Exploring the Role of Generative AI in Constructing Knowledge Graphs for Drug Indications with Medical Context

NeuroDKG predicates

Predicate	Frequency	Definition
contains	2	defines the drug contents of a combination drug
disease	176	The disease which belongs to the DrugDiseaseTargetGroup context. The disease which the aforementioned drug treats.
drug	174	the drug which belongs to the DrugDiseaseTargetGroup context
hasAgeGroup	43	relates to the age group of the target group that is appropiate for drug prescription e.g. between 6-12 years old
hasComorbidity	8	defines a comorbidty that the target group must have to be prescribed the drug of context
hasCurrentMedication	18	defines the current medication the target group must be taking in order to be prescribed the drug of context
hasMinAge	23	relates to the minimum age of the target group which can be prescribed this drug of context e.g. adult minimum age is 18
hasSymptom	149	defines a symptom that the target group presents which is also treated by the drug of context
hasTherapy	9	defines an alternative therapy which is not pharmacological that the target group needs to be receiving in order to be eligible for the Drug of the context
hasTreatment	17	defines a treatment that the targetGroup needs to be receiving in order to be eligible for the Drug of the context.
is_a	173	
responseStatus	14	relates to treatment response to other medications indicated for the disease of context. E.g. if the drug of context is a second-line treatment that is only administered when the target group is treatment resistant to prior medication
targetGroup	133	defines the target group which belongs to the DrugDiseaseTargetGroup context.
treatmentDuration	7	relates to the time period in which the drug should be adminstered for e.g. acute, maintenance
type	175	type of treatment i.e. indication, symptomatic relief, prevention, adjunctive therapy

Prompt 1

Paragraphs (separated by a full stop): {" ".join(newSentences)}

Instructions:

Inside each Paragraph, identify all triples and output a JSON dictionary with Paragraphs as keys and lists of triples formatted as [SUBJECT, PREDICATE, VALUE] as values.
Only output the resulting JSON.

Prompt 2

Paragraphs (separated by " | "): {" | ".join(newSentences)}

Instructions:
Inside each Paragraph, identify all context words/phrases and output a JSON dictionary with Paragraphs as keys and lists of {{[context type]: [corresponding words/phrases as a list of values (more than one possible)]}} as values.
Other context types are self-explanatory.
Only output the resulting JSON.

Prompt 3

Paragraphs (separated by " | "): {" | ".join(newSentences)}
Instructions:
Inside each Paragraph, identify all context words/phrases (allowed context types: "age group", "co-morbidity", "symptom", "co-therapy", "adjunct therapy", "past therapies", "treatment duration", "conditional", "co-prescribed medication", "genetics", "temporal aspects") and output a JSON dictionary with Paragraphs as keys and lists of {{[context type]: [corresponding words/phrases as a list of values (more than one possible)]}} as values.
Definition of context types:
"conditional" - a statement about when the medication is appropriate to use
"target" - the condition (symptom or illness) that is intended to be treated
"co-prescribed medication"- drugs commonly prescribed together with the given drug (not therapeutic procedures! -> for that use "co-therapy")
"co-therapy" - procedures or therapies that should be applied in combination with the drug (not medications or substances! -> for that use "co-prescribed medication")
"co-morbidity" - diseases or conditions that commonly occur together (with a target condition) in the same patients
"genetics"- particular genetic strains of a disease
"temporal aspects"- information which explains at what life stage, disease stage, or treatment phase a drug should be administered
Other context types are self-explanatory.
Only output the resulting JSON.

Prompt context types

Medical Context	Description
age group	Refers to information related to the age group of individuals affected by a medical condition or receiving treatment.
co-morbidity	Refers to diseases or conditions that commonly occur together with a specific target condition in the same patients.
symptom	Refers to specific indications or symptoms associated with a medical condition.
co-therapy	Refers to procedures or therapies that should be used in combination with a particular treatment.
adjunct therapy	Refers to additional therapies or treatments used alongside the primary treatment.
past therapies	Refers to previous or historical treatments that a patient has undergone.
treatment duration	Refers to the duration or length of time for which treatment is recommended or prescribed.
conditional	Refers to statements or conditions that determine when the medication or treatment is appropriate to use.
co-prescribed medication	Refers to medications that are commonly prescribed together with a specific drug.
genetics	Refers to genetic factors or specific genetic strains associated with a disease.
temporal aspects	Refers to information related to the timing or stages of a disease or treatment, such as at what life stage or disease stage a drug should be administered.

Context Types

Label	NeurodKG Mapping	Frequency	Prompt Freq.
Age group	hasAgeGroup	43	
co-morbidity	hasComorbidity	8	
symptom	hasSymptom	149	
co-therapy	hasTherapy	9	
Adjuct therapy	hasCurrentMedication	18	
Past therapies	—	—	
Treatment duration	treatmentDuration	7	
conditional	—	—	
Co-prescribed medication	—		
genetics	—		
Temporal aspects	—		
target	disease	176	

