UNIVERSITY OF SARAJEVO
FACULTY OF ELECTRICAL ENGINEERING

MODELING OF FUZZY SYSTEM FOR PROSTATE CANCER RISK ESTIMATION USING COEVOLUTIONARY ALGORITHM

Researchers:
PhD Zikrija Avdagic
MScEE Samir Omanovic

Project timetable:

- 2004-2006/Jan-Dec ,Sarajevo

Theoretical research :

- Field: Artificial Intelligence
- Field: Bioinformatics (Genomics&Proteomics)
- Field: Biomedical (Prostate&Cancer)

Practical implementation :

- MATLAB-Fuzzy Logic Toolbox
- MATLAB-Genetic Algorithm Toolbox

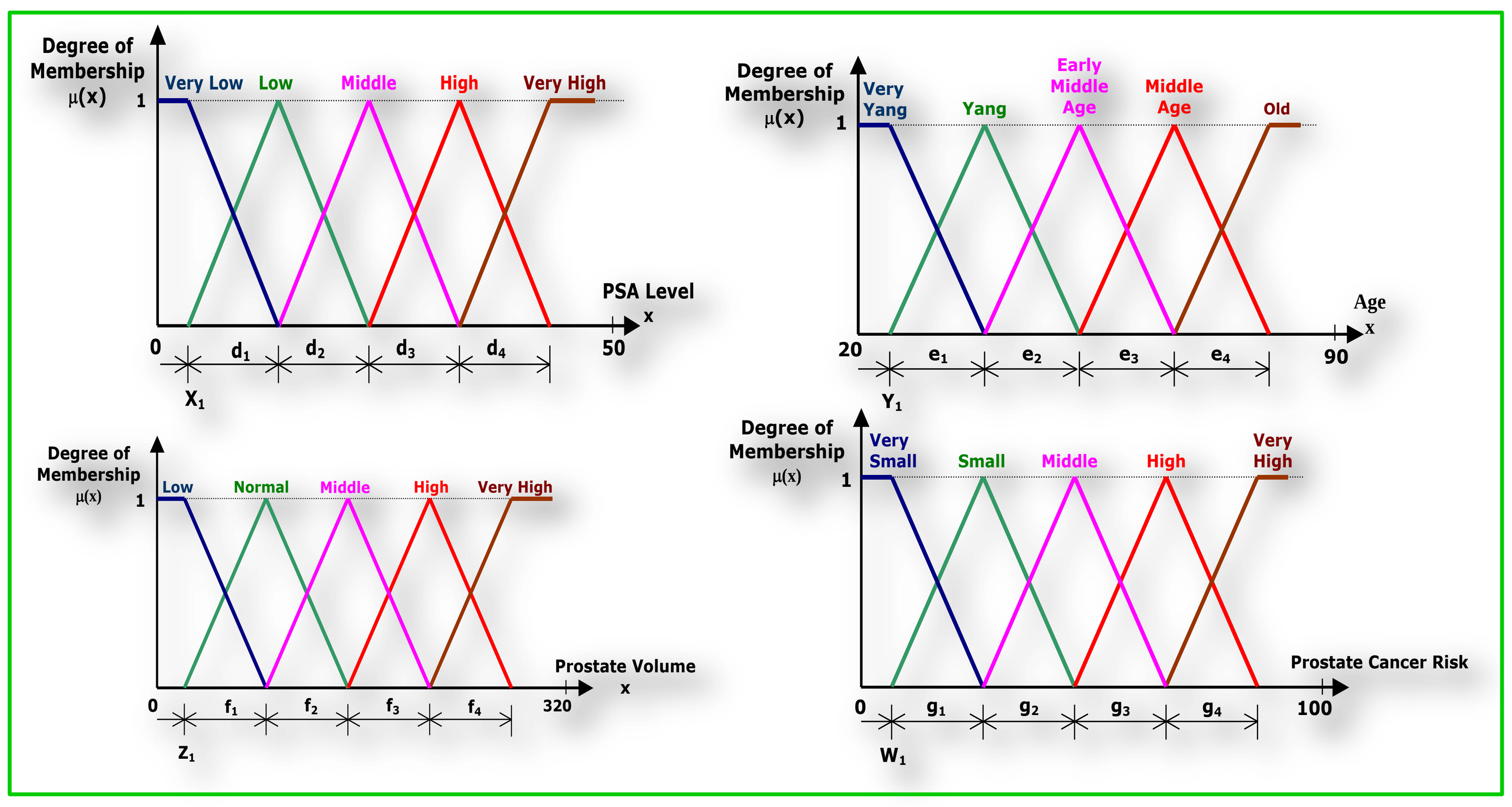
In this project, an attempt was made to explain modeling of fuzzy system for prostate cancer risk estimation using optimization based on genetic algorithm. Prostate cancer risk estimation is a real and complex problem that can be solved by building fuzzy system. Fuzzy modeling in this case is long and hard process, but it is possible to apply evolution algorithm to evolve fuzzy system and speed up fuzzy modeling process. The aim of this project is to present advantages and disadvantages of such modeling process and to discuss possibility of its practical application. As the result of this research, example of fuzzy system, which is obtained through evolution process, is presented.

Key words **— fuzzy, genetic algorithm, coevolution, prostate cancer, modeling.**

