Artificial Intelligence algorithms and natural language processing for the recognition of syncope patients on Emergency Department medical records.

Franca Dipaola, Mauro Gatti[1], Anna Giulia Bottaccioli, Dana Shiffer, Maura Minonzio, Franca Barbic, Roberto Menè, Marco Anastasio[2], Elena Sini[2], Enrico Brunetta, Raffaello Furlan

Internal Medicine, Department of Biomedical Sciences, Humanitas University, Humanitas Research Hospital, Rozzano, Milan, Italy.

[1] IBM Italy

[2] ICT Department, Humanitas Research Hospital, Rozzano, Milan, Italy

Background. Hospital discharge diagnoses are frequently used to identify a study population. Currently, the International Classification of Disease, 9th revision, Clinical Modified (ICD-9-CM) is one of the most widespread coding system available. By means of language uniformity, ICD-9-CM enables data extraction from both administrative and clinical database thus permitting to perform large sample size studies or administrative analyses, quickly and inexpensively. However, robustness of results depends on the accuracy of ICD-9-CM code association with the disease. Syncope diagnosis is often an exclusion process, not supported by any biomarker or imaging test. Such sources of imprecision may reduce the diagnostic accuracy of ICD-9-CM 780.2 coding for syncope ("syncope and collapse"). Thus, we reasoned that an artificial intelligence based approach, i.e. the use of cognitive algorithms capable to automatically categorize and identify elements, would be remarkably useful in identifying true syncope patients from an administrative dataset.

Aim. To develop machine learning algorithms based on natural language processing which may enable the identification of patients affected by syncope in Emergency Department (ED), by using administrative data.

Methods. *Study design*: Retrospective, observational. *Study population*: Medical records of all patients evaluated at Humanitas Research Hospital ED between 1st December 2013 and 31st March 2014 (2013 dataset) and 1st December 2015 and 31st March 2016 (2015 dataset) were analyzed. Each dataset was extracted from the Humanitas Hospital electronic repository by IT department experts and was made available in the form of a Microsoft Excel spreadsheet.

By manual evaluation of medical records, syncope was identified if either the term "syncope" was clearly reported in the ED discharge diagnosis description or if the description of the episode agreed with the European Society of Cardiology guidelines definition of syncope. Clinical judgment of emergency physicians was considered the "gold standard" for syncope diagnosis. Subsequently, the 2013 and 2015 datasets were combined in a single dataset which was used to feed the machine learning algorithms. Data were split in training, validation and test data to prevent overfitting and ensure reliability. The methodology used is the Nested Cross Validation. Classification of records was performed by developing multiple feature extractor algorithms (both manual and automatic) and testing their performance in combination with multiple classifiers. For each combination of feature extractor and classifier we measured algorithm accuracy on the validation set using multiple accuracy measures. To get a *single measure* that could encompass the need to give more relevance to sensitivity (small false negative numbers) without glossing over precision (small false positives numbers), the F_3 was selected for giving triple importance to sensitivity over precision.

Results.

Table 1. Study population

	ED admissions (n)	**Pts with syncope (n)**	**Pts without syncope (n)**
2013 dataset	16070	251	15819
2015 dataset	16087	320	15767

Table 2. Algorithms performance

Classifier	**Feature Extractor**	**F_3**	**Sensitivity**	**Precision (PPV)**
Naïve Bayes	Humanitas key grams	68.10	70.58	51.80
Support Vector Machines	Humanitas key grams	78.79	92.43	33.84
Naïve Bayes	Normalized Gini Index	81.91	89.31	46.91
Support Vector Machines	Normalized Gini Index	82.14	91.4	42.8

We calculated that, to achieve 100% of sensitivity, that means no loss of any syncope patient, a 222.3 days/person time would be required for manually analyzing every single medical record out of the total 32157. In contrast, the present algorithm, with the best F_3 value and a 91.4% sensitivity, would enable to identify syncope patients

("true positives") by analyzing only 4% of medical records with an expected analysis time of 8.9 days/person.

Conclusions. The use of the present algorithm might result in a significant reduction in the required time to analyze large databases with low costs and high reproducibility (in Italian language). However, external validation on a different patient cohort is necessary to confirm and strengthen these results. Finally, a similar approach might be used to quickly address diagnosis and prognosis of other diseases in the ED.