seperti di dalam Jadual 4.5.

Jadual 4.5: Keputusan Pengkelasan.

Kumpulan		Kumpulan Ramalan					Jumlah
		FC	FM	FP	BM	BP	
Kumpulan	FC	41	6	8	0	0	55
	FM	0	55	6	1	1	63
	FP	0	20	68	3	0	91
	BM	0	1	1	46	0	48
	BP	0	2	6	0	55	63
Peratusan, %	FC	74.5	10.9	14.5	0	0	100
	FM	0	87.3	9.5	1.6	1.6	100
	FP	0	22.0	74.7	3.3	0	100
	BM	0	2.1	2.1	95.8	0	100
	BP	0	9.5	9.5	0	87.3	100

Jadual 4.5 menunjukkan peratus kejituuan pengkelasan yang dilakukan oleh analisis diskriminan. Analisis diskriminan mampu memberikan kejituuan pengkelasan sebanyak 74.50% jenis air tawar dengan tahap kualiti bersih (diwakili oleh FC). Bagi analisis diskriminan pengkelaskan jenis air tawar dengan tahap kualiti pertengahan (diwakili oleh FM) pula adalah 87.30%. Kumpulan jenis air tawar dengan tahap kualiti

tercemar (diwakili oleh FP) dikelaskan dengan tepat sebanyak 74.70%. Bagi kumpulan jenis air payau dengan tahap kualiti pertengahan (diwakili oleh BM) dan jenis air payau dengan tahap kualiti tercemar (diwakili oleh BP) masing-masing data dikelaskan dengan tepat sebanyak 95.80% dan 87.30%. Prestasi ramalan keseluruhan yang diberikan oleh analisis diskriminan adalah 82.80%. Keputusan ini menunjukkan bahawa di samping menentukan pembolehubah-pembolehubah tak bersandar yang memberikan impak yang tinggi, analisis diskriminan ini juga mampu melakukan pengkelasan data dengan baik.

4.3 Keputusan Analisis Rangkaian Neural

Terdapat 78 spesis alga yang telah dicerap di sepanjang lembangan Sungai Pinang. Jika kesemua spesis alga digunakan sebagai parameter masukan kepada rangkaian neural, rangkaian yang dibangunkan menjadi sangat kompleks di mana rangkaian neural memerlukan 78 nod masukan. Berdasarkan keputusan yang diberikan oleh ujian univariat dan ujian multivariat, didapati bahawa terdapat 21 spesis alga memberikan impak yang tinggi atau spesis yang dominan.

Kesemua rangkaian neural ditetapkan bilangan lelaran data latihan dan bilangan nod tersembunyi optimum bagi memberikan kejituuan pengkelasan yang maksimum. Di samping itu juga, nilai ralat kuasa dua purata (*Mean Square Error, MSE*) juga ditentukan. Analisis bagi mendapatkan bilangan lelaran data latihan yang optimum ditentukan terlebih dahulu dengan menetapkan bilangan nod tersembunyi sebanyak 10. Bilangan lelaran data latihan yang optimum ditentukan dengan menambah lelaran data latihan satu persatu sehingga prestasi kejituuan yang maksimum dan nilai MSE yang minimum diperolehi. Setelah ditetapkan lelaran data latihan, bilangan nod tersembunyi akan ditambah satu persatu sehingga prestasi kejituuan yang optimum diperolehi. Analisis-analisis dilakukan kepada rangkaian RBF dan MLP konvensional, rangkaian neural hibrid dan rangkaian neural berhirarki.

Berdasarkan keputusan-keputusan yang diperolehi, bilangan lelaran data latihan dan bilangan nod tersembunyi yang optimum serta nilai MSE terkecil bagi setiap rangkaian neural dipaparkan di dalam Jadual 4.6. Perbandingan prestasi kejituhan bagi mengkelaskan jenis dan tahap kualiti air sungai pula ditunjukkan di dalam Jadual 4.7.

Jadual 4.6: Perbandingan bilangan lelaran data latihan dan bilangan nod tersembunyi yang optimum serta nilai MSE di antara rangkaian RBF, MLP dilatih menggunakan algoritma latihan BP, LM dan BR, HMLP, HiMLP dilatih menggunakan algoritma latihan BP, LM dan BR serta H^2 MLP.

Rangkaian	Bilangan Lelaran Data Latihan	Bilangan Nod Tersembunyi	MSE pada Fasa Latihan	MSE pada Fasa Ujian
RBF	1	60	0.0282	0.8961
MLP dengan BP	83	48	0.5685	0.6180
MLP dengan LM	33	6	0.1606	0.1581
MLP dengan BR	36	7	0.1681	0.1708
HMLP	8	4	0.1445	0.1545
HiMLP dengan BP	16	11	0.1477	0.1585
HiMLP dengan LM	16	5	0.1420	0.1511
HiMLP dengan BR	19	7	0.1593	0.1621
H^2 MLP	5	2	0.0085	0.0096

Jadual 4.7: Perbandingan prestasi kejituuan ramalan di antara rangkaian RBF, MLP dilatih menggunakan algoritma latihan BP, LM dan BR, HMLP, HiMLP dilatih menggunakan algoritma latihan BP, LM dan BR serta H^2 MLP.

Rangkaian	Fasa Latihan, %	Fasa Ujian, %	Keseluruhan, %
RBF	90.00	65.00	77.50
MLP dengan BP	41.00	35.00	38.00
MLP dengan LM	89.00	73.00	81.00

MLP dengan BR	74.00	70.00	72.00
HMLP	95.16	89.40	92.28
HiMLP dengan BP	72.67	61.33	67.00
HiMLP dengan LM	89.67	87.00	88.34
HiMLP dengan BR	89.00	84.00	86.50
H^2 MLP	97.13	97.09	97.11
DA	-	-	82.80

Jadual 4.6 menunjukkan rangkaian RBF hanya memerlukan satu lelaran data latihan sahaja untuk mencapai prestasi optimum tetapi bilangan nod tersembunyi yang tinggi iaitu sebanyak 60 diperlukan. Nilai MSE yang diperolehi adalah sebanyak 0.0282 fasa latihan dan 0.8961 pada fasa ujian. Jadual 4.7 menunjukkan rangkaian RBF hanya berupaya mengkelaskan data-data ujian sehingga 65.00% sahaja. Keputusan ini menunjukkan rangkaian RBF mempunyai keupayaan pengitlakan yang rendah kerana memerlukan bilangan nod tersembunyi yang tinggi walaupun kadar penumpuan yang pantas ditunjukkan.

Jadual 4.6 dan 4.7 menunjukkan rangkaian MLP yang dilatih menggunakan BP, LM dan BR masing-masing mencapai prestasi kejituuan optimum pada lelaran data ke 83, 33 dan 36 dan nod tersembunyi ke 48, 6 dan 7. Pada masa ini, keupayaan pengkelasan bagi setiap algoritma latihan BP, LM dan BR masing-masing sebanyak 35.00%, 73.00% dan 70.00%. Ralat pengkelasan yang diberikan oleh rangkaian MLP yang dilatih menggunakan BP, LM dan BR pula masing-masing sebanyak 0.5685, 0.1606 dan 0.1681 pada fasa latihan dan 0.6180, 0.1581 dan 0.1708 pada fasa ujian.

Daripada Jadual 4.7, didapati rangkaian MLP yang dilatih menggunakan algoritma latihan LM mampu melakukan pengkelasan lebih baik berbanding rangkaian MLP yang dilatih oleh algoritma latihan BR dan BP dengan perbezaan keupayaan pengkelasan masing-masing adalah 38.00% dan 3.00%.

MLP dengan BR	74.00	70.00	72.00
HMLP	95.16	89.40	92.28
HiMLP dengan BP	72.67	61.33	67.00
HiMLP dengan LM	89.67	87.00	88.34
HiMLP dengan BR	89.00	84.00	86.50
H ² MLP	97.13	97.09	97.11
DA	-	-	82.80

Jadual 4.6 menunjukkan rangkaian RBF hanya memerlukan satu lelaran data latihan sahaja untuk mencapai optimum tetapi bilangan nod tersembunyi yang tinggi iaitu sebanyak 60 diperlukan. Nilai MSE yang diperolehi adalah sebanyak 0.0282 fasa latihan dan 0.8961 pada fasa ujian. Jadual 4.7 menunjukkan rangkaian RBF hanya berupaya mengkelaskan data-data ujian sehingga 65.00% sahaja. Keputusan ini menunjukkan rangkaian RBF mempunyai keupayaan pengitikan yang rendah kerana memerlukan bilangan nod tersembunyi yang tinggi walaupun kadar penumpuan yang pantas ditunjukkan.

Jadual 4.6 dan 4.7 menunjukkan rangkaian MLP yang dilatih menggunakan BP, LM dan BR masing-masing mencapai prestasi kejituhan optimum pada lelaran data ke 83, 33 dan 36 dan nod tersembunyi ke 48, 6 dan 7. Pada masa ini, keupayaan pengkelasan bagi setiap algoritma latihan BP, LM dan BR masing-masing sebanyak 35.00%, 73.00% dan 70.00%. Ralat pengkelasan yang diberikan oleh rangkaian MLP yang dilatih menggunakan BP, LM dan BR pula masing-masing sebanyak 0.5685, 0.1606 dan 0.1681 pada fasa latihan dan 0.6180, 0.1581 dan 0.1708 pada fasa ujian. Daripada Jadual 4.7, didapati rangkaian MLP yang dilatih menggunakan algoritma latihan LM mampu melakukan pengkelasan lebih baik berbanding rangkaian MLP yang dilatih oleh algoritma latihan BR dan BP dengan perbezaan keupayaan pengkelasan masing-masing adalah 38.00% dan 3.00%.

Rangkaian MLP dengan menggunakan algoritma latihan LM hanya memerlukan 33 lelaran data sahaja untuk menempuh dengan baik berbanding rangkaian MLP yang dilatih menggunakan algoritma BP dan BR. Rangkaian MLP dengan menggunakan algoritma latihan LM juga memberikan struktur rangkaian yang ringkas dengan nod tersembunyi yang diperlukan sebanyak 6, lebih baik berbanding bilangan nod tersembunyi yang diperlukan oleh rangkaian MLP yang menggunakan algoritma latihan BR dan algoritma latihan BP. Rangkaian MLP yang menggunakan algoritma latihan LM juga mampu memberikan ralat pengkelasan yang lebih kecil berbanding rangkaian MLP yang menggunakan algoritma latihan BP dan BR dengan perbezaan masing-masing sebanyak 0.4079 dan 0.0075 fasa latihan manakala 0.4599 dan 0.0199 pada fasa ujian.

Jadual 4.6 menunjukkan rangkaian HMLP hanya memerlukan 8 lelaran data dan 4 nod tersembunyi sahaja bagi mencapai prestasi kejituhan yang optimum berbanding rangkaian MLP terbaik (iaitu yang dilatih menggunakan algoritma LM) dan rangkaian RBF yang mana masing-masing memerlukan 33 dan 1 lelaran data latihan dan 6 dan 60 nod tersembunyi bagi mencapai prestasi yang optimum. Keputusan yang diperolehi menunjukkan rangkaian HMLP mampu menempuh dengan pantas di samping rangkaian yang ringkas berjaya dibentuk berbanding rangkaian MLP yang dilatih oleh algoritma LM dan rangkaian RBF. Jadual 4.8 pula menunjukkan keupayaan pengkelasan yang diberikan oleh rangkaian HMLP adalah 89.40%, iaitu 16.40% lebih baik berbanding rangkaian MLP dan 24.40% lebih baik berbanding rangkaian RBF. Walaubagaimanapun, ralat pengkelasan yang diberikan oleh rangkaian HMLP adalah lebih besar berbanding rangkaian MLP yang dilatih menggunakan algoritma LM dan rangkaian RBF dengan perbezaan prestasi MSE masing-masing adalah 0.0039 dan 0.1163 pada fasa latihan tetapi lebih baik berbanding rangkaian MLP yang dilatih menggunakan algoritma LM dan rangkaian RBF masing-masing sebanyak 0.0036 dan 0.7416 pada fasa ujian.

Jadual 4.6 dan 4.7 menunjukkan rangkaian HiMLP dengan menggunakan algoritma latihan BP mampu memberikan keupayaan pengkelasany sebanyak 61.33% pada lelaran data latihan ke 16 dan nod tersembunyi sebanyak 11 berbanding rangkaian MLP menggunakan algoritma latihan BP pada lelaran data latihan ke 83 dan nod tersembunyi sebanyak 48 dengan perbezaan prestasi pengkelasany sebanyak 26.33%. Rangkaian HiMLP yang dilatih oleh algoritma latihan LM berjaya memberikan pengkelasany yang optimum pada lelaran data latihan ke 16 dan nod tersembunyi sebanyak 5 dengan keupayaan pengkelasany sebanyak 87.00%, iaitu 14.00% lebih baik berbanding rangkaian MLP yang dilatih menggunakan algoritma latihan yang sama. Keupayaan pengkelasany sebanyak 84.00% pula ditunjukkan oleh rangkaian HiMLP yang dilatih menggunakan algoritma latihan BR pada lelaran data latihan ke 19 dan nod tersembunyi sebanyak 7. Keputusan ini menunjukkan rangkaian HiMLP ini berupaya melakukan pengkelasany dengan lebih baik daripada rangkaian MLP yang menggunakan algoritma latihan yang sama dengan perbezaan prestasi pengkelasany sebanyak 14.00%. Selain itu rangkaian HiMLP juga menghasilkan ralat pengkelasany yang lebih kecil berbanding rangkaian MLP. Rangkaian HiMLP yang menggunakan algoritma latihan BP, LM dan BR masing-masing memberikan nilai MSE sebanyak 0.1477, 0.1420 dan 0.1593 pada fasa latihan iaitu 0.4208, 0.0186 dan 0.0088 lebih baik berbanding rangkaian MLP yang dilatih menggunakan algoritma BP, LM dan BR. Nilai MSE yang diperolehi pada fasa ujian bagi rangkaian HiMLP yang dilatih menggunakan algoritma BP, LM dan BR pula masing-masing adalah sebanyak 0.1585, 0.1511 dan 0.1621, iaitu 0.4100, 0.0070 dan 0.0087 lebih baik berbanding rangkaian MLP yang dilatih menggunakan algoritma BP, LM dan BR.

Walaupun pengubahsuai yang dilakukan kepada rangkaian MLP dengan menghirarkikan rangkaian tersebut mampu memberikan prestasi kejituany yang tinggi, tetapi pengubahsuai yang dilakukan masih belum mampu menandingi prestasi keupayaan pengkelasany yang dilakukan oleh rangkaian HMLP. Rangkaian HiMLP

terbaik (iaitu yang dilatih menggunakan algoritma LM) memberikan keupayaan pengkelasany lebih kecil iaitu perbezaan sebanyak 2.40% berbanding rangkaian HMLP. Walaubagaimanapun, keupayaan pengkelasany rangkaian HiMLP ini adalah lebih baik daripada keupayaan pengkelasany yang diberikan oleh rangkaian RBF dengan perbezaan sebanyak 32.00%.

Jadual 4.6 menunjukkan rangkaian H^2 MLP hanya memerlukan 5 lelaran data latihan dan nod tersembunyi sebanyak 2 sahaja bagi mencapai prestasi pengkelasany sebanyak 97.09% seperti yang ditunjukkan di dalam jadual 4.7. Rangkaian ini mampu mencapai kejituany pengkelasany yang tinggi kerana pengkelasany dilakukan secara berhirarki di samping sambungan hibrid ditambah ke dalam rangkaian ini. Ini bermakna rangkaian HMLP tidak lagi melakukan pengkelasany data-data secara satu persatu tetapi mengkelaskan data secara berhirarki. Rangkaian H^2 MLP berupaya melakukan pengkelasany dengan lebih baik berbanding rangkaian HMLP. Rangkaian H^2 MLP menumpu pada lelaran data latihan ke 5, lebih pantas berbanding rangkaian HMLP yang memerlukan 8 lelaran data latihan. Struktur rangkaian H^2 MLP juga lebih ringkas dengan memerlukan 2 nod tersembunyi sahaja berbanding 4 nod tersembunyi pada rangkaian HMLP. Perbezaan prestasi pengkelasany yang diberikan oleh rangkaian H^2 MLP adalah 7.69% lebih baik berbanding rangkaian HMLP. Selain itu, rangkaian H^2 MLP juga mampu melakukan pengkelasany dengan lebih tepat di mana nilai MSE yang diberikan adalah 0.0085 pada fasa latihan dan 0.0096 pada fasa ujian.

Rangkaian H^2 MLP ini jelas mampu mengurangkan saiz rangkaian selain mampu memberikan prestasi pengkelasany yang terbaik di antara rangkaian-rangkaian RBF, MLP, HiMLP dan HMLP. Rangkaian H^2 MLP memberikan struktur teringkas iaitu hanya memerlukan lima lelaran data latihan dan dua nod tersembunyi sahaja. Rangkaian RBF, MLP, HiMLP dan HMLP masing-masing menumpu pada lelaran data latihan ke 1, 33, 16 dan 8 manakala bilangan nod tersembunyi yang diperlukan adalah 60, 6, 5 dan 4. Perbezaan kejituany pengkelasany yang ditunjukkan oleh rangkaian

H^2MLP berbanding rangkaian RBF, MLP dan HiMLP yang dilatih oleh algoritma LM serta HMLP masing-masing adalah sebanyak 32.09%, 24.09%, 10.09% dan 7.69%. Rangkaian H^2MLP juga mampu mengkelaskan data-data ujian dengan tepat di mana nilai MSE adalah 0.0085 pada fasa latihan dan 0.0096 pada fasa ujian. Nilai MSE yang diberikan oleh rangkaian H^2MLP adalah lebih baik berbanding rangkaian RBF, MLP dan HiMLP yang dilatih oleh algoritma LM serta HMLP dengan perbezaan sebanyak 0.0184, 0.1511, 0.1349 dan 0.1324 pada fasa latihan manakala 0.8870, 0.1490, 0.1420 dan 0.1454 pada fasa ujian.

Di samping itu juga, Jadual 4.7 juga menunjukkan keupayaan yang diberikan oleh analisis diskriminan di dalam mengkelaskan data-data ujian. Prestasi yang diberikan oleh analisis diskriminan adalah sebanyak 82.80% di mana prestasi yang diberikan ini adalah lebih baik berbanding kejituhan pengkelasan yang diberikan oleh rangkaian RBF dan MLP tetapi kurang baik berbanding rangkaian HiMLP, HMLP dan H^2MLP . Perbezaan prestasi pengitlakan yang diberikan oleh analisis diskriminan terhadap rangkaian RBF dan rangkaian MLP yang dilatih menggunakan algoritma LM masing-masing adalah 17.80% dan 9.80%. Perbezaan prestasi pengkelasan yang diberikan oleh rangkaian HiMLP yang dilatih menggunakan algoritma LM, HMLP dan H^2MLP terhadap analisis diskriminan pula masing-masing adalah 4.20%, 6.60% dan 14.29%.