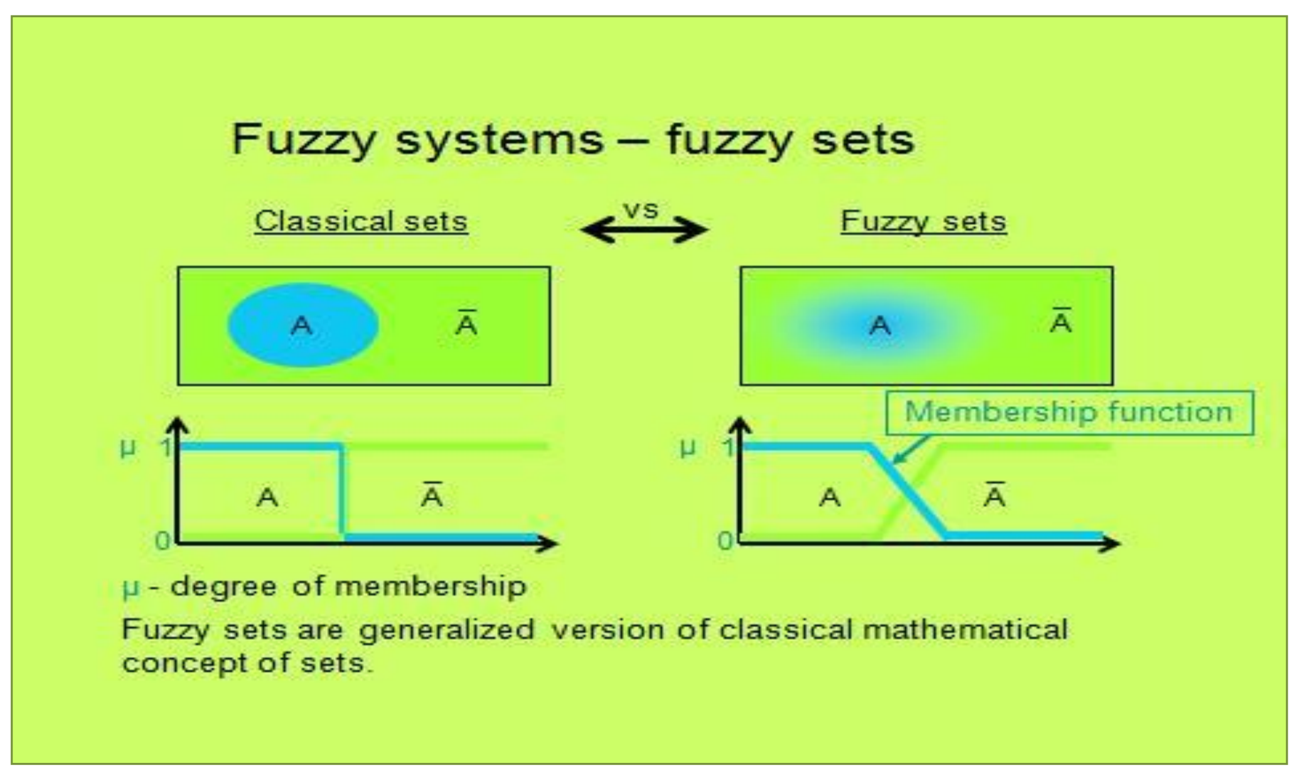
Fuzzy systems – fuzziness

- Modeling of fuzzy systems is based on imprecise concepts and imprecise dependencies among them.
- Example: If **Temperature** is High and **Sunshine** is Strong then **Vacation Conditions** are Perfect.
- For every real value (e.g. T = 20°C) is possible to define our estimation that it fits into some concept (e.g. **Temperature** is High), and that is called *fuzziness*. So, ***fuzziness*** **is a measure of our believe that some value (instance) fits into semantically ideal concept.**

Fuzzy system for prostate cancer risk estimation has 3 input fuzzy variables: PSA level, Age and Prostate Volume (PV), 1 output fuzzy variable: Prostate Cancer Risk (PCR) and collection of fuzzy rules. For presenting fuzzy variables, triangle, L and Γ types of fuzzy sets are used. They are choiced for simpler computations. Each variable is defined with 5 fuzzy sets, as shown on figures. Parameters X_1, d_1, d_2, d_3, d_4, Y_1, e_1, e_2, e_3, e_4, Z_1, f_1, f_2, f_3, f_4, W_1, g_1, g_2, g_3 and g_4 define the look of fuzzy variables.



Prostate cancer risk estimation can be based on three parameters: PSA (Prostate Specific Antigen) level, patient age and prostate volume (PV). PSA is a protein produced by the prostate that may be found in an increased amount in the blood of men who have prostate cancer, benign prostatic hyperplasia, or an infection of the prostate gland. A blood sample is measured in an assay and the amount of PSA is reported as [ng/ml]. Older patients have higher risk of prostate cancer, but also have higher prostate volume for normal PSA concentration. Prostate volume is determined by ultrasound imaging and is measured in [ml]. Many research projects are made to relate these parameters with prostate cancer risk estimation. Criteria for measuring the quality of one fuzzy system are based on the results of these projects.



Fuzzy systems – fuzzy operators

Operators in fuzzy theory are not uniquely defined because of membership gradation.

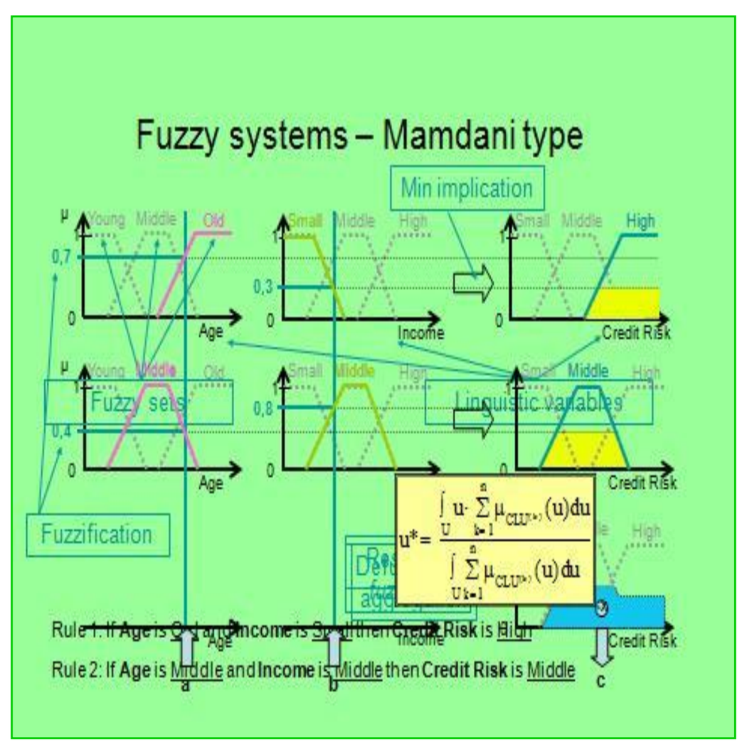
∧ operator ("AND" operator) is most commonly defined in one of the following ways:

Minimum	$\mu_A(x)\wedge\mu_B(x) = \min(\mu_A(x)+\mu_B(x))$
Product	$\mu_A(x)\wedge\mu_B(x) = \mu_A(x)\cdot\mu_B(x)$
Bounded sum	$\mu_A(x)\wedge\mu_B(x) = \max(0, \mu_A(x)+\mu_B(x)-1)$

∨ operator ("OR" operator) is most commonly defined in one of the following ways:

Maximum	$\mu_A(x)\vee\mu_B(x) = \max(\mu_A(x)+\mu_B(x))$
Product	$\mu_A(x)\vee\mu_B(x) = \mu_A(x)+\mu_B(x)-\mu_A(x)\cdot\mu_B(x)$
Bounded sum	$\mu_A(x)\vee\mu_B(x) = \min(1, \mu_A(x)+\mu_B(x))$

~ operator ("NOT" operator) is defined as:
$\mu{\sim}A(x) = 1 - \mu A(x)$



Evolutionary algorithms

Evolutionary algorithms (EA) are group of methods used in searching and optimization, inspired by natural evolution. They belong to group of so cold random search methods.

