Review of: "Feature Selection and Classification of Type II Diabetes on High Dimensional Dataset"

Manpreet Kaur[1]

1 Sant Longowal Institute of Engineering and Technology

Potential competing interests: No potential competing interests to declare.

The authors have proposed a feature selection and classification method that is applicable to Pima Indian Type II Diabetes data. There are some issues:

1. The dataset, as mentioned above, is not high-dimensional data. Whether this method is applicable to high-dimensional data, and what is innovative in this method that makes it useful for high-dimensional data, needs to be addressed.

2. The authors have given a discussion of the Naive Bayes classifier, while others have also been used. If the authors recommend this classifier based on the results, what is so different about this algorithm that it gives better results than the other classifiers?

3. Figure 1 is without a caption. Also, in the figure, it is mentioned that the Genetic Algorithm is used for feature selection, while in the discussion part, there is no mention of this algorithm. Instead, the authors have given the names of some other algorithms which are available in a Python module.

4. Where Data Mining is used is not clear from the paper; it is mentioned in figure 3.

5. If the authors have used GA or any other feature selection method, how the results at table 1 for dimensionalities 3, 5, 7, and 9 are taken is not clear from the paper. Also, why the values 2, 4, 6, and 8 are not considered is not explained.

6. What the significance of table 2 is is not clear.