Secara keseluruhan, didapati bahawa rangkaian H^2MLP mencapai prestasi kejituhan optimum sebanyak 97.09%, iaitu prestasi kejituhan yang terbaik di antara kesemua rangkaian neural yang digunakan. Begitu juga dengan struktur rangkaian yang dibangunkan, rangkaian H^2MLP mampu menumpu dengan pantas pada 5 lelaran data latihan sahaja dan mampu membentuk rangkaian yang ringkas dengan hanya memerlukan 2 nod tersembunyi sahaja. Ralat pengkelasan yang diberikan oleh rangkaian H^2MLP adalah paling minimum berbanding rangkaian neural yang lain sebanyak 0.0085 pada fasa latihan manakala 0.0096 pada fasa ujian. Berdasarkan

prestasi kejituhan pengkelasan yang dilakukan, rangkaian H^2MLP mempunyai kemampuan untuk melakukan pengkelasan yang tepat dengan kejituhan pengkelasan yang tinggi. Oleh itu, rangkaian H^2MLP dipilih untuk digunakan ke dalam sistem penentuan jenis dan tahap kualiti air sungai yang dibina.

4.4 Kesimpulan

Keputusan yang diperolehi menunjukkan analisis diskriminan dengan ujian univariat dan ujian multivariat yang dijalankan telah berjaya mengurangkan penggunaan 78 spesis alga kepada 21 jenis sahaja. Dengan penggunaan 21 spesis alga sahaja, rangkaian neural yang dibangunkan menjadi lebih ringkas berbanding penggunaan kesemua 78 spesis alga sebagai parameter masukan yang akan menjadikan rangkaian neural terlalu kompleks.

Berdasarkan kepada keputusan lain yang diperolehi, didapati rangkaian RBF dan MLP kurang sesuai untuk digunakan dalam penentuan jenis dan kualiti air sungai. Prestasi rangkaian RBF bergantung kepada penempatan pusat-pusat rangkaian manakala rangkaian MLP menumpu dengan perlahan. Penghirarkian yang dilakukan kepada rangkaian MLP mampu membentuk rangkaian HiMLP yang berupaya mengitlik dengan lebih baik lagi berbanding rangkaian MLP. Selain melakukan pengkelasan secara berhirarki, rangkaian MLP juga diubahsuai dengan menambah sambungan lelurus menghubungkan lapisan masukan rangkaian terus ke lapisan keluaran tanpa melalui lapisan tersembunyi. Hasilnya, satu rangkaian HMLP berjaya dibentuk serta memberikan keupayaan pengkelasan yang lebih baik berbanding rangkaian RBF, MLP dan HiMLP. Satu keputusan yang lebih baik diperolehi berbanding keputusan yang diberikan oleh rangkaian RBF, MLP, HiMLP dan HMLP apabila pengkelasan secara berhirarki yang dilakukan kepada rangkaian HMLP. Rangkaian yang dikenali sebagai rangkaian H^2MLP berjaya mengurangkan saiz rangkaian di samping berupaya mengitlik sebanyak 97.09%.

Keputusan yang diperolehi di dalam Bahagian 4.2 juga telah membuktikan bahawa analisis diskriminan berjaya mengurangkan penggunaan 78 jenis alga kepada 21 jenis sahaja dan penggunaan spesis ini sudah mencukupi untuk memberikan keupayaan pengitlakan yang baik kepada rangkaian H²MLP bagi menentukan jenis dan kualiti air sungai. Analisis diskriminan ini juga memberikan kejituhan pengkelasannya sebanyak 82.80%. Keputusan yang diperolehi juga telah menunjukkan bahawa objektif penyelidikan ini untuk menentukan jenis dan kualiti air sungai telah tercapai. Selain itu, satu sistem bagi menentukan jenis dan kualiti air sungai telah berjaya dibangunkan dengan menggunakan rangkaian H²MLP sebagai alat pengkelasannya. Sistem yang dibangunkan ini adalah mesra pengguna dan mampu memberikan kejituhan pengkelasannya yang tinggi.

Secara keseluruhannya, penggunaan spesis alga ke dalam rangkaian neural jelas menunjukkan bahawa spesis-spesis alga sesuai dijadikan sebagai parameter masukan. Berdasarkan kepada keputusan yang diberikan, penggunaan rangkaian neural ke dalam sistem pintar telah berjaya memberikan kejituhan pengkelasannya yang baik. Di samping itu juga, penghirarkian yang dilakukan kepada rangkaian neural bagi mengkelaskan data-data juga telah berjaya memberikan prestasi kejituhan yang tinggi.

BAB 5

KESIMPULAN

5.1 Kesimpulan

Di dalam penyelidikan ini, satu sistem penentuan jenis dan kualiti air sungai berdasarkan komposisi alga menggunakan rangkaian neural telah berjaya dibangunkan. Sistem penentuan kualiti air sungai yang dibangunkan ini menggunakan spesis alga yang terdapat di dalam air sungai sebagai masukan kepada sistem. Sistem ini berkemampuan untuk melakukan pengkelasannya kualiti air sungai dengan kejituhan yang tinggi. Sistem penentuan kualiti air sungai mampu mengkelaskan air sungai kepada jenis air terlebih dahulu kemudian jenis air sungai ini akan dikelaskan kepada tahap-tahap kualiti air. Pembangunan sistem ini dapat membantu memudahkan bidang tugas ahli-ahli limnologi dalam menentukan tahap kualiti air sungai.

Penyelidikan ini mencadangkan penggunaan analisis diskriminan bagi mengenalpasti spesis-spesis alga yang dominan iaitu memberikan impak yang tinggi dalam menentukan jenis dan kualiti air sungai. Penggunaan analisis diskriminan dalam penyelidikan ini amat penting kerana keputusan yang diberikannya akan digunakan semula kepada rangkaian neural sebagai parameter masukan dalam mengkelaskan air sungai kepada jenis dan tahap kualitinya. Penyelidikan ini mencadangkan penggunaan 78 jenis alga sebagai parameter masukan kepada rangkaian tetapi penggunaan kesemua spesis-spesis alga ini adalah terlalu banyak dan akan menyebabkan rangkaian neural menjadi terlalu kompleks. Perkara ini dapat diatasi dengan aplikasi analisis diskriminan dalam mengenalpasti spesis alga yang dominan yang akan digunakan sebagai parameter masukan kepada rangkaian neural. Keputusan yang diperolehi daripada analisis diskriminan menunjukkan bahawa 21 jenis alga iaitu *Acnanthes Minutissima*, *Anomeoneis Brachysira*, *Coccconeis Placentula*, *Coccconeis Species*, *Coscinodiscus Granii*, *Cyclotella Comta*, *Cymbella Inaequalis*, *Diatoma Species*, *Diploneis Decipiens*, *Eutonia Pectinalis*, *Fragilaria Crotonensis*, *Frustulia*

Saxonica, Ghomponema Acuminatum, Hantzschia Amphioxys, Navicula Radiosa, Neidium Affine, Hitzschia Amphibia, Pinnularia Viridis, Psammothidium Bioretii, Stauroneis Obtusa dan Suriella Parma Sovereign sudah memadai digunakan sebabai masukan kepada rangkaian neural.

Selain itu, penyelidikan ini telah mengambil satu langkah ke hadapan dengan mencadangkan penggunaan rangkaian neural bagi menentukan jenis air sungai terlebih dahulu kemudian disusuli oleh penentuan tahap kualiti air sungai. Berbanding dengan kebanyakan penyelidikan terdahulu, mereka hanya menentukan tahap kualiti air sungai sahaja tanpa menentukan jenis-jenis air sungai. Kajian ini telah menggunakan rangkaian fungsi asas jejarian (RBF) dan perseptron berbilang lapisan (MLP) yang dilatih oleh tiga jenis algoritma latihan iaitu algoritma perambatan belakang (BP), Lavenberg Marquardt (LM) dan Bayesian Regularization (BR) tetapi keputusan yang diberikan adalah kurang baik. Maka, penyelidikan ini mencadangkan penggunaan rangkaian perseptron berbilang lapisan hibrid (HMLP) bagi mendapatkan keputusan yang lebih baik. Rangkaian HMLP yang dibangunkan berjaya memberi keupayaan pengkelasan yang lebih baik berbanding rangkaian RBF dan MLP di samping ralat pengkelasan yang kecil. Bagi mengoptimumkan prestasi kejituhan pengkelasan, penyelidikan ini mencadangkan pengkelasan dilakukan secara berhirarki. Pengkelasan secara berhirarki ini dilakukan terhadap rangkaian MLP dan HMLP. Oleh itu, terdapat empat rangkaian neural baru yang dibangunkan iaitu rangkaian perseptron berbilang lapisan berhirarki (*Hierarchical Multilayered Perceptron*, *HiMLP*) yang dilatih menggunakan algoritma latihan BP, algoritma LM dan algoritma BR serta rangkaian perseptron berbilang lapisan hibrid berhirarki (H^2MLP).

Pengkelasan yang dilakukan secara berhirarki memberikan keputusan yang lebih baik berbanding pengkelasan yang dijalankan tanpa penghirarkian. Kesemua rangkaian HiMLP yang dibangunkan berjaya memberikan prestasi yang lebih baik berbanding rangkaian MLP. Di samping itu juga, saiz rangkaian bagi setiap rangkaian berjaya dikurangkan. Walaupun rangkaian yang dibentuk adalah ringkas tetapi

keupayaan pengkelasan adalah lebih tinggi berbanding rangkaian MLP dengan ralat pengkelasan yang diberikan oleh rangkaian HiMLP adalah lebih kecil berbanding rangkaian MLP. Keputusan yang sama juga diperolehi apabila rangkaian HMLP dihirarkikan bagi membentuk rangkaian H^2MLP . Saiz rangkaian H^2MLP berjaya dikurangkan serta keupayaan pengkelasan adalah semakin bertambah berbanding rangkaian HMLP. Penyelidikan ini memilih rangkaian H^2MLP untuk digunakan di dalam sistem penentuan jenis dan kualiti air sungai kerana rangkaian ini mampu mengkelaskan jenis dan kualiti air sungai dengan lebih baik berbanding rangkaian-rangkaian neural yang lain. Rangkaian H^2MLP ini juga mampu untuk menumpu dengan berkebolehan untuk belajar dengan pantas di samping ralat pengkelasan adalah paling minimum berbanding rangkaian-rangkaian lain.

Keputusan yang diperolehi di dalam Bahagian 4.2 hingga 4.4 di dalam Bab 4 telah membuktikan bahawa penggunaan 21 jenis spesis alga yang hadir di dalam air sungai yang digunakan sebagai data masukan kepada rangkaian neural sudah memadai bagi rangkaian H^2MLP menghasilkan prestasi kejituhan pengkelasan yang tinggi bagi menentukan jenis dan tahap kualiti air sungai. Walaupun rangkaian H^2MLP menggunakan 21 nod masukan bagi data masukan, namun rangkaian ini masih mampu mengekalkan struktur rangkaian yang ringkas dengan keupayaan pengitlakan yang tinggi. Rangkaian H^2MLP ini hanya memerlukan lima lelaran data sahaja untuk menumpu dengan baik serta mempunyai struktur rangkaian yang ringkas dengan memerlukan hanya satu nod tersembunyi bagi mencapai prestasi yang optimum. Keputusan yang diperolehi di dalam Bahagian 4.4 menunjukkan rangkaian H^2MLP mampu mengkelaskan sehingga 97.13%, 97.09% dan 97.11% masing-masing pada fasa latihan, fasa ujian dan keseluruhan dan ralat pengkelasan pula adalah xxx pada fasa latihan manakala 0.0091 pada fasa ujian.

Secara keseluruhannya, keputusan yang diperolehi telah berjaya mencapai objektif-objektif penyelidikan ini iaitu:

- (i) rangkaian neural yang dicadangkan berjaya mengkelaskan tahap kualiti air sungai.

Rangkaian neural yang dicadangkan juga mampu mengkelaskan air sungai kepada jenis air dan kualiti air secara serentak. Jenis air akan dikelaskan kepada air tawar dan air payau, manakala kualiti air akan dikelaskan kepada bersih, pertengahan dan tercemar.

- (ii) penyelidikan ini berjaya melakukan analisis penentuan jenis dan tahap kualiti air sungai menggunakan komposisi alga sebagai parameter masukan kepada rangkaian neural.
- (iii) penyelidikan ini mendapati bahawa hanya 21 jenis alga daripada 78 yang dicadangkan memberikan impak yang tinggi kepada pengelasan yang dilakukan. Alga-alga yang memberikan impak yang tinggi adalah *Acnanthes Minutissima*, *Anomoeoneis Brachysira*, *Cocconeis Placentula*, *Cocconeis Species*, *Coscinodiscus Granii*, *Cyclotella Comta*, *Cymbella Inaequalis*, *Diatoma Species*, *Diploneis Decipiens*, *Eutonia Pectinalis*, *Fragilaria Crotonensis*, *Frustulia Saxonica*, *Gomphonema Acuminatum*, *Hantzschia Amphioxys*, *Navicula Radiosa*, *Neidium Affine*, *Nitzschia Amphibia*, *Pinnularia Viridis*, *Psammothidium Bioretii*, *Stauroneis Obtusa*, dan *Surirella Parma Soveriegne*.
- (iv) satu sistem pintar berjaya dibangunkan bagi menentukan jenis dan kualiti air sungai secara automatik menggunakan ciri-ciri dari (i), (ii) & (iii).

5.2 Cadangan-cadangan

Walaupun penyelidikan ini telah membuktikan bahawa sistem penentuan jenis dan kualiti air sungai yang dibina mempunyai keupayaan pengitlakan yang tinggi, namun terdapat beberapa cadangan yang wajar dilakukan untuk memudahkan lagi proses ramalan dilakukan. Ahli biologi perlu menjalankan pemerhatian kepada sampel air di bawah mikroskop berkuasa tinggi dan mencatat kehadiran alga di dalam sampel air tersebut sebelum data-data yang diperolehi dimasukkan ke dalam sistem secara manual. Perkara ini akan menyebabkan ralat pemerhatian berlaku. Bagi mengatasi masalah ini, dicadangkan supaya data-data kehadiran alga-alga ditentukan

menggunakan teknik pemprosesan imej. Sampel air akan diletakkan di bawah mikroskop dan imej air akan dibesarkan dengan skala tertentu kemudian dirakam. Imej ini kemudiannya akan diekstrak menggunakan teknik pemprosesan imej bagi menentukan alga-alga yang hadir dan seterusnya alga-alga ini akan dijadikan data masukan kepada rangkaian neural.

Setelah cadangan pertama berjaya dilakukan, cadangan kedua mampu menjadikan sistem penentuan jenis dan kualiti air sungai ini berfungsi secara automatik. Pembangunan satu mikro pengawal yang menggabungkan pemprosesan imej dan rangkaian neural mampu memberikan nilai komersil kepada sistem ini. Mikro pengawal ini disertakan dengan sebuah kamera kecil kalis air yang akan digunakan bagi merakam imej-imej alga di dalam sungai. Imej-imej yang diperolehi akan diekstrak di dalam mikro pengawal seterusnya pengelasan jenis dan tahap kualiti air dilakukan di dalam mikro pengawal ini. Keputusan pengelasan akan dipaparkan pada paparan kristal cair (*Liquid Crystal Display, LCD*). Sistem ini mudah dibawa ke mana-mana di samping dapat menjimatkan masa pengelasan serta mesra pengguna.

RUJUKAN

- Aggarwal, K.K., Singh, Y., Chandra, P., Puri, M., Indraprastha, G.G.S. & Patil, T.Y. (2005) Bayesian regularization in a neural network model to estimate lines of code using function points. *Journal of Computer Sciences*, 4, Pp 505-509.
- Backer, L.C. (2002) Cyanobacterial harmful algal blooms (cyanoHABs): Developing a public health response. *Journal of Lake and Reservoir Management*, 18, Pp 20-31.
- Boyra, A., Nascimento, J.A., Tuya, F., Sanchez-Jerez, P. & Haroun R.J. (2004) Impact of sea-cage fish farms on inertial macrobenthic assemblages. *Journal of the Marine Biological Association of the UK*, 84, Pp 665-668.
- Bradley, D.C. & Ormerod S.J. (2001) Community persistence among stream invertebrates tracks the North Atlantic Oscillation. *Journal of Animal Ecology*, 70 (6), Pp 987-996.
- Broomhead, D. S. & Lowe, D. (1988). Multivariable Functional Interpolation and Adaptive Networks. *Journal of Complex System*. 2, Pp 321-355.
- Chang, H. (2005) Spatial and Temporal Variations of Water Quality in the Han River and Its Tributaries, Seoul, Korea, 1993–2002. *Journal of Water, Air and Soil Pollution*, 161(1/4), Pp 267-284.
- Choudhury, K., Saha, D.K., & Chakraborty, P.(2001) Geophysical Study for Saline Water Intrusion in a Coastal Alluvial Terrain. *Journal of Applied Geophysics* 46(3), Pp 189–200.
- Coyne, K.J., Handy, S.M., Demir, E., Whereat, E.B., Hutchins, D.A., Portune, K.J., Doblin, M.A. & Cary, S.C. (2005) Improved quantitative real-time PCR assays for enumeration of harmful algal species in field samples using an exogenous DNA reference standard. *Journal of Limnology & Oceanography*, 3, Pp 381–391.
- Crone, S. F., Kausch, H. & Prebmar, D. (2004) Prediction of the CATS benchmark using a Business Forecasting Approach to Multilayer Perceptron Modeling. Proceedings of IEEE International Joint Conference on Neural Networks, 4, Pp 2783-2788.
- Cuffney, T.F., Meador, M.R., Porter, S.D. & Gurtz, M.E. (2000) Responses of physical, chemical, and biological indicators of water quality to a gradient of agricultural land use in the Yakima River Basin, Washington. *Journal of Environmental Monitoring and Assessment*, 6, Pp 259-270.
- DOE (Department of Environment, Malaysia) (1998) *Biomonitoring of Sungai Langat: A case study*. Environment Management: Water Quality Monitoring – Biomonitoring, 2, Pp 23-27.
- Duda, R.O., Heart, P.E. & Stork. (2001) *Pattern Classification*. 2nd edition. New York: John Wiley & Sons.
- Dzeroski S, Todorovski L (2003) *Natural Resources and Water: Managing Queensland's natural resources for today and tomorrow water series*. Queensland Government: Natural Resource Sciences.
- Gallant, S.I. (1995) *Neural Network Learning and Expert Systems*. Massachusetts: The MIT Press.
- Gentleman, W.E. (1973). Least Squares Computations by Givens Transformation with Square Roots. *Journal of The Institute of Mathematics and Applications*, 12, Pp 329-339.
- Golland, P., Eric, W., Grimson, L. & Kikinis, R. (1999) Statistical Shape Analysis Using Fixed Topology Skeletons: Corpus Callosum Study. *Lecturer Notes in Computer Science*, 1613, Pp 382-394.
- Guo, C., Wang, J. & Yin, F. (2004) Advances in neural networks. Proceedings of the International Symposium on Neural Networks, Dalian, Pp 305-312.
- Haykin, S. (2001) *Kalman Filtering and Neural Networks*. New Jersey: Prentice Hall.
- Hush, D. R. & Horne, B. G. (1993). Progress in Supervised Neural Networks. *IEEE Signal Processing Magazine*, Pp 8-39.