Basic mechanisms of EA are: crossover, mutation and selection.

These mechanisms manipulate genetic information about individuals contained in chromosome (genotype), but not with physical representation of individuals (*phenotype*).

Crossover combines chromosome parts of parents, transferring part of genetic information from both parents to child.

Mutation produce small variance in chromosome during reproduction.

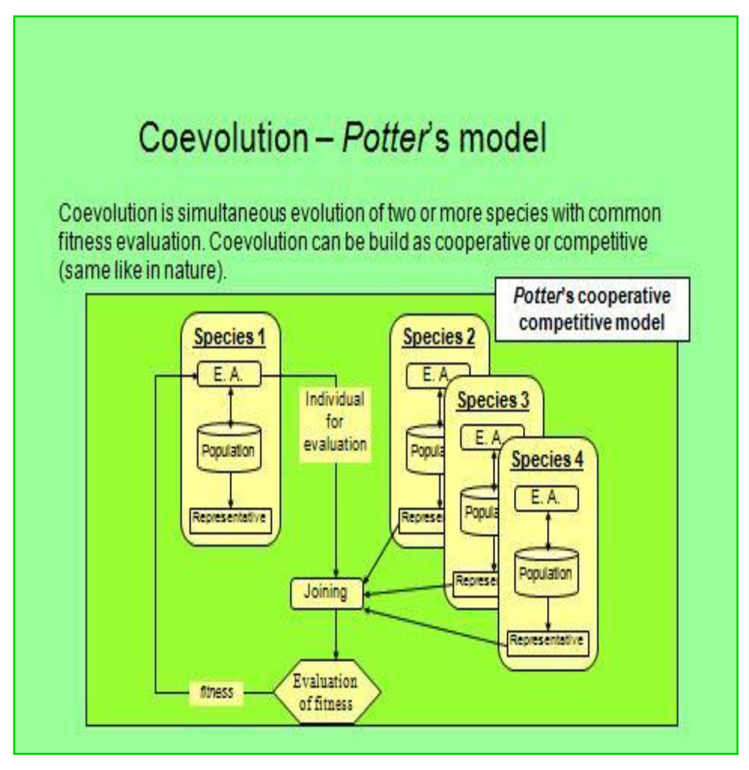
Selection choose individuals that will be involved in reproduction based on their fitness (measure of quality). Individuals with higher fitness have better chances to be involved in reproduction.

Evolutionary algorithms – basic structure

```
begin EA
    t = 0
    Generate initial population P(t)
    while not End of evolution do
        Evaluate P(t)
        P'(t) = Select [P(t)]
        P''(t) = Reproduce [P'(t)]
        P(t+1) = Merge [P''(t), P(t)]
        t = t+1
    end while
end EA
```

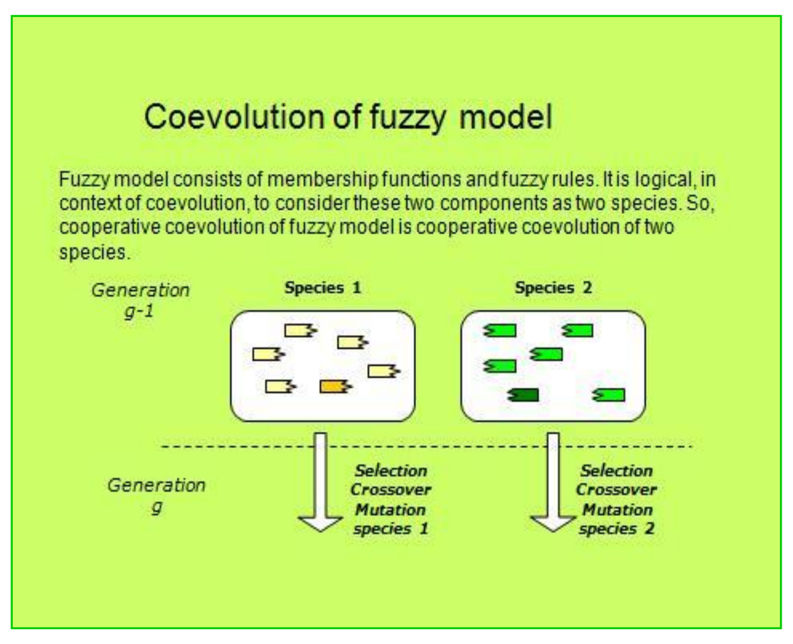
Cooperative coevolution is parallel evolution process of two or more species, where species helps each other in reaching some common goal. Each species has its own selection, crossover and mutation, but in the point of fitness calculation they join to create one whole whose quality is possible to measure.

Fuzzy model consists of membership functions and fuzzy rules. For applying cooperative coevolution on optimizing fuzzy model we create two species: membership functions and fuzzy rules. Number of rules and number of fuzzy sets for variables are not changed during evolution process, but they are set at the evolution start. As presented on figures fuzzy variables are defined with 5 fuzzy sets, each. Number of rules must also be defined and in this case is m = 14. Species has chromosomes like on figure. For all parameters of fuzzy model, which are in chromosomes, are intuitively defined ranges of values: X1∈[0,8], d1∈[2,8], d2∈[2,8], d3∈[2,8], d4∈[2,9], Y1∈[1,15], e1∈[1,10], e2∈[1,10], e3∈[1,10], e4∈[1,15], Z1∈[0,65], f1∈[5,65], f2∈[5,65], f3∈[5,65], f4∈[5,60], W1∈[0,20], g1∈[1,20], g2∈[1,20], g3∈[1,20], g4∈[1,20], A∈{Very Low – 1, Low – 2, Middle – 3, High – 4, Very High – 5}, B∈{Not in rule – 0, Very Yang – 1, Yang – 2, Early Middle Age – 3, Middle Age – 4, Old – 5}, C∈{Not in rule – 0, Low – 1, Normal – 2, Middle – 3, High – 4, Very High – 5}, D∈{Very Small – 1, Small – 2, Middle – 3, High – 4, Very High – 5}. Names of fuzzy sets are defined intuitively and are just names with no any meaning.

