



INFORMATICS
INSTITUTE OF
TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**IntegraBNER: Integrating Multimodal Data Sources for
Enhanced Biological Named Entity Recognition**

Final Thesis by

Saveen Dinuwara Haupe Liyanage

Supervised by

Mr. Dilum De Silva

Submitted in partial fulfilment of the requirements for the BEng (Hons) Software Engineering degree at the University of Westminster.

April 2025

ABSTRACT

Biomedical studies and clinical reporting often involve enormous amounts of unstructured text data and diagnostic images, which pose the challenge of extracting informative biological named entities such as disease, anatomical site, and clinical findings. Traditional Named Entity Recognition (NER) models are directed toward textual data only, which fails to recognize the potentially crucial visual information embedded in radiologic imaging. This research, IntegraBNER, bridges this gap by suggesting a multimodal NER solution that integrates radiology reports and chest X-ray images for more accurate and context-aware entity recognition in the biomedical field.

IntegraBNER integrates Bio_ClinicalBERT for extracting text features and ResNet-50 for visual representation of the chest X-rays. A fusion model having an attention-based MLP classifier integrates the two modalities to perform multi-label biomedical entity prediction. The system is trained and evaluated on the OpenI Indiana Chest X-ray dataset and automatically labeled using a hybrid autolabeling strategy with synonym dictionary matching and the output of two fine-tuned BioBERT NER models for better coverage.

To test usability and interpretability, IntegraBNER offers a Gradio-based GUI for users to type in a radiology report and/or an X-ray image. The UI returns highlighted entities and a tabular output with predicted terms, confidence values, and entity types from a carefully curated vocabulary. IntegraBNER has been successfully hosted in Hugging Face Spaces to facilitate public accessibility and expert validation.

Performance analysis exhibits excellent performance in text-based prediction (micro F1-score: 0.94) and illustrates the complementary nature of visual information. While fusion-based accuracy is still being improved (F1-score: 0.77), high recall on common medical conditions like pneumothorax and pleural effusion is achieved. Error analysis in depth revealed that prediction over rare entities and weak image features require improvement in fusion techniques and labeling depth.