Natural language processing and machine learning to assist radiation oncology incident learning

F. MATHEW [1], H. WANG[2] , L. MONTGOMERY[1] and J. KILDEA[1]

[1]Medical Physics Unit, McGill University, Montreal, QC, Canada [2]Unaffiliated, Montreal, QC, Canada

BACKGROUND

- Radiation oncology incident learning systems (ILSes) are tools to identify, report and learn from radiotherapy incidents.
- Staff report incidents in ILS with a free-text incident description.
- Investigators manually classify incidents according to oncology-specific taxonomies to find patterns that can facilitate follow-up actions.
- But the manual classification of such reports is a time-consuming and resource-intensive process and can hinder incident learning.
- Therefore, strategies to reduce the burden of manual incident classification are of interest to the radiation oncology community.

Incident description (example):

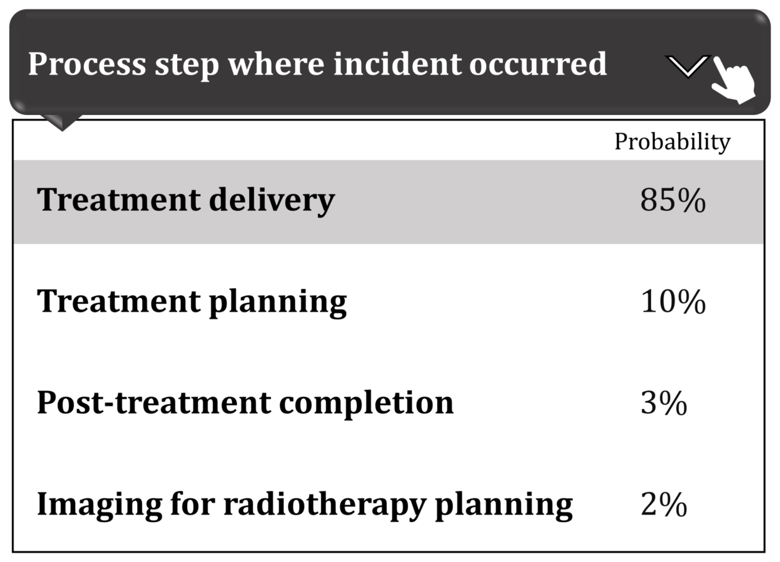
"Plan not ready. Pt was scheduld for 8:45 for plan 2, plan was not ready . Pt was called at 8:00 to come for 11:00. Plan ready @ 12:15."

Data elements of interest:

1. Process step where incident occurred (**8 label options**)
2. Problem type of the incident (**16 label options**)
3. Contributing factors of the incident (**25 label options**)

OBJECTIVE

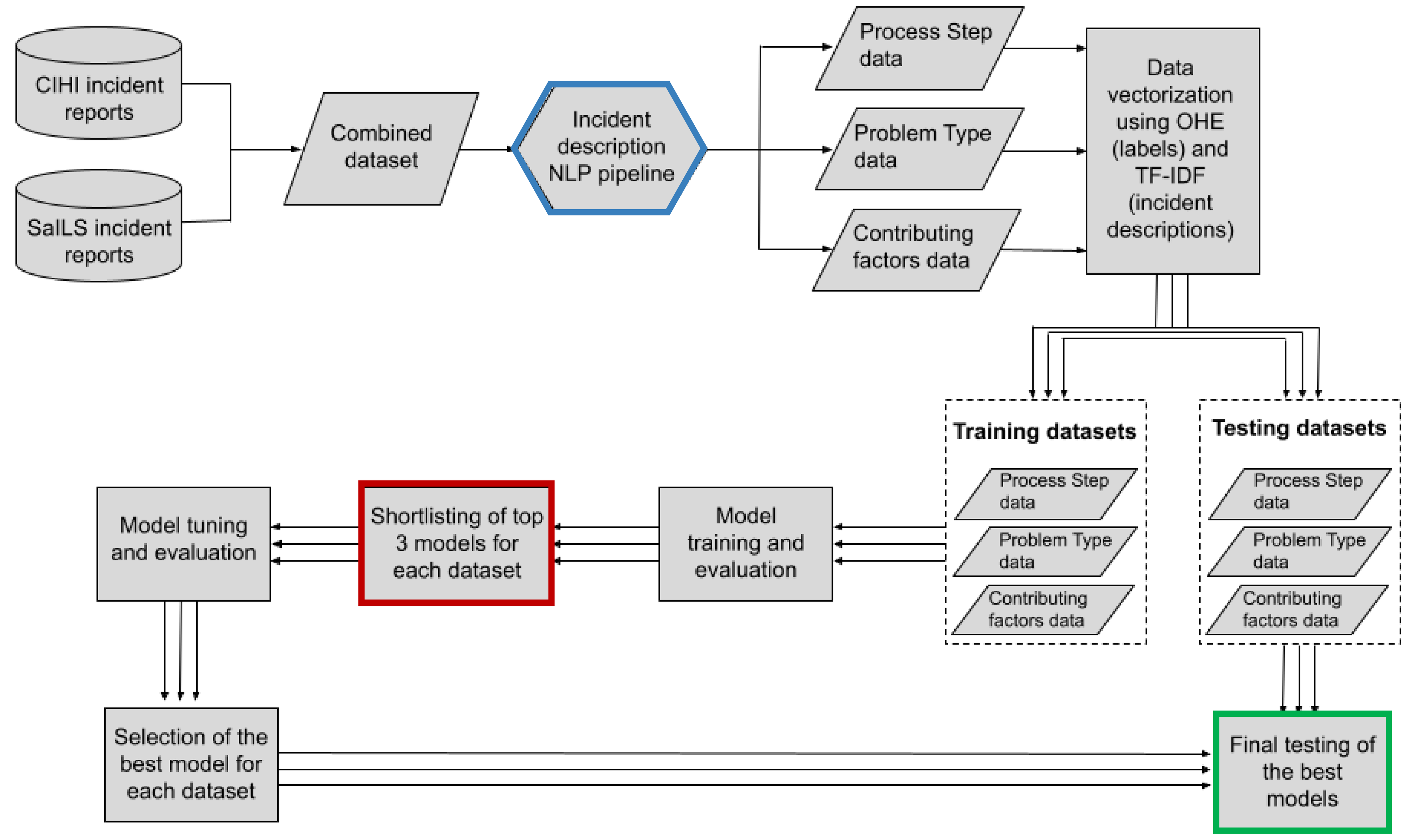
To assist radiation oncology incident classification using Natural Language Processing (NLP) and Machine Learning (ML) techniques by generating a drop-down menu of label recommendations, arranged according to their probabilities



