15TH ITU ACADEMIC CONFERENCE

ITU KALEIDOSCOPE
NEW DELHI 2024

Innovation and digital transformation for a sustainable world

21-23 October 2024
New Delhi, India

Enhancing Oncology Care with Federated Learning and Foundation Models

- Gagan N and Sanand Sasidharan

23 October 2024

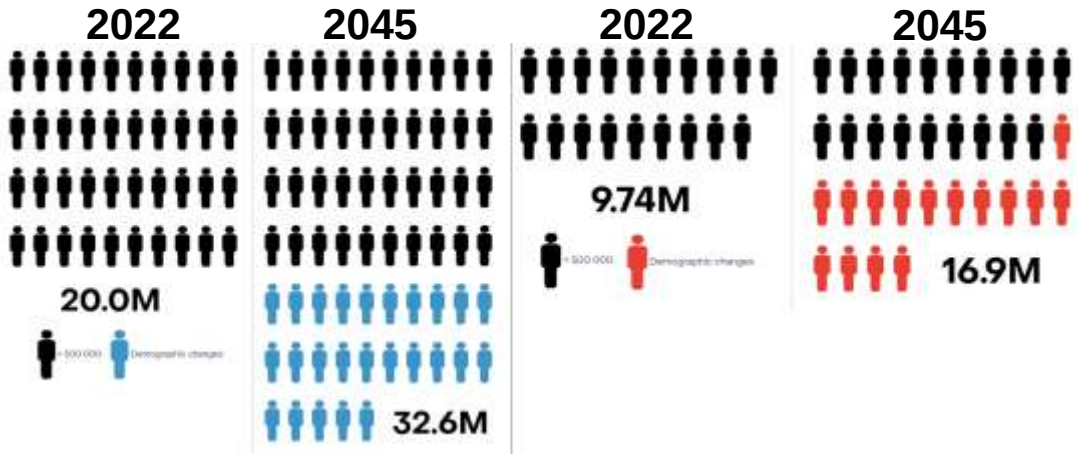
Gagan N

GE Healthcare

Session 6 – Enabling technologies

AI Enabled Health Care – Challenges in Cancer Care

Estimated number of new cases and deaths from 2022 to 2045



Imagine trying to solve a puzzle where:

- The pieces (patient data) are scattered across the world
- Each piece is locked in a separate vault (data privacy)
- The picture keeps changing (new research and treatments)
- Every puzzle is slightly different (unique for every patient)
- We need to see the big picture without opening the vaults

We need innovative solutions to piece together this complex puzzle while respecting privacy and adapting to new discoveries.

Key Challenges in Advancing Oncology Care through AI-Enabled Technologies

- Vast and Complex Data
- Privacy and Data Silos
- Personalized Treatment Complexity

Our work supports UN Sustainable Development Goals

SDG 3
Good Health and Well-being
- Improved Cancer Care
- Better Health Outcomes

SDG 9
Industry, Innovation, and Infrastructure
- Advanced healthcare AI
- Privacy-preserving data infrastructure

SDG 10
Reduced Inequalities
- Global knowledge sharing
- Improved care in underserved areas

Our Solution

We combine two cutting-edge technologies to improve cancer care

Large Language Models

Think of it as a super-smart digital oncologist that can read and understand medical texts

- Reads Medical Records
- Identifies Key Cancer Related Information
- Understands Complex medical Terminologies
- Learns from millions of medical records

+

Federated learning

Imagine a global team of researchers collaborating without sharing sensitive data

- Learns from data worldwide
- Protects patient privacy
- Breaks down data silos between hospitals
- Allows smaller hospitals to contribute and benefit

=

Faster & Global Insights

How They Work Together

- LLMs can lean to understand oncology text
- Federated Learning allows LLMs to learn from hospitals worldwide
- Each hospital's data stays private, but the LLMs gets smarter

A globally-informed Foundational Model that respects patient privacy

Real-World Impact

- Faster, more accurate cancer diagnoses
- Personalized treatment recommendations
- Equal access to AI-powered insights for all hospitals
- Accelerated cancer research without compromising patient privacy

From Foundations to Specialization

- **Foundation Models :** Machine learning models that is trained on broad data such that it can be applied across a wide range of use cases.
- **Large Language Models (LLMs) :** Type of foundation model specifically designed to understand and generate human language.
- **BERT :** (Bidirectional Encoder Representations from Transformers) An LLM that understands context by looking at words before and after.
- **BioBERT :** A version of BERT trained on biomedical text, giving it a strong foundation in medical terminology.

