SPCVerf: Medication Related Concern Extraction and Verification Pipeline

Veronika Kukk, 2nd year Computer Science MSc
Supervisors: Hendrik Šuvalov (MSc), Uku Kangur (MSc)

Medication Safety Monitoring

Medication safety monitoring through voluntary reporting systems **suffers from under-reporting**, making it challenging to estimate the frequency and severity of adverse drug reactions. People often share medication related concerns online, which could be used for safety improvement. Due to the large scale of this **online data**, automatic information extraction and analysis methods are required.

Proposed Solution

Pipeline SPCVerf extracts medication related concerns from patients' texts, and verifies whether these concerns are mentioned in medication leaflets (SPC).

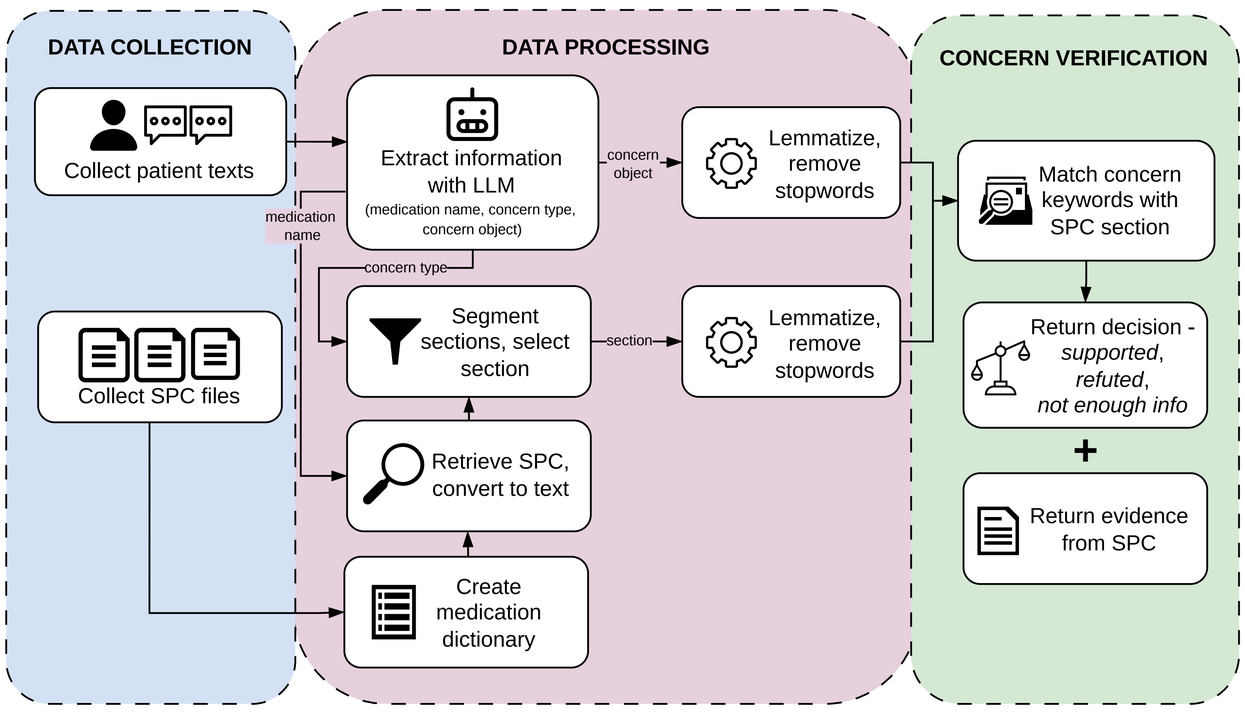

Figure 1. Design overview of SPCVerf.

Data

Patients' texts are extracted from kliinik.ee, a popular Estonian health counseling portal, where anonymous users submit questions that healthcare professionals publicly answer.

Summary of product characteristics (SPC) is a medication leaflet for healthcare professionals. Collected through the Register of Medicinal Products.

Figure 2. Illustration of SPC file contents.

Design

SPCVerf combines large language model based information extraction with automated concern verification using keyword searching.

Information extraction: Using gpt-4.1-mini model.

Concern Verification: Detemine whether the SPC supports, refutes, or does not contain the concern. The concern keyword search allows slight misspellings, word order changes and small gaps in phrases, uses disease synonyms from Estonian Wordnet, and checks for nearby negation.

