Available online at www.joac.info

ISSN: 2278-1862

Journal of Applicable Chemistry

2014, 3 (3): 919-989

(International Peer Reviewed Journal)

State-of-Art-Review (SAR-Invited)

Mathematical Neural Network (MaNN) Models Part III: ART and ARTMAP in OMNI_METRICS

K RamaKrishna[1], V. Anantha Ramam[2], R. Sambasiva Rao[2*]

1. Department of Chemistry, Gitam Institute of Science, Gitam University, Visakhapatnam, 530 017, **INDIA**
2. School of Chemistry, Andhra University, Visakhapatnam 530 003, **INDIA**

Email: karipeddirk@gmail.com, rsr.chem@gmail.com

Accepted on 25th April 2014

(Dedicated to Dr U Murali Krishna, former Special officer, A U P G center, Nuzvid, former Professor of [Analytical, engineering] department, Andhra University, on his *sahasra chandra darsanam* (thousand lunar months of life on the lap of mother earth).

ABSTRACT

Adaptive resonance theory (ART) proposed by Grossberg in 1976, is a self-organizing (SO) unsupervised learning approach. It balances stability vs. plasticity dilemma in learning new traits without forgetting the old ones. Another popular SO mapping (SOM) of Kohonen introduced in 1990s subtly differs from ART and based on neighborhood influence. ART1-NN is the start of a new era of unsupervised-data-driven models using resonance, ordinary differential equations (ODEs) and backward connections. The main functioning of ART1 is in feature and category layers which are connected both ways. It accepts only binary input and makes use of winner-takes- all (WTA) and vigilance approaches. ART2 and fuzzy-ART are modifications of ART1 to handle analogue and floating point input values. ART3 performs a parallel search and incorporates a term similar to chemical transmitters playing a key role in biological system. The NNs reported over two decades in this category include coupled-, probabilistic-, projection-, performance-guided-, lateral-priming-, efficient-, fusion-ARTs. The philosophy of Grey relational analysis inspired from human brain is used in Grey-ART. Multiple channel data is analysed with multi-ART. Under hybrid category, SOM is used in fully-organized-SOM-ART and RBF in RBF-ART.

ARTMAP is a supervised learning procedure retaining the structure/function of ART as it is and contains ARTx and ARTy modules for binary input and output. Fuzzy-ARTMAP (fam) accepts floating point input as well as binary strings. Thus, ARTMAP and fuzzy-ARTMAP inherit advantages as well as shortcomings /pitfalls/limitations of ART. Both these NNs were put forward by Grossberg, Carpenter and their associates in early nineteen nineties. The efforts in the past 20 years are aimed at surmounting the ill effects of order of presentation of input patterns, noise and category proliferation which are the inherent limitations of ART. FAM is extended to function approximation task in FasART and FasBak and Marriott-FAMs. Instance counting-NN was proposed by Carpenter for medical diagnosis with inconsistent data. Order-FAM surmounts the negative consequences of random order of presentation of input. In hyper-spherical- and Gaussian-FAMs, a kernel is used instead of a rectangular transfer function (TF). Parallel FAM is a fast training version for pipeline approach. Instar-, out-star-, distributed-, WTA, winner-takes-most (WTM), unsupervised-, semi-supervised-, self-supervised- learning methods are in practice. Auto-

rule-discovery (ARD) and self-supervised-ARTMAP, a knowledge extraction and discovery tool are applied to satellite images. Other improvements in this category of NNs are Georgiopenlos-, Granger-, Charalampidis- Cascade-, Hierarchical-, probabistic-, adaptive- ARTMAPs and Bayesian-, Distributed-, GA-FAMs. The applications of ART-based NNs are widespread in chemistry, medical diagnosis, and engineering. The technology transfer to Boeing Corporation, MIT Lincoln laboratory for mega tasks is a gold standard test for its versatility and applicability and occupies a niche in the arsenal of data-driven non-parametric/AI paradigms.

Keywords: Adaptive resonance theory (ART), Classification, Supervised-ART (ARTMAP), Binary, floating point data, fuzzy-ARTMAP, Inconsistent/contradictory input, Medical diagnosis, Chemical analysis, Auto-rule-discovery (ARD), Knowledge-extraction-discovery-satellite-images

---

Contents_ **ART** and **ARTMAP** in **OMNI_METRICS**

3.6 **Hybrid-ART-NNs (ART + $$$]**

- Fully Organized self-organizing-ART (FO-SO-ART)
- Fusion Architecture for Learning, COgnition and Navigation (FALCON)
- Cerebellar model articulation controller (CMAC) + ART

4 Supervised (self organizing) NNs based on ART

4.1 **ARTMAP**

4.2 **Evolution of ARTMAP [Evol.ARTMAP]**

- ARTMAP-Georgiopenlos
- Cascade-ARTMAP
- Hierarchical ARTMAP
- Default_ARTMAP
- ARTMAP Rule Discovery (ARTAP.Rule_Discov, ARD)
- Self-organizing-ARD (SO-ARD)
- Self-supervised-ARTMAP
- Information fusion-ARTMAP
- Deriving knowledge hierarchy from a trained network

4.3 **Fuzzy-ART-MAP (FAM)**

4.3.1Applications.FAM
4.3.2 Advantages and limitations of FAM
4.3.3 Match tracking anomaly
4.3.4 Comparison of performance of FAM with other paradigms

4.4 **Evolution of fuzzy ARTMAP ($FAM)**

- Ordered fuzzy ARTMAP
- ARTMAP-Instance counting (ARTMAP-IC)
- Gaussian- ARTMAP
- Hyper-spherical FAM
- Hyper-sphere-(or RBF-)-FAM
- Bayesian-fuzzy-ARTMAP
- Simplified FAM and its modifications
- Parallel-FAM
- prob-FAM
- Fuzzy adaptive system (Fas) -ARTMAP
  - Fas-ARTMAP_ Cano
  - Distributed Fas–ARTMAP-Carpenter
  - Distributed Fas–ARTMAP-Parrado
  - FAS-ARTMAP with STORE
- Boosted-FAM
- Mu-FAM
- Safe-mu-FAM
- GA-FAM
- Modifications of fuzzy-ARTMAP
  - Fuzzy-ARTMAP-Granger_2001

INTRODUCTION

Nearly three-quarters-of-a-century of evolution of mathematical NNs (MaNN) and neuro-physiological science made a landmark in brain research, modeling, information /knowledge extraction, classification/clustering, pattern recognition and function approximation of real time spacio-temporal data [1-182]. The multivariate/multidimensional data is about worldly objects, satellite images and statistically designed experimental output in medical science, engineering, physico-chemical processes etc. Humans and other primates effortlessly recognize objects in the world as they move their eyes, heads and bodies [25].

1.1 Human brain

The bliss of perception/recognition of objects exposed to light in the visible region is a characteristic of advanced (human being) brain. The ability of a child to distinguish males from females, monkeys from gorillas/chimpanzees and that of an adult not to forget faces of his parents even after a long gap of time are the gifts of nature. In the words of Carpenter and Grossberg [147] , 'As we move freely throughout the world we can

- attend to both familiar and novel tasks
- rapidly learn to recognise
- test hypothesis
- learn to name novel objects without unselectively disturbing memories of familiar objects.

In yester years, the broad perspective sustained was 'brain is bliss or a bag of tricks played randomly or with an intention'. Now, neurobiology and computational model based on nature inspired paradigm is in a state of maturity. In the language of computer and medical sciences fusion, the hardware and software of a human (or monkey) brain recognizes, stores and retrieves the objects, seen at multiple positions/sizes and from different viewpoints. This invariance is gradually built up in V1, V2 and V4 through infero temporal and prefrontal cortex of human brain. The size of neocortex of humans is three times more compared to that of apes. It covers the surface of brain and organized into six layers which spread along cortical sheet horizontally. Many columns of diameter 300-500 mm cut through these layers vertically. The brain learns effortlessly, processes visual/ audio and tactile sensory data streams. A brain of even a common man/expert of any age or child prodigy receives continuously large information through sensory organs from a variety of external sources over learning/practicing period and emergency time [22].

The brain somehow implements real-time probability theory, hypothesis testing, prediction and decision-making. It is mysterious to comprehend on molecular level how local computations enable fast autonomous adaptation to a non-stationary (world of) environment whose rules change in unexpected manner through time. The brain neither contains terminology nor concept but a huge cloud of charges spread over.

The medical and surgical sciences now employ fMRI, computer interfaces and surgical-robots along with the state of ART procedures to probe into brain function at organ/sub-sections/cell/molecular level. The high resolution IR/microwave cameras on land and satellites capture the pictures/images not visible to eye of a man. The insects view in UV-region, a passive one for human eye. The next generation robotic vision includes the blend of these regions. For autonomous land vehicles (ALVs), the objects in the shadow regions also do not miss the attention. Now night vision is essential for plane landing and in defense operations. Non-clinical studies include para-psychology and effect of environment on human brain.

1.2 Biological learning

Learning in living species (called biological learning) is multifold; briefly perceptual/cognitive and spatial/motor processes play a role in predictive mechanisms and in turn partially control learning. The spatial representations and motor gains at infant stage and for adults are different. Thus, catastrophic forgetting is also an invited feature which arises in right proportion with growth. It has a beneficial effect as it is totally irrelevant for an adult to store the traits learnt during infant stage of life. Stability–plasticity dilemma is solved through matching and mismatch-based learning by refining continuously spatial maps and sensory–motor gains.

1.3 Mathematical model of brain: A simple mathematical model of human brain started in 1940s with rigorous scientific enquiry. Since then multifold progress resulted in mathematical models implemented in software, but their function is not even a fraction of a subsection of the brain. However, this technology excelled almost many of the century old techniques in solving many real life problems. Grossberg [159] reviewed the state-of-art of NNs in 1988. The knowledge and intelligence bits of Helmholtz, Maxwell gave birth to the base of 20th century science. The next leap is in natural computation algorithms/global multiple-object tools on theoretical front, imaging and micro-array experiments steered brain science to the state of maturity. The non-linear, non-stationary and non-local characteristics of memory, verbal/physical communication/behavior/thinking are miracles of (human) brain and mimicking even partially is awesome. The definition of intellect and its measures vary with time both qualitatively and quantitatively. Similarly what were considered as intellectual trends in the last century are now a common sense base and this profile continues. Physical/mental stress/strain/health/design/challenges for the theory of expert and risk/rescue operations bring forth a stir in the information processing. It is still a marvel that the brain makes sense out of that what exists/faded/deep buried/apparently unavailable, verbal/figurative ocean of data/information/knowledge [22]. Finally what humans feel/express appear to others in a structured form.

NNs as models of brain [1-10]: In 1943, McCulloch and Pitts [1] proposed small units (called artificial neurons) connected to explain the functioning of nervous system. It is the first simple as possible (SAP) model based on mathematical logic. The idealistic artificial neurons with accumulation (confluence) operator is an abstraction of biological neurons. The truth table of Boolean logic gates (AND, OR) were successfully reproduced with fixed connection weights. Hebb [7] introduced a learning law in 1949 for biological synapses and could explain psychological results. Rosenblatt [2] implemented perceptron model and ADALINE/MEDALINE of Widrow is a breakthrough for linear NNs. The learning rules in these models are, of course, different. The historical perspective of NN research was reviewed in a monograph 'linear learning machines', which arose interest in NNs with high expectations. But, the publication of Papert and Minsky [3] proving that linear NNs cannot explain even XOR gate was a death blow to the discipline resulting in financial cut and consequent reduction in publications. But, dedicated researchers, Anderson [7], Grossberg, Amari, Hopfield [5], Fukushima [4] et al. [Appendix 1b] continued their studies from a different perspective.

1.4 Limitations of NNs of 1970s

Learning algorithms of 1970s
- Unstable in a non-stationary world
- Cannot learn rare cases important
- Unstable if learning is fast

The weight (W) matrix of a feed forward multiple-layer NN with a chosen architecture was fixed when the training with data set is completed. The model performs fine as long as the prediction/test samples are a subset of training lot i.e. the trend is same with similar noise structure and there are no data from other types of clusters. But, the real world data sets come up with new class of entities which are not a part of training sample data. Other natural phenomena that crop up are the existing classes change (some classes disappear, new ones come up, class characteristics vividly change (drift away) to form another sub-class and so on). The only way out is to repeat entire training process with all samples. Yet, the information is in the architecture and weight or synaptic strength of connections (W) matrix. A pure stable network does not absorb additional knowledge from even new pattern with preservable important information. On the other hand, a pure plastic NN has no scope of preserving the accumulated information/knowledge. Hence, a compromise for this stability/plasticity dilemma is a NN flexible enough to learn from new patterns and sufficiently rigid to retain/assimilate the trend/unique data of the trained/test patterns. How the brain of an outstanding scientist reorganizes and discovers is still a pie in the sky? However, ART networks exhibit many complementary properties embodied in laminar cortical circuitry for vision, audition, speech, language and cognitive information processing.

2. Era of Adaptive Resonance theory

Grossberg [171] introduced Adaptive Resonance Theory (ART) as a modeling and means of understanding human cognitive (vision, learning, audition etc.) information processing. The essence of it is pattern matching. In a physical system resonance is said to occur when the small vibrations of suitable frequency causes a large amplitude vibration. In ART-NN, the information (output of PEs) reverberates back and forth between layers of the network. When a proper pattern is inputted, stable oscillations occur. This is how the name ART goes in to the name of neural networks of this category. The vital role of ART in resonant states responsible for fast and stable learning certifies the name adaptive resonance.

2.1 Evolution of Adaptive resonance theory (ART)

Inspiration for ART from Nature
- ◎ Recurrent structure of information processing
  - ▶ in the cortex and deeper lying structures

Grossberg inspired by fascinating functions of human brain and animal behavior, put forward short term memory (STM) during his graduate studies [181]. The concepts of neural resonance and neural adaptation are derived from cybernetic explanation of learning in natural (animal and human) brains. After a doctoral degree in mathematics, he pursued mathematical models to explore brain from Psychophysical, physiological and neurological data, with a focal point of his auto/adapted resonance theory (ART)

ART_throws light on
- Prediction of behavioral/ neurobiological data
- About human cognition
- Dynamics of spiking laminar cortical networks

proposed in 1976 with unique features described in Chart 1. The normal and abnormal aspects of human and animal perception and cognition, the spiking and oscillatory dynamics of hierarchically-organized laminar thalamo-cortical networks in multiple modalities necessitated the evolution of the hypotheses/theories of brain research. A subset of them refined evolution of ART with added components. It is considered as right approach in at least partially solving the classical mind/body problem. It is a mathematical model for fundamental behavior of functions viz. learning, forgetting, parallel and distributed information storage, long term memory (LTM), short term memory (STM) and pattern recognition (PR) of biological brain. Its' philosophy is altogether different from hierarchical cluster analysis (HCA), which is based on mathematical and statistical methods.

Chart 1: Adaptive resonance theory -Unique features

- Offers solution for stability–plasticity dilemma
- Predicts how brain learns quickly without forcing catastrophic Forgetting of already learned, still successful, memories.
- Carries out fast, yet stable incremental learning under both Unsupervised and supervised learning Conditions in response to a complex non-stationary world
- Models
  - Real-time prediction, search, learning, recognition.
  - Human cognitive information processing
- Autonomous incremental learning under both unsupervised and Supervised conditions in response to stimulate from a Complex non-stationary world
- A cognitive neural theory explaining quick learning of brain
- Basis of set of evolving computational algorithms for large scale applications
- Predicts functional link between processes of CLEARS
- Probes into clarification of brain processes, conscious experience
- Predicts function of top-down-weight (WTD) as learnt expectations reflecting Coherent representations of dynamic real world

3 Unsupervised (self organizing) NNs based on ART

The category to which a pattern belongs is not known a priori and thus the learning is called unsupervised.

3.1 Data Structure

The data (tensor of any order) may be contaminated with noise (fuzzy, any type of statistical distribution, data from another process) and may also contain single or multiple outliers with or without a trend. Some data sets contain inconsistent, contradictory, missing values or may be of sparse in nature. Significant target oriented modifications of Fuzzy ART yields reliable results even in such situations.

Complimentary code: I is the normal positive binary valued pattern of m (length) bits. The compliment (negative) of the pattern I is (IC = one-I) where one is patterns of all 1s. By concatenating I with IC we form the pattern C = [IC,I] of length 2*m. If the length of I is n, Ic is m-n and length(I) + length(IC) = m (Table-1). Complimentary code is used in ART-x and ARTMAP NNs.

Table-1: Complimentary code of binary string

4-bit binary string			8-bit binary string		
I	=	1001	I	=	1001 0000
IC	=	0110	IC	=	0110 1111

I+IC	=	1111
[I,IC]	=	[1001,0110]

I+IC	=	1111 1111
[I,IC]	=	[1001 0000,0110 1111]

3.2 ART-1

The first artificial mathematical (Ma) neural network (NN) incorporating the ground-breaking ART philosophy was in a seminal publication of ART1-NN (accepting only binary input) by Carpenter and Grossberg [162] in 1987. It is an on-line and self-organizing-continuous learning procedure. It learns prototypes of patterns rather than the given data as it is. ART-1, though introduced primarily to classification tasks, spread its wings beyond leaps and bounds in NN literature over the last quarter century. It is competing even for function approximation (curve fitting) parameterization and novelty detection is the corner stone. It influenced many other neuro-computational experts and resulted in hybrid NNs viz. RBF-ART. The parallel computation processing of ART1-NN in hardware on a ring of processors was implemented in 1998 wherein the communication was divided in two bidirectional rings, one for the F1 and other for the F2 layers.

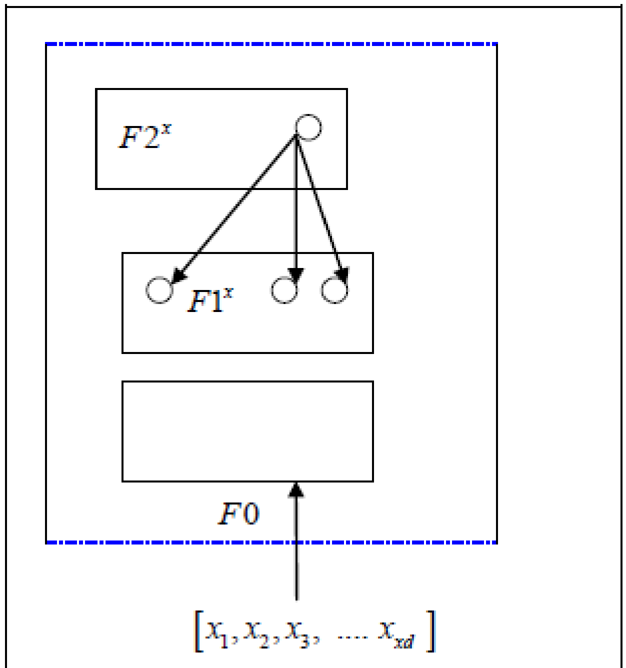


Fig. 1: Components of ART-1

Architecture.ART1: The architecture of ART1 is a layered structure with vectors of neurons. It can be considered as a recurrent multi-level NN consisting of three layers F0, F1 and F2. I(nput) layer accepts only binary digits and is passed into F0 (preprocessing) layer. F0 calculates complimentary code and the concatenated input and complimentary one is inputted to F1, the feature layer (input field/comparison/matching field). The number of neurons in F0 and F1 layers is equal and thus there is one-to-one mapping. It can be deemed that WF0F1 is unity. Thus, the connection between them is non-modifiable. The number of neurons in F1 is equal to the twice the length of binary input string. F2 is classification (clustering/recognition/competition/choice/output) layer. Each neuron in F2-layer represents a prototype of patterns selecting category during self organizing activity. F1 and F2 layers are fully connected in both forward and backward directions. The bottom-up/feature-to-category (forward) connections with weights (WBU or WFC or WF1F2) and top-down/ category–to-feature (backward) connections with weights (WTD or WFC or WF2F1) are shown in Fig. 1. The top down weights represent learned patterns or expectations. The bottom-up weights have ability to add a class when a data point falls outside of a present proximity of all existing classes. Each of the output neurons has a strong excitatory connection to itself and strong inhibitory connections to all other neurons. The advantage of this approach is identifying new categories (KB-1) of data as they emerge without disturbing the existing categories.

Knowledge base (KB 1): New category in ARTx

If	Match is good enough between input and learnt
Then	Synchronous resonance &
fast learning

If	Mismatch is big enough
Then	Hypothesis testing, or memory search to detect new category
If	familiar pattern satisfing the limiting conditions of previous patterns presented to NN
Then	network recognizes category & incorporates new information in new input by adapting weights
If	a novel input is presented, i.e. not satisfying limiting conditions of previous examples
Then	structure is adapted and the novel input is identified as first representative of a new class

It accepts only binary string as input. The output is distinct classes of training patterns.

Choice function: The function $T_j(i)=\dfrac{|I \wedge W_j|}{\alpha+|W_j|}$ determines winner category and shows preference for the hyper box with smaller size $\left(\text{or larger } |w_j|\right)$. It needs smaller change to cover the categories.

The activation function is a bitwise Boolean operator.

$\cap$	:	Bitwise Boolean 'AND' operator
$\lvert binary\ string \rvert$	:	Number of ones in binary string

Functioning of ART-1

F0-layer: It converts input into complimentary code and serves as a storage site of patterns. The output of F0 is transmitted to F1. A set of binary digits (I) of any length, for example [1,0,0,0,1,1] is inputted to the F0 layer. The complimentary code (IC) is developed and concatenated input and its compliment is the output of F0 layer, which is transmitted to the feature (F1) layer. The user chosen parameters are given **table-2.**

Complimentary code

- Allows learning process to encode not only features present but also absent
- Equal basis for features that are consistently absent with features that are consistently present
- **If** system is leaning a category &
  Features (Xz) are sometimes absent and sometimes present
  Then regarded as uninformative with respect to that category

Table-2: Parameters of ART type NNs

Professional II- ART-1 input

Input File Name		
(Training)	**trainART1.nna**	
Testing	**testART1.nna**	

#Input	**2**
F2	**50**

			# PEs	
Vigilance	0.8		Input	
Learning rate	1.0		F2	60
Choice parameter	2.0			

Professional II- fuzzy-ARTMAP input

Input File	**Name**
(Training)	**trainFAM.nna**
Testing	**testFAM.nna**

#Input	**2**
# output	**1**
F2	**50**

			# PEs	
Choice parameter	0.1		Input	
Vigilance parameter			F2	60
Base line vigilance	0.0			
Recode rate	0.5			
Error tolerance	0.01			

F1-layer: It carries out clustering or partitioning of input space. The basis is Euclidean distance measure of the features. The outputs of F1 layer form the input to the category (F2) layer **(Alg. 1).**

F2-layer: Winner takes all (WTA) principle governs the categories.

Uncommitted node in F2: A node in F2 layer is called uncommitted if it didn't learn any input pattern or not selected before by any input pattern.

Committed node in F2: A node in F2 is committed, if it already learnt one or more input patterns in the preceding learning steps [88] or a node selected already by an input pattern

The neurons in the output (class) layer correspond to long-term memory (LTM). The creation of new categories is based on user chosen vigilance (threshold) parameter, which corresponds to short term memory (STM). It self-scales its computational units to learn context dependent values of signal and noise. The match tracking concept is used in learning process. After the training of all the patterns, the output of ART-1 corresponds to the distinct categories. If a test pattern is now presented, it identifies to which class it belongs. If the new pattern is not one among the already learnt classes, the test case will be added as a new class to the existing list of classes. The features of ART-1 are in chart 2.

Alg 1: ART-1

Input : A sequence of binary values $x \in [0,1]^d$

Vigilance parameter (vig.par) [0 to 1] : [0.8]

Initiation of weight vectors $w_j \in [0,1]^D$

Repeat For all training samples

Phase : 1 ***Preprocessing***

Calculate Complementary code (Ic) of input pattern (I)

values of [I,Ic] from F0 are the input to F1 (STM) neurons

While Similarity > = vig.par

Phase : 2 ***Competing stage***

Non-reset (uncommitted) F2 neuron (j)

If first pattern

Then first F2 node is committed

Else Analyze pattern for a category

Cal maximum(response) activated by bottom up activity of WBU

It is set as winner neuron

Phase : 3

Matching stage

Similarity :Top Down (TD) activity of LTM (v) denoted by F2 neuron J is compared with input pattern

If Wk matches with xi by similarity test
Then xi is added to the cluster represented by Wk
Modify Wk
Else xi becomes a new cluster

If winner unit passes through two tests for a given input

Then network is in resonance & W is upgraded

If Kth unit passes through first test and second test fails
Then kth unit is deactivated i.e.
output clamped to zero till a new input is trained

If kth unit fails for first test
Then Wk is described to be far from the input.
Initialize a new unit

End while

Phase : 4

Refine Wj of v(LTMs) -- F2 neuron j

End repeat

Chart 2: Advantages of ART-1-NN

+ Stability/plasticity dilemma is partially solved
  + Novelty detection
+ Learns continuously new associations without disrupting previous information
+ Self-organising
  + does not require human intervention
+ Operates in non-stationary environment
+ Does not collapse even if the classification environment drastically changes
+ Robust to unexpected response changes of profiles in the feature (variable) space
+ Theoretical proof exists for self stabilization and convergence of W, as ART system is represented by ordinary differential equation (ODE)
+ Proofs of Convergence of learning, Pattern diversity, N-N-N conjecture, are reported for ART-1
+ Intelligently finds clusters in data with large number of features
+ Creates new classes online which are not the part of training phase
+ Input activity is normalized like in competitive NN
+ Contrast enhancement of input patters
+ Distinction between STM and LTM

Chart 2(b): Limitations and remedial procedures of ART-1

− Category proliferation
− Does not accept analog input
  Remedy : ART-2

- Does not accept floating point input
  - **Remedy** : Fuzzy-ART
- Sensitive to the order of presentation of training patterns.
- Fails to categorize patterns with one vigilance parameter
  - **Remedy :** Multi-ART
- Unnecessary classes are generated since ART processes noise as an informative signal
- Information is ignored as it mistakes some times to be noise
  - **Remedy :** Coupled ART-NN

The fact that ART was inspired by recurrent brain structures is altogether different with issues of implementing ART on a computer.