- Jeong, K.S., Joo, G.J., Kim, H.W., Ha, K. & Recknagel, F. (2001) Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network. *Journal of Ecological Modeling*, 146, Pp 115-129.
- Joergensen, S. E. & Bendoricchio, G. (2001) *Fundamentals of Ecological Modeling*, 3th Edition. Amsterdam: Elsevier Science Ltd.
- Kambe, J., Aoyama, T., Yamauchi, A. & Nagashima, U. (2007) Extraction of a Parameter as an Index to Assess Water Quality of the Tamagawa, Tokyo, Japan, by Using Neural Networks and Multivariate Analysis. *Journal of Computer & Chemical Japan*, 6(1), Pp 16- 26.
- Komarek J. (2005): Studies on the cyanophytes (Cyanobacteria, Cyanoprokaryota) of Cuba. *Journal of Freshwater Anabaena species*. 77, Pp 211–234.
- Kromhout, G. A. (2005) A comparison of the protonephridia of fresh-water, brackish-water and marine specimens of *Gyratrix hermaphroditus*. *Journal of Morphology*, 72(1), Pp 167-181.
- Lu, S., & Hodgkiss, I.J. (2004) Harmful algal bloom causative collected from Hong Kong waters. *Journal of Hydrobiologia*, 512(1/3), Pp 231-238.
- Mashor, M. Y. (1999). Nonlinear System Identification Using HMLP Networks. *Journal-Institution of Engineers, Malaysia*. 60(2). Pp 33-40.
- Mashor, M. Y. (2000a). Structure Detection of RBF Networks Using DRR Algorithm. *International Journal of The Computer, The Internet and Management*. 8(1), Pp 48-57.
- Mashor, M. Y. (2000b). Hybrid Multilayered Perceptron Networks. *International Journal of System Science*. 31(6). Pp 771-785.
- Mat-Isa, N.A. (2002) *Sistem diagnosis awal barah pangkal rahim berdasarkan rangkaian neural*. Tesis Phd, Universiti Sains Malaysia.
- Maznah, W. & Mansor, M (2002) Aquatic pollution assessment based on attached diatom communities in Pinang River Basin, Malaysia. *Journal of Hydrobiologia*, 487, 229-241.
- Maznah, W., Mansor, M. & Ho, S.C. (2000) Periphyton biomass related to water pollution in Pinang River basin, Malaysia. *International Journal on Algae*, 2(4). Pp 50-70.
- Milner, A.M., Sarah C.C. & Lee, E.B. (2006) Persistence and stability of macroinvertebrate communities in streams of Denali National Park, Alaska: implications for biological monitoring. *Journal of Freshwater Biology*, 51(2), Pp 373-387.
- Mitra, S., & Pal, S.K. (2002) Data Mining in Soft Computing Framework: A Survey. *IEEE Trans. Neural network*, 13, Pp 3-14.
- Nelson, M.M. & Illingworth, W.T. (1992). *A Practical Guide to Neural Nets*. Reading: Addison-Wesley.
- Nichols, S., Steven, W. & Byron S. (2000) A Proposed Aquatic Plant Community Biotic Index for Wisconsin Lakes. *Journal of Environment Management*, 26(5), Pp 491-502.
- Norgaard, M., Ravn, O., Poulsen, N. K. & Hansen, L. K. (2000). *Neural Networks for Modeling and Control of Dynamic Systems*. London: Springer-Verlag London Limited.
- Pareek, N., Jat, M. K. & Jain, S. K. (2006) The utilization of brackish water, protecting the quality of the upper fresh water layer in coastal aquifers. *Journal of The Environmentalist*, 26(4), Pp 237-246.
- Poggio, T. & Girosi, F. (1990). Network for Approximation and Learning. Proc. of The IEEE. 78(9). Pp 1481-1497.
- Powell, M. J. D. (1985). Radial Basis Functions for Multivariable Interpolation: A Review. Proc. of IMA Conf. on Algorithms for The Approximation of Functions and Data. Pp 143-167.
- Quevauviller, P. (2006) Chemical Monitoring Activity under the Common Implementation Strategy of the WFD. *Journal of Soils and Sediments*, 6(1), Pp 2-3.
- Recknagel, F., French, M., Harkonen, P. & Yabunaka, K.I. (1997) Artificial neural network approach for modeling and prediction of algal blooms. *Journal of Ecological Modeling*, 96, Pp 11-28.

Rosenberg, D.M. (1998) A National Aquatic Ecosystem Health Program for Canada: We should go against the flow. *Journal of Bull Entomol Society, Canada*, 30(4), Pp144-152.

Sawyer, C.N., Perry L.M., & Gene F.P. (2003) *Chemistry for Environmental Engineering and Science, 5th edition*. New York: McGraw-Hill.

Sheel, J. S., Vrooman, D., Renner, R. S. & Dawsey, S. K. (2001) A Comparison of Neural Networks and Classical Discriminant Analysis in Predicting Students' Mathematics Placement Examination Scores. ICCS: International Conference, 28/30 May, San Francisco, CA, USA. 2074 Pp 952-959.

Suen, J.P. & Eheart, J.W. (2003) Evaluation of neural network for modeling nitrate concentration in rivers. *Journal of Water Resources Planning & Management*, 129(6), Pp 505-510.

Tang, T. C. & Chi, L. C. (2005) Neural networks analysis in business failure prediction of Chinese importers: A between-countries approach. *Journal of Expert Systems with Applications*, 29(2), Pp 244-255.

Wilamowski, A. & Pener, H. (2003) Efficacy of microencapsulated against the sand fly, Phlebotomus papatasi Scopoli. *Journal of Vector Ecology*, 28 (2), 229-233.

Wilson, H. & Recknagel, F. (2001) Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes. *Journal of Ecological Modeling*, 146, Pp 69-84.

Windelspecht, M. (2002). *Backyard Water Monitoring*. New River News, 2(2), Pp 5-7.

Yoon, H.S., Hackett, J.D., Dolah, F.M.V., Nosenko, T., Lidie, K.L. & Bhattacharya, D. (2005) *Tertiary Endosymbiosis Driven Genome Evolution in Dinoflagellate Algae*. MBE Advance Access.

Zaheer, I. & Bai, C.G. (2003) Application of artificial neural network for water quality management. *Journal of Lowland Technology International*, 5(2), Pp. 10-15.

Zhang, Y., Jouni, T.P., Sampsa S.K. & Martti T.H. (2003) Water Quality Retrievals From Combined Landsat TM Data and ERS-2 SAR Data in the Gulf of Finland. *IEEE Transactions on Geoscience and Remote Sensing*, 41(3). Pp 1-8.

SENARAI PENERBITAN DAN SEMINAR

Seminar

Isa, N.A.Mat., Hashim, F.R., Mei, F.W., Ramli, D.A., Omar, W.M.W & Zamli, K.Z.
(2006) Predicting Quality of River's Water Based on Algae Composition Using Artificial Neural Network. IEEE International Conference on Industrial Informatics, Pp 1340-1345.

Jurnal

Fakroul Ridzuan Hashim, Nor Ashidi Mat Isa, Wan Maznah Wan Omar, Dzati Athiar Ramli, Kamal Zuhairi Zamli & Fauzi Baharudin. Penentuan Kualiti Air Sungai Menggunakan Rangkaian Neural Berbilang Lapisan Hibrid. *Jurnal Teknologi, UTM*. (dalam penelitian)

Nor Ashidi Mat Isa, Fakroul Ridzuan Hashim, Wan Maznah Wan Omar, Dzati Athiar Ramli & Kamal Zuhairi Zamli. Predicting Quality of Water By Using H²MLP Networks. *Elsevier ScienceDirect Journal*. (dalam penelitian)

LAMPIRAN C

PENERBITAN

SENARAI PENERBITAN DAN SEMINAR

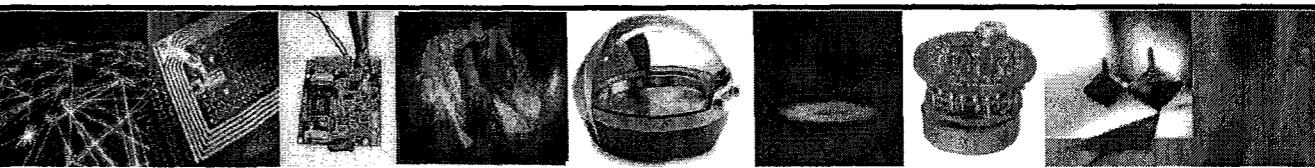
Seminar

Isa, N.A.Mat., Hashim, F.R., Mei, F.W., Ramli, D.A., Omar, W.M.W & Zamli, K.Z. (2006) Predicting Quality of River's Water Based on Algae Composition Using Artificial Neural Network. IEEE International Conference on Industrial Informatics, Pp 1340-1345.

Jurnal

Fakroul Ridzuan Hashim, Nor Ashidi Mat Isa, Wan Maznah Wan Omar, Dzati Athiar Ramli, Kamal Zuhairi Zamli & Fauzi Baharudin. Penentuan Kualiti Air Sungai Menggunakan Rangkaian Neural Berbilang Lapisan Hibrid. *Jurnal Teknologi, UTM.* (dalam penelitian)

Nor Ashidi Mat Isa, Fakroul Ridzuan Hashim, Wan Maznah Wan Omar, Dzati Athiar Ramli & Kamal Zuhairi Zamli. Predicting Quality of Water By Using H²MLP Networks. *Elsevier ScienceDirect Journal.* (dalam penelitian)



2006 IEEE International Conference on Industrial Informatics

Integrating Manufacturing and Services Systems

16 - 18 August 2006
Grand Copthorne Hotel, Singapore

IEEE Catalog Number: 06EX1434C
ISBN: 1-4244-9701-0
Library of Congress: 2006927229



© 2006 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

For technical inquiries, please contact:
Research Publishing Services
Tel: +91-44-24350149; Fax: +91-44-24720542
E-mail: enquiries@researchpubonline.com



© Research Publishing Services

Message from General Chairs
Committees
General Conference Information
Conference Venue Floor Plan
Keynote Speakers
Tutorial Speakers
Program-at-a-Glance
Technical Program
Author Index
International Reviewers
Welcome to Singapore

Predicting Quality of River's Water Based on Algae Composition Using Artificial Neural Network

Nor Ashidi Mat Isa¹, Fakroul Ridzuan Hashim², Fong Wai Mei³, Dzati Athiar Ramli⁴, Wan Maznah Wan Omar⁵ and Kamal Zuhairi Zamli⁶.

^{1,2,3,4,6} School of Electrical & Electronic Engineering,
Universiti Sains Malaysia, Engineering Campus,
14300 Nibong Tebal, Penang, Malaysia.

¹ Tel: +604-5996051, Fax: +604-5941023, Email: ashidi@eng.usm.my

² fakroulridzuan@yahoo.com

³ cheryl_fong82@yahoo.co.uk

⁴ Tel: +604-5996028, Email: dzati@eng.usm.my

⁶ Tel: +604-5996079, Email: eekamal@eng.usm.my

⁵ School of Biology,
Universiti Sains Malaysia, 11800 Minden,
Penang, Malaysia

Tel: +604-6533533, Email: wmaznah@usm.my

Abstract—Artificial neural networks which are inspired by the concept of the biological neurons are commonly used in many applications including in the field of water quality management. The neural network approaches have provided an educated solution to aid in the decision-making process for river system as well as a viable means of the forecasting for water quality parameters. This paper attempts to determine the suitability and the applicability of artificial neural networks for detecting quality of river's water based on algae composition. 21 different types of algae have been used as input data and the river's water was classified into 4 categories, namely clean, polluted, brackish and moderate. Multilayered perceptron network with three different learning algorithms have been studied. The multilayered perceptron trained using Bayesian Regularization algorithm has been proven to produce the best results with high accuracy percentage (93.50%) as compared to the Lavenberg Marquadt (93.00%) and back propagation (63.505%). Further analysis (i.e. more testing data, new architecture of neural network) will be carried out to further improve the system.

Index Terms— algae composition, multilayered perceptron network, river's water quality prediction.

I. INTRODUCTION

A neural network can be seen as a machine that is designed to model the way in which the brain performs a particular task or function of interest [1]. By offering useful properties and capabilities, i.e. nonlinearity, input-output mapping and neurobiological analogy, neural network has successfully been applied in various fields including financial, engineering and mathematics. In those fields, neural networks have been used among others to identify systems, as digital image

processor, in pattern recognition system as well as data classification techniques with significant success rate.

Recently, the applications of the neural networks have been expanded in the field of environment and pollution issues including the water quality management. These are due to the successful of previous researches in this field [2]-[6]. The researches cover the water quality prediction and management, and planning for water pollution control of river systems, wetlands and low land. In terms of water quality management, the neural network approach to multi-objective optimizing has provided an educated solution to aid in the decision-making process for river systems [6]. It is also proven that it has provided a viable means of forecasting of water quality parameters [6].

The water quality management includes the researches in an aquatic ecosystem. Aquatic ecosystems are very complex due to the diversity and connections of the components governing the system's dynamics. A single species or substance which rapidly increases in number or concentration will become a pollutant for the ecosystem and can subsequently affect their surroundings drastically [7]. One of the common environmental hazardous is the blue-green blooms which can reduce the quality of water in rivers, lakes and reservoirs.

The certain species of algae can cause water discolorations [8]. Recreational of aquatic systems can be impaired by resulting water discoloration. Blooms can grow occasionally in shallow waters, so dense that caused fish and invertebrate mortality due to oxygen depletion. This species can also affect human health by the toxins [9]. The toxins can find their way to humans through either drinking water or food chain. As reported by Falconer [10], algae toxins will pass through the normal water treatment and are resistant to boiling. They can cause gastroenteritis and toxic injury to the liver. Additionally, the occurrence of fish and shellfish

poisonings as reported by Todd [11] can considerably reduce consumption and export of seafood. Some algae species can even cause serious damage in aquaculture systems by damming or clogging fish gills. Besides, they can also impair taste and odor problems in drinking water and give rise to customer complaints [10]. Most of these could be prevented or minimized if algae blooms were determined in an early stage. Determination process which is based on microbial in freshwater bodies is useful for developing management responses aimed at reducing the negative consequences of algae blooms to the community [9].

Currently, the quality of water prediction is done by the existence of algae in the water [3][4][5]. The quality of the water is determined by the quantity of algae composition lying in the water. The quality is either very clean, clean, dirty or very dirty in condition [6][7][12]. Using microscope is the current analysis to determine the number of algae and their composition which is time consuming and low accuracy. Hence, the artificial neural networks for detecting the quality level of river's water are currently proposed by researchers to overcome the problems. Nevertheless, it must be stressed that implementation of computer assisted programs using neural networks into conventional water quality management procedures are intended to be only as supporting tools. They merely assist biologists in decision-making and are by no means to replace human expert.

This study focuses on the development of an intelligent detection system for river's water quality data based on the Multilayered Perceptron (MLP). In doing so, this study employs three different learning algorithms, namely the Bayesian Regularization, the Lavenberg Marquadt and back propagation. Based on 21 types of algae the system will be trained to classify the river's water into four classes, which are clean, polluted, brackish and moderate.

II. THE PROPOSED RIVER'S WATER QUALITY PREDICTION SYSTEM

Computer-aided systems based on neural networks are a promising tools in assisting biologists to produce an accurate and faster prediction result for river's water quality in grit to reduce the time required for the conventional determination procedure in laboratory [6]. Although computer-aided screening system based on neural networks has been studied over the last two decades, most of the previous studies still focuses on several common issues. These include:

- 1) Some of previous studies focuses on using neural network for modeling and predicting of population of algae species [4][7][9]. By using the obtained model, the water quality will be determined conventionally.
- 2) The previous study only classified the water quality into two classes, which are dirty and clean [2][6].

Building from the previous work described earlier, this study attempts to enhance the intelligent classification system for river water quality. The current developed system has the following characteristics:

1) The classification of river water quality will be carried out based on 21 different species of algae. These species of algae have been proven [13][14] to be the input data that can be used to classify the quality of river water in laboratory test manually. Thus, this study proposes the use of these species of algae as the input data for neural network to classify the quality of river water automatically.

2) The current study will identify the suitability and applicability of Multilayered Perceptron network in further classifying the quality of river water into four categories, namely clean, polluted, brackish and moderate.

III. MULTILAYERED PERCEPTRON NETWORK

Artificial Neural Networks are computational systems, which are inspired by the concept of biological neural cells which are known as neurons. Neurons are the tiny cells that our brains are composed of. A biological neuron is defined as a collection of a few billion of neurons. Based on the concept, artificial neural networks are proposed, which are capable to model biological neurons structures in terms of both their architecture and operation. They function as mathematical computational models for non-linear function approximation, data classification, and clustering / non-parametric regression. They also function as simulations for the behavior of a collection of model biological neurons in the human. Artificial neural networks are capable in providing performance reliability especially in decision-making in placing of the human brain. Hence, that can act as an alternative to intelligent systems. One of the commonly used artificial neural networks for that purpose is the Multilayered Perceptron (MLP) network.

