Artificial Intelligence and Machine Learning Club (AIML C), IIT Delhi, 01 February 2021.

Pattern Recognition, Machine Intelligence to Data Science: Evolution and Challenges

A journey in brief over 45 years

Sankar K. Pal
National Science Chair
Indian Statistical Institute
Calcutta
http://www.isical.ac.in/~sankar

Contents

- Journey started in March 1975
- Pattern Recognition and Data Mining
- Fuzzy Sets and Uncertainty Analysis
- Crisis: Integrations and Soft Computing & MI
- Rough Sets and Granular Computing
  - Relevance to Data Mining
- Generalized Rough Sets & Entropy: Uncertainty modeling
- Video Processing: Object extraction and tracking
- Bioinformatics: miRNA selection
- Challenging issues: CTP, NC, BDA, Deep Learning
- Evolution of Data Science + Caution

Journey started in 1975 -

- I joined ISI on March 01, 1975 as a CSIR-SRF for PhD (worked over 45 yrs.)

Resource

ISI Library (with no IEEE Journal)

- K.S. Fu, (Ed.), *Sequential Methods in Pattern Recognition and Machine Learning*, Academic Press, London, 1968
- K.S. Fu (Eds.), *Syntactic Methods in Pattern Recognition*, Academic, London 1974
- G.S. Sebestyen, *Decision Making Processes in Pattern Recognition*, The Macmillan Co. N.Y., 1972 - (**PhD Thesis**)
- A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets: Fundamental Theoretical Elements*, vol. 1, Academic Press, N.Y.,1975.

Purchased from CSIR Cont. Grant

- L. A. Zadeh, K. S. Fu, K. Tanaka, and M. Shimura (Eds.), *Fuzzy Sets and Their Application to Cognitive and Decision Processes*, Academic, London, 1975

INRAPHEL, Calcutta University Library

- L.A. Zadeh, Fuzzy sets, *Inform. Control*, 8, 338-353, 1965
- L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Trans. Syst., Man, Cyberns.*, SMC-3, 28-44, Jan. 1973

Pattern Recognition System (PRS)

Measurement $\rightarrow$ Feature $\rightarrow$ Decision

Space Space Space

– Uncertainties arise from deficiencies of information available from a situation

- Deficiencies may result from incomplete, imprecise, ill-defined, not fully reliable, vague, contradictory information in various stages of a PRS

PR Tasks & Challenges

- Classification: Sampled data (incomplete information) is given about the pattern space And the Challenge is to estimate the unknown regions of the pattern space based on the sampled data $\Rightarrow$ Abstraction + Generalization

  (Supervised Learning)

- Clustering: Entire data is given And Challenge is to partition it into meaningful regions. No. of regions may be known or unknown

  (Unsupervised Learning)

Fuzzy Sets:
Flexibility & Uncertainty Analysis

(Lotfi Zadeh, *Inform. Contro1*, 1965)

(~ 96,000 citations)

Prof. Lotfi A Zadeh, UC, Berkeley who first explained the theory of fuzzy sets passed away on Sep 06, 2017 at the age of 96+.

- Fuzzy Sets are nothing but Membership Functions
- Membership Function: Context Dependent

↓

- Concept of Flexibility & Uncertainty Analysis (overlapping data/ concept/ regions)
  - Probability $p(x)$ *vs*. Membership $\mu(x)$

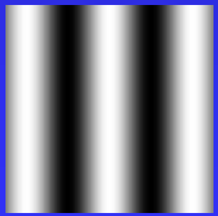
Challenge

- Bringing out the root relation between -

Abstract concept of *Fuzzy Sets* & Tasks of Pattern Recognition and Image Processing

- Notion of multi-class belonging of a pattern

- A grayscale image with sinusoidal gray value gradation ⇨ *Fuzzy (ill-defined)* boundaries, regions, edges, corners relation, properties



Note:

- E. Ruspini (SRI), "A new approach to clustering", *Inform. Control*, 15, 22-32, 1969
  - Clustering should be fuzzy, NOT crisp
  - Patterns may have origin from > 1 class

- J.C. Dunn and J.C. Bezdek (fuzzy ISODATA 1974, fuzzy c-means 1974) – initiated a new direction to fuzzy cluster analysis

- Pal and Dutta Majumder (ISI, 1975) IEEE T-SMC 1977 (FS in speech recognition)

- J.M.B. Prewitt (1970) – Image segmentation should be fuzzy sub-sets of image

- A. Rosenfeld and his group (UMD, College Park) 1979 (extending dig. geometry to fuzzy subsets)

- Pal + King (Imperial College, London 1979) Electron Lett 1980 (enhanc)

Example Applications:

- Speech recognition
- Medical image (MRI, X-rays)
- Remote sensing image (Defence applications)
- Natural language processing

Crisis in FL Research

- FS research got stuck little in mid ‘80s (as in many other areas)
- Determining membership functions (criticism)

- Japanese products on FL Control
- Re-appearance/Revival of ANN & Learning (1987/1988)
- Introduction of Genetic Algorithms (searching/optimiz)

- FL research flourished again at a higher gear
- Funding agencies (India + abroad) came forward
- Conferences held in conjunction with other paradigms
- IEEE Transactions and CI Societies + other journals $

In late eighties scientists thought –

Why NOT Integrations ?

Fuzzy Logic + ANN

ANN + GA

Fuzzy logic + GA

Fuzzy Logic + ANN + GA

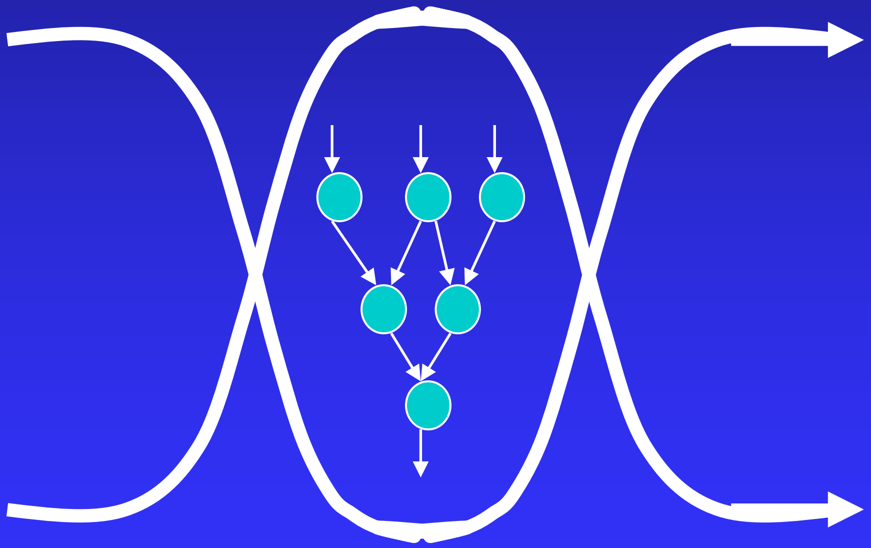
Fuzzy Logic + ANN + GA + Rough Set

Neuro-fuzzy hybridization was the first and most visible integration realized

Why N-F Fusion ?