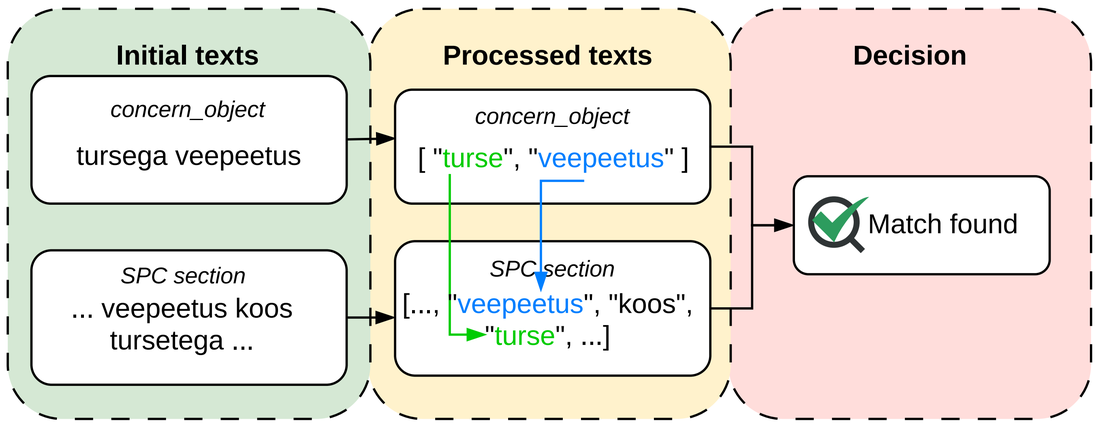

Figure 3. An example of concern matching where the words are in different order, contain additional words, and are in different declensions.

Majority of the keywords within three words => Match found
If no nearby negation => *Supported*
If nearby negation => *Refuted*
In other cases => *Not found majority concern*

Evaluation

Information Extraction

Manually annotated testset of 200 concerns.

Table 1. Evaluation of GPT model's information extraction ability for *concern_type*.

Concern type correctness	Count
CORRECT	120
INCORRECT	57
OTHER	23

Table 2. Evaluation of GPT model's information extraction ability for *concern_object*.

Concern object correctness	Count
CORRECT	105
INCORRECT	95

concern_type (enum) accuracy 60%
concern_object (string) accuracy 52.50%

Concern Verification

Medical expert annotated testset of 250 concerns.

Differences in SPCVerf and medical expert were due to concern objects that use non-medical language (e.g. "worms"), narrower expressions of a condition (headache instead of ache), and the inclusion of multiple concerns in a single instance (e.g. "back pain and flu").

Table 3. SPCVerf decision labels and distribution of the checkable concern types on the whole dataset.

	concern_type					
Decision label	undesirable effect	pregnancy	interaction	indication	contraindication	Total
MED_NOT_FOUND	3786	1184	484	2591	1811	9856
NOT_FOUND_MAJORITY_CONCERN	2986	989	567	1427	989	6958
SUPPORTED	744	6	93	258	58	1159
REFUTED	0	174	0	17	1	192
SPC_CANNOT_PARSE	45	22	14	50	22	153
SPC_SECTION_NOT_FOUND	92	1	8	4	2	107
Total	7653	2376	1166	4347	2883	18425

Table 4. Concern verification decision results on reduced testset for SPCVerf, GPT model and MedGemma model.

	SPCVerf			gpt-4.1-mini			medgemma-27b-text-it			
Decision label	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Support
NOT_ENOUGH_INFO	0.68	0.95	**0.79**	0.69	0.73	0.71	0.73	0.80	0.77	128
SUPPORTED	0.70	0.21	0.33	0.47	0.42	0.45	0.56	0.50	**0.53**	66
REFUTED	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6

Ablation Study

Testing LLMs for concern verification. The concerns, which were extracted in the data processing part, were used in a new prompt without the patient's original text. Each prompt contained information for one concern in JSON format: *medication_name*, *concern_type*, *concern_object*.

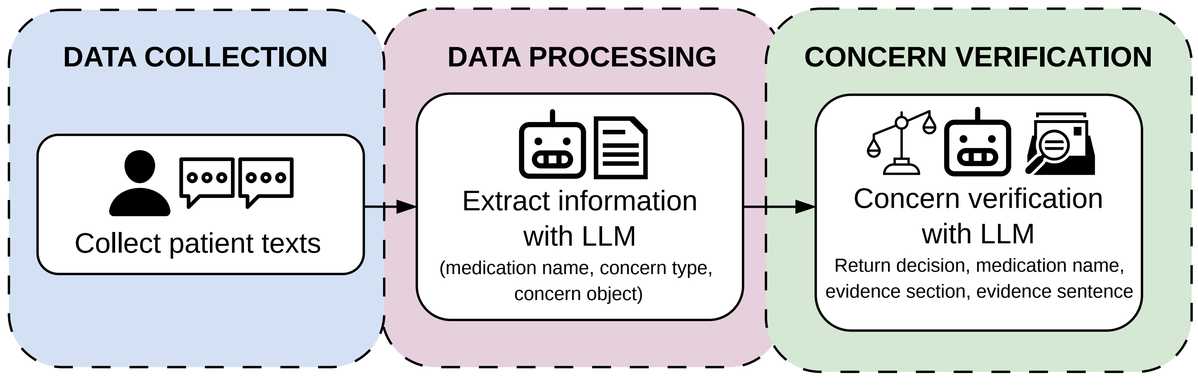

Figure 4. Overview of the concern verification with LLMs.

+ Pros:
- not tied to one source,
- generalizing ability,
- verifying non-specific or non-medication related concerns,
- semantics.

- Cons:
- hallucinations,
- medically inaccurate,
- contradictions for evidence and decision,
- black-box, not explainable.

Conclusions

Overall, a robust NLP-based concern verification system grounded in medication information leaflets can achieve performance comparable to LLM-based approaches while relying on verifiable, authoritative medical sources rather than implicit model knowledge.

Future research is needed to improve information extraction and concern matching methods, such as using multiple evidence sources, local LLMs for information extraction, phonetic spellings, and multiword phrase search.