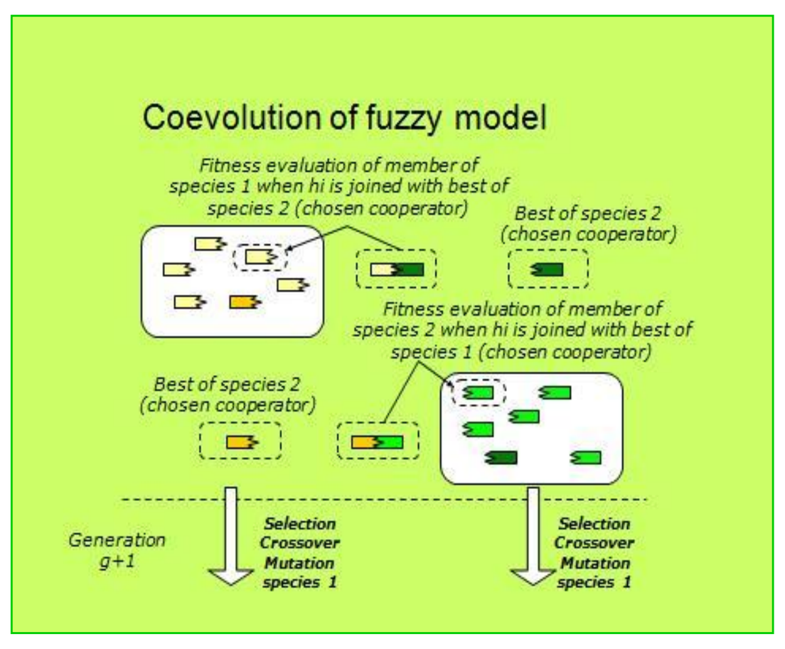


Coevolution – Potter's model
Coevolution is simultaneous evolution of two or more species with common fitness evaluation. Coevolution can be build as cooperative or competitive (same like in nature).
Potter's cooperative competitive model
Species 1
Species 2
Species 3
Species 4
Individual for evaluation
Joining
Evaluation of fitness

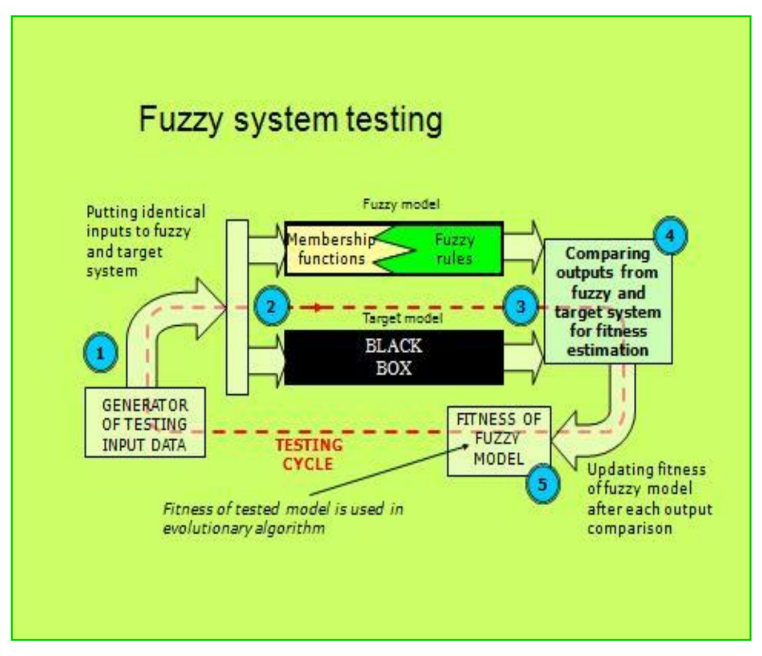
Cooperative coevolution of fuzzy system is presented on figure. Selection, crossover and mutation are independent for both species. But, when is necessary to compute the fitness of newly created member from one of species it is necessary to couple him with the best-known member from the other species to make one whole.



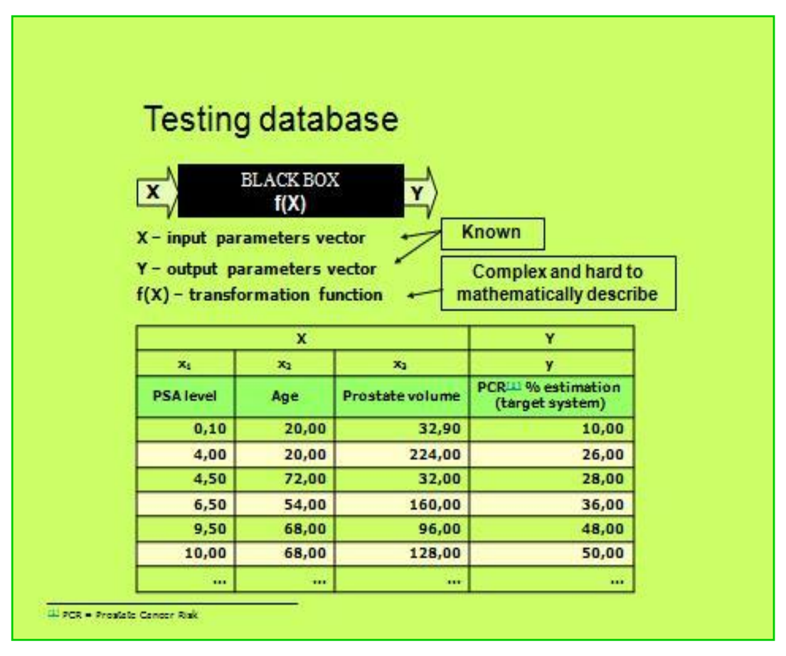
Coevolution of fuzzy model
Fuzzy model consists of membership functions and fuzzy rules. It is logical, in context of coevolution, to consider these two components as two species. So, cooperative coevolution of fuzzy model is cooperative coevolution of two species.
Generation g-1
Species 1
Species 2
Generation g
Selection Crossover Mutation species 1
Selection Crossover Mutation species 2