Specializing in Oncology

Step 1 : Start with BioBERT

We begin with BioBERT, which already understands general medical language.

It's like a student who has completed general medical training.

Step 2 : Learning Oncology Patterns

The model is further finetuned on a comprehensive oncology corpus, including Peer-reviewed cancer research, Clinical trial reports, Oncology guidelines, Anonymized patient records

Masked Language Modeling : Improves the model's understanding of oncology terminology and context

- Randomly masks 15% of words in each input
- Model predicts the masked words
- Enhances comprehension of cancer-specific terms

Next Sentence Prediction : Improves understanding of relationships between sentences

- Model predicts if two sentences are consecutive
- Enhances grasp of document structure
- Improves contextual understanding in oncology texts

This is like student going through intensive oncology training and connecting the dots in cancer care

Step 3: Fine-tuning for Specific Tasks

Adapt the model for oncology-specific tasks

- **Named Entity Recognition (NER) for identifying cancer types, treatments, etc.**
- Relation Extraction to understand connections between entities
- Text Classification for categorizing medical documents

It's like having an AI oncologist who is expert in one particular task which will assist human oncologists.

The Federated Learning Process

Hospitals share raw patient data

Data sent to central repository

AI model trained on combined data

Trained model sent back to hospitals

- Privacy concerns with data sharing
- Legal and ethical issues
- Centralized data vulnerable to breaches
- Excludes institutions unable to share data

Conventional Approach

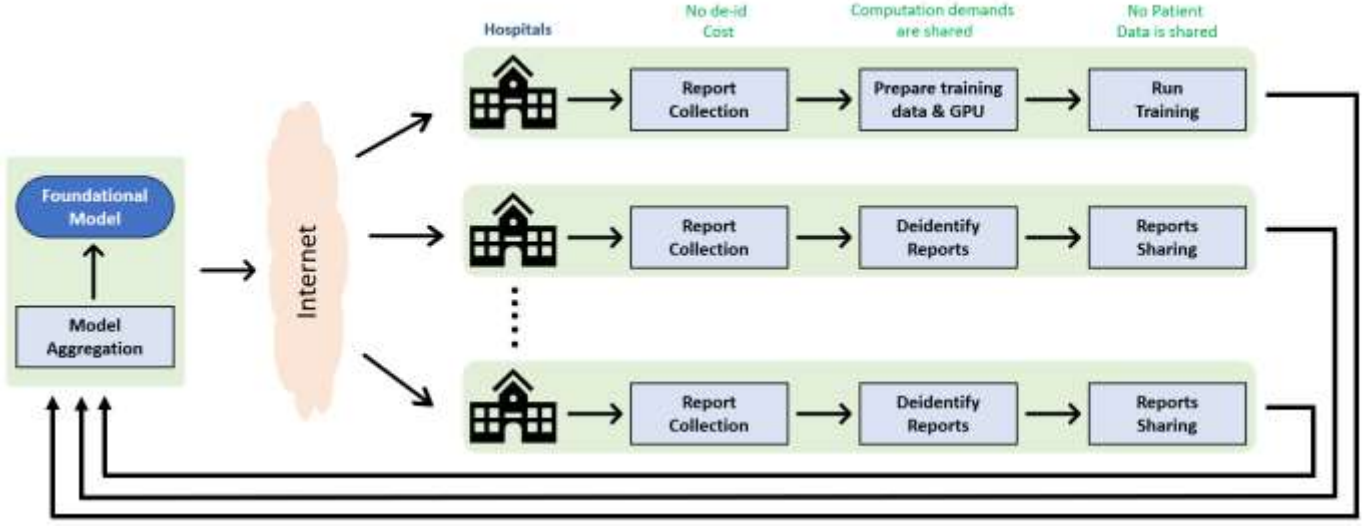
Initial Model is shared with Hospitals

Hospitals will train model locally on private data

Only model updates sent to central server

Central Server Combines Updates

- Preserves patient data privacy
- Complies with data protection laws
- Enables global collaboration



Federated Approach

It's like a global team of doctors sharing their insights without revealing patient details

