

Review of: Comparison of machine learning and deep learning techniques in promoter prediction across diverse species

Reviewed by: Dr. Manal Kalkatawi

1. Basic Reporting:

- The report is well-written in a clear language and a good structure and flow of information
- In the introduction, the paragraph that starts at line 89, it is mentioning some of the genomics applications and not covering all of them; such information should be clearly state

2. Experimental Design:

- Data extraction was not explained well and needed more details
- 90/10 data division for training and testing is subject to overfitting, and the performance of the testing, in this case, is questionable
- Random Forest (RF) is not a representation of ML models, so elaborate on why you selected it?
- Ceasing your experiments, as mentioned in section Comparison of Feature Encoding Techniques at line 226, because the training time was very high is not enough to draw your conclusion
- Regarding comparing with existing techniques, it is not a fair comparison in which you should unify the datasets and no just report the published results

3. Validity of The Findings:

- You have mentioned one of the challenges that were in previous studies is choosing the negative data, yet I do not think shuffling it would solve the
- I think having 1000 long sequence is too long to the extent that you might lose some of the important features and include irrelevant ones

4. General Comments:

- You have mentioned one of the challenges that were in previous studies is choosing the negative data, yet I don't think shuffling it would solve the
- Have you checked the sequences and removed the duplicates or 100% identical sequences?
- I would recommend a visual representation of your ML/DL models with the parameters that were used
- The subtitle "Feature extraction" might be changed to data representation or data preprocessing

5. Confidential notes to the editor:

- None

6. Decision:

- Minor accept, with handling the issues mentioned above