Review of: "NER Sequence Embedding of Unified Medical Corpora to Incorporate Semantic Intelligence in Big Data Healthcare Diagnostics"

Nitsa Herzog[1]

1 Northumbria University

Potential competing interests: No potential competing interests to declare.

The paper represents an NLP approach to the diagnosis of diabetes mellitus.

I recommend improving the quality of the publication and clarity of the information.

The issues mainly found relate to machine learning approaches' research methodology and representation.

Can you please address the following questions:

What comorbidities of DM were used in the research?

How many comorbidities (classes) were included in machine learning tasks?

It needs to be clarified from your description what primary and secondary diseases are. It looks like, in one case you used multiclass and, in another one - binary classifications.

It needs to be made clear how many ML algorithms you use. Initially, you mentioned Bi-LSTM, but later, you compared the results of Naïve Bayes, logistic Regression and GBM.

What is the proposed NER structure? The section “Sequential NER Embedded Techniques” has minimal representation.

Best wishes with your research.