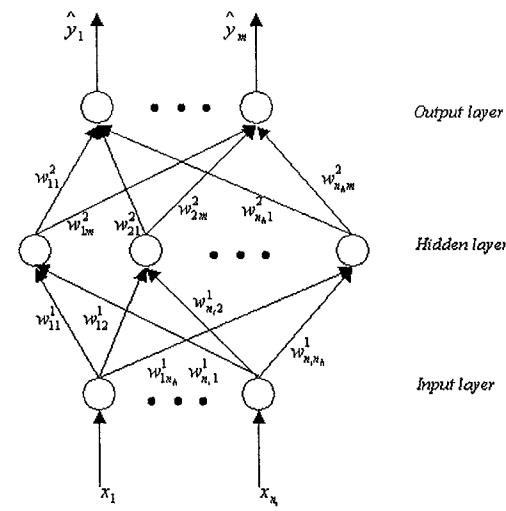


Fig 1: Architecture of the MLP network

In 1958, Rosenblatt introduced a perceptron model [15]. Cascading of several perceptron models in a layer produce a network as shown in Figure 1. This neural network is referred as MLP network. Based on Figure 1, a MLP network with m outputs and n_h hidden nodes and n_i input nodes can be expressed as:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 F\left(\sum_{i=1}^{n_i} w_{ij}^1 x_i^0(t) + b_j^1\right) \text{ for } 1 \leq k \leq m \quad (1)$$

where w_{ij}^1 and w_{jk}^2 denote the weights of the connection between input and hidden layer, and weights of the connection between hidden and output layer respectively. b_j^1 and x_i^0 denote the thresholds in hidden nodes and inputs that are supplied to the input layer respectively. $F(\bullet)$ is an activation function and is normally being selected as sigmoid function.

From Equation (1), the values of w_{ij}^1 , w_{jk}^2 and b_j^1 must be determined using appropriate algorithm. Back propagation (BP) algorithm is commonly used to find optimum values for those parameters due to its easy implementation and produced good performance. This algorithm is based on the error-correction learning rule. This study will employ another two common learning algorithms to train the MLP network, namely the Levenberg-Marquardt and the Bayesian Rule algorithms. The performance of these training algorithms will be compared.

A. Back Propagation Algorithm

Back-propagation, the most commonly used training algorithm, is a gradient descent procedure that computes the derivatives' values in an efficient way, and modifies the weights according to a parameter known as the learning rate [16]. Back propagation is the steepest decent type algorithm where the weight connection between the j -th neuron of the hidden layer and the i -th neuron of the input layer are respectively updated according to:

$$\begin{aligned} w_{ji}(t) &= w_{ji}(t-1) + \Delta w_{ji}(t) \\ b_j(t) &= b_j(t-1) + \Delta b_j(t) \end{aligned} \quad (2)$$

The increment $\Delta w_{ij}(t)$ and $\Delta b_j(t)$ given by:

$$\begin{aligned} \Delta w_{ij}(t) &= \eta_w \rho_j(t) x_i(t) + \alpha_w \Delta w_{ij}(t-1) \\ \Delta b_j(t) &= \eta_b \rho_j(t) + \alpha_b \Delta b_j(t-1) \end{aligned} \quad (3)$$

where the subscripts w and b represent the weight and threshold respectively, α_w and α_b are momentum constants which determine the influence of the past parameter changes on the current direction of movement in the parameter space, η_w and η_b represent the learning rates and $\rho_j(t)$ is the error signal of the j -th neuron of the hidden layer which is back propagated in the network.

Since the activation function of the output neuron is linear, the error signal at the output node is

$$\rho_j(t) = y_k(t) - \hat{y}_k(t) \quad (4)$$

where $y_k(t)$ is the expected output. For the neurons in the hidden layer

$$\rho_j(t) = F'(x_i(t)) \sum_j \rho_j^k(t) w_{jk}^2(t-1) \quad (5)$$

where $F'(x_i(t))$ is the first derivative of $F(x_i(t))$ with respect to $x_i(t)$. Since back propagation algorithm is the steepest decent type algorithm, the algorithm suffers from a slow convergence rate. The search for the global minima may become trapped at local minima and the algorithm can be sensitive to the user selectable parameters [17].

B. Levenberg-Marquardt Algorithm

The Levenberg-Marquardt algorithm is a gradient-based, deterministic local optimization algorithm. Once employed to train the MLP model, the advantage of the Levenberg-Marquardt over the traditional Back Propagation algorithm is that it can provide a faster (second-order) convergence rate and keep relative stability [18][19].

Like the quasi-Newton methods, the Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feed forward networks), then the Hessian matrix can be approximated as:

$$H = J^T J \quad (6)$$

and the gradient can be computed as:

$$g = J^T \rho \quad (7)$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. The Jacobian matrix can be computed through a standard back propagation technique that is much less complex than computing the Hessian matrix [20]. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$\Delta w = -[J^T J + \mu I]^{-1} J^T \rho \quad (8)$$

where Δw is a differential weights and μ is a control parameter.

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift towards Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step increases the

performance function. In this way, the performance function will always be reduced at each iteration of the algorithm [20].

C. Bayesian Rule Algorithm

Given the Baye's Rule as [21]

$$p(\theta | D) = \frac{p(D | \theta)}{p(D)} \quad (9)$$

where $p(\theta)$ is the prior probability of a parameter θ before having seen the data and $p(\theta | D)$ called the likelihood were the probability of the data D .

Bayes' Rule was used to determine the posterior probability of θ given the data D [20]. In general this will provide an entire distribution over possible values of θ . This process was applied to neural networks and came up with the probability distribution over the network weights, w , given the training data $p(w | D)$. When finding a posterior distribution over weights,

$$\begin{aligned} p(w | D) &= \frac{p(D | w)p(w)}{p(D)} \\ &= \frac{p(D | w)}{\int p(D | w)p(w)dw} \end{aligned} \quad (10)$$

In the Bayesian formalism, learning the weights means changing our belief about the weights from the prior, $p(w)$, to the posterior, $p(w | D)$ as a consequence of seeing the data as illustrated by Figure 2.

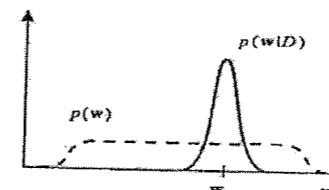


Fig 2 : Changing prior weights to posterior weights

IV. DATA SAMPLES

Aquatic population gives a variety resulting of alterations in the biological integrity of aquatic systems. Algae could serve as an indicator of the water quality degree, and many algae indicator have been used to classify environmental status. Kolkwitz & Marsson [21] were the pioneer to classify algae species. Palmer had published a composite rating of algae species, which can be used to indicate the clean and polluted water. In Malaysia, development has inevitably resulted in adverse changes in the hydrology and ecology of river ecosystems. Recently, only the biological aspects of the integrated study have been included in the water quality monitoring program, and a few algae studies have been

conducted in relation to the water polluted. In related study, Wan Maznah & Mansor [13] reported the occurrence of clean, polluted and brackish diatom species collected from artificial substrate along Sungai Pinang Basin and its tributaries. The Sungai Pinang Basin is the biggest river basin in Penang Island and one of the major polluted river basins in Malaysia. Twelve sampling sites were designed along the Pinang River system. Each sampling site was visited monthly between March 1998 and March 2000 and sampled for water quality values of the six variables and periphytic algae. This study uses the data obtained in [13] to classify the river's water quality into four categories ; namely clean, polluted, moderate and brackish using the MLP networks. The samples consist of 21 types of algae ; namely Acanthans, Amphora Libica, Anomoeoneis Brachysira, Cocconeis, Coscinodiscus, Cyclotella, Cymbella, Diatoma, Diploneis Decipiens, Eunotia, Fragilaria, Frustulia, Gomphonema Acuminatum, Hantzschia Amphioxys, Navicula, Neidium Affine, Nitzschia, Pinnularia, Psammothidium Bioretti, Stauroneis Obtusa and Surirella. The existence of these algae species in the river water samples will be used as input data for the MLP networks. Thus, the MLP network consists of 21 input nodes which each input node represents species of algae respectively. 120 data (32 clean, 55 polluted, 11 moderate and 22 brackish) are used to train the MLP network to determine the optimum structure of network, while 80 data (23 clean, 36 polluted, 7 moderate and 14 brackish) are used to test the MLP network's capability.

V. RESULT

Similar to the procedure used by [22][23] for data classification, this study applies two analysis. Firstly, the optimum structure analysis is implemented to determine the optimum number of hidden nodes that can produce the best river's water quality classification. For this analysis, the MLP network is trained with 3600 epochs. The time taken during each training are 9.859s, 5.078s and 4.641s for the MLP network trained using BP, LM and BR algorithms respectively. Table 1 shows the result of the analysis for different types of training algorithm.

TABLE I
RESULTS FOR THE OPTIMUM STRUCTURE ANALYSIS OF THE MLP NETWORK TRAINED USING DIFFERENT TRAINING ALGORITHMS.

MLP Network Training Algorithm	Optimum number of hidden nodes
BP	9
LM	3
BR	3

The second analysis, (i.e the performance comparison analysis) is then applied using the optimum structure of the MLP network obtained in the first analysis. The analysis is done based on the accuracy of correct

classifications of river's water quality. Table 2 shows the result of the analysis for training and testing phase.

TABLE 2
RESULTS FOR PERFORMANCE ANALYSIS OF THE MLP NETWORK TRAINED USING DIFFERENT TRAINING ALGORITHMS.

MLP network trained using different training algorithm	Accuracy Percentage		
	Training (%)	Testing (%)	Overall (%)
BP	61.67	66.25	63.50
LM	95.83	88.75	93.00
BR	95.83	90.00	93.50

This study further employs an analysis to find the minimum number of epochs and the shortest time to reach predetermined accuracy. As the MLP network trained using BP can only produce the highest accuracy of 63.50%, this study sets the accuracy level for all algorithms to 63.50%. Table 3 shows the results of the minimum number of epochs and the shortest time to reach predetermined accuracy.

TABLE 3
RESULTS FOR MINIMUM NUMBER OF EPOCHS AND MINIMUM CONVERGENCE TIME ANALYSIS OF THE MLP NETWORK TRAINED USING DIFFERENT TRAINING ALGORITHMS.

MLP network trained using different training algorithm	Convergence Rate	
	Number of Epochs	Time (s)
BP	3600	9.859
LM	10	5.078
BR	10	4.641

VI. DISCUSSION

The results tabulated in Table 1 show that the MLP network trained using the Bayesian Regularization (BR) forms the simplest network architecture as it only requires 3 hidden nodes as compared to other training algorithms (back propagation (BP) and Lavenberg Marquadt (LM)).

For the performance analysis, the results obtained as tabulated in Table 2 show that the MLP network trained using BR algorithm produces the best performance of accuracy with 95.83% and 90.00% for training and testing phase respectively, which gives overall accuracy of 93.50%.

The results obtained as tabulated in Table 3 show that the MLP network trained using BR and LM algorithms needs the lowest number of epochs (i.e. only 10 epochs) to reach the predetermined accuracy as compared to BP which requires 3600 epochs. But the result favors the MLP network trained using BR as compared to LM and BP in term of minimum time required to obtain the predetermined accuracy. The MLP network trained using BR only needs 4.641s to reach the predetermined

accuracy with different of 0.637s and 5.218s compared to LM and BP respectively. This shows that the MLP network trained using BR has lower convergence rate as compared to LM and BP.

The promising results obtained suggest four conclusions:

- 1) The MLP network is suitable as intelligent classifier to classify the quality of river water into clean, polluted, brackish and moderate.
- 2) Algae composition is significantly possible to be used as input data for the MLP network to determine the river water quality.
- 3) The MLP network using BR as the training algorithm is more intelligent and has lower convergence rate as compared to LM and BP.
- 4) Also the MLP network trained using BR algorithm produce the highest accuracy of river water quality determination as compared to the MLP network trained using LM and BP.

VII. CONCLUSION

This paper analyzes neural network method in terms of water quality monitoring and the result comparison with the biological monitoring approach. The earlier concept of the usage of algae in water monitoring is based on the indicator species concept. As a pilot study, this system proves that artificial neural network is suitable and capable to classify the river's water quality based on algae comparison.

Further improvement can be made to upgrade the performance of the proposed system. In addition, more case studies should be done to test the system in order to establish its capability and reliability. A good suggestion is by using different type of neural networks and learning algorithms could be employed.

REFERENCES

- [1] S. Haykin. *Neural Network: A Comprehensive Foundation*, Prentice Hall: New Jersey, 1994.
- [2] C. G. Wen. & C. S. Lee. "A neural network approach to multi-objective optimization for water quality management in a river basin," *Wat. Resources Res.*, 34(3), 1998, pp 427-436.
- [3] H. R. Maier, G. C. Dandy & M. D. Burch. "Use of artificial neural networks for modeling cyanobacteria Anabaena spp. In the River Murray, South Australia," *Ecol. Model.*, 105, 1998, pp 257-272.
- [4] F. Recknagel. "ANNA - Artificial Neural Network Model for predicting species abundance and succession of blue-green algae," *Hydrobiologia*, 349, 1997, pp 45-47.
- [5] F. Recknagel, T. Fukushima, T. Hanazato, N. Takamura & H. Wilson. "Modeling and predicting phyto and zooplankton dynamics in Lakes Kasumigaura by artificial neural networks," *Lakes Reserve. Res. Manag.*, 3, 1998, pp 45-47.
- [6] R. M. Holger & C. D. Graeme. "The use of artificial neural networks for the prediction of water quality parameters," *Wat. Res. Res.*, 32(4), 1996, pp 1013-1022.
- [7] F. Recknagel, M. French, P. Harkonen & K. I. Yakabuna. "Artificial Neural Network approach for modeling and prediction of algal blooms," *Ecological Modeling*, 96, 1997, pp 11-28.
- [8] G. M. Hallegraeff. "Algal blooms and their apparent global increase," *Phycologica*, 32(2), 1993, pp 79-99.
- [9] H. Wilson & F. Recknagel. "Toward a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes," *Ecological Modeling*, 01, 2001, pp 69-84.
- [10] I. R. Falconer. *Alga toxin in seafood and drinking water*, Academic Press: London, 1993.
- [11] E. C. D. Todd. "Domoic acid and amnesic shellfish poisoning," *J. Food Protection*, 56(1), 1993, pp 69-83.
- [12] N. Takamura, T. Iwakuma & M. Yassuna. "Primary production in Lake Kasumigaura, 1981-1985," *Jpn. J. Limnol.*, 48, 1987, pp 19-38.
- [13] W. Maznah & M. Mansor. "Aquatic pollution assessment based on attached diatom communities in the Pinang River Basin, Malaysia," *Hydrobiologia*, 487, 2002, pp 229-241.
- [14] W. Maznah & M. Mansor. "Benthic diatoms in the Pinang River (Malaysia) and its tributaries with emphasis on species diversity and water quality," *International Journal on Algae*, 1(4), 1999, pp 103-108.
- [15] D. R. Hush & B. G. Horne. "Progress in Supervised Neural Networks," *IEEE Signal Processing Magazine*, 1993, pp. 8-39.
- [16] A. A. El-Falahi, R. A. Marti & L. B. Lasdon. "Path relinking and GRG for artificial neural networks," *European Journal of Operational Research*, 169(2), 2006, pp 508-519.
- [17] S. Jayavant, M. Singh. "Artificial neural network analysis of malaria severity through aggregation and deformability parameters of erythrocytes," *Clinical Hemorheology and Microcirculation*, 29(3-4), 2003, pp 457-468.
- [18] R. Battiti. "First and second-order method for learning between steepest descent and Newton's method," *Neural Computation*, 4, 1992, pp 141-166.
- [19] D. Wang, W. Z. Lu. "Forecasting of ozone level in the time series using MLP model with a novel hybrid training algorithm," *Atmosphere Environment*, 40(5), 2006, pp 913-924. Citing internet source URL: <http://www.mathworks.com/access/helpdesk/help/toolbox/hnet/backpr12.html>
- [20] R. Kolkwitz & M. Marsson. "Oekologie der pflanzlichen Saprobien," *Berichte der Deutschen Botanischen Gesellschaft*, 26(2), 1908, pp 505-519.
- [21] N. A. Mat Isa., M. Y. Mashor, N. H. Othman. "Classification of Cervical Cancer using HMLP Network with Confident Level Analysis," *International Journal of Computer, the Internet and Management*, 11(1), 2003, pp 17-29.
- [22] P. Mitra, S. Mitra, & S. K. Pal. "Staging of Cervical Cancer with Soft Computing," *IEEE Transaction on Biomedical Engineering*, 47 (7), 2000, pp 934-940.

Penentuan Kualiti Air Sungai Menggunakan Rangkaian Perceptron Berbilang Lapisan Hibrid

Fakroul Ridzuan Hashim¹, Nor Ashidi Mat Isa², Wan Maznah Wan Omar³,
Dzati Athiar Ramli⁴, Kamal Zuhairi Zamli⁵, dan Fauzi Baharudin⁶.

^{1,2,4,5} Pusat Pengajian Kejuruteraan Elektrik & Elektronik,

⁶ Pusat Pengajian Kejuruteraan Awam,

Universiti Sains Malaysia, Kampus Kejuruteraan,
14300 Nibong Tebal, Pulau Pinang, Malaysia.

¹ Tel: +604-5996051, Faks: +604-5941023, Emel: ashidi@eng.usm.my

² Tel: +604-5996093, Faks: +604-5941023, Emel: fakroulridzuan@yahoo.com.my

Abstrak - Rangkaian neural buatan berkonsepkan neuron biologi telah digunakan secara meluas dalam pelbagai bidang dan antaranya dalam bidang pengurusan kualiti air. Pendekatan kepada penggunaan rangkaian neural telah menghasilkan kaedah penyelesaian masalah yang pintar bagi membantu proses penyelesaian masalah berkaitan sistem sungai terutama dalam menentukan kualiti air sungai. Kajian ini dijalankan bertujuan untuk menguji sama ada rangkaian neural buatan berkemampuan dan bersesuaian untuk mengesan kualiti air sungai berdasarkan komposisi alga. 21 jenis alga yang berbeza telah digunakan sebagai parameter masukan. Kualiti air sungai dikelaskan kepada empat kategori iaitu bersih, sederhana, tercemar dan payau. Rangkaian Perceptron Berbilang Lapisan Hibrid (HMLP) yang dilatih menggunakan Ralat Ramalan Jadi Semula Ubahsuai (MRPE) digunakan dalam kajian ini. HMLP yang telah dilatih menggunakan algoritma MRPE algorithm telah terbukti menghasilkan keputusan yang sangat baik dengan peratusan kejituhan yang tinggi iaitu 98.58%. Keputusan yang telah diperolehi, jelas menunjukkan rangkaian HMLP sesuai digunakan untuk menentukan kualiti air sungai dan berkeupayaan tinggi dalam menentuan kualiti air sungai dengan.

