

## KOMPARASI ANTARA ALGORITMA SUPPORT VECTOR MACHINE DAN ARTIFICIAL NEURAL NETWORK UNTUK PEMODELAN DATA TIME SERIES

Dadan Saepul Ramdan  
Teknik Informatika Politeknik TEDC Bandung  
E-mail: [ramdanplusplus@gmail.com](mailto:ramdanplusplus@gmail.com)

### Abstrak

Pemodelan *time series* merupakan salah satu subjek yang fundamental dan penting bagi beberapa penelitian. Banyak peneliti yang melakukan penelitian dalam subjek *time series* dalam beberapa tahun terakhir (Ratnadip & AgrawalAn, 2009). *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) telah diaplikasikan dengan sukses dalam banyak subjek penelitian dengan hasil generalisasi yang baik seperti *rule extraction*, *classification* dan *evaluation*. *Support Vector Machine* (SVM) adalah metode yang menjanjikan untuk prediksi *time series* karena menggunakan fungsi resiko dari *emperical error* dan *regularized term* yang berasal dari prinsip minimalisasi resiko struktural (Kyoung-jae, 2003). *Artificial Neural Networks* (ANN) adalah alat yang *powerful* yang didesain supaya mampu menyediakan solusi akurasi yang tinggi untuk permasalahan pemodelan di dunia nyata (Felicia, 2014). Penelitian ini bertujuan untuk membandingkan antara algoritma SVM dan ANN untuk pemodelan data *time series*.

Kata kunci : *time series*, *artificial neural networks* , *support vector machine*, *ann*, *svm*

### Abstract

*Time series modeling is one of the fundamental and important subject in some researches. Recently many researchers who do research in time series subject (Ratnadip & AgrawalAn, 2009). Support Vector Machine (SVM) and Artificial Neural Network (ANN) have been applied successfully in many research subjects with good generalization results as rule extraction, classification and evaluation. Support Vector Machine (SVM) is a promising method for the prediction of time series for use functionality emperical risk of error and the regularized term derived from the structural risk minimization principle (Kyoung-jae, 2003). Artificial Neural Networks (ANN) is a powerful tool that is designed to be able to provide high accuracy solution for modeling problems in the real world (Felicia, 2014). The purpose of this study is compare between SVM and ANN algorithms for modeling time series data.*

*Key words: time series, artificial neural networks , support vector machine, ann, svm*

### I. Pendahuluan

Pemodelan *time series* merupakan salah satu subjek yang fundamental dan penting bagi beberapa penelitian. Banyak peneliti yang melakukan penelitian dalam subjek *time series* dalam beberapa tahun terakhir. Banyak pemodelan dan metode yang diajukan untuk meningkatkan akurasi, sfisiensi dari pemodelan *time series* (Ratnadip & AgrawalAn, 2009).

Pemodelan *time series* adalah area penelitian dinamis yang telah menarik perhatian dari komunitas para peneliti lebih dari dua dekade terakhir. Tujuan utama dari pemodelan *time series* adalah untuk secara hati-hati dan ketat/teliti mempelajari observasi sebelumnya mengenai *time series* untuk mengembangkan model yang tepat yang menjelaskan struktur yang melekat atau karakteristik pada sebuah

*time series*. Model ini kemudian digunakan untuk menghasilkan *future value* untuk *time series*, seperti untuk membuat ramalan/*forecasts* (Ratnadip & AgrawalAn, 2009).

*Time series forecasting* ini dapat disebut sebagai tindakan prediksi masa depan dengan memahami masa lalu. Dikarenakan, sangat diperlukan dan pentingnya *time series forecasting* dalam berbagai bidang praktik, seperti bisnis, ekonomi, *finance*, *science* dan *engineering*, perhatian yang tepat seharusnya diambil agar cocok dan sesuai untuk memperhatikan *time series*. Jelas sekali bahwa suksesnya *time series forecasting* bergantung pada model yang sesuai (Ratnadip & AgrawalAn, 2009).

