Fuzzy-based Approach for Filling the Metabolic Pathway Hole

Ahmed Farouk Al-Sadek[1,3], Laila ElFangary[2] and Alaa Eldin Abdallah Yassin[1]

[1] Central Lab for Agricultural Expert Systems,
Giza, Egypt
[2] Faculty of Computers and Information - Helwan University ,
Cairo, Egypt
[3] Faculty of computer science, October University for modern science and Arts,
Giza, Egypt

Abstract

A key challenge in metabolic pathway hole problem is the reconstruction of the pathway data, reactions, enzymes and genes to use all this data to set the missing genes to the pathway which suffer from missing some genes in its reactions, we mean by the reconstruction here the relation between the enzymes and the genes in the pathway, However, most organism specific metabolic networks are left with a number of unknown enzymatic reactions, that is, many enzymes are missing in the known metabolic pathways, and these missing enzymes are defined as metabolic pathway holes , Although all reactions in some pathways are known, but also this pathways have a holes, the hole in this case means that we do not know the genes behind this reactions.

Results: In this paper we propose a new method to solve the second type of pathway holes using fuzzy logic approach. We applied fuzzy on our published database, RGBMAPS which consists of 100 pathways, 338 reactions and nearly 200,000. The system achieved an accuracy of 84%, where the correct genes which the system sets were 59 genes from the 70 genes.

Keywords: *Metabolic pathway. Bioinformatics. Pathway hole. Fuzzy logic. RGB MAPS database.*

1. Introduction

Metabolic network is one of the important classes of biological networks, consisting of enzymatic reactions involving substrates and products. Recent developments in pathway databases enable us to analyze the known metabolic networks. However, most organisms specific metabolic networks are left with a number of unknown enzymatic reactions, that is, many enzymes are missing in the known metabolic pathways, and these missing enzymes are defined as metabolic pathway holes [2], Although all reactions in some pathways are known, but also this pathways have a holes, the hole in this case means here that, we do not know the gene(s) that produce this enzyme.

Soft computing Technologies promise to become a powerful computational methodology for solving problems accurately and acceptably.

In this paper we propose a new method to solve the second type of pathway holes using fuzzy logic approach, where fuzzy logic is one of the soft computing components that could deal with uncertainty in real problem [2], due to the nature of continued data of our problem, because of the vagueness of boundaries between the concepts we preferred to use the fuzzy logic approach to overcome this problem.

2. Background

In recent years, a large number of metabolic databases have been developed to cover the huge amount of genome sequencing, where a key challenge in systems biology is the reconstruction of an organism's metabolic network from its genome sequence [9].Once the sequences are obtained, functions must be assigned to these new sequences [10], so the researchers do their efforts to solve the problems of the metabolic network. In this section we will present some related works which deal with metabolic pathway problems as prediction of EC form the chemical transformation and a Bayesian method [4] for identifying missing enzymes, where the prediction of pathways is hard for many reasons two of them are:

- The noise that introduced by the set of metabolic enzymes in the genome which mean that we have an errors and omissions because of this noisy.
- Reactions that share in multiple pathways are vague in supporting the presence of more than one pathway.

The first approach for predicting potential EC numbers from the chemical transformation pattern of substrates-product pairs called: E-zyme [8] solve one type of pathway hole which is missing enzyme of the reactions. In this method they focus on the prediction of the first three digits of the EC numbers (EC sub-subclass) by developing a new algorithm consist of three steps, in the first step they applied a graph alignment between the left-hand side and right-hand side of the reaction, the second step they compare the predicted pattern reaction with known EC numbers, in the last step of the algorithm they applied a voting scheme for selecting appropriate EC numbers [8].

In the second approach, the research team of this work has developed a method that efficiency combines homology and pathway-base evidence to identify candidates for filling holes in pathway/Genome database. They merge between sequence similarity based and statistical based approach for classifying proteins using a Bayes classifier , to do that they developed algorithm consist of three steps , the first one they acquired the known proteins for the given EC from the solved organisms , in the second step they done the similarity process between the known proteins of the first step and the target organism using BLAST to obtain the candidate proteins, in the last step they used a Bayes classifier to determine the probability that the candidate protein has the function required to fill the pathway hole [1].