Katakunci — komposisi alga, rangkaian neural, rangkaian Perceptron Berbilang Lapisan Hibrid, Ralat Ramalan Jadi Semula Ubahsuai, penentuan kualiti air sungai.

I. PENGENALAN

Satu rangkaian neural boleh diibaratkan sebagai mekanisma yang dapat merekabentuk satu model menyamai kaedah otak bekerja dalam pelbagai situasi [1]. Kelebihan dan kebolehan rangkaian neural merangkumi pelbagai aplikasi seperti aplikasi ketidaklinearan [2][3] dan pemetaan masukan & keluaran [4]. Disamping itu juga, rangkaian neural berjaya diaplikasikan dalam bidang kewangan [5], kejuruteraan [6][7] dan statistik [8]. Aplikasi rangkaian neural seperti pengenalpastian sistem [9][10], pemprosesan imej [11][12] dan pengenalpastian corak [13] telah membawa kejayaan di peringkat yang tinggi.

Kini, pendekatan rangkaian neural telah diaplikasikan dalam bidang berkaitan isu-isu pencemaran terutamanya isu pengurusan kualiti air. Aplikasi rangkaian neural kini telah mendapat perhatian setelah melihat kepada kejayaan kajian-kajian terdahulu [14]-[18]. Kajian yang dijalankan merangkumi peramalan dan pengurusan kualiti air serta perancangan kawalan saliran sungai yang telah tercemar. Dalam menguruskan kualiti air, pendekatan menggunakan rangkaian neural telah memberikan penyelesaian yang baik bagi mengatasi permasalahan berkaitan saliran sungai [18]. Pengurusan kualiti air merangkumi kajian-kajian dalam satu ekosistem akuatik yang besar. Ekosistem akuatik adalah terlalu kompleks disebabkan oleh pembiakan spesis yang berterusan. Pembiakan spesis dengan cepat dan dalam kuantiti yang besar akan mendatangkan pencemaran kepada ekosistem dan memberikan kesan buruk kepada persekitaran jika tidak dikawal [19]. Populasi spesis alga biru kehijauan merupakan alga yang sangat berbahaya kepada persekitaran dan mampu menurunkan kadar kualiti air di sungai-sungai, tasik-tasik dan kawasan-kawasan tadahan [19].

Spesis *Cylindrospermopsis Raciborskii* boleh menyebabkan mengaratan air terjadi [20]. Sistem rekreasi akuatik akan menjadi tidak seimbang disebabkan oleh pengaratan air. Populasi ini boleh

membiaik dalam air yang cetek dan menyebabkan kematian kepada hidupan akuatik akibat kekurangan oksigen terlarut yang terkandung di dalam air. Disamping itu juga, toksin yang terhasil daripada spesis *Cylindrospermopsis Raciborskii* boleh mengancam nyawa manusia [21]. Toksin ini mungkin terkandung di dalam air minuman dan lebih malang lagi jika terdapat di dalam sistem rantaian makanan manusia. [22] membuktikan toksin-toksin daripada alga mampu melepas kaedah rawatan air yang dilakukan secara normal dan tegar kepada suhu didih air. Ia akan mengakibatkan gastrik dan keracunan kepada hati. Tambahan pula, Todd [23] melaporkan toksin daripada spesis *Alexandrium*, *Dinophysis* dan *Pseudo-nitzschia* multiseries menyebabkan hidupan laut pupus akibat keracunan. Perkara ini menyebabkan pengurangan kepada permintaan eksport makanan laut. Spesis seperti *Heterosigma* *Akashiwo* dan *Cryptosporidium Parvum* mampu membunuh spesis-spesis ikan dengan akibat insang yang tersumbat teruk. Selain itu, spesis *Cyanophytes* menyebabkan bau, rasa dan warna kepada bekalan air minuman berubah sama sekali seterusnya meningkatkan rungutan di kalangan pelanggan [22]. Masalah-masalah ini dapat dihalang atau dikurangkan sekiranya populasi pembiakan populasi spesis alga ini dapat dikesan pada peringkat awal. Kehidupan mikro di dalam air amat berguna dan mampu memberikan sebarang tindakbalas dalam pengurusan kualiti air. Perkara ini dapat mengurangkan kesan negatif akibat kehadiran spesis-spesis alga ini di dalam komuniti hidupan air [21].

Berdasarkan kepada fakta yang dibentangkan, banyak penyelidikan telah dilakukan untuk menentukan kualiti air kehadiran alga di dalam air [15][16][17]. Tahap kualiti air bergantung kepada kuantiti komposisi alga yang terdapat di dalamnya. Kualiti air terbahagi kepada beberapa kategori iaitu sangat bersih, bersih, kotor atau sangat kotor [18][19][24]. Penggunaan mikroskop bagi menentukan komposisi dan bilangan alga yang terdapat pada sampel air memerlukan masa yang lama serta tahap kejituannya adalah rendah. Bagi mengatasi masalah ini, penggunaan rangkaian neural buatan dalam menentukan tahap kualiti air sungai telah dicadangkan. Penggunaan sistem bantuan komputer melalui rangkaian neural dalam prosedur pengurusan kualiti air hanyalah sekadar sokongan kepada kaedah konvensional sahaja. Ia mampu membantu ahli biologi dalam membuat penyelesaian masalah tetapi bukan menggantikan tugas manusia secara keseluruhan.

Fokus utama kajian ini adalah membangunkan satu sistem pintar untuk menentukan kualiti air sungai menggunakan rangkaian HMLP. Dalam menjalankan kajian ini, algoritma latihan yang dinamakan MRPE telah digunakan. Berdasarkan 21 jenis spesis alga, sistem ini akan dilatih untuk mengelaskan kualiti air kepada empat kategori iaitu bersih, sederhana, tercemar dan payau.

II. SISTEM YANG DICADANGKAN

Bantuan daripada sistem perkomputeran menggunakan rangkaian-rangkaian neural berupaya membantu ahli-ahli biologi bagi mendapatkan keputusan yang jitu. Ramalan yang telah diberikan adalah tepat dan pantas dalam menentukan kualiti air sungai. Pendekatan ini dapat mengatasi masalah batasan masa seperti yang dialami oleh kaedah konvensional yang dijalankan di dalam makmal pada masa kini [18]. Walaupun penggunaan rangkaian neural telah digunakan sejak dua dekad yang lalu, tetapi kebanyakan kajian terdahulu hanya memfokuskan kepada beberapa isu-isu semasa. Antaranya ialah :

- 1) Kajian terdahulu hanya memberi penekanan dalam meramal kehadiran populasi spesis-spesis alga [16][19][21]. Dengan menggunakan model ramalan itu, kualiti air ditentukan secara konvensional dan manual di dalam makmal.
- 2) Bagi kajian terdahulu, kualiti air hanya ditentukan kepada dua kategori sahaja sama ada bersih atau kotor [14][18].

Berdasarkan kajian terdahulu, kajian ini memfokuskan pengubahsuaian kepada sistem penentuan pintar bagi meramal kualiti air sungai. Sistem terkini yang dibangunkan adalah memenuhi ciri-ciri yang telah ditetapkan seperti di bawah :

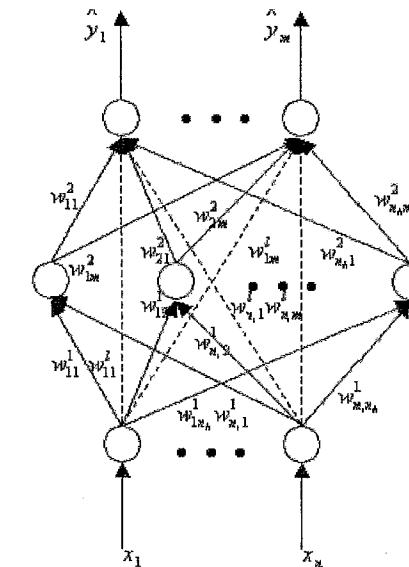
- 1) Kualiti air sungai akan ditentukan berdasarkan 21 jenis spesis alga yang berlainan. Penggunaan spesis-spesis ini telah disahkan oleh [25][26] sebagai penunjuk bagi menentukan kualiti air sungai secara konvensional di dalam makmal. Oleh itu, kajian ini mencadangkan supaya spesis-spesis alga

ini dijadikan parameter-parameter masukan kepada rangkaian neural bagi menentukan kualiti air sungai secara automatik.

- 2) Kajian ini akan mengenalpasti kesesuaian dan kebolehan rangkaian neural dalam menentukan kualiti air sungai kepada empat kategori iaitu bersih, sederhana, tercemar atau payau secara automatik.

III. RANGKAIAN PERCEPTRON BERBILANG LAPISAN HIBRID

Rangkaian neural buatan merupakan salah satu sistem pengkomputeran yang diinspirasikan daripada konsep sel neural biologi yang juga dikenali sebagai neuron. Neuron adalah sel-sel yang sangat kecil yang terdapat di dalam otak manusia. Satu neuron biologi terdiri daripada berbilion-bilion neuron. Berdasarkan konsep itu, rangkaian neural buatan telah dicadangkan dan berupaya memodelkan struktur neuron biologi baik dari segi senibina mahupun aplikasi. Ia berfungsi sebagai model perkomputeran matematik kepada anggaran fungsi tak lurus, pengklasifikasian data dan regresi pengelompokan. Ia adalah simulasi kepada himpunan model neuron-neuron biologi dalam sistem otak manusia. Rangkaian neural buatan sangat berkebolehan dan mampu menghasilkan prestasi yang tepat terutamanya menyelesaikan masalah. Oleh sebab itu, ia menjadi alternatif kepada sistem-sistem pintar. Dalam kajian ini, rangkaian HMLP telah dicadangkan sebagai peramal kepada kualiti air sungai. Rangkaian HMLP dengan satu lapisan tersembunyi adalah seperti yang ditunjukkan pada Rajah 1.



Rajah 1. Rangkaian HMLP dengan satu lapisan tersembunyi

Keluaran pada satu nod keluaran ke- k , \hat{y}_k diberikan oleh [27]:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 F(\sum_{i=1}^{n_i} w_{ij}^1 x_i^0(t) + b_j^1) + \sum_{i=0}^{n_i} w_{ik}^l x_i^0(t) \quad \text{for } 1 \leq k \leq m \quad (1)$$

dimana w_{jk}^2 mewakili sambungan-sambungan pemberat di antara lapisan tersembunyi dengan lapisan keluaran. w_{ik}^l pula mewakili sambungan-sambungan pemberat di antara lapisan masukan dengan lapisan keluaran. w_{ik}^l mewakili sambungan-pemberat tambahan yang menghubungkan lapisan-lapisan masukan terus ke lapisan-lapisan keluaran. x_i^0 dan b_j^1 mewakili masukan yang telah dihubungkan kepada pengambangan nod tersembunyi. n_i , m dan n_h masing-masing adalah bilangan nod masukan, nod keluaran dan nod tersembunyi. $F(\bullet)$ adalah satu fungsi pengaktifan dan penyelidikan ini memilih fungsi sigmoid sebagai fungsi pengaktifan. w_{jk}^2 , w_{ik}^l , w_{ij}^1 dan nilai ambang b_j^1 tidak diketahui dan dikemaskini untuk meminimumkan ramalan ralat yang diberikan oleh :

$$\varepsilon_k(t) = y_k(t) - \hat{y}_k(t) \quad (2)$$

di mana $y_k(t)$ dan $\hat{y}_k(t)$ masing-masing adalah keluaran sebenar dan juga keluaran rangkaian.

IV. RALAT RAMALAN JADI SEMULA UBAHSUAI

Merujuk kepada persamaan (1), nilai w_{ij}^1 , w_{jk}^2 , w_{ik}^l dan b_j^1 perlu dipertimbangkan dengan menggunakan algoritma tertentu. Algoritma Perambatan Suap Belakang (Back Propagation, BP) telah diketahui umum mampu memberikan nilai optimum kepada parameter-parameter tersebut. Walaubagaimanapun, algoritma ini dapat memberikan prestasi yang baik tetapi kadar penumpuan adalah perlahan. Bagi mengatasi masalah ini, Chen *et al.* [28] telah mencadangkan penggunaan Ralat Ramalan Jadi Semula (Recursive Prediction Error, RPE) bagi menggantikan algoritma BP. Algoritma RPE memberikan kadar penumpuan yang lebih cepat di samping mampu memberikan nilai pemberat dan nilai ambang yang lebih baik. Dalam [27], Mashor telah mencadangkan versi ubahsuai kepada algoritma RPE, yang dikenali sebagai Ralat Ramalan Jadi Semula Ubahsuai (Modified Recursive Prediction Error, MRPE). Dengan mengoptimumkan momentum dan kadar pembelajaran, algoritma MRPE mampu memperbaiki kadar penumpuan yang telah ditunjukkan oleh algoritma RPE.

Algoritma RPE yang telah diuahsuai oleh Chen *et al.* [28] meminimumkan kos fungsi berikut :

$$J(\hat{\Theta}) = \frac{1}{2N} \sum \varepsilon^T(t, \hat{\Theta}) \Lambda^{-1} \varepsilon(t, \hat{\Theta}) \quad (3)$$

dengan mengemaskinikan parameter vektor penganggaran, $\hat{\Theta}$ (terdiri daripada $w-w$ dan $b-b$), yang terhasil semula menggunakan algoritma Gauss-Newton :

$$\hat{\Theta}(t) = \hat{\Theta}(t-1) + P(t)\Delta(t) \quad (4)$$

dan

$$\Delta(t) = \alpha_m(t)\Delta(t-1) + \alpha_g(t)\psi(t)\varepsilon(t) \quad (5)$$

di mana $\varepsilon(t)$ dan Λ masing-masing adalah ralat ramalan dan satu matrik positif yang simetri $m \times m$, dan m adalah bilangan nod-nod keluaran; $\alpha_m(t)$ dan $\alpha_g(t)$ masing-masing adalah momentum dan kadar pembelajaran. $\alpha_m(t)$ dan $\alpha_g(t)$ boleh diwakilkan oleh sesuatu nilai diantara 0 hingga 1, manakala nilai-nilai tipikal $\alpha_m(t)$ dan $\alpha_g(t)$ masing-masing menghampiri nilai 1 dan 0. Kajian semasa telah

menunjukkan $\alpha_m(t)$ dan $\alpha_g(t)$ mampu memperbaiki kadar penumpuan yang telah dihasilkan oleh algoritma RPE berdasarkan kepada:

$$\alpha_m(t) = \alpha_m(t-1) + a \quad (6)$$

dan

$$\alpha_g(t) = \alpha_m(t)(1 - \alpha_m(t)) \quad (7)$$

di mana a adalah nilai tetap (dengan nilai tipikal $a = 0.01$), $\psi(t)$ mewakili gradian keluaran ramalan satu langkah kehadapan, \hat{y} dengan mematuhi parameter rangkaian :

$$\psi(t, \Theta) = \left[\begin{array}{c} \hat{y}(t, \Theta) \\ \frac{d\hat{y}(t, \Theta)}{d\Theta} \end{array} \right] \quad (8)$$

$P(t)$ dalam persamaan (4) telah dikemaskinikan jadi semula berdasarkan :

$$P(t) = \frac{1}{\lambda(t)} \left[P(t-1) - P(t-1)\psi(t)(\lambda(t)I + \psi^T(t)P(t-1)\psi(t))^{-1}\psi^T(t)P(t-1) \right] \quad (9)$$

di mana $\lambda(t)$ adalah faktor lemati, $0 < \lambda(t) < 1$, yang telah dikemaskinikan menggunakan skim berikut :

$$\lambda(t) = \lambda_0 \lambda(t-1) + (1 - \lambda_0) \quad (10)$$

Di mana λ_0 adalah faktor lemati awalan. Nilai awalan matrik $P(t)$, iaitu $P(0)$ telah ditetapkan kepada αI di mana I adalah matrik identiti dan α adalah nilai tetap, bertipikal di antara 100 dan 10000.

Matrik gredian, $\psi(t)$ diubah dengan meletakkan satu sambungan linear tambahan kepada model rangkaian HMLP satu lapisan tersembunyi dengan membezaan persamaan (1) terhadap parameter θ_c , untuk menghasilkan :

$$\psi_k(k) = \frac{dy_k(t)}{d\theta_c} = \begin{cases} u_j & \text{jika } \theta_c = w_{jk}^2 \quad 1 \leq j \leq n_h \\ x_i & \text{jika } \theta_c = w_{ik}^l \quad 0 \leq i \leq n_l \\ u_j(1-u_j)w_{jk}^2 & \text{jika } \theta_c = b_j^1 \quad 1 \leq j \leq n_h \\ u_j(1-u_j)w_{jk}^2 x_i & \text{jika } \theta_c = w_{ij}^1 \quad 1 \leq j \leq n_h, 1 \leq i \leq n_l \\ 0 & \text{lain-lain} \end{cases} \quad (11)$$

Pengubahsuaihan kepada algoritma RPE untuk rangkaian HMLP satu lapisan tersembunyi boleh diimplementasikan seperti berikut [27]:

1. Berikan nilai awalan kepada pemberat-pemberat dan pengambangan, $P(0)$, a , b , $\alpha_m(0)$, λ_0 dan $\lambda(0)$. (b adalah parameter yang mempunyai nilai tipikal di antara 0.8 dan 0.9).
2. Masukkan data masukan pertama kepada rangkaian dan kira keluaran rangkaian berdasarkan kepada persamaan (1).
3. Kira ralat ramalan berdasarkan kepada persamaan (2).
4. Kira matrik $\psi(t)$ berdasarkan kepada persamaan (11). (Nota : Elemen $\psi(t)$ perlu dikira daripada lapisan keluaran dan seterusnya ke lapisan tersembunyi).
5. Kira matrik $P(t)$ dan $\lambda(t)$ masing-masing berdasarkan persamaan (9) dan (10).
6. Sekiranya $\alpha_m(t) < b$, kemaskinikan $\alpha_m(t)$ berdasarkan persamaan (6).
7. Kemaskinikan $\alpha_g(t)$ dan $\Delta(t)$ masing-masing berdasarkan kepada persamaan (7) dan (5).
8. Kemaskinikan parameter vektor $\hat{\Theta}(t)$ berdasarkan kepada persamaan (4).
9. Ulang langkah (3) hingga (8) bagi setiap sampel data latihan.