Coevolution of fuzzy model
Fitness evaluation of member of species 1 when hi is joined with best of species 2 (chosen cooperator)
Best of species 2 (chosen cooperator)
Fitness evaluation of member of species 2 when hi is joined with best of species 1 (chosen cooperator)
Generation g+1

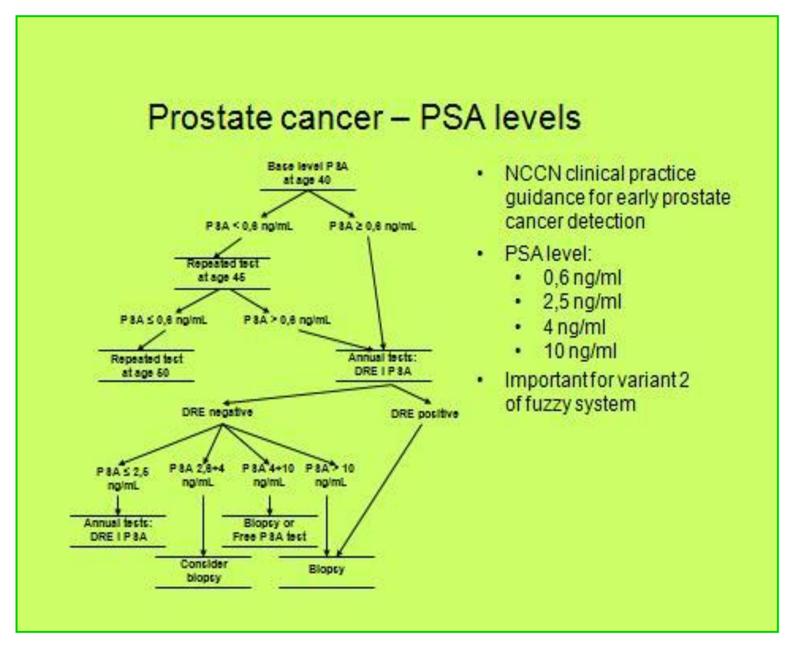


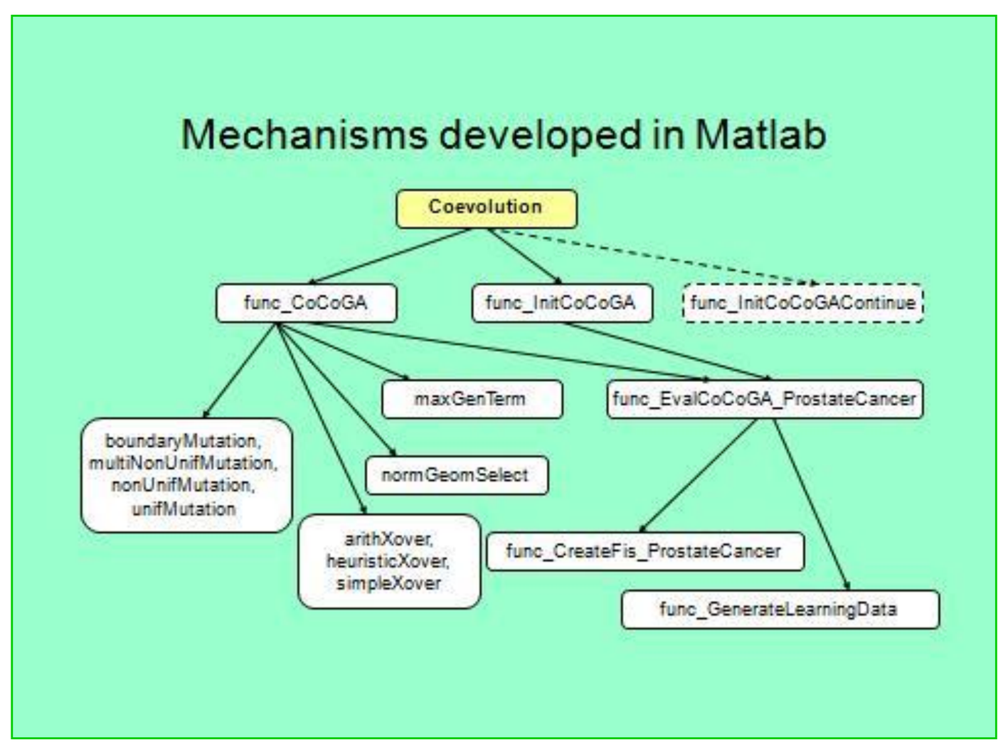
Fuzzy system testing
Putting identical inputs to fuzzy and target system
Membership functions
Fuzzy rules
BLACK BOX
Comparing outputs from fuzzy and target system for fitness estimation
GENERATOR OF TESTING INPUT DATA
TESTING CYCLE
FITNESS OF FUZZY MODEL
Updating fitness of fuzzy model after each output comparison
Fitness of tested model is used in evolutionary algorithm