Fuzzy Set theoretic models try to mimic human reasoning and the capability of handling uncertainty – (SW)

Neural Network models attempt to emulate architecture and information representation scheme of human brain – (HW)



NEURO-FUZZY Computing

(for More Intelligent System)

Major Characteristics of ANN

- **Adaptivity:** Adjusts to change in environment (new data/ information)
- **Speed:** via massive parallelism
- **Fault tolerance:** to missing, confusing, noisy data
- **Ruggedness:** to failure of components (nodes/links)
- **Optimality**: as regards to error rates in classification

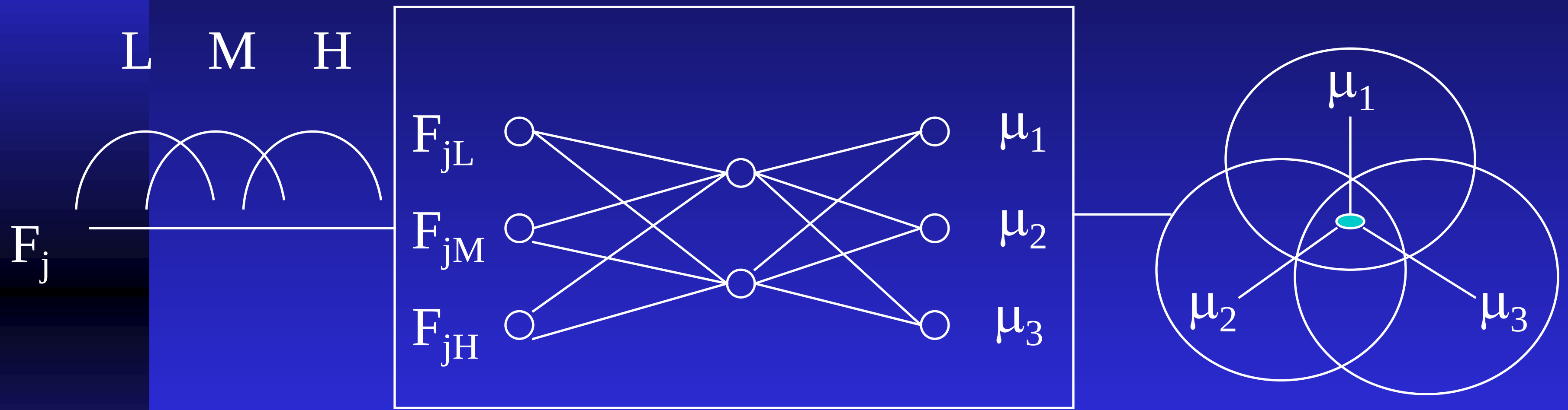
- Learns from Examples (If Input is A then Output is B)
- Encodes the Input-Output relation, however complicated, into network parameters w_is

Example:

IEEE Trans. Neural Networks, 3, 683-697, 1992

1994 Outstanding Paper Award - IEEE Neural Net Council

Fusion: Nonlinear boundary + Uncertainty handling



- Handling imprecise input
- Handling uncertainties arising from overlapping classes
- Back-propagated errors are assigned appropriate weightage depending on μ values at corresponding output nodes

➢ Can handle linguistic input in addition to those by conv. MLP

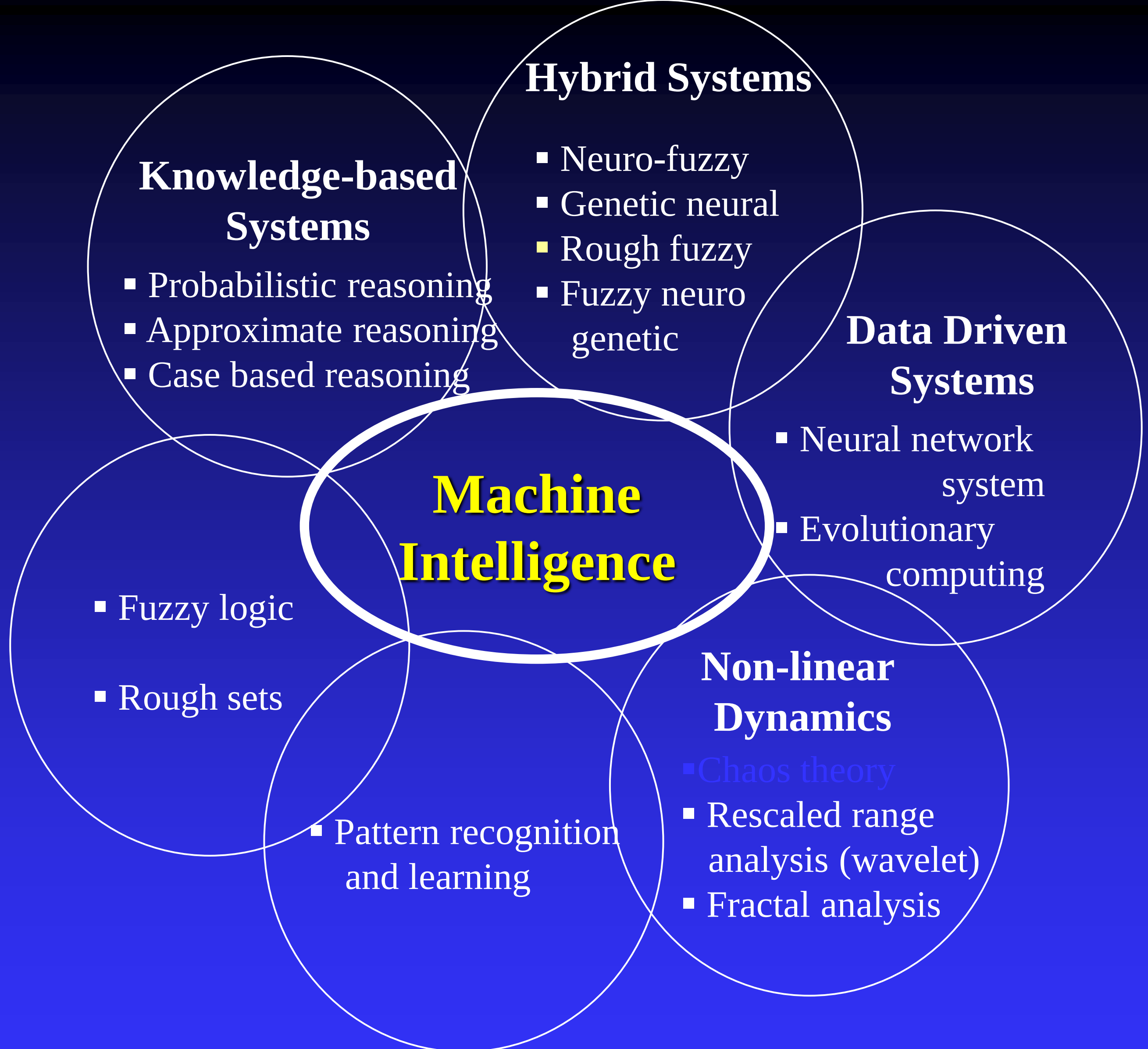
Speech Recognition (# vowel classes 6)