3. Methodology

We designed a fuzzy based model to fill the metabolic pathway hole. From the Poole of soft computing we select fuzzy logic according to its ability to deal with vagueness and imprecise classes of the data. Due to the nature of continued data of our problem, because of the vagueness of boundaries between the concepts we preferred to use the fuzzy logic approach to overcome this problem.

As we know, Fuzzy logic is an "approach to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) Boolean logic on which the modern computer is based.", this does not mean that Fuzzy logic don't 0 or 1, no Fuzzy already includes 0 and 1 as extreme cases of truth (or "fact") but also includes the various states of truth in between so that, for example, the result of a comparison between two things could be not "tall" or "short" but ".60 of tallness."[6][7], any fuzzy logic system (FLS) consists of four main parts, Fuzzifier, Rules, Inference engine and Defuzzifier.

3.1 Overall algorithm

Figure 1 shows the overall algorithm used for filing pathway hole. The steps of the algorithm applied on each reaction from the 338 reactions of RGBMAPS database.

a) ***Gene's retrieval*** – retrieve from RGBMAPS database genes that catalyze the desired reaction in other organisms, then we retrieve all possible genes of the organism of interest using BLAST. RGBMAPS gives us directly all genes of a specific EC in the target organism human after passing three phases of collecting data, (i) collect the pathways of the interest organism with its reactions and ECs, (ii) retrieve all genes that act with this EC in the different organisms, (iii) using BLAST to retrieve all possible genes of the target organism which are similar to the genes of other organisms [3].

b) ***Candidate genes to fill hole*** – in this step of the algorithm we feed our proposed fuzzy model with all genes of the EC, to filter all these genes and candidate the genes only that can fill the pathway hole, we illustrated this step in the later section.

c) ***Candidate evaluation*** – in this step we applied shot-gun score to set the correct gene form these candidate genes that obtained from step2.

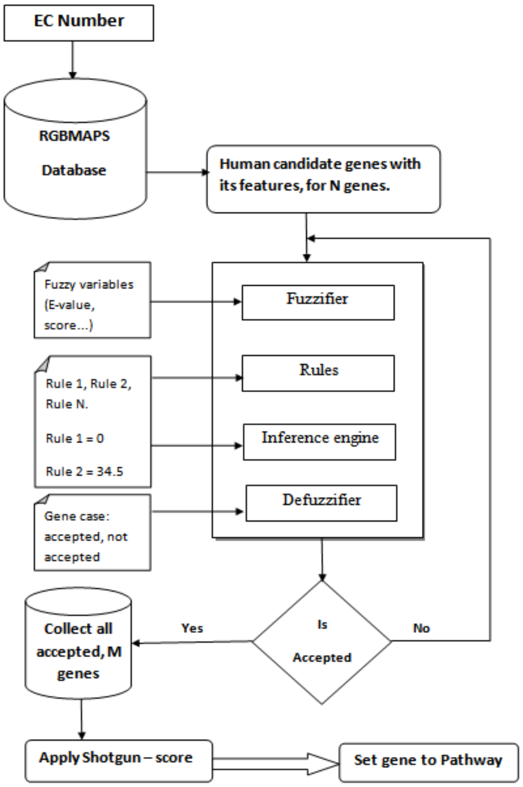
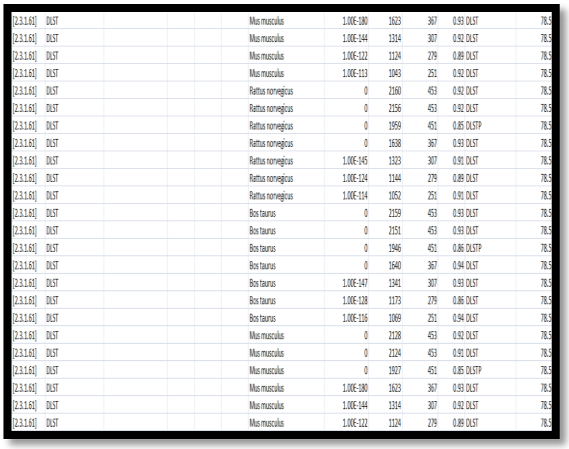


Figure 1: The block diagram for the proposed algorithm

a) Gene's retrieval

As we present above, in this step we need to obtain all possible genes to a specific EC in different organisms to be the input data to the fuzzy system, we select RGBMAPS database [3] to be our source of data. To do that we select the EC, for example 2.3.1.61 then RGBMAPS candidate all genes that act with this EC, as shown in figure 2.



