

1 **Video Article:**

2 **Selecting multiple biomarker subsets with similarly effective**
3 **binary classification performances**

4
5 **AUTHORS AND AFFILIATIONS:**

6 Xin Feng¹, Shaofei Wang¹, Quewang Liu¹, Han Li², Jiamei Liu², Cheng Xu², Weifeng Yang², Yayun
7 Shu², Weiwei Zheng¹, Bingxin Yu³, Mingran Qi⁴, Wenyang Zhou¹, Fengfeng Zhou¹.

8 ¹College of Computer Science and Technology, and Key Laboratory of Symbolic Computation
9 and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin, China.

10 ²College of Software, Jilin University, Changchun, Jilin, China.

11 ³Ultrasonography Department, China-Japan Union Hospital of Jilin University, Changchun, Jilin,
12 130012, China.

13 ⁴Department of Pathogenobiology, College of Basic Medical Science, Jilin University,
14 Changchun, Jilin, 130012, China.

15

16 **# CORRESPONDING AUTHOR:**

17 Fengfeng Zhou (email: FengfengZhou@gmail.com)

18

19 **KEYWORDS:**

20 Biomarker detection, feature selection, OMIC, binary classification, filter, wrapper, extreme
21 learning machine, ELM

22

23

24

25 How to install kSolutionVis

26 In order to run the kSolutionVis program properly, users should read the follow instructions.

27
28 (1) Firstly, users should download the Python version 3.6.0 or above from the website:
29 <https://www.python.org/downloads/release/python-362/>.

30
31 Or for the users' convenience, it is highly recommended to download a Python IDE such as Anaconda
32 and Pycharm from the following websites:

33 <https://www.anaconda.com/distribution/>
34 <http://www.jetbrains.com/pycharm/>

35
36 (2) After successfully installing Python, the next step is to install the packages needed by the
37 kSolutionVis program, including *pandas*, *abc*, *numpy*, *scipy*, *sklearn*, *sys*, *PyQt5*, *sys*, *math* and
38 *matplotlib*. For example, users may use the command line to install one of the above mentioned
39 packages - *pandas* by inputing the following command in the Windows command line:
40 *pip install pandas*

41
42 The other packages could be installed likewise. The command pip will check whether a package is
43 installed and only install those unavailable packages.

44
45 (3) Download the software kSolutionVis from: <http://www.healthinformatics.org/supp/resources.php>

46
47 (4) Run the software kSolutionVis by the following command in the directory of kSolutionVis:
48 *python kSolutionVis.py*

49

50 Error messages

51 Here is the list of messages that some errors occurred when using the software kSolutionVis.

- 52 1. The number of samples doesn't match the number of class labels!
53 When the number of samples is not equal to the number of sample class labels.
- 54 2. xxx are not in the class labels!
55 When samples do not match correctly with the class labels.
- 56 3. Please make sure the label set contains the "Class/class" column
57 When the class label file does not have the column with the name "Class" or "class".
- 58 4. No matched feature!
59 When no feature was found with the user-given feature name.
- 60 5. There is no qualified Triplet! The max value=XXX.
61 When the cutoff piCutoff is higher than the classification performance of any triplets. The
62 maximal classification performance was given, so that the user may choose a smaller cutoff.

63

64 Parallel running time

65 Parallel running of kSolutionVis using different CPU cores on the dataset ALL1. Column
66 "top_x" gives the values of the parameter top_x. Column "Windows" gives the running time of
67 kSolutionVis for each "top_x" using one CPU core. The columns "Core-1"/"Core-2"/"Core-
68 4"/"Core-6"/"Core-8" are the running time of kSolutionVis on a MacBook laptop computer
69 using the numbers of CPU cores 1/2/4/6/8, respectively. Due to the technical limitation of the

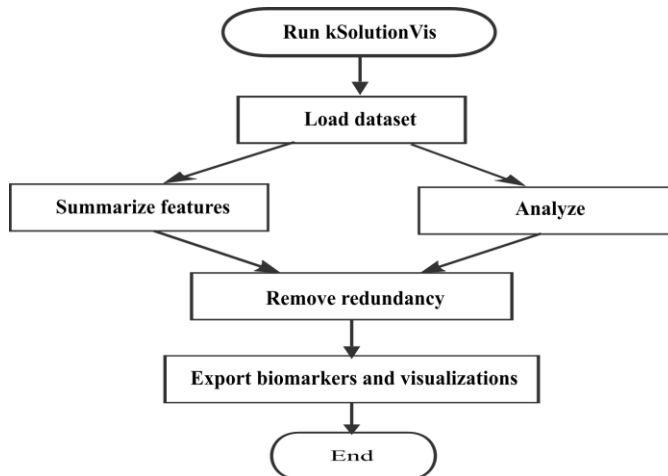
70 Python parallel libraries on Windows, the parallel version of kSolutionVis cannot run on
 71 Windows. The Windows user may choose the single-core version of kSolutionVis.

