

INTERNATIONAL JOINT  
CONFERENCE ON  
NEURAL NETWORKS

VOL. I

# IJCNN-91-SEATTLE

International  
Joint  
Conference on  
Neural Networks

July 8-15 1991  
Washington State  
Convention & Trade Center  
Seattle, WA

***Volume I***

---

Co-sponsored by the  
International Neural Network Society  
and the  
Institute of Electrical and Electronics Engineers, Inc.

# TABLE OF CONTENTS

---

## VOLUME I

### APPLICATIONS I *Oral Presentations*

- The Early Diagnosis of Heart Attacks: A Neurocomputational Approach** 1-1  
R. Harrison, S. Marshall, R. Kennedy, *University of Sheffield*
- Particle Tracking by Deformable Templates** 1-7  
A. Yuille, K. Honda, C. Peterson, *Harvard University Div. of Applied Sciences*
- The PI-sigma Network: An Efficient Higher-Order Network for Pattern Classification and Function Approximation** 1-13  
J. Ghosh, Y. Shin, *The University of Texas at Austin*
- A Dynamic Thesaurus and Its Application to Associated Information Retrieval** 1-19  
H. Kimoto, T. Iwadera, *Nippon Telegraph and Telephone*
- Data Association in Multi-Target Tracking: A Solution using a Layered Boltzmann Machine** 1-31  
R. Iltis, P. Ting, *University of California, Santa Barbara*
- Self-organization of Non-numeric Data Sets** 1-37  
D. Bradburn, *TRW Financial Systems*
- Associative Switch for Combining Multiple Classifiers** 1-43  
L. Xu, A. Krzyzak, C. Suen, *Universite Concordia*

### APPLICATIONS II *Oral Presentations*

- Study of Continuous ID3 and Radial Basis Function Algorithms for the Recognition of Glass Defects** 1-49  
K. Cios, R. Tjia, N. Liu, R. Langenderfer, *University of Toledo*
- Higher Order Data Compression with Neural Networks** 1-55  
A. Namphol, M. Arozullah, S. Chin, *The Catholic University of America*
- Neural Network Training Schemes for Non-Linear Adaptive Filtering and Modelling** 1-61  
O. Nerrand, P. Roussel-Ragot, L. Personnaz, G. Dreyfus, *ESPCI*; S. Marcos, O. Macchi, C. Vignat, *Laboratoire des Signaux et Systemes*
- A Parallel Non-Neural Trigger Tracker for the SSC** 1-67  
R. Farber, W. Kennison, A. Lapedes, *Los Alamos National Laboratory*
- Robustness Test of an Incipient Fault Detector Artificial Neural Network** 1-73  
M. Chow, S. Yee, *North Carolina State University*
- Self-Organizing Network for Regression: Efficient Implementation and Comparative Evaluation** 1-79  
V. Cherkassky, H. Lari-Najafi, Y. Lee, *University of Minnesota*
- Probing the Critic: Approaches to Connectionist Pattern Synthesis** 1-85  
J. Lewis, *New York Institute of Technology*

### APPLICATIONS III *Oral Presentations*

- A New Approach to the Design of Neural Network Classifiers and Its Application to the Automatic Recognition of Handwritten Digits** 1-91  
S. Knerr, L. Personnaz, G. Dreyfus, *Ecole Superieure de Physique et de Chimie Industrielles*
- Classification of Hand-Written Digits and Japanese Kanji** 1-97  
T. Vogl, K. Blackwell, G. Barbour, *Environmental Research Institute of Michigan*; S. Hyman, D. Alkon, *Laboratory for Molecular and Cellular Neurobiology*

# TABLE OF CONTENTS CONTINUED

## VOLUME I

- Distorted Handwritten Kanji Character Pattern Recognition by a Learning algorithm Minimizing Output Variation** I-103  
Y. Kimura, *Nippon Telegraph & Telephone Corporation*

- Unsupervised Context Learning in Natural Language Processing** I-107  
J. Scholtes, *University of Amsterdam*

- Classification of Acoustic Emission Signals Via Hebbian Feature Extraction** I-113  
J. Yang, G. Dumont, *The University of British Columbia*

- A Hybrid Neural Network Classifier of Short Duration Acoustic Signals** I-119  
S. Beck, L. Dreuser, R. Still, J. Whiteley, *Tracor Applied Sciences*

- Sound Localization with a Neural Network Trained with the Multiple Extended Kalman Algorithm** I-125  
F. Palmieri, M. Datum, A. Shah, A. Moiseff, *The University of Connecticut*

## APPLICATIONS IV *Oral Presentations*

- An Effective Steiner Tree Construction for VLSI Routing** I-133  
A. Kahng, *University of California, Los Angeles*

- Using a Semi-Asynchronous Hopfield Network to Obtain Optimal Coverage in Logic Minimization** I-141  
P. Chu, *Cleveland State University*

- A Time-varying Recurrent Neural System for Convex Programming** I-147  
J. Wang, *University of North Dakota*

- An Algorithm-structured Neural Net for the Shortest-path Problem** I-153  
G. Fahner, *University of Dusseldorf*

- Solving the Assignment Problem with Statistical Physics** I-159  
J. Kosowsky, A. Yuille, *Harvard University*

- Solving Two-point Boundary Value Problems in Trajectory Formation** I-165  
H. Li, C. Chen, G. Lakhani, *Texas Tech University*

- Approximation Property of Multi-layer Neural Network (MLNN) and its Application in Nonlinear Simulation** I-171  
Y. Yu, W. Zhang, *South China University of Technology*; Q. Jia, *Xi'an Institute of Space Radio Technology*

## APPLICATIONS I *Poster Presentations*

- Application of Neural Computation to Sound Analysis for Valve Diagnosis** I-177  
K. Arai, *Mitsubishi Kasei Corporation*; H. Shimodaira, Y. Sakaguchi, K. Nakano, *University of Tokyo*