[2.3.1.61]	DLST	Mus musculus	1.00E-180	1623	367	0.93	DLST	78.5
[2.3.1.61]	DLST	Mus musculus	1.00E-144	1314	307	0.92	DLST	78.5
[2.3.1.61]	DLST	Mus musculus	1.00E-122	1124	279	0.89	DLST	78.5
[2.3.1.61]	DLST	Mus musculus	1.00E-113	1043	251	0.92	DLST	78.5
[2.3.1.61]	DLST	Rattus norvegicus	0	2160	453	0.92	DLST	78.5
[2.3.1.61]	DLST	Rattus norvegicus	0	2156	453	0.92	DLST	78.5
[2.3.1.61]	DLST	Rattus norvegicus	0	1959	451	0.85	DLSTP	78.5
[2.3.1.61]	DLST	Rattus norvegicus	0	1638	367	0.93	DLST	78.5
[2.3.1.61]	DLST	Rattus norvegicus	1.00E-145	1323	307	0.91	DLST	78.5
[2.3.1.61]	DLST	Rattus norvegicus	1.00E-124	1144	279	0.89	DLST	78.5
[2.3.1.61]	DLST	Rattus norvegicus	1.00E-114	1052	251	0.91	DLST	78.5
[2.3.1.61]	DLST	Bos taurus	0	2159	453	0.93	DLST	78.5
[2.3.1.61]	DLST	Bos taurus	0	2151	453	0.93	DLST	78.5
[2.3.1.61]	DLST	Bos taurus	0	1946	451	0.86	DLSTP	78.5
[2.3.1.61]	DLST	Bos taurus	0	1640	367	0.94	DLST	78.5
[2.3.1.61]	DLST	Bos taurus	1.00E-147	1341	307	0.93	DLST	78.5
[2.3.1.61]	DLST	Bos taurus	1.00E-128	1173	279	0.86	DLST	78.5
[2.3.1.61]	DLST	Bos taurus	1.00E-116	1069	251	0.94	DLST	78.5
[2.3.1.61]	DLST	Mus musculus	0	2128	453	0.92	DLST	78.5
[2.3.1.61]	DLST	Mus musculus	0	2124	453	0.91	DLST	78.5
[2.3.1.61]	DLST	Mus musculus	0	1927	451	0.85	DLSTP	78.5
[2.3.1.61]	DLST	Mus musculus	1.00E-180	1623	367	0.93	DLST	78.5
[2.3.1.61]	DLST	Mus musculus	1.00E-144	1314	307	0.92	DLST	78.5
[2.3.1.61]	DLST	Mus musculus	1.00E-122	1124	279	0.89	DLST	78.5

Figure 2: all possible genes of EC: 2.3.1.61

As we see the first column represent the EC, the second column is the gene(s) of the different organisms that act with this EC, the third is the organism name, the fourth, fifth, six and seventh columns represents E-value, score, length and identity respectively, the eight column represent the candidate human gene which is similar to the genes in the different organisms, and the last column represent the fuzzy value of this genes after applying the rules.

b) Candidate genes to fill hole

The purpose of this step is to filter these entire possible genes which are 86 genes in our example EC: 2.3.1.61, throw our fuzzy system to produce the candidate genes for this EC. To do that we need to determine the fuzzy input sets and the rules that the system will use .In the following internal sections we will illustrate that in some details.

- **Fuzzy input sets**

  We chose three variables to be our fuzzy sets, E-value, score and identity.

  We designed our fuzzy variables by setting the variables ranges and its mathematical shape.

- **Fuzzy variables ranges**

  As presented in table 1, we show the different ranges of fuzzy variables that we assign to this system, as shown in table we assign three sets to each variable, high, medium and low.