V. SAMPEL DATA

Populasi akuatik mampu meramal sebarang fenomena ke atas perubahan yang berlaku dalam kombinasi biologi sistem akuatik. Alga dapat dijadikan penunjuk kepada tahap kualiti air dan populasi pelbagai spesis alga telah digunakan bagi penentuan keadaan alam sekeliling. Kolkwitz & Marsson [29] merupakan individu yang merintis penemuan penentuan kualiti air berdasarkan komposisi spesis-spesis alga. Di Malaysia, pembangunan pesat menyebabkan perubahan yang drastik dalam hidrologi dan ekologi ekosistem sungai. Kini, hanya aspek-aspek biologi daripada kajian yang bersepudau mengandungi program pemerhatian kualiti air. Beberapa kajian terhadap alga telah dijalankan berkait-rapat dalam hubungan berkaitan pencemaran air. Dalam satu kajian, Wan Maznah & Mansor [25] telah

mengumpulkan spesis-spesis alga yang berada di dalam air di sepanjang kawasan tадahan Sungai Pinang. Kawasan tадahan Sungai Pinang adalah kawasan tадahan yang terbesar di bahagian pulau di negeri Pulau Pinang dan adalah salah satu kawasan tадahan yang teruk tercemar. Terdapat 12 stesen pensampelan yang telah dikenalpasti di sepanjang Sungai Pinang. Pensampelan air sungai telah dilakukan pada setiap bulan bermula Mac 1998 hingga Mac 2000. Dalam kajian ini, data-data yang terkandung di dalam [25] telah digunakan untuk menentukan kualiti air secara automatik kepada beberapa kumpulan kualiti air sama ada bersih, sederhana, tercemar ataupun payau. Proses menentukan kualiti air ini dilakukan menggunakan rangkaian HMLP. Sampel-sampel air sungai mengandungi 21 jenis alga seperti Acnanthes, Amphora Libica, Anomoeoneis Brachysira, Coccconeis, Coscinodiscus, Cyclotella, Cymbella, Diatoma, Diploneis Decipiens, Eunotia, Fragilaria, Frustulia, Gomphonema Acuminatum, Hantzschia Amphioxys, Navicula, Neidium Affine, Nitszchia, Pinnularia, Psammothidium Bioretti, Stauroneis Obtusa and Surirella. Kehadiran spesis-spesis alga di dalam sampel-sampel air sungai akan dijadikan sebagai parameter masukan kepada rangkaian HMLP. Oleh itu, rangkaian HMLP akan mengadungi 21 nod masukan yang mewakili spesis-spesis alga. 1200 data telah digunakan bagi melatih rangkaian HMLP manakala 800 data akan digunakan untuk menguji kebolehan rangkaian tersebut, pada fasa ujian.

VI. KEPUTUSAN

Prosedur pengklasifikasian rangkaian neural telah dijalankan seperti [30][31], di mana dua jenis analisis telah dilakukan. Pertama, analisis terhadap struktur optimum rangkaian neural dengan mengambil kira bilangan lelaran data dan bilangan nod tersembunyi. Keputusan yang telah diperolehi menunjukkan bahawa sistem yang telah dicadangkan mampu memberikan kejituuan yang tinggi dengan hanya menggunakan 3 lelaran data dan juga menggunakan 26 bilangan nod tersembunyi sahaja. Analisis kedua dijalankan dengan menggunakan struktur optimum yang telah diperolehi daripada analisis pertama untuk menentukan kualiti air. Jadual 1 dan jadual 2 masing-masing menunjukkan prestasi klasifikasi yang telah

dijalankan pada fasa latihan dan ujian. Keputusan ini terhasil selepas rangkaian HMLP dilatih menggunakan struktur optimum yang telah diperolehi pada analisis terdahulu.

Jadual 1 : Keputusan ramalan kualiti air sungai bagi fasa latihan.

Klasifikasi	Betul	Salah	Jumlah	Prestasi
Bersih	328	0	328	100.00%
Sederhana	108	1	109	99.08%
Tercemar	542	1	543	99.82%
Payau	205	15	220	93.36%
Keseluruhan	1183	17	1200	98.58%

Jadual 2 : Keputusan ramalan kualiti air sungai bagi fasa ujian.

Klasifikasi	Betul	Salah	Jumlah	Prestasi
Bersih	214	8	222	96.39%
Sederhana	67	6	73	94.37%
Tercemar	364	1	365	99.63%
Payau	122	18	140	87.20%
Keseluruhan	767	33	800	95.88%

Penyelidikan ini juga membandingkan prestasi pengklasifikasian kualiti air sungai menggunakan rangkaian HMLP dengan rangkaian Perceptron Berbilang Lapisan (Multilayered Perceptron, MLP) konvensional. Rangkaian MLP dilatih menggunakan dua jenis algoritma latihan yang berlainan iaitu perambatan belakang (BP) dan Lavenberg Marquardt (LM). Jadual 3 menunjukkan perbandingan bilangan lelaran data dan bilangan nod tersembunyi optimum bagi rangkaian neural yang digunakan.

Jadual 4 pula menunjukkan perbandingan prestasi ramalan bagi semua rangkaian neural.

Jadual 3. Keputusan bagi analisis struktur optimum

Jenis Rangkaian Neural	Bilangan Lelaran Data		Bilangan Nod Tersembunyi
	3	26	3
HMLP using MRPE	3	26	3
MLP using BP	3600	3	3
MLP using LM	100	3	3

Jadual 4 : Keputusan analisis prestasi bagi rangkaian neural

Jenis Rangkaian Neural	Prestasi Kejituuan		
	Latihan (%)	Ujian (%)	Keseluruhan (%)
HMLP dengan MRPE	98.58	95.88	97.23
MLP dengan BP	70.67	64.25	67.46
MLP dengan LM	97.83	93.15	95.49

VI. PERBINCANGAN

Keputusan yang telah diberikan oleh jadual 1 dan 2 menunjukkan rangkaian HMLP memberikan prestasi ramalan dengan kejituuan yang tinggi. Jadual 1 menunjukkan keputusan bagi fasa latihan, di mana ramalan pada klasifikasi bersih adalah sangat baik dengan kadar ketepatan adalah 100%. Keputusan 99.08% telah diberikan oleh klasifikasi sederhana manakala keputusan 99.82% pula adalah bagi klasifikasi tercemar. Keputusan bagi klasifikasi sederhana dan tercemar adalah sangat baik walaupun tidak menyamai kadar keputusan klasifikasi bersih. Keputusan bagi klasifikasi payau memberikan keputusan dengan ketepatan ramalan sebanyak 93.36%. Walaupun keputusan klasifikasi payau paling rendah, tetapi dengan peratusan ketepatan ramalan yang tinggi menjadikan keputusan ramalan itu masih

baik. Keputusan keseluruhan kejituhan prestasi ramalan bagi fasa latihan adalah sangat baik dengan kejituhan 98.85%.

Jadual 2 pula menunjukkan keputusan prestasi ramalan rangkaian HMLP bagi fasa ujian. Keputusan prestasi ramalan yang diberikan oleh klasifikasi bersih adalah 96.39% tepat. Keputusan yang ditunjukkan oleh klasifikasi sederhana pula adalah 94.37%. Keputusan yang telah diberikan oleh klasifikasi tercemar adalah amat tinggi dengan 99.63%. Keputusan yang diberikan oleh klasifikasi tercemar adalah sangat baik dan hampir mencapai ketepatan maksimum. Bagi prestasi ramalan klasifikasi payau pula, keputusan ketepatan adalah hanya 87.20% sahaja. Keputusan ini kurang baik berbanding klasifikasi-klasifikasi lain dengan ketepatan lebih dari 90.00% tetapi ketepatan yang diberikan oleh klasifikasi payau masih lagi baik. Keputusan keseluruhan kejituhan prestasi ramalan bagi fasa ujian pula adalah baik dengan kejituhan 95.88%.

Jadual 3 menunjukkan keputusan bagi analisis struktur optimum manakala jadual 4 menunjukkan analisis prestasi kejituhan. Kedua-dua jadual adalah perbandingan di antara rangkaian HMLP yang menggunakan algoritma latihan MRPE dengan rangkaian MLP yang menggunakan algoritma latihan BP dan LM. Daripada keputusan yang telah diberikan, didapati rangkaian HMLP dengan algoritma latihan MRPE mampu menghasilkan rangkaian yang paling ringkas disamping kadar penumpuan yang cepat. Struktur yang paling optimum dihasilkan pada lelaran data ketiga dan bilangan nod tersembunyi yang ke-26. Disamping itu juga, kejituhan pada keseluruhan prestasi ramalan adalah sangat tinggi iaitu 97.23%. Keputusan yang diberikan oleh rangkaian MLP dengan algoritma latihan BP menunjukkan rangkaian itu adalah satu struktur yang menumpu dengan lebih perlahan berbanding rangkaian HMLP. Struktur optimum mampu dicapai pada lelaran data ke-3600 dan bilangan nod tersembunyi ketiga. Walaupun nod tersembunyi yang diperlukan oleh rangkaian MLP dengan algoritma latihan BP adalah sedikit berbanding dengan rangkaian HMLP menggunakan algoritma latihan MRPE tetapi kejituhan keseluruhan prestasi ramalan yang ditunjukkan hanya 67.46% sahaja. Bgitu juga dengan keputusan yang diberikan oleh rangkaian MLP dengan algoritma latihan LM, pada lelaran data ke-100 dan bilangan nod tersembunyi ke-3 sudah mampu mencapai struktur optimum. Kadar penumpuan bagi rangkaian MLP

dengan algoritma latihan LM adalah pantas tetapi kejituhan keseluruhan prestasi ramalan yang ditunjukkan adalah 95.49% kurang berbanding rangkaian HMLP menggunakan algoritma MRPE sebagai algoritma latihan.

Berdasarkan prestasi kejituhan pada jadual 4 pula, analisa peratus perbezaan bagi kejituhan prestasi ramalan dapat dilakukan. Dengan menggunakan rangkaian MLP dengan menggunakan BP sebagai algortima latihan, keputusan yang diberikan adalah 70.67% bagi fasa latihan diikuti dengan 64.25% bagi fasa ujian. Rangkaian MLP dengan menggunakan LM sebagai algortima latihan pula memberikan peningkatan keputusan yang lebih baik iaitu 97.83% pada fasa latihan manakala 93.15% bagi fasa ujian. Daripada keputusan ini dapat diperhatikan rangkaian MLP yang dilatih menggunakan algoritma LM memberikan keputusan yang sangat baik berbanding rangkaian MLP yang dilatih menggunakan algoritma BP dengan peratus perbezaan adalah 27.16% bagi fasa latihan dan 28.90% bagi fasa ujian. Rangkaian HMLP dengan menggunakan algoritma latihan MRPE pula memberikan keputusan prestasi ramalan yang terbaik. Rangkaian ini mampu berikan keputusan dengan ketepatan yang tinggi iaitu 98.85% bagi fasa latihan manakala 95.88% bagi fasa ujian. Perbezaan keputusan yang ketara ditunjukkan berbanding MLP yang dilatih menggunakan algoritma BP dengan 27.91% pada fasa latihan manakala 31.67% pada fasa ujian. Perbezaan keputusan yang ditunjukkan terhadap rangkaian MLP yang dilatih menggunakan algoritma LM pula amat kecil disebabkan oleh rangkaian MLP telah sedia ada baik dan mampu memberikan keputusan yang tinggi dengan perbezaan peratusan adalah 0.75% pada fasa latihan manakala 2.73% pada fasa ujian. Bagi keseluruhan prestasi ramalan, rangkaian HMLP dengan menggunakan algoritma latihan MRPE memberikan ketepatan 97.23%. Perbezaan peratusan adalah 29.77% berbanding rangkaian MLP yang dilatih menggunakan algoritma BP dan 2.72% bagi rangkaian MLP yang dilatih menggunakan algoritma LM. Perbezaan keputusan kejituhan rangkaian MLP yang dilatih menggunakan algoritma LM berbanding rangkaian MLP yang dilatih menggunakan algoritma BP adalah 22.03%.

Keputusan-keputusan yang telah diberi jelas menunjukkan rangkaian HMLP berkebolehan untuk melakukan ramalan dengan baik di samping dapat menghasilkan kejituhan yang tinggi. Keputusan-

keputusan ini jelas menunjukkan rangkaian HMLP berkebolehan tinggi dalam menentukan kualiti air sungai berdasarkan komposisi alga malah lebih baik daripada rangkaian MLP.

Daripada keputusan-keputusan yang diperolehi, terdapat tiga kesimpulan yang dapat dibuat :

- 1). Rangkaian HMLP sesuai digunakan dalam sistem pintar bagi mengklasifikasikan kualiti air sungai.
- 2). Komposisi alga mampu dijadikan sebagai parameter masukan kepada rangkaian-rangkaian neural bagi meramal dan menentukan kualiti air sungai.
- 3). Rangkaian HMLP mampu mengklasifikasikan kualiti air sungai kepada empat kategori iaitu bersih, sederhana, tercemar dan payau berbanding pengklasifikasian kepada dua kategori sebelum ini iaitu bersih dan tercemar.

VI. KESIMPULAN

Kajian ini menganalisis kaedah penggunaan rangkaian neural dalam pemerhatian kualiti air dan perbandingan keputusan telah dilakukan terhadap pendekatan melalui pemerhatian secara biologi. Konsep pemerhatian kualiti air bagi sistem yang dicadangkan adalah berdasarkan kehadiran spesis-spesis alga di dalam air. Kombinasi kehadiran spesis alga di dalam air digunakan sebagai penunjuk dalam menentukan kualiti air tersebut. Sebagai satu kajian yang baru, sistem ini telah membuktikan rangkaian neural buatan sesuai dan mampu untuk mengklasifikasikan kualiti air sungai kepada empat kategori iaitu bersih, sederhana, tercemar dan payau berdasarkan komposisi alga. Penentuan kualiti air secara automatik telah diperkenalkan bagi menggantikan penentuan kualiti air secara konvensional. Penambahbaikan kepada sistem yang dibangunkan ini masih boleh dijalankan. Dalam pada itu, kes kajian perlu diperbanyakkan lagi bagi merealisasi dan mengoptimumkan penggunaan sistem ini secara global.

RUJUKAN

- [1] Haykin, S. *Neural Network: A Comprehensive Foundation*, Prentice Hall: New Jersey, 1994.
- [2] Billings, S. A. "Identification of Nonlinear System – a survey", *IEEE proceedings*, Vol 127, No. 6, pp 272-285, 1980.
- [3] Chen, H. W. "Modelling and Identification of Parallel Nonlinear System : Structural Classification and Parameter Estimation Method", *Proceeding of IEEE*, Vol 88, No. 1, 1995.
- [4] Lee, H. S. & Han, D. "Implementation of Real Time Color Gamut Mapping using Neural Network", *IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications*, University of Technology, Espoo, Finland, 2005.
- [5] Kim, S. H. & Hak, S. K. "Graded Forecasting using An Array of Bipolar Predictions : Application of Probabilistic Neural Networks to a Stock Market Index", *International Journal of Forecasting*, Vol 14, No. 3, pp 323-337, 1998.
- [6] Yeh, I. C., "Structural Engineering Application with Augmented Neural Networks", *Computer-Aided Civil and Infrastructure Engineering*, Vol 13, Issue 2, pp 83-96, 1998.
- [7] Cai, J. G., Zhao, J. & Hudson J. A. "Computerization of Rock Engineering Systems using Neural Network with an Expert System", *International Journal of Rock Mechanics and Rock Engineering*, Vol 31, No. 3, pp 135-152, 1998.
- [8] Press, S. J. & Wilson, S., "Choosing between Logistic Regression and Discriminant Analysis", *Journal of The American Statistic Association*, Vol 73, pp 699-705.
- [9] Mashor, M. Y. "Nonlinear System Identification using HMLP Networks", *Journal of Institution of Engineers Malaysia*. Vol 60(2), pp 33-40, 1999.
- [10] Mashor, M. Y. "Some Properties of RBF Networks with Applications to System Identification", *International Journal of The Computer, The Internet and Management*. Vol 7(1), pp 34-56, 1999.
- [11] Chen, C. W., Luo, J. & Parker, K. J. "Image Segmentation via Adaptive K-Mean Clustering and Knowledge-Based Morphological Operation with Biomedical Application", *IEEE Transaction on Image Processing*, Vol 7(12), pp 1673-1683.

- [12] Craine, B. L., Craine, E. R., O'Toole, C. J. & Qiang, J. "Digital Image Colcoscopy : Corrected Area Measurement using Shape-Form-Shading", *IEEE Transaction on Medical Imaging*. Vol 17(6), pp 1003-1010, 1998.
- [13] Chai, B. B., Zhang, X., Zhao, Y. & Sklansky, J. "Binary Linear Decision Tree with Genetic Algorithm", *Proceeding of The 13th International Conference on Pattern Recognition*. Vol 3, pp 530-534, 1996.
- [14] Wen, C. G. & Lee, C. S. "A neural network approach to multi-objective optimization for water quality management in a river basin," *Wat. Resources Res*, 34(3), 1998, pp 427-436.
- [15] Maier, H. R., Dandy, G. C. & Burch, M. D. "Use of artificial neural networks for modeling cyanobacteria Anabaena spp. In the River Murray, South Australia," *Ecol. Model*, 105, 1998, pp 257-272.
- [16] Recknagel, F. "ANNA – Artificial Neural Network Model for predicting species abundance and succession of blue-green algae," *Hydrobiologia*, 349, 1997, pp 45-47.
- [17] Recknagel, F., Fukushima, T., Hanazato, T., Takamura, N. & Wilson, H. "Modeling and predicting phyto and zooplankton dynamics in Lakes Kasumiagaura by artificial neural networks," *Lakes Reserve. Res. Manag*, 3, 1998, pp 45-47.
- [18] Holger, R. M. & Graeme, C. D. "The use of artificial neural networks for the prediction of water quality parameters," *Wat. Res. Res*, 32(4), 1996, pp 1013-1022.
- [19] Recknagel, F., French, M., Harkonen, P. & Yakabuna, K. I. "Artificial Neural Network approach for modeling and prediction of algal blooms," *Ecological Modeling*, 96, 1997, pp 11-28.
- [20] Hallegraeff, G. M. "Algal blooms and their apparent global increase," *Phycologica*, 32(2), 1993, pp 79-99.
- [21] Wilson, H. & Recknagel, F. "Toward a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes," *Ecological Modeling*, 01, 2001, pp 69-84.
- [22] Falconer, I. R. *Alga toxin in seafood and drinking water.*, Academic Press: London, 1993.
- [23] Todd, E. C. D. "Domoic acid and amnesic shellfish poisoning," *J.Food Protection*, 56(1), 1993, pp 69-83.
- [24] Takamura, N., Iwakuma, T. & Yassuna, M. "Primary production in Lake Kasumigaura, 1981-1985," *Jpn.J.Limnol*, 48, 1987, pp 19-38.
- [25] Maznah, W. & Mansor, M. "Aquatic pollution assessment based on attached diatom communities in the Pinang River Basin, Malaysia," *Hydrobiologia*, 487, 2002, pp 229-241.
- [26]. Maznah, W. & Mansor, M. "Benthic diatoms in the Pinang River (Malaysia) and its tributaries with emphasis on species diversity and water quality," *International Journal on Algae*, 1(4), 1999, pp 103-108.
- [27] Mashor, M.Y. "Hybrid Multilayered Perceptron," *International Journal of Systems Science*, Vol. 31 No.6, 2000, pp. 771-785.
- [28]. Chen, S., Cowan, C. F. N., Billings, S. A. & Grant, P. M. "A parallel recursive prediction error algorithm for training layered neural networks," *International Journal of The Computer, The Internet and Management*, Vol.11, No.1, 2004, pp 17-29.
- [29]. Kolkwitz, R. & Marson, M. "Oekologie der pflanzlichen Saprobien," *Berichte der Deutschen Botanischen Gesellschaft*, 26(2), 1908, pp 505-519.
- [30]. Mitra, P., Mitra, S. & Pal, S. K. "Staging of Cervical Cancer with Soft Computing," *IEEE Transaction on Biomedical Engineering*. 47 (7), 2000, pp 934-940.
- [31]. Mat Isa N. A., Mashor, M. Y. & Othman, N. H. "Classification of Cervical Cancer using HMLP Network with Confident Level Analysis," *International Journal of Computer, the Internet and Management*. 11(1), 2003, pp 17-29.

