# Development of a dataset and prospective evaluation of a model to identify high quality papers on the clinical impact of pharmacist interventions.

*Maxime Thibault, B.Pharm, M.Sc*[1] *and Cynthia Tanguay, B.Sc, M.Sc*[2]
[1] *CHU Sainte-Justine Department of Pharmacy and Research Center* [2]*CHU Sainte-Justine Department of Pharmacy*

**Background.**
The clinical role of pharmacists is expanding. They aim to implement evidence-based interventions that improve patient outcomes. The number of studies evaluating the impact of pharmacists is overwhelmingly increasing. This wealth of information hinders the identification of quality papers that describe their impact on patient outcomes. In addition, there is a lack of a widely accepted controlled vocabulary in this field.[1,2] The objectives of this project were to **1)** create a dataset that can be used to train machine learning models capable of identifying high quality papers on pharmacist interventions **2)** develop a model for this task and **3)** prospectively evaluate the model.

**Methods.**
**1) Dataset creation.** Starting in November 2021, every week, an automated PubMed search was performed. The search query was developed to retrieve as much papers as possible about pharmacist interventions. To select the highest quality papers, we developed selection criteria based on guidelines for the reporting of studies on this topic.[3-6] Two reviewers independently evaluated the abstracts to determine if they met the criteria. Discrepancies were resolved by discussion until consensus was reached.
**2) Model training and selection.** After one year, this dataset was used to train models. The titles and abstracts of all papers were used as a feature and the labels were whether each paper met the criteria. The dataset was chronologically split (to mimic the chronological publication of papers in real life) into training and validation (80%) and test (20%) partitions. The models were non-neural classifiers (logistic regression, support vector machines, random forests, AdaBoost and k-nearest neighbors) applied to the output of either Latent Dirichlet Allocation or Latent Semantic Indexing, as well as a multinomial naive Bayes classifier applied on word count vectors. A pre-trained transformers-based model (BioBERT-Base v1.2)[7], fine-tuned on the dataset, was compared to these models. Because the objective was to avoid missing relevant papers, maximum recall was defined as the evaluation objective. The training and validation partition was used to perform 5-fold cross validation with a grid search to optimize hyperparameters for the non-neural models. The hyperparameters of the transformers-based model were optimized on the training and validation partition using Ray Tune.[8] The decision threshold of the models was adjusted to maximize recall by performing a 5-fold procedure in which, for each fold, the threshold that resulted in a recall of 0.99 or more on the validation data was calculated, and then averaged across all folds. Performance on the test partition for the best non-neural and neural models was calculated using a 0.5 decision threshold (non-adjusted) and the adjusted threshold to select the final model.
**3) Prospective evaluation.** The final model was used prospectively from October 30, 2022, to March 19, 2023. Abstracts returned by the automated search continued to be evaluated every week by the reviewers and predictions were collected in the background. The model was updated weekly with new data. The decision threshold was also updated weekly by performing the same 5-fold procedure. The updated threshold was used to make predictions on new abstracts the following week. The precision and recall of the model during prospective use were calculated. A comparison to various manual PubMed searches made on the same period was performed as a baseline, to determine if a similar paper selection could be achieved simply by tailoring the search strategy to identify high quality papers.

**Results.**
**1)** On March 20, 2023, the dataset contained 5679 papers of which 395 (7%) met the criteria. The kappa between reviewers was 0.777. **2)** The data used to train and evaluate models in October 2022 contained 3933 papers of which 275 (7%) met the criteria. The best non-neural model was the multinomial naive Bayes classifier with an alpha of 0.2. This model showed a test precision of 0.406 and recall of 0.796 before threshold adjustment. After threshold adjustment, precision was 0.205 and recall was 0.840. The transformers-based model, with 4 training epochs and a $1x10^{-5}$ learning rate, showed a test precision of 0.717 and recall of 0.796 before threshold adjustment, and a precision of 0.134 and a recall of 0.981 after threshold adjustment. **3)** The transformers-based model was selected for prospective evaluation. During this period, 1631 papers were evaluated and 114 (7%) met the criteria. The best tailored PubMed search showed a precision of 0.068 and a recall of 0.798. The model showed a precision of 0.310 and a recall of 0.974. If the reviewers had only evaluated the papers selected by the model, 1273 abstracts (78.1%) would not have needed review, representing a considerable workload reduction. Only 3 papers out of 114 (2.6%) would have been incorrectly excluded.

**Conclusion.**
We have developed a new dataset containing high-quality papers about the impact of pharmacist interventions on patient outcomes. The papers were selected in accordance with published guidelines and independently evaluated by two reviewers. From this dataset, we trained a transformers-based model and showed that it can be used to significantly reduce the workload of labeling new papers, while incorrectly excluding a minimal number of papers. The dataset is updated weekly and freely available online, as well as the code used to perform all experiments in this project.[9] With the help of this model, we have implemented a weekly newsletter ( https://impactpharmacy.net/ ) to broadcast papers meeting the criteria to clinical pharmacists. This will help clinical pharmacists determine which interventions will result in the best outcomes for their patients.

References

1. Tonin FS, Gmünder V, Bonetti AF, Mendes AM, Fernandez-Llimos F. Use of 'Pharmaceutical services' Medical Subject Headings (MeSH) in articles assessing pharmacists' interventions. Explor Res Clin Soc Pharm. 2022 Aug 20;7:100172. doi: 10.1016/j.rcsop.2022.100172.
2. Fernandez-Llimos F, Desselle S, Stewart D, Garcia-Cardenas V, Babar ZU, Bond C, Dago A, Jacobsen R, Nørgaard LS, Polidori C, Sanchez-Polo M, Santos-Ramos B, Shcherbakova N, Tonin FS. Improving the quality of publications in and advancing the paradigms of clinical and social pharmacy practice research: The Granada Statements. Explor Res Clin Soc Pharm. 2023 Jan 20;9:100229. doi: 10.1016/j.rcsop.2023.100229.
3. Bougioukas KI, Bouras E, Apostolidou-Kiouti F, Kokkali S, Arvanitidou M, Haidich AB. Reporting guidelines on how to write a complete and transparent abstract for overviews of systematic reviews of health care interventions. J Clin Epidemiol. 2019 Feb;106:70-79. doi: 10.1016/j.jclinepi.2018.10.005.
4. Clay PG, Burns AL, Isetts BJ, Hirsch JD, Kliethermes MA, Planas LG. PaCIR: A tool to enhance pharmacist patient care intervention reporting. J Am Pharm Assoc (2003). 2019 Sep-Oct;59(5):615-623. doi: 10.1016/j.japh.2019.07.008.
5. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, Schulz KF; CONSORT Group. CONSORT for reporting randomised trials in journal and conference abstracts. Lancet. 2008 Jan 26;371(9609):281-3. doi: 10.1016/S0140-6736(07)61835-2.
6. Page M J, McKenzie J E, Bossuyt P M, Boutron I, Hoffmann T C, Mulrow C D et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021; 372:n71. doi:10.1136/bmj.n71.
7. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020 Feb 15;36(4):1234-1240. doi: 10.1093/bioinformatics/btz682.
8. Liaw R, Liang E, Nishihara R, Mortiz P, Gonzalez JE, Stoica I. Tune: A Research Platform for Distributed Model Selection and Training. arXiv:1807.05118 [cs.LG]. doi: 10.48550/arXiv.1807.05118.
9. https://github.com/urppchusj/impactpharmacie/