A mock-up of the ranked drop-down list of labels for the process step data element of ILS

OUR NLP-ML PIPELINE

We gathered more than 6500 incident reports from Canadian Institute for Health Information (CIHI) and our local ILS (SaILS) databases.



A flowchart describing the stages of our NLP-ML model development. Simultaneous procedures for each data element are represented by parallel lines/arrows.

- ML algorithms (Estimators) were obtained from Scikit-learn library.
- Not easy to determine which model is best suited for our dataset; We tested them all.
- Mostly designed for binary classification.
- But we have a multi-label classification problem.

We extended them to support multi-label compatibility using two techniques: MultiOutputRegressor and RegressorChain

TrueLabelIndex score

Model prediction: [Label 5, Label 3, Label 1, Label 4, Label 2]

If True label (expert labelled value) = Label 3,

Then, TrueLabelIndex score = 2

NLP OF INCIDENT DESCRIPTIONS

1.Line-break removal
2.Translation
3.Punctuation and whitespace removal
4.Lowercase normalization
5.Autocorrection
6.Entity replacement
7.Stopword removal
8.Lemmatization

Original Incident description (example):

"INCCURATE TARGET . CTV WAS BIGGER THAN PTV, WAS NOTICED ONLY AT THE END OF PLANNING PROCESS, TARGET HAD TO BE CORRECTED AND PLAN REDONE."

Processed incident description (example):

"inaccurate target ctv big ptv notice end planning process target correct plan redone"

TOP 3 ML MODELS FOR EACH DATASET

Models trained on **process step** dataset	TrueLabelIndex score (1-8; best score = 1)
MultiOutputRegressor + Ridge	1.57
MultiOutputRegressor + Linear SVR	1.71
MultiOutputRegressor + SGD Regressor	2.07

TrueLabelIndex scores for the three best-performing models that were evaluated for classification of the process step data element on the training data

Models trained on **contributing factors** dataset	TrueLabelIndex score (1-25; best score = 1)
MultiOutputRegressor + SGD Regressor	4.32
MultiOutputRegressor + Lasso Lars	4.88
MultiOutputRegressor + Linear SVR	7.62

TrueLabelIndex scores for the three best-performing models that were evaluated for classification of the contributing factors data element on the training data.

Models trained on **problem type** dataset	TrueLabelIndex score (1-16; best score = 1)
MultiOutputRegressor + SGD Regressor	2.96
MultiOutputRegressor + Linear SVR	2.98
MultiOutputRegressor + Passive Aggressive Regressor	3.38

TrueLabelIndex scores for the three best-performing models that were evaluated for classification of the problem type data element on the training data.

FINAL TEST RESULTS OF THE BEST MODELS

Data element	Optimal ML model	**TrueLabelIndex score obtained with ML model** (Best score =1)
Process step	MultiOutputRegressor + Linear SVR	1.48 ± 0.03
Problem type	MultiOutputRegressor + Linear SVR	1.73 ± 0.05
Contributing factors	MultiOutputRegressor + Linear SVR	2.66 ± 0.08

The final test - TrueLabelIndex scores of the optimal, trained models for each of the three data elements, after hyperparameter tuning. Uncertainties are the standard error of the corresponding mean value.

CONCLUSIONS

- We built three different NLP-ML models (MultiOutputRegressor + Linear SVR) that can generate lists of label recommendations for the process step, problem type and contributing factors data elements in our ILS.
- On average, these models place the most appropriate label within the top three label suggestions.
- The trained models will be used to generate dropdown menus in our local ILS to semi-automate the incident investigation process.

ACKNOWLEDGEMENTS

McGill

CIHI ICIS Canadian Institute for Health Information / Institut canadien d'information sur la santé

CPQR Canadian Partnership for Quality Radiotherapy / PCQR Partenariat canadien pour la qualité en radiothérapie

Centre universitaire de santé McGill / McGill University Health Centre