Classification of River Water Quality using Artificial Neural Network

Based on Biological Indicator

Nor Ashidi Mat Isa¹, Fakroul Ridzuan Hashim², Kamal Zuhairi Zamli³
and Wan Maznah Wan Omar⁴

^{1,2,3} School of Electrical & Electronic Engineering,
Universiti Sains Malaysia, Engineering Campus,
14300 Nibong Tebal, Seberang Perai Selatan, Penang, MALAYSIA.

¹ Tel: +604-5996051, Fax: +604-5941023, Email: ashidi@eng.usm.my

² Tel: +604-5996093, Fax: +604-5941023, Email: fakroulridzuan@yahoo.com.my

³ Tel: +604-5996079, Fax: +604-5941023, Email: eekamal@eng.usm.my

⁴ School of Biological Sciences, Universiti Sains Malaysia,

11800 Minden, Penang, MALAYSIA

Tel: +604-6533533, Fax: +604-5941023, Email: wmaznah@usm.my

Abstract: Biological neurons' concept has inspired the development of artificial neural network. Its application covers a wide range of field including the water quality management. One of the advantages in applying the neural network approach is the educated solution providing in assisting the decision-making process for river system and also for forecasting the water quality parameters. The applicability and suitability of this artificial neural network in detecting the river's water quality based on algae composition

is determined in this research. Water type classification and river's water quality prediction is expected as the outcome of this study. As the input data, 21 different types of algae are used. As for the river's water, it will be classified either as freshwater or brackish-water. For each freshwater and brackish-water, it will then be categorized into 3 different quality levels which are clean, moderate and polluted. Hence, the hierarchical hybrid multilayered perceptron (H^2MLP) with modified recursive prediction error (MRPE) learning algorithm have been studied. The H^2MLP network is proven to produce results with accuracy percentage of 99.13%. This is yet to be improved as soon as the further analysis (i.e. testing data, new architecture of neural network) is carried out.

Keywords: Algae composition, hierarchical hybrid multilayered perceptron network, river's water quality prediction, modified recursive prediction error.

1. INTRODUCTION

The researches of water quality management have been done including the research on aquatic system. The balancing of aquatic ecosystems is usually very complex because of the diversity and the connections amongst the components in the ecosystem influence the system's dynamic. This is due to the increasing number or concentration of any single species or substances will pollute the ecosystem. These pollutants will then drastically affect the system's surroundings [1]. One of the components that can reduce the quality of water in rivers, lakes and reservoirs is the blue-green blooms [1].

Cylindrospermopsis Raciborskii species can cause water discolorations [2]. This water discoloration can impair the recreational of aquatic system. Blooms grow can

cause fish and invertebrate mortality because bloom growth tends to cause the oxygen depletion when they fill up the density, especially in shallow waters. *Cylindrospermopsis* species endanger human health by the toxicity [3]. This often occurs to human either by directly drinking the water or indirectly via the food chain. According to Falconer's report [4], algae toxins are resistant to normal water treatment and also to boiling. This is how algae attack human health, specifically by causing gastroenteritis and toxic injury to the liver. As reported by Todd [5], *Alexandrium*, *Dinophysis* dan *Pseudo-nitzschia multiseries* species could poison many species of fish and shellfish which can lead to reduction of seafood export and consumption. Serious damage that occurs in aquaculture systems such as damming and clogging fish grills are caused by the *Heterosigma Akashiwo* dan *Cryptosporidium Parvum* species. Furthermore, *Cyanophytes* species will lead to taste and odor problems for drinking water where this will cause increasing complaints amongst the water consumers [4]. Due to these problems, early detection of the algae blooms presence in aquaculture ecosystem is essential. The microbial-in-freshwater-bodies-based detection process is important in order to develop management responses in which to reduce the negative effects from the algae blooms to the community [3].

To predict the water quality, the existence of algae in the water is determined [6][7][8]. Basically, there are 4 condition categories for the water quality which are very clean, clean, dirty and very dirty [1][9][10]. The most common and current method to determine the algae contents is by using the microscope. However, this method is inefficient due to time and accuracy wise. River's water quality determination is done

traditionally through biological observation towards the algae types contained in the water. The river's water sample is taken from identified stations every month. The algae observation is done in laboratory. Each algae will be computed and recorded for determination results. For more accurate, some tests using certain chemicals need to be done. The results gained from the test will then be compared to the table prepared by Department of Environmental (DOE). To obtain the results from the test using chemicals, it requires longer time besides the possibility of less accuracy due to human error.

Generally, several parameters are used as the indicator in order to determine the river's water quality. The values such as *dissolved oxygen* (DO), *biological oxygen demand* (BOD), *chemical oxygen demand* (COD) and *suspended solids* (SS) will be recorded and compared to the table prepared by Department of Environmental [11][12]. The table only covers all areas in Malaysia and the nearby countries. This is because of the river's water quality level and temperature varied in different countries. However, the equipments used are usually expensive but able to produce accurate results. The parameters that often used are only able to determine water quality without taking into account the presence of algae. Basically, algae presence in the river's water is not the ultimate reason of pollution, but the pollution has caused several algae species to be able to live in the area. This will make the pollution worse and more critical with the presence of algae species as the main agent of pollution. Waste release from the living area to the river without going through treatment process will make the water pollution even worse.

In order to overcome this problem, the artificial neural network is suggested by the researchers to detect the quality level of river's water. However, the implementation

of computer assisted programs using neural networks into conventional water quality management procedures are intended to be only as supporting tools. This is basically to assist the biologists in making the decisions, not to replace human experts.

A neural network is like a machine which is developed to model the function of brain in a particular task or function of interest [13]. Being such useful properties and capabilities of neural network i.e. nonlinearity [14][15][16] and input-output mapping [16][17], the neural network has taken place in various fields including financial [18], engineering [19][20][21] and statistical application [22].

The application of the neural network has expanded wider recently in the environment and pollution issues including the water quality management. This is because of the succeeded previous research of neural network [1][6] – [9][23]. The research has covered the field from water quality prediction and management to planning for water pollution control of river systems, wetlands and low land. The neural network approach is useful in optimizing multi-objective as it is packaged with educated solution to aid in the decision-making process for river systems [9]. Besides, it also provides a viable means of forecasting the water quality parameters [9].

The application of algae composition increment using neural network has been done [7]. The composition of bad algae which gives negative effects to its surroundings will be controlled while as for the good algae which improves the quality of water will be increased. Since the algae growth is very rapid, the fast determination of water quality is needed. A research on developing a system which can predict and determine water quality in a faster rate but user-friendlier has been done. In our previous study, neural network is proven to produce results with very high accuracy percentage for water quality

determination [24]. The quality of river's water is also categorized into four more detailed groups compared to the previous researches which are clean, moderate, polluted and brackish. The research produces up to 98.85% of accuracy. As for the latest research, there has been an upgrading where determination system does not only determine the water quality, but also the types of the river's water.

The development of an intelligent detection system for river's water quality based on H²MLP is emphasized on this research. Hence, the learning algorithm named MRFE is employed. The system will be trained according to the 21 types of algae in order to classify the river's water into 2 types either it is fresh or brackish. Each of water type is then classified further based on its quality, such as clean, moderate or polluted.

2. THE PROPOSED SYSTEM

In assisting the biologists to produce an accurate and faster result for river's water quality, the computer-aided systems based on neural networks is applied as an aid tool in which to reduce the time required for the conventional determination procedure in laboratory [9]. Over the last two decades, the computer-aided screening system based on neural network has been studied. However, the previous studies are emphasizing on several common issues which include:

1. Using neural network for modeling and predicting algae species population [1][3][7]. Hence, the water type and quality will have to be determined manually and conventionally.
2. Only classifying the river's water into two classes (i.e. dirty and clean [9][23]).

The classification is improved by adding more groups (i.e. clean, moderate,

polluted and brackish) in our previous study [24].

This study is improving the previous studies on developing an intelligent classification system for river's system which are described earlier. The development of the system focuses on the following matters:

1. Water quality classification will be done based on the 21 different types of algae species. These species are proven [25][26] to function as the input data for manual river's water classification laboratory test. As for the application of neural network, these algae species are proposed as the input data to classify river's water type and quality automatically.
2. Identifying the suitability and applicability of an artificial neural network (i.e. especially the H²MLP network) to classify the river's water quality into 2 types, namely fresh water and brackish water. For fresh water type, it will be further classified into 3 classes, such as clean, moderate or polluted. For brackish-water, it will be further classified into two classes, which are brackish or polluted.

3. INTELLIGENT CLASSIFICATION SYSTEM FOR RIVER WATER QUALITY

From literature review no attempt was carried out to use the conventional multilayered perceptron (MLP) network to classify the river water quality based on biological indicator (i.e. algae composition). In our previous studies [24], the standard MLP network trained using back propagation algorithm produced 63.51% accuracy when it was used to classify the river water quality into clean, polluted, brackish and moderate. With

additional linear connection between input nodes and output nodes, we proved that a hybrid version of the MLP network called HMLP network, improved the overall accuracy up to 92.13 percent.

Hybrid Multilayered Perceptron

It has been shown in [27] that modeling a linear model using the standard nonlinear MLP network is not the best solution. An optimum performance of modeling both linear and nonlinear systems could be achieved using Hybrid Multilayered Perceptron (HMLP) network [27]. Nonlinear system is modeled by the standard connections (i.e. represented by line connection in Figure 1) as of the standard MLP network, and the linear system could be modeled by additional direct connections between input nodes to output nodes (i.e. represented by dotted line connections in Figure 1). For m output nodes, the output of the HMLP network is given by:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 F\left(\sum_{i=1}^{n_l} w_{ij}^1 x_i^0(t) + b_j^1\right) + \sum_{i=1}^{n_l} w_{ik}^l x_i^0(t) \quad \text{for } 1 \leq k \leq m \quad (1)$$

where w_{ij}^1 , w_{jk}^2 and w_{ik}^l denote the weights of the connection between input and hidden layer, weights of the connection between hidden and output layer, and weights of the linear connection between input and output layer respectively. b_j^1 and x_i denote the thresholds in hidden nodes and inputs that are supplied to the input layer respectively. $F(\bullet)$ is an activation function and is normally be selected as sigmoid function.

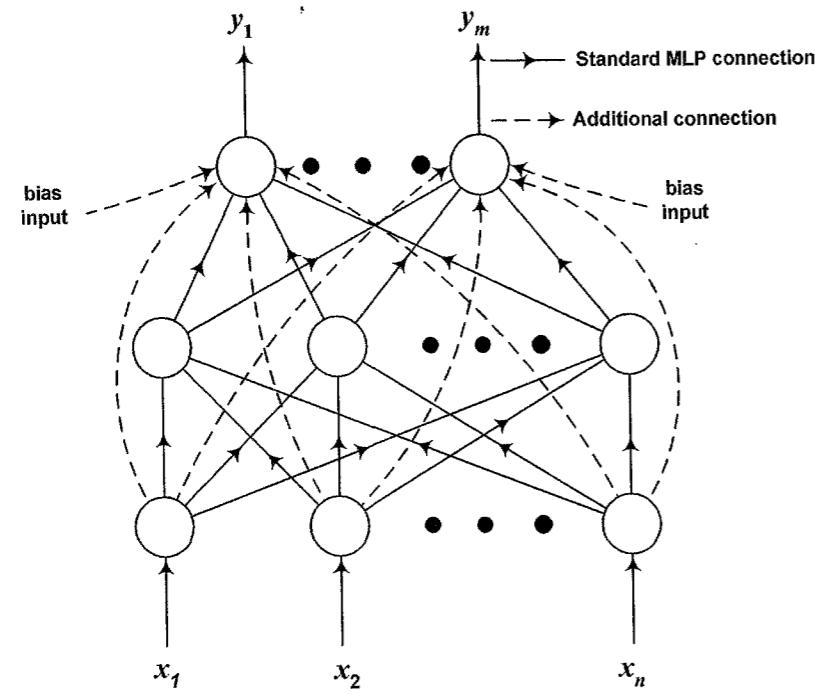


Figure 1: One hidden layer HMLP network

Learning algorithm for the HMLP network to determine the values of w_{ij}^1 , w_{jk}^2 , w_{ik}^l and b_j^1 have been proposed in [27]. To handle the additional linear connections, a modified version of Recursive Prediction Error (RPE) (the detailed RPE algorithm could be found in [28][29]), namely Modified Recursive Prediction Error (MRPE) is introduced [27]. By optimizing the way the momentum and the learning rate are assigned, the MRPE algorithm is able to improve the convergence rate of the RPE algorithm. This section will briefly explain the MRPE algorithm. The detailed MRPE algorithm can be found in [27].

The standard RPE algorithm proposed in [28] minimizes the following cost function:

$$J(\hat{\Theta}) = \frac{1}{2N} \sum \varepsilon^T(t, \hat{\Theta}) \Lambda^{-1} \varepsilon(t, \hat{\Theta}) \quad (2)$$

by updating the estimated parameter vector, $\hat{\Theta}$ (consists of w 's and b 's), recursively using the Gauss-Newton algorithm:

$$\hat{\Theta}(t) = \hat{\Theta}(t-1) + P(t) \Delta(t) \quad (3)$$

and

$$\Delta(t) = \alpha_m(t) \Delta(t-1) + \alpha_g(t) \psi(t) \varepsilon(t) \quad (4)$$

where $\varepsilon(t)$ and Λ are the prediction error and a $m \times m$ symmetric positive definite matrix respectively, and m is the number of output nodes; $\alpha_m(t)$ and $\alpha_g(t)$ are the momentum and the learning rate respectively. $\alpha_m(t)$ and $\alpha_g(t)$ can be arbitrarily assigned to some values between 0 and 1, and the typical values of $\alpha_m(t)$ and $\alpha_g(t)$ are closed to 1 and 0 respectively. In [27], $\alpha_m(t)$ and $\alpha_g(t)$ are varied to further improve the convergence rate of the RPE algorithm according to:

$$\alpha_m(t) = \alpha_m(t-1) + a \quad (5)$$

and

$$\alpha_g(t) = \alpha_m(t)(1 - \alpha_m(t)) \quad (6)$$

where a is a small constant (typically $a = 0.01$); $\psi(t)$ represents the gradient of the one-step-ahead predicted output, \hat{y} with respect to the network parameters:

$$\psi(t, \Theta) = \left[\frac{d \hat{y}(t, \Theta)}{d \Theta} \right] \quad (7)$$

$P(t)$ in equation (3) is updated recursively according to:

$$P(t) = \frac{1}{\lambda(t)} \left[P(t-1) - P(t-1) \psi(t) (\lambda(t) I + \psi^T(t) P(t-1) \psi(t))^{-1} \psi^T(t) P(t-1) \right] \quad (8)$$

where $\lambda(t)$ is the forgetting factor, $0 < \lambda(t) < 1$, and has been updated using the following scheme:

$$\lambda(t) = \lambda_0 \lambda(t-1) + (1 - \lambda_0) \quad (9)$$

where λ_0 and the initial forgetting factor, $\lambda(0)$ are the design values. The initial value of the $P(t)$ matrix, $P(0)$ is set to αI where I is the identity matrix and α is a constant, typically between 100 and 10000.

The gradient matrix, $\psi(t)$ can be modified to accommodate the extra linear connections for a one-hidden-layer HMLP network model by differentiating equation (7) with respect to the parameters, θ_c , to yield:

$$\psi_k(k) = \frac{dy_k(t)}{d\theta_c} = \begin{cases} u_j & \text{if } \theta_c = w_{jk}^2 \quad 1 \leq j \leq n_h \\ x_i & \text{if } \theta_c = w_{ik}^l \quad 0 \leq i \leq n_i \\ u_j(1 - u_j)w_{jk}^2 & \text{if } \theta_c = b_j^1 \quad 1 \leq j \leq n_h \\ u_j(1 - u_j)w_{jk}^2 x_i & \text{if } \theta_c = w_{ij}^1 \quad 1 \leq j \leq n_h, 1 \leq i \leq n_i \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The MRPE algorithm to determine the output $y_k(t)$ for a one-hidden-layer HMLP network can be implemented as follows [27]:

1. Initialize weights, thresholds, $P(0)$, a , b , $\alpha_m(0)$, λ_0 and $\lambda(0)$. (b is a design parameter that has a typical value between 0.8 and 0.9).
2. Present inputs to the network and compute the network outputs according to equation (1).
3. Calculate the prediction error according to:

$$\varepsilon_k(t) = y_k(t) - \hat{y}_k(t) \quad (11)$$

where $y_k(t)$ is the actual output.

4. Compute matrix $\psi(t)$ according to equation (10). Note that, elements of $\psi(t)$ should be calculated from the output layer down to the hidden layer.
5. Compute matrix $P(t)$ and $\lambda(t)$ according to equations (8) and (9) respectively.
6. If $\alpha_m(t) < b$, update $\alpha_m(t)$ according to equation (5).
7. Update $\alpha_g(t)$ and $\Delta(t)$ according to equations (6) and (4) respectively.
8. Update parameter vector $\hat{\Theta}(t)$ according to equation (3).
9. Repeat steps (2) to (8) for each training data sample.

