Academic Year 2023/2024

Biomedical Informatics

Development of a Biomedical Ontology

Deep-learning-based personalized prediction of absolute neutrophil count recovery and comparison with clinicians for validation Choo et al. [2023]

Group

Andreas Brummer

Lorenzo Leuzzi

Michele Simeone

1 Article Analysis

For this assignment, we have chosen the research article titled "Deep-learning-based personalized prediction of absolute neutrophil count recovery and comparison with clinicians for validation" as the basis for developing an ontology that encapsulates the integration of deep learning techniques in medical diagnostics, specifically focusing on the prediction of neutrophil count recovery in pediatric oncology. The research emphasizes the critical need for accurate prediction of neutrophil recovery in cancer patients, especially considering the shortcomings of current pharmacokinetic-pharmacodynamic models (PK-PD models). The study employed a deep learning approach, using data from 525 pediatric patients with solid tumors to train a Google's Temporal Fusion Transformer model Lim et al. [2021] and validating it with data from 99 patients. This model outperformed clinicians, achieving a 76.76% accuracy in predicting neutrophil recovery within a one-day margin, compared to the 58.59% by specialists and 32.33% by the resident group. Notably, approximately 80% of clinicians adjusted their predictions after viewing the model's output, with 90.53% of these adjustments enhancing prediction accuracy. The study also revealed a broad acceptance of the deep learning model among clinicians, as indicated by questionnaire results. The discussion in the paper highlights the innovative use of deep learning in predicting neutrophil recovery and its influence on clinical decision-making, albeit with some hesitancy among clinicians to fully trust the model. The study concludes that deep learning models are not only more accurate than human experts in predicting neutrophil recovery but also useful in assisting clinicians to make more precise predictions, thus potentially improving patient care in clinical settings. This research stands as a pioneering effort in integrating AI into pediatric oncology, showcasing the potential of deep learning in augmenting clinical decision-making processes.

2 Ontology Objective

The objective of this ontology is to establish a comprehensive framework that encapsulates the intricate aspects and relationships involved in the treatment of cancer patients undergoing chemotherapy, with a focus on using deep learning models for the prediction and management of neutropenia, a common and critical side effect. This ontology integrates diverse classes ranging from patients and their specific cancer types to the details of chemotherapy regimens and the chemotherapeutic agents used. It addresses the condition of neutropenia, emphasizing its severity, duration, and the crucial process of neutrophil count recovery.

The ontology incorporates both traditional mathematical models and advanced deep learning models, highlighting their roles in predicting neutrophil count recovery. It encompasses the perspectives of clinicians, detailing their experience and specialization, and their interaction with both the deep learning model and traditional methods for better patient care.

Infections, as a significant risk for immunocompromised patients, are included, underlining the importance of early detection and management. Medical datasets form a vital part of this ontology, as they are essential for training and validating predictive models. The ontology also considers guidelines and criteria for defining severe neutropenia, ensuring adherence to clinical standards.

Additionally, the ontology elaborates on attributes like patient demographics, cancer stages, drug properties, model performance metrics, and the severity of infections. It also describes relationships such as the impact of chemotherapy on neutropenia, the use of datasets in model training, and the role of clinicians in evaluating model predictions and managing patient care according to established guidelines. Axioms in the ontology provide logical propositions that underline critical insights, such as the importance of accurate predictions for patient management and the relevance of experience in clinical decision-making. Instances in this ontology represent specific real-world examples of each class, providing concrete context to the abstract concepts.