Banyak usaha yang telah dilakukan oleh para peneliti selama bertahun-tahun untuk mengembangkan model yang efisien untuk meningkatkan akurasi dari *time series forecasting*. Salah satu yang populer adalah *Autoregressive Integrated Moving Average* (ARIMA) (Box & Jenkins, 1970) (Hipel & McLeod, 1994) model. Asumsi dasar dibuat untuk menerapkan model ini adalah bahwa *time series* dianggap linier dan mengikuti statistik distribusi tertentu sebelumnya, seperti distribusi normal. *Time series forecasting* untuk musiman, Box dan Jenkins (Hamzacebi, 2008) telah mengajukan variasi dari model ARIMA yang cukup sukses, yang dikenal dengan *Seasonal ARIMA* (SARIMA) (Box & Jenkins, 1970) (Hamzacebi, 2008) (Hipel & McLeod, 1994).

Akhir-akhir ini *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) mendapatkan perhatian dari para peneliti untuk *time series forecasting* sehingga telah banyak diaplikasikan dengan sukses dalam banyak subjek penelitian dengan hasil generalisasi yang baik seperti *rule extraction*, *classification* dan *evaluation*.

Akhir-akhir ini SVM, sebuah algoritma jaringan saraf tiruan baru telah di kembangkan oleh Vapnik (Vapnik, 1988). *Support Vector Machine* (SVM) adalah metode yang menjanjikan untuk prediksi *time series* karena menggunakan fungsi resiko dari *empirical error* dan *regularized term* yang berasal dari prinsip minimalisasi resiko struktural (Kyoung-jae, 2003). Kebanyakan model dari jaringan saraf tiruan mengimplementasikan *empirical risk minimization principle*, sedangkan SVM mengimplementasikan *structural risk minimization principle* (Kyoung-jae, 2003). Penelitian sebelumnya untuk meminimalisir *mis-classification error* atau deviasi dari solusi yang sebenarnya

dengan data training, tetapi penelitian selanjutnya untuk meminimalisir batas atas dari *generalization error* (Kyoung-jae, 2003).

*Artificial Neural Networks* (ANN) adalah alat yang *powerful* yang didesain supaya mampu menyediakan solusi akurasi yang tinggi untuk permasalahan pemodelan di dunia nyata. Dalam hal ini, area penerapan *Artificial Neural Network* sangat luas. Diantaranya, seperti: *computer science*, *medical diagnosis*, *robotics* dan *heavy industry*, *quantitative finance*, *astronomy*, *physical science*, *quantum mechanics*. *Artificial Neural Networks*, contohnya telah digunakan untuk berbagai tujuan, seperti : *data processing*, *filtering*, *clustering*, *adaptive control*, *data mining*, *algorithm optimisation*, *systems engineering*, *data classification*, *pattern* dan *sequence recognition* (Felicia, 2014).

Penelitian ini bertujuan untuk membandingkan antara algoritma *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) untuk pemodelan data *time series*.

## II. Penelitian terkait

Banyak usaha yang telah dilakukan oleh para peneliti selama bertahun-tahun untuk mengembangkan model yang efisien untuk meningkatkan akurasi dari *time series forecasting*. Salah satu yang populer adalah *Autoregressive Integrated Moving Average* (ARIMA)[5, 7]model. Asumsi dasar dibuat untuk menerapkan model ini adalah bahwa *time series* dianggap linier dan mengikuti statistik distribusi tertentu sebelumnya, seperti distribusi normal. *Time series forecasting* untuk musiman, Box dan Jenkins (Hamzacebi, 2008) telah mengajukan variasi dari model ARIMA yang cukup sukses, yang dikenal dengan *Seasonal ARIMA* (SARIMA) (Box & Jenkins, 1970) (Hamzacebi, 2008) (Hipel & McLeod, 1994).

(Mingbao, Jiawei & Shiqiang, 2007) melakukan penelitian menggunakan SVM untuk melakukan *time series forecasting* yang diterapkan pada kepadatan cincin pertumbuhan kayu. Hasil dari penelitian tersebut membuktikan bahwa SVM memiliki performa yang lebih baik. Dalam aplikasinya SVM menerapkan prinsip *structural risk minimization* yang membuat generalisasinya lebih baik daripada teknik konvensional yang lain.

Wang Weiwei (Wang, 2007) melakukan penelitian untuk *time series prediction* menggunakan SVM yang digabungkan dengan GA (*genetic algorithm*). Ada beberapa keuntungan dari penerepan metode tersebut. Pertama, *high prediction performance*, karena *input space* dipartisi kedalam banyak region berdasarkan pada karakteristik data. Kedua, GA di

perkenalkan untuk menyelesaikan kombinasi parameter yang dilibatkan pada SVM. hasil dari simulasi memperlihatkan penggabungan SVM dan GA lebih efektif dalam *noisy forecasting* dan *non-stationary time series* daripada penggunaan *single SVM model*.