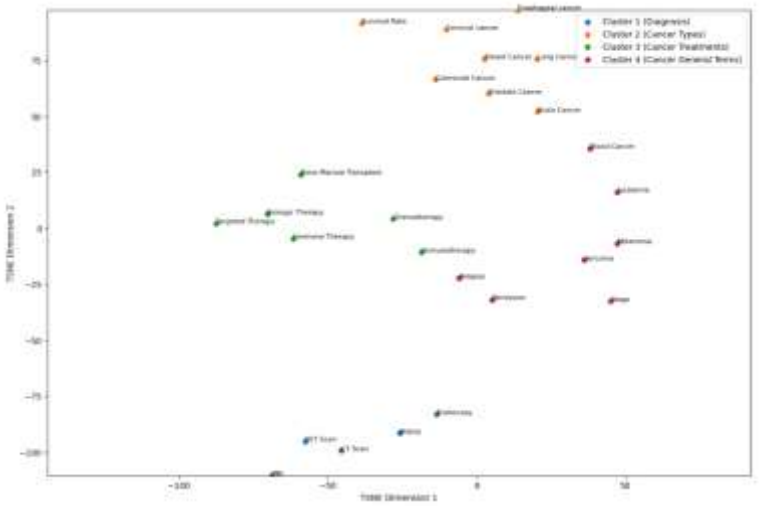
Results : Embedding Visualization

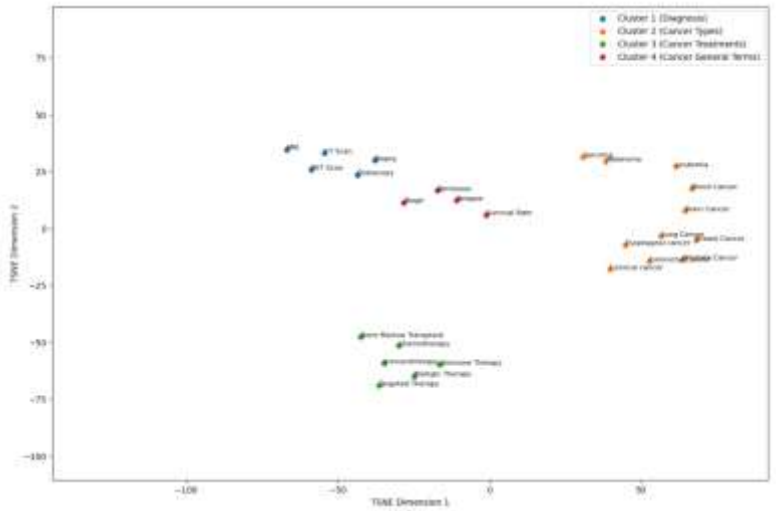
What are embeddings?

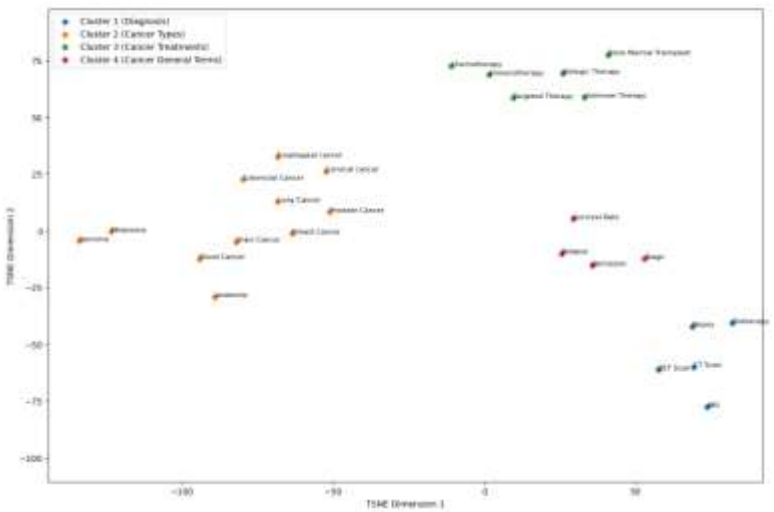
Embeddings are like the AI's understanding of words. Words with similar meanings are placed closer together in a multi-dimensional space.

Why this matters ?