No-match tracking: Anagnostopoulos [57] introduced no match tracking concept in ART-NNs. A category that maximizes the bottom up weights for a given input pattern is selected.

If	Vigilance test is passed & mapped on to an incorrect output
Then	Chosen category is deactivated and a new (uncommtted) category is activated

This new category will be used to encode other inputs. When the situation corresponds to the rule cited, this module deactivates the chosen category. Further, the vigilance threshold is increased. The search for another committed/uncommitted category is continued.

Dataset.binary-ART-1: The binary patterns analysed by ART1 and ART-type NNs are in Table-3.

Table-3: Datasets. Binary

#Input	Input
1	*000 111*
2	*001 100*
3	*001 111*
Orthogonal;	
CC =0.016	

#Input	Input
1	*000 10*
2	*001 01*
3	*000 01*
Orthogonal;	
CC =0.4	

#Input	Input
1	*000 111*
2	*001 100*
3	*111 100*
4	*001 111*
5	*000 011*

#Input	Input
1	11110
2	01001
3	00111
Vig.par = 0.3	

3.3 Evolution of ART based NNs

The outcome of neuro-/physiological-/pharmaceutical-/medial-diagnostic-/ICU-care-/multi-organ-therapy-research is a set of large numerical-/attribute-/symbolic-databases of varying complexity, reliability, precision and accuracy. These databases contain consistent-/ inconsistent-/ conflicting- data-/information-/knowledge-/intelligence-bits.

Evolution of architectures of ART and ARTMAP NNs (vide infra) was to accept different categories of inputs (binary, symbolic, graph structures), combat with limitations of basic modules, imbibing the positive features from cross disciplinary modules etc. The evolution of '$ARTx' NN-models are result of refinement of the earlier theory and need based in category learning, recognition and prediction.

Adaptive Resonance Associative Map (ARAM): Tan [130] reported ARAM in 1995 with the inspiration of auto resonance theory of Grossberg and Art-x models of Carpenter. The core structure of ARAM consists of an instar-outstar system. It has two overlapping ART modules sharing a single category field. ARAM is rapid, stable with hetero associative learning in real time environment. It produces better generalization than BPNN (chart 3). ARAM-NN performs two different memory tasks viz. classification and hetero associative recall. It compares with counter propagation NN. For a set of three visual patterns (plane, tank, helicopter) and an extended set of 10 patterns, size of ARAM is significantly smaller (1704 and 2030) compared to BAM (80640 and 9240).

Chart 3. Positive features of ARAM-NN

+ Stronger noise immunity than BAM
+ Better in storage and recall of small sets of pattern pairs
+ Fast learning
+ Guaranteed perfect storage
+ Full memory capacity

Fuzzy ARAM: Tan [130] extended ARAM for analogue input patterns in fuzzy-ARAM.

Dataset.Fuzzy ARAM: The SONAR return data contains 208 patterns and 60 features with floating point values. Two classes, 97-rough cylindrical rocks and 111-metal cylinders are surfaces for the SONAR reflection. Fuzzy- ARAM is superior to k-NN, MLP-BP and linear discriminant analysis (LDA) (Table-4).

Table-4: Performance of different methods for SONAR return data set

Model	H#	% Accuracy
I/O	0	73.1
SLP	12	90.4
SLP	24	89.2
k-NN	--	91.6
Fuzzy-ARAM	22 to 42	91.6
Fuzzy-ARAM	68 to 72	92.9
Fuzzy-ARAM	Voting across 5 simulations	94.2

ART-2

Carpenter and Grossberg [161] proposed ART2. It is a stable unsupervised pattern learning NN to categorize either analog or binary input patterns presented in an arbitrary order.

Architecture.ART-2: The macroscopic architecture of ART-2 is same as that of ART1. But, F1 layer is divided into six sub layers viz. w, x, u, v, p and q. The rth layer of orienting subsystem and six sub layers of F1, have equal number of neurons. They are equal to the length of binary string. The layers are connected as w→ x → v → u → p → q in the forward direction and the reverse flow is in q → v and u →w. G is the gain control unit and is present between the pairs of layers w,x ; u,v ; p,q ; u,r; and p,r. A non-specific inhibitory signal is sent to each unit on the layer G. Full connections exist between bottom up connections to F2 and top down connections between F2 and F1. Typical floating point datasets studied are in table-5.

Table-5: Dataset.floating_point

#Input	Input
1	0.2
2	0.7
3	0.1
4	0.5
5	0.4
ART/02	

#Input	Input	
	x1	x2
1	1.0	0.1
2	1.3	0.8
3	1.4	1.8
4	1.5	0.5
5	0.0	1.4
6	0.6	1.2
7	1.5	1.9
8	0.7	0.4
9	1.9	1.4
10	1.5	1.3

#Input	Input	
	x1	x2
1	*0*	
2		
3		
4		
5		
6		
7		
8		
9		
10		

ART-2a

In 1991, Carpenter, Grossberg and Rosen [150,151] modified ART-2 for rapid category learning and recognition and named it as ART-2a. It performs an unsupervised classification of n-multivariate-m-dimensional data vectors into data clusters of higher internal similarity. It is three times faster than ART-1. ART-2a is similar to k-nearest neighbors (k-NN), but faster compared to KNN. The numbers of pair wise vector comparisons are $\frac{n^2 - n}{2}$ and $ntest\_vector * trained\_weight\_vectors$.

The limitations of ART-2a leading to remedial measures are

Chart 4: Negative features of Art-2

- Accepts binary/analog data, but not floating point values
    - Remedy : Fuzzy-Art
- Sensitive to noise and inaccuracy of input
    - Remedy : The remedial measures are auto scaled feature
        - pre-processing (used ART 2a), PCA of sub matrix of circular and
        - octahedral samples and Fisher weighting
- The efficiency depends upon the nature of the features of input patterns.
- For patterns of different categories of features, the classification is inefficient
- The noise is considered as informative signal and vice versa

Wienke [994] adapted Minkowski similarity metric based on Eucledian distance in ART-2a. Scaling to unit length allows the use of vector angle as a measure of dissimilarity. This extension enables ART-2a to handle peak ratios in chromatography, spectroscopy and concentration profiles of emission sources in environmental analysis.

ART3

ART in Function Approximation

MIN-MAX	1993
RFALCON	1996

Prob-ART	1996
Fas Back	1997
dART MAP	1998
Prob-ART	1995
Fas Art	1996

Carpenter and Grossberg [152] proposed ART3 in 1990 using hierarchical search employing chemical transmitters in self organising pattern recognition architectures. It performs parallel search or hypothesis testing of distributed recognition codes in a multi-level networks. It incorporates 'chemical transmitters' to control the search process. In general ART-x [: ART-1, ART-2, ART-2a, ART-3, ARAM] NNs are designed for classification tasks and have advantages over MLP (Chart 5).

Chart 5: Advantages and limitations of ARTx architectures compared to r MLP

- + Autonomous learning in non-stationary environment
- + Novelty detection
- + Overcomes plasticity vs. stability dilemma
- + Dynamic allocation of nodes without network destruction
- + Faster convergence
- + Guaranteed convergence
  Reason: monotonically decreasing weights
  - + ensure stable learning
- + The linear prototype units are allotted dynamically based on the novelty detection

- — Does not accept supervised or floating data
- — Results differ with order of presentation

FuzzyART

Carpenter, Grossberg and Rosen put forward FuzzyART [150] in 1991. It accepts floating point (real continuous) data unlike ART-x NNs. In fact, binary data can be inputted as well. It is an unsupervised classification NN yielding good results (Alg 2) even with a few training samples. Like ART-x models, it does not require a priori information regarding the number of clusters.

Architecture.Fuzzy ART: The architecture is identical with that of ART1. The set theory intersection operator ($\cap$) is replaced by fuzzy set theory conjunction or MIN operator ($\wedge$). The activation function and connection between neurons are replaced by fuzzy-membership functions and Fuzzy rule in fuzzy-ART neural network. If the input is binary, it reduces to ART-1 with low CPU time. In the fuzzy-ART both bottom-up and top-down weights are subsumed in weight (WFART). Typical rules regarding hyperbox, matching etc are given in KB-2. Although, there is no loss or gain in information content, the advantages and limitations of Fuzzy-ART are in chart 6.

Alg 2: Fuzzy-ART

Step	:	-1	Input : Real or binary data
			All categories are set to be uncommitted WBU(i,j) =winit(j,n) j = 1,2, ….,N
		Repeat	For all training samples
Step	:	1	***Preprocessing***

Normalization with Euclidean distance or
complimentary code of input
→ It avoids category proliferation

While Similarity >=vigilance

Step : 2 ***Competing stage***
Input to jth node in the category (F2) layer

Calculate neuron satisfying winner takes all
(WTA) rule
If more than one uj is maximal
Then output node with smallest index is chosen
tie is broken

Step : ***Matching stage***
Similarity :Top Down (TD) activity of LTM (v)
(F2 neuron J) is compared with input pattern

Step : 3 ***Vigilance test***
If Mismatch occurs
Then Weighted sum of uj of the winner is reset to -1,
till algorithm returns to step 3
If Category is selected
Then It is committed

If Winner is j &
J passes through vigilance test
Then Up gradation of weights
End while

End repeat

KB 2: fuzzy-ART-NN

If granularity is higher
Then pattern match gets finer
$vig.par \rightarrow 1$

If There is no match found in existing patterns during classification
Then reset unit is activated

If data are not uniformly distributed within hyper box
Then Fuzzy ART predicts (infer) existence of data in corners of rectangular regions support where no evidence occurs
Remedy : GART

Chart 6a: Positive features Fuzzy-ART

- Non-binary input vectors can be processed
- Single weight vector connection
- Hyperbox shaped input regions generated by Fuzzy-ART-NN are interpretable
- Activation and match functions
- Fuzzy-ART works correctly

Chart 6b: Short comings Fuzzy-ART

- Euclidian norm looses information regarding vector length
  → alteration of the information content of non-normal data set
- Choice parameter, learning rate
- Fuzzy Art stores only minimum values and thus information regarding the maximum value is lost

Growing fuzzy topology ART:

The growing fuzzy topology ART (Grow.FuzTopol.ART) infuses the advantages of fuzzy-topology, growing cell structure and adaptive resonance theory. Here, training is with push–pull learning strategy (Chart 7).

Chart 7: Advantages of Growing fuzzy topology ART

- **Architecture**
  - Reduces proliferation of nodes in the F2 layer
  - Enhances class- dependent clustering ability of the GCS model through fuzzy-ART mode
- **push–pull training algorithm**
  - Enhances the cluster discriminating property
  - partially improves disadvantages of forgetting

Alg.

- create GCS with 3 nodes (GCS3)
- Initialise first node of F2 with GCS3
- replacing each F2 node with a GCS by including categorization property of the conventional fuzzy ART model

3.4.1 Applications.ARTx

The applications of ART-NN models are omni-fold encompassing science, engineering, medical-diagnosis/treatment/surgery, industry, defense and human-brain-mind-consciousness processes. The remote sensing, airplane design, and the control of autonomous adaptive robots etc. are other typical areas of utility.

Technology transfer of ART-NNs

The technology transfer of ART-NN to (Boeing Corporation) industry and Governmental agencies (MIT) Lincoln Laboratory is a testimony of not only academic excellence but also utility of these models for real life tasks. The application of neural networks in Neural Information Retrieval System [136] in Boeing is probably still on the largest scale. ART performs clustering of binary templates of aero plane parts in a complex hierarchical network covering over 100,000 items, grouped into thousands of self organised clusters. The savings to a tune of millions of dollars per annum in manufacturing costs are claimed.

[Environemntal , industrial, time-dependent] chemistry

Wienke and Buydens [8] , in mid 1990s , reviewed an overview of ARTx-NNs in chemical applications. The pattern vectors from autoregressive (AR) model of process analytical time series are classified by ART-2. The detection of 'not-normal-states' for continuous control of chemical reactors and monitoring their normal health was modelled with ART2-NN. Kateman et.al [8] employed ART-1 in classifying and interpretation of UV/VIS and IR spectra. The ART-weights after decoding were shown of chemical significance and correspond to spectral absorbance. The monitoring of ambient air in environmental studies was modeled with ART-2a and weights are chemical-perceptors outputting chemical-environmental-phenomenon. ART-2a is comparable to SIMCA and better than MLP-BP-NN in rapid sorting of post-consumer plastics by remote NIR-data.

Miscellaneous applications

ART-x improved the analysis of aerial images, aircraft databases, nuclear power activity , hand written characters, speech, rule extraction from databases, optic-electronic processes, VLSI, radar signal information, pattern recognition in turbulent fluids, classification of chemical complexes – UV-Vis-IR data, automatic monitoring of signals chemical plants, semantic associations between terms in textual databases, extraction of rules from trained NNs, weather forecasting and prediction of stay in hospitals/ICU and post-surgery survival period.

3.5 $ART

Coupled-ART

It is an unsupervised real time learning system [110] introduced by Lee and directly accesses the learned patterns. The architecture of coupled ART consists of upper and lower layers. The lower layer has dual modules (ART-L, ART-R) running in parallel and is controlled by F3 layer. They correspond to the function of left and right cerebral hemispheres. The vigilance parameters (vig.par1 and vig.par2) control their function. The function of F3 is comparable to corpus callosum of the human brain.

Dataset. Coupled-ART: Coupled-ART-NN categorizes three alphabets (B, E, F) with local and global features better than ART-1. The replacement of ART-1 modules by ART-2 or ART-3 was considered as worth pursuing.

Prob-ART

Prob-ART [123] derives its name as this NN uses the theory of probability. It filters noise, learns only

$$\mu_{jm} = \frac{1}{\left|W_j^{xy}\right|} * \sum \varepsilon_{nm} * W_{jn}^{xy} \qquad m = 1, 2, ..., 2 * m_b$$

…… Eqn. 25

μ_{jm}	Predicted value for mth component of output vector associated with node j of ART-x
$\left\|W_j^{xy}\right\|$	Total number of associations between nodes in ART-y and node j in ART-x
ε_{nm}	mth component of vector associated with nth category in ART-y
W_{jn}^{xy}	Frequency associated with nth category of ART-y and node j of ART-x

basic signal in a better fashion. Thus, category proliferation, an inherent menace of ART-x models, is reduced for noisy signals. In Prob-ART-NN, reset mechanism and inter-ART-vigilance parameters are not used. Hence, prob-ART looses some of salient features of fuzzy-ART-NN. The size is governed by ρ^x, since ρ_{xy}. Inter-ART map weights records the number of times that two categories are counted or associated. The vigilance parameter remains constant as inter-ART reset is not used. In other words, the size of the categories in ART-a is kept constant. The absence of inter-ART vigilance parameter allows one-to-many relations between neurons in ART-a and ART-b modules. Here, the weights reflect the importance of each relation. Chart 8 compares features of Prob-ART with Fuzzy-ART NN.

Chart 8: Comparison between Prob-ART and Fuzzy-ART

Prob-ART	Fuzzy-ART
Suppression of inter-ART set and corresponding mechanism of match tracking	---
$F2^x$ It is totally unsupervised as in $F2^y$	Generation of categories in $F2^x$ is guided by equivalent process in $F2^y$
Exception handling included reflecting the frequency of association between categories in the inter-ARTMAP weights If low frequency Then exception	Contradiction between past-knowledge and new input pattern generates a new category

Dataset.simulated.Prob ART: A set of 1000 random values (x) are generated and the response (y) is simulated with and without noise. The results of modeling of datasets show FasArt and FasBak, are superior to fuzzy-ARTMAP in the size of neurons in ART-x, ART-y, RMSE and maximum error. Prob-ART requires less number of neurons in ART-x module.

Correlation based ART

Yavas [16] introduced correlation based ART (corr-ART) NN, which uses correlation analysis for category matching. The output of the first level corr-ART is fed to the input layer of second corr-ART NN. This two layer hierarchical structure performs one more categorization operation over the output of first layer categories. The back propagation of matching information from the second layer of NN to the first is implemented to analyze the application in Robots. The envisaged application is that the observer robot can detect abnormal conditions arising in a machine operated in a factory. Of course, all forms of machine operations possible are learned during the training secession.

Positive features of Correlation based art

- **Architecture**
  - Architecture is expandable with reusable modules
  - Model is hierarchical
- **Classification**
  - Improves the quality of classes
- **Learning**
  - System learns different forms of the behavior
  - Detects the relation between them.

Projected-ART (proj-ART) [83]

The clustering algorithms are not effective for high dimensional data sets. Sparsity of data hinders to search for clusters in full feature space. On the other hand, pruning of full feature space results in information loss. It is referred as feasibility-reliability dilemma.

In the competition/learning phase only committed nodes of F2 layer accept input signals from F1 layer. The non-committed nodes do not participate in the learning competition. If previously committed node is not chosen for a current pattern, naturally a uncommitted node becomes the winner at that step. Similarity test performed with a hidden layer for each of F1 nodes is the key procedure of projected-ART.

Architecture.projectedART : The structure of F0, F1 and F2 layers is same as that of ART-1. But, each node in F1 layer has a hidden one . It is for similarity check to assess whether the node (v_i) in F1 is active relative to a F2 node (v_j). F1 layer selectively sends signals to the node in F2 layer. A node in F1 layer can be active with respect to some F2 nodes, while inactive relative to others.

Dataset.simulated.Proj-ART: A twenty dimensional data set with 5 clusters and 10,000 data points is simulated. Each cluster is in a subspace of 7-dimensions. Proj-ART successfully found the number of clusters and centers of original clusters with a small error.

Dataset.simulated.Proj-ART: The data set with five clusters in 29,23,15,26 and 33-dimensional subspaces and a total of 10,000 points is analysed. The number of outliers is 500. Projected ART correctly outputs the number of clusters and sufficiently large subspaces where the clusters are formed.

Dataset.simulated.Proj-ART: Two data sets with eight and twelve patterns in four dimensional space are simulated and analysed.

Proj-ART-buffer NN: The projective ART-buffer-NN is an extension by introducing buffer management and a new similarity (degree) function. The buffer does not allow immediate clustering of datasets into a single cluster. The average similarity degree successfully works with high similar noise datasets. Further, it results in an order-independent objective without correct parameters. The buffer checkout process handles huge amounts of input by a small buffer space.

Proj-ART-buffer

+ Does not depend on precise choice of input parameters Proj-ARTNN

— Requires accurate parameters and good input data

Remedy: Proj-ART-buffer NN

P(erformance) guided -ART (perform-ART)

P(erformance)-ART is a semi-supervised modular learning NN. The ART and LVQ learning are combined in consecutive stages. Further, there is a performance feed-back in this network. The advantage of snap-drift learning method incorporated in ART-learning module is that it relearns and destabilizes with changes in input or performance. It punishes lower performance of NN by reducing the weights of connections while, at the same time rewarding winning nodes. .

If	Performance is poor and slower
Then	ART-learning (minimal list or snap)
If	Performance is good
Then	LVQ with continuous learning or drift

Architecture. perform-ART: It has a distributed ART architecture and thus has more than one winning neuron and here three nodes are employed. The three F21 nodes with highest bottom-up-activation are selected. At the converged stage of distributed output category, three highest activation level F21 nodes are in resonance state.

Functioning of *perform* -ART

When an input pattern is presented at F01, it is transmitted to distributed p-ART for feature selection by snap-drift algorithm [77]. Here, ART reset mechanism is operative. The size of F0 is fixed and is equal to the pattern width (maxmimum of 50 bits). The three (D=3) winning F2 nodes best matching the input pattern are the input for selection p-ART, which selects only one appropriate output. It is also called proxylet in telephone communication terminology.

Like in SOM, if no single output node explains the pattern, one of its nearest neighbors is selected. The number of F22 nodes is less than 100.

dp-ART: It learns the key features in the patterns. They are the input to SP-ART which selects appropriate output, which references a desired location on application specific SOM net. It learns key features and becomes input to selection performance (SP)-ART.

Alg 3: Alg. *perform-ART*

Initialise parameters.

For each new input pattern

For top D distributed winning nodes:

Find the candidate F21 node to learn the current input pattern i.e. F21 (which is not inhibited) whose prototype is most similar to the current input pattern.

Test reset condition

If reset is true

Then the current candidate F21 node is rejected; inhibit node and return to 1.

If reset is false

Then current candidate F21 node is accepted for learning;

Learning

Weights of dP-ART are adapted according to the snap-drift learning style indicated by the external performance feedback value

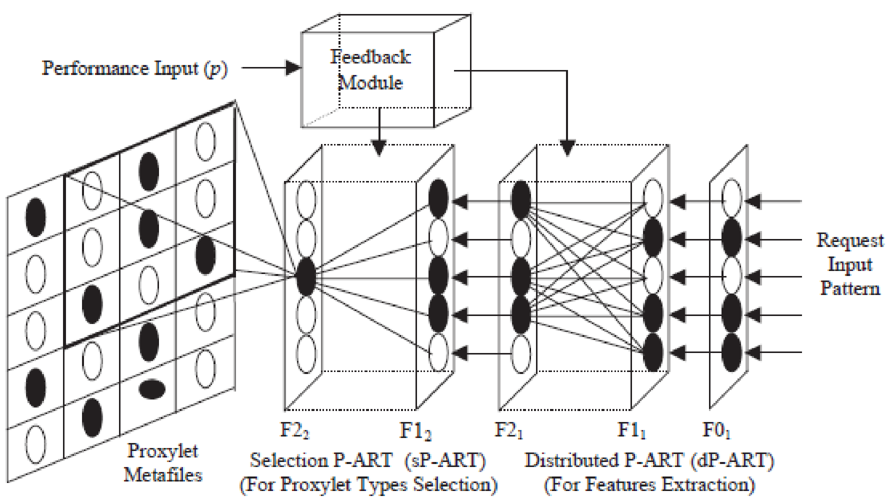
Process the output pattern of F21 as an input pattern to F12

Find winning node in F22, using the same procedure

Adjust sP-ART weights using the snap-drift determined by external performance feedback level

End for

End for



Grey-ART

Liu et al. [78] and Yeh et al. [71] used grey system analysis along with ART resulting in Grey-ART, a hybrid module (Alg.4). Grey-ART accepts binary as well as multi-valued inputs.

Grey relational analysis refers to an incomplete, uncertain low quality data/information/model. For instance, the functioning of human brain, prediction of economy/traffic, self organization of data into classes and inverse modeling are instances of grey-tasks. Grey relational analysis calculates a similarity measure for a reference and test sequence of finite length with incomplete information.

Alg. 4: Grey-ART

Step : -1 Input : X(i ,j)

Step : 0 Weight vector of jth neuron

Step : For i = 1 : NP

_______ **Grey relation** ________________

1 Calculate similarity measure

selection of winner largest relational grade

$$\max_j \{g(x_i, y_i)\}$$

Let the winner be denoted by jwinner

______________ART__________________

Step : 2

If $\{g(x_i, y_i*)\} \geq \rho$

Then jth neuron passed vigilance test

Else jth neuron failed vigilance test

Step : 3 .

If jth neuron failed vigilance test

Then a new neuron unit k is created with weight vector

$$y_k = x_i$$

Step : 4 If jth neuron passed vigilance test

Then Upgrade the weight vector of the winner (Eq.)

End repeat

Dataset,simulated.Grey-ART: Yeh [71] considered 2-D data with three non-overlapping clusters with centers [0.2,0.3],[0.7,0.7], [0.4,0.3]. The results with Grey-Art at maximum ρ (=0.6 to 0.8) are consistent with the actual centers of the clusters.

Dataset,simulated.Grey-ART: A data set (NP = 200) of four clusters of different shapes and sizes was analysed at a maximum ρ of 0.7. The performance of Grey-ART (0.63) is greater than ART-2 (0.61).

Dataset,simulated.Grey-ART: Lu et al. [78] analysed the pattern (x) with fuzzy variation and variation of intensity of background with Grey-ART model.

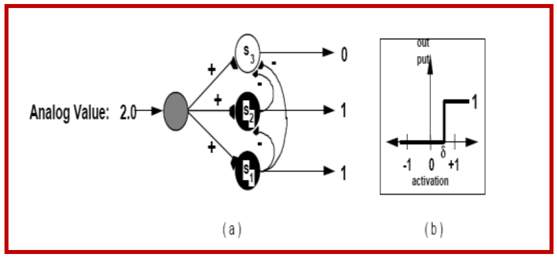
LAteral Priming-ART(LAPART)

LAPART-1 is an adaptive–inference-NN with a specific purpose of extracting rules or logical inference relationship in classification tasks.

Architecture.LAPART-1: This NN consists of two ART1 modules which are coupled laterally with adoptive connections (Fig. 2). LAPART has fixed vigilance parameter unlike ARTI. It doesn't implement complement coding.