- Fuzzy Associative Memory in Conceptual Design** I-183  
A. Bahrami, C. Dagli, *University of Missouri-Rolla*; B. Modarress, *Boeing Commercial Airplane Group*

- Neural Network Classification of Metal Surface Properties Using a Dynamic touch Sensor** I-189  
D. Brenner, J. Principe, K. Doty, *University of Florida*

# CONTINUED TABLE OF CONTENTS

## VOLUME I

<b>Neural Networks for Second-order Medical Tasks</b> D. Brown, <i>Food and Drug Administration</i>	I-195
<b>Feature Deactivation Using Partial Inhibitory Networks During Multiple Object Recognition</b> L. Chan, <i>The Chinese University of Hong Kong</i>	I-199
<b>A Neural Network Approach to Machine Vision Systems for Automated Industrial Inspection</b> T. Cho, R. Conners, <i>The Bradley Dept of Electrical Engineering</i>	I-205
<b>An Auto-adaptive Synthetic Neural Network for Real-time Separation of Independent Signal Sources</b> M. Cohen, P. Pouliquen, A. Andreou, <i>The Johns Hopkins University</i>	I-211
<b>A Neural Network Approach to the Partial Shape Classification Problem</b> L. Gupta, A. Upadhye, <i>Southern Illinois University at Carbondale</i>	I-215
<b>Sensor Failure Detection and Recovery by Neural Networks</b> T. Guo, J. Nurre, <i>NASA</i>	I-221
<b>Improved Detection of Biological Substances Using a Hybrid Neural Network and Infrared Absorption Spectroscopy</b> F. Ham, G. Cohen, B. Cho, <i>Florida Institute of Technology</i>	I-227
<b>Classification of Japanese Kanji Using Principal Component Analysis as a Preprocessor to an Artificial Neural Network</b> S. Hyman, D. Alkon, <i>Laboratory for Molecular and Cellular Neurobiology</i> ; T.P. Vogl, K.T. Blackwell, G.S. Barbour, J.M. Irvine, <i>Environmental Research Institute of Michigan</i>	I-233
<b>Comparison of Neural Network Models for Process Fault Detection and Diagnosis Problems</b> P. Jokinen, <i>NESTE Technology</i>	I-230
<b>Prototype Neural Network Processor for Multispectral Image Fusion</b> J. Kagel, J. Reeder, <i>McDonnell Douglas Electronics Systems Co.</i>	I-245
<b>Acoustic Diagnosis for Compressor with Hybrid Neural Network</b> M. Kotani, H. Matsumoto, <i>Kobe University</i> ; T. Kanagawa, <i>Osaka Gas Co.</i>	I-251
<b>Integration of Neural Networks and Decision Tree Classifiers for Automated Cytology Screening</b> J. Lee, J. Hwang, D. Davis, A. Nelson, <i>University of Washington</i>	I-257
<b>A Neural Piecewise Linear Classifier for Pattern Classification</b> Z. Lo, B. Bavarian, <i>University of California, Irvine</i>	I-263
<b>Fast Diagnosis of Integrated Circuit Faults using Feedforward Neural Networks</b> J. Meador, A. Wu, C. Tseng, T. Lin, <i>Washington State University</i>	I-269
<b>A New Self-organizing Method and Its Application to Handwritten Digit Recognition</b> T. Shimada, K. Nishimura, K. Haruki, <i>Toshiba Corp., Systems &amp; Software Engineering Lab.</i>	I-275
<b>A Neural Network Approach to Cloud Detection in AVHRR Images</b> O. Slawinski, J. Kowalski, P. Cornillon, <i>University of Rhode Island</i>	I-283

# TABLE OF CONTENTS CONTINUED

---

## VOLUME I

### APPLICATIONS II *Poster Presentations*

<b>A Commodity Trading Model Based on a Neural Network-expert System Hybrid</b> K. Bergerson, D. Wunsch, <i>Neural Trading Company</i>	I-289
<b>Power Series Analyses of Back-propagation Neural Networks</b> M. Chen, M. Manry, <i>University of Texas at Arlington</i>	I-295
<b>Recurrent Neural Networks and Time Series Prediction</b> J. Connor, L. Atlas, <i>University of Washington</i>	I-301
<b>Linearized Least-squares Training of Multilayer Feedforward Neural Networks</b> S. Douglas, T. Meng, <i>Stanford University</i>	I-307
<b>Development of a Hopfield Network for Solving Integral Equations</b> I. Elshafey, L. Upda, S.S. Udpa, <i>Iowa State University</i>	I-313
<b>Use of Neural Networks to Predict Lightning at Kennedy Space Center</b> D. Frankel, I. Schiller, J. Draper, <i>KTAADN, Inc</i> ; A. Barnes, <i>Geophysics Laboratory</i>	I-319
<b>Team Theory and Back-propagation for Dynamic Routing in Communication Networks</b> G. Frisiani, T. Parisini, L. Siccardi, R. Zoppoli, <i>DIST - University of Genova</i>	I-325
<b>A Diffusion Process for Global Optimization in Neural Networks</b> T. Guillermin, N. Cotter, <i>The University of Utah</i>	I-335
<b>Genetic Optimization of Self-organizing Feature Maps</b> T. Samad, S. Harp, <i>Honeywell SSDC</i>	I-341
<b>Forecasting Economic Turning Points with Neural Nets</b> R. Hoptroff, T. Hall, <i>Kings College London</i> ; M.J. Bramson, <i>Clarendon Consultants</i>	I-347
<b>Should Back Propagation Be Replaced by More Effective Optimization Algorithms?</b> J. Hsiung, W. Suewatanakul, D. Himmelblau, <i>University of Texas</i>	I-353
<b>Self-organizing Multi-layer Semantic Maps</b> H. Ichiki, M. Hagiwara, M. Nakagawa, <i>Keio University</i>	I-357
<b>Recurrent Neural Network to Acquire the Grammatical Competence</b> R. Kamimura, <i>Tokai University</i>	I-361
<b>A Continuous-time Inference Network for Minimum-cost Path Problems</b> K. Lam, <i>The University of British Columbia</i>	I-367
<b>Application of Feedforward and Recurrent Neural Networks to Chemical Plant Predictive Modeling</b> R. Hecht-Nielsen, <i>HNC, Inc</i> ; J. Lambert, <i>Institut Francais du Petrole</i>	I-373
<b>A Maximum Neural Network for the Max Cut Problem</b> K. Lee, Y. Takefuji, N. Funabiki, <i>Case Western Reserve University</i>	I-379
<b>Self-organization Via Competition, Cooperation and Categorization Applied to Extended Vehicle Routing Problems</b> Y. Matsuuyama, <i>Ibaraki University</i>	I-385
<b>A Comparison of Two Neural Network Architectures for Vector Quantization</b> M. Naraghi-Pour, M. Hegde, F. Bourge, <i>Louisiana State University</i>	I-391