top_x	Windows	Core-1	Core-2	Core-4	Core-6	Core-8
10	12.98	8.56	8.48	9.41	10.59	10.53
20	32.30	20.37	18.14	21.59	24.84	23.70
30	96.56	56.41	43.43	54.80	62.33	68.62
40	223.23	128.74	95.20	123.45	141.05	166.43
50	453.75	254.54	187.55	253.10	276.08	318.10

72
 73

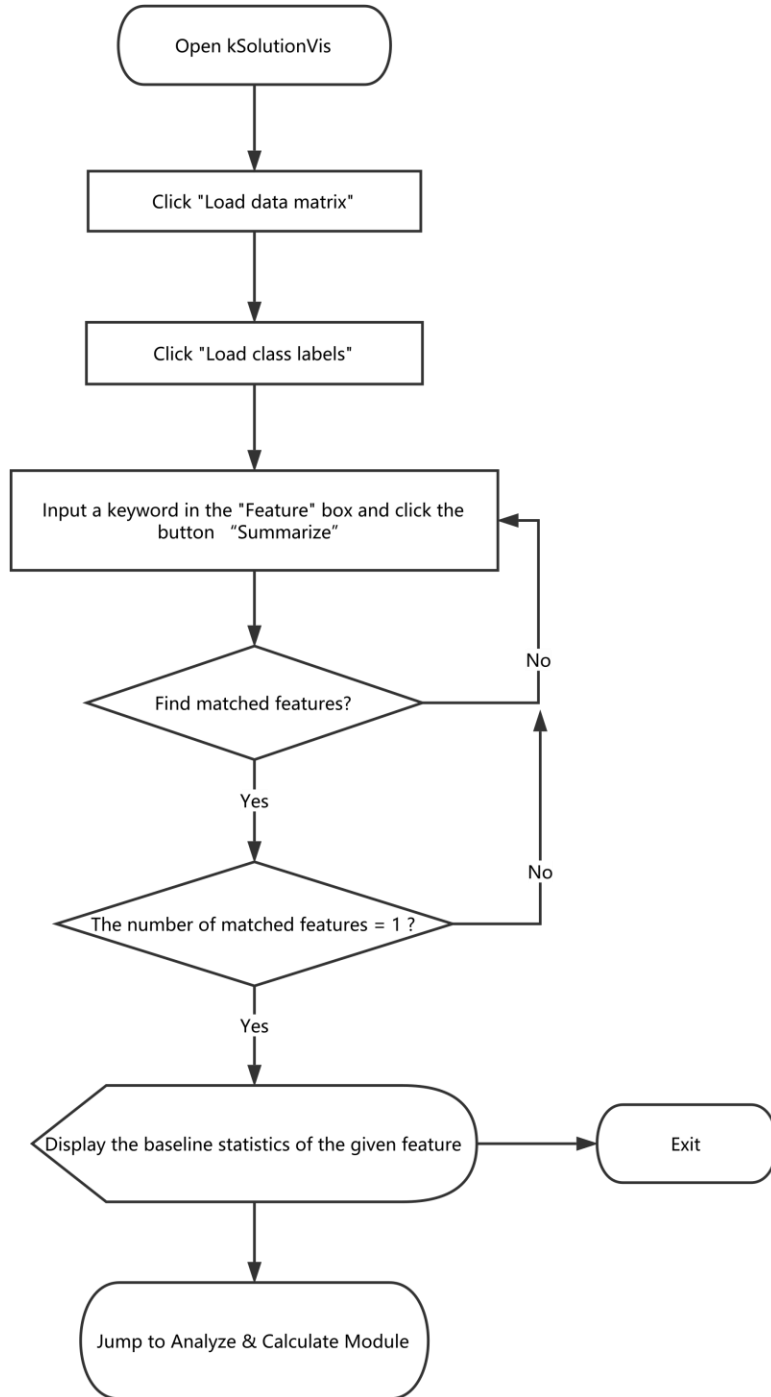
74 Workflows of kSolutionVis

75 1. Overall workflow



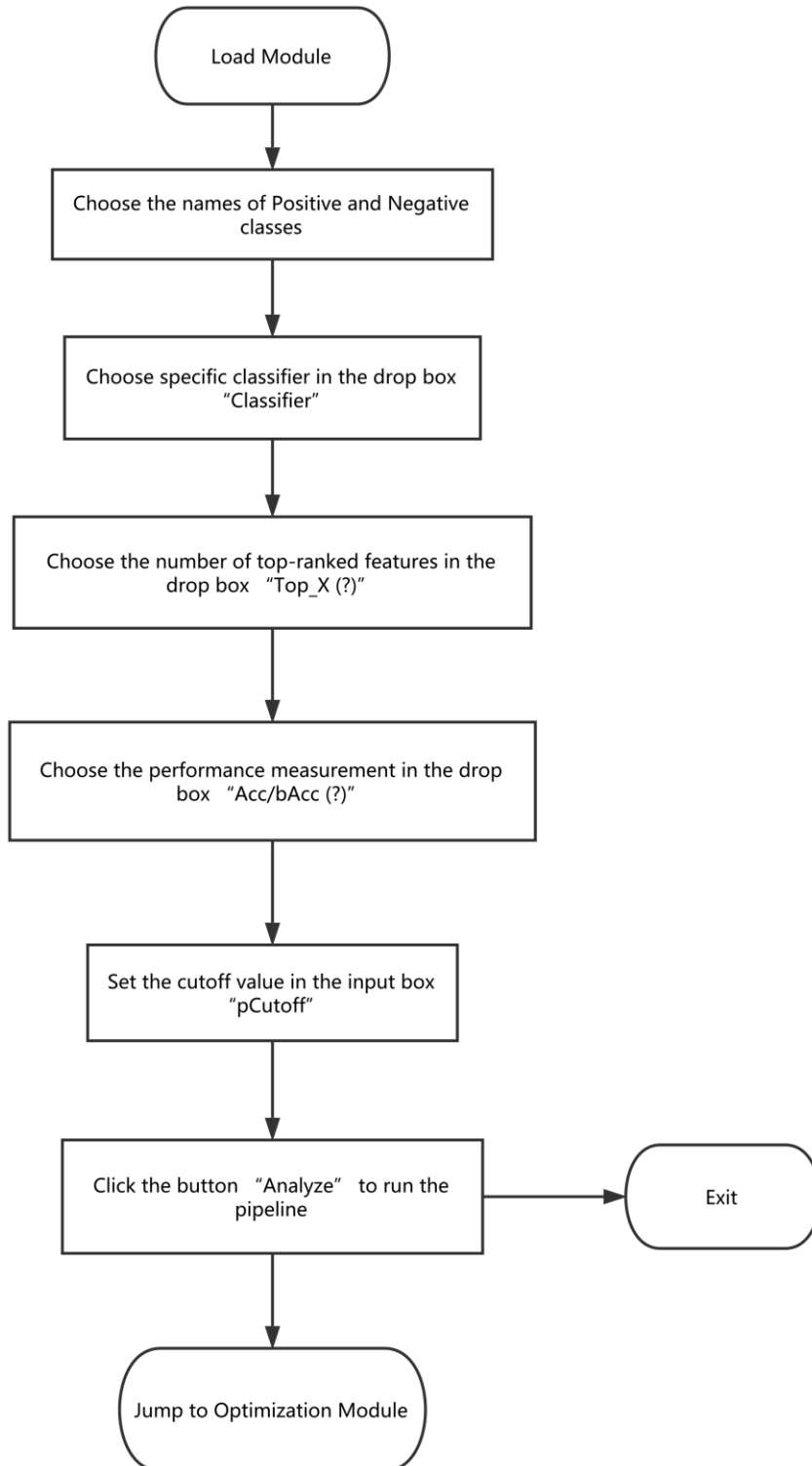