Stack net: It is a pre-processing NN converting a real value into a string of binary bits used in

Fig.2: Conversion of floating point value into binary bits



LAPART-NN. The maximum and minimum of the real values of the data vector is determined. The range is quantified into m-subunits, depending upon the accuracy of measurement. Each subinterval produces 1 if the value > δ, and zero otherwise, where the threshold (δ>0) is a user defined parameter.

Functioning of LAPART-1: To start with there are no learned inferences and connections of F2A → F2B are weak. For the first input pair, the first node in F2A as well as that in F2B are committed and results in first inference rule F2A(1) → F2B(1). From then onwards, a new rule is formed or tests the previously learned ones. The algorithm is described in Alg-4. In LAPART, original ARTMAP rules are slightly modified to define end of learning process and positive/negative features are incorporated in Chart 9.

Alg 4: *LAPART1*

There are no learned inferences
The connections of F2A → F2B are weak

If First pattern
Then Neuron in F2A and F2B are committed like in ART1
Due to the strong connection between F2B ← F2A
Inferencing relation is developed
Classa(1) ← IA ; classb(1) ← IB; class(1) ← IB
Rule(1) = IA(1)→IB(1)

For i=2:NP

Input a pair of patterns IAi,IBi

If ith pattern resonates with nth pattern &
There exists the rule Rule(m) = IA(k)→IB(j)

Then A strong connection occurs between F2Ai → F2Bj
Read templateTB(j)
The vigilance performance test is now performed

If [IB/TBj]/[IB] < vig.par.B

Then Inference is discarded &
Reset in subnet B occurs Lateral reset i.e. vig.par.A of the subnet A undergoes reset
It overrides subnet A's decision and searches for alternate class.
At this point nodes of F2A and F2B are reduced
End for

Chart 9: Features of LAPART-1

+ Converges in two epochs
[yet maintains good generalization]
+ Lapart-1 overcomes the template proliferation
+ Results are independent of the order of presentation of input

− There is no guarantee that a LAPART-1 network reaches convergence on the training set
Remedy : LAPART-2

LAPART-2

For binary inputs, LAPART-2 converges within two epochs. It appears that it is the first fixed pass procedure in ART-family of NNs. It is worth to emphasize that ART-1, ART-2 etc are finite-n(umber)-pass methods. In an epoch, the complete list of input patterns is presented and is referred as on-line learning. It enables extraction of rules from both input and output spaces of classification. The noteworthy feature is that it partitions both input and output spaces. Generally, many other methods simply label the classes.

Architecture. LAPART-2: The architecture is same as LAPART1.

Functioning. LAPART-2: The procedure for a lateral reset is modified which results in a rule extraction neural network that converges in two epochs to train a data set. In LAPART-2 only uncommitted F2A nodes are susceptible for reset. By recoding IA to uncommitted template, the current knowledge of new class is stored. Another possibility is through forming strong connection between Ai' → Bj'.

Dataset.simul.LAPART-2: Simulated datasets consisting of two overlapping classes following rectangular and normal distribution of points are analysed with LAPART-2 (Table-6). The learning experiments were conducted with varying vig.par(0.1,0.2, … , 0.9,0.95). The results are compared with Bayesian classification procedure.

Table-6: Classification with LAPART-2 of simulated overlapping classes

NP_Class1 = 500; NP_Class2 = 500		Performance	
		LAPART-2	Bayesian
rectangular uniformly	50% overlap two equal sized	75	75
normal	Mean :0.333, 0.5) and (0.666, 0.5) sigmas (0.166, 0.166),	81	84
overlapping normally	Mean (0.5, 0.5), (0.5, 0.5), sigmas (0.166, 0.166) and (0.333, 0.333)	65	73

Here, non-statistical NN is compared with Bayesian classifier, a statistical information system. In LAPART-2, the ratio of the number of learned A sub network templates to the total number of training samples is 0.25, reflecting very little memorization, consistent with the good generalizability. Yet there is a little memorization due to the architectural and procedural implications.

Multi-ART (MART)

Multiple channel ART [119] is an unsupervised on-line learning NN. It uses the principles of ART-x models. The potential weakness of ART1, ART2 and traditional statistical classification procedures is category proliferation. Multi-ART with channel weighting mechanism minimizes the proliferation of categories even for noisy data. It accepts data from multiple sensors/channels/input paths simultaneously. The concept of channel credit, radii and global vigilance are introduced which enable learning of different classes by adaptive and individual discrimination philosophy. Some of the unique features of MART are

- one pair of F1 and F2 layers for each channel (They are represented by F1c and F2c where c = 1 to number of channels
- integration of the categories through F3-layer
- Direct communication of orienting subsystem with F3 and F2s
- No direct connection between F2s and F3 layer
- Dynamic creation and suppression of classes
- Use of city block distance as a measure of similarity

Architecture.Multi-ART: In ART-1 there is only one F1(feature) and a single F2(category) layer, which are connected through orienting subsystem. But, MART has a pair of F1 and F2 layers for each channel. A pair in cth channel (block) is represented by F1c and F2c which are fully connected in both ways. The number of PEs (or neurons) in each F1c is equal to the length of the pattern in that channel. The size of F2c and F3 is dictated by maximum number of distinguishable classes. The block thus has pairs equal to number of channels. The output of each node of F1c is connected to cth neuron of orienting system. The outputs of c-neurons of orienting system are fed to c+1 neuron representing global vigilance.

Each node in F2i,k is connected by an unweighted two-way link to the nodes in F3k. The output of each node of F1 is connected to ith node of orienting system in addition to those from F3. The single output line to the global node Re sends signal to nodes in F3.

Class selection

The local similarities between the input pattern and the learnt classes (template) is determined for each channel in F3 layer. They are the input to F4.

F2 layer: The input to F2 layer is the output from F1i. The output at F2 from the channel ' i ' represents one of the categories it distinguishes. The dissimilarities between two vectors v1 and v2 is within the range zero to one. A value of zero means total coincidence.

F3 layer : It implements WTA principle for outputs of F3 (local similarities) applying competitive learning rule. The global similarity index (pK) is the maximum of the local similarities for ' b blocks' and is calculated. The output of F4 (template of classes learnt) is passed downwards through each channel to determine the local differences for a new input pattern.

Orientation system: It calculates the difference between the current input and set of all templates for the global classes established. It decides whether there is resonance or there is a need to include a new class. The prime function is to arrive at a global result of all local comparisons. It is instrumental in increasing the number of categories during on-line learning of new patterns. The channel credit is a measure of the quality of signals of a channel and represents the weight during global classification. Class radii correspond to the weights inside the class radii block. *Class manager* controls the dynamic creation and suppression of classes

Data flow in the forward (bottom up) direction: The classifications of multi-channel patterns in MART can be visualized through the data flow in forward (bottom up) and reverse (top down) steps (Alg-6).

Alg 6: Data flow in the forward pass MART

For Channel = 1 to K
 Input enters classifier via F1ij
 Compares it with the templates associated with F1i-to-F2i connections
 Total upward output for F2ik(as a measure of similarity) between Ii and zik
 Summation in F3k from all inputs from other channels
 Horizontal connections in F3 implement competition process
 The result is only one node with highest similarity to
 input pattern is active and all others are inactive
endfor

Data flow in reverse (Top-down) propagation: The vector output (out.F4) is propagated to F5 layer. Its output is zero except for the winning neuron. The output (di) of the ith single channel represents the difference between input pattern and that of class k1 in the channel. This output is propagated towards orienting system.

The result of computations in F3 is propagated back to F2s. All the nodes in F2i are inhibited except the winning node ($F2_{i,winning}$). The weights corresponding to the winning node are transmitted to F1i. The weighted input corresponds to the NN calculated pattern and its difference from the observed pattern is the deviation. Then it is sent to orientation system. All the channel differences are weighted depending upon the current credibilities and summed. The global vigilance parameter is compared to check for resonance or creation of a new category (KB-3).The application of MART for simulated patterns and ECG data proved its superiority in classification of multi-channel patterns.

KB 3: Refinement of global parameter Multi-ART

If $d > rk''$
Then $\Delta\rho$ increases

If $d < rk$ &
$\rho < \rho_{min}$
Then $\Delta\rho$ decreases

While new class is not formed
 □*kg* remains same
Endwhile
 □*ref* : initial value

Dataset.simul.ECG.Multi-ART:

A typical ECG of a healthy person is given in Fig. 3. A set of 2000 patterns from five basic morphologies in four channels are simulated and analysed with MART. The length of each pattern is 125 binary bits and Fig. 4 depicts typical patterns.

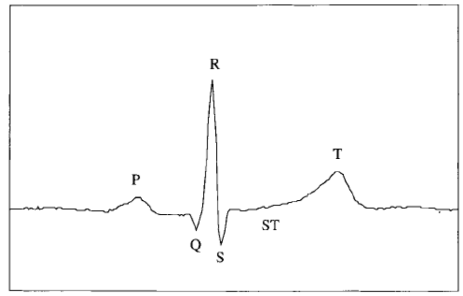


Fig. 3. ECG elements of a typical cardiac cycle (Courtesy ref

Data set.ECG.Multi-ART: The two channel ECG traces from MIT-BIH database are typical in the sense that they contain relatively large number of ectopic ventricular beats with anomalous QRS morphology. A trace belonging to one out of a maximum of ten categories has 125 bits (0s and 1s). MART is implemented on PC with 200 MHz Pentium Pro processor with 32 MB RAM under OS Solaris x86 v2.5. Typical analyses of six ECGs with MART agree with expert cardiologist's diagnosis. The average sensitivity and predictability exceed 97% (Table 7). Further, the effect of addition of normal/sine wave noise (Fig. 5) showed that MART is robust as revealed by the result that misclassification rate is zero for four out of six traces. The proliferation of categories in presence of noise is minimum. MART is integrated within an intelligent cardiac (multichannel ECG) monitoring system and is in operation in ICUs.

permission

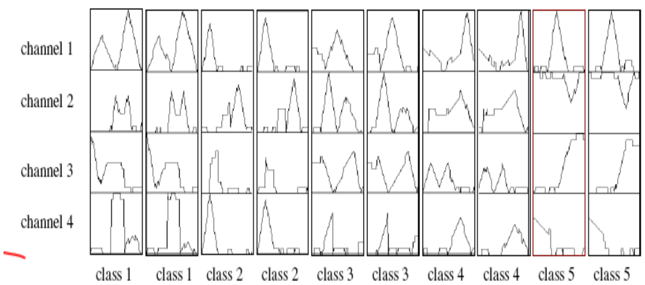


Fig. 4: Simulated patterns for MART from five classes in four channels

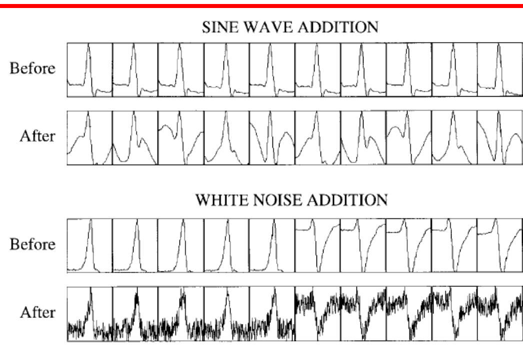


Fig. 5. upper rows : Noise-free beats ;
corrupted signal after addition of sinusoidal noise and white noise (courtesy)

Table 7: Classification of *ECG* traces by MART

Trace	Se%	+P%	No.. of beats
105	100	100	59
213	1--	100	75
219	87.5	96.58	64
221	100	100	35
228	98.56	96.88	105
234	100	100	72
Average	97.68	98.96	

Efficient-ART

In ART, mismatch reset condition and repeated search process are time-consuming. Baraldi and Alpaydin proposed Efficient-ART-1 [86,87] eliminating these hurdles. The two versions viz. Effic-ART-1 and Effic-ART-2 differ in the properties of activation and match functions. Effic-ART version 1 (Alg.7) is equivalent to the sequential

KB-4: Efficient-ART

If	match function does not increase monotonically with activation function
Then	EART-1
If	Activation function increases monotonically with match function
Then	EART version 2 holds

version of the parallel Adaptive Hamming NN (AHN). EART-1 implements ART 1, improved ART1 (improv-ART1), Fuzzy ART, or AHN. The EART-2 is more efficient than EART-1 and KB-4 depicts heuristic-rules for their applicability.

Alg 7: Efficient-ART-1

Initialization
PE counter and presentation counter are set to zero

For Pattern = 1 to NP
Present input pattern

Detect PEs eligible for Resonance-vigilance test
Apply vigilance test (Eqn. 3) Orienting subsystem employs
Put set of PEs which passed through the test,
into ensemble (number >1)

If size(PE-ensemble) = 0
Then go to step 4b

Resonance Domain Detection—Activation Value
Best-Matching unit selection
Select PE with largest activation in PE-ensemble

Resonance condition—Reinforcement learning
-hard-competitive weight adaptation
adjusted W^i by ART adaption law

If soft-competitive learning strategies are adopted
Then other prototypes are considered suitable for adaptation

Nonresonance condition—new processing element allocation
If size(PE-ensemble) = 0 i.e. no resonance
condition is achieved
Then allocate one new processing unit dynamically
to match external expectations

Go to step 1
End For

In Efficient-ART-2 , Step 3 of is Efficient-ART-1 replaced by

Step : 3 **Resonance Domain Detection—Activation Value Computation and Best-Matching Unit Selection (1):**
If vigilance test passed
Then go to step 4a
Else go to step 4b

The results of IRIS and Simpson data with Efficient-ART are compared with S-fuzzy-ART, fuzzy-ART, FOSOM-ART.

3.6 Hybrid-ART-NNs (ART + $$$]

Fully Organized self-organizing-ART (FO-SO-ART)

Balardi and Alpaydin [86,87] proposed FO-SO-ART (alg. 8) based on Extended- ART-2 to overcome the difficulties in Fuzzy-ART. The prime object is to overcome the shortcomings of Fuzzy-ART. It is a

constructive, on-line learning, topology-preserving, soft-to-hard competitive, minimum-distance-to-means clustering network. It belongs to SART type and imbibes all good characteristics of Fuzzy ART, self-organizing map (SOM) and neural gas (NG). The input to FOSOM-ART is finite set of analogue values in d-dimensional space.
The application domain of FOSOM-ART (like growing NG) extends (Chart 10) to vector quantization, entropy maximization and structure detection in multi-dimensional input data and the positive and negative features (Chart 11) are worth consideration.

Alg 8: FOSOM-ART

Repeat for all patterns
  Choose a new pattern sequentially or randomly

  Select best and next best matching units
  If Vigilance test passed
  Then same category
  Else no resonance , search for a new PE
  Learning
  Resonance Condition—Reinforcement Learning:
  Increase the best matching counter unit
  Learning though constrained competitive Hebbian rule
  Apply soft-to-hard competitive update to output units belonging to the same map of best-matching unit

  Nonresonance Condition
  Allotting of new Processing Element dynamically

End repeat
superfluous cells are removed
Superfluous lateral connections are removed

Check for Convergence:
If not converged go to Repeat

Chart 10: Advantages of FO-SO--ART	**Not implemented in**	
Minimizes a ESS via a soft-to-hard competitive model transition	SOM	GNG
parameters are not affected by outliers In fact they are instead mapped onto noise categories	SOM	NG
requires no randomization of the initial template vectors	SOM	NG
requires no *a priori* knowledge of the topology of the network	SOM	
Explicitly deals with lateral connections	SOM	NG
Addresses all constraints required to make the **CCHR** guarantee perfect topology-preserving mapping	GNG	
Fuzzy ART requires no complement coding of the input data	Fuzzy ART	
removing noise categories to avoid overfitting	Fuzzy ART	

Chart 11: Features of FO-SO--ART

- The computation time increases linearly with number of links
- The network size is automatically adjusted with complexity of task because of feedback interaction between attentional and orienting Subsystems
- It is robust to noise and avoids over fitting as the neuron removal strategy is in operation

- Chances of dead unit formation are meager in FOSOM-ART Compared to random or splitting by other initialization techniques. It distributes initial reference vectors in the input-manifold uniformly.
- FOSOM-ART is less prone to be trapped in local minima, compared to hard competitive clustering techniques
- The competition strategy among lateral connections (of FOSOM-ART and GNG) in a constructive framework, is superior to that of traditional constructive clustering systems
- FOSOM-ART implements better strategy compared to NG and SOM which are non-constructive clustering systems and also have no lateral connections explicitly
- The expressive power of FOSOM-ART and GNG with competition among lateral connections in a constructive framework, is superior to that of traditional constructive methods
- The domain of an applicability is broader than that of Fuzzy ART, S-Fuzzy ART, and GART
- FOSOM-ART trains faster than NG, SOM, and Fuzzy-LVQ

- FOSOM-ART does not minimize any known objective function, i.e., its termination is not based on optimizing any model of the process or its data.
- The results are dependent up on the order of presentation of patterns as a consequence of example driven generation of reference vectors and lateral connections
- Example-driven learning strategy is more sensitive to the noise in the input
- Uses two parameters er and λ_{init} in addition to row and minimum number of epochs

Dataset.simul.FO-SO-ART: Three Gaussian clusters and circular ring (Fig.6) with a total number of 160 points forms non-convex dataset. With two epochs FOSOM-ART trained successfully data using 11 templates.

Dataset.spiral.FO-SO-ART: 194 noiseless patterns belong to two concentric spirals (Fig 6). The points in the outer spiral are farther away from points in the inner spiral. FOSOM-ART generated two spirals (Fig 6) with 148 templates and 16 maps. The performance is superior to growing cell structure. This task is a hard nut for MLP-BP.

Dataset.HumanFace FO-SO-ART: A 3D-digitized human face with 9371 vectors (Fig. 6) is trained with FOSOM-ART. Tthe output of FOSOM-ART is with 1745 nodes and 19 maps with MSE of 2.98 and 15 epochs. The performance excels that of NG.

Dataset.IRIS. FO-SO-ART: IRIS data set and Simpson cluster data set are also analyzed and found FOSOM-ART is competitive with other clustering algorithms for 3 clusters.

(a) ***(b)***

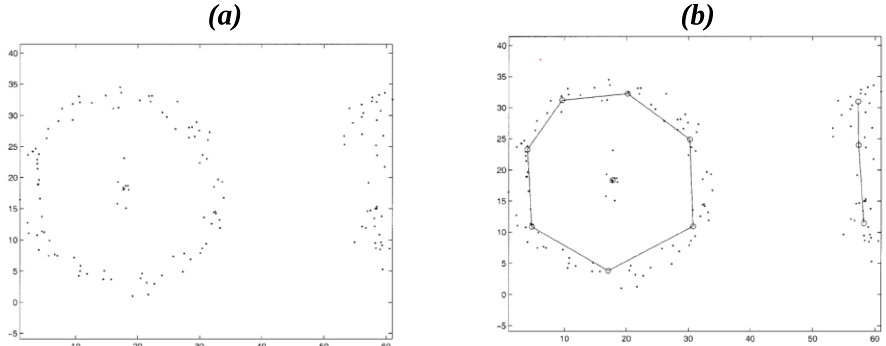


Fig. 6: Output of FO-SO-ART
(6a) circular ring and three Gaussian clusters. (b) FO-SO-ART Ouput

(a) ***(b)***

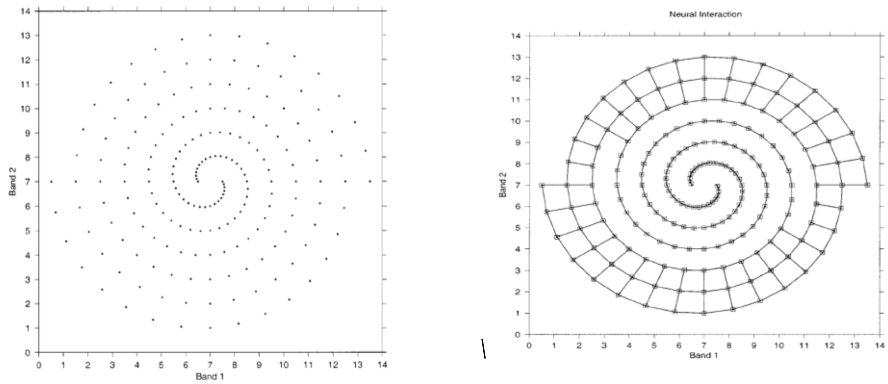


Fig. 6b. Two-spiral problem (a) Dataset (b) FO-SO-ART Ouput

(a) ***(b)***

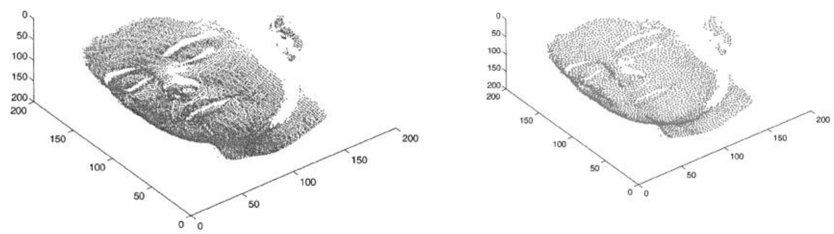


Fig. 6c. 3-D digitized human face , (e) Dataset (f) FO-SO-ART Output (courtesy Ref: 86,87)

Fusion Architecture for Learning, COgnition and Navigation (FALCON)

Tan et al. [50] proposed a fusion NN architecture based on fuzzy ART with SARSA(state-action-reward-state-action)- learning. It is used for learning mine-field navigation and the results are much better than standard-gradient descent based methods. It accepts bounded input in the range 0 to 1 with discrete action set.

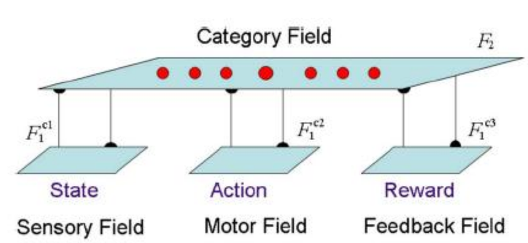
.

Architecture. FALCON**:** It consists of three F1 (feature) layers corresponding to sensor response (state), actions (motor field) and feed back is connected to all the three F1s. Each F1 layer has separate WBU and WTD.

— FALCON inherited category proliferation from classical fuzzy ART

Dataset.artificial_mine.FALCON: The task is navigation in an artificial mine field between two chosen locations within specified time frame without collision with any mine positions. In this study 16 x 16 mine filed containing 10 mines is considered. The five sonar sensors cover 180^{o} forward view and the distance in the ith direction to an obstacle is based on measured sonar signal. The action comprises of move-left, move-right, move-straight, move-diagonally-left and right.

Fig 7: Fusion Architecture for Learning, Cognition and Navigation ((FALCON)



Cerebellar model articulation controller (CMAC) + ART

A hybrid NN with ART-2 CMAC components has multiple best features in learning (**Kb-5**). The results for data sets IRIS, Wine, breast cancer, liver and lung are compared with ART 2, SO-CMAC, hierarchical – CMAC and SO-HMAC.

Kb-5: Cerebellar model articulation controller CMAC + ART

If Winner fails a vigilance test
Then A new state is created
Else Memory contents of winner
Updated based on learning information

Learning

- Unsupervised art2
- Distributed learning information Into associative memory locations
- Som neighborhood

+ Implicit neighborhood structure in nn
+ No need not user choice

4. Supervised (self organizing) NNs based on ART

4.1 ARTMAP

In 1991, Carpenter, Grossberg et al. [147] introduced ARTMAP, a supervised auto resonance (ART) based learning system for a set of paired binary data. ARTMAP is a self organizing expert system. It cojointly maximizes predictive generalization and minimizes predictive error. The internal controller ensures autonomous operation of the system in real time. It learns quickly, efficiently and accurately compared to other algorithms. With only 50% data (for many bench mark datasets) 100% accuracy is obtained.

Architecture.ARTMAP: It consists of two ART modules namely ARTx (for input X) and ARTy (for output y) respectively for data transmission from external world to the software (Fig. 8). The two ART modules are connected by inter-ART-module. Match tracking allows sub-categories to be resolved in classification. Match tracking mechanism ensures that the selected category in the input space is correctly mapped into the corresponding category of the output space during training. In the training phase, binary input patterns (x) are given to ART-x and binary observed/correct (corresponding to x) patterns (y) to ART-y. The I/O pairs presented to ARTMAP are held at the network inputs for such a long time so that the weights converge to their steady values. Carpenter extensively treated the fast learning scenario. The advantages and limitations are briefed in Chart 12.