Sheng Lu, Zhong-jian Cai dan Xiao-bin Zhang (Lu, Cai & Zhang, 2009) menerapkan GA-SVM untuk *tax gross forecasting*. Pada model GA-SVM ini, GA digunakan untuk memilih parameter SVM yang cocok. Hasil percobaanya memperlihatkan bahwa GA-SVM dapat mencapai akurasi *forecasting* yang tinggi daripada *artificial neural network*.

**III. landasan teori**

**A. Time Series**

*Time series* adalah sebuah *sequential set data*, biasanya diukur beberapa kali. Dalam matematika, kita bisa definisikan sebagai sebuah vektor  $x(t)$ ,  $t=0,1,2,\dots$  dimana  $t$  adalah waktu,  $x(t)$  adalah random variabel (Hipel & McLeod, 1994).

Sebuah *time series* terdiri dari beberapa *record*, jika hanya terdiri dari satu variabel disebut *univariate*, dan jika terdiri dari lebih dari dua variabel atau lebih disebut *multivariate*. Sebuah *time series* bisa menjadi *continuous* atau *discrete*. Dalam *continuous*, observasi *time series* diukur disetiap *instance* waktu, sedangkan dalam *discrete*, observasi *time series* diukur di titik *discrete* waktu. Contohnya pembacaan cuaca, aliran sungai dapat di kategorikan sebagai *continuous time series*. Pada sisi lain, populasi dari kota tertentu, produksi dari sebuah perusahaan, nilai tukar antara dua mata uang dikategorikan sebagai *discrete time series*. Biasanya pada *discrete time series* observasi yang berturut-turut dicatat pada interval waktu, seperti jaman, harian, mingguan, bulanan, tahunan (Ratnadip & AgrawalAn, 2009).

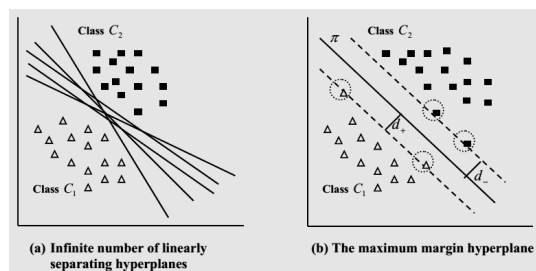
**B. Support Vector Machine (SVM)**

SVM dikembangkan oleh Vapnik dan kawan-kawan di laboratorium AT & T Bell pada tahun 1995 (Vapnik, 1988). Pada dasarnya SVM didesain untuk memecahkan masalah *pattern classification*, seperti *optimal character recognition*, *face identification*, *text classification* dan lain-lain. Tetapi sekarang ini, para peneliti menemukan aplikasi yang lebih lebar pada domain lain, seperti *function approximation*, *regression estimation* and *time series prediction problems* (Cao & Francis, 2003)

SVM bekerja berdasarkan pada prinsip *Structural Risk Minimization* (SRM), tujuan dari SVM adalah

untuk mencari decision rule dengan kemampuan generalisasi yang bagus melalui pemilihan beberapa subset tertentu dari *training data*, yang disebut *support vectors* (Vapnik, 1988). Karakteristik penting yang lain dari SVM adalah bahwa training process sama dengan pemecahan *linearly constrained quadratic programming problem*. Jadi, sebaliknya untuk *network training* lainnya, SVM selalu unik dan secara umum optimal. Bagaimanapun juga salah satu kekurangan dari SVM adalah ketika ukuran training besar, memerlukan jumlah komputasi yang besar yang meningkatkan jumlah waktu waktu (Cao & Francis, 2003).

Ide utama dari SVM ketika diaplikasikan ke masalah binary classification adalah canonical hyperplane yang secara maksimal memisahkan dua class yang diberikan dari training sample[8, 9]. Mari kita memperhitungkan dua set data point yang dapat dipisahkan secara linier dalam  $\mathbb{R}^n$  yang harus di klasifikasikan kedalam satu dari dua class  $C_1$  dan  $C_2$  menggunakan linier hyperplane (straight lines). Dari jumlah yang tak terbatas dari pemisahan hyperplane, satu dengan margin maksimum harus dipilih untuk classification dan generalization terbaik.