76
 77

2. Loading datasets and summarizing the features



78
79
80
81
82
83
84
85

3. Detecting biomarkers



86

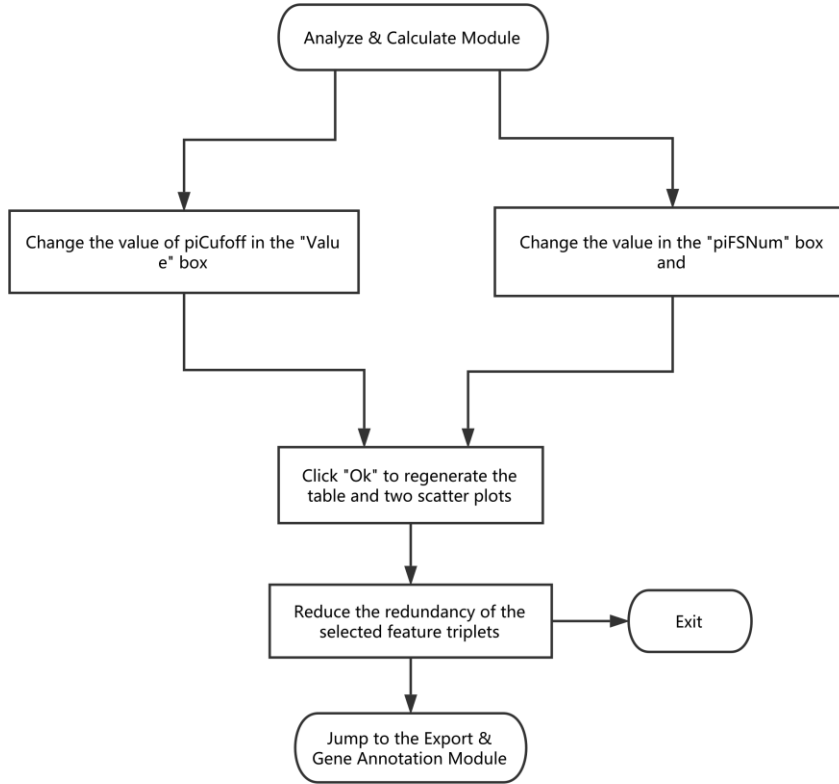
87

88

89

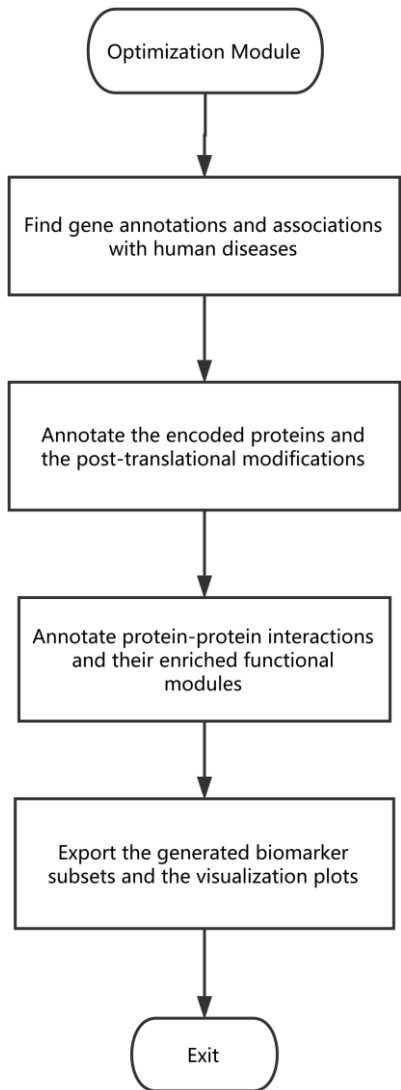
90

4. Visualizing the results



91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115

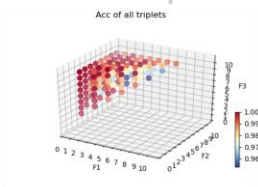
116 **5. Exporting the results**



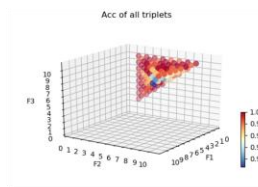
117

118 **Manual tuning of the 3D dot plots**

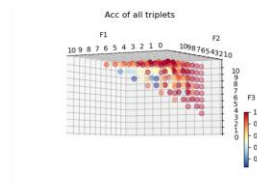
119 The user may tune the viewing angles of these 3D dot plots.



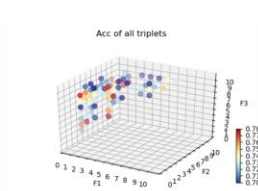
(a)



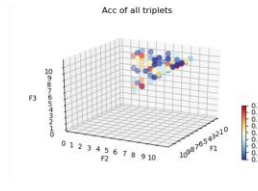
(b)



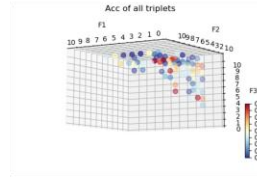
(c)



(d)



(e)



(f)

120
 121