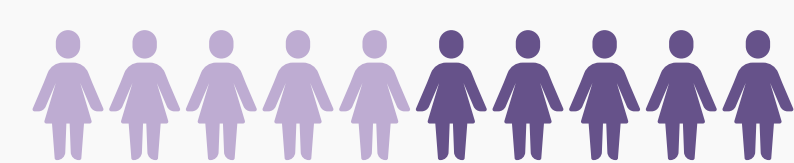
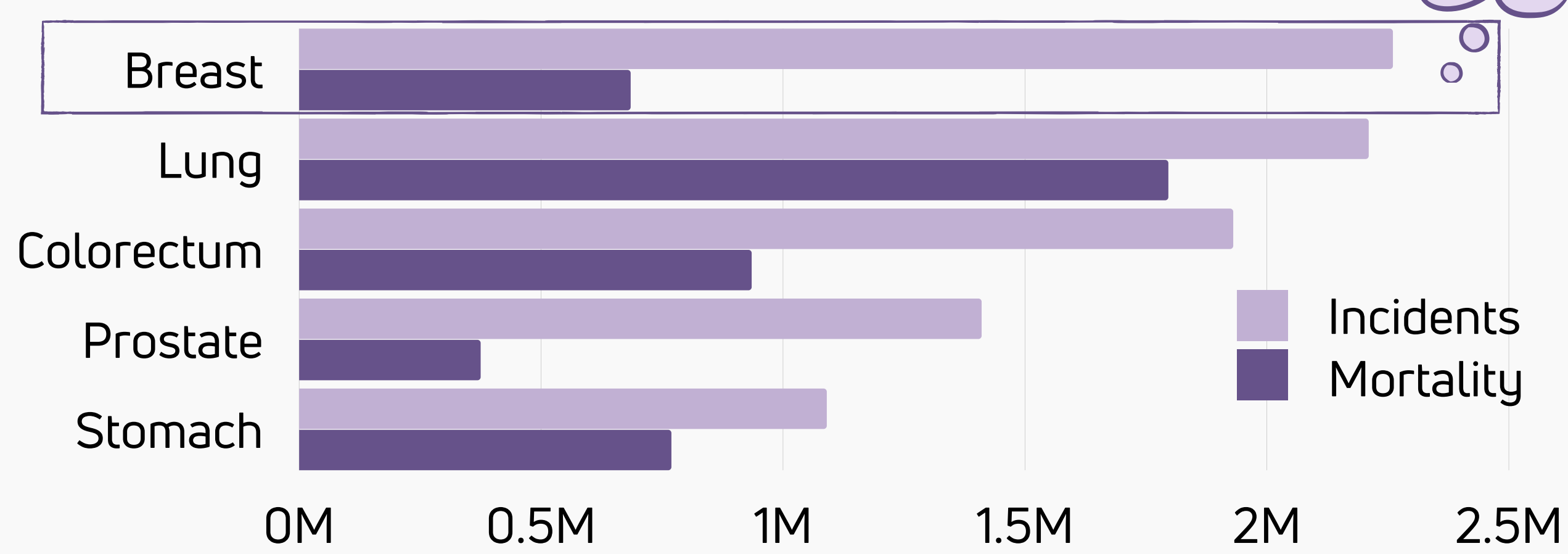


Obtaining insights for the emotional well-being of cancer patients using NLP on their social media and browser history

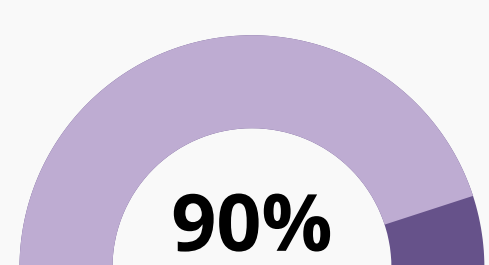
Georgia Pantalona, Filareti Tsalakanidou, Lazaros Apostolidis and Symeon Papadopoulos - CERTH/ITI

ABOUT BREAST CANCER

7.8B diagnosed in last 5 years



50% are less than 65 years old



90% live for 5 years or more

OBJECTIVES

Cancer can affect the patient's emotional well-being by causing:



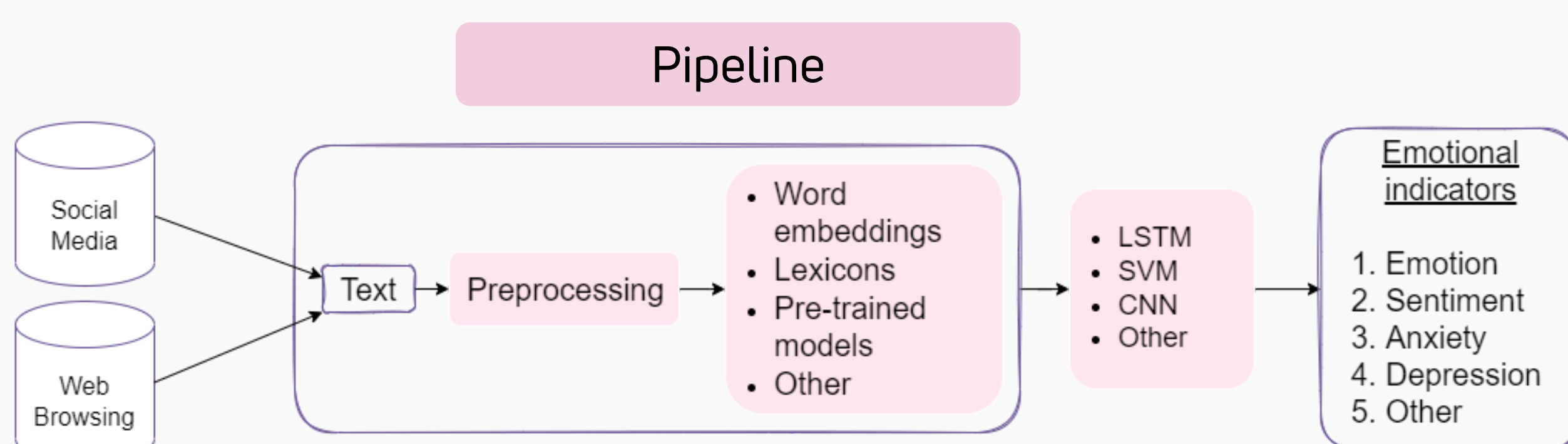
Good mental health is important for the **treatment trajectory**

Cancer **recurrence** is associated with **poor mental health**

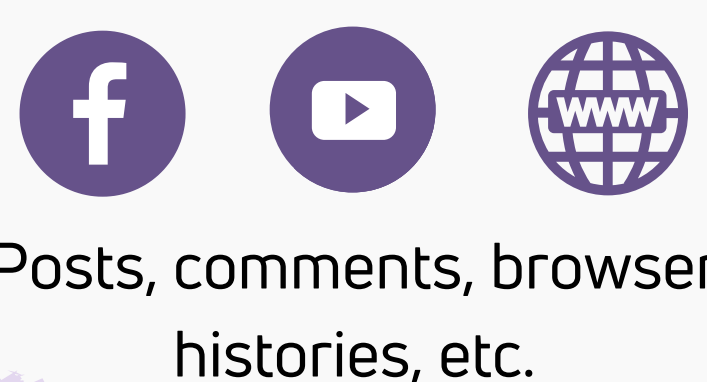
Proper monitoring of the emotional state can assist experts to be **aware and intervene** early if necessary

Goal: Extract insights on emotional well being.