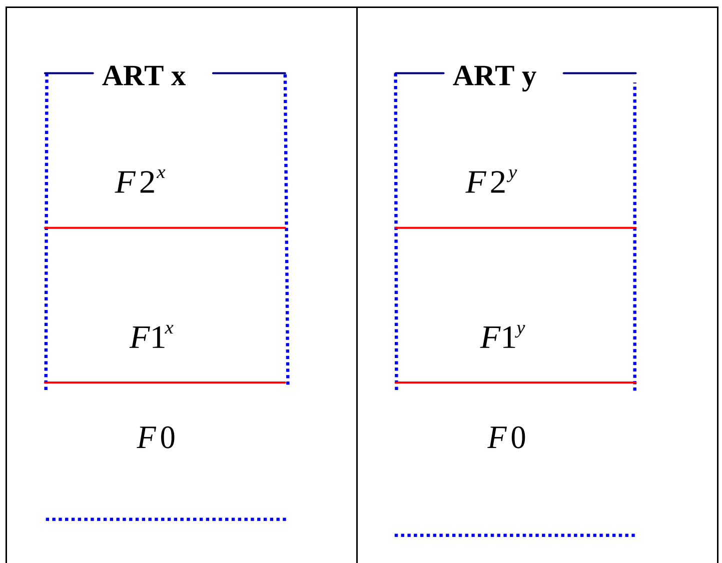


Fig. 8: Architecture of ARTMAP

Chart 12: Features of ARTMAP

+ Output of classification task is interpretable
+ Fast training compared to other NNs
+ Heuristic (IF_THEN) rules comparabl to FF NNs
+ Smaller memory is sufficient, [compressed representation of data → less CPU time compared to kNN]

Limitations of ARTMAP

- Category proliferation
  - **Remedy:** B(ootested) ARTMAP
- Large or too many rules for noisy data ~~(7)~~
  Output likelihood of an adverse outcome following an operation surgery

Inter-ARTMAP: The associative memory connecting the two fuzzy art modules, ART-x and ART-y, is called Inter-ARTMAP [95]**.** It stores the relations between the clusters in the input and output spaces. The relations (connections) are many-to-one. It's function is to determine whether correct I/O mapping is achieved [95]. The reset ensures that the learned categories are retained while new categories are learnt. It allows supervised learning in Fuzzy-ARTMAP, Fas-ART and Fas-back also. Inter-ART vigilance presetting is a form of back propagation of information.

Vigilance parameters: If the network performs a wrong prediction because of the already learned associative link, the vigilance parameter is increased (KB 6) by the minimal amount to correct the predictive error at ART-y module. ART-x then will search for another category for the current input until it searches a correct prediction. In case of an unsuccessful (failure) event, a new category in ART-x and its associative link to the corresponding ART-y category is created. Vigilance parameter controls the maximum allowed size of the category. In Fas-ARTMAP it indicates the maximum support for the fuzzy set.

Physical significance of weights in ARTMAP

Each category weight $F2^x$ roughly corresponds to a rule [85]. Each synaptic W can be translated into verbal or algorithmic description of antecedent of corresponding rule.

KB 6: Heuristic rules of ARTMAP

If	Noisy pattern is given to ARTX
Then	Unnecessary categories will be committed to account for noisy pattens
If	Selected category in artx predicts a wrong label
Then	Vigilance is raised $\longrightarrow$ finer or newer category selected $(\equiv unnecessary\ category)$
If	I/o mapping is many to one
Then	FAM operation can be described by artx alone. i.e. ARTY is not necessary
If	Identification of model is insensitive to noise in the learning phase
Then	No need of special data pre-processing techniques
If	Data is noisy
Then	Inter-art reset mechanism fires incorrectly
	Reason : Two points belonging to same category form a new category due to noise

Algorithm

Pre-processing

scaling X data in the range 0 to 1

complimentary code developed in $F0^x$

Learning phase

Presentation of I to $F1^x$
F2x categories developed
Expected categories in F2y through learned F2x→F2y association

Target input is fed to ART2y module

Testing/prediction phase

One or a set of test patterns presented to ARTx
inter-art map activates a unit of ARTy
Ws correspond to the predicted output

4.2 Evolution of ARTMAP [Evol.ARTMAP]

The functionality of ARTMAP-NNs has been improved in different directions over last two decades by Carpenter, Grossberg et al., the initiators of these innovative models in NN arsenal and also by other researchers. The basic philosophy [59, 70] of ARTMAP-NNs involves category choice, category match, learning (up gradation of W) and performance with test dataset and novelty detection. Although, WTA theme is retained, reset is guided by ρ_{xy} in ARTMAP operation. The modifications of ARTMAP or Fuzzy-ARTMAP are primarily to decrease prediction error, effect of order of presentation of patterns, network complexity, over fitting and category proliferation. The other objective is to increase generalization, detection/learning of unfamiliar patterns not present in the training set and enhancing the robustness to noise and outliers. Mostly, the basic architecture is kept intact and changes are suggested in learning and activation functions. The remedial measures are modification of architecture, complete/slow/distributed training or a priori (user chosen) maximum number of categories. A small set of highly productive categories is selected to implement complete training.

The category proliferation is less in fas-ART and fas-back compared to that in ARTMAP. Ordered fuzzy ARTMAP circumvents the ill effects of presentation order of training samples. Mu-(μ)-ARTMAP [85] and FAM-Charalampidis [95] deal with noisy data. ARTMAP-IC [115] handles missing/inconsistent or sparse data in classification task. FasART, FasBak , FAM-Marriott [131] are applied for function approximation of noisy data.

ARTMAP-Georgiopoulos_1994

Michael Georgiopoulos et. al [138] considered paired binary data set consisting of many explanatory variables and single y (MISO). The original rules of ARTMAP are modified and the number of epochs needed ranged between 200 and 999, depending upon the task. The complementary coding is used. Here, the results are independent of order of presentation and well suited for classification with fast learning of many-to-one mappings

Cascade-ARTMAP

Tan [124] incorporated symbolic knowledge into learning and recognition phases of neural computation. Multi-step inference is achieved explicitly through intermediate variables of rule-based knowledge. Later, a hierarchical ARTMAP NN for classification in probabilistic environment was proposed.

ARTMAP Rule Discovery (ARTAP.Rule_Discov, ARD)

Based on the learning strategy of ARTMAP, Carpenter in 2005 [67] designed ARTMAP rule discovery (ARD) system. After an initial learning phase, rules are extracted offline based on local rules integration. It is more informative compared to know just input/output behavior to understand how the system classified the objects into classes.

Self-Organizing (SO) ARD (Self_org. ARTMAP.Rule_Discov, SO_ARD)

Carpenter and Olivera [22] implemented on-line-rule-discovery in Self-organizing (SO) ARTMAP-rule-discovery (SO. ARTMAP. Rule_Discov) system (Fig. 9). It is applicable to small-scale as well as large-scale spatio-temporal data sets. Every time (either in training or test phase) a datum is inputted to SO_ARD, it extracts information and adds/upgrades confidence/ refines / the knowledge (bits) continuously and incrementally. Rules (mostly if-then-else type) are explicit functional relationships extracted from inter- and intra- implicit connectivity (consequences) of data. The confidence in self organized rule is encoded as the weight in a path from one class node to the other. The learning takes place in two stages. The first phase is a supervised training and in the second part only unlabeled feature variables (unsupervised X) data is analysed. The rate of learning of rule depends upon experience (how many times the similar/nearly equal I/O transactions occurred). The restricted SO_ARD simplifies to ARTMAP-default (Alg. 8). The rule based activation adds missing classes and deletes incorrect ones.

Experience-based rule learning	+ Minimizes creation of false rules during early Learning → inaccurate predictions

Alg.8: ARTMAP-default
Input : [Feature vector, class]
Initiation

Rules are excitatory or inhibitory	+ Rules discovered online from the start of learning + Rule-based activation + Adds missing classes + Removes incorrect ones	**Learning**_Stage_1 Match tracking Credit assessment Distributed ARTMAP learning **Learning**_Stage_2 Unlabelled input Extended_ARD

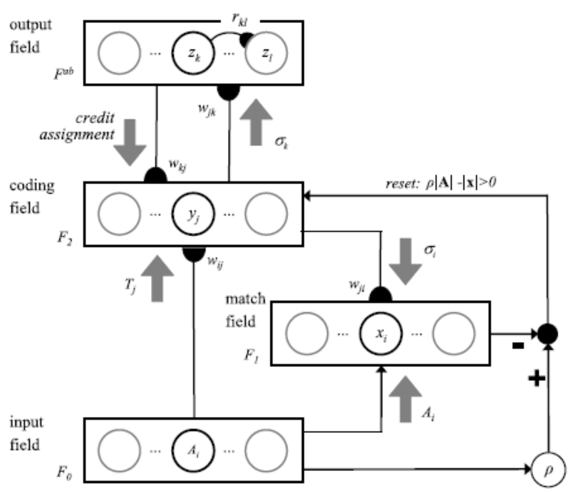
Rule self-organization: The system tries to predict output classes directly for each input (alg. 9). This self-organization mechanism continuously updates rules in the form of associative weights joining neurons (or nodes) representing the classes

Rule learning in SOARD

Supervised learning system: It learns many-to-one map recognizing different inputs corresponding to one class. Ex. Input features (Spot, Rex) belong to dog.

ARTMAP style: It also learns one-to-many maps retaining stability-plasticity advantage. That is, it learns a new class without forgetting the previous learnt class. Also it corrects erroneous predictions
Ex: For input 'spot' it learns a new class animal without forgetting dog. It also corrects erroneous prediction cat.

Fig. 9: Architecture of SOARD (Courtesy Ref : 22)



Alg.9: SO_ARD

Input
Set of [input samples & noise]

Stage_1
[ARTMAP_D_Stage_2] +
[Rule_based_inference]

Stage_2
[ARTMAP_D_Stage_2] +
[Rule_self_organization]
Voting

For j=1:nrules
If > 50% voters vote rule(j) & Confidence level > 50%
Then rule(j) valid
Else rule(j) invalid
end for

It is tested with datasets viz. new (audio) voices/letter and number datasets (Table 8). Simulations using small-scale and large-scale datasets demonstrate functional properties of the SOARD system in both spatial and time-series domains.

Table 8: Characterstics of datasets tested with SOARD

I/O pairs	Dataset	# I features	#O classes	Class levels
45,437	Voices	256	16	3

Table 9: Rules for Letter/number data with SOARD

Input	Class	Rule	%C
A, a	ay	ay→letter	99

34,714	Boston Large	36	13	3
23,327	Fenway	20	14	3
9,600	Boston Reduced	8	7	3
6,400	Letter and Number	8	6	2

B,b	be	be→letter	98
		one→number	100
		two→number	97

Dataset.voices.SOARD: The male and female adult voices are 16-bit PCM mono-channel sampled at 11,025 Hz. The comman noises present are sirens, dogs barking, construction, highway, and engines. The analysis of 45,437 preprocessed input vectors with SOARD showed a success of more than 90% on average in recognizing voice, male/female and noise.

Dataset. Letter/number.*SOARD:* This dataset contains lower and upper case English alphabets (A,a,B,b) and Roman and Arabic numerical (I,1,II,2).The input is a 8-D feature input vector and associated output class (letter, number). The knowledge discovery system produced the intra-class rules with high confidence (Table 9).

Dataset.Boston_satellie_image.SOARD: The rules, change in their confidence image data with ARD and SOARD NNs are briefly documented in table 11 for animals and water bodies against terrane constructions.

Table 10: Rule generation for Boston_satellie_image with SOARD

Input	class activated	Rule	Confidence	Rule	Confidence	instance
spot	Dog					1
spot	animal	Dog → animal	grows	animal → Dog	grows	2
Jinx	Cat					3
Jinx	animal	cat → animal	grows	animal → Dog	decreases	4
		Not learned before				
Rex	dog			Rex → animal		5

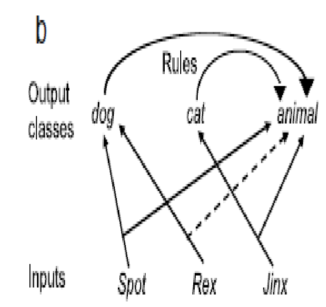


Table 11: Abridged self organized rules of ARTMAP rule discovery NN with ARD and SOARD

Correct rules				
Level 1	Level3	Level4	SOARD	ARD
beach→	open space		73	
beach→		natural	75	
ocean→	water		81	
ocean→		natural	80	
ice→	water		77	36
ice→		natural	80	36
river→	water		88	82
river→		natural	87	82
road→		manmade	38	

Missed (omitted) correct rules
SOARD (confidence)
road → manmade (38% confidence)
ARD
ice →natural/water road→manmade

Incorrect (wrong rules)		
SORAD		% Confidence
beach→	park	73
ARD		
river→	ocean	82

Data: Train : Boston image strips 1-4
Test : image strip 4

Self-supervised ARTMAP

Carpenter addressed and successfully adapted self-supevised-learning-ARTMAP-NN-model (Alg.10), which has unique positive features (Chart 13). It integrates information/knowledge from supervised (i.e. labeled patterns with some features: teacher), unsupervised (unlabled patterns with more features) and internal model activated bits(self-labeled patterns).

Alg. 10: Self-supervised ARTMAP

Phase 1: Supervised learning with toy dataset
Phase 2: Unsupevised learning with real world task

If	Self-supervised ARTMAP & Stage 2 [tr and tet]
Then	Default(Contrast parameter) = 2.0
If	Self-supervised ARTMAP & Stage 1 [supervised]
Then	Default(Contrast parameter) >large

Chart 13: Self-supervised-ARTMAP

- System trained on labeled data with limited features continues to learn on an expanded but unlabeled feature set
- Learns about novel features from unlabeled patterns
- Accuracy may improve dramatically compared to that of the initial trained system
- Avoiding performance deterioration from unlabeled data
- Does not destroy (partial-) knowledge previously acquired from labeled patterns
- Improves test accuracy
- Slow distributed learning on unlabeled patterns focuses
  - On novel features
  - Confident predictions,
  - Defines classification
  - boundaries for ambiguous objects in the labeled patterns

Extended_Self-supervised-ARTMAP

Repeat

Self-supervised-ARTMAP

- Stage I
  - winner-take-all fast learning → WBU ~= WTD
- Stage II : F2 code y is distributed →
  - WBU & WTD diverge

End repeat

Information fusion-ARTMAP

Carpenter et al. [69] considered a task wherein information from sensors and experts is reliable. Image fusion concerns with synergistic combination of information provided/ generated from various sensors or by the same single sensor in many measuring contexts. But, the category of an object based on evidence renders earlier information inconsistent. Further, inherent relationships among classes although known, they are not used either by human user or even by the automated system. A new information- fusion-ARTMAP-system using distributed code representation is reported. ARTMAP has the inherent capability of learning one-to-many relations. This feature results a self-organizing expert system leading to produce (or discover) hierarchical knowledge structures. The results of Boston and Monterey image analysis are briefed in table 12.

Table 12: Rule generation using information-fusion-ARTMAP

Boston-map	
Rules generated by information-fusion-ARTMAP from	
High confidence rules x → y	C(%)
Beach →open space	88
Beach → natural	87
Ice → water	86
Ice → natural	94
Industrial → built-up	90
Industrial → man-made	96.

Three marginal equivalence relations
Park → open space (86%)/ open space → park (85%)
residential → built-up (82%)/ built-up → residential (78%)
water → natural (99%)/ natural → water (68%)

KB for information fusion	
If	Independent sensors label a vehicle car or truck or airplane & one of the labels are correct
Then	Weigh confidence/reliability of each source & merging complementary information & Gathering more data if conflict still exists

Open space → man-made (38%)	
Park → manmade (36%)	

Monterey image	
Grass → tree (42%)	car → roof (34%)
vehicle0 → road (27%)/	
Two marginal equivalence relations	
Road → pavement (91%) / pavement → Road (83%)	
Tree → natural (98%) / natural → Tree (83%)	

Deriving knowledge hierarchy from a trained network [67] : ARTMAP-fusion-system assigns each input an arbitrary number of output classes in a supervised learning mode. The distributed predictions of a trained ARTMAP network generate a hierarchy of output class relationships.

This approach is made use in inferring patterns of drug resistance, creating a hypothetical set of relationships among protease inhibitors from resistance patterns of genome sequences of HIV patients and improving marketing suggestions to individual consumers. The apparent contradictions in input pixel labels of Monterey and Boston images are resolved by assigning output classes to different levels in knowledge hierarchy chart.

4.3 Fuzzy-ART-MAP (FAM)

In 1992, Carpenter and Grossberg [143] of MIT, USA proposed fuzzy-ARTMAP for supervised learning of paired floating point or binary data. The objective, finding cojoint(max(prediction generalization) & (min(prediction error)) is same as that in ARTMAP. FAM maps an analog input of any dimension to an analog output space. There is a striking similarity between Binary FAM with binary inputs/outputs and ARTMAP. It is one of the celebrated NN-models with a diverse philosophy from feed forward/recurrent NNs. This model and its clones have a niche in the arsenal of natural computational tools. It is incorporated in only in a few sought after commercial software packages (Professional II).

Architecture.Fuzzy_ARTMAP: The architecture is same as that of ARTMAP except that ART-1 is replaced by fuzzy-ART module. The subtlety is employing fuzzy operations in place of classical set theory computations (used in ARTMAP). Another way of looking at it is, that fuzzy-ARTx and fuzzy-ARTy are linked by a mapfield. The learning in FAM, Ellipsoidal ARTMAP (EMAP) or Gaussian ARTMAP (Gauss-ARTMAP) [126] is semi-supervised [92,93, 126] due to the presence of ART-1 architecture. The categories correspond to the hidden units. In fact, each neuron is mapped to a specific class. All active units in F2 contribute to output through inter-ART weights reflected in defuzzification function. Fuzzy-ARTMAP [97] requires a priori knowledge of choice and vigilance parameters. The training and testing phases completes the learning and probing into unknown/unseen data set prediction. A simple heuristic is setting vigilance parameters to zero values, but the flavor of the philosophy of ARTMAP is lost. A tiny advantage claimed is easy comparison of performance with MLP and other similar networks, which have compact representation of data.

If $\rho_{xy} = 1$

Then each neuron in $F2^x$ is linked to just one neuron in $F2^y$

Applications.FAM

The data analysis which was obvious with tabular form, plotting as X- Y graphs or from look up tables in eighteen century, underwent renaissance in 1900s and now consumes large CPU time with state-of-the-art theoretical/ empirical /nature-inspired tools. The challenge ahead is intelligent computing, may be, adaptive-/self-learning-/ auto-correcting and Auto-discovery paradigm. The tit-bits of omnimetrics (183-190) vocabulary are briefed in Appendix-3.

Dietetometrics

Fuzzy-ARTMAP model was used for early detection of growth of fungus in bakery products. The seven fungal speices isolated from bakery products were analysed with mass spectrometry using the two sampling head space techniques viz. static headspace and solid phase microextraction. The SPME−MS-based e-nose is better than static headspace and fuzzy-ARTMAP (Table 13). It had lower predictive errors compared to principal component analysis (PCA), coupled to discriminant function analysis (DFA).

Table 13: Fuzzy-ARTMAP in prediction

Hours of inoculation	Prediction of fungal growth
24	98
24	78
96	88

Environmental waste models

A rapid sorting of post-consumer plastic waste was modelled (8) with FuzzyARTMAP after preprocessing IR data by scaling and features selection. The results are superiror to MLP-NN and PLS. The network architecture weights have spectroscopic interpretability.

Structure property relationships (SPropR): A few typical studies using quantum chemical properties, molecular descriptors for solubility, partition coefficient (61), boiling point are briefed in table 14. Here, every sub-goal is higher order and becomes exemplary for physic-chemical research.

Table 14: ARTMAP in Structure property relationships

ARTMAP >MLPBP >MLR	
Molecular descriptors (PM3-SEMO-QC)	
1) molecular polarizability 2) dipole moments 3) total point charge 4) total hybridization 5) total sum 6) ionization potential 7) heat of formation	Henry's Law constant (495 org compounds) NN | Tr | Te Fuzzy ARTMAP | **0.03** | **0.28** 7-17-1 MLPBP | 0.13 | 0.27
log*K*ow, 442 compounds 1) average polarizability	NN | Tr | Te

<table>
<tr>
<td>
2) dipole moments
3) exchange energy
4) total electrostatic interaction energy
5) total two-center energy
6) ionization potential
……. …..
</td>
<td>
<table>
<tr><td>Fuzzy ARTMAP</td><td>0.03</td><td>0.14</td></tr>
<tr><td>12-11-1 BP</td><td>0.23</td><td>0.27</td></tr>
</table>
</td>
</tr>
<tr>
<td>
▶ topological descriptors
1) first-
2) second-
3) third-
4) fourth-order molecular connectivity indices
▶ CQC descriptors
5) average polarizability
6) dipole moment
7) resonance energy
8) exchange energy
9) electron−nuclear attraction energy
10) nuclear−nuclear (core−core) repulsion energy
</td>
<td>
<table>
<tr><td colspan="3">Solubility
(515 organic compds)</td></tr>
<tr><td>NN</td><td>Tr</td><td>Te</td></tr>
<tr><td>Fuzzy ARTMAP</td><td>0.02</td><td>0.14</td></tr>
<tr><td>11−13−1 MLPBP</td><td>0.29</td><td>0.28</td></tr>
</table>
</td>
</tr>
<tr>
<td>
boiling points, aliphatic hydrocarbons
▶ four valance molecular connectivity indices
1) 1χv
2) 2χv
3) 3χv
4) 4χv
5) second-order Kappa shape index (2κ)
6) molecular weight
▶ CQC
7) dipole moment
</td>
<td>
<table>
<tr><td colspan="3">Boiling point</td></tr>
<tr><td></td><td colspan="2">^{o}K</td></tr>
<tr><td>NN</td><td>Tr</td><td>Te</td></tr>
<tr><td>Fuzzy ARTMAP</td><td>0.81</td><td>1.30</td></tr>
<tr><td>7−4−1 MLPBP</td><td>1.65</td><td>1.73</td></tr>
</table>
</td>
</tr>
<tr>
<td>
1168
sum of atomic numbers; five valence connectivity indices; and the second-order kappa shape index, without or with the dipole momen
</td>
<td>
<table>
<tr><td>#Comp</td><td>property</td><td colspan="2">Prediction error fuzzyARTMAP</td></tr>
<tr><td>1168</td><td>BP</td><td>2</td><td>^{o}K</td></tr>
<tr><td>530</td><td>Critical T</td><td>1.4</td><td>^{o}K</td></tr>
<tr><td>463</td><td>Critical P</td><td>0.02</td><td>MPa</td></tr>
</table>
</td>
</tr>
</table>

Structure toxicity relationships (SToxR): Here, the results of toxicity in fish (piscimetrics) and rats (biometrics) are cited. Espinosa et al. (61) developed fuzzy ARTMAP models to predict acute toxicity (LC50) of benzene derivatives for fathead minnow and the oral toxicity (LD50) organic compounds for rats. SOM is employed to probe in to similarities of chemical compounds and select relevant and informative molecular descriptors.

Medicinometrics

Diagnostics.Carcinoma

Over 25,000 malignancy cases of mature B lymphocytes of which around 40% are non-Hodgkin's lymphoma category are reported every year. Yet, current taxonomy of cancer still lumps together molecularly different diseases with distinct clinical phenotypes. The variable presence of chromosomal translocations, deletions of tumor suppressor genes and numerical chromosomal abnormalities are responsible for molecular heterogeneity within individual cancer diagnostic categories. In 1832, Thomas Hodgkin recognized human lymphomas. Since then, classifications are based on both morphologic and molecular parameters. Revised European-American Lymphoma (REAL) classification is the recent taxonomical classification based on distinct clinical-pathological entities. It is chemical eye at molecular level for carcinoma [191-207].

Informatrics

The noisy as well as noise free patterns in monochrome satellite images and color patterns were recognized with ART1 along with moment-based feature extractor. The invariant values to rotation, translation and scaling change are converted to analog values and inputted to ART1.

Classification_without a priori knowledge: FAM is designed for multi-class tasks with no a priori knowledge of distribution of patterns and even number of classes.

Function approximation_FAM: FAM is used for function approximation for data sets without/low noise. However, a large amount of noise/outliers lead to over learning problem, thus diminishing generalization. The prob-ARTMAP was introduced for function approximation in noisy environment. The typical applications of fuzzy-ATMAP-NN are in automatic target recognition based on radar range profile, 3D-object understanding and prediction from the series of 2D-views, recognition of speaker independent oval/ online hand written alphabets / ECG signal and diagnosis of genetic abnormality/ breast cancer and heart disease.

4.3.1 Advantages and limitations of FAM: The limitations/remedial measures reported and positive features resulting in advantages of ARTMAP and FAM are shortlisted in chart 14.

Chart 14a: Advantages_FAM

+ Learns off-line as well as on-line data
+ Learns new data without forgetting the past (i.e. Circumvents plasticity-stability dilemma)
+ Excels for tasks involving fast incremental learning in non-stationary environment
+ Fast convergence
+ Requires less memory as it uses a compressed representation of the data
+ High stability unlike Clustering algorithms with Euclidean distance
+ Detects data with novelty
+ Easy explanation of output

Chart 14b : Limitations & remedial measures_FAM

− Sensitive to noise
  - Remedy : mu-ARTMAP, Boosted- ARTMAP
    - Robust to noise, as they allow small error on training data

− Category proliferation
  - Remedy:
    - Slow learning
    - User chosen priori maximum number of categories
    - Incremental learning
    - On-line training
    - D-ARTMAP
    - Prob-ART
    - GA-FAM which performs simultaneous evolution of topology and refinement of W

− Performance is affected by order presentation training samples
  - Remedy
    - Safe-mu-ARTMAP, Smaller size FAM (ss-FAM), ss- Ellipsoidal ARTMAP (ss-EAM),
    - Genetic Algorithm-FAM (GA-FAM), ss-GA-FAM
    - Ordered-fuzzy-ARTMAP
    - Presentation to maximize the performance measure
    - Maximum clustering algorithm to select order of patterns
      - Training FAM with different orders of presentation Large CPU time

- Compliment coding leads to category proliferation
- Over learning in presence of noise
- Fuzzy categories of hyper rectangular form
  - Remedy : hyper spheres handles noisy patterns
- Sensitivity to statistical overlapping of the two classes
- Large CPU time and storage
- Poor generalization in presence of noise i.e. Erroneous prediction
  - Remedy : GA- ARTMAP
- Slow convergence
  - Remedy: Choice par =0 ; vigilance par = 0
- Depends upon tuning (Choice and vigilance) parameters
  - Remedy: Choice par =0 ; vigilance par = 0
    - Philosophy of ARTMAP is lost

4.3.3 Match tracking anomaly: If ART-x input matches its category prototype perfectly, but the network results in a wrong prediction, it is an anamolous situataion. The word match tracking anomaly (MTA) was coined for the incident during FAM training. According to Carpenter [144, 147-148], there is no chance for MTA for binary input patterns with same number of ones (1s). Later, Bartfai [127] also reported that it will never arise if each input (with same number of 1s) is presented to FAM before the corresponding non-contradictory target is presented. Raising the vigilance parameter above the matching level no doubt prevents the network to find another ART-x category. It is desired to limit the category proliferation, but there is no corrective action for the wrong prediction. The proofs of convergence of learning, pattern diversity, N-N-N conjecture, pattern clustering are reported for ART-1. However, these properties are not directly extendable to FAM because of increase of vigilance parameter

4.3.4 Comparison of performance of FAM with other paradigms: Heinke and Hamker [118] compared performance of Fuzzy ARTMAP, Growing Neural Gas (GNG), Growing Cell Structures (GCS) and the multi-layer perceptron (MLP) on a number of benchmark datasets. In the case of FAM, size and time of computation are small, but, inferior to GNG, GCS and MLP.