## CONTINUED TABLE OF CONTENTS

### VOLUME I

<b>Dynamic Programming Approach For Multilayer Neural Network Optimization</b> <i>P. Saratchandran, Nanyang Technological Institute</i>	I-397
<b>Data Normalization with Self-organizing Feature Maps</b> <i>A. Ultsch, G. Halmans, Universitat Dortmund</i>	I-403
<b>Optimization by Extended LVQ</b> <i>T. Yoshihara, T. Wada, Olympus Optical Co., Ltd.</i>	I-407

### ELECTRONIC & OPTICAL IMPLEMENTATIONS I *Orals*

<b>An Analog Neural Network Processor and its Application to High-Speed Character Recognition</b> <i>B. Boser, E. Sackinger, J. Bromley, Y. LeCun, R. Howard, L. Jackel, AT&amp;T Bell Laboratories</i>	I-415
<b>A Neural Architecture for the Assignment Problem: Simulation and VLSI Implementation</b> <i>S. Eberhardt, T. Daud, D. Kerns, R. Tawel, A. Thakoor, Jet Propulsion Laboratory, California Institute of Technology</i>	I-421
<b>A Real-time VLSI Neuroprocessor for Adaptive Image Compression Based Upon Frequency-sensitive Competitive Learning</b> <i>W. Fang, B. Sheu, S. Chen, University of Southern California at Los Angeles</i>	I-429
<b>CMOS Implementation of Analog Hebbian Synaptic Learning Circuits</b> <i>C. Schneider, H. Card, The University of Manitoba</i>	I-437
<b>Architecture of a General-purpose Neural Signal Processor</b> <i>U. Ramacher, J. Beichter, N. Bruls, Siemens AG</i>	I-443
<b>A Parallel Analog CCD/CMOS Neural Network IC</b> <i>C. Neugebauer, A. Yariv, California Institute of Technology</i>	I-447
<b>Neural Network LSI Chip with On-chip Learning</b> <i>T. Furuta, H. Eguchi, H. Horiguchi, S. Oteki, T. Kitaguchi, RICOH Co.</i>	I-453

### ELECTRONIC & OPTICAL IMPLEMENTATIONS II *Orals*

<b>Learning Algorithms for Optical Multilayer Neural Networks</b> <i>Y. Qiao, D. Psaltis, California Institute of Technology</i>	I-457
<b>Design and Performance of a Prototype General Purpose Analog Neural</b> <i>P. Mueller, J. Spiegel, V. Agami, P. Aziz, D. Blackman, P. Chance, University of Pennsylvania and Corticon Inc.</i>	I-463
<b>A Multiprocessor Machine for Large-scale Neural Network Simulation</b> <i>J. Elias, M. Fisher, C. Monemi, University of Delaware</i>	I-469
<b>"Bump" Circuits for Computing Similarity and Dissimilarity of Analog Voltages</b> <i>T. Delbruck, California Institute of Technology</i>	I-475
<b>A Neural-net Board System for Machine Vision Applications</b> <i>H. Graf, R. Janow, C. Nohl, J. Ben, AT&amp;T Bell Laboratories</i>	I-481
<b>Analog Programmable Chips for Implementing ANNs Using Capacitive Weight Storage</b> <i>F. Kub, K. Moon, J. Modolo, Naval Research Laboratory</i>	I-487



# TABLE OF CONTENTS CONTINUED

## VOLUME I

- A VLSI Implementation of a Neural Network Car Collision Avoidance Controller** I-493  
J. Nijhuis, B. Hofflinger, S. Neuber, A. Siggelkow, L. Spaanenburg, *Institute for Microelectronics  
Stuttgart*

## ELECTRONIC & OPTICAL IMPLEMENTATIONS *Posters*

- Discarding Outliers Using a Nonlinear Resistive Network** I-501  
S.-C. Liu, B. Mathur, *Rockwell Science Center*, J.G. Harris, *California Institute of  
Technology*

- An Adaptive CMOS Matrix-vector Multiplier for Large Scale Analog Hardware Neural Network  
Applications** I-507  
G. Cauwenberghs, C. Neugebauer, Amnon Yariv *California Institute of Technology*

- The Lockheed Probabilistic Neural Network Processor** I-513  
T. Washburne, M. Okamura, D. Specht, W. Fisher, *Lockheed Missiles & Space Co.*

- Finite Precision Error Analysis of Neural Network Electronic Hardware Implementations** I-519  
J. Hwang, J. Holt, *University of Washington*

- Design of a Scalable Opto-electronic Neural System Using Free-space Optical Interconnects** I-527  
A. Krishnamoorthy, G. Yayla, S. Esener, *University of California, San Diego*

- Experimental Demonstration of Large-scale Holographic Optical Neural Network** I-535  
T. Lu, F. Lin, *Physical Optics Corporation*

- An Optoelectronic Adaptive Resonance Unit** I-541  
D. Wunsch, T. Caudell, C. Capps, R. Falk, *The Boeing Company*

- Into Silicon: Real Time Learning in a High Density RBF Neural Network** I-551  
C. Scofield, D. Reilly, *Nestor Inc*

- Neural Circuit Architectures for Real-time Signal Processing in Video Rate Communication  
Systems** I-557  
S. Bibyk, R. Kaul, K. Adkins, Z. Bhatti, *The Ohio State University*

- A Continuous Time Synapse Employing a Refreshable Multilevel Memory** I-563  
P. Hasler, L. Akers, *Arizona State University*

- Distributing Back Propagation Networks Over the Intel iPSC/860 Hypercube** I-569  
D. Jackson, *Mentor Graphics Corporation*; D. Hammerstrom, *Adaptive Solutions, Inc.*

- Priform Model Execution on a Neurocomputer** I-575  
E. Means, D. Hammerstrom, *Adaptive Solutions Inc.*

- A CMOS UV-Programmable Non-volatile Synaptic Array** I-581  
R. Tawel, R. Benson, A. P. Thakoor, *California Institute of Technology*

- Extensible Linear Floating Point SIMD Neurocomputer Array Processor** I-587  
R. Means, L. Lisenbee, *HNC, Inc.*

## IMAGE PROCESSING & VISION I *Oral Presentations*

- Character Recognition with Selective Attention** I-593  
K. Fukushima, T. Imagawa, E. Ashida, *Osaka University*



# CONTINUED TABLE OF CONTENTS

## VOLUME I

**Dynamic Image Segmentation and Optic Flow Extraction** I-599  
H. Tunley, *University of Sussex*

**An Unsupervised Training Rule for Dynamic Information Processing** I-605  
S. Highighi, *Intel Corporation*; L.A. Akers, *Arizona State University*

**Adaptive Fuzzy System for Transform Image Coding** I-609  
S. Kong, B. Kosko, *University of Southern California*

**A Neural Net for Reconstruction of Multiple Curves with a Visual Grammar** I-615  
E. Mjolsness, A. Rangarajan, C. Garrett, *Yale University*

**Real-Time Early Vision Neurocomputing** I-621  
A. Waxman, *MIT Lincoln Laboratory*; D. Fay, *Boston University*

**A Continuous-Time Global Theory of Visual Perception** I-627  
H. Ogmen, *University of Houston*

## IMAGE PROCESSING & VISION II *Oral Presentations*

**Invariant Recognition of Cluttered Scenes by a Self-Organizing ART Architecture: Figure-Ground Separation** I-633  
S. Grossberg, L. Wyse, *Boston University*

**Texture Analysis via Unsupervised and Supervised Learning** I-639  
H. Greenspan, R. Goodman, *California Institute of Technology*; R. Chellappa, *University of Southern California*

**Error Back Propagation with Minimum-entropy Weights: A Technique for Better Generalization of 2-D Shift-Invariant NNs** I-645  
W. Zhang, A. Hasegawa, K. Itoh, Y. Ichioka, *Osaka University*

**Computational Theory and Neural Network Model of Perceiving Shape from Shading in Monocular Depth Perception** I-649  
H. Hayakawa, T. Inui, M. Kawato, *ATR Auditory and Visual Perception Research Laboratories*

## IMAGE PROCESSING & VISION *Poster Presentations*

**Preattentive Texture Segmentation and Grouping by the Boundary Contour System** I-655  
D. Cruthirds, S. Grossberg, A. Gove, E. Mingollo, *Boston University*

**A Biologically Inspired Neural Network Model for 3-D Motion Detection** I-661  
J. Liaw, M. Arbib, *University of Southern California*

**Perceptual Organization in Motion Correspondence** I-667  
R. Mohan, *IBM Thomas J. Watson Center*

**A Neural Network with Multiple Hysteresis Capabilities for Short-term Visual Memory (STVM)** I-671  
M. Gupta, G.K. Knopf, *University of Saskatchewan*

**A Neural Model of Image Velocity Encoding** I-677  
K. Gurney, M. Wright, *Brunel University*

**Texture Segmentation by Clustering of Gabor Feature Vectors** I-683  
S. Lu, J. Hernandez, G. Clark, *Lawrence Livermore National Lab*

# TABLE OF CONTENTS CONTINUED

## VOLUME I

<b>Non-linear Prediction Using a Three-layer Neural Network</b> N. Nasrabadi, <i>Worcester Polytechnic Institute</i> ; S. Dianat, S. Venkataraman, <i>Rochester Institute of Technology</i>	I-689
<b>Methods for Enhancing Neural Network Handwritten Character Recognition</b> M. Garriss, R. Wilkinson, C. Wilson, <i>National Institute of Standards and Technology</i>	I-695
<b>Detection of Dim Targets in High Cluttered Background Using High Order Correlation Neural Network</b> R. Liou, M. Azimi-Sadjadi, <i>Colorado State University</i> ; R. Dent, <i>IBM Corporation</i>	I-701
<b>A Neural Network for the Detection of Rotational Motion</b> I. King, J. Liaw, M. Arbib, <i>University of Southern California</i>	I-707
<b>Interpreting Line Drawings with Higher Order Neural Networks</b> G. Salem, <i>IBM Corporation</i> ; T. Young, <i>University of Miami</i>	I-713
<b>Working Memory Networks for Learning Multiple Groupings of Temporally Ordered Events: Applications to 3-D Visual Object Recognition</b> G. Bradski, G. Carpenter, S. Grossberg, <i>Boston University</i>	I-723
<b>Generalization of the Harris "Coupled Depth-slope" Analog Visual Reconstruction Network</b> D. Suter, <i>La Trobe University</i>	I-729
<b>A Neural Network Based on Differential Gabor Filters for Computing Image Flow from Two Successive Images</b> T. Tsao, <i>Vitro Corporation</i> ; V. Chen, <i>Naval Research Laboratory</i>	I-741
<b>Sensitivity Enhancement of Elementary Velocity Estimators with Self and Lateral Facilitation Mechanisms</b> S. Courellis, V. Marmarelis, <i>University of Southern California</i>	I-749
<b>LEARNING &amp; GENERALIZATION I <i>Oral Presentations</i></b>	
<b>Error Surfaces for Multi-layer Perceptrons</b> D. Hush, J. Salas, B. Horne, <i>University of New Mexico</i>	I-759
<b>The Effect of Initial Weights on Premature Saturation in Back-Propagation Learning</b> Y. Lee, S. Oh, M. Kim, <i>Electronics &amp; Telecommunications Research Institute</i>	I-765
<b>Decoupled Extended Kalman Filter Training of Feedforward Layered Networks</b> G. Puskorius, L. Feldkamp, <i>Ford Motor Company</i>	I-771
<b>A Stochastic Training Model for Perceptron Algorithms</b> J. Shynk, <i>University of California, Santa Barbara</i> ; N. Bershad, <i>University of California, Irvine</i>	I-779
<b>An Asymptotic Singular Value Decomposition Analysis of Nonlinear Multilayer Neural Networks</b> S. Goggin, K. Gustafson, K. Johnson, <i>University of Colorado</i>	I-785
<b>Linear Neural Networks Which Minimize the Output Variance</b> F. Palmieri, J. Zhu, <i>University of Connecticut</i>	I-791
<b>Learning Algorithms and Fixed Dynamics</b> N. Cotter, P. Conwell, <i>University of Utah</i>	I-799

## CONTINUED TABLE OF CONTENTS

### VOLUME I

#### LEARNING & GENERALIZATION II *Oral Presentations*

- A Model of Symbol Grounding in a Temporal Environment** I-805  
B. Bartell, G. Cottrell, *University of California, San Diego*

- An Unified Approach for Integrating Explicit Knowledge and Learning by Example in Recurrent Networks** I-811  
M. Gori, P. Frasconi, M. Maggini, G. Soda, *Universita di Firenze*

- Fast Temporal Neural Learning Using Teacher Forcing** I-817  
N. Toomarian, J. Barhen, *Jet Propulsion Laboratory, Calif. Institute of Tech.*

- Neural Networks Learning in a Changing Environment** I-823  
T. Heskes, B. Kappen, *University of Nijmegen*

- Learning Pulse Coded Spatio-temporal Neurons with a Local Learning Rule** I-829  
A. Klaassen, *LIMS/cnrs*; A. Dev, *Delft University of Technology*

- Generalization by Weight-elimination Applied to Currency Exchange Rate Prediction** I-837  
A. Weigend, D. Rumelhart, B. Huberman, *Stanford University*

- Generalization Capabilities of Minimal Kernel-based Networks** I-843  
M. Kraaijveld, R. Duin, *Delft University of Technology*

#### LEARNING & GENERALIZATION III *Oral Presentations*

- RCE Networks: An Experimental Investigation** I-849  
M. Hudak, *Siemens Corporate Research, Inc.*

- Optimal Adaptive K-Means Algorithm with Dynamic Adjustment of Learning Rate** I-855  
C. Chinrungrueng, C. Sequin, *University of California at Berkeley*

- ARTMAP: A Self-organizing Neural Network Architecture for Fast Supervised Learning and Pattern Recognition** I-863  
G. Carpenter, S. Grossberg, J. Reynolds, *Boston University*

- Projection Pursuit Learning** I-869  
Y. Zhao, C. Atkeson, *Massachusetts Institute of Technology*

- Incremental Learning with Rule-based Neural Networks** I-875  
R. Goodman, C. Higgins, *California Institute of Technology*

- Objective Functions for Probability Estimation** I-881  
J. Miller, R. Goodman, *California Institute of Technology*; P. Smyth, *Jet Propulsion Laboratory*

- Generalization Accuracy of Probabilistic Neural Networks Compared with Back-propagation Networks** I-887  
D. Specht, P. Shapiro, *Lockheed Missiles & Space Co., Inc.*

## The Early Diagnosis of Heart Attacks: A Neurocomputational Approach

Robert F Harrison, Stephen J Marshall and R Lee Kennedy  
Department of Control Engineering  
University of Sheffield  
Sheffield S1 3JD  
England

### Abstract

A multi-layered Perceptron is trained to diagnose the presence, or otherwise, of acute myocardial infarction (heart attack) in patients admitted to an emergency unit with acute chest pain. Two learning algorithms, based on mean-square-error and the log-likelihood function, are compared. Their performance does not differ significantly but the latter rule converges much more rapidly.

Performance in excess of that of the admitting clinicians is achieved for a number of performance indicators, and a protocol for combining the network's diagnosis with that of the clinician is proposed. This results in further improvements in performance, indicating that the MLP can act as a useful decision aid in an emergency context.

### Introduction

The motivation behind this work is to demonstrate the effectiveness of neural networks in clinical decision support. To this end we have chosen a very widespread problem which demands urgent diagnosis. Suspected heart attack is the most common cause of admissions to Accident and Emergency Departments in the UK and the cause of over 1.5 million admissions annually, in the US [1]. In order that the resulting network should be a useful clinical aid we have restricted the input data to only those items which are collected in a normal case history.

The diagnosis of Acute Myocardial Infarction (AMI), or heart attack, is a complex decision making process. Patients are admitted to a hospital's emergency department suffering from chest pain, and the clinician must form a diagnosis accurately and quickly if the patient is to receive the correct treatment. Typically 20-30 % of emergency hospital admissions are of patients with possible AMI and fewer than half of those admitted to a Coronary Care Unit prove to have AMI.