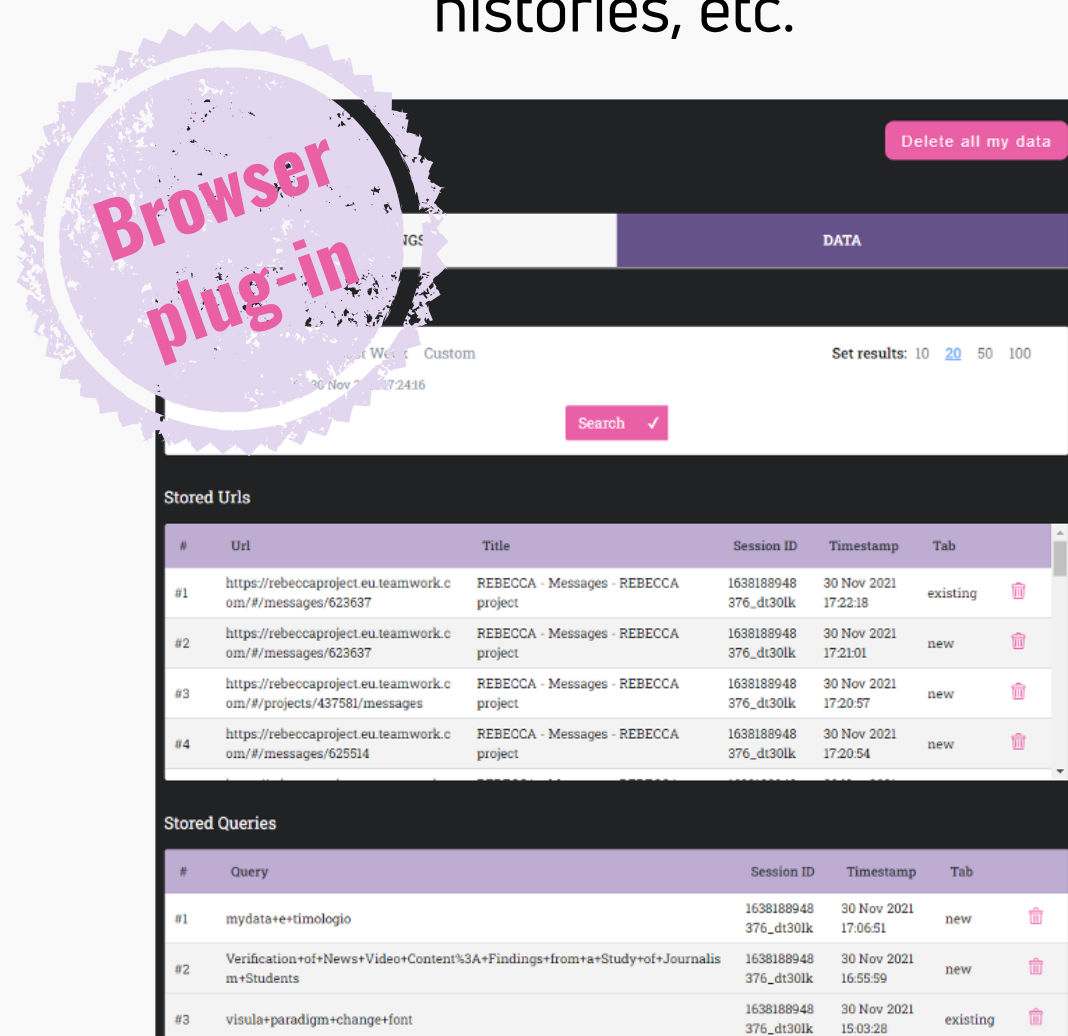
METHODOLOGY



Data Sources & Collection



Posts, comments, browser histories, etc.



1. Data are pseudonymized
2. Pause user tracking feature
3. Delete records feature

Preprocessing

Traditional Features

Presence based, frequency based, syntax related, lexicon based

Word/Sentence Representations

Static: Word2Vec, Glove, FastText, other
Contextualised: BERT, GPT, ELMo, other

Related work

- Classical machine learning has been utilised with good results
- Pre-trained models (e.g. BERT) show potential and surpass existing benchmarks
- Results highly depend on dataset and pre-processing

OUR PRELIMINARY WORK

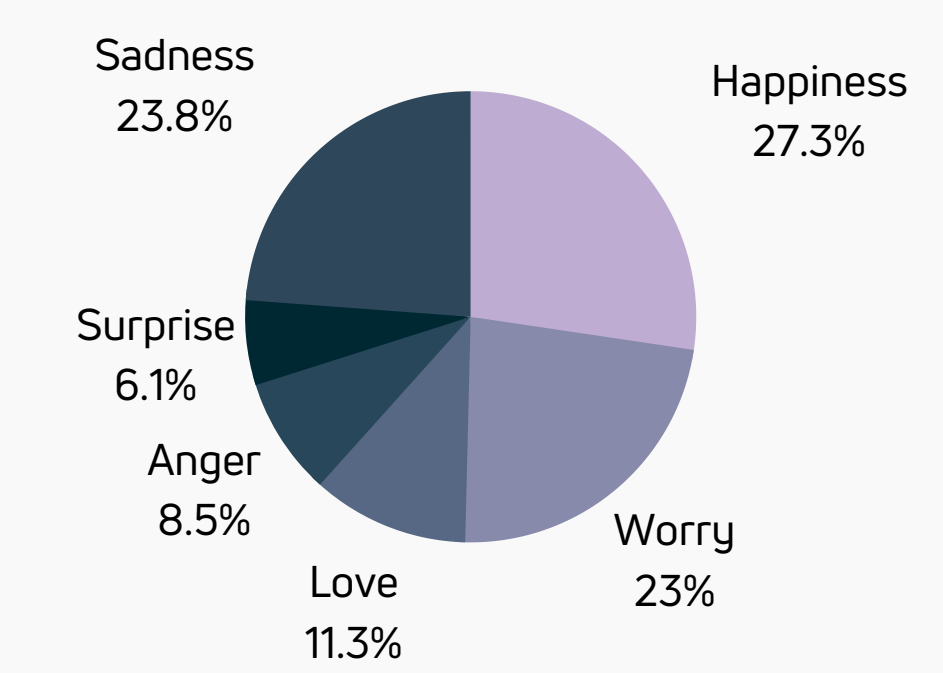
Multi-class problem:

Created a unique dataset by merging existing datasets

Total of 49K instances

Emotions: Sadness, Worry, Surprise, Love, Happiness, Anger

Class Distribution



Model	Accuracy	Macro - F1	Macro - Recall	Macro - Precision
Naive Bayes	49.39%	0.2940	0.3311	0.5399
Logistic Regression	57.68%	0.4837	0.6443	0.6410
Random Forest	53.06%	0.4554	0.4320	0.5971
Linear SVM	60.69%	0.5586	0.5379	0.6142
CNN	64.87%	0.6146	0.6012	0.6719
LSTM	67.10%	0.6398	0.6209	0.6954
Bi - LSTM	66.42%	0.6250	0.6102	0.6929
BERT	67.29%	0.6406	0.6400	0.6400
BERT + LSTM	67.84%	0.6522	0.6400	0.6700

Future steps involve the text analysis in order to detect depression or anxiety signs and the introduction of other online pattern indicators.

ISSUES & CHALLENGES

Data Gathering

- Privacy issues - Ethical issues - invasiveness into private life
- Low or no use of the Internet, especially by older women
- Annotation Difficulties

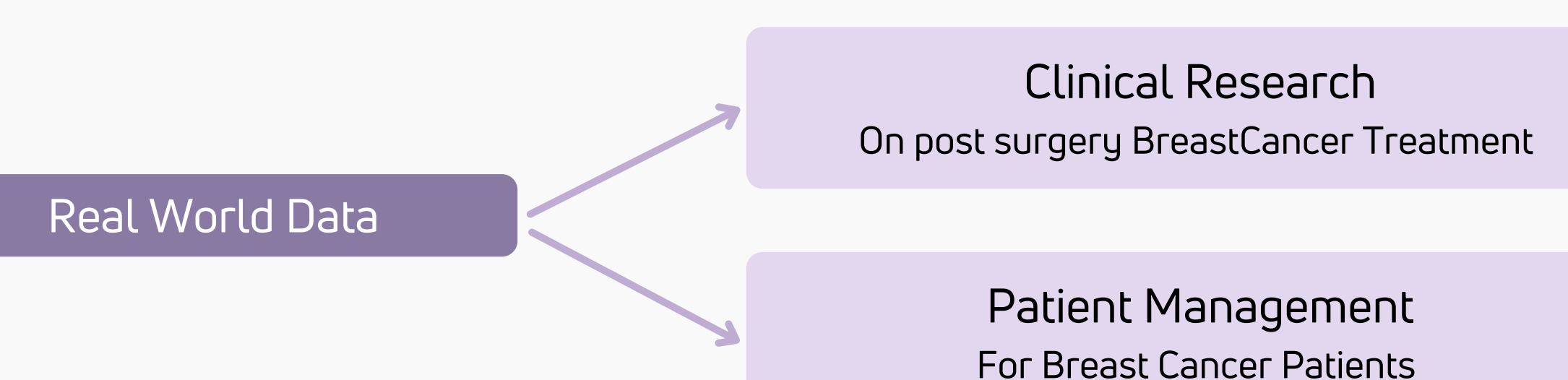
Emotional well-being identification

- Many confounding factors - no clear association between online behavior and emotional well-being
- Several languages
- Domain adaptation for cancer survivors
- Short text is limiting

ABOUT THE REBECCA PROJECT

Horizon 2020 project - REsearch on BrEaSt Cancer induced chronic conditions supported by Causal Analysis of multi-source data.

The vision



360° patient monitoring using multi-source RWD

Sources

- EHRs
- Wearables
- Mobile Apps
- Social Media
- Online activities
- more...

Functional & Emotional indices

- Behaviour
- Lifestyle
- Emotional Status
- Patient Environment
- Medical History
- more...

[1] Carreira H, Williams R, Dempsey H, Stanway S, Smeeth L, Bhaskaran K. (2021). Quality of life and mental health in breast cancer survivors compared with non-cancer controls: a study of patient-reported outcomes in the United Kingdom. *Journal of Cancer Survivorship*. 15: 1-12
[2] Burgess C, Cornelius V, Love S, Graham J, Richards M, Ramirez A. (2005) Depression and anxiety in women with early breast cancer: five year observational cohort study. *BMJ*. 330: 702. *BMJ* (Clinical research ed). 330: 702.