(Classification Score %)

- Bayes' classifier: 79.2
- Neural Nets (nodes in each hidden = 10, 20; five layers)
  - Testing (Training)
  - Conventional: 84.6, 82.2 (86.0, 87.6)
  - Hard linguistic input: 72.5, 70.2 (77.7, 71.5)
  - Fuzzy version: **84.2, 83.6 (92.2, 92.2)** – with 20% linguistic patterns

Accordingly defined –

Machine Intelligence **(1993)**



Machine Intelligence: A core concept for grouping **various advanced technologies** with **Pattern Recognition and Learning**

IAS are physical embodiments of Machine Intelligence

Soft Computing

- While different challenges of synergistic integrations between FL, ANN and GAs were being addressed with application specific merits, Zadeh defined the concept of Soft Computing consolidating them under one umbrela.

- L.A. Zadeh, “Fuzzy logic, neural networks, and soft computing”, *Comm. ACM*, 37, 77-84, 1994

SOFT COMPUTING
(L. A. Zadeh)

Aim :

- To exploit the tolerance for imprecision uncertainty, approximate reasoning and partial truth to achieve **tractability, robustness, low solution cost,** and **close resemblance** *with human like decision making*

- To find an approximate solution to an imprecisely/precisely formulated problem.

- High precision carries a high cost

Roles of Principal Constituents of SC

FL : Algorithms for dealing with imprecision and uncertainty

NC : Machinery for learning and curve fitting

GA : Algorithms for search and optimization

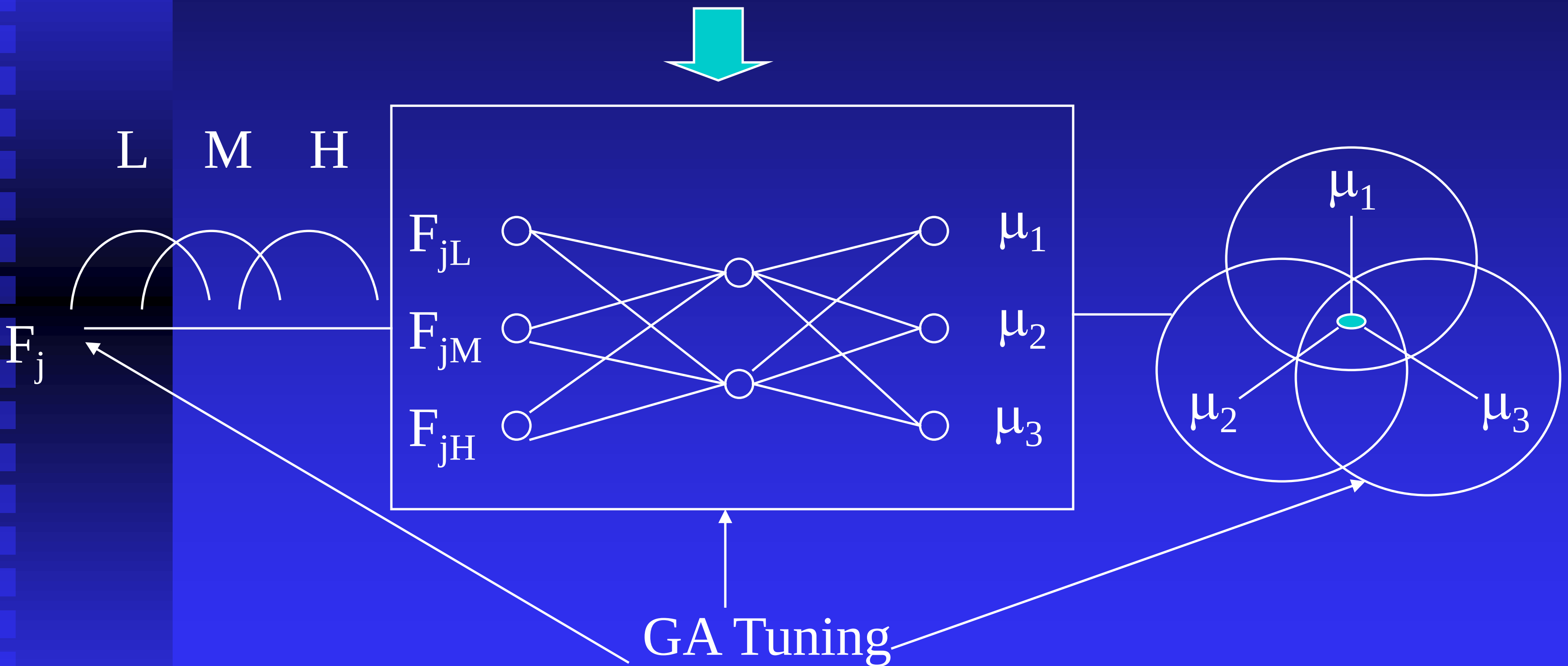
RS → Handling uncertainty arising from granularity in the domain

- Within Soft Computing FL, NC, GA, RS are **Complementary** rather than **Competitive** **$**