3 Ontology Development

Creating an ontology for a medical paper, such as "Deep-learning-based personalized prediction of absolute neutrophil count recovery and comparison with clinicians for validation", involves a structured process to effectively capture and represent the knowledge within the paper. In the process, the initial step involves identifying key concepts and entities directly from the text. This is done by thoroughly reading the medical paper and paying particular attention to nouns or noun phrases. This step is crucial in categorizing primary classes and instances within the ontology, such as types of cancers, neutropenia situations, and deep learning models. Following this, relationships and properties of each entity are defined, establishing how these entities interact with each other and their specific attributes. For instance, the relationship between a chemotherapy regimen and its resultant neutropenia is established, along with defining properties like the severity of neutropenia or the type of chemotherapeutic agents involved. The next critical phase is developing axioms, which are essentially logical rules applied to the entities and relationships, ensuring consistency and aiding in new information inference. For example if neutrophil count is below the threshold, then severe neutropenia is diagnosed. The actual construction of the ontology involves selecting a suitable tool, such as Protégé, and methodically inputting the entities, their properties, and relationships into a structured format, often requiring hierarchical organization and subclass creation. This step is vital for visualizing and validating the ontology's framework, ensuring it accurately represents the knowledge domain of the medical paper.

This detailed and systematic approach ensures the created ontology effectively encapsulates the complex domain of medical science.

4 Ontology

The following **classes** have been identified along with their **descriptions**.

1. **Patient**: Individuals receiving chemotherapy.
2. **Cancer/Tumor**: Different types of cancer or tumors in patients.
3. **Chemotherapy Regimen**: Specific treatment plans involving chemotherapeutic agents.
4. **Chemotherapeutic Agents**: group of drugs used in chemotherapy
5. **Neutropenia**: A condition of abnormally low neutrophil count.
6. **Neutrophil Count Recovery**: The process of neutrophil count returning to normal levels.
7. **Deep Learning Model**: The AI model used for predicting neutrophil count recovery.
8. **Mathematical Models**: Traditional models used in medical predictions.
9. **Clinician**: Healthcare professionals involved in patient care and treatment decisions.
10. **Infections**: Diseases caused by pathogens, particularly in immunocompromised patients.
11. **Medical Datasets**: Collections of patient data used for training and validating models.
12. **Guidelines checklist**: Standards or criteria for clinical practices or research.
13. **Criterion of severe neutropenia**: specific standards or benchmarks used to identify and classify severe neutropenia.
14. **Loss function**: A mathematical function used to quantify the difference or error between predicted values by a model and the actual values in the data

The **relationships** between classes are as follows.

1. Patient undergoes Chemotherapy Regimen.
2. Chemotherapy Regimen causes Neutropenia.
3. Chemotherapeutic Agents are components of broader Chemotherapy Regimen for cancer patients.
4. Neutropenia affects Neutrophil Count Recovery.
5. Deep Learning Model predicts Neutrophil Count Recovery.
6. Clinician evaluates Neutrophil Count Recovery.
7. Deep Learning Model assists Clinician.
8. Mathematical Models were previously used for predicting recovery.
9. Patients with Neutropenia are more susceptible to Infections.
10. Medical Datasets are utilized by Models.
11. Clinician follows Guidelines checklist.
12. Criterion is monitored during chemotherapy
13. The Loss Function is applied to train a Deep Learning Model / Mathematical Model using a Medical Dataset

The logical **axioms** are defined as follows.

1. If a Patient has severe Neutropenia, then Neutrophil Count Recovery is critical.
2. If a Deep Learning Model has high accuracy, it is more reliable for predicting Neutrophil Count Recovery.
3. Clinicians with more experience might provide better assessment than those with less.
4. Compliance with Guidelines checklist enhances the quality of clinical decision-making.
5. If Neutrophil count is below threshold, then severe neutropenia is diagnosed
6. Lower Loss implies better Model

Table 4.1 provides details on the **attributes** and **instances** associated with these classes.

5 Ontology Implementation

The implementation of this ontology began with a comprehensive review of the current literature and existing ontologies in the field of cancer treatment and artificial intelligence. This process involved iterative refinement, ensuring alignment with both clinical practices and research advancements. Additionally, the ontology was enriched by incorporating relevant elements from established vocabularies and standards available on platforms like BIOPORTAL. The resulting ontology offers a detailed and structured representation of this two complex domains.

Class	**Attributes**	**Instances**
Patient	Age, Sex, Type of Cancer, Chemotherapy History	Specific Patient
Cancer/Tumor	Type, Stage	Brain tumor, