Research Highlights

Manuscript title: Predicting ICD-9 Code Groups with Fuzzy Similarity based Supervised Multi-Label Classification of Unstructured Clinical Nursing Notes.

Highlights of this research include:

- Design of a fuzzy token-based similarity matching approach for unstructured clinical data. This is used for deriving optimal data representations and eliminating anomalous or redundant data, due to which the cognitive burden is reduced, and an improvement in the clinical decision-making process is observed.
- Leveraging vector space and topic modeling to extract the rich patient-specific information available in unstructured clinical nursing notes to predict ICD-9 code groups accurately. Experimental results show that our proposed supervised learning models consistently outperformed the state-of-the-art models built on structured data.
- Design of an approach that utilizes unstructured clinical text for the development of CDSSs, thus eliminating the dependency on the availability of structured EHRs. This can be crucial in countries where structured EHR adoption is not widespread.
- Exhaustive benchmarking experimentation revealed that the proposed fuzzy similarity model consistently outperformed the state-of-the-art structured data based disease prediction model by 7.79% in terms of AUPRC and 1.24% in terms of AUROC.