

CASE STUDY 1: KEYWORD: epilepsy

Metric	Value
GEO datasets processed	749
PubMed IDs linked	429
NCT trial IDs extracted	2
Average time required	~10 minutes
Linkage rate	~57%

CASE STUDY 2: triple negative breast cancer

Metric	Value
GEO datasets processed	1518
PubMed IDs linked	784
NCT trial IDs extracted	11
Average time required	~20 minutes
Linkage rate	~52%

CASE STUDY 3: placebo

Metric	Value
GEO datasets processed	1066
PubMed IDs linked	686
NCT trial IDs extracted	108
Average time required	~20 minutes
Linkage rate	~64%

COMPARISON WITH EXISTING TOOLS

Features	Trialstreamer	GEO2R	PubMed	ClinicalTrial.gov	Our Streamlit web-app
Data Input	Text query	GEO accession IDs	PubMed IDs	NCT IDs	Keywords
Links GEO to PubMed	No, it doesn't link.	No, it doesn't link	Yes, it has in the abstract	No, it doesn't link	Yes, it links
Links PubMed to NCT IDs	Yes, it links	No, it doesn't link	No, it doesn't link	Yes, it links	Yes, it links
User Interface	Web-based	web-based	web-based	Web-based	Streamlit web-based
Combines GEO-PubMed-NCT	No, it doesn't combine.	No, it doesn't combine	No, it doesn't combine	No, it doesn't combine	Yes, it combines
User Interface	Yes	Yes	Yes	Yes	Yes
Focus Area	Clinical trial mining only	Gene expression and methylation expression	Biomedical literature	Clinical trial informations	It includes gene expression, methylation, and clinical trial information. All in one platform

Github

<https://github.com/shweta-rai11/deploy>

Streamlit Web Application

<https://extractbiodata.streamlit.app/>