4.4 Evolution of fuzzy ARTMAP

ARTMAP-Instance counting (ARTMAP-IC)

Carpenter and Markuzon [117] introduced ARTMAP-IC in 1998. It is an extension of ARTMAP for inconsistent cases in medical diagnosis.

Architecture.ARTMAP-IC: To the fuzzy-ARTMAP skeleton, distributed prediction and category instance counting systems are added. The popular match tracking (MT^+) algorithm is replaced by negative match tracking procedure (MT^-). This new algorithm predicts even with inconsistent (i.e. cases with different outcomes but with identical inputs) or sparse data (fig. 10). In the testing phase distributed category representation, gives rise to probabilistic prediction. The results of training the data with different orders of presentations are analysed by voting strategy

If	Q = 1
Then	ARTMAP-IC reduces to ARTMAP
If	Q = NP
Then	entire trained system is used in the prediction

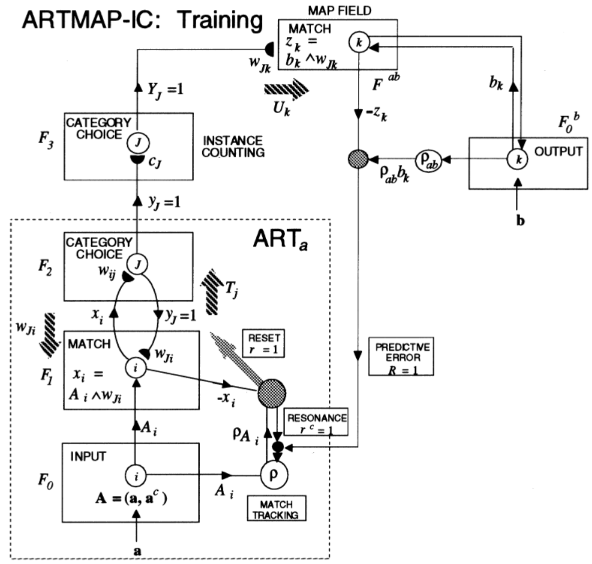


Fig. 10: Architecture of ARTMAP –IC
(courtesy from Ref. 117)

which improves the prediction.

Dataset.Medical.ARTMAP-IC: The medical data sets tested with ARTMAP-IC are concerned with Pima Indian diabetes, Wisconsin breast cancer, heart disease and gall bladder removal (cholecystectomy). The performance of ARTMAP-IC is superior to that of logistic regression, k-nearest-neighbors, perceptron-like ADAP, Multi-surface pattern separation, unsupervised CLASSIT, instance-based (IBL)- IB1,IB2 and IB3- C4 and ART-EMAP.

Table 15: Comparision of ARTMAP-IC results for medical data sets with Neural and statistical models

Model	Heart disease	PIMA-Indian diabetes	Breast cancer
Logistic Reg	79	77	97
k-NN	67	77	96
ARTMAP	74	66	96
ARTEMAP	76	76	97
ARTMAP-IC	78	79	96

Gaussian- ARTMAP

It is an on-line constructive clustering procedure. Baraldi (86) proved that Gaussian-ARTMAP also belongs to symmetric-ART (Sym.ART) class of networks and can also be implemented with Extended-ART-1 (Ext-ART-1) algorithm. The basis of Gaussian-ARTMAP is maximum likelihood probability distribution function (pdf) estimators for mixtures of Gaussian functions with noteworthy features (Chart 15).

Chart 15: Features of Gaussian- ARTMAP

- + Learns internal representation more efficiently
- + Less category proliferation
  - + Results in reduced number of categories
- + More resistant to noise compared to fuzzy-ARTMAP
  - + Performs better than ARTMAP for noisy data

– Geometric interpretation changes

– Extraction of If-then rules is not possible

Architecture.Gaussian-ARTMAP: The architecture is same as that of ARTMAP. The match function is discriminant function of a Gaussian classifier. Initially, memory consists of neurons with unassigned weights. During training the weights are assigned to the relevant neurons. Each category in Gaussian-ARTMAP is an hyper-ellipse defined by Gaussian distribution function. It is usual that in ARTMAP and fuzzy ARTMAP NNs, a category is interpreted as hyper-rectangle. The soft-competitive version gained popularity, in spite of the fact initially it was introduced in hard-competitive form. The performance of Gaussian-ARTMAP is far better than Fuzzy-ARTMAP and EM-of-mixture modeling.

Hyper-sphere-(or RBF-)-FAM or RBF-FAM

Su, Lee and Hsich (64) proposed Hyper-sphere-FAM in 2006.

Architecture-Hyper-sphere-FAM: Fuzzy-ARTx is replaced by RBF and the training method by successive approximations. But, fuzzy-ARTy and MAP-field remain the same. It replaces MT with negative MT and employs voting strategy. Hyper-sphere is used as a TF in the competitive layer of RBF-FAM. It is a new basis function, which overcomes the problem for constant response value (Fig. 11). During training, only the center vectors of the Gaussian basis functions are tuned but not the standard deviations. In the prediction phase, the output is the weights connecting the winning neuron of RBF module to the outputs of MAPfield. It is easy to derive the deterministic updating rule to expand a hyper-rectangle or hyper-sphere to include a new data pattern (Fig. 12 and Fig. 13). Many ways of expansion have been reported, but any choice depends upon the complexity. $\theta_{recruited}$ indicates the number of neurons responsible for the value of output. For instance, $\theta_{recruited} = 0$ shows that all neurons contribute, while a large value of $\theta_{recruited}$, indicate a small number responsible for the output. It is a user chosen parameter reflecting approximation accuracy. θ_d, is a measure of similarity threshold. For, k-class pattern recognition, it is set to unity. In case of function approximation, it gives an idea of ESS or asymptotic accuracy of approximation. NNs in off-line learning mode utilize the gradient information completely in optimizing the values of parameters. On the other hand, incremental learning does not use gradient information. Alpha is used in the denominator to break the ties. It selects the hyper-sphere with a smallest γ_j, when the data falls inside the several hyper spheres simultaneously. Chart 18 depicts the highlights and shortcomings of *Hyper-sphere-FAM.*

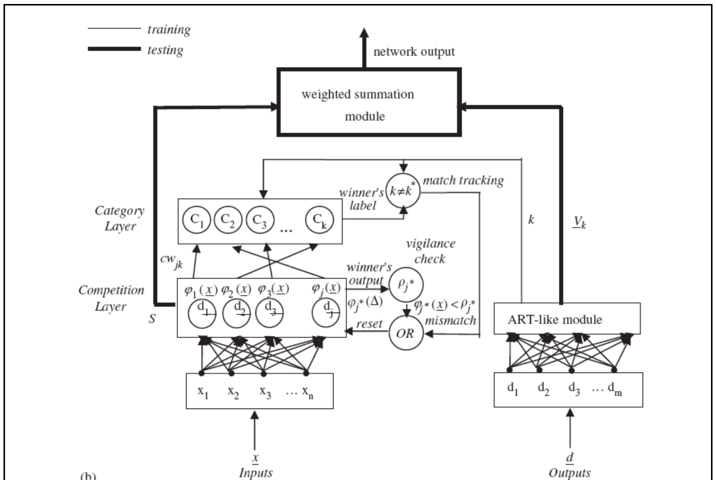


Fig. 11: Architectures (a) Fuzzy ARTMAP (b) HS-ARTMAP (courtesy Ref: 64)

Fig. 12. comparisons of ease of expansion of hyper rectangle, hyper-sphere, and hyper-ellipse

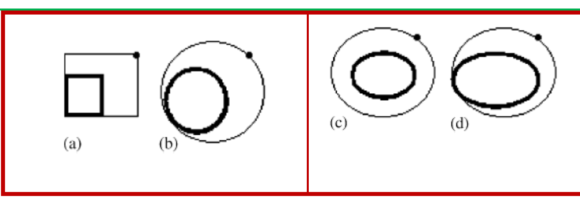
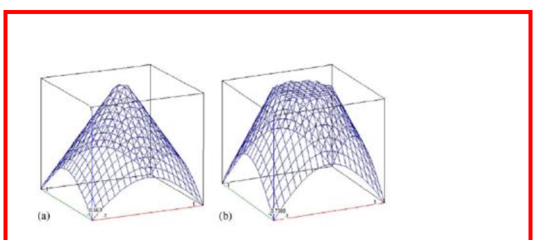
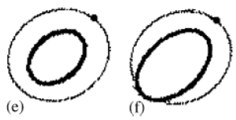


Fig. 13. Basis function: (a) $r_j = 0.1.$ (b) $r_j = 0.5.$





(e) (f)

Dataset.simul.Hyper-sphere-ARTMAP: A three class, 2D-dataset consisting of 579 points was analyzed with Hyper-sphere-ARTMAP, where vigilance parameter was varied from 0.7 to 0.0. The training (70% of data) resulted in a classification which was 100% accurate after several epochs. The test data classification accuracy was also close to 100%.

Chart 16: characteristics of MISO dataset

	Range	NP
Training	X1 : [-2 to 2] X2 [-2 to 2]	441
testing	X1 : [-1.9 to -1.9] X2 = [-1.9 to -1.9]	400

Dataset.IRIS. Hyper-sphere-ARTMAP: IRIS data was analyzed varying the number of training samples (75 and 150). The training accuracy was 100%, while the testing varied between 93.6 to 95.5%.

Dataset.Simul.MISO-Hyper-sphere-ARTMAP: The Hyper-sphere-ARTMAP approximates MISO function data (Chart 16) with $\theta_d = 0.0$. The other models with $\theta_d = 0.5$ or 0.1 are inadequate.

Dataset.Simul.MIMO-Hyper-sphere-ARTMAP: The MIMO function used by Srinivasa (123) is employed as test function. The outputs, y1 and y2 are perturbed by 40% of the simulated output (chart 17) with Gaussian noise. The results of Hyper-sphere-ARTMAP are compared with modified-PROB-ARTMAP.

Chart 17: characteristics of MIMO dataset

U1(t) = sin (2*pi*t)/250

U2(t) = cos(2*pi*t)/250 0<=t <=800

	Time steps	NP
Training	0 to 800	100 random
testing	0 to 800	t= [0.5,1.0, …799.5]

Dataset.BUPA.Hyper-sphere-ARTMAP: The results of Hyper-sphere-ARTMAP are compared with S-FAM. It does not degrade greatly when new information is learnt. It is superior compared to SFAM, which was unable to retain fully the previously learnt knowledge.

Physical significance of cluster learning

The neurons compete with each other using the maximum output criteria, to be a winner one.

Chart 18: **Advantages and Limitations of** Hyper-sphere-FAM

- Hyper-sphere-FAM models successfully even if the response output is constant over a certain region.
    - The inherent limitation RBF-NN to model Gaussian profile with flat maximum is overcome. The other alternative is use of raised cosine RBF as TF. MLP with sigmoid TF fails to model
- It integrates advantages of RBF as well as FAM
- Robust on-line learning
- Less sensitive for order of presentation
- Incremental learning capability
    - Retains most of the information already learnt, even during incremental training
- Performs well for function approximation in presence of high noise
- No need of complimentary coding
- Enhances predictive capability in medical diagnostic systems.
- Low number of parameters
    - If hyper rectangle, then 2*n

- If hyper ellipsoid, then (2*n +1)
- If hyper ellipsoid and principal axes are parallel to co-ordinate axes, then [n + n * (n+1)]/2
- If hyper-sphere, then n+1

— The whole of information already learnt is not retained during incremental training i.e. part of it is lost
— It cannot achieve correct performance after one training epoch

Bayesian-fuzzy-ARTMAP

In 2007, Vigdor and Lerner [59] proposed Bayesian-fuzzy-ARTMAP-NN (popularly known as Bayesian-ARTMAP) in which hyper-rectangular basis function (for activation) is replaced by Gaussian categories. It is useful in medical diagnosis of rare diseases. Bayesian framework is introduced into Fuzzy-ARTMAP to increase classification accuracy of the model as well as diminishing the menace of category proliferation. The volumes of selected category are a variable, resulting in shrinking or growth of categories. The categories are represented by multidimensional Gaussian distributions. Bayes' decision theory is employed to associate patterns with categories probabilistically (Alg. 11) for the purpose of ART learning. It is followed by association of categories to classes in executing ART map learning, using theory of probability again. The prediction of a class is done with all the categories associated with it. Bayesian-ART estimates class and category posterior probabilities in ARTMAP and ART stages repeatedly. Still, it is an open problem whether voting strategy improves generalization of this NN.

Architecture.Bayesian-ARTMAP: The architecture is same as FAM. The deterministic fuzzy rules for matching function and learning rule are replaced with statistical learning and inference. Category choice function in FAM is based on fuzzy set theory, while that in Baysian- ARTMAP is based on Bayes' set theory. The match function is of limiting size in FAM, where as it is volume in Baysian-ARTMAP. In FAM, WTA principle operates, while concept of many categories in Bayesian-ARTMAP. Complimentary coding is not required in Bayesian-ARTMAP and typical positive features are in Chart 19.

Bayes decision theory accounts not only for the distance of category to a pattern, but also the dominance of category over others quantified as category a priori probability. Volume is more appropriate compared to sum of sides in discriminating clusters in high dimensional space. Each Gaussian category is defined by mean, covariance and prior probability. It is superior to the weight vector of fuzzy-ART hyper rectangular category. Gaussian category of Bayesian-ARTMAP is clearly identified by its center of mass, shape of distribution and dominance with respect to other clusters. In FAM, the category is denoted by weight vector consisting of category of extreme corners. The distribution of data within the category of FAM is completely unknown.

Alg 11: Bayesian-ARTMAP

Step : 1 **Category Choice**
All existing categories compete in representing an input pattern
Category_winning = arg max (posterior probability).

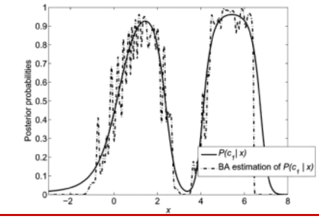
Step : 2 **Category matching (vigilance test)**
Hyper volume is defined as the determinant of Gaussian co-variance matrix.
vigilance test for the category hyper volume
Hyper_volume (j) = det(cov(.))
Hyper_volume_max : maximum of hypervolume allowed for category

If covariance matrix is diagonal

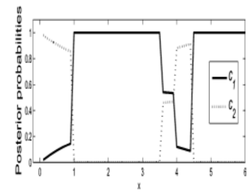
Then $$\text{hyper\_volume} = \det(sd)j = \prod_{d=1}^{\dim} \left(sd\right)_j^2$$

Posterior class probability (---) : Bayesian-Fuzzy-ARTMAP ;___ : data [courtesy Ref : 59]

If	Bayesian-fuzzy-ARTMAP & Full covariance matrix is used
Then	Each class is modeled more accurately compared to the model using only variances Example: Model (cov[0.8 0.5;0.5,0.9]) is superior to model(var [0.8 ;0.9])
If	If cov(i,j) = 0, $i \neq j$
Then	Results of the models with cov or only variances are same Example : Model (cov[0.8 0.0;0.0,0.9]) is equal to model(var [0.8 ;0.9])
If	A category has small number of patterns
Then	Calculation of full covariance is poor Classification accuracy is undermined



Posterior probabilities of classes



Dataset. NonGaussian-simul.Bayesian-fuzzy-ARTMAP: A two class problem with non-Gaussian clusters is simulated. The first cluster comprises of a mixture of uniform and Rayleigh distributions. The second class is a mixture of two uniform distributions.

Dataset.Gaussian- simul.Bayesian-fuzzy-ARTMAP: Two thousand points comprising two classes in a single dimension are generated by a mixture of two Gaussian components.

Dataset.realWorld-Bayesian-fuzzy-ARTMAP: Bayesian ARTMAP outperformed FAM in modeling synthetic and real data sets (Table 16). The performance was judged in terms of classification accuracy, learning curves, extent of overlap of the classes, number of categories and expected loss. The category proliferation is insignificant (Table 17) compared to Gaussian-ARTMAP and FAM. The results are averaged values for all real world databases.

Table 16: Comparison of performance of BA-FAM with FAM

	Liver	PIMA	Breast	Down syndrome	IRIS	Glass	Image	Wine	Hayes
BA-FAM	**61.2**	**72.8**	**95.8**	**87.1**	**95.7**	**64.7**	**84.6**	**93.6**	**71.8**
FAM	58.6	66.9	95.2	85.8	90.3	45.6	77.9	92.6	

	Mushroom	Ionosphere	Car
BA-FAM	94.3	87.4	89.6
FAM	**99.4**	**89.2**	**95.7**

Table 17: comparison of category proliferation of Bayesian-FAM with other-FAMs

NN	% accuracy (test data set)	Category proliferation
Bayesian-FAM	82.9	0.004
Gaussian-FAM	81.0	0.02
FAM	76.7	0.12

Chart 19: Bayesian-Fuzzy-ARTMAP

- It overcomes category proliferation
- Less sensitive to the order of data presentation compared to fuzzy-ARTMAP
- Preserves the advantages of FAM
- Bayesian- Fuzzy-ARTMAP is an alternate to fuzzy-ART

Simplified FAM and its modifications [68]

SFAM 1.0 and SFAM 1.1 are applicable for classification tasks only. It has low architectural redundancy and reduces overheads in computation. Yet, there is no loss of predictive capabilit. Some of the unnecessary computations in SFAM 1 are eliminated resulting in SFAM 2.0 and SFAM 2.1. There are off-line and on-line versions of SFAM-2.0 with and without match tracking. In probabilistic simplified FAM, Bayesian classification is introduced into SFAM. In fast simplified FAM, partitioning of data set speeds up to 'p' times on sequential machine and 'p^2 ' on a parallel computer. It (FS-FAM) is slow for large data bases. Further improvement is through combination of partitioning of both data and network, which reduce redundancies and speeds up convergence. Castro et al. (57, 68) trained subsets of training data and produced cumulatively a combined FAM called Hilbert-fast-simplified-FAM. It reduces the CPU time or increases the convergence speed. The generalization is definitely equivalent to Fast-SFAM and better in some tasks.

Parallel-FAM

The parallel version of the algorithm for online SFAM 2.0 [57] without match tracking is referred as parallel-FAM. Castro [57, 68] reported coarse grain parallelization based on pipe-line approach. This speeds up the training process. The FAM without match tracking version is rendered suitable for parallel computation. Theoretical proofs are derived for parallel-FAM without match tracking.

Architecture.Parallel-FAM: The communications between F1 and F2 layers are divided into two bidirectional rings. The increase in number of processors speeds up the computation close to linearity. Learning examples are pipelined through the ring. It optimizes the network utilization. The elimination of all computations in ART-y and simplification of ARTxy modules increases the speed of computation. 96 AMD nodes (Dual Athlon MP 1500+) each with 512MB RAM are connected through a fast ether-network. It has high latency i.e. the optimum performance is obtained when the number of communication packets is large in size but small in number. Parallel implementation on this platform (Beowlf parallel) is different from shade memory or vector machines. Hilbert space filling curve method is used to partition the data for analysis by parallel-FAM.

Dataset.simul.Parallel-FAM: A two class and 16-dimensional Gaussian data sets of size 532,000 patterns were generated. All, but the discriminating dimensions are identically distributed with the same mean and variance. The discriminating dimension has offset means such that the overlapping between Gaussian curves is set at 5%. For another data set it is fixed at 15% (Table 18).

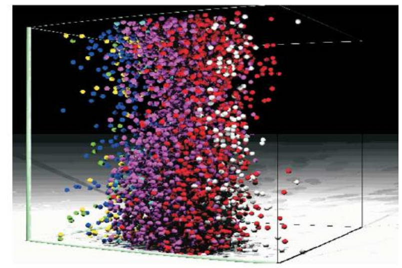
Table 18: Effect number of patterns on accuracy of classification

Patterns in thousands	5% overlap		15% overlap	
	Classification accuracy	Average number of templates created	Classification accuracy	Average number of templates created
32	92.50	7,032.83	79.25	10,608.83
128	92.91	25,740.5	80.10	40,319
512	93.21	92,365.66	80.54	152,827.91

Table 19: Effect number of patterns on Forest cover data classification

Fig. 14. A subset (5000) of Forest covertype data with Classes in different colours

32	70.29	5,148.83
128	75.05	22,831
512	79.28	100,720.75



Dataset. Forest cover.Parallel-FAM:: It contains 581,012 patterns and 54 variables. Each pattern belongs to one of seven forest tree cover types. Blackard donated this database to UCI. The first three dimensions of the database is depicted in the Fig. 14.

prob-FAM

The difficulties faced in applying FAM under noisy conditions motivated [123] to promote implementation of probabilistic concepts into FAM. The result is probabilistic-fuzzy-ARTMAP (or in short prob-FAM) and later developments (Chart 20).

*Architecture.*Prob-FAM: Prob-FAM contains two-Fuzzy-ART modules linked by a map-field. It accepts complementary coded input. Here, the similarity ends with FAM. The map-field has a different dynamics. It uses probability information combined from various neurons. The probability relations are stored in Inter-ARTMAP. The subtle difference between FAM and prob-FAM is the different inter-module linkage dynamics mediated through the map field. The probabilistic information about interlayer node association is based on it's the frequency, which generates expected value of predicted output. Srinivasa [123] proposed a modification of prob-FAM to surmount the difficulties in noisy environment (prob-FAM-Srinivasa). Typical limitations, remedial measures along with advantages are described in non-elaborated format in chart 21.

Chart 21: Advantages and limitations of Prob-FAM- Srinivasa

- + Reduces proliferation (or ART-x) of categories form noisy data for
  - function approximation
  - Classification
- + Filters noise in training data
- + Reduction in RMSE
- + Increased generalization

- — In the testing phase, the inputs which are not within the range of any relevant ART-x category are rejected.
  - *Remedy:* Delete ART-x vigilance & WTA during testing

Chart 20: Hybrid-FAM with probabilistic implications

Fuzzy + ARTMAP	FAM + probability
FAM →	prob-FAM → Boosted-FAM
	mu- FAM → safe-mu- FAM

Equivalence of prob-FAM with FAM

If noise free data &
Training performed for optimal learning
Then Results of FAM and prob-FAM are similar
& One epoch is sufficient

Table 20: Comparison of number of categories in FAM and prob-FAM

Module	Noisy		Noise free	
	FAM	Prob-FAM	FAM	Prob-FAM
ART-x	806	**112**	312	**110**
ART-y	61	**61**	53	**53**

Dataset.simul.FunctionAprrox.prob-FAM: 1000 SISO data points are simulated using the functional relation. Gaussian noise N(0,1.0), multiplied by 0.02 is added, producing noisy data Eqn. 01. Prob-FAM resulted in less number of categories, even for noisy data. The vigilance parameter is varied ([0.2 to 0.9], 0.999) for training data set of size 1000, 5000 and 10,000. Prob-FAM resulted in less number of categories of explanatory matrix (X) compared to conventional FAM (table 20)

Eqn. 01: Equation for simulation of periodic profile with Gaussian noise
$ysim \; = \dfrac{\sin(10*x)+\sin(20*x)+\sin(30*x)+\sin(40*x)+\sin(50*x)+\sin(60*x)+\sin(70*x)+10}{20}$
$ynoisy = ysim + \left[N(0,1.0)\right]*0.02$

Dataset.simul.Classification.prob-FAM: One thousand uniform random floating point values in the interval [0 to 1] are simulated. The first category is in the ranges -- [0, 0.3], [0.7, 1.0], [0.45, 0.55] --, while second one is around [0.3, 0.7]. FAM is successful between vigilance parameter values of 0.2 to 0.5. The error rate for FAM and prob-FAM are same at a value of vigilance parameter equal to 0.99.

Fuzzy adaptive system (Fas)-ARTMAP

Izquierdo (97) introduced Fuzzy-ARTMAP (popularly referred to as Fas-ART) and fas-BACK-NNs to surmount the shortcomings of fuzzy ARTMAP and Prob-ARTMAP. They are equivalent at test phase. These neuro-fuzzy models are for function approximation with noisy input patterns. Fas-FAM is a minimum-distance-to-means clustering network (chart 22). It uses fuzzy-mf as an activation function. The theoretical literature on fuzzy systems proves Fas-FAM has a universal function approximation property. A distributed version of Fas-FAM called Distributed Fas-FAM was also proposed.