Prostate cancer

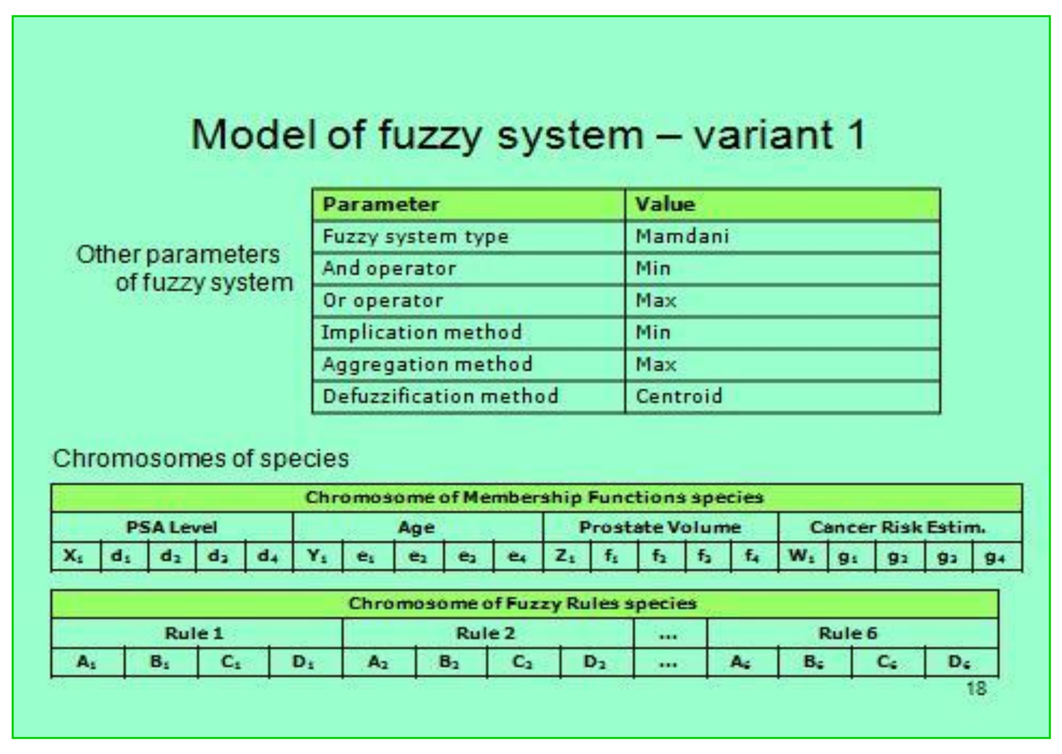
- one of most common mail malign disease
- early detection ⇒ successful healing
- different detection methods: PSA, DRE, ultrasound, biopsy,...
- PSA test – simple blood test; lack of specificity
- biopsy – complicated; unpleasant for patient; can cause depression, infection, bleeding and pain
- avoiding biopsy for people who has small prostate cancer risk
- simple factors that can be used for pretty precise prostate cancer risk estimation and decrease number of unnecessary biopsies:
  - PSA
  - age
  - prostate volume



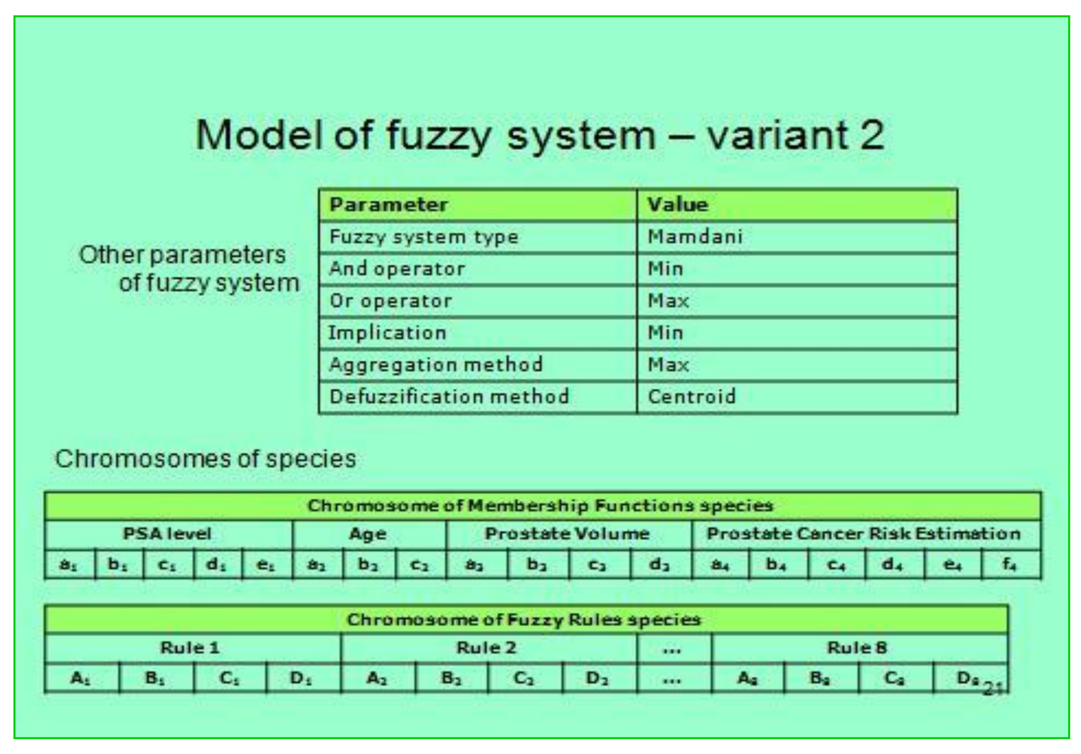


Mechanisms developed in Matlab
Coevolution
func_CoCoGA
func_InitCoCoGA
func_InitCoCoGAContinue
maxGenTerm
func_EvalCoCoGA_ProstateCancer
boundaryMutation, multiNonUnifMutation, nonUnifMutation, unifMutation
normGeomSelect
arithXover, heuristicXover, simpleXover
func_CreateFis_ProstateCancer
func_GenerateLearningData

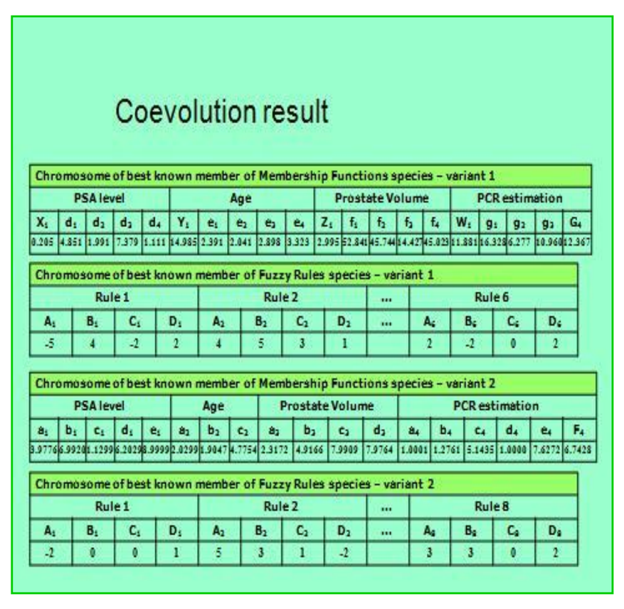
Most important moment in coevolution is fitness computation. It is shown on figure. When two members, each from one of two species (the one whose fitness we compute and the chosen cooperator – best in other species) are coupled, they represent one fuzzy system. That fuzzy system we can test using data from the database of known diagnoses. At the first (step 1) we generate matrix of inputs (input parameters for PCR estimation). Sequentially (one by one) set of input values is put on input of fuzzy system (step 2) and fuzzy output that represents PCR estimation is computed (step 4). That is repeated for all data in input matrix and matrix of corresponding outputs is created. At the same time we put same inputs at the entrance to the target system (step 3) and take outputs (step 5) – this is not actually performed because at the start, in our database, we have these outputs. Then in step 6, corresponding outputs from both systems are compared and percentage of approximately equal answers (PCR estimations) is computed. That percentage is fitness of that fuzzy system. Actually it is the fitness of the member from one of species, since in this step fitness of chosen cooperator is not changed. All mechanisms for evolving fuzzy system including coevolution algorithm are implemented as Matlab functions.