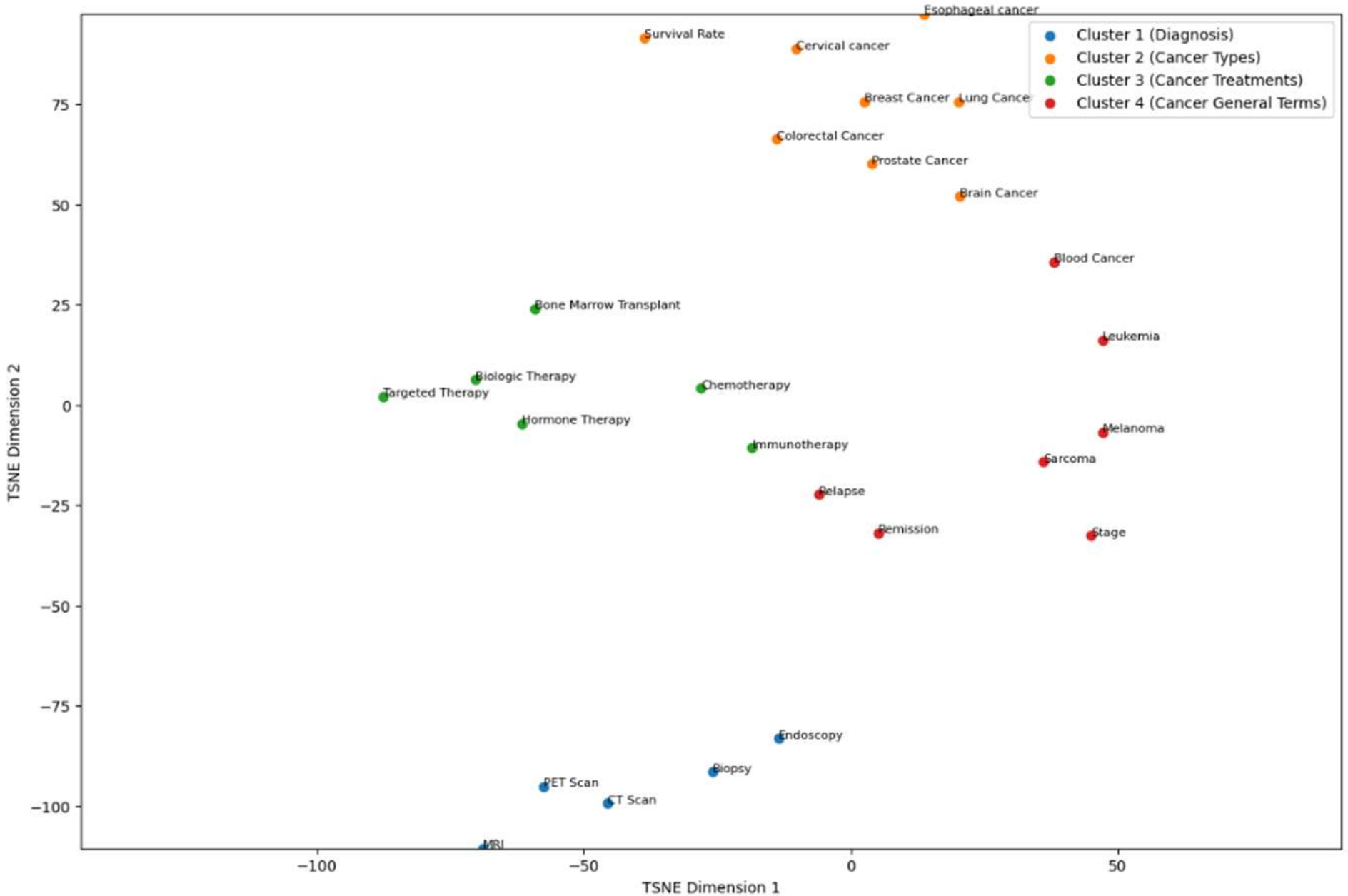
Better grouping of related terms means our AI can make more accurate connections in medical texts, leading to improved information extraction and decision support in oncology care.



