



FSC  
Fondo per lo Sviluppo  
e la Coesione



# Progetto LIFEMap

## Seminario scientifico

14 marzo 2025

12:00 - 13:00

# Performance Comparison of Fine-tuned BERT-Based Genomic Language Models

**Relatore:** Payel Patra – Università de L'Aquila

## Abstract

The evolution of artificial intelligence can empower accurate medical diagnostics (e.g., disease, Symptoms, and treatment), personalized treatment plans, and medical data analysis, speeding up the diagnosis or improving treatments. As a primary study, we conducted a comparison among several BERT-based models fine-tuned on unstructured Clinical Summary Mammary Malignancy (CSMM) data to annotate clinical notes with breast cancer-specific concepts. Our research transforms unstructured clinical text data into a structured format (CSV), simplifying the identification and recognition of specific entities (namely, Clinical Named Entity Recognition -NER), such as age, gender, disease, symptoms, medication, doses, medical history, and cancer stage. The considered BERT-based models are BioBERT, BioClinicalBERT, RoBERTa, PubMedBERT and BlueBERT. The comparative analysis reveals that our fine-tuned BioBERT transformer model demonstrates more robust performance, achieving a high F-score of 96% in extracting medical entities, which confirms their effectiveness in handling domain-specific clinical texts.

**Link -** <https://meet.google.com/vaa-oczm-yjy>