Difficulties in the accurate diagnosis of AMI arise for many reasons. Patient symptoms are often atypical and can have many different causes, *eg* strained chest muscles or stomach ulcers. There is no quick reliable test to aid diagnosis, the electrocardiograph (ECG) is the only diagnostic tool available at presentation, and this often fails to show diagnostic changes. There is, however, a standard test that does give the correct diagnosis, but this involves the measurement of enzyme and ECG changes over a period of 24 to 48 hours. This gives a reference against which the clinician's performance can be measured, but is of no use in the early stages of diagnosis.

In addition to deciding whether to admit the patient to hospital for observation, or to a Coronary Care Unit (CCU) a third decision of whether or not to administer thrombolytic therapy can be made. To be of use this must be administered within about 6 hours of the heart attack. It benefits patients by reducing mortality and heart tissue damage [2], but can be very expensive and there are risks associated with inappropriate use [3]. In the US it has recently been estimated that of those requiring thrombolytic therapy only 10% receive it [4].

There is an obvious need therefore for a clinical decision aid to assist the clinician in making a fast, accurate diagnosis. A neurocomputational approach is well suited to this problem. First, the frequent occurrence of AMI enables a large amount of patient data to be collected relatively easily. Second, the standard test gives a 'correct' diagnosis for each patient, this can be used as a target for training a network such as the Multi-Layer Perceptron (MLP) used here. Third, the difficulty of extracting explicit rules for the diagnosis makes the application of more traditional expert system techniques difficult and time consuming [5] [6]. Such methods also suffer from portability problems in that refining a system to work well in a different geographical location where, for instance referral procedures make be markedly different, may take a long

time. Conventional statistical techniques have also been applied to this problem but none has found wide acceptance. To the authors' knowledge, only one other study of this problem has been conducted using neural networks [7] but it achieved an unacceptably low level of accuracy. We are unable to ascertain a reason for this.

In our work we interpret the network inputs as probabilities of the presence of a particular symptom and its output as the probability that AMI is present. Baum and Wilczek [8] propose that by treating the external variables in such a way and varying the weights in the gradient direction of log likelihood during training, the probabilistic interpretation is allowed. In addition learning is made more efficient. To determine any advantages of this method experiments were performed comparing the performance of networks trained using the least-mean-square backpropagation algorithm and the maximum likelihood modification.

In order to provide the clinician with information about how a particular diagnosis was arrived at, we have investigated the sensitivity of the network output to individual symptoms. We believe that such "contribution analyses" will be invaluable in overcoming objections to the use of neural networks on the grounds that they are "black-boxes".

### The log-likelihood cost function

It is claimed by Baum and Wilczek [8] that backpropagation can be put on a satisfactory conceptual footing and probably be made more efficient by defining the values of output and input neurons as probabilities, and replacing the mean-square-error (MSE) by the log-likelihood function (LL) resulting in a maximum likelihood (ML) rather than a least mean-square (LMS) optimization. From our point of view, treatment of symptoms and diagnoses as probability of occurrence is most appealing and has led us to compare the two learning algorithms.

Given a set of  $N$  input vectors,  $s^\mu$ ;  $\mu \in [0, N]$ , and their associated target values  $t^\mu$ , backpropagation attempts to adjust the network weights so as to minimize the MSE,  $E_{MS}$ , between the output and target values as defined by  $E_{MS} = \sum_{\mu,j} (t_j^\mu - o_j^\mu)^2$  where  $o_j^\mu$  is the output of the  $j$ -th node when  $s^\mu$  is presented as input to the network. This formulation for MSE can be replaced with the LL function,  $E_{LL}$  given by  $E_{LL} = -\sum_{\mu,j} \{t_j^\mu \log o_j^\mu + (1 - t_j^\mu) \log(1 - o_j^\mu)\}$ . By varying the weights in the gradient direction of this function (found by the chain-rule of differentiation) the backpropagation algorithm generalizes immediately from minimizing MSE to maximizing LL and hence likelihood. Holt and Semani [9] have also investigated this and conclude that the use of the LL formulation can significantly reduce training times in the network, and can increase the dynamic ranges of the initial weights over which proper convergence can be achieved. They also note that despite these apparent advantages the LL approach has not found wide acceptance in applications. Here we investigate any practical differences between the two methods when applied to the AMI problem.

### The network structure

The network architecture used here is the multi-layered Perceptron (MLP). One intermediate layer of neurons is used which is fully interconnected with the 53 neurons of the input layer and the single neuron of the output layer. The optimum size of the hidden layer is notoriously difficult to define and here guidance is obtained using the algorithm proposed in [10], with a figure of 18 being chosen.

### Contribution analysis

One of the major criticisms of the MLP is its inability to explain the decision it has reached. The large number of connections and the nonlinearities in the hidden layers obscures the rôle of any particular input in the final diagnosis. In an attempt to overcome this problem we have devised a method to analyse the effect of each input factor on the output. This is done simply by calculating the sensitivity of the output to any individual input ie the effect of a unit change in that input. In fact we calculate the sensitivity of the net input to each output unit to preserve dynamic-range. This is justified by the monotonicity of the output units' sigmoid functions,  $\sigma(\cdot)$ . This is easily done using the chain-rule of differentiation yielding

$$S_{i,m}^\mu = \frac{\partial \text{net}_i^\mu}{\partial s_m^\mu} = \sum_{j \in H} w_{i,j} \sigma' \left( \sum_{k \in I} w_{j,k}^\mu s_k^\mu \right) w_{j,m}$$

where  $H$  and  $I$  denote the units of the hidden and input layers respectively.

The sensitivity of the  $i^{\text{th}}$  output unit to the  $m^{\text{th}}$  input unit is a function of all the inputs as we would expect. This makes for difficulties in deriving a global measure of sensitivity. However, because the sign of  $\sigma'$  is always positive, the sign of  $S_{i,m}^{\mu}$  remains unchanged for any set of symptoms. This means that we can average over all cases. This gives a number for each input unit relating to its contribution for or against the diagnosis, serving both to highlight the important factors used to make the decision and allowing clinicians to check any suspect factors in the input to the network.