Table.1: value ranges of fuzzy set parameters

Parameter	max	High			Medium			Low		
e-value	10	0	0.2	1	0.8	1.4	2	1.7	3	10
Score	6000	1000	3000	6000	500	1500	2500	0	400	800
Identity	1	0.6	0.8	1	0.3	0.5	0.7	0	0.2	0.4

The three ranges high, medium and low are arranged according to the value ranges of each fuzzy variable, where E-value the value of it is between 1 and 10, so we suggest that the high from 0 and 1 where medium between 0.8 and 2 and low is from 1.7 and upper, but score has a big value ranges because its value ranges is huge than E-value and identity.

- **Fuzzy variables mathematical shape**.

each variable take a shape in the fuzzy sets, these shapes are left shoulder, right shoulder and triangle, where left shoulder represent in E-value the high value but in score and identity represent the low value, and right shoulder represent in E-value the low value but in score and identity represent the high value, but triangle represents in all parameters the medium value, all this shapes presented in figure 3.

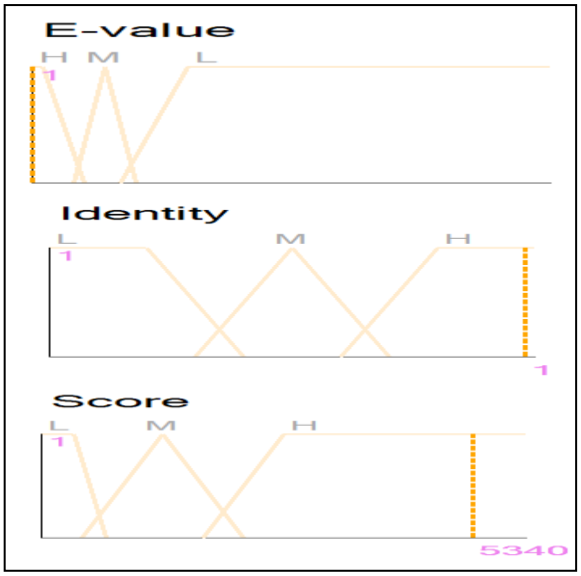


Figure 3: fuzzy set variables.

- **Fuzzy rule system**

  In our work we create three models of rules and applied it on the data to elect the more efficient one, the election processes are discussed in "evaluation" section and the election model is presented in the figure 4.

Rule 1: If score is high and identity is high then accepted is very.
Rule 2: if E-value is low and score is low and identity is low then accepted is not.
Rule 3: if E-value is high and score is medium and identity is high then accepted is very.
Rule 4: if E-value is high and score is low and identity is low then accepted is not.
Rule 5: if E-value is high and score is medium and identity is medium then accepted is may be.

Figure 4: Fuzzy rule system.

- **Applying fuzzy rules**
  In this step we apply the elected fuzzy rule model on the data that obtained from the first step of the algorithm, which is the all possible genes of the EC which was in our example 86 genes of EC: 2.3.1.61. After applying the rules on 86 genes, the fuzzy system classify the genes into three classes accepted genes, may be genes and not accepted genes and this according to the ranges of the fuzzy variables and the different values of each genes. As presented in table 2, present sample to three different cases form the 86.

Table 2: Example of three genes of the possible genes

Organism	gene	Human gene	E-value	score	identity	Fuzzy value	case
Mus musculus	DLST	DLST	0	2128	0.92	78.5	Accepted
			High	medium	High	Rule 1 , 3	
Arabidopsis	DLST	DLST	1E-100	937	0.48	59.5	May be
			High	medium	medium	Rule 5	
Mus musculus	DLST	DLAT	1E-35	376	0.28	21	Not accepted
			High	low	low	Rule 4	

After applying the rules on all possible genes, we have 42 accepted genes, 14 not accepted, 20 may be gens, 10 NAN, we select only the accepted genes in all organisms, so we have 42 candidate genes, all this genes can be the correct gene that can fill the hole, so now we need another level of filtration to obtain only the correct gene.

c) Candidate evaluation

As we present above we need to evaluate the candidate genes produced by the fuzzy model, to elect one gene only to set the pathway hole, we applied s Shotgun-score on these genes, and ranking from the biggest score to the lowest, where shotgun-score is “the number of query sequences whose fuzzy system output included the candidate sequences”[1].

In our example we have 42 hits (only accepted genes); we need to decide which one of them is the correct gene to fill the hole.