Gambar 2.1 Support Vector Machine Model

Pada gambar 2.1(a) dapat dilihat bahwa ada jumlah yang tak terbatas dari *hyperplanes*, pemisahan *training data point*. Seperti terlihat pada gambar 3.1(b)  $d_+$  dan  $d_-$  menunjukkan jarak tegak lurus dari pemisahan *hyperplanes*, ke arah *data point* paling dekat dari  $C_1$  dan  $C_2$  secara masing-masing. Kemudian kedua dari jarak tegak lurus  $d_+$  dan  $d_-$  disebut sebagai margin dan total margin tersebut adalah  $M = d_+ + d_-$ . Untuk klasifikasi yang akurat dan juga untuk generalisasi terbaik, *hyperplanes* yang memaksimalkan total margin dianggap sebagai yang optimal dan dikenal sebagai Maximum Margin Hyperplane (Vapnik, 1988). Hyperplanes ini  $d_+ = d_-$ . *Data point* dari kedua *class* yang terdekat ke maximum margin *hyperplanes* dikenal sebagai *Support Vector*[33]. Pada gambar (b)  $n$  menunjukkan optimal *hyperplanes* dan *circulated data point* dari kedua class mewakili *support vectors*.

### C. Artificial Neural Network (ANN)

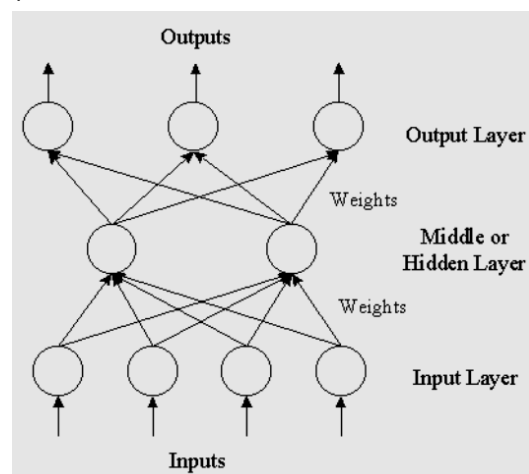
Algoritma *Artificial Neural Networks* (ANN) telah diusulkan sebagai teknik alternatif untuk *time series forecasting* dan memperoleh popularitas yang besar dalam beberapa tahun terakhir (Ratnadip & AgrawalAn, 2009). Tujuan dasar dari ANN adalah untuk membangun sebuah model untuk meniru kecerdasan otak manusia kedalam mesin (Kamruzzaman, Begg & Sarker). Mirip dengan kerja otak manusia, ANN mencoba untuk mengenali keteraturan dan pola pada *input data*, belajar dari pengalaman dan menyediakan hasil dasar yang digeneralisasi pada pengetahuan sebelumnya (Ratnadip & AgrawalAn, 2009). Walaupun pengemabangan ANN termotivasi secara biologis, tapi setelah itu, ANN telah diaplikasikan dalam banyak area yang berbeda, khususnya untuk berbagai tujuan *forecasting* dan *classification* (Kamruzzaman, Begg & Sarker). Dibawah ini beberapa fitur yang menonjol dari ANN yang membuatnya cukup paforit untuk analisa *time series* dan *forecasting*.

Pertama, ANN adalah *data-driven* dan *self-adaptive* yang natural (Ma & Naimin, 2007). Tidak perlu untuk menentukan sebuah bentuk model tertentu atau untuk membuat *priori assumption* apapun mengenai *statistical distribution* dari sebuah data; model yang diinginkan terbentuk berdasarkan *features presented* dari data. Pendekatan ini cukup berguna untuk beberapa situasi tertentu, dimana tidak ada panduan teori yang tersedia untuk proses pembuatan data yang sesuai.

Kedua, ANN adalah *non-linier*, yang membuatnya lebih praktis dan akurat dalam pemodelan pola data yang kompleks, sebagai lawan dari berbagai pendekatan tradisional yang *linier*, seperti metode ARIMA (Kamruzzaman, Begg & Sarker) (Zhang, Patuwo & Hu, 1998). Ada banyak *instance* yang mengusulkan bahwa ANN membuat analisa dan *forecasting* yang cukup lebih baik daripada berbagai model *linier*.