### The data

The data for the study were obtained from 300 consecutive emergency referrals, with a complaint of chest pain, to the Medical Department of the Northern General Hospital, Sheffield, England. Demographic, clinical and electrocardiographic data was recorded from all patients, as was the admitting clinician's estimate of the likelihood of the patient having suffered AMI (as a percentage). Myocardial infarction was diagnosed according to the standard criterion thus supplying the "teacher" for the network. From each patient record 38 symptoms were coded as a 53 dimensional, binary vector and the target vector (diagnosis) was coded as 1 for AMI, 0 otherwise. Continuous valued variables such as age and duration of pain were coded as binary vectors eliminating logically redundant elements as described by Widrow *et al* [11]. The first 150 symptom/diagnosis pairs were used to train the network while the second 150 were used to test it.

## Results

### Speed of learning

To confirm the predicted increase in learning speed both the LMS and ML algorithms were trained for 1000 epochs. The learning rate of the ML network was set at four times smaller than that of the LMS net (0.01). This takes account of the fact that changes in gradient in the ML net can be more than four times greater [8] than those taking place in the LMS variant so we might expect the two systems to learn at similar rates. Figure 1 clearly demonstrates that the MSE for both nets falls off very much more quickly for the ML net. Similar, although mirror-image, curves (not shown) can be seen if LL is calculated for each network. This obviously has implications for the required computational power.

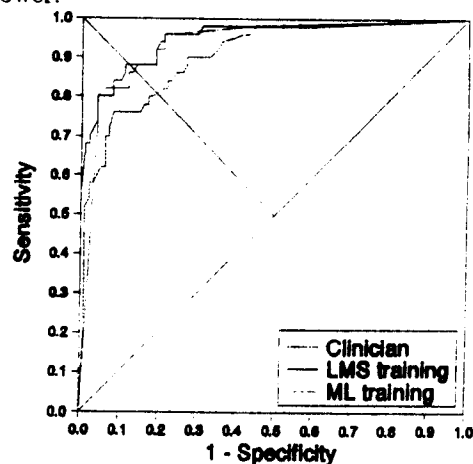
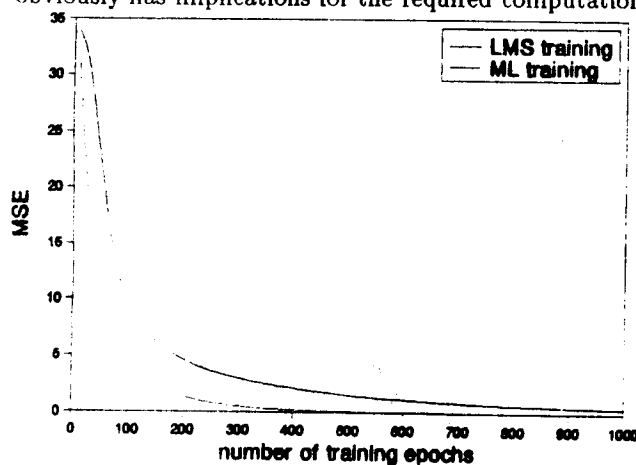


Fig. 1: MSE v no. epochs for LMS and ML training

Fig. 2: ROC curve for LMS & ML nets & clinician

### Diagnostic capability

The diagnostic power of the trained MLP was assessed by calculating the following performance indicators for the test set of (previously unseen) data. They are defined as follows [12]. Diagnostic accuracy: ratio of number of correct diagnoses to total number of cases; Sensitivity: ratio of number of correct *positive* diagnoses to the total number of patients *with* the disease; Specificity: ratio of number of correct *negative* diagnoses to the total number of patients *without* the disease.

In the language of decision theory sensitivity is the probability of detection and specificity, 1 - probability of a false alarm. It is easy, therefore, to plot the Receiver Operating Characteristic (ROC) curve for each of the experiments [13]. This is shown in figure 2 and includes the ROC curve for the admitting clinicians, for comparison. The intersection of the ROC curve with the semi-diagonal indicates the optimum threshold

level,  $T_o$ , (*ie* sensitivity = specificity = accuracy) that each system can operate at for a simple yes/no diagnosis. Thresholds and performance are tabulated in table 1. It is clear from figure 2 that there is no significant difference between the ML and LMS algorithms and that the clinicians performed worse than either of the networks.

The objective of this work is not to develop an alternative to the clinician, but rather an aid which can make optimal use of all available data at presentation *including* the clinician's opinion. For instance, we might make the combined network and clinician system more *specific* by insisting that both diagnoses should be positive if AMI is to be diagnosed (logical AND). Such a protocol might be valid when trying to decide whether or not to administer a thrombolytic agent. A less critical decision, that of whether or not to admit the patient to CCU might benefit from the combination of network and clinician using logical OR. This says if either diagnosis is positive then AMI is present and increases *sensitivity*. A third option arises which admits a "grey" area of opinion, representing probabilities of AMI lying between an upper and a lower limit, say, 10% and 90%. Table 2 indicates a possible scoring system for combining diagnoses. Using this weighted combination of clinical opinion and output from the neural network, MI is diagnosed if a combined score of 3 or more is obtained. Notice how responsibility ultimately devolves to the clinician. It is not suggested that this protocol is in any way optimal. It was chosen arbitrarily to demonstrate how a workable system might be achieved. Clearly, a good deal of effort is required to define the way in which machine and clinician should best co-operate.