Table 3: Shotgun-score result summary

Organism	DLST	DLSTP	total
Rattus norvegicus	12	2	14
Bos taurus	12	2	14
Mus musculus	12	2	14
Sum	36	6	42

From above, after applying shotgun-score, the DLST gene has the biggest shotgun-score so DLST is the correct gene to fill pathway hole.

4. Evaluation

As we illustrate in the previous section, we elect our fuzzy rules from three models, here we will show this three models and how we elect the fuzzy one.

To build our three rule models we implement a tool to build these models in easy way and to give the ability to test another models in future.

a) Fuzzy rule evaluation

We summarize the three models in the following tables 4, 5 and 6, where H mean high ,M mean medium and L mean low , which represent fuzzy variables ranges as presented before.

Table 4: Fuzzy model 1

Rule	E-value	Score	identity	Case	#Gene Cases
Rule 1	H	H	H	very	29
Rule 2	L	L	L	Not	9
Rule 3	L	M	M	May be	
Rule 4	M	H	H	very	1
Rule 5	H	M	H	very	29
Rule 6	M	L	L	Not	1
Rule 7	H	L	L	Not	27
Rule 8	H	M	M	May be	22

Table 5: Fuzzy model 2

Rule	E-value	Score	identity	Case	#Gene Cases
Rule 1	H	H	H	very	31
Rule 2	L	L	L	Not	10
Rule 3	L	M	M	May be	3
Rule 4	M	H	H	Very	6
Rule 5	M	L	L	Not	1
Rule 6	H	M	H	Very	26
Rule 7	H	L	L	May be	28
Rule 8	H	M	M	very	19
Rule 9	H	L	M	May be	8

Table 6: Fuzzy model 3

Rule	E-value	Score	identity	Case	#Gene Cases
Rule 1	-	H	H	very	27
Rule 2	L	L	L	Not	9
Rule 3	H	M	H	very	26
Rule 4	H	L	L	Not	30
Rule 5	H	M	M	May be	26

we need to elect one of the previous sets to be our rule fuzzy sets, we apply these different models on 100 genes and make statistics to measure the efficiency of each one; the fowling flow charts illustrate this statics of each set.
From above we select model 3 to be our fuzzy rule because we look at the number of rules in the set and the number of NAN cases, which mean that no rule from this set face the case, so the model 3 is the best one because it have 1 NAN and consists of 5 rule only.

b) Filing hole evaluation.

In this section we evaluate the shotgun-score level which is the last step in our algorithm to set the correct gene to fill the pathway hole, so this evaluation reflect the evaluation of our proposed system at all, as presented in table 7 we applied our proposed system on 70 sample to set the correct gene the total percent is 84%.

5. Observations

- Row 9, 10: Note: the difference between CHSY1 and CHSY2 is the last number only, as we mention that in the conclusion section. The same note in rows 36, 37 and 42.
- Row 11: Note: the difference between PCYT1B and PCYT1A is the last number only, as we mention that in the conclusion section, the two genes act at the same EC, so the answer considered correct.
- But in row 53 the difference between the correct gene and the candidate genes is also the last letter CHKB and CHKA, but we consider the answer is wrong because the two gens don't act on the same EC in the pathway.
- In Row 50 the system set two genes with the same shotgun score 4 , the correct one CHKA and PCYT1B , so in the percent of correction we consider this case as wrong case , because the system don't give us a definitive answer.
- In row 55, there is new note, the system candidate gene decrease from the correct gene by one letter EPT1 and CEPT1.

6. Interpretation

When the difference between the two genes is in the last position of the name, we looked if this position is number, so the similarity between the two genes is very close, and if the difference which is in the last position is letter, so the two genes are similar but less than the first one.

7. Conclusion

- The proposed system has the ability to solve pathway hole problem using the proposed database RGBMAPS and the proposed fuzzy system.
- In data collection phase of our database, to do this task in manually way, that is very hard, waste effort and time; we have overcome the problem by writing a small Perl program to make these steps easier.
- We make BLAST with the amino acid sequence (AA Seq.) not by the nucleotide sequence (NT Seq.) because the hole in pathway interested in the function (AA Seq.) not by the NT Seq.
- Some Reactions acts with the same gene, (Ex: 1.14.15.4 and 1.14.15.5) act with the same gene, (1.1.1.145 and 5.3.3.1) act with the same gene, (2.1.2.2, 6.3.3.1 and 6.3.4.13) act with the same gene, this note may be very useful in gene therapy.
- We note that our system give a good result with the genes that differ from the correct gene in the last number of the gene name like GFPT1 and GFPT2 where this note is very promising in gene therapy.
- Using fuzzy system is very favorable in pathway problems, because it's gaining strength through its seemed closer to the way our brains work, which make the researchers closer to the data.
- In future woks we will use machine learning to build the fuzzy system rule and also we will reevaluate our proposed system after changing the fuzzy variables ranges, (E-value, score and identity).