Terakhir, ANN telah menunjukkan bahwa jaringan dapat mengira-ngira *continuous function* apapun untuk berbagai akurasi yang diinginkan (Zhang, Patuwo & Hu, 1998). ANN menggunakan proses paralel untuk memproses informasi dari data untuk mengira-ngira sejumlah *class* dari fungsi dengan nilai akurasi yang tinggi. Lebih jauh lagi, ANN dapat sepakat dengan situasi, dimana *input* datanya *erroneous*, *incomplete* atau *fuzzy* (Kamruzzaman, Begg & Sarker).

ANN yang paling banyak digunakan dalam masalah *forecasting* adalah *multi-layer perceptrons* (MLPs) yang menggunakan *single hidden layer feed forward network* (FNN) (Zhang, Patuwo & Hu, 1998). Model ini tandai oleh network dari tiga *layer*; *input*, *hidden* dan *output*, yang dihubungkan oleh *acyclic link*. Ada kemungkinan lebih dari satu *hidden layer*. *Node* di berbagai layer juga dikenal sebagai *processing element* (Ratnadip & AgrawalAn, 2009). Arsitektur *three-layer feed forward* dari ANN dapat digambarkan seperti dibawah ini.



Gambar 2.2 Arsitektur *Artificial Neural Network*

## IV. Metodologi penelitian

### A. Metode Penelitian

Penelitian ini adalah penelitian eksperimen dengan metode penelitian sebagai berikut :

#### 1) Penentuan Masalah

Penentuan masalah penelitian adalah dengan menggunakan studi literatur dan studi lapangan berdasarkan pada jurnal-jurnal serta buku-buku yang berhubungan dengan *time series data*, *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN).

#### 2) Penentuan *Computing Approach*

Pada penelitian ini pendekatan yang digunakan adalah *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) karena dari hasil studi literatur, penerapan kedua metode tersebut cocok diterapkan pada data *time series*.

### B. Metode Pengumpulan Data

Penelitian ini menggunakan dataset yang sudah ada (Akbilgic, Bozdogan & Balaban, 2013) yang berkaitan dengan *Istanbul Stock Exchange* dan dapat download secara gratis dengan mengunjungi halaman berikut:

<https://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE>.

C. Eksperimen dan Pengujian

Experimen dan pengujian pada penelitian ini terdiri dari dua tahapan, diantaranya :

- 1)Melakukan pengujian data set dengan menggunakan *tool RapidMiner 5.2*, kemudian memilih algoritma SVM untuk melakukan proses pemodelan data.
- 2)Melakukan pengujian data set dengan menggunakan *tool RapidMiner 5.2*, kemudian memilih algoritma ANN untuk melakukan proses pemodelan data.

D. Evaluasi dan Validasi Hasil

Penelitian ini menggunakan sebuah *tool* yakni *RapidMiner*, *tool* ini akan memperlihatkan hasil dari penelitian. Evaluasi dan validasi hasil dari penelitian ini akan dilakukan dengan melihat presentase RMSE, SE dan *execution time* dari SVM dan ANN.

V. Pengukuran

A. Dataset yang Digunakan

Penelitian ini menggunakan dataset yang sudah ada (Akbulgic, Bozdogan & Balaban, 2013) yang berkaitan *Istanbul Stock Exchange* dan dapat download secara gratis dengan mengunjungi halaman berikut:

<https://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE>. Date set ini mempunyai jumlah instance sebanyak 536 *instance*, dan jumlah atribut sebanyak 8 atribut. Data set ini mengandung informasi *returns of Istanbul Stock Exchange* dan 7 international *index* lainnya: SP, DAX, FTSE, NIKKEI, BOVESPA, MSCE\_EU, MSCI\_EM dengan rentang waktu 5 Juni 2009 sampai 22 Februari 2011. Dibawah ini table yang menjelaskan mengenai data set tersebut.

Tabel 5.1 Variabel/Atribut Dataset

Nama Variabel/Atribut	Deskripsi
ISE	<i>Istanbul stock exchange national 100 index</i>
SP	<i>Standard &amp; poor's 500 return index</i>
DAX	<i>Stock market return index of Germany</i>
FTSE	<i>Stock market return index of UK</i>
NIKKEI	<i>Stock market return index of Japan</i>
BOVESPA	<i>Stock market return index of Brazil</i>
EU	<i>MSCI European index</i>
EM	<i>MSCI emerging markets index</i>

B. Pengukuran

Pengukuran dilakukan dengan menggunakan sebuah *tool* yakni *RapidMiner 5.2* dengan menggunakan 2 *operator* yang berbeda SVM dan ANN untuk melakukan proses pemodelan data. Dibawah ini adalah tabel hasil pengukuran tersebut.