Architecture.Fas-FAM: The general structure of FAM is retained in Fas-FAM. By introducing mfs for features, fuzzy nature is incorporated. The inter-art-map stores relations between $F2^x$ and $F2^y$ units. The advantage is that equivalence between activation function (ACF) of a unit and mf to fuzzy set is defined by the equation $AMF = \pi\, amf\,(I_i)$. The training is same as that of fuzzy-ARTMAP. The activations of neurons are calculated and that with maximum activation is selected.

In fas-BACK, the learning is guided by minimization of prediction error and learning is through matching. A few more additional learning rules are added. It reduces the network complexity and yet the same time maintains a similar performance index as that in Fas-FAM. Dynamic Fas-ART, FEMA, pipe lining FAM without match tracking are hybrid systems with notable characteristics.

Chart 22: Features of Fas-ARTMAP

- + Successful for noisy data unlike Fuzzy-ARTMAP
  - + Less network complexity
  - + Low prediction error
- + Suitable for incremental learning
- + Superior to Prob-ARTMAP
- + Inherited advantages/ theoretical merits of FIS
- + Dual membership/activation function
- + Defuzzyfication produces numeric outputs
- — large increase in CPU time

Fas-ARTMAP_ Cano

A supervised NN based on FAM was introduced with Boolean (ARTMAP) or Fuzzy (Fuzzy-ARTMAP) logic operators in the activation function of neurons. These operators are replaced in Fas-ARTMAP by a membership function. Each category is associated with a fuzzy set and triangular/ Gaussian/bell-shaped MFs are used. The activation function for each neuron in F2-x is equal to the product of fuzzy MF degrees on each input feature.

Distributed Fas–ARTMAP-Carpenter

Carpenter (1998) proposed distributed – ARTMAP including distributed code in learning as well as test phases of pattern recognition (PR). It has the same architecture of Fuzzy-ARTMAP.

Distributed Fas–ARTMAP-Parrado-Hernandez_2003

Parrado introduced distributed learning Fas-ARTMAP resulting in a hybrid distributed Fas-ARTMAP (Chart 23). More than one category is activated to represent the input pattern unlike in FAM where only one category is committed through WTA mechanism. Instance counting before prediction of output class, switching to WTA when the prediction is incorrect and assignment of credit before attaining resonance state have been included. The learning rate of each neuron is controlled by its study state activation.

Chart 23: Positive facets of distributed -Fas-ARTMAP

- Patterns closer to the learnt categories will be learned faster compared to those far off
- LTM is equivalent to the support of that fuzzy set in Fas ARTMAP
- STM component depends on the current pattern presented and competitive activation of other neurons

FAS-ARTMAP with STORE

The attempt to mimic the memory by ordinary differential equations (ODEs) [164] is the basis of STORE memory module. A method to discretize the solution of DE was proposed. The inclusion of STORE in FAM (or FAS-ARTMAP) architecture permits some patterns to form clusters in ART-x independently from each other. The sequence of winner units is stored and related to the predictions of ART-y through inter- ARTMAP Ws,

Boosted-FAM

Boosted-FAM [66] is a modification of prob-FAM and it improves generalization. It is proposed to reduce category proliferation in FAM. Boosted-FAM takes care of error in training data. But, limitations become critical with exponential increase in the number of input features.

Eqn. 02: Equation for activation function in Boosted_FAM

$$Error_{Total} = \frac{\sum_{j=1}^{Nx} \left| w_j^{ab} \right| - \left| w_{kj}^{ab} \right|}{\sum_{j=1}^{Nx} \left| w_j^{ab} \right|}$$

Boosted-FAM

If Error-total > error-max
Then $\rho_j^{new} = \rho_j^{old} + \Delta\rho$

If there is no match found in existing patterns during classification
Then reset unit is activated

Architecture.Boosted-FAM: The two unsupervised fuzzy-ART modules are linked by a map field [85]. A unit in $F2^x$ may be linked to many in $F2^y$ with different probabilities. There is no change in the structure of fuzzy-ARTMAP. The activation function is same as that used in prob-ART. In ARTx, each category (j) has a different vigilance parameters ρ_j^0. The inter-art-map stores probabilistic relations. So, categorization of input space

is determined by vig.par.x. The inter-ART reset mechanism [85] is suppressed. Thus, an unsupervised on line learning cycle is performed. Probabilistic setting is introduced to evaluate the need for committing new categories. In off-line evaluation, Tr error will determine whether a new cycle with higher vigilance is

Chart 24: Advantages of *Boosted-FAM*

- Solves category proliferation
- Optimizes size of categories
- Robust to noise in training data
- Suppresses match tracking
- Preserves the association of each category to a hyperbox
- If-then-else rule extraction from learned weight is straight forward procedure

required or not to create finer categories. The vigilance parameter ρ_j^0 is initiated to a low value during training . The total training error is averaged sum of error contributions of all categories in ARTx. The advantages are briefed in chart 24.

Mu-FAM

Gomez-Sanchez [85] reported mu-FAM in 2002. It reduced category proliferation, allows a little error in training data and avoids over fitting. It combines probability information and inter-Art reset mechanism. Learning in two stages with modification in the strategy of mu-FAM enhances predictive accuracy and generalization performances.

Architecture. mu-FAM: The architecture is same as that of FAM. The inter-art reset mechanism is governed by probabilistic equation like in prob-FAM. But the vigilance of ART-x does not increase. ART-x vigilance is not raised when inter-ART reset mechanism is fired. Therefore, does not result in category proliferation. In the second phase, some more patterns are presented again with increased ART-x vigilance parameter value and total prediction entropy is calculated. This two stage learning avoids new categories and permits placing hyperboxes inside other hyperboxes. An off-line learning strategy guarantees prediction accuracy (**Alg. 12**) It also allows one-to-many F2x → F2y relations, just like in Prob-FAM. W^{xy}_{jk} stores probabilistic information. Each category of ART-x has a vigilance parameter like in Boosted-FAM such that category size is determined by the underlying distribution. The limitations and remedial measured are given chart 25.

Chart 25: Shortcomings and remedies of Mu-*FAM*

- Lacks tracking population exception
- Category proliferation with noisy data
  - Reason : Lack of inter ART mechanism, becomes critical in high dimensional space
    - Exceptions are not handled properly
    - Hyperbox cannot be created inside another hyperbox
- Poor generalization
  - Remedy: Batch training instead of match tracking mechanism

Alg. 12: **Algorithm of mu-FAM** [60]

```
For   Epoch = 1 to epoch_max
      If epoch = 1, NPP= NP, end
      For   i=1:NPP

            Input ith pattern and corresponding label(y)

                          Learning
            If    label of input passes through resonance test &
                  entropy test
            Then update the W of the category
            Else  reset the winner
                  Next winner is tried
            End if
      End for
End for
```

***Off line evaluation*_mu-FAM**

```
For i = 1: npattern
      Input ith pattern without changing W vector
End for

Calculate total entropy

If    (total entropy < threshold) or (epoch > epoch_max)
Then training is completed

Eliminate category which contributes highest amount
to total entropy
Pick up the patterns concerned
```

Parameters of mu-FAM

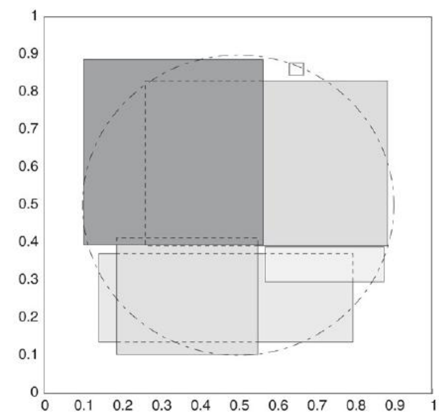
Rule 9

If hmax = $\log_2 N^b$ i.e. at the limit
Then mu-FAM suppresses inter-ART reset & thus behaves like Boosted-ART

Rule 99

If Hmax = $\log_2 N^b$
Then Off line storage is not necessary

If Rule 9 & rule 99
Then mu-FAM reduces to prob-ART-NN



Circle in a square with centers at the same point (Courtesy Ref: 60)

Dataset.simul.circle_in_a_square.mu-FAM: 1000 uniform random points are used for training and another 1000 data as test cases. The task is to decide whether these points are inside or outside the circle lying within a square of twice its area.

Dataset.simul.Gaussians_overlapping.mu-FAM: A two class dataset with 1000 data points is generated. The probability of class-1 is 1/8, while that for second (class-2) is ½. The test data contains 10,000 data points and mu-FAM has lower error compared to Boosted-FAM and FAM (table 21).

Dataset.Character_handwritten.mu-FAM: 2106 upper case letters are trained from UNIPEN database containing 5-million characters. The features extracted for a character include stroke length, three-angles describing the curvature of the stroke, ratio between the sides of the box containing the whole character. mu-FAM produced less number of rules (four per character) while FAM generated 10-rules per character.

Table 21: Performance of mu-FAM, Boosted- FAM and Fuzzy-ARTMAP for simulated and real data sets

Method	Circle-in-a-square		Overlapping Gaussians		hand-writing-recognition	
	Rules	% Error	Rules	% Error	Rules	% Error
mu-FAM	**10.**	**5.2**	**12.**	**5.5**	**105**	7.0
Boosted-FAM	39.	6.9	16.	6.3	489	6.7
Fuzzy-ARTMAP	25.	5.7	27.	6.4	254	6.0

Extraction of rules of mu-FAM: Like fuzzy-ARTMAP, mu-FAM also extracts rules of the type
If a is bj, then output is Ck from the weights of trained NN.

Relation of mu-FAM *with other techniques:* It is related to ID3 algorithm of Quinlan [163] proposed in 1986.

Safe-mu-FAM [60]

Gomez-Sanchez et al. [85, 97] proposed safe-mu(or micro)-FAM to reduce category proliferation. It also allows encoding patterns mapped to different classes. The number of patterns with mixtures of labels is controlled by the entropy of the categories. Here a mixture of labels is allowed in a category or categories. Zhong et al. [60] identified good default parameter values needed for the NN.

Datasets.simul.Gaussian.classification.safe-mu-FAM: The simulated data sets with two features varying number of classes (2, 4, 6) and different extents of overlap (5, 15, 25, 40%) are used for safe-mu-FAM. Each dataset consists of 500 training, 5000 verification and 5000 test samples.

Datasets.IRIS.safe-mu-FAM: From the IRIS dataset, the data corresponding to linearly separable class and the first two features are deleted. For each of the remaining (100) points, 100 noisy data with N(0,small) is generated with a total of 10,100 patterns.

Datasets.Abalone.safe-mu-FAM: Abalone database shows the functional relationship with 8-features, of which, the categorical attribute is eliminated. The output (y:age) is divided into three classes.

Datasets.Page.safe-mu-FAM: The database, Page (from UCI repository) contains the blocks of layout of the pages in a document. It is a classification task and is typical in the sense that this dataset contains 80% probability of occurrence of major class. The performance of safe-mu-FAM is superior to fuzzy-ARTMAP, elliposoidal-ARTMAP, distributed-ARTMAP and Gaussian-ARTMAP.

GA-FAM

Al-Daraiseh et al. [54] tackled category proliferation and generalizability by ternary hybrid NN comprising of genetic algorithm, fuzzy set using ART philosophy. GA is applied to modify the number of neurons in the hidden layer and Ws. The concept of GA-FAM can be extended to EAM.

GAs to design topology of NNs: In direct coding, each connection of neurons in the NN is represented in a chromosome. This method, effectively prunes the ineffective neurons as well as Ws based on chosen set of performance criteria. The pruned network is optimal and noteworthy results have been obtained. In indirect encoding method, the topology and learning algorithm are fixed (ex. FFNN or Rec-NN) and the parameter values are refined by GA. The algorithm is depicted in alg. 13.

Alg 13: GA-Fuzzy-ARTMAP

Initial population
Base line vigilance parameter
Choice parameter

Order of training changes for each candidate of population

While not converged
  A set of FAMs are trained using ART's chaining rule
  Two parents are chosen from pop-1 by tournament selection
  Cross over operation
  Cat.del operation
  Mutation : The chromozomes from the previous step are considered. For each category l(p) or u(p) are selected randomly or by 50% probability. They are mutated by adding a random number (N(0,0.01))
  Test for convergence
End while

Fitness function.GAFAM:
FitFn(i) = PCC(c) – α * (ncat(i-mincat)

α	:	user controlled parameter reflecting accuracy against complexity
cat-del	:	This operator destroys categories in ART-NN during evolution and finally stabilizing at a smaller ART-structure.
ncat(i)	:	Number of categories
PCC(i)	:	Percent of correct classification
mincat	:	Number of classes if known (otherwise user chosen)

Datasets.GA-FAM: Twenty seven data bases are studied with GA-FAM and the results are compared with FAM, EAM, GAM and micro-ARTMAP. The simulated data sets pertain to 2D-classification with 2, 4, or 6 classes and one/two/four circles in a square. The experimental data sets are IRIS, abalone, PAGE, optidigits, pendigits, satellite image, image segmentation, waveform, glass, Pima Indian diabetes etc.. The classification performance between GA-FAM and other methods (ratio of categories by GA-FAM and number of categories by another method) is greater than 10%.

Fuzzy-ARTMAP-Georgiopulos_1999

Daghar and Georgiopoulos [109] showed that the general performance and number of iterations of ordered-fuzzy-ARTMAP is superior to FAM. Max-min clustering algorithm is used to identify the order of presentation of training patterns. The training phase is identical with FAM. It is applicable to other ART-type NNs including LAP-ART, ART-EMAPQ and ARTMAP-IE. Nine classification datasets -- IRIS, wine, SONAR, DIABETES, BREAST, BALANCE, BUPA, Cars and glass-- are used to test the performance.

Fuzzy-ARTMAP-Granger_2001

Granger modified fuzzy-ARTMAP in 2001 [96] as a hybrid system utilizing modified FAM and well proven statistical/mathematical tools.

Architecture-Fuzzy-ARTMAP-Granger: Fuzzy-ARTMAP is combined with on-line clustering algorithm (Nearest neighbor matching with a band of Kalman filters) and evidence accumulation module. Negative match tracking (MT^-) implemented in ARTMAP-IC is used resulting in high level accuracy and compression of radar pulse data set of limited size. Further, the modules executing familiarity, discrimination, indicator vector strategy, and category instance counting, on-line clusters algorithms (k-NN) and Kalman filters are employed for analysis of radar data. Indicator vector strategy processes partial input (patterns with missing components during training/testing). Familiarity discrimination allows detecting patterns belonging to unfamiliar classes presented during testing.

Fuzzy ARTMAP-Charalampidis _2001:

Charalampidis [95] modified the testing phase of FAM, where features capture only shape characteristics of signal and not the actual/average amplitude. It is applied to classification of noisy signals of image segmentation in grey scale like encephalographs, electrocardiograph, and recognition of speech signal and satellite photos. Some of the modifications in FAM are elimination of match tracking, replacing WTA by distributed takes all/winner takes most, modifying match tracking/activation equations, overtraining, validation by CV and adopting slow learning especially for noisy data. (chart 26)

Chart 26 : Features of **Fuzzy artmap-charalampidis _2001**

+ Generalization of this NN is superior to FAM
+ Generalization of this NN is superior to FAM
+ Correct classification independent of illumination

— Noise distorts output

Ordered fuzzy ARTMAP

This method does not require (vigilance and choice) parameters. The size of the network is comparable to those employing a random order of presentation of patterns. Further, the computational overheads are also a little.

Castro [68] partitioned the training data into smaller sets by Hilbert space filling method. Each of these sets is trained with a different FAM. This approach is in contrast to the long practiced single FAM with all the data in one shot. It can be extended to Prob-NN, RBF and other competitive classification NNs. Parallel implementation of match-tracking of FAM is a more involved task.

5. ART&ARTMAP in research mode (2013): Auto resonance theory (ART), the brain child of Grossberg is an expert now in solving classification/pattern recognition tasks for binary/floating point/symbolic/image input and had an impact even in autonomous-robots' pursuits. Many of modification, advances and noteworthy applications are culmination of three decades of research results of Carpenter and Grossberg. The contributions from other groups are responsible in enhancing fragrance of a thousand petal blossomed flower of neural network models. The state-of-ARTMAP NNs in the method-base format are described in Chart 27.

Chart 27 State-of-art-of-ARTMAP in research mode

XARTY	
Unsupervised	**ART**
Supervised	**ARTMAP**

Software packages
Matlab
Professional II
….

Method Base_ ART

ARTX
None
ART-1
ART-2
ART-2a
ART-3

Hybrid ART
None
Binary
Ternary

(Binary) Hybrid ART
None
Fully_Organized-SO- ART

Ternary hybrid ART
None
Fuzzy + ART + growing cell SOM

XART
None
Fuzzy-
Coupled-
Prob-
Projected-
P(erformance) guided-
Grey-
Multi-ART
Efficient-ART
Dynamic FASART
Biased ART
S-ART

Parallel- ART
None
Parallel-alg

Miscelleneous-ART
None
ARAM

Method Base_ ARTMAP

XFuzzyARTMAP
None
Ordered fuzzy
Hyper-spherical-
RBF-
Bayesian-
prob-
Modified-
Mu-
Safe-mu-
GA-
Granger_200 -
Georgiopenlos_1999-
Charalampidis-
FEMA
pipe lining FAM
Dynamic

XARTMAP
None
Gaussian-
Instance counting-
Georgiopenlos_1994
Cascade-
Hierarchical-

Rule Discovery	Fusion NNs
ARD ARTMAP Rule Discovery	FAM
SOARD Self organizing ARD	

Ensemble
FAM

Basis Functions

Hyperbox
Adaptive Basis functions

Abbreviations and definitions

Abbreviations		Definition
ARAM	:	AR Associative Map
ART	:	Adaptive resonance theory
SO	:	Self-organized
STM	:	Short-term memory
MTM	:	Medium-term memory
LTM	:	Long-term memory

DIRECT	:	Direction-to-Rotation Transform
Road	:	Reactive obstacle avoidance direction

MISO	:	Multiple input single output
ARTMAP	:	ART mapping
k-NN	:	k-Nearst neighbor

LGN	:	Lateral geniculate nucleus
PPC	:	Posterior parietal cortices
PFC	:	Prefrontal cortices
SEF	:	Supplementary eye fields
SEER	:	Surveillance, Epidemiology, End Results
NCI	:	National Cancer Institute

6. Current state of MaNNs and artificial brains

In 1986, the publication of a book by Rumelhart proposing back-propagation, a learning strategy used earlier by Werbos [7,8] reawakened surge of feed-forward-multilayer-NN models. This revitalized moral/financial support for ignored (worthwhile) discipline for over one and half decades (also called dark period). This new learning rule allowed the construction of networks which were able to overcome the problems of the perceptron-based linear networks. They could solve more complicated (non-linear) tasks even in multiple dimensions with numerical-, symbolic-, black-and-white-/grey-/colored image (pixel-, voxel-) data.

NNs on hardware: Recently, the progress in implementation of NN models in hardware during the last two decades is reviewed. All major NNs (MLP, RBF, SOFM, CNN, Rec.NN etc) are realized in digital, analog, neuromorphic and FPGA (field programmable gate array) electronic chips. Reconfigurable FPGA based designs are available for spiking NN and cellular NN. CMOS/ Nano-wire/ nano-device ("CMOL") technology is used for large scale hard-NN implementation. It has the advantage of flexibility and high density of molecular scale nano-devices. Generalized Hebbian algorithm was implemented on hardware for a classification task. The images (4 x 4) of 32,000 training samples of five textures are modeled with 90% success. In the hardware, PCA and W-training run in parallel. The success rate can be improved with increase in number of species or increasing the dimension of the vector.

Hardware mimicking -- traits of human brain

The success of IBM –Blue chip in winning over human grandmaster in chess, soccer game with robots and today's humanoid robots are a testimony for fruits of cross fertilized research of physiology, computer science and instrumentation. Now we are able to peep through a science window of evolving intelligence (E-I and E-senses) only a subsection of infinite nature. It is realized that even micro process(es) in the living cell plays role to the overall function at organ stage and finally species level. Years-ahead research will enable experts to visualize, simulate and reorder finally the wayward chemical reactions of bio-molecules and non-bonded interactions. It is still an open question to what extent order/disorder/ chaos/ probability/possibility/ fuzziness, making/breaking bonds and mutations all have their role.

(Partial) Rat brain on hardware

Blue – brain – project by Henry Markram is conceived as the best artificial brain projects on the planet. An IBM "Blue Gene" super computer with 10,000 computer-chips is used to simulate/ mimic the part of neural signaling of cortical column of a 15days old rat brain. Each processor (CPU) functions exactly like a physiological neuron with details at ion-channel level. The dream of artificial brain became partially true in this project.

Artificial human brain

Human brain is most complex in humanly known universe. The limitations in the last century for the expected progress in AI to realize artificial brain are inadequate hardware and knowledge of neuroscience. John Taylor discussed NN based artificial brain architecture and its implications for consciousness and AI. Henningsen described a hierarchical network of NNs model for the human brain. This model explains lower level brain simulation dynamics with higher level cognitive aspect in a thorough manner.

To put it in a nut shell, the diverse hot topics of expert researchers are

- Mimicking a part of the brain through software. For exmple object recognition using V1,V2 and V4 etc.
- Implementing very large scale NNs on hardware chips.
- Embedded NNs in an industrial process or F15 war planes along with other mathematical models, solutions, methods etc.

Cohen-Grossberg NNs: Traditional hierarchical portioning clustering algorithms always fail to deal with very large databases. The stability and existence of periodic solutions to delayed Cohen-Grossberg-BAM-NN using continuous coincidence theory are reported. New criteria for exponential stability for interval-Cohen-Grossberg-NN with time varying delay is the basis of a study wherein the methods used are linear matrix inequality, matrix norm and Haloney inequality procedures. Recently, global asymptotic stability of equilibrium solution to generalized Kohen-Grossberg-BAM is reported in which the boundedness on activation function is removed.

7. Future track (2015-) prospects-Afterword

The expert opinion is that time has come in the 21st century to think of artificial general intelligence including artificial brain. The future single-minded focus on bull's eye is hardware system realization of realistic brain of human level and best sensory processing of animal world. In the human domain, brain-mind-consciousness complex is another crucial dimension to jump into super intelligence, artificial species (life) etc. The knowledge extraction in complex equilibria to decipher protonated/ hydroxylated/ mononuclear metal-ligand systems, alternate mechanisms in kinetics of reactions, deciphering IR/Raman active molecules are under active consideration using ID3 algorithm and fuzzy ART based approaches. A better understanding of psycho-physiological processes not only revolutionizes neuro-surgery, organ-substitutes, but unravels secretes of the nature of human brain to combat with the eradication of dreaded diseases with complimentary knowledge of genes. The subsystems comprise of mathematical/informaion sciences, intrinsic-cross- scientific-disciplines, cloud computing and so on. The data/ noise/ outliers are from deterministic, probabilistic, fuzzy, possibilistic, difference trends viz. wavelets, ridgelets, rational-polynomials. The data structure ranges from binary, multi-valued, floating point, imaginary/ quaternion numericals, symbolic/text, pixels/voxels in 3-way and 4-way tnsorial form. The innovations hitherto realized torch curiosity in exploratory research snowballing mind-blowing products with reassuring outcome. The firm contemplation of Markram for simulation of full cortex of human brain by 2018 is a navigator for the research with nature-evolution.

Acknowledgements

Prof UMK introduced 'Chemometrics' to me in the year 1980. My first contribution was an article 'Chemometrics-Social relevance' in a symposium in 1982. We thank Prof W U Malik, Dr G Saxena for promoting the special symposium on 'Chemometrics' in annual meeting of Indian Council of Chemists at Madurai in 1986. It was a motivating and steering conference for young researchers towards chemometrics under the sectional presidentship of Prof Murali Krishna. I am ever indebted for his unstinted encouragement, tit-bits of astuteness and also supporting my research with chemicals.

RSR thank Prof. G Nageswara Rao in appreciation of timely execution of even tough research programs for over a decade during the beginning of his research career. Dr GNR posses a positive attitude in molding himself without a second thought even to blind-fold-tasks with no fruits except learning experience.

8. Appendices

Appendix-Ia: [Animal/human] Learning

Natural learning: Human beings (viz. common man, wise expert, target oriented person) and even animals perceive (hear, see, feel by touch, smell the odor, taste the material) the surrounding world with their sense organs on the fly and spend more time only in typical circumstances. Learning, unlearning, life-long learning, forgetting, selective-forgetting and catastrophic-forgetting are buzz words in the realm of experiences of people. Although they learn fast, the memories persist for a long time (days, months, years or decades), in spite of continuous learning of new traits. They learn sequences of data in real time mostly in unsupervised manner wherein nature itself is a teacher. The other route of learning is an unpredictable concoction of unsupervised and supervised integration trials.

The ability of humans to vividly remembering adventurous/exciting movies is a common example of fast learning in an unfamiliar environment. Recognition of familiar objects, learning new ones, internal representation of external world as seen and communicating information to other members of the same species are usual not only to human beings, but to many living creatures. A few typical species are sniffing dogs/pigs, ants, dolphins and dancing-honeybees. We know and feel it to be natural that human beings need not start from 'square A' to learn a new thing. Also, one need not forget what all we learnt/know to learn a new aspect, an exception or extension to the learnt lot.