Results Statistics

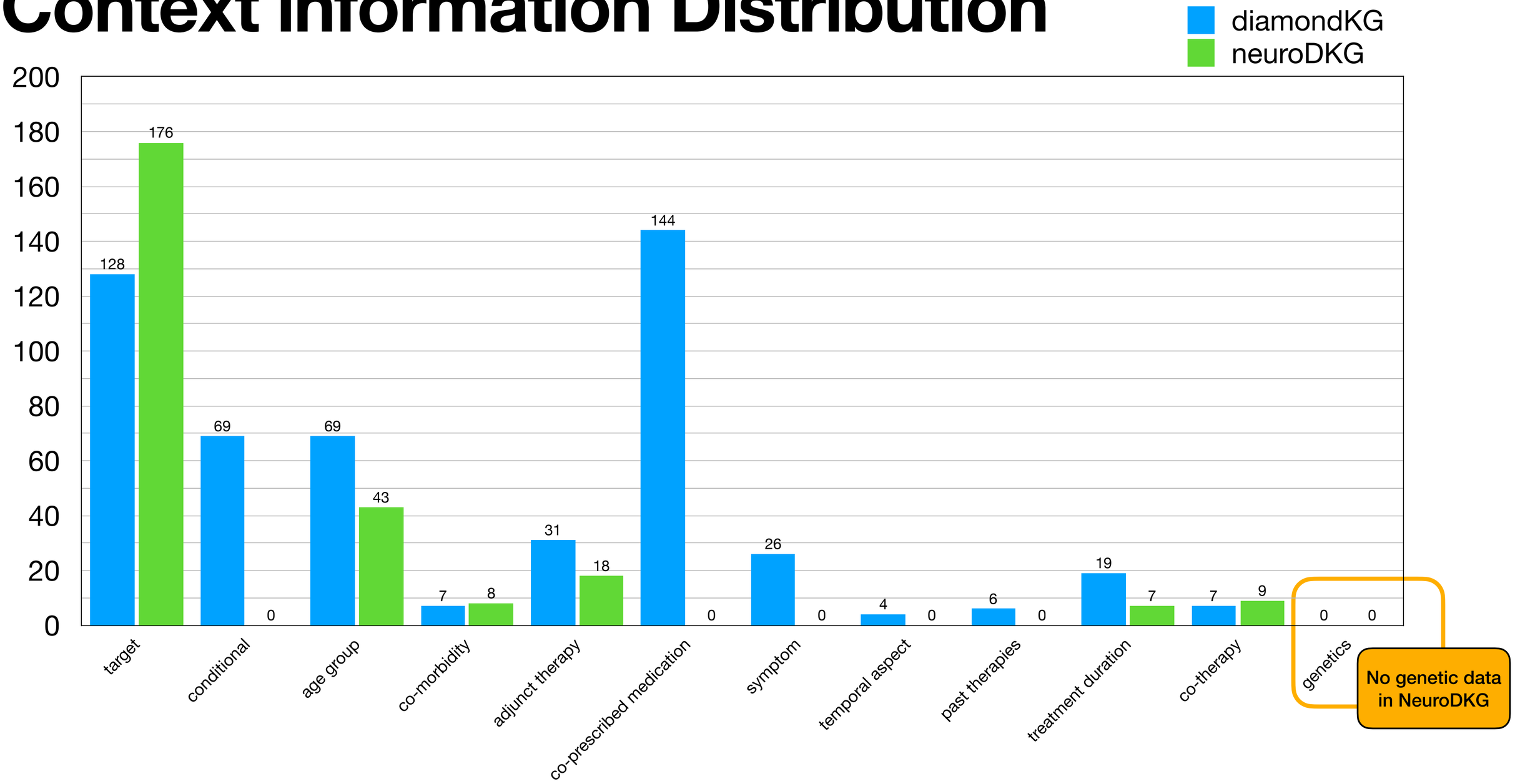
Prompt	Total Triples	Average Triples P/S	Sentence with context
NeuroDKG*	510	3.0	64.11%
Prompt 1	442	2.11	- - -
Prompt 2	911	4.82	- - -
Prompt 3	621	3.32	71.12%

* Restricted to only triples of interest (i.e., drug, disease and context types)

Results Statistics - Notes

- NeuroDKG has been restricted, we only consider triples with predicates related to context types, diseases or drugs
- Prompt results split sentences differently (fixed for prompt_3 output)
- Prompt 3 results do not comprise “drug” information

Context information Distribution



Manual Annotation - Results

- Total number of (context_type, context_value) pairs: 614 (502 with one correct context_type)

- Total number of correct pairs: 390 (63.52%)
- Total number of wrong pairs: 112 (18.24%)
- Total number of pairs with wrong context_type: 60 (9.77%)
- Total number of missed pairs: 52 (8.47%)

Manual Annotation - Results

Accuracy (exact match)	0.6351791530944625
Hamming loss	0.038545059717698 16

	precision	recall	f1-score	Support
target	**0.81**	**0.79**	**0.80**	214
age group	**0.91**	**0.96**	**0.93**	71
co-morbidity	0.58	0.28	0.38	25
symptom	0.67	0.70	0.68	66
co-therapy	**0.73**	**1.00**	**0.84**	8
adjunct therapy	0.39	1.00	0.57	15
past therapies	0.14	0.17	0.15	6
treatment duration	0.59	0.76	0.67	17
conditional	0.56	0.77	0.65	65
co-prescribed medication	0.43	0.83	0.57	12
genetics	0.00	0.00	0.00	0
temporal aspects	0.40	0.67	0.50	3
micro avg	0.69	0.78	0.73	502
macro avg	0.52	0.66	0.56	502
weighted avg	0.72	0.78	0.74	502
samples avg	0.64	0.64	0.64	502