Tabel 4.2 Hasil Pengukuran

Metode	RMSE	SE	Time
SVM	0.014 +/- 0.002	0	6 detik
ANN	0.015 +/- 0.002	0	10 detik

Dari tabel 4.2 diatas kita bisa melihat bahwa SVM mempunyai RMSE = 0.014 +/- 0.002, SE = 0 dan waktu selama 6 detik. Sedangkan ANN mempunyai RMSE = 0.015 +/- 0.002, SE = 0 dan waktu selama 10 detik.

VI. Penutup

A. Kesimpulan

Kesimpulan dari penelitian ini adalah bahwa algoritma *Support Vector Machine (SVM)* mempunyai performa lebih baik dibandingkan dengan algoritma *Artificial Neural Network (ANN)* ketika diterapkan pada *time series dataset* dengan jumlah *instance* 536 dan jumlah atribut 8. Berdasarkan pada hasil pengukuran (table 4.2) SVM memiliki RMSE, SE dan *execution time* yang lebih kecil dibandingkan dengan ketika menggunakan ANN.

B. Saran

Saran dari penelitian ini, untuk kedepannya diharapkan dapat menggunakan *dataset* yang lebih kompleks dan lebih besar, sehingga didapatkan nilai perbandingan yang lebih akurat. Untuk mendapatkan *performance* yang lebih akurat, pada penelitian selanjutnya bisa menggunakan teknik penggabungan algoritma. Sebagai contoh penggabungan SVM dan *Genetic Algorithm (GA)*, dalam hal ini GA bisa digunakan untuk memilih parameter terbaik pada SVM.

Daftar Pustaka

Ratnadip Adhikari, R. K. AgrawalAn, "Introductory Study on Time Series Modeling and Forecasting", 2009.

Kyoung-jae Kim, "Financial Time Series Forecasting Using Support Vector Machines", Elsevier, 2003

Felicia Ramona Birau, "Forecasting Financial Time Series Based On Artificial Neural Networks", Impact Journal, 2014.

Akbulgic, O., Bozdogan, H., Balaban, M.E., "A novel Hybrid RBF Neural Networks Model As A Forecaster", Statistics and Computing, 2013 DOI 10.1007/s11222-013-9375-7.

- G.E.P. Box, G. Jenkins, "*Time Series Analysis, Forecasting and Control*", Holden-Day, San Francisco, CA, 1970
- C. Hamzacebi, "*Improving Artificial Neuralnetworks' Performance In Seasonal Time Series Forecasting*", *Information Sciences* 178 (2008), pages: 4550-4559.
- K.W. Hipel, A.I. McLeod, "*Time Series Modelling of Water Resources and Environmental Systems*", Amsterdam, Elsevier 1994
- V.N. Vapnik, "*Statistical Learning Theory*", Wiley, New York, 1998.
- L.J. Cao and Francis E.H. Tay "*Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting*", *IEEE Transaction on Neural Networks*, Vol. 14, No. 6, November 2003, pages: 1506-1518.
- Joarder Kamruzzaman, Rezaul Begg, Ruhul Sarker, "*Artificial Neural Networks in Finance and Manufacturing*", Idea Group Publishing, USA.
- G. Zhang, B.E. Patuwo, M.Y. Hu, "*Forecasting With Artificial Neural Networks: The State Of The Art*", *International Journal of Forecasting* 14 (1998), pages: 35-62.
- Mingbao Li, Jiawei Zhang, Shiqiang Zheng, "*Comparison Study on Forecasting of Timber Growth Ring Density with SVM and Neural Networks*", *IEEE*, 2007.
- Wang Weiwei, "*Time Series Prediction Based on SVM and GA*", *IEEE*, 2007.
- Sheng Lu, Zhong-jian Cai, Xiao-bin Zhang, "*Application of GA-SVM Time Series Prediction in Tax Forecasting*", *IEEE*, 2009
- L. Ma and L. Naimin, "*Texture Feature Extraction and Classification for Iris Diagnosis*," *ICMB*, 2007.