What a treachery it is, if we were to forget (erase/overwrite) old learned traits/skills/intelligence bits to learn new ones? This dilemma was called stability-plasticity and the brain has a capability to circumvent it. A utopian wish is the man made system also should be adaptive (plastic) to new relevant information and yet be capable of being stable (rigid) too for irrelevant input changes. A way out of the stability–plasticity dilemma depends upon bottom-up data patterns as well as learned expectations or top-down ones and rules of match/mismatch. Thus, ART is a self-organizing heuristic system. It is a concept well penetrated in modeling and is a corpus inside both supervised/unsupervised/fused bunch of Carpenter and Grossberg's computational niche.

Formal learning: In this knowledge/ information based society, humans do learn from explicit scheduled/unscheduled training from prescribed text book or from a qualified teacher followed by oral/ written examination on-line or off line. This is start of process for one and half decades succeeded by learning from prototype/novel/unpredicted-exceptional/rare-catastrophic incidents. This is a gateway for same/similar/new/novel information promoting integration in the brain at individual level. The idealistic motto of education, learning, tutorials are motivating learned individuals for attention-grabbing towards a lifetime-independent-learning. It facilitates to start the relevant profession and to continue to grow up the ladder. It is just not to bring forth skilled/routine teachers/practitioners sustaining continuity of education/transfer of technology. The fringe benefit is, of course, earning livelihood while revising/retaining the tidbits of learnt.

Traditional data analysis: When all the data for analysis are available, information/knowledge extraction in offline or online mode is performed. Further, batch /sequential approaches are again method/algorithm dependent and personal choice. The analysis of data is again under the same umbrella, pros and cons and information/knowledge extraction/ discovery in yester years was also academic interest. It was mostly implemented off-line and only when all the possible data are available. The realistic constraints and limitations in data analysis (classification/discrimination/clustering/PR) of natural/man-made objects, satellite maps and mimicked industrial products are multi-fold at varying levels of hierarchy. The samples are sparse; data is measured over varied time periods (in static/dynamic mode), in distant locations on different scales/ precision/accuracy with distinct sub-goals/goals and time/finance requirements/constraints.

Going one more step ahead, the information may be conflicting/ refuting/ supporting earlier well-trodden hypotheses/findings. The recent advances in information fusion tools promoted resolution of conflicts and set right of inconsistencies at any/all phases of information to intelligence repetitive-closed-loop directing experimental/ simulation/ computational activity towards enhancing truth value of truth and decreasing falsehood value of truth.

Unsupervised/ supervised Models

The classical unsupervised method learns/ models causative variables from (explanatory, environment) features-data. In the case of supervised approach, for each datum of features [X(i,:)], a corresponding response [y(i,1)] scalar for uni-response case or a vector of response [y(i,:)] for multi-channel response data acquisition system are available. When response is not available for some feature combinations, semi-supervised learning using all X vectors and those containing both X and y is the choice. Auto-/adaptive-/self-judgement-machine intelligent and (expert) human choices include self-directed learning/integration taking into consideration of features not previously incorporated.

Appendix Ib: Innovations in NN-research during 1970 to 1986

Ackley	Boltzman machines	1985
Amari	Neural theory-assoicaation and concept formation	1977
Anderson J A	Probability learning-categorical perception	1977
Anderson J A	Distributed memory storage	1984
Barto	Neuron like adaptive elements	1983
Farhat et al.	Hopfield model-optical implementation	1985
Fukushima et al.	Neo-cognitron	1983
Grossberg	How brain builds cognitive code	1980
Hopfield	Collective computative abilities- physical systems-NNs	1982
Hopfield	Neurons with graded neurons	1984
Kohonen	Memories-correlation matrix	1972
Kohonen	Self organized formation of topologically correct feature maps	1982
McClelland Rumelhart	Letter perception	1981
Rumelhart et al.	Parallel distributed processing	1986
Rumelhart et al.	Learning representations by back-propagating errors	1986

Sejnowski et al.	NETtalk- parallel network that learns to read aloud	1986

Appendix Ic: [Machine] learning -ART-based

Hebbian, anti-Hebbian, Widrow-Hoff rules, back propagation, different functions [wavelet,Trigonometric, polynomial], nature inspired approaches are in wide use.

Typical training algorithms, their limitations and remedy

ॐ Feed forward clustering machine learning algorithms
- Competitive learning
- Self organizing maps
- Neocognitron
- Back propagation

— Catastrophic forgetting for non-stationary (dynamic) trend

Remedy: ART

Types of learning based on input [X ,y] and features

ॐ Learning
- Un- supervised [X]
- supervised [X] [y]
- Semi-supervised

— fixed number of features throughout training and testing

Remedy: Self- supervised

If	Synthetic or planned experimental data
Then	number of features fixed throughout training and testing &
methods	[unsupervised/supervised/semi-supervised]

Winner-takes-all (WTA)

In winner-takes-all mechanism, only one category is allowed to learn or dominates this competition process. In real nervous system, there are parallel pathways from multiple sensory cells and also to the multiple motor cells. The competition of the path ways with each other includes inhibitory connections. Goldfish has a pair of Mauthner cells in the region of hindbrain connected to the right and left audio-tactile system. Each of them is connected to the motor nerves which cause the contraction of the muscle, in the opposite side. The fish on hearing a loud sound turns the body to the opposite direction. In essence, primordial motor action is the output of surviving path ways in presence of inhibitory competition. This is the biological inspiration of WTA concept in ART type NNs. The other successful approach is winner-takes-most (WTM).

Hyperbox

The categories arrived in Fuzzy Art are seen as hyperboxes. A hyper box starts as (isolated) point in multi-dimensional space. It increases its size up to a maximum restricted by vigilance threshold. When a new pattern is presented, the current pattern may be inside the already existing hyperbox. Otherwise the hyperbox is enlarged just to include the new pattern. If the pattern cannot be fit into the exiting hyperbox even after maximum expansion, then a new hyperbox is invoked reflecting a new category. Their associated weight vector Wj defines the corners.

Vigilance parameter

Vigilance parameter reflects the degree of match like a photograph resembling an object. It controls the degree of mismatch (which the system can tolerate) between the new patterns and the learned (stored) patterns in ART-type NNs. The range of vigilance parameter (vig.par) is 0 to 1. It sets the upper limit for the size of the hyper box. For example, if vig.par is set to a low value (0.001), every pattern will be considered to belong to a class and thus, all data points results in singleton clusters. At the other extreme, with a value of 0.999 for vig.par, even a distinct pattern from the learned ones, appear similar to it and thus, the net result is a single cluster of all heterogeneous patterns. Vigilance parameter determines the activity of the reset unit and is activated when there is no match between a new pattern and existing ones. A few heuristics for learning with variation of vigilance parameters are given here.

ART-1 is stable irrespective of vigilance parameter. In other words, final clusters do not change even if a few more patterns are trained. It controls granularity of clusters produced by ART-1. A small numerical value allows for large deviation from cluster centers. Hence, it leads to small number of clusters. In fuzzy-ART, vig.par is $R_j \leq M * (1 - vig.par^x)$, while in FAM it is relaxed as match tracking raises base line vigilance parameter. However, in mu-ARTMAP, vig.par is set to zero.

Matchtracking-ARTMAP

- ◎ Resets its mistakes
- ◎ Learns from mistakes
- ◎ Guarantees WTA network passes the Next Input Test

— **Does not guarantee a distributed prediction**

Memories ARTMAP

- ◎ Include broad and specific pattern classes
- ◎ Specific pattern classes formed as exceptions to general ``rules"

ARTMAP learning

- ◎ Produces mixtures broad and specific classes
  - + Composition depends on order of presentation of training patterns

If	Vigilance parameter is large
Then	narrow (broad) generalization and prototypes representing fewer exemplars (Larger lot of tight clusters)
If	Vigilance parameter is small
Then	It leads to broad (narrow) generalization and abstract prototypes
If	Vigilance parameter nearer to 1 (0) i.e. in the limit
Then	Prototype learning reduces to exemplar learning
If	(I U B)/│I│ < Vigilance parameter
Then	Insufficient matching
If	(I U B)/│I│ > Vigilance parameter
Then	Matching sufficient to the extent of Vigilance parameter & modify weights

If	Novel features from which to learn & Semi-supervised algorithm
Then	Unlabeled data used to refine model parameters

Learning. Distributed_ART

The objective is to support stable fast learning with arbitrarily distributed F2 code patterns of y. It involves dynamic weight, distributed instar/ distributed outstar learning and increased gradient content-addressable memory.

Semi-supervised and self- learning systems: An unsupervised learning system clusters unlabeled input. The semi-supervised system learns from unlabeled as well as labeled inputs during training. The self-supervised paradigm models two learning stages. During the first stage, the system receives all output labels, but only a subset of possible feature values for each input. In the second stage of learning, the system receives all feature values for each input, but no output labels. The self-supervised learning system introduced here models life-long experiences

If	Distributed ART & Fast learning
Then	Beta2 = 1
If	Distributed ART & Fast learning
Then	Wbu ~= wtd
If	Contrast parameter increases
Then	T → y transformation becomes Increasingly sharp
If	Contrast parameter → infinity
Then	→ WTA

Numerical, classical-AI and NN computations: Turing or von Neumann architecture is a general framework for computing. In the words of Grossberg, the architectures (for audition, cognition, cognitive–emotional interactions, sensory–motor control and the like) developed by his school are far more general than traditional AI algorithms for vision.

Machine Learning: The consolidated analysis of human cognitive information processing and sTable coding in a dynamic real environment is the inspiration in promotion of ART mathematical models using set theory, complimentary coding and solving differential equations under constrained assumptions. This class of NNs viz. ART and ARTMAP developed in collaboration with Carpenter is now a paradigm by itself. It is competitive in numerical modeling of classification/PR and excelled many statistical and other time tested algorithms. Information fusion [69], hierarchical knowledge discovery and fusion of neural networks set a new trend in white box approach of NNs. This promotes automatic knowledge/intelligence detection/up gradation/discovery in cross-disciplinary research pursuits. The multi-modal /self-organizing/ hierarchical type of information from multiple sensors is the central core of these investigations. Ensembles_of_FAM generate robust and reliable clusters and rule set. The trend setting NNs in this series are self-supervised ARTMAP [40], biased ART [41] and impact of adaptive basis functions [107] with continuous SOM.

If	Human and machine learning
Then	Scope for self-directed learning incorporating features not previously experienced in traditional learning method : [Self-supervised ARTMAP]

Appendix-Id: Rule/knowledge extraction

Although rule discovery systems replaced the yester-years knowledge engineer extracting rules from subject expert, the current trend is to generate rules from numerical processing of data. Rule generation from Ws of trained FF-NNs for classification and later for function approximation is successful. The rule extraction procedures raised the status of non-linear (NL-) feed forward (FF-) NNs from black-box to explicit transparent and expert systems. A detailed account of rule generation systems will be core of another tutorial review.

Rule generating systems

- algorithm
  - Eclat
  - Frequent Pattern Growth
  - Evolutionary
  - Apriori
- Fuzzy NNs
- discovery and inference
- GEPCLASS
- Knowledgetron

Appendix-Ie: [Robot] learning with FAM

DIRECT

- \+ Handles a wide range of perturbations in robot
  - \+ Joint locking
  - \+ Tools-of- no- prior-experience
- \+ Fault tolerant
- — Obstacles in environments
- **Remedy** : **Direct-Road**

DIRECT-ROAD

- \+ Avoid obstacles in environments with simple obstacle configurations
- — Not all configurations of obstacle

The commands for movements in an obstacle-free space of a redundant robot arm are self-generated. These correlated visual, spatial and motor information bits are the input to Direction-to-Rotation Transform (DIRECT) - learning-model- module. It controls forward and inverse kinematics and employs online-fuzzy-ARTMAP-NN. DIRECT-ROAD, self-organized process of mental rehearsals of movements (with FAM) and futuristic implementation of latest ART/ARTMAP/ARD/SOARD will allow flying into finding truth of grazy-reality and transparent-illusionary-falsehood of visible and invisible universe at physical/mental comprehension perspectives.

ॐ **Remedy :** self-organized process of mental rehearsals of movements (with FAM)

SO(FuzzzyARTMAP)-Mental-Rehearsals-Of-Movements

- Accounts for all the obstacles in its environment
- Retrieves multiple-solutions to reach each target
- — Human and animal experiments on reaching complex environment

Application

- Real robot platforms
- Successful obstacle avoidance during reaching tasks in real-world environments

Appendix-II: ART-Current status

Short-term memory (STM): The patterns emerging in attention system belong to this category. The activity of pattern in the resonance state is for a short period and hence called STM. The STM traces exist only in a single application of the input vector. Learning (change of Ws) in ART occurs only during the resonance period.

Long-term memory (LTM): The connection weights between the input layers and output layer represent traces of long-term memory. The time required for refining weights between PEs is much longer. Thus, traces (WTDs) encode information for an extended period and remains with the NN. Grossberg developed over years a number of equations for STM, MTM , LTM and also a large number of modules/micro-circuits which comprise of shunting on-center off-surround networks, gated dipole opponent processing networks, associative learning/ reinforcement learning/ spectral timing networks etc.

Complementary Computing: The global brain is organized into complementary units where parallel processing streams interact generating biologically intelligent functions/behaviors.

Laminar Computing: Cerebral cortex is organized into layered circuits. The specializations in these layers support all forms of higher-order biological intelligence. The domain of laminar computing concerned with the details here. It embodies similar to laminar computing. It is an unknown blend of best characteristics of many processes. In the language of modelling, we understand and communicate them as digital/analog input, feedforward and feedback NN architectures and data driven bottom-up and hypothesis-driven top-down processing.

Probabilistic models , ART, Brain

◎ **Bayesian statistics**

- Nonlocal computations compute entire probability distribution
  - Work best in dealing with stationary data
  - Fails for dynamic patterns
    - Remedy: ART

◎ **ART**

- Only local computations
- Principles/mechanisms go beyond Bayesian analysis
- Embodied parsimoniously in the laminar circuits of cerebral cortex.

In an algorithmic mode, singular approximation to the full dynamics of an ART system is used. STM activations are solved at equilibrium for fast dynamics. In spite of the fact, these simplifications are far away from even near reality, the results of mathematical development of the cognitive and neural theory resulted in reliable results for many mega tasks in industry, science and defense. In nature, evolutionary the pressures over millennia implicitly equivalent to specific cases and extensions of now developed equations and modules. It is right place to remember genetic programming produces several equivalent equations for the same task which appear as very diverse for a naked mathematical eye.

Appendix-III: Omnimetrics

Prior to nineteen seventies, elite chemists employed the then advanced statistics and the scientific term in use was 'Chemical statistics'. At that time, Chemometrics did not exist. Swante wold (son of a famous economist Herman Wold, who introduced PLS) coined the word 'Chemometrics' in a project for financial assistance at the suggestion of his friend. Kowalski from USA, Svante wold from Sweden and Massart started Chemometric society. During the first phase of Ph.D. training at Andhra university, India, myself along with another co-researcher, published on statistics for chemists (190d). The ACS Symposium Series on Chemometrics by Kowalski in 1977 is a noteworthy all time reference volume. It spread its wings all over the world and in all disciplines and the two journals (Journal of Chemometrics from John Wiley and Chemometrics and intelligent Laboratory systems from Elsevier) since 1986 are testimony for the growth of this interdisciplinary research pursuit. The biannual reviews on Chemometrics in Analytical chemistry, ACS journal, during the period 1980 to 1996 by Kowalowski and his research group and since 1998 by Lavine is a comprehensive state of art compendium of this discipline in depth and breadth. The contributions of Kateman and Buydens from Univ of Nijmegan, Netherlands formed another corner stone of the castle of Chemometrics. Gesteinger won ACS laurels for Chemoinformatics, where in hundreds of molecular descriptors for a compound, conformer etc is used. In early seventies advanced clustering, classification and pattern recognition methods

Chart A3a: Subbranches of Chemometrics

Chemo_metrics : [	Enviro _metrics	Pharmaco_metrics
	Specio _metrics	Kineto _metrics
	Dieteto _metrics	Nano _metrics
	Quali _metrics]	
Nano _metrics:[	nano_bio_metrics	nano_chemo-metrics
	Nano_medicino_metrics]	

Discipline_metrics :	[Discipline_acronym] + [_metrics]
Omni_metrics :	[Discipline1, Discipline2 …, Disciplinei] + [_metrics]

and multiple linear regressions were used in economics, geology, and psychology with buzz words of econometrics and biometrics. The new phase of biometrics is with the continued progress of the "omics" era with a dedicated objective of the curation and interpretation of biological data. The advances in electronics, optics and instrumentation resulted in hyphenated technique with 3way-data with first order and second order advantages. It necessitated new tools in multi-variate calibration/curve resolution and pattern recognition. The effects of non-adherence of normal distribution of errors for many data sets were circumvented with non-parametric/distribution free methods and neural networks. With growing importance of information theory, need for analysis of conflicting/ missing/interval datasets with no a priori knowledge of model functions, high noise, outliers, breaks, non-convex and multi-modal trends, a new era of computations emerged. Now, with experiments in genomics, particle physics and chemical sciences, information extraction does not fit within the frame of yester year's data-mining perspectives.

Chart A3b: Object mode OmniMetrics (OmOM)

omni_metrics: [	Bio _metrics
	Chemo _metrics
	Pisci _metrics
	Econo _metrics
	Techno _metrics
	Informato_metrics
	software _metrics
	Medicino _metrics
	Biblio _metrics
	performance _metrics
	Metrics _metrics]
-omics	[Gen_omics
	prote_omics
	metabon_omics
	metabol_omics]
	[cheminformatics]

In a nutshell, terabytes of data, thousands of variables from diverse disciplines with multiple vague objectives spurred the acronym 'Omni-metrics' for attribute, binary, numerical (floating point, complex, quaternary), symbolic, imagery (pixels, voxels) data. Omni-metrics (Om) (Chart A3b) comprises of a hybridization/ fusion of measurements/ information/ knowledge/ technology and 'metrics' with thoroughly grown 'scientific disciplines' for generating intelligent bits from experiments /simulation activity and prediction for critical foresight/control/ realization of far off goals.

Appendix-IV: Keyword generation from titles

The title of a research publication is parsed into words making use of separation characters viz. space (blank), coma, semicolon etc. The verbs (is, has), prepositions (of), articles (a, an, the) are pruned. The major key words (here ART, ARTMAP, fuzzy, neural network, SOM) are separated forming a major category. The remaining words are sorted into a class title_keywords . Here and there, combination into compound words (with hyphen or object oriented format) and removal (local discretion) is affected to increase the context based information/knowledge measure. The titles of around 300 publications and full papers are available on ART series, but only the output of typical publications of Grossberg and Carpenter are given in Table A4.1.

This tool is under rigorous testing in combining abstracts/keywords/highlights/ legends of figure/table, column/row headings of tables/algorithms/charts of inter disciplinary publications. The ultimate goal is to arrive at Pareto-optimal reduction in the size of textual information of cross-disciplinary research enabling machine comprehension and also enhance a human researcher to probe into more published research output.

	Table A4-1: Typical hall mark contributions of Grossberg et al. in ART probing into brain research			
Year	**Keywords_generated_from_Title_of_researchPapers**			
2013	How a brain learns	Consciously attend, recognize	Changing world	ART
2012	Stochastic fuzzy delayed Cohen–Grossberg -NNs	Exponential and almost sure exponential stability		
2012	Neural dynamics	Saccadic and smooth pursuit eye movement coordination	Visual tracking	Unpredictably moving targets
2012	Stereopsis and 3D surface perception	Laminar cortical circuits	Spiking neurons	Neural rate models → spiking models
2012	Supplementary eye fields	Item-order-rank working memory	Saccade selection	
2012	Real robots	Obstacle avoidance	Bio-inspired kinematic controller	
2011	How brain rapidly learn & reorganize	View invariant and Position-invariant object representations	In inferotemporal cortex	
2011	Invariant recognition	Multiple-scale task-sensitive attentive Learning	Morph properties of cells in inferotemporal cortex	
2011	Neuromorphic model	Spatial lookahead		
2011	Invariant recognition	3D-vision	Cortical area V2	Transforms Absolute to relative disparity
2010	NN-community (scientists)			
2010	How children learn	To follow gaze, share joint attention, imitate their teachers	Use tools during social interactions	
2009	Cortical dynamics	Natural scenes	Motion-based object segmentation	Obstacle avoidance
2008	SOVEREIGN	Autonomous neural system	Incrementally learning	Rewarded goal
2007	Hippocampus	Learning	Space time	Behavioral control
2007	Consciousness mind	CLEARS		
2004	Neuromorphic model	Chromatic achromatic	Natural images	
2004	Laminar Visual cortical circuits	Perceptual grouping	Synchronization	
2004	Laminar frontal cortex	Basal ganglia circuits	Interaction	Control planned reactive saccades
2004	ARTSTREAM NN-model	Auditory scene analysis source segregation		

2003	Laminar	Cortical dynamics	Visual Perception	
2003	Brain representing & comparing	Multi-digit numbers	Spatial and categorical processes	
2002	Corticogeniculate interactions	Temporal dynamics	Binocular disparity processing	
2001	Spiking neurons	Technology	Neuroscience	
2000	Visual cortical mechanisms	Interacting layers, Networks, columns, maps		
2000	Cortico-cerebellar interactions	Attentiveimitation & predictive learning	Sequential handwriting movements	
2000	Hebbian pairing in cortical pyramidal neurons	Frequency-dependent synaptic potentiation,	depression, spike timing	
1999	NN	Synthetic aperture radar images	Enhancing	Boundaries surfaces
1998	Reticular formation	Saccade generator		
1998	NN model	Simple and complex cell Receptive fields	Cortical maps of orientation, ocular dominance	
1997	Cortical Dynamics	3D- Surface Perception	Binocular and Half-Occluded Scenic Images	
1995	Radar	Synthetic Aperture	Multiple Scale Boundary surface	
1995	VIEWNET Architectures	Fast-Learning Recognizing 3D- objects	[Input] Multiple 2D-Views	
1994	Recognition Segmentation	Connected Characters	Selective Attention	
1993	Autonomous behavior	Changing world	Stable control	
1993	Learning a Head-Centered Visuomotor Representation	3-D Target Position		
1993	Biological Vision	Figure-Ground Problem	Solution	
1991	Vector associative maps	Error-based learning	Movement trajectories Control	Real-time
1991	Visual Perception	Cortical Model	Cooperative Feature Linking	Synchronized oscillations
1989	Cort-x boundary segmentation	Invariant recognition	Cluttered scenes	
1989	Visual	motion perception	Neural architecture	
1989	Cortical Complex Cells	Stereo Boundary Fusion	Multiplexing distributed Data	Maps, Filters, Feedback nets
1988	Nonlinear NNs	Architectures Principles, Mechanisms		
1976	Adaptive pattern classification	Universal recoding	Neural feature detectors	Parallel development

REFERENCES

[1] W.S. McCulloch, W. Pitts, *Bull. Math. Biophy.*, **1943,** 5, 115–133.

[2] F. Rossenblatt, *Proc. Nat. Phys. Lab.*,**1958**, 1, 421-456.
(b) *Principles of neurodynamics*, **1962**, New York, Spartan.

[3] M. Minsky, S. Papert, *Perceptrons*, **1969,** Cambridge, MA, MIT press.

[4] K. Fukushima, Biological cybernetics, **1975**, 20,121-136.

[5] J.J. Hopfield, *Proc. Nat. Acad. Sci.*, **1982**, 79, 2554-2558.

[6] D.E. Rumulhart, Parallel distributed processing, MIT, **1986**, Cambridge, Massachusetts, England (b) J.L. McClelland, D.E. Rumulhart, Parallel distributed processing, MIT, Cambridge, Massachusetts, England, 1987.

[7] J. A. Anderson, E. Rosengeld (Edited), **1988,** *Neuro-computing, Foundations of Research,The MIT press, Cambride, Massachusetts, London, England.* (Full papers and References therein).

[8] L.M.C.Buydens, W.J. Melsses (editors), *Chemometrics-Exploring and exploiting chemical information*,**1994**, Katholick Univ., Nijmegan, Netherlands.
[9] S. Kaski, J. Kangas, T. Kohonen, Bibliography of Self-Organizing Map SOM) papers: 1981–1997, *Neural Computing Surveys*, **1998,** *1, 102–350.*
[10] *Neural computing manual*, **2000**,Professional II/plus, 200, 230 East main street, Suite 200, Carnegie PA 15106 page 36.

ART- Brain

[11] A. Ul Haque, M.H. Nehrir, *IEEE Trans. Power. Syst.* **2014,** (in press).
[12] M. Eastwood, C. Jayne, *Neurocomput.*, **2014,** 133, 249–257.
[13] M. Jin, R. Li , Z. Xu , X. Zhao, *Neurocomput.*, **2014,** 133,309–316.
[14] S. Grossberg, *Neural Netw.*, **2013,** 37, 1–47.
[15] G. A. Carpenter, S. E. Chelian, *Neural Netw.*, **2013,** 37, 93–102.
[16] M. Yavas, F.N. Alpaslan, *Neurocomput.*, **2012,,** 77, 71–81.
[17] N. Srinivasa, R. Bhattacharyya, R. Sundareswaraa, C. Lee, S.Grossberg, *Neural Netw.*, **2012,** 35, 54–69.
[18] M.R. Silver, S. Grossberg, D. Bullock, M. H. Histed, E. K. Miller, *Neural Netw.*, **2012,** 26, 29–58.
[19] Y. Cao, S. Grossberg, *Neural Netw.*, **2012,** 26, 75–98.
[20] S. Grossberg, K. Srihasam, D.Bullock, *Neural Netw.*, **2012,** 27, 1–20.
[21] Q. Zhua, X. Li, *Fuzzy Sets and Systems* **2012,** 203,74–94.
[22] G. A. Carpenter, S. Oliver, *Neural Netw.*, **2012,** 25, 161–177.
[23] S. Grossberg, K. Srinivasan, A.Yazdanbakhsh, *Neural Netw.*, **2011,** 24, 686–692.
[24] R. Ivey, D. Bullock , S. Grossberg, *Neural Netw.*, **2011,** 24, 257–266.
[25] S. Grossberg, J. Markowitz, Y. Cao, *Neural Netw.*, **2011,** 24, 1036–1049.
[26] Y. Cao, S. Grossberg, J. Markowitz, *Neural Netw.*, **2011,** 24, 1050–1061.
[27] G. A. Carpenter, A. K. Ravindran, *Neural Netw.*, **2011,** 24, 208–216.
[28] E. Berghöfer,D. Schulze, M. Tscherepanow, S. Wachsmuth, *IFIP Adv. Inf. Commun.Tech.* ***2011,*** 363, 1-10.
[29] H. deGaris, C. Shuo, B. Goertze, L.Ruiting, *Neurocomput.*, **2010,** 74, 3–29.
[30] B. Goertzel, R. Lian, I. Arel, H. Garis, S. Chen, *Neurocomput.*, **2010,** 74, 30–49.
[31] B. Goertzel, J. Pitt, M. Ikle, C. Pennachin, L. Rui, *Neurocomput.*, **2010,** 74, 84–94.
[32] K. Thomsen, *Neurocomput.*, **2010,** 74, 121–128.
[33] J. Liu, D. Perez-Gonzalez, A. Rees, H. Erwin, S. Wermter, *Neurocomput.*, **2010,** 74, 129–139.
[34] J. Misra, I. Saha, *Neurocomput.*, **2010,** 74, 239–255.
[35] R. Lian, B. Goertzel, R Liu, M. Ross, M. Queiroz, L.Vepstas, *Neurocomput.*, **2010,** 74, 95–103.
[36] M. Jiang, C. Zhou, S. Chen, *Neurocomput.*, **2010,** 74, 113–120.
[37] H. deGaris, B. Goertzel, *Neurocomput.*, **2010,** 74, 1–2.
[38] S. Grossberg, T. Vladusich , *Neural Netw.*, **2010,** 23, 940–965.
[39] S. Grossberg, *Neural Netw.*, **2010,** 23, 1135–1138.
[40] G.P. Amis, G.A. Carpenter, *Neural Netw.*, **2010,** 23, 265-282.
[41] G.A. Carpenter, **2010,** 23(3), 435-451.
[42] N. A. Browning, S. Grossberg, E. Mingolla, *Neural Netw.*, **2009,** 22, 1383-1398.
[43] G.A. Keskin, C. Ozkan, *Quality And Reliability Engineering International Qual. Reliab. Engng. Int.*, **2009,** 25, 647–661.
[44] Z. Xu, X. Shi, L. Wang, J. Luo, C. Zhongb, S. Lua, *Sensors and Actuators B* **2009,** 141, 458–464.
[45] M. Grana, *Neurocomput.*, **2009,** 72, 2065–2066.
[46] J.Cano-Izquierdo , M. Almonacid, M. Pinzolas, J. Ibarrola, *Neural Netw.*, **2009,** 22, 472-487.
[47] A. Kaylani M. Georgiopoul, M. Mollaghasem, G.C.Anagnostopoulos, *Neurocomput.*, **2009,** 72, 2079–2092.
[48] N.G. Brannona, J.E. Seiffertt, T.J. Draelos, D.C. Wunsch II, *Neural Netw.*, **2009,** 22, 316-325.
[49] W. Gnadt, S. Grossberg, *Neural Netw.*, **2008,** 21, 699–758.
[50] A.Tan, N. Lu, D. Xiao, *IEEE Trans. Neural netw.*, **2008,** 19, 230-244.

[51] S. Grossberg, *Neural Netw.,* **2007,** 20, 1040–1053.
[52] A. Gorchetchnikov, S. Grossberg, *Neural Netw.,* **2007,** 20, 182–193.
[53] A. Gorchetchnikov, S. Grossberg, *Neural Netw.*, **2007,** 20, 182–193.
[54] A.Al-Daraiseh, A. Kaylani, M. Georgiopoulos, M. Mollaghasemi, A.S. Wu, G. Anagnostopoulos, *Neural netw.*, **2007,** 20, 874-892.
[55] W. Pedrycz, *Neurocomput.*, **2007,** 70, 1403–1413.
[56] N. El Barbri, A. Amari, M. Vinaixa, B. Bouchikhi, X. Correig, E. Llobet, *Sensors and Actuators B: Chemical*, **2007,** 128(1), 235-244.
[57] J. Castro, J. Secretan, M. Georgiopoulo, R. DeMara, G. Anagnostopoulosc, A. Gonzalezd, *Neural netw.*, **2007,** 20, 109–128.
[58] G.A. Carpenter, C.S. Gaddam, E. Mingolla, *Neural Netw.*, **2007,** 20(10), 1109-1131.
[59] B. Vigdor , B. Lerner, *IEEE Trans. Neural Netw.* **2007,** 18(6), 2007-1628.
[60] M. Zhong, B. Rosander, M. Georgiopoulosa, G.C. Anagnostopoulos, M. Mollaghasemi, S. Richie, *Neural netw.*, **2007,** 20, 245–259.
[61] C. Valencia, G. Espinosa, J. Giralt, F. Giralt, *Comput. & Chem. Eng.*, **2007,** *31(9), 1131-1140.*
(b) J. Chem. Inf. Comput. Sci., **2003**, 43 (1), 85–112; **2002**, 42 (2), 162–183; **2001**, 41 (5), 1177–1207, **2000**, 40(3), pp 859–879.
(c) Ind. Eng. Chem. Res., **2002**, 42 (2), 343–359; **2001**, 40(12), 2757–2766.
[62] R. Bhatt, , G. A. Carpenter, S. Grossberg, *Vision Research*, **2007,** 47, 3173-3221.
[63] H. Ohta, Y.P. Gunjib, *Neural netw.*, **2006,** 19, 1106–1119.
[64] M. Su, J. Lee, K. Hsieh, *Neurocomput.*, **2006,** 69, 2284–2300.
[65] I. Dagher, *J. Discrete Alg.*, **2006,** 4, 538–553.
[66] S. J Verzi, Gregory L Heileman , M. Georgiopoulos, *Neural netw.*, **2006,** 19, 446–468.
[67] G. A. Carpenter, S. Martens, O.J. Ogas, *Neural netw.*, **2005,** 18, 287–295.
[68] J. Castro, M. Georgiopoulos, R. Demara, A. Gonzalez, *Neural netw.*, **2005,** 18, 967–984.
[69] G.A. Carpenter, S. Martens, O. J Ogas, *Neural netw.*, **2005,** 18, 287–295.
[70] B. Vigdor , B. Lerner, *IEEE Trans. Evolut. Comput.*, **2005,** 9.
[71] M. Yeh, S. Chiang, *Neurocomput.*, **2005,** 67 , 313–320.
[72] S. Grossberg, K.K. Govindarajan, L. L. Wysec, M.A. Cohena, *Neural Netw.*, **2004,** 17, 511–536.
[73] J.W. Browna, D. Bullock, S. Grossberg, *Neural Netw.*, **2004,** 17, 471–510.
[74] A.Yazdanbakhsh, S. Grossberg, *Neural Netw.*, **2004,** 17, 707–718.
[75] S Hong, S. Grossberg, *Neural Netw.*, **2004,** 17, 787–808.
[76] X. Liao, C. Li, K. Wong, *Neural netw.*, **2004,** 17, 1401–1414 .
[77] S.W. Lee, D. Palmer-Browna, C.M. Roadknight, *Neurocomput.*, **2004,** 61, 5 – 20.
[78] T. Liu , J. Yeh , C. Chen , Y.T. Hsu, **2004,** 40256 ,407 – 414.
[79] N. Kopco, G.A.Carpenter, *Int. J. Hybrid Int. Syst.*, **2004,** 1(2), 57-71.
[80] S. Grossberg, D.V. Repin, *Neural Netw.*, **2003,** 16, 1107–1140.
[81] S. Grossberg, *Neural Netw.*, **2003,** 16, 925–931.
[82] O.Parsons, G. A. Carpenter, *Neural Netw.*, **2003,** 16(7), 1075-1089.
[83] Y. Cao, J. Wu, *Neural netw.*, **2002,** 15,105-120.
[84] S.G.A. Grunewald, *Neural Netw.*, **2002,** 15,181-200.
[85] E. Gomez-Sanchez, Y.A. Dimitriadis, J.M. Cano-Izquierdo, J. lopez coronado, *IEEE Trans. Neural Netw.*, **2002,** 13, 58-69.
[86] A. Baraldi, E. Alpaydin, *IEEE Trans. Neural Netw.*, **2002,** 13(3), 645-661.
[87] A. Baraldi, E. Alpaydin, *IEEE Trans. Neural Netw.*, **2002,** 13(3), 662-67.
[88] Y. Cao, J. Wu, *Neural netw.*, **2002,** 15, 105-120.
[89] G. A. Carpenter, B.L. Milenova, *Neural Comput.* **2002,** 14, 873-888.
[90] A. Baraldi , E. Alpayd, *IEEE Trans. Neural Netw.*, **2002,** 13(3) 662-677.
[91] S. Grossberg, *Neural Netw.*, **2001,** 14, 587-58.
[92] A. Koufakou, M. Georgiopoulos, , G.C. Anagnostopoulos, T. Kasparis, *Neural Netw.*, **2001,,** 14, 1279–1291.

[93] G.C. Anagnostopoulos, , M. Georgiopoulos, *Int. joint conf. on Neural netw.,* **2001,** 1221–1226.
[94] M.M. Campos, G.A. Carpenter, *Neural Netw.*, **2001,** 14, 505-525.
[95] D. Charalampidis, T. Kaspairs,M. Georgiopoulos, *IEEE Trans. Neural Netw.*, **2001,** 12, 1023-1036.
[96] E. Granger, M.A. Rubin,S. Grossberg, P. Lavoie, *Neural netw.*, **2001,** 14, 325-344.
[97] J. manuel, C. Izquierdo, Y.A. Dimitriadis, E. Gomez Sanchez, J.L. Coronado, *Neural netw.,* **2001,** 14 , 407-425.
[98] D. Muchoney, J. Williamson, *IEEE Trans. Geosci. Remote Sens.* **2001,** 39,101-109.
[99] H.Y. Zhang, C.W. Chan, K.C. Cheung, Y.J. Ye, *Automatica*, **2001,** 37(7)**,** 1065-1070.
[100] G.A. Carpenter, *Trends in Cognitive Sci.*, **2001,** 5(3), 114-118.
[101] S. J Zottile , D. S. Faber, *The Neuroscientist*, **2000,** 6, 25–37.
[102] Z, Yi, P. Heng, P. Fung, *IEEE Trans. Circuits And Systems—Analog And Digital Signal Processing,* **2000,** 47(12).
[103] M. Okatan, S. Grossberg, *Neural Netw.*, **2000,** 13, 699-708.
[104] S. Grossberg, R.W. Paine, *Neural Netw.*, **2000,** 13, 999–1046.
[105] W.D. Ross, S. Grossberg, E. Mingolla, *Neural Netw.*, **2000,** 13, 571–588.
[106] G.A. Carpenter, *Behavioral and Brain Sciences* **2000,** 23, 473-474.
[107] M.M. Campos, G. A. Carpenter, *Neural Processing Letters*, **2000,** 11, 59-78.
[108] E. Mingolla, W. Ross, S. Grossberg, *Neural Netw.*, **1999,** 12, 499–511.
[109] I. Dagher, M.Georgiopoulos, G.L. Heileman, G. Bebis, *IEEE Trans. Neural Netw.*, **1999,** 10, 768-778.
[110] N.H. Lee, *Pattern recognition letters*, **1999,** 20, 131-140.
[111] Y. Oda, K. Kawasaki , M. Morita , H. Korn, H. Matsui, *Nature*, **1998,** 39,182–185.
[112] S.J. Olson, S. Grossberg, *Neural Netw.*, **1998,** 11, 189–208.
[113] S.J. Olson, S. Grossberg, *Neural Netw.*, **1998,** 11, 189–208.
[114] G. Gancarza, S. Grossberg, *Neural Netw.*, **1998,** 11, 1159–1174.
[115] Xin-Hua Song, Philip K Hopke, MaryAnn Bruns, Deborah A Bossio, Kate M Scow, *Chemomet. Intel. Lab. Syst*, *41(2),* **1998,** *161-170.*
[116] L.Y. Cai, H.K. Kwan, *IEEE Trans. Systems, Man, And Cybernetics—Part B,* **1998 ,**28(3) 1998-324.
[117] G.A. Carpenter, N Markuzon, *Neural Netw.*, **1998,** 11 323–336.
[118] D. Heinke , F.H. Hamker, *IEEE Trans. Neural Netw.* **1998,** 9.
[119] Manuel Fern´andez-Delgado and Sen´en Barro Ameneiro, *IEEE Trans. Neural netw.*, **1998** 9, 139-150.
[120] G. A. Carpenter, B.L. Milenova, B.W. Noeske, *Neural Netw.*, **1998,** 11(5), 793-813.
[121] E. Gómez Sánchez, J.A. Gago González, Y.A. Dimitriadis, J.M. Cano Izquierdo, J. López Coronado *Pattern Recog. Let.*, **1998,** 19(3–4), 357-364.
[122] S. Grossberg, N.P. Mcloughlin, *Neural Netw.*, **1997** 10(9) 1583-1605.
[123] N Srinivasa, *IEEE Trans. Sign. Proc.* **1997,** 45 (10), 2533–2550.
[124] A.H. Tan, *IEEE Trans. Neural netw.*, **1997,** 8, 237–250.
[125] G.A. Carpenter, *Neural Netw.*, **1997,** 10(8), 1473-1494.
[126] J. Williamson, *Neural Netw.*, **1996,** 9(5), 881-897.
[127] G. Bartfai, *Neural Netw.*, **1996,** 9, 295-308.
[128] G. Bradski, S. Grossberg, *Neural Netw.*, **1995,** 8 (7/8) 1053-1080 .
[129] S. Grossberg, E. Mingolla, J. Williamson, *Neural Netw.*, **1995,** 8(7/8) 1005-1028.
[130] A. Tan, *Neural Netw.*, **1995,** 8(3), 437-446.
[131] S. Marriott, R.F. Harrison, *Neural Netw.*, **1995,** 8, 619-641.
[132] J. Huang, M. Georgiopoulos, G.L. Heileman, *Neural netw.*, **1995,** 8, 203-213.
[133] G. A. Carpenter, W.D. Ross, *IEEE Trans. Neural Netw.* **1995,** 6(4), 805-818.
[134] G. A. Carpenter, A H Tan, *Connection Science*, **1995 ,***7, 3-27.*
[135] M. A. Rubin, *Neural Netw.*, **1995,** *8(7–8*), *1109-1116.*
[136] T.P. Caudell, , S. D. G. Smith, R. Escobedo, M. Anderson, *Neural Netw.*, **1994,** 7 , 1339-1350.

[137] S. Grossberg, *Neural Netw.*, **1994,** 7(1), 205-207.
[138] M. Georgiopoulos, J. Huang, G.L. Heileman, *Neural Netw.*, **1994,** 7,495-506.
[139] S. Grossberg, *Neural Netw.*, **1993,** 6, 463-483.
[140] S. Grossberg, F Guenther, T D Bullock, D Greve, *Neural Netw.*, **1993,** 6, 43-67.
[141] S. Grossberg, *Mathematical Approaches to Neural Networks ,J.G. Taylor (Editor),* **1993,** *Elsevier Science Publishers B.V.*
[142] G. A. Carpenter, S. Grossberg, J.H. Reynolds, *Neural Netw ,* **1992,** 565-588.
[143] G. A. Carpenter, S. Grossberg, N Markuzon, J.H. Reynolds, D.B. Rosen, *IEEE Trans. Neural netw.*, **1992,** 3, 698-712.
[144] G. A. Carpenter, S. Grossberg, *Neural Netw.*, **1991,** 4, 759-771.
[145] S. Grossberg, D. Somers, *Neural Netw.*, **1991,** 4(4) 453-466.
[146] P Gaudiano, S. Grossberg, *Neural Netw.*, **1991,** 4, 147-183.
[147] G. A. Carpenter, S. Grossberg, J.H. Reynolds, *Neural Netw.*, **1991,** 4, 565–588.
[148] G. A. Carpenter, S. Grossberg, J.H. Reynolds, *Neural Netw.*, **1991,** 4 , 565-588.
[149] S.W. Tsay, R W Newcomb, *IEEE Trans. Neural netw.*, **1991,** 2214.
[150] G. A. Carpenter, S. Grossberg, D.B. Rosen, *Neural Netw.*, **1991,** 4, 759–771.
[151] G. A. Carpenter, S. Grossberg, D.B. Rosen, *Neural Netw ,* **1991,** 4, 493-504.
[152] G. A. Carpenter, S. Grossberg, *Neural Netw.*, **1990,** 3, 129-152.
[153] G.A. Carpenter, *Neural Netw.*, **1989,** 2(4), 243-257.
[154] S. Grossberg , J A Marshall, *Neural Netw.*, **1989,** 2, 29-51.
[155] S. Grossberg, M E. Rudd, *Neural Netw.*, **1989,** 2, 421- 450.
[156] G. A. Carpenter, S. Grossberg, C Mehanian, *Neural Netw.*, **1989,** 2, 169-181.
[157] G. A. Carpenter, S. Grossberg, C Mehanian, *Neural Netw.*, **1989,** 2, 169-181.
[158] J.H. Gennari, P. Langley, D.H. Fisher, A.K. Bunke, *Art. Intel* **1989,** 40, 11–60.
[159] S. Grossberg, *Neural Netw.*, **1988,** 1, 17-61.
[160] G. A. Carpenter, S. Grossberg, *Computer,* **1988,** 21, 77-87.
[161] G. A. Carpenter, S Grossberg, *Appl. Optics* **1987,** 26 (23), 4919–4930.
[162] G. A. Carpenter, S. Grossberg, *Computer Vision, Graphics, and Image Processing,* **1987,** 37,54–115.
[163] J. R Quinlan, *Machine Learning,* **1986,** 1(1) , 81-106.
[164] S. Grossberg, *Studies of mind and brain, pp. Boston, MA: Reidel Press,* **1982,** 229-295,.
[165] G.A. Carpenter, *In S. Grossberg (Ed., Mathematical Psychology and Psychophysiology, AMS/SIAM Symposium Series,* **1981,** 13, *49-90.*
[166] A. Carpenter, *SIAM J. Applied Math.*, **1979,** 36, 334-372.
[167] G. A. Carpenter, V Knapp, *J. Mathematical Biology,* **1978,** 6, 305-316.
[168] G. A. Carpenter, **1978, 3**, 58-76.
[169] G. A. Carpenter, *J. Mathematical Analysis and Applications* **1977** , **58**(1), 152-173.
[170] G. A. Carpenter, *J. Differential Equations,* **1977, 23**(3), 335-367.
[171] S. Grossberg , *Biol. Cybernet.* **1976,** 23,121–134.
[172] G. A. Carpenter, *In P. Hilton (Ed.), Structural Stability, The Theory of Catastrophes, and Applications in the Sciences, Springer Series Lecture Notes in Mathematics,* **1976,** 525,58-76.
[173] G. A. Carpenter, *Newsletter* **1976,** 42-44.
[174] S. Grossberg, *Int. Rev. Neurobiology,* **1975,** 18, 263-327.
[175] S. Grossberg, *In R. Rosen and F. Snell (Eds., Progress in theoretical biology.* **1974,** *New York: Academic Press.* .
[176] S. Grossberg, *Studies in Applied Math.*, **1973,** 52,217-257 .
[177] S. Grossberg, *Mathematical Bioscience,* **1972,** 15, 253-28.
[178] S. Grossberg, *Mathematical Bioscience,* **1972,** 15, 39-67.
[179] S. Grossberg, *Kybernetik,* **1972,** 10, 49-57.
[180] O. L. Mangasarian, *IEEE Trans. Inf. Theory,* **1968,** 14, 801–807.
[181] S. Grossberg, *Undergraduate results,***1957**.

[182] H.P. Henningsen, *Neurocomput.*, **2010,** 74,57–75.

Omnimetrics-Chemometrics

[183] B.R. Kowalski, Anal.chem., **1980**,52, 112R-122R ; **1982**,54, 232R-243R ; **1986**,58, 294R-314R.(b) M.F. Delancy, Anal.chem.**1984**,56, 261R-277R; (c)S.D. Brown et al. Anal.chem. **1988,** 60,252R-273R; **1990,** 62,84R-101R; **1992,** 64, 22R-49R; **1994,** 66,315R-359R; **1996,** 68, 21R-61R. (d) B.K. Lavine, J. Workmann Jr, Anal. Chem. **1998,** 70, 209R-228R; **2000,,** 72, 91R-97R**; 2002,** 74, 2763-2770 ; **2004,** 76, 3365-3372; **2006,** 78, 4137; **2008,** *80,* 4519–4531 ; **2010,** *82,* 4699–4711; **2013**, 85, 705−714.

[184] Windows on chemo metrics, The Royal Society of Chemistry U.K.

[185] B.R. Kowalski Pattern recognition in chemistry (Handbook of statistics)**1982, Amazon.com.** (b) M.A. Sharaf, D.L.. Illman, B.R. Kowalski, Chemometrics, **1986**,

[186] B.R. Kowalski, Non-Linear Calibration Using Projection Pursuit Regression, Application to an Array of Ion Selective Electrodes,1988, Amazon.com.

[187] S.D. Brown, R. Tauler, B. Walczak Comprehensive Chemometrics, Elsevier B.V.

[188] D.L. Massart et al. Handbook of Chemometrics and qualimetrics, **1997**, Elsevier Science B.V., The Netherlands.

[189] A.Holtona,Vaishnavi Vijayakumara,b,1 andNora Khaldia,c Trends in Food Science & Technology, **2013**, 34, 5-17.

[190] G.Nageswara Rao, K.Sambasiva Rao, R.Sambasiva Rao and LH. Abdel Rehman, *J**.** Indian Chem.Chem*, **2000**, 17, 48.(b) K RamaKrishna, R Sambasiva Rao et. al., Journal of Applicable Chemistry (Online), **2012**, 1,109-124; **2013**, 2, 698-713, 1007-1034, 1413-1458; **2014**, 3(1), 06-029; **2014**, 3(2), 449-492, 834884 .(c) M.V. Subaba Rao, R. Sambasiva Rao et. al., Asian J Chem., **2010**, 22, 5937-5950; Research J. Science and Tech., 2013, 5(4): 412-420.(d) K.V. Bapanaiah, R. Sambasiva Rao, *Science Reporter*, **1977**, 12, 315

Cancer research

[191] R. Jagsi, CA Cancer J. Clin., **2014,** 64, 135–152.

[192] E. Ward, C. DeSantis, A. Robbins, B. Kohler, A. Jemal, CA Cancer J. Clin., **2014,** 64, 83-103.

[193] R. Siegel, C. DeSantis, A. Jemal, CA Cancer J. Clin., **2014,** 64, 104–117.

[194] J. Cullen, CA Cancer J. Clin., **2014,** 64, 2 ,79-82.

[195] M. K. Barton, CA Cancer J. Clin., **2014,** 64, 2 ,75-78.

[196] C. L. Wolfgang, , J. M. Herman, D. A. Laheru, A. P. Klein, M. A. Erdek, E. K. Fishman, R.H. Hruban, CA Cancer J. Clin., **2013,** 63, 318–348.

[197] M.D. Galsky, A. C. Small, C. Tsao, W.K. Oh, CA Cancer J. Clin., **2012,** 62, 299-308.

[198] R. Siegel, C. DeSantis, K. Virgo, K. Stein, A. Mariotto, T. Smith, D. Cooper, T. Gansler, C. Lerro, S. Fedewa, C. Lin, C. Leach, R. S. Cannady, H.Cho, S. Scoppa, M. Hachey, R. Kirch, A. Jemal, E. Ward, CA Cancer J. Clin., **2012,** 62, 220-241.

[199] L. H. Kushi, C. Doyle, M. McCullough, C. L. Rock, W. Demark-Wahnefried, E.V. Bandera, S. Gapstur, A. V. Patel, K. Andrews, T. Gansler, CA Cancer J. Clin., **2012,** 62, 30–67.

[200] K. K. Turaga, L. K. Kvols, CA Cancer J. Clin., **2011,** 61, 113-132.

[201] A. Jemal, F.Bray, M. M. Center, J. Ferlay, E. Ward, D. Forman, CA Cancer J. Clin., **2011,** 61, 69-90.

[202] Blood. **1997,** 89, 3909-3918.

[203] S. A Rosenberg, Blood **1994,** 84, 1359-1360.

[204] N. L. Harris, et al., Blood **1994,** 84, 1361-1392.

[205] D. M Reed, Johns Hopkins Hosp. Rep. **1902,** 10, 133-196.

[206] C.Sternberg, Heilk **1898,** 19, 21-90.

[207] T.Hodgkin, Med.-Chir. Trans. **1832,** 17, 68-114.