Table 7: system evaluation

#	EC	Pathway genes	System gene	note
1	2.3.1.61	DLST	DLST	√
2	1.2.4.1	PDHA1	PDHA1	√
3	1.8.1.4	DLD	DLD	√
4	2.3.1.12	DLAT	DLAT	√
5	4.2.1.47	GMDS	GMDS	√
6	1.2.4.4	BCKDHA	BCKDHA	√
7	2.3.1.168	DBT	DBT	√
8	2.4.1.174	CSGALNAC	CSGALN	√
9	2.4.1.175	CHSY1	CHSY2	X
10	2.4.1.226	CHSY1	CHSY2	X
11	2.7.7.15	PCYT1B	PCYT1A	√
12	2.7.8.2	CEPT1	CEPT1	√
13	3.1.4.4	PLD1	PLD1	√
14	1.14.13.39	NOS1	NOS1	√

Table 7 : continued

15	6.3.4.5	ASS1	ASS1	√
16	4.3.2.1	ASL	ASL	√
17	2.5.1.21	FDFT1	FDFT1	√
18	1.14.99.7	SQLE	SQLE	√
19	1.1.1.1	ADH1B	CACNA1	X
20	1.2.1.3	ALDH2	ALDH2	√
21	6.2.1.1	ACSS1	ACSS1	√
22	1.11.1.6	CAT	CAT	√
23	5.3.3.2	IDI1	IDI1	√
24	2.5.1.1	FDPS	FDPS	√
25	2.5.1.10	FDPS	FDPS	√
26	4.1.1.15	GAD1	GAD1	√
27	1.2.1.24	ALDH5A1	ALDH5A1	√
28	2.6.1.19	ABAT	ABAT	√
29	1.11.1.9	GPX1	GPX1	√
30	1.8.1.7	GSR	GSR	√
31	1.11.1.12	GPX4	GPX4	√
32	1.4.4.2	GLDC	GLDC	√
33	2.1.2.10	AMT	AMT	√
34	1.8.1.4	DLD	DLD	√
35	١,١,١,٣٠	BDH1	BDH1	√
36	2.8.3.5	OXCT**1**	OXCT**2**	**X**
37	2.3.1.9	ACAT**1**	ACAT**2**	**X**
38	6.4.1.3	PCCB	PCCB	√
39	5.1.99.1	MCEE	MCEE	√
40	5.4.99.2	MUT	MUT	√
41	2.7.1.23	NADK	NADK	√
42	3.1.3.2	ACP**6**	ACP**2**	**X**
43	1.6.1.2	NNT	NNT	√
44	1.1.1.49	G6PD	G6PD	√
45	3.1.1.31	PGLS	PGLS	√
46	1.1.1.44	PGD	PGD	√
47	1.14.16.1	PAH	PAH	√
48	4.2.1.96	PCBD1	PCBD1	√
49	1.5.1.34	QDPR	QDPR	√
50	2.7.1.32	CHKA	CHKA=	X √
51	2.7.7.15	PCYT1A	PCYT1A	√
52	2.7.8.2	CHPT1	CHPT1	√
53	2.7.1.82	CHK**B**	CHK**A**	**X**
54	2.7.7.14	PCYT2	PCYT2	√
55	2.7.8.1	**C**EPT1	EPT1	X
56	3.5.4.16	GCH1	GCH1	√
57	4.2.3.12	PTS	PTS	√
58	1.1.1.153	SPR	SPR	√
59	1.3.1.2	DPYD	DPYD	√
60	3.5.2.2	DPYS	DPYS	√
61	3.5.1.6	UPB1	UPB1	√
62	1.2.1.18	ALDH6A1	ALDH6A1	√
63	2.6.1.1	GOT1	GOT1	√
64	1.1.1.37	MDH2	MDH2	√
65	1.2.1.8	ALDH7A1	ALDH7A1	√
66	1.1.99.1	CHDH	ALDH7A1	x
67	2.3.1.38	FASN	FASN	√
68	2.3.1.41	OXSM	FASN	X
69	2.7.1.26	RFK	RFK	√
70	2.7.7.2	FLAD1	FLAD1	√