Model of fuzzy system – variant 1
Other parameters of fuzzy system
Parameter | Value
Fuzzy system type | Mamdani
And operator | Min
Or operator | Max
Implication method | Min
Aggregation method | Max
Defuzzification method | Centroid
Chromosomes of species
Chromosome of Membership Functions species
Chromosome of Fuzzy Rules species



Model of fuzzy system – variant 2
Other parameters of fuzzy system
Parameter | Value
Fuzzy system type | Mamdani
And operator | Min
Or operator | Max
Implication | Min
Aggregation method | Max
Defuzzification method | Centroid
Chromosomes of species
Chromosome of Membership Functions species
Chromosome of Fuzzy Rules species

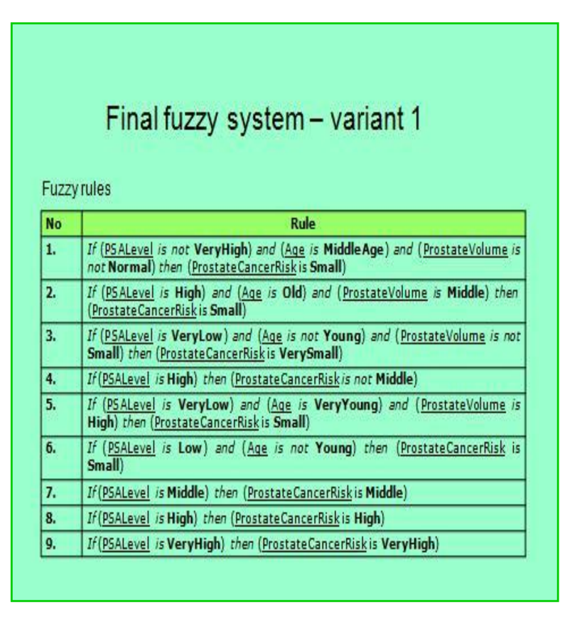


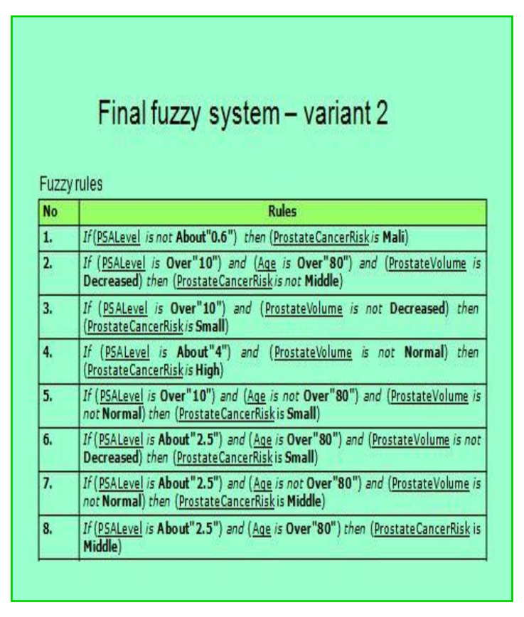
Membership functions chromosome

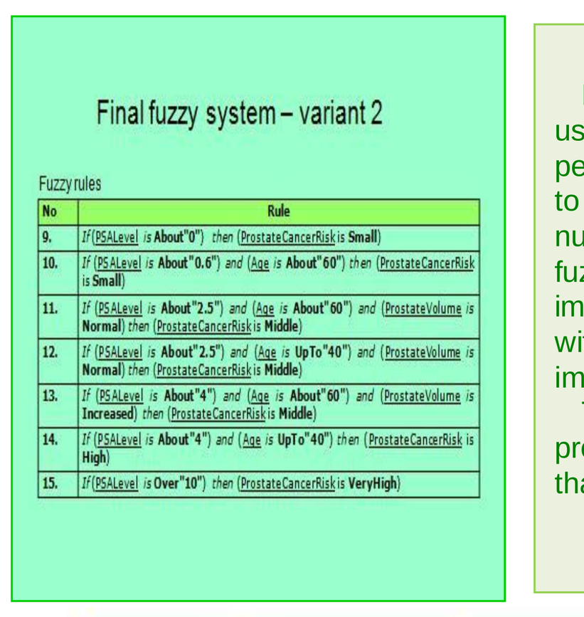
PSA Level					Age					PV					PCR				
X_1	d_1	d_2	d_3	d_4	Y_1	e_1	e_2	e_3	e_4	Z_1	f_1	f_2	f_3	f_4	W_1	g_1	g_2	g_3	g_4

Fuzzy rules chromosome

Rule 1				Rule2				...	Rule14			
A_1	B_1	C_1	D_1	A_2	B_2	C_2	D_2	...	A_{14}	B_{14}	C_{14}	D_{14}