	$T_o$	acc	sens	spec
LMS net	0.57	0.88	0.88	0.88
ML net	0.74	0.87	0.86	0.87
clinician	0.47	0.81	0.80	0.82
LMS + clin	—	0.89	0.96	0.86
ML + clin	—	0.87	0.94	0.84

Table 1: Performance

Clinician		Network	
Prob MI (%)	Score	Prob MI (%)	Score
0—10	0	0—10	0
11—50	1	11—50	1
51—90	2	51—90	2
91—100	3	91—100	2

Table 2: Scoring system

Figure 3 shows the average contribution over all cases for each input for both the LMS and ML networks. Positive contributions add to the overall output while negative ones subtract. The six most important positive and four most important negative contributors are listed in table 3. All agree with clinical opinion, for example, the presence or absence of ECG changes are some of the major factors used by the clinician. Unfortunately space does not allow us to discuss them in detail here. The order of importance for both networks is similar although the absolute values of  $S_{i,m}$  for the ML network tend to be lower. This can be explained by the fact that we have observed that the final weights in the ML net tend to be smaller than those in the LMS net.

$m$	$S_{i,m}$		Symptom
	ML	LMS	
48	4.9	8.2	New ST segment elevation
50	2.8	4.1	ST or T wave suggest isch.
47	2.1	3.4	Patient hypoperfused
49	1.9	2.7	New pathological Q waves
13	1.7	2.4	Other risk factor
26	1.7	2.9	Associated with Sweating
31	-2.2	-3.2	Episodic pain
42	-1.9	-3.1	History of Angina
10	-1.6	-2.4	Family history of isch. HD
21	-1.2	-1.8	Pain worse on inspiration

Table 3: Most important contributors

### Conclusions

For the present study we conclude that the performance of the ML and LMS trained networks does not differ significantly. However, there is a distinct advantage in terms of speed of learning when the ML algorithm is used.

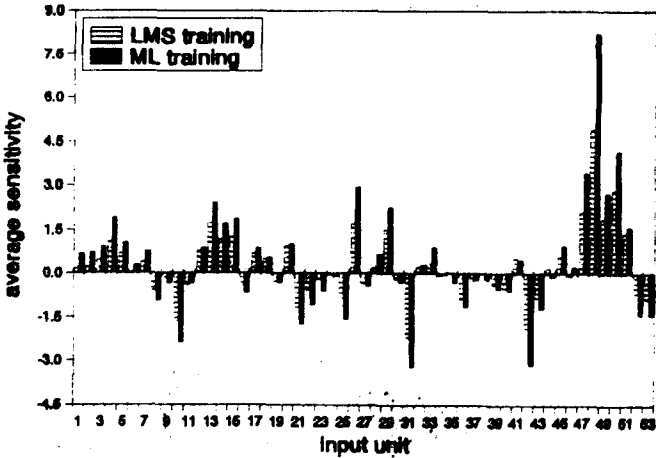


Fig. 3: Average sensitivity of output on inputs



For a computer-based diagnostic aid to be of value it must be statistically valid, diagnostically accurate and its use must enhance the clinician's diagnostic performance. In addition it must be readily usable in the clinical setting. In this study we have not rigorously addressed the question of statistical validity but we have developed a system whose performance can exceed that of experienced physicians. Our system is readily usable on a portable computer, could be developed as a hand held instrument and gives an instant prediction of the likelihood that a patient has sustained AMI.

We have also shown how our system could be used to enhance the clinician's judgement and to indicate the importance of those factors which influence the decision making process. The preliminary results in this latter area are encouraging and we believe that this will be a fruitful line of enquiry. As with all clinical decision aids though, the true performance of the system can only be assessed by a formal, clinical trial.

## References

- [1] Pozen, M *et al* (1984) A predictive instrument to improve coronary care unit admission practices in acute ischaemic heart disease: a prospective multi-centre clinical trial *New England J Med* **310** 1273-1278.
- [2] ISIS-2 (Second International Study of Infarct Survival) Collaborative Group (1988) Randomized trial of intravenous streptokinase, oral aspirin, both or neither, among 17,187 cases of suspected acute myocardial infarction *Lancet* **ii** 349-360.
- [3] Rawles, J and Julian, D (1989) Role of the general practitioner in managing patients with myocardial infarction - impact of thrombolytic treatment *Brit Med J* **299** 555-557.
- [4] Kennedy, J W. (1990) Expanding the use of thrombolytic therapy for AMI *Annals of Internal Med* **113** 907-908.
- [5] Fattu, J, Blomberg, D and Patrick, E (1987) CONSULT learning system applied to early diagnosis of chest pain *Proc 11<sup>th</sup> Annual Symposium on Computer Applications in Medical Care* 71-77.
- [6] Emerson, P *et al* (1988) The development of ACORN, an expert system enabling nurses to make admission decisions about patients with chest pain in an accident and emergency department *Medical Informatics in Clinical Medicine, Nottingham* 37-40.
- [7] Hart, A and Wyatt, J (1989) Connectionist models in medicine: an investigation of their potential *Proc European Conf on AI in medicine, London* 115-124.
- [8] Baum, E B and Wilczek F (1988) Supervised learning of probability distributions by neural networks *IEEE Conf. on Neural Information Processing Systems, ed. Anderson, New York: American Inst. Physics* 52-61.
- [9] Holt, M J J and Semnani, S (1990) Convergence of back-propagation in neural networks using a log-likelihood cost function *Electronic Letters* **26** 1964-1965.
- [10] Gutierrez, M, Wang, J and Lipschik, G (1989) Estimating hidden units for two-layer Perceptrons *Proc IEE 1st Int Conf on Artificial Neural Networks* 120-124.
- [11] Widrow, B, Gupta, N and Maitra, S (1973) Punish/reward: learning with a critic in adaptive threshold systems *IEEE Trans on Systems, Man and Cybernetics* **SMC** **3** 455-465.
- [12] Harrison, R F, Kennedy, R L, Marshall, S J (1991) The multi-layer perceptron as an aid to the early diagnosis of myocardial infarction *Proc Third European Conference on Artificial Intelligence in Medicine* (To appear).
- [13] Meistrrell, M L (1990) Evaluation of neural network performance by receiver operating characteristic (ROC) analysis: examples from the biotechnology domain *Computer Methods and Programs in Biomedicine* **32** 73-80.