The accuracy of proposed system = (the correct genes / total genes) = 59/70 = **84%.**

Reference

[1] Green ML, Karp PD,”A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases”, BMC Bioinformatics 2004, 5:76.

[2] Marc S, Bakk r “Metabolic Pathway Visualization Using Gene-Expression Data”, Master’s Thesis 2007, Institute for Computer Graphics and Vision Graz University of Technology ,Graz.

[3] Alaa.yassin, Laila.Elfangary,A. Sadek : “RGB MAPS: a Proposed Database for solving Metabolic Pathway Hole”, The 8th International Conference on Informatics and Systems (INFOS2012) – 14-16 May, 2012.

[4] Dale JM, Popescu L, Karp PD:”Machine learning methods for metabolic pathway Prediction”, BMC Bioinformatics 2010, 11:15.

[5] Hui-Huang Hsu, “Advanced data mining technologies in bioinformatics”, Copyright © 2006 by Idea Group Inc.

[6] A.Sharaf Eldin, M. Hana, S. Kassim and S. Rashad, "Associations System for Breast Cancer Microarray Data", International Journal of Intelligent Computing and Information Science, Vol. 9, No. 2, July, 2009.

[7] Manjula Kurella,"DNA Microarray Analysis of Complex Biologic processes",2001.

[8] Yamanishi Y, Attori M, Kotera M,Goto S, Kanehisa M:”E-zyme:predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs”, Bioinformatics2009.

[9] Dale JM, Popescu L, and Karp PD:”Machine learning methods for metabolic pathway Prediction”, BMC Bioinformatics 2010, 11:15.

[10] Karp PD, Caspi R:”A survey of metabolic databases emphasizing the MetaCyc family”, Arch Toxicol2011.

Bibliography

[1] Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PD," Computational prediction of human metabolic pathways from the complete human genome", Genome Biology 2004, 6:R2.

[2] Yamanishi Y, Attori M, Kotera M,Goto S, Kanehisa M:"E-zyme:predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs", Bioinformatics2009.

[3] Bedri Karakas, Ashani T. Weeraratna, Abde M. Abukhdeir, Hiroyuki Konishi, John P.Gustin, Michele. Vitolo, Kurtis E. Bachman, and Ben Ho Park, "p21 Gene Knock Down Does Not Identify Genetic Effectors Seen with Gene Knock Out", Cancer Biol Ther. 2007 July ; 6(7): 1025–1030.

[4] "Gene Therapy" Produced by the Centre for Genetics Education. Internet: http://www.genetics.edu.au, The Australasian Genetics Resource Book – © 2007.

[5] Gregory Stephanopoulos "Metabolic Fluxes and Metabolic Engineering", Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, Copyright © 1999 by Academic Press, 1999.

[6] Dale JM, Popescu L, and Karp PD:"Machine learning methods for metabolic pathway Prediction", BMC Bioinformatics 2010, 11:15.

[7] http://blast.ncbi.nlm.nih.gov/Blast.cgi? , July, 2011.

[8] Uniprot, http://www.uniprot.org/, July, 2011.

[9] Whatis.com: Definitions: fuzzy logic, http://whatis.techtarget.com/definition/0,,sid9_gci212172,00.html, Juli, 2012.

Ahmed Farouk Al-Sadek is the head of Agricultural Expert System Development in Central Laboratory for Agricultural expert System, CLAES and also works as program leader of computer science department in Faculty of computer science, October University for modern science.

Laila Mohamed ElFangary is member staff of Faculty of Computers and Information - Helwan University.

Alaa Eldin Abdallah Yassin is a member in research staff of Central Laboratory for Agricultural expert System, CLAES, works as research assistant in Knowledge Engineering and Expert System Building Tools.