Hierarchical Hybrid Multilayered Perceptron

The MRPE algorithm can only be applied to one output node HMLP network. Although in the MRPE algorithm, it is denoted that it can be applied to any output node (i.e. $y_k(t)$) but from the algorithm, the output for each output node must be implemented separately. Thus, for the proposed classification system for river water quality, five separate HMLP networks are required to detect the fresh water with clean quality, fresh water with moderate quality, fresh water with polluted quality, brackish water with moderate quality and brackish water with polluted quality respectively.

The original structure of the HMLP network has to be modified for effective classification process. Four HMLP networks are cascaded together to form hierarchical HMLP (H^2MLP) network. Figure 2 shows the first HMLP network classifies the river water into its types; i.e. fresh and brackish water. The input data for the brackish water will be fed into the second HMLP network, which will further classify the quality of brackish water into moderate and polluted. The input data for the fresh water will be fed

into the third HMLP network. The data will be classified into clean/moderate and polluted. Then, the input data for the clean/moderate quality will be fed into the fourth HMLP network, which will classify the data into clean and moderate quality. The H^2 MLP network approach reduces the number of the neural networks adopted if the standard HMLP network is used for the classification process.

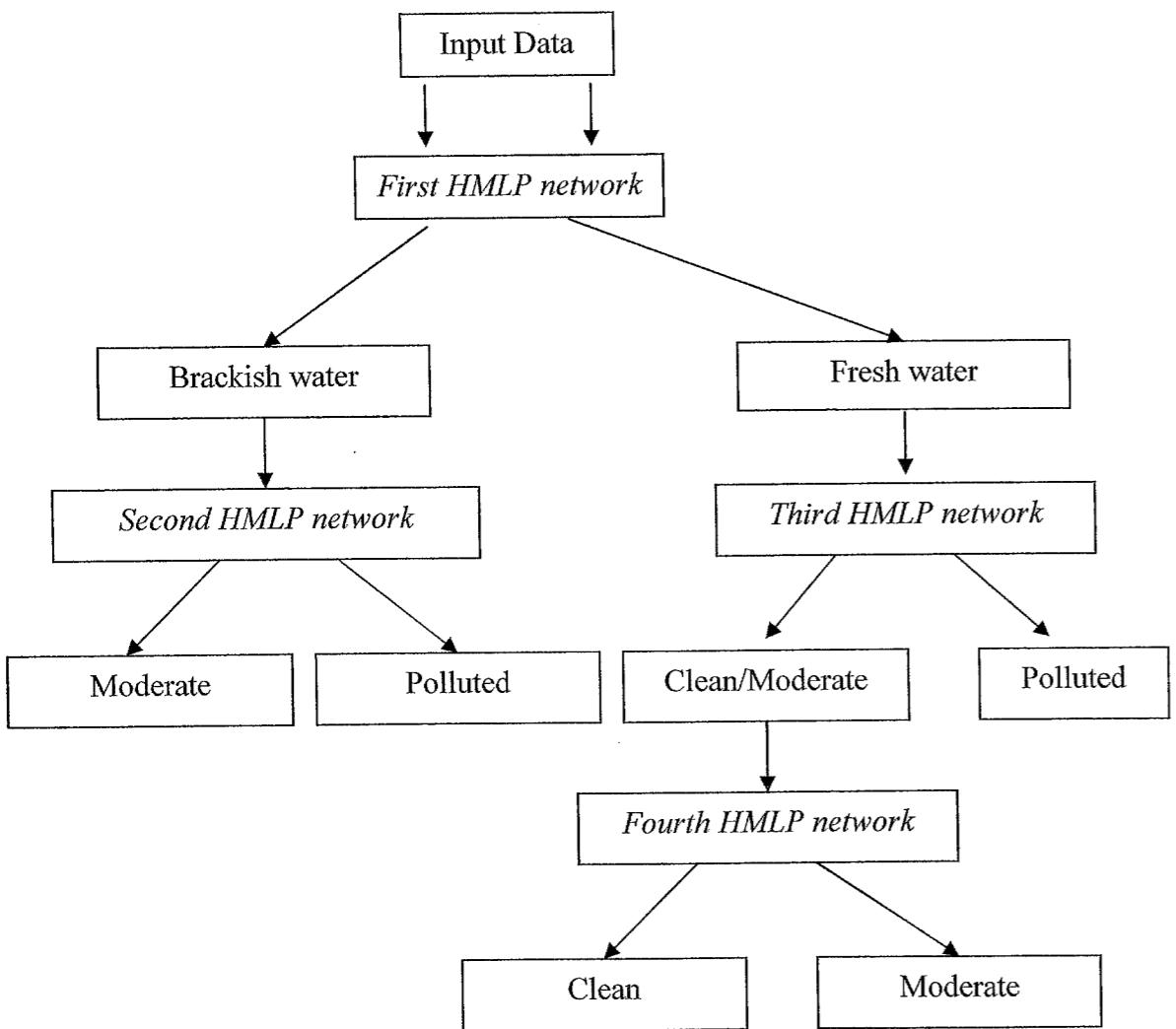


Figure 2: The H^2 MLP network

4. DATA SAMPLES

The effectiveness of the proposed intelligent classification system has been evaluated using 320 data (55, 63, 91, 48 and 63 data of fresh water with clean quality, fresh water with moderate quality, fresh water with polluted quality, brackish water with moderate quality and brackish water with polluted quality respectively). The H^2 MLP network was trained using 180 data (32, 32, 52, 32 and 32 data of fresh water with clean quality, fresh water with moderate quality, fresh water with polluted quality, brackish water with moderate quality and brackish water with polluted quality respectively) and tested using the remaining data (23, 31, 39, 16 and 31 data of fresh water with clean quality, fresh water with moderate quality, fresh water with polluted quality, brackish water with moderate quality and brackish water with polluted quality respectively). The data were taken from published paper by one of co-authors of this paper [25][26]. The number of cases for training was arbitrarily chosen and it was proven in this study and our previous study [24] that 180 data were significant enough to train the H^2 MLP network in order to produce the good results. The performance was compared with four standard ANNs (i.e. the MLP network trained using back propagation (BP), the MLP network trained using Lavenberg Marquadt (LM), the MLP network trained using Bayesian Regularization (BR) and the HMLP network trained using MRPE). Five separate HMLP networks were used to classify the fresh water with clean quality, fresh water with moderate quality, fresh water with polluted quality, brackish water with moderate quality and brackish water with polluted quality respectively. The capability of the proposed classification system was analyzed based on accuracy and mean squared error (MSE) analysis. The definition and procedure of the MSE analysis in [16][27] was closely followed.

As the classification process is done based on algae composition, 21 types of algae are used as input data for the ANNs. The 21 algae species are Acnanthes, Amphora Libica, Anomoeoneis Brachysira, Cocconeis, Coscinodiscus, Cyclotella, Cymbella, Diatoma, Diploaneis Decipiens, Eunotia, Fragilaria, Frustulia, Gomphonema Acuminatum, Hantzschia Amphioxys, Navicula, Neidium Affine, Nitzschia, Pinnularia, Psammothidium Bioretti, Stauroneis Obtusa and Surirella.

5 RESULTS AND DISCUSSION

The results for the H^2 MLP network were compared with the MLP and HMLP networks. Six tests were conducted with the number of hidden nodes (which form the structure of ANNs) ranged from 1 to 50. Table 2 depicts the percentage of accuracy, while Table 3 depicts the result for the MSE analysis. The H^2 MLP network performed better than other ANNs for each number of hidden nodes. It was able to achieve 97.11% accuracy with MSE value of 0.0093 even with small number of hidden nodes (i.e. one hidden node). Table 4 shows the optimum diagnostic performance for each ANNs when implemented with their optimum hidden nodes. The H^2 MLP network produced the best classification performance (i.e. 97.11% and 0.0091 of MSE value) and yielded the simplest network structure (i.e. needed 3 hidden nodes).

From Table 2, it is shown that the proposed hierarchical ANN (H^2 MLP) network was able to achieve better classification performance than the standard MLP and HMLP networks. For example, in each test, the H^2 MLP network outperformed the MLP network in term of the percentage of accuracy by more than 25.0%. In addition, the H^2 MLP network outperformed the HMLP network with different of accuracy percentage by more

than 4.5%. From Table 3, the H^2 MLP network was proven to converge to the smallest final MSE value as compared to the MLP and HMLP networks. The H^2 MLP network was capable to produce as small as 0.0091 of MSE value which was 1697.8%, 1736.3%, 1876.9% and 5229.7% smaller than that of the HMLP network, the MLP network with LM algorithm, the MLP network with BR algorithm and the MLP network with BP algorithm. The results also show that the H^2 MLP produced the similar percentage of accuracy and converged to the similar MSE value for different numbers of hidden nodes. These results indicate that the H^2 MLP network did not require a large number of hidden nodes. Typically, the optimum number of hidden nodes for the H^2 MLP network was less than ten and the performance of the H^2 MLP network was stable for different number of hidden nodes (i.e. from one to fifty hidden nodes). This is one advantage of the combination between hierarchical and hybrid structure of the H^2 MLP network as compared to conventional ANNs. The results demonstrate that the same findings as in our previous studies [24]. Table 2, 3 and 4 depict the overall results.

Table 2: Classification performance comparison in term of accuracy percentage for different number of hidden nodes between the MLP, HMLP and H^2 MLP networks.

Number of hidden nodes	Type of ANN				
	MLP with BP	MLP with BR	MLP with LM	HMLP	H^2 MLP
1	22.50	69.50	56.50	92.28	97.11
5	22.00	72.00	81.00	92.28	97.11
10	23.00	72.00	81.00	92.28	97.11
15	16.00	72.00	81.00	92.28	97.11

20	17.00	72.00	81.00	92.28	97.11
25	24.50	72.00	81.00	92.28	97.11
30	12.50	72.00	81.00	92.28	97.11
35	18.00	72.00	81.00	92.28	97.11
40	25.00	72.00	81.00	92.28	97.11
45	20.50	72.00	81.00	92.28	97.11
50	18.00	72.00	81.00	92.28	97.11

Table 3: Classification performance comparison in term of MSE analysis for different number of hidden nodes between the MLP, HMLP and H^2 MLP networks.

Number of hidden nodes	Type of ANN				
	MLP with BP	MLP with BR	MLP with LM	HMLP	H^2 MLP
1	0.6050	0.1915	0.1950	0.2050	0.0093
5	0.5825	0.1720	0.1650	0.1545	0.0091
10	0.5725	0.1708	0.1580	0.1545	0.0091
15	0.5600	0.1708	0.1580	0.1545	0.0091
20	0.5350	0.1708	0.1580	0.1545	0.0091
25	0.5225	0.1708	0.1580	0.1545	0.0091
30	0.5057	0.1708	0.1580	0.1545	0.0091
35	0.5000	0.1708	0.1580	0.1545	0.0091
40	0.4950	0.1708	0.1580	0.1545	0.0091
45	0.4850	0.1708	0.1580	0.1545	0.0091
50	0.4650	0.1708	0.1580	0.1545	0.0091

Table 4: Optimum classification performance for the MLP, HMLP and H^2 MLP networks.

	Type of ANN				
	MLP with BP	MLP with BR	MLP with LM	HMLP	H^2 MLP
Optimum hidden nodes	40	10	10	6	3
Accuracy	25.00	72.00	81.00	62.28	97.11
MSE analysis	0.4950	0.1708	0.1580	0.1545	0.0091

6 CONCLUSION

In this paper, an intelligent classification system has been proposed to determine the type and the stage of river water quality based on biological indicator (i.e. algae composition). The system employs the Hierarchical Hybrid Multilayered Perceptron (H^2 MLP) network to classify the river water into five categories, which are fresh water with clean quality, fresh water with moderate quality, fresh water with polluted quality, brackish water with moderate quality and brackish water with polluted quality. 21 species of algae are used as input data. The effectiveness of the proposed system has been demonstrated empirically using 320 data. The system successfully classifies the river water quality and type with accuracy up to 97.11% with small value of MSE (i.e. 0.0091).

Although the results obtained so far are encouraging, more investigations on both theoretical and practical aspects are needed to further vindicate the applicability of the proposed system to classify the river water quality.

ACKNOWLEDGEMENT

This work is supported under eScience Fund Grant entitled ‘Development of A New Architecture and Learning Algorithm of Artificial Neural Network for Determination of Potential Drug in Herbal Medicine’ and USM Short Term Grant entitled ‘Development of An Intelligent System for Classifying Quality of Water Based on Artificial Neural Network’.

REFERENCES

- [1]. F. Recknagel, M. French, P. Harkonen & K. I. Yakabuna. “Artificial Neural Network approach for modeling and prediction of algal blooms,” *Ecological Modeling*, 96, 1997, pp 11-28.
- [2] G. M. Hallegraeff. “Algal blooms and their apparent global increase,” *Phycologica*, 32(2), 1993, pp 79-99.
- [3] H. Wilson & F. Recknagel. “Toward a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes,” *Ecological Modeling*, 01, 2001, pp 69-84.
- [4] I. R. Falconer. *Alga toxin in seafood and drinking water.*, Academic Press: London, 1993.
- [5] E. C. D. Todd. “Domoic acid and amnesic shellfish poisoning.” *J.Food Protection*, 56(1), 1993, pp 69-83.
- [6] H. R. Maier, G. C. Dandy & M. D. Burch.. “Use of artificial neural networks for modeling cyanobacteria Anabaena spp. In the River Murray, South Australia,” *Ecol. Model*, 105, 1998, pp 257-272.
- [7] F. Recknagel. “ANNA – Artificial Neural Network Model for predicting species abundance and succession of blue-green algae,” *Hydrabiologia*, 349, 1997, pp 45-47.
- [8] F. Recknagel, T. Fukushima, T. Hanazato, N. Takamura & H. Wilson. “Modeling and predicting phyto and zooplankton dynamics in Lakes Kasumiagaura by artificial neural networks.,” *Lakes Reserve. Res. Manag*, 3, 1998, pp 45-47.
- [9] R. M. Holger & C. D. Graeme. “The use of artificial neural networks for the prediction of water quality parameters,” *Wat. Res. Res*, 32(4), 1996, pp 1013-1022.
- [10] N. Takamura, T. Iwakuma & M. Yassuna. “Primary production in Lake Kasumigaura”, 1981-1985,” *Jpn.J.Limnol*, 48, 1987, pp 19-38.
- [11] Department of Environment, Malaysia, “Environmental Management : Water Quality Monitoring-Biomonitoring”, *Biomonitoring of Sungai Langat : A Study Case*, Vol 2, 1998.
- [12] M. Mansor, L. P. Eng, S. C. Eng, N. W. Chan, S. M. Rawi, R. Rainis, O. H. Kadir & E. Tan, “Classification of Malaysian Rivers : Project on Water Pollution Control”, *A Study Case To Classify Rivers In Malaysia*. Vol 9, 1999.
- [13] Haykin, S. *Neural Network: A Comprehensive Foundation*, Prentice Hall: New Jersey, 1994.
- [14] Billings, S. A., “Identification of Nonlinear System – a survey”, *IEEE proceedings*, Vol 127, No. 6, pp 272-285, 1980.

- [15] Chen, H. W., "Modelling and Identification of Parallel Nonlinear System : Structural Classification and Parameter Estimation Method", *Proceeding of IEEE*, Vol 88, No. 1, 1995.
- [16] M. Y. Mashor, "Some Properties of RBF Networks with Applications to System Identification", *International Journal of The Computer, The Internet and Management*. Vol 7(1), pp 34-56, 1999.
- [17] H. S. Lee & D. Han., "Implementation of Real Time Color Gamut Mapping using Neural Network", *IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications*, University of Technology, Espoo, Finland, 2005.
- [18] S. H .Kim & S. K. Hak, "Graded Forecasting using An Array of Bipolar Predictions : Application of Probabilistic Neural Networks to a Stock Market Index", *International Journal of Forecasting*, Vol 14, No. 3, pp 323-337, 1998.
- [19] I. C. Yeh, "Structural Engineering Application with Augmented Neural Networks", *Computer-Aided Civil and Infrastructure Engineering*, Vol 13, Issue 2, pp 83-96, 1998.
- [20] J. G. Cai, J. Zhao & J. A. Hudson, "Computerization of Rock Engineering Systems using Neural Network with an Expert System", *International Journal of Rock Mechanics and Rock Engineering*, Vol 31, No. 3, pp 135-152, 1998.
- [21] M. Y. Mashor, "Nonlinear System Identification using HMLP Networks", *Journal of Institution of Engineers Malaysia*. Vol 60(2), pp 33-40, 1999.
- [22] S. J. Press & S. Wilson, " Choosing between Logistic Regression and Discriminant Analysis", *Journal of The American Statistic Association*, Vol 73, pp 699-705.
- [23] C. G. Wen. & C. S. Lee. "A Neural Network Approach to Multi-Objective Optimization for Water Quality Management in A River Basin," *Wat. Resources Res*, 34(3), 1998, pp 427-436.
- [24] N. A. Mat Isa, F.R .Hashim, F.W. Mei, D.A.Ramli, W. M. Wan Omar & K. Z. Zamli " Predicting Quality of River's Water Based on Algae Composition using Artificial Neural Network", IEEE International Conference on Industrial Informatics, 2006, pp 1340-1345
- [25] W. Maznah & M. Mansor. "Aquatic Pollution Assessment Based on Attached Diatom Communities in the Pinang River Basin, Malaysia," *Hydrobiologia*, 487, 2002, pp 229-241.
- [26] W. Maznah & M. Mansor. "Benthic Diatoms in the Pinang River (Malaysia) and Its Tributaries with Emphasis on Species Diversity and Water Quality," *International Journal on Algae*, 1(4), 1999, pp 103-108.
- [27]. M.Y. Mashor, "Hybrid Multilayered Perceptron," *International Journal of Systems Science*, Vol. 31 No.6, 2000, pp. 771-785.
- [28] Chen, S., Cowan, C.F. N., Billings, S. A., & Grant, P. M. "A Parallel Recursive Prediction Error Algorithm for Training Layered Neural Networks." *Int. J. of Control.* 51(6). pp. 1215-1228. 1990.
- [29] S. Chen, C.F.N. Cowan, S. A. Billings & P. M. Grant. "A parallel recursive prediction error algorithm for training layered neural networks, " *International Journal of The Computer, The Internet and Management*, Vol.11, No.1, 2004, pp 17-29.