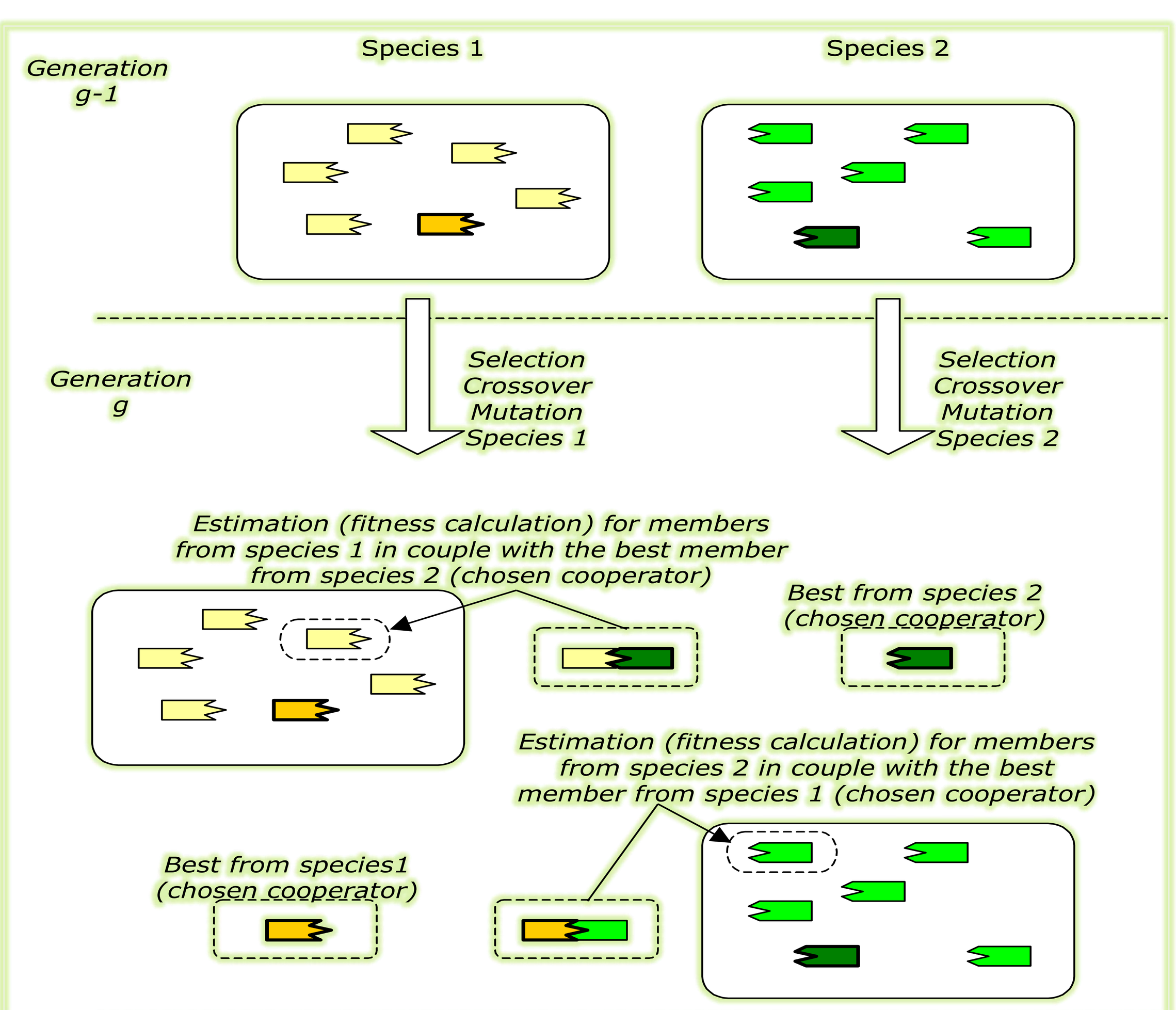






Example model has accuracy of approximately 70%, when tested with database of 3375 diagnoses used during coevolution, but is possible to create model with better results. That can be achieved by performing longer coevolution, for what is necessary more time and computer resources. Also is possible to evolve more structure elements of the model during coevolution, such as: type of fuzzy sets used, number of fuzzy sets for one fuzzy variable, number of fuzzy rules, etc. All that should give better final fuzzy system, but it complicates coevolution process and need more time and computer resources. Very important is quality of the created fuzzy model. That quality depends primarily on the size of the database with diagnoses and the accuracy of the diagnoses. It is possible to involve more diagnostic parameters to improve the accuracy of diagnoses.

This shows that application of cooperative coevolution on optimizing fuzzy model for this concrete problem gives good results. It also shows possibility of applying similar approach on solving any problem that we see as black box with known inputs and known outputs.



Generation g-1
Species 1
Species 2
Generation g
Selection Crossover Mutation Species 1
Selection Crossover Mutation Species 2
Estimation (fitness calculation) for members from species 1 in couple with the best member from species 2 (chosen cooperator)
Best from species 2 (chosen cooperator)
Estimation (fitness calculation) for members from species 2 in couple with the best member from species 1 (chosen cooperator)
Best from species1 (chosen cooperator)

COEVOLUTION RESULTS

- During experiments some elements are manually changed, like size of population for both species, size of database with known diagnoses, used for fitness computation, number of rules, etc., to try to influence evolution process and gain better result. On figures 8 to 11 is presented an example of membership functions after evolution process of 150 generations, where size of database with diagnoses used was 3375 and number of rules was 14.

Set of fuzzy rules is:

- *if* ***PSA level*** *is* Very Low *and* ***Age*** *is* Yang *and* ***PV*** *is* Middle *then* ***PCR*** *is* Small
- *if* ***PSA level*** *is* Very Low *and* ***Age*** *is* Early Middle Age *and* ***PV*** *is* Middle *then* ***PCR*** *is* Small*....*
- *if* ***PSA level*** *is* High *then* ***PCR*** *is* High

No	PSA Level	Age	Prostate Volume	PCR[1] estimation (target syst.)	PCR[1] estimation (*fuzzy* syst.1)	PCR[1] estimation (*fuzzy* syst.2)
1.	0,10	20,00	32,90	10,00	11,70	7,81
2.	0,50	20,00	37,15	10,00	11,94	8,00
3.	0,50	66,00	32,00	10,00	12,34	7,95
4.	1,00	20,00	40,30	10,00	12,50	8,00
5.	1,50	74,00	40,00	10,00	13,32	15,10
6.	2,50	64,00	320,00	20,00	15,84	14,92
7.	3,00	35,00	320,00	22,00	17,35	19,81
8.	3,00	60,00	43,90	10,00	17,35	7,76
9.	3,00	60,00	64,00	22,00	17,35	21,25
10.	4,00	20,00	256,00	26,00	20,70	31,44
11.	5,50	58,00	288,00	32,00	27,28	30,01
12.	7,00	56,00	224,00	38,00	35,38	36,92
13.	9,50	68,00	96,00	48,00	48,87	53,30
14.	10,00	68,00	160,00	50,00	49,41	54,85