Review of: "Rules Extraction, Diagnoses and Prognosis of Diabetes and its Comorbidities using Deep Learning Analytics with Semantics on Big Data"

Mohammed Smaoui[1]

1 Kuwait University

Potential competing interests: No potential competing interests to declare.

The research seems to revolve around the use of machine learning, specifically automated ML model simulations, to evaluate functionalities of different cloud platforms when dealing with varying sizes of datasets for the diagnosis of diabetes mellitus (DM) and its comorbidities. The paper emphasizes:

1. Extraction of Customized Rules: These rules are specific to the diagnosis of diabetes and its associated diseases.
2. Evaluation of Data: Datasets, from large to small, were pruned and processed. The datasets varied in size and complexity, with some representing a single patient's profile and others representing a multitude of patient records.
3. Machine Learning Techniques: Deep Multinomial Learning was a primary method, with up to 100% accuracy being achieved for multitask classification.
4. Representation & Visualization: A series of figures and tables (like Figure 52, Figure 35, etc.) present results, diagnoses, and other crucial findings.
5. Case Analysis: A few patients were profiled in detail, showcasing how the tools and methods were applied in practice and the outcomes achieved.
6. Future Directions: The paper highlights potential future enhancements to the methodology, integrating new analytics models and accounting for missing or newly pertinent data, like that related to Covid-19.

Strengths of the Paper

The paper is thorough in its exploration of different datasets, from individual patient records to larger compilations. Utilizing advanced ML and NLP tools, especially the use of Deep Multinomial Learning, is commendable. The pruned big dataset and the detailed patient profiles provide clear insights into the research process and findings. Multiple figures are cited throughout the paper, suggesting that there's a strong emphasis on visual representation of findings, which can be extremely useful for readers. The paper doesn't just present findings; it also looks ahead, pointing out what's next for this research field.

Proposed Improvements

- Some terminologies and methodologies, like "LMHFL" or the relationship between the datasets, can be better elucidated for readers who might not be familiar with these specifics.
- Comparison with Existing Methods: While your techniques are innovative, a comparison or contrast with existing

methodologies can highlight the novelty and efficacy of your approach.

- Limitations Section: All studies have limitations. Whether it's the size of the dataset, the tools used, or the scope of the study, it's always good to clearly outline these.
- Real-world Implications: Discuss how this research can be applied in real-world settings. Can it be used in hospitals or clinics? How will it change the current diagnostic scenario?

  Discussion Section: Elaborate more on the findings. For instance, what does achieving up to 100% accuracy mean in a broader sense? Are there concerns about overfitting?
- Error Analysis: Delve deeper into the errors or anomalies found, like the "strong association between DM and its comorbidities". It's essential to understand not just where the model performs well, but also where it doesn't.

In summary, your paper seems to provide a good contribution to the realm of machine learning in medical diagnosis, especially related to diabetes and its comorbidities. However, like all research, there's always room to further refine and clarify to make the work more accessible and valuable to a broader audience.