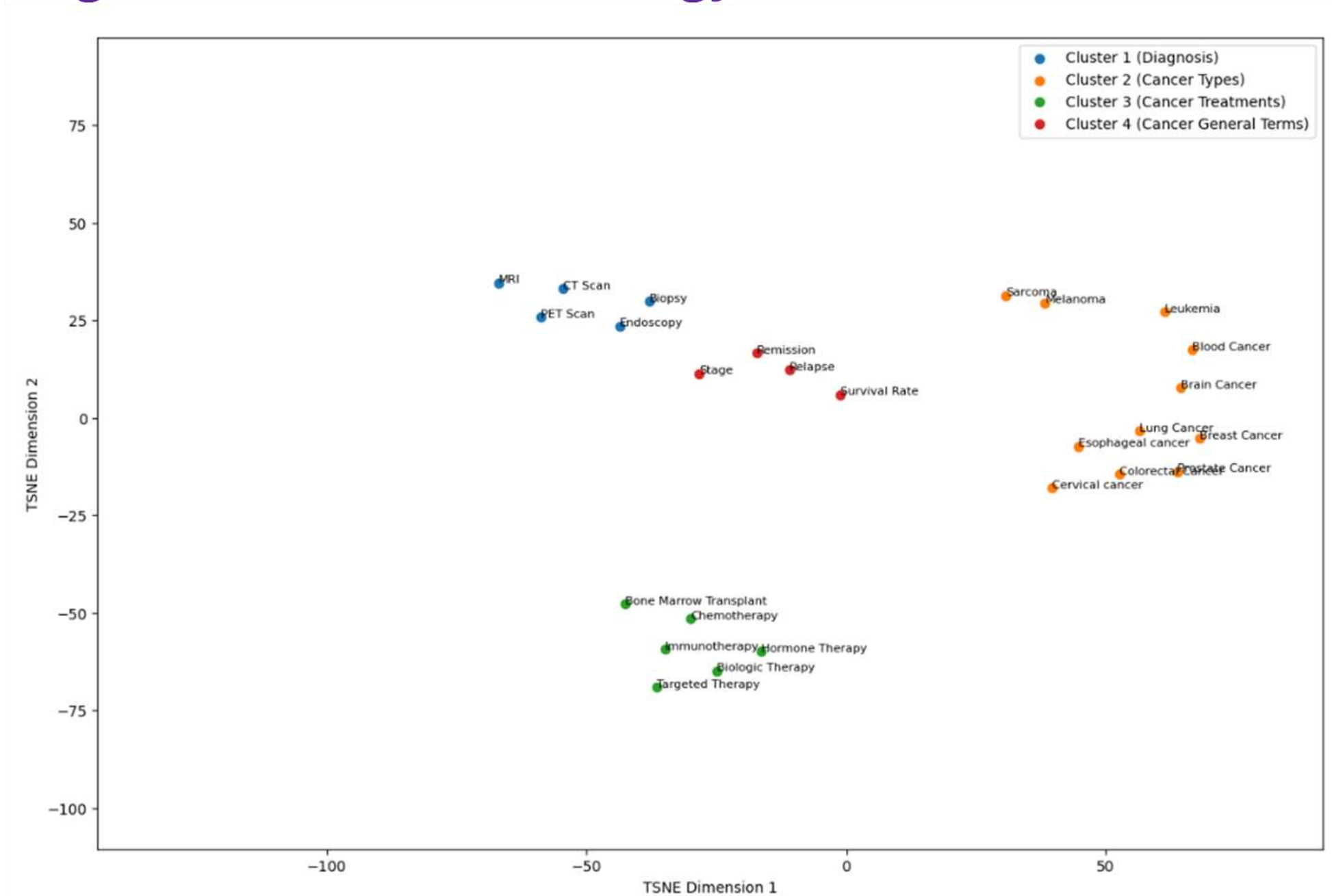




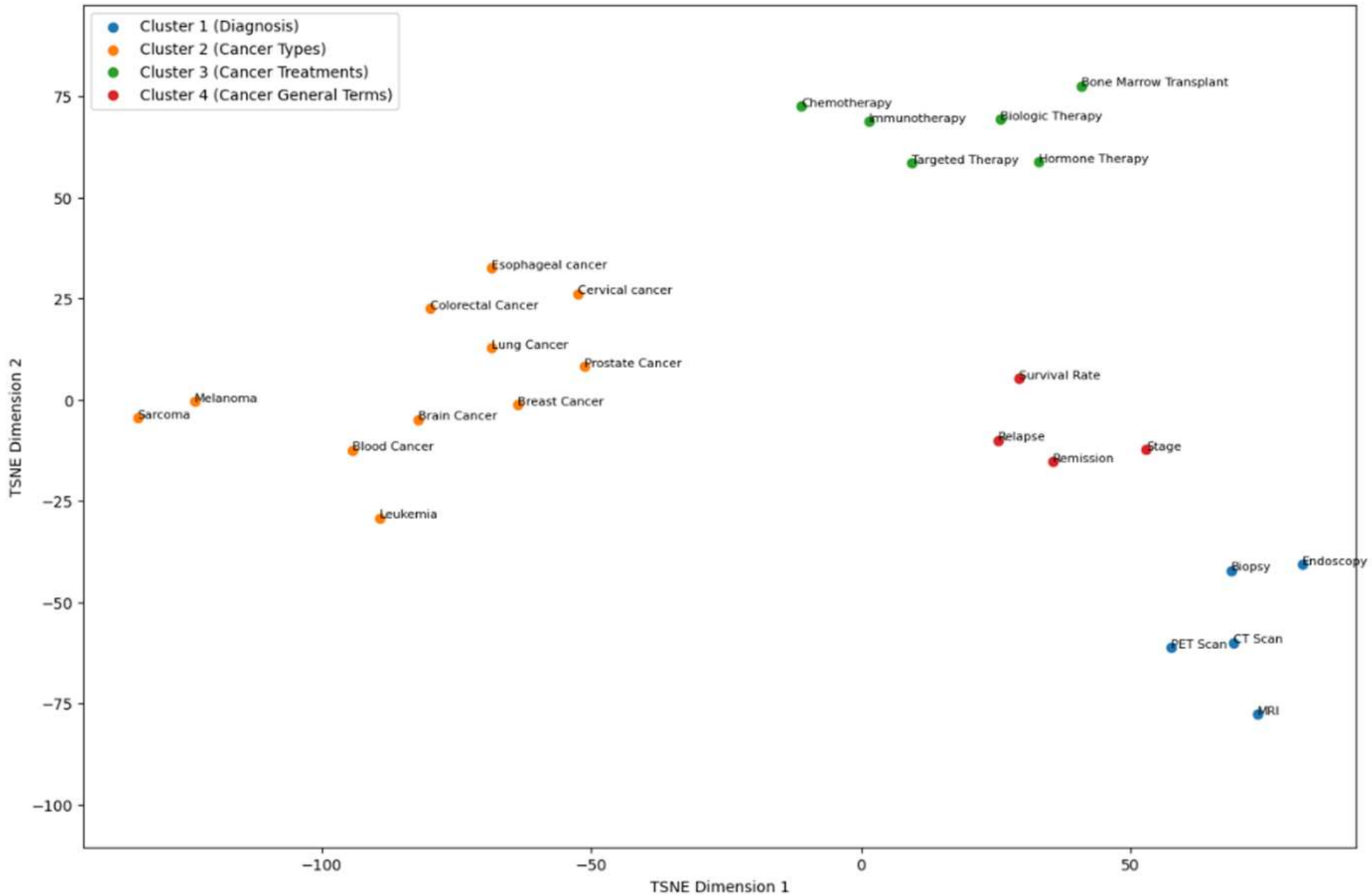
Embedding Visualization of BioBERT



Embedding Visualization of Oncology Fine Tuned BioBERT



Embedding Visualization of Oncology Fine Tuned BioBERT Through Federated Learning



Results: Named Entity Recognition (NER) Task

What is Named Entity Recognition?

NER is the AI's ability to identify and classify key information in text, like recognizing "chemotherapy" as a cancer treatment in a medical report.

Why this matters ?

Better entity recognition means our AI can more accurately extract crucial information from medical texts. This leads to improved support for diagnosis, treatment planning, and overall patient care in oncology.

Tag	BioBERT	Fine-tuned BERT	Federated Fine-tuned BERT
Caner Treatment	7	269	307
Prosthetic	504	601	655
Drug Regiment	1885	2095	2239
Pathological Finding	118	254	356

Our work paves the way for

- More equitable access to advanced cancer care
- Personalized treatment plans powered by global data
- Continuous improvement of AI models

> “Creating a world where HealthCare has no limits.”

Thank you!