IEEE Trans. Neural Networks, 9, 1203-1216, 1998

Example: Synergistic Integration of ANN, FL, GA and RS

Incorporate Domain Knowledge using Rough Sets



X X | 0 0 0 | X X

0 0 | X X X | 00

IEEE Trans. Knowledge Data Engg., 15(1), 14-25, 2003

Merits

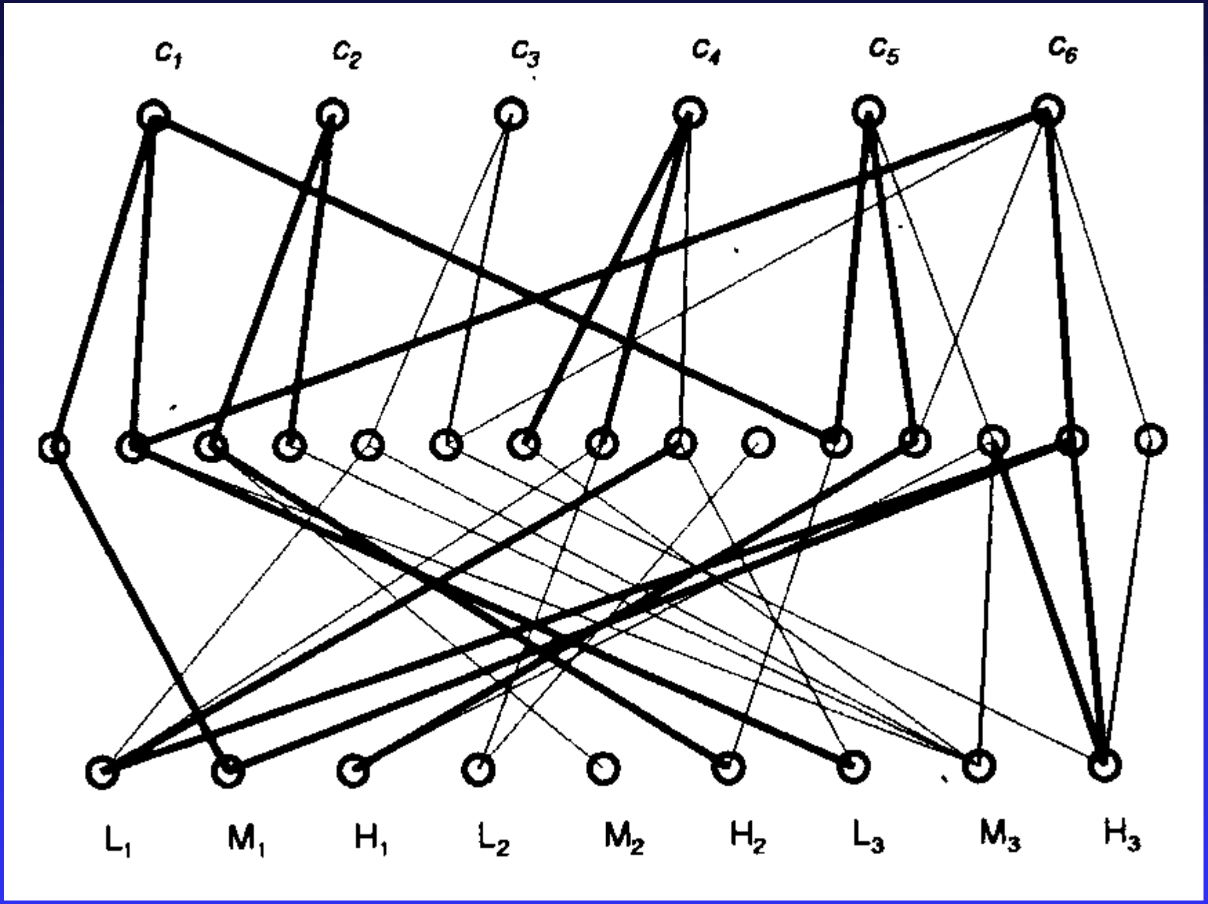
Enhances

- Classification Performance
- Training time
- Network compactness

Generates Rules of

- Higher accuracy
- Smaller size
- Less confusion

Example of Compact Network



Network Connectivity obtained for 6-class vowel recognition using Modular Rough Fuzzy MLP *(IEEE Trans. Knowledge Data Engg., 15(1), 14-25, 2003)*

Around 2000, **Data Mining** became a buzz word (primarily for www and Genome project producing large & heterogeneous data)

Pattern Recognition and *Machine Learning* principles applied to a very large (both in size and dimension) heterogeneous database

≡ *Data Mining*

Data Mining + Knowledge Interpretation

≡ Knowledge Discovery

⇩

Process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data

Rough Sets and Granular Computing

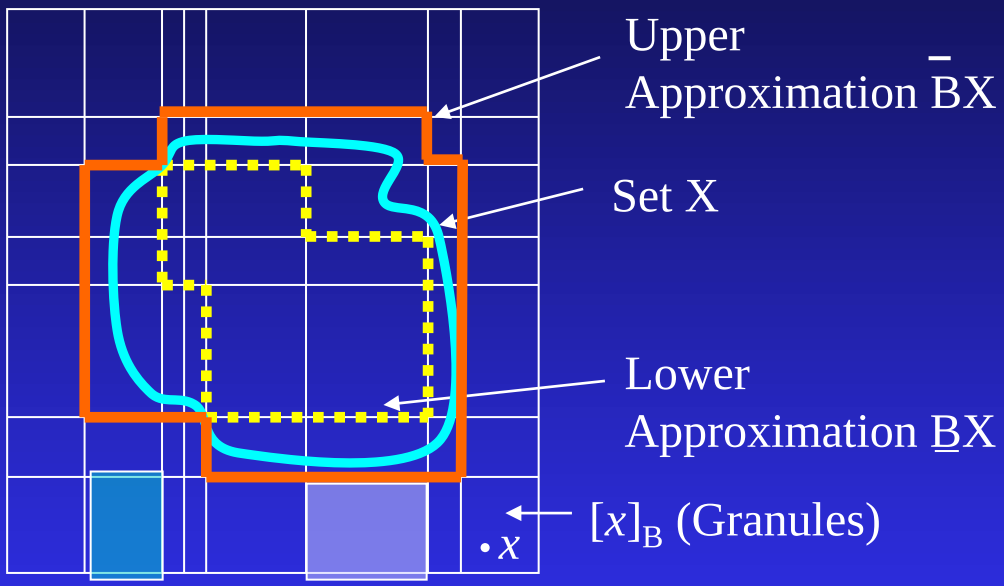
RS – Crisp set defined over a Crisp granulated domain

NASA: RS, GA - PPW, MB, students

Rough Sets

Z. Pawlak 1982, Int. J. Comp. Inf. Sci.

$\Omega_B \subseteq U$



$[x]_B$ = set of all points belonging to the same granule as of the point x in feature space Ω_B

⇨ $[x]_B$ is the set of all points which are *indiscernible* with point x in terms of feature subset B

Rough Sets are Crisp Sets, but with rough description

- The vague definition of X in U (in terms of lower and upper approxs.) signifies the incompleteness of knowledge about U

Minimize Incompleteness to Make Decision

Rough Sets

Uncertainty Handling

(Using lower & upper approximations)

Granular Computing

(Using information granules)

Two Important Characteristics

Granular Computing (GrC): An information processing paradigm

- that works with the process of *information granulation/ abstraction*, and

- where computation is performed using *information granules* and not the data points (objects)

