Reading and Labeling Medical Discharge Summaries Using Artificial Intelligence to Improve Medical Discharge Process

Moraes C[1], Ribas G[1], Vilela DF[1], Cunha PLT[1]
[1]Unimed-BH, Belo Horizonte, MG, Brazil

Objectives

Discharge summaries of hospitalized patients often contain unstructured information that is difficult to represent and analyze.

In this study, we propose a machine learning solution for the automatic reading and labeling of medical instructions given to the patient at the time of hospital discharge. Elderly aged 90 years or older were selected as a case study, given their special need for medical assistance.

Methods

The discharge summaries of 642 patients in 2021 were analyzed, containing 294K sentences (Figure 1) with a long tail distribution and a mean of 55 words per sentence (Figure2).

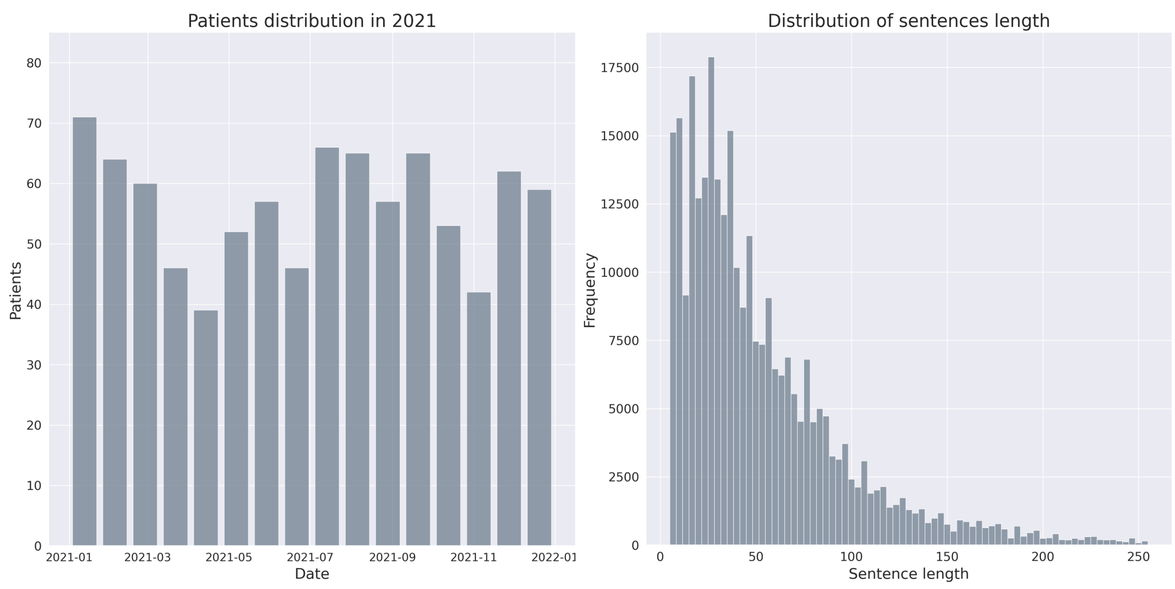

Figure 1. Data evenly distributed throughout the year collected for the work, ruling out the possibility of current seasonality in the sampled data. **Figure 2.** Distribution of number of words for sentence.

The database was submitted to a labeling process by a healthcare professional, classifying them into seven medical action items: appointment, lab, home care, medication, rehabilitation, recommendation, and transfer. 137K sentences (335 patients) were used in the models training and validation process.

The model's architecture was built from a pre-trained deep learning algorithm for natural language processing, connecting its outcome to the input of a classification algorithm for each label (Figure 3).

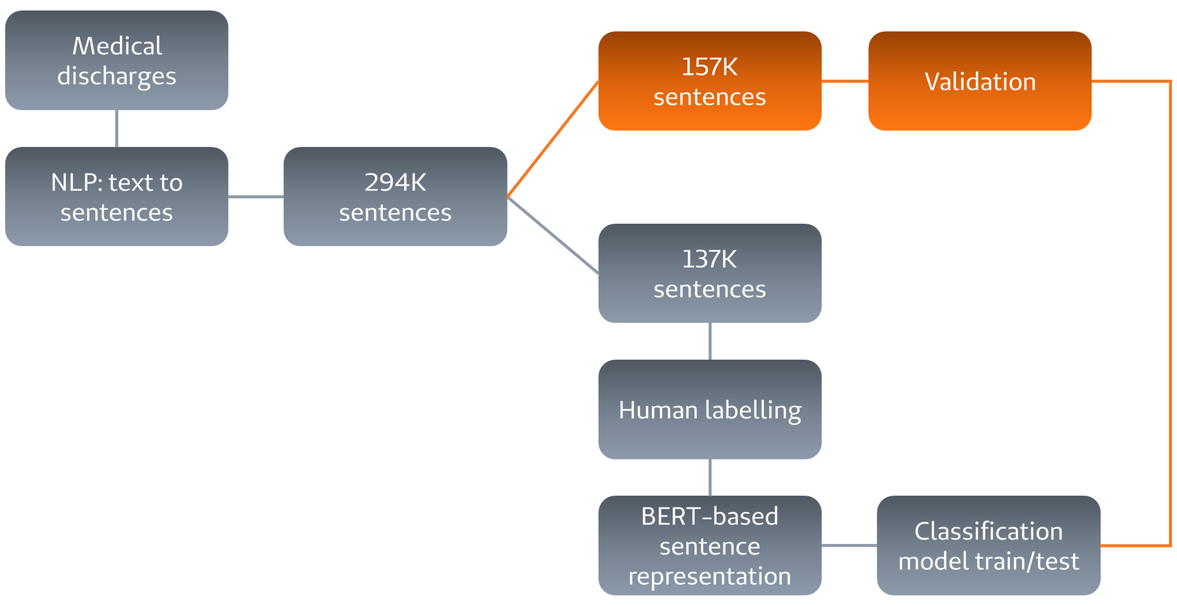

Figure 3. Conceptual model of the design of algorithms for training, testing and validating medical discharges.

Methods

In the test dataset (157K sentences, 335 patients), the labels, highly unbalanced (11.6% of positive instances), were distributed in: medication (7.35%), recommendation (2.06%), appointment (1.34%), home care (0.72%), lab (0.08%), rehabilitation (0.04%) and transfer (0.001%) (Figure 4).

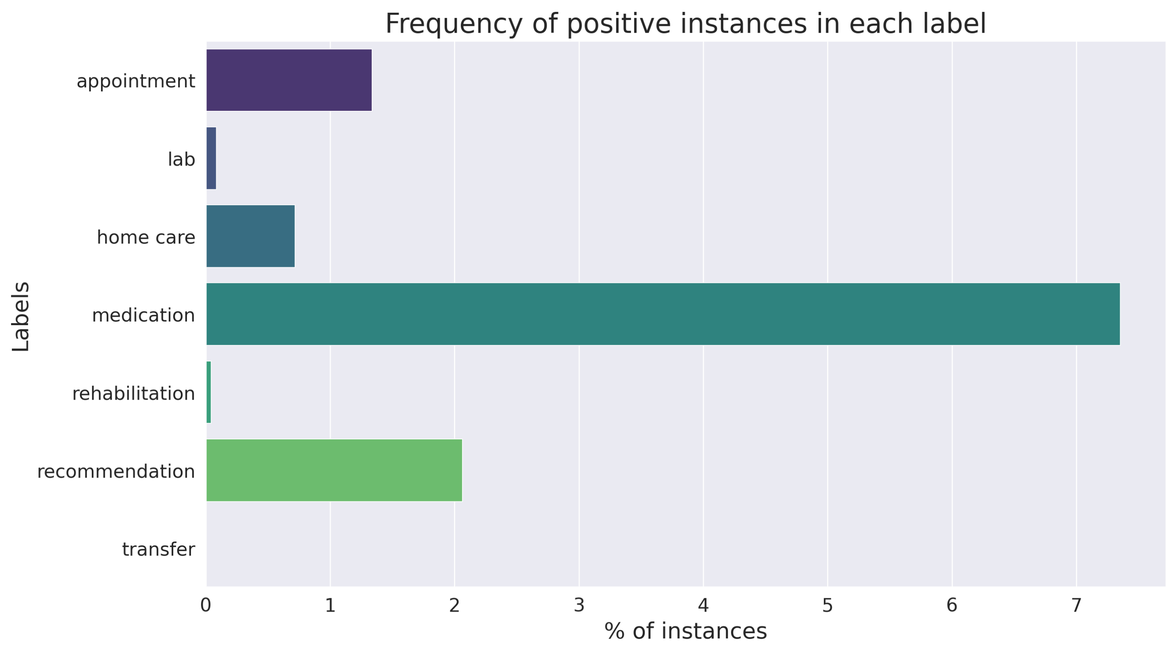

Figure 4. The graph shows the percentage frequency of each class within the 11.6% of positive instances found in the dataset.

Given the problem of disproportionate classes, sensitivity and area under the ROC curve were the main accurate metrics. The highest sensitivity was observed in transfer at 100% (AUC 82%), followed by lab at 95% (AUC 84%) due to the lower number of instances.

However, the wide vocabulary variation showed a lower sensitivity in recommendation at 83% (AUC 79%), although the lowest sensitivity was observed in home care at 72% (AUC 77%). The models showed an average sensitivity of 89% and AUC of 81% (Figure 5).

	Sensivity	Specificity	AUC ROC	Accuracy	Precision	Recall
Appointment	**0,93**	0,75	0,84	0,86	0,86	0,84
Lab	**0,95**	0,73	0,84	0,75	**0,60**	0,84
Home Care	0,72	0,82	0,77	0,79	0,74	0,77
Medication	**0,92**	0,63	0,77	0,85	0,79	0,77
Rehabilitation	0,86	0,77	0,81	0,77	**0,60**	0,81
Recommendation	0,83	0,75	0,79	0,80	0,79	0,79
Transfer	**1,00**	0,64	0,82	0,65	**0,51**	0,82
Average	0,89	0,73	0,81	0,78	0,70	0,81

Figure 5. Table containing the main metrics to evaluate each classification model. The color scale shows the best performance values of the models in strong purple decreasing in performance in green to the median values in orange.

Conclusion

The case study led to the creation of a labeling tool for actionable items in discharge summaries, which supported a team of health professionals to follow the care trajectory of the elderly patient, providing better assistance to their health.

Presented at ISPOR Europe 2022; Vienna, Austria and virtual; November 7-9, 2022