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

Kim Gaik Tay[1]

1 Universiti Tun Hussein Onn Malaysia

Potential competing interests: No potential competing interests to declare.

Recommendation: Reject submission due to poor writing quality across all sections, including the abstract, introduction, methodology, results, discussion, and conclusion.

Detailed Comments:

1. **Abstract:**

   - The abstract should include an introduction, problem statement, objectives, methodology, and results.
   - Currently, the abstract lacks clarity and does not effectively summarize the paper's content.

2. **Introduction:**

   - The first paragraph is confusing and does not provide a clear context for the study.
   - The introduction should introduce diabetes, differentiate between type 1 and type 2 diabetes, review related work, identify the research gap, and state the objectives and contributions of this work.

3. **Methodology:**

   - **Dataset Source:** The dataset's source must be described in detail.
   - **Feature Selection:**

     - There is a discrepancy in the methodology: Figure 1 mentions Genetic Algorithm (GA) for feature selection, while Section III refers to a statistical test. This needs to be clarified.
     - Specify if and how GA was used for feature selection, as the previous section mentions statistical tests.
     - Clearly state and describe the feature selection process.

   - **Classification Algorithms:**

     - Section IV lists SVM, NB, RF, LR, KNN, and GBC, but Figure 1 lists NB, J48 Graft, MLP NN, and MOE Fuzzy. Additionally, the results section only discusses NB. Clarify which classifiers were used.

   - **Classification Rate:** Define and discuss the classification rate.
   - **Performance Evaluation:** This section should follow the discussion on the classifiers (NB, J48 Graft, MLP NN, MOE Fuzzy) in Fig 1.

4. **Sections:**

   - **Section VI:** This section appears to be unused in the study. Remove or integrate relevant content.
   - **Section III:** This section should focus on the methodology. Move the coding details found in Section VII to the Appendix.

5. **Results:**

   - Define terms such as "support" using kfold cross-validation, "exactness," "review," and "backing" (ensure correct terminology is used).
   - Explain the criteria for selecting 2, 4, and 6 features.
   - The feature matrices are inconsistent: 2 feature spaces show 3 rows, 4 feature spaces show 5 rows, and 6 feature spaces show 7 rows. Provide a clear explanation for this inconsistency.
   - Discuss the comparative performance of different feature selection methods.

6. **Figures and Captions:**

   - Figure 4 is blurred and needs to be clear and legible.
   - Add captions to all figures generated by Python.

7. **Verification Results:**

   - Include verification results to support the study's findings.

Summary: The paper needs substantial revisions to improve clarity, methodology description, consistency in reported results, and overall writing quality. The current submission is difficult to follow due to poor organization, unclear methodology, and inconsistent terminology.