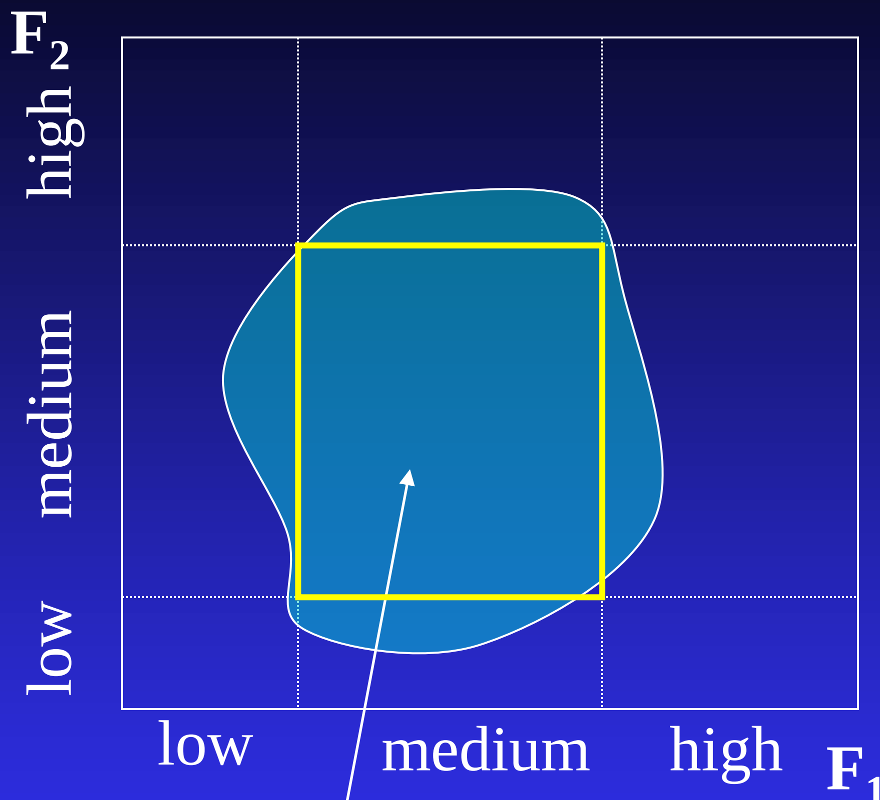
- Information compression
- Computational gain

➢ Suitable for Mining Large Data

Concept of -

f- Information Granules using Rough Rules

Information Granules and Rough Set Theoretic Rules



- Elongated cluster
- Case generation with reduced dim
- Variable dim reduction

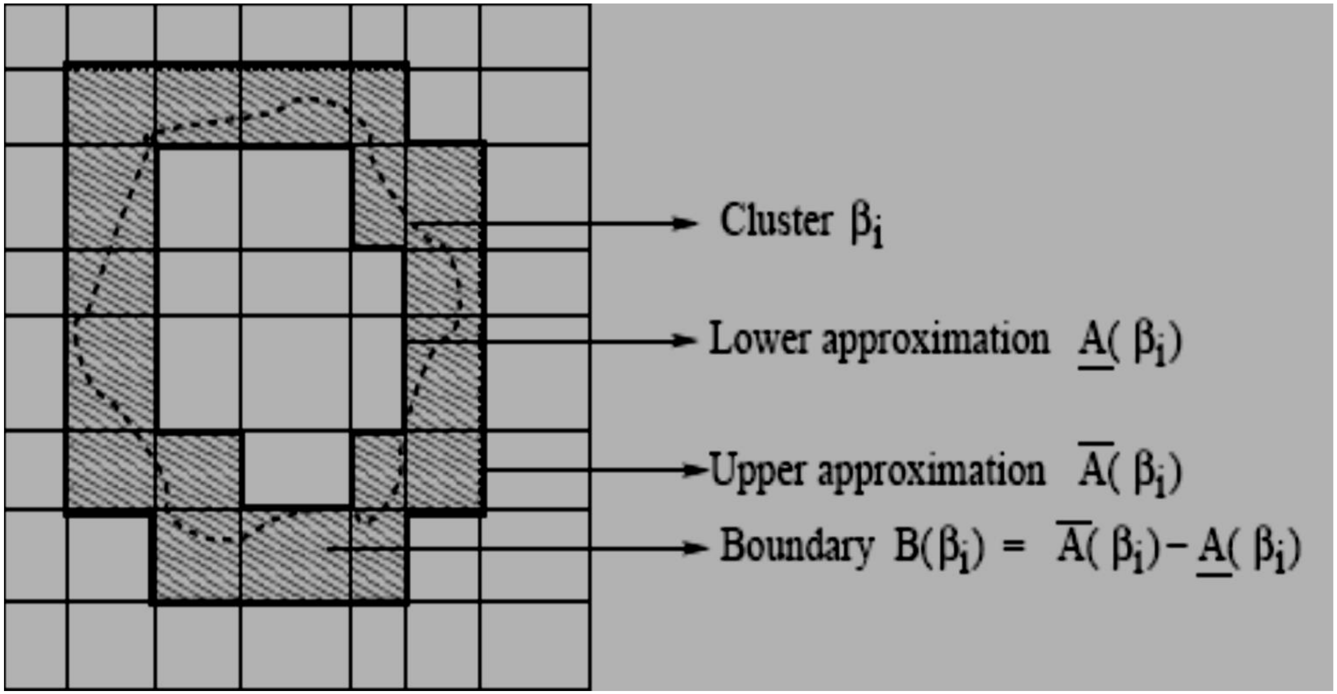
$$\text{Rule} \leftarrow M_1 \wedge M_2$$

- Rule provides crude description of the class using granule □

What is
Lower & Upper Approximations of Clusters
and associated Uncertainty Modelling?

IEEE Trans. Syst., Man and Cyberns. Part B, 37(6), 1529-1540, 2007

Cluster definition using rough lower & upper approx



Roughness in β :
1 - |lower|/ |upper|

- Sets and Granules can *either or both* be fuzzy (in real life)
  - Lower and upper approximate regions could be fuzzy *(m.func)*

➡ Generalized Rough Sets – Stronger model of uncertainty handling

(uncertainty due to overlapping regions + granularity in domain)

Applications of Granular Mining (& Uncertainty Modelling)

- Video tracking (Image analysis)
- miRNA selection (Bioinformatics)
- Link prediction, Comm. detection (Social Net)
- Neural learning and network formation

Role of -

- Granules (window, quad-tree, arbitrary shape)
- Lower-upper approximation

Handling Overlapping/ Occlusion

(Unsupervised Video Tracking)

- Spatio-color neighborhood granules in τ-space used to design NRS Filter
- **NRS Filter:** Rough estimate (US) of location and color model of objects
- These information are used to model the nature of variation in size, speed/ direction of objects so as to locate objects in next frame
- Object regions with min. roughness & intuitionist entropy is *Tracked*
- Handles partial or total occlusion

Inputs-Outputs of NRS Filter: Design

- Union of changed regions among current to all P previous frames **(in granular level)**

  $\delta_P = \cup\delta_p$: p = t-1, t-2, ,..., t-P Input

- P changed regions between consecutive frames

  δ_i, i = t, t-1, …, t-P Input

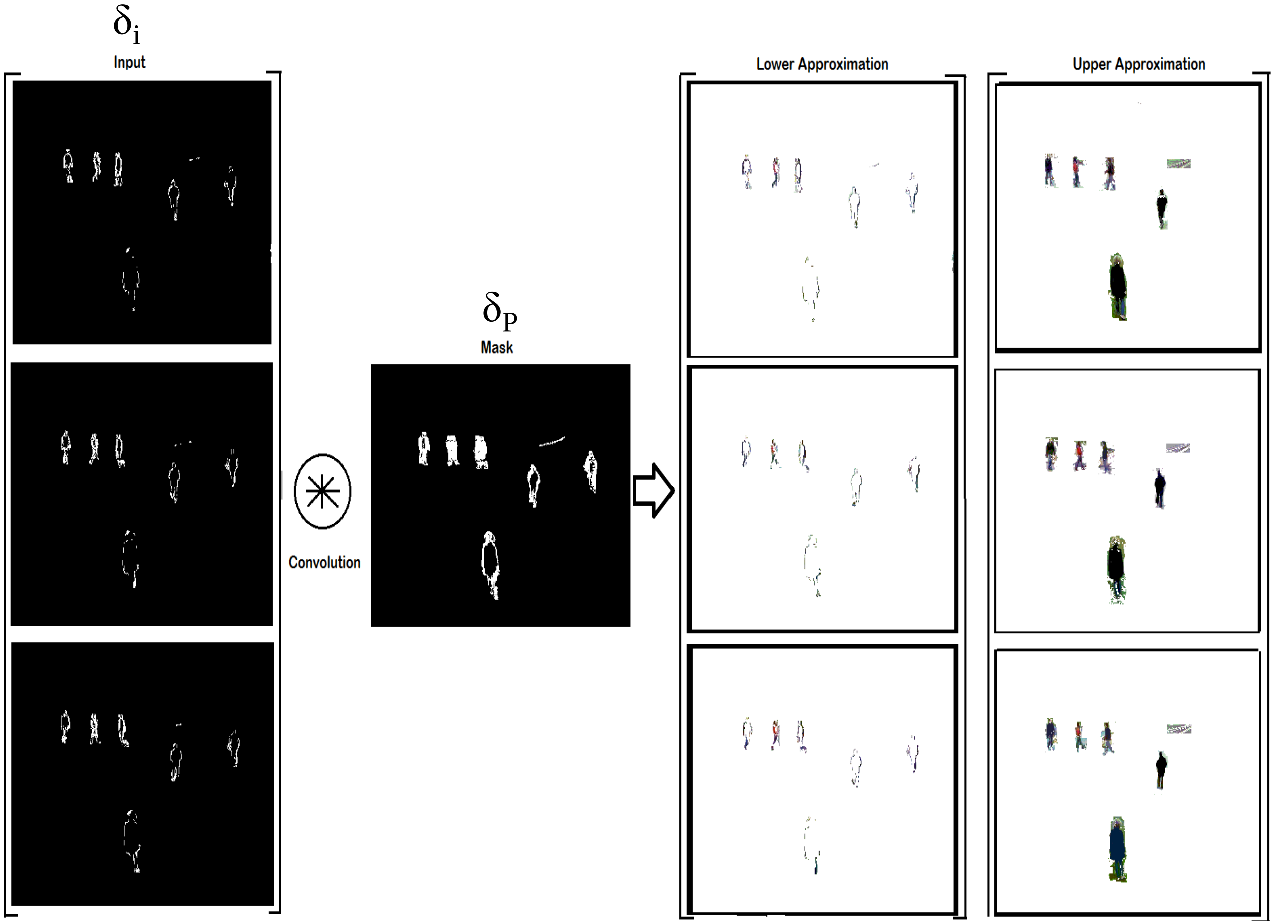
❖ **Input and Output Relation**

$$\{\delta_i : i = 1, \ldots, P\} * \delta_P = \{\underline{O_c} : c = 1, \ldots, P\}$$

$$\{\overline{O_c} : c = 1, \ldots, P\}$$

- ❖ A "one to P-point convolution" takes place in the filter to result in a P-point matrix.
- ❖ There will be 2 P-point **Output** matrices representing 2 approximated (lower and upper) decision spaces of objects defined over the filter
- ❖ $*$: Convolution operators used are ∪ and ∩

NRS Filter (P = 3)



Convolution results in two 3 × 1 matrices: Two types of approximated regions of the objects as the output of NRS filter

AVSS Tracked

Frames per sec = 15, P = 6

Cam 132 Tracked

Frames per sec = 15, P = 6

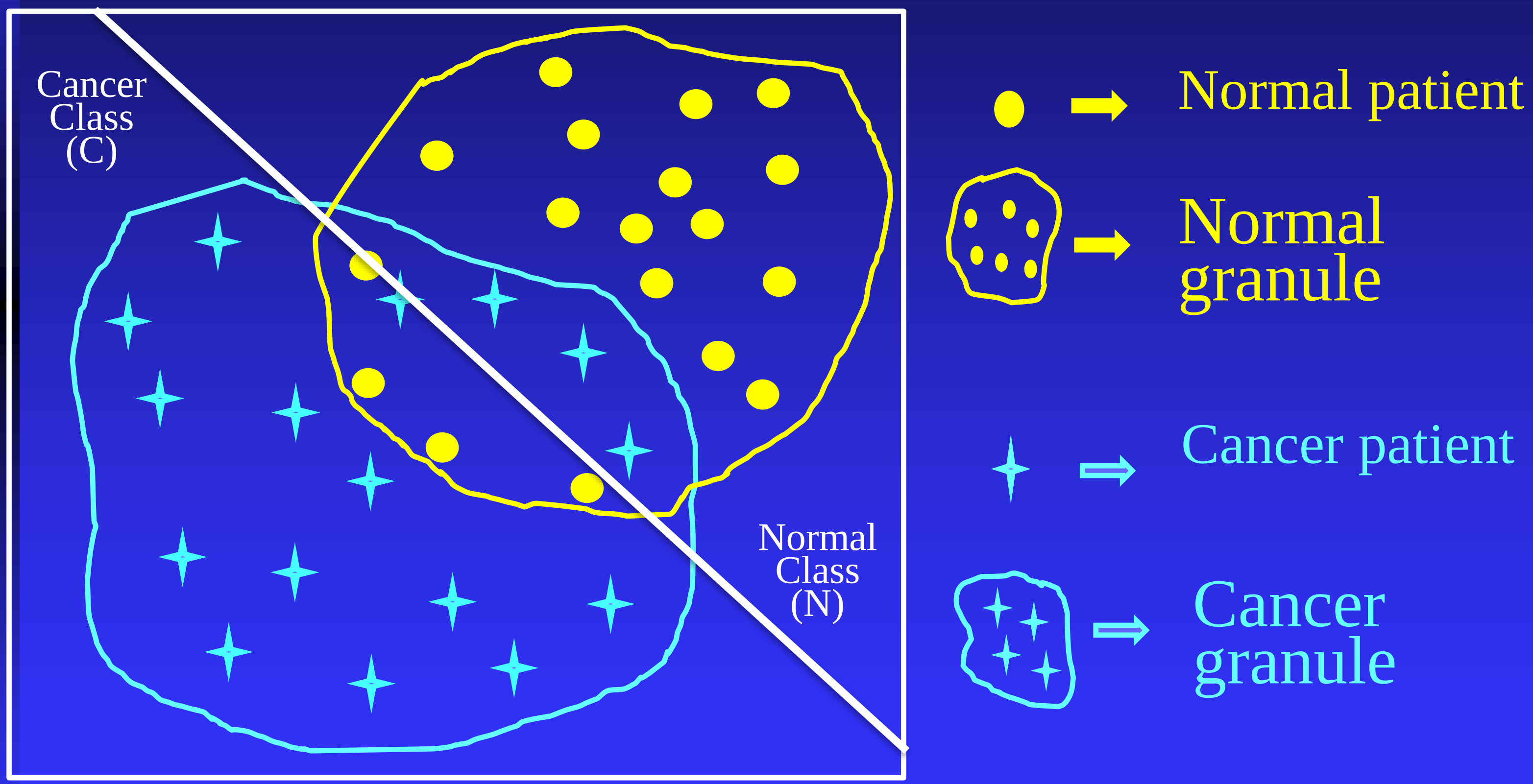
Example:

miRNA Ranking in Cancer Detection

- Small sample, Large dimension
- Set is crisp (C or N) & Granules are fuzzy: *Fuzzy-rough entropy*
- *Fuzzy Lower Approx*. of N & C classes: Degree of being sure to be in N & C classes ⟹ Used to find prob. (relative frequency) of definite and doubtful regions for entropy computation

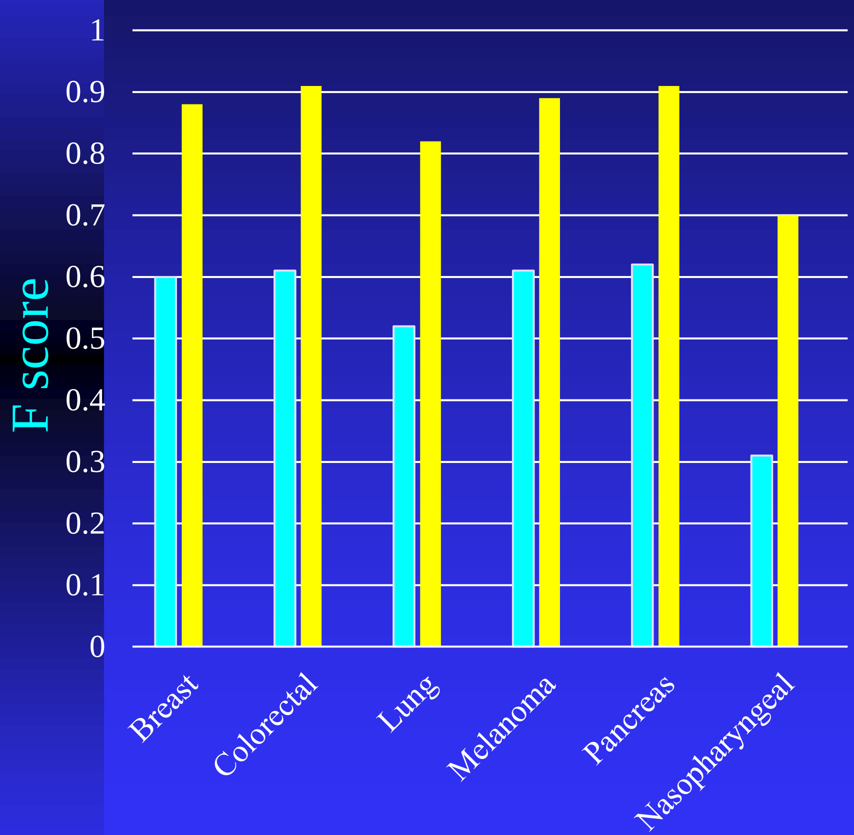
- *Entropy* minimization implies higher *Relevance* of a miRNA
- *Top 1% miRNAs provide significant improvement over entire set in terms of F-score*

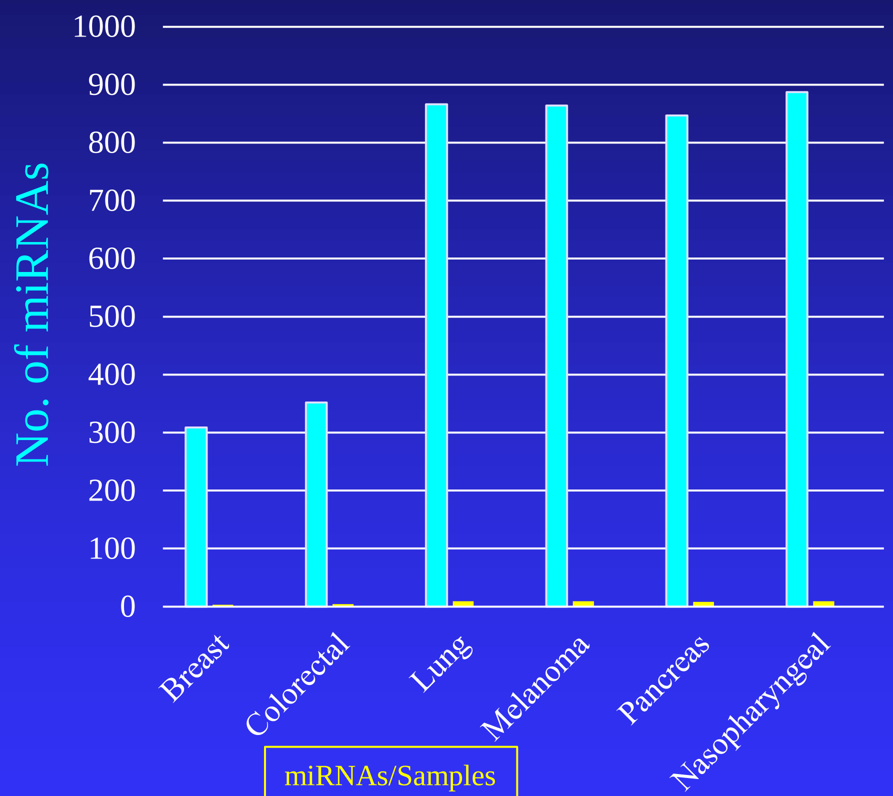
Crisp Classes & Fuzzy Granules of Patients



IEEE/ACM Trans. Comput. Biology & Bioinformatics, 15(2), 659-672, 2018

Results: Relevance (All 1% selected)





miRNAs/Samples
309/98
352/66
866/36
864/57
847/158
887/50

- Classification using SVM

Summary

- Granular Views Over 45 years in ISI
- Different Machine Learning tools
- Significance in Video Analytics + BI – Examples
  - Identifying drug resistant miRNAs *IEEE/ACM Trans. Comput. Biology & Bioinformatics, 2019, DOI: 10.1109/TCBB.2019.2933205*
  - Video conceptualization *Inform Sci., 543, 488-503, 2021.*
  - Social link prediction *IEEE Trans. Computational Social Systems, 5 (3), 841-853, 2018*
  - Neural network generation *IEEE Trans. Neural Networks and Learning Systems,* 27(9), 1890-1906, 2016 (use lower inform granules, form basic net, then grow by upper set)

Where are these leading to ?

(CTP, NC, BDA)

Relevance to BIG Data Analytics

- Uncertainty handling and Granular mining points of view (Covid-19 detection & screening from X-ray or CT-scan images)
- Granulated Deep learning – reducing comp time

• Instead of scanning the entire image pixel by pixel in the Convolution layer of deep learning, we jump over the granules only. ⇨ For a 32x32 image with N granules, sliding the filter is done only N times instead of over 32x32 pixels where N << 32x32.

• Hence a *significant speed up* is observed, compromising some accuracy

Neural Computing and Applications, 32(21), 16533-16548, 2020

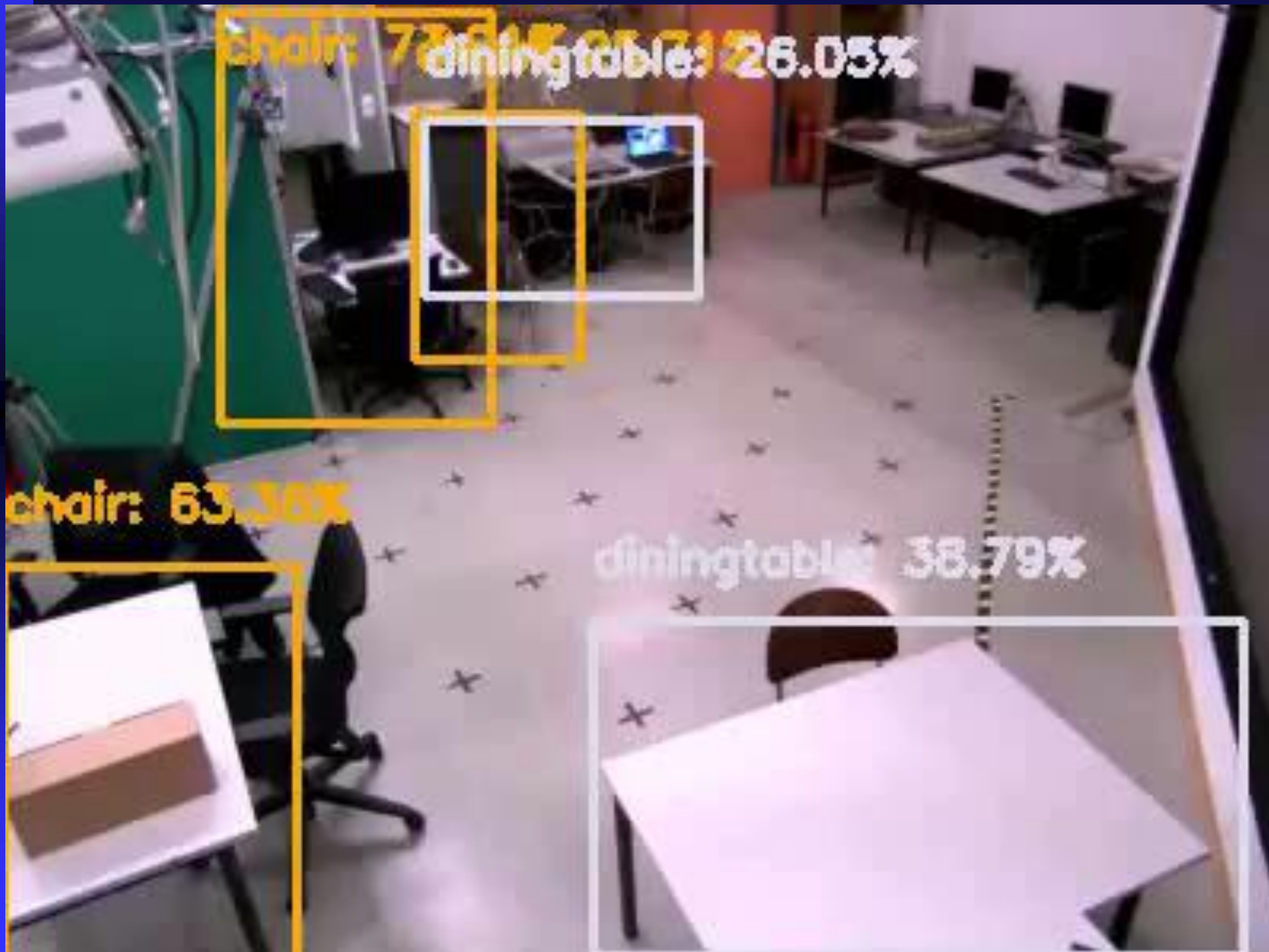
Neural Comp. and Appl., 32(21), 16533-16548, 2020

Comparative Result

- **Time and accuracy comparison for object recognition and tracking** on Cam 131 Sequence of ICG Lab data set (Chap scenario)
- Granulated Deep Learning (using 3x3 granules, rectangular granules by quad-tree decomposition, and arbitrary shaped granules) ***Vs***. Deep Learning without Granulation

Method	Speed	Track	Accuracy of detection	Processor
Granulated Deep Learning using 3x3 granules	2.2 fps	74.6%	62.11%	CPU
Granulated Deep Learning using rectangular granules	2 fps	80.1%	67.11%	CPU
Granulated Deep Learning using arbitrary shaped granules	1.89 fps	81.67%	68.56%	CPU
Deep learning without granulation	1.6 fps	82.25%	70.2%	CPU

Results **with arbitrary shaped granules** on Cam 131 Sequence of ICG Lab data (Changing appearance (chap) scenario)



Granulated RCNN for Multi-object detection

- Object detection in a Region based CNN (RCNN) has 2 stages: object localization (extracting RoIs) and classification.
- G-RCNN (developed on AlexNet architecture) is an improved version of Fast RCNN and Faster RCNN for extracting RoIs by incorporating Spatio-temporal granulation in a deep CNN.
- Compared to Fast and Faster RCNNs, G-RCNN uses
  - (i) only granules formed over the pooling feature map (instead of its all feature values) in defining RoIs, (ii) only the positive RoIs (i.e., RoIs denoting only objects, instead of the whole RoI-map) as input to FC1 while retraining GRCNN, (iii) only the f_t-regions corresponding to RoIs (instead of the entire f_t feature map) for performing object classification, and (iv) videos directly as input, rather than static images. **All these improve the real time detection speed and accuracy.**

Evolution

PR (1960's) → IP (1970's) → AI+ML+Expert Systems (1980's) → Knowledge Based System (1990's) → DM (2000's) → Big Data (2010) → Data (driven) Science

- New approaches for different tasks of PR to handle varying nature of data and decision-making
  (Feature Selection, *IEEE Trans. Pattern Anal. Machine Intell., 24(3), 301-312, 2002*)

- New terms & technologies coined – BIG Hope -- (DL)

- **Caution**

Acknowledgement

- Students and younger colleagues/ collaborators
- National Science Chair, SERB-DST, GoI

Thank You!!

Please Stay Safe and Keep Well

Giant Panda from Chengdu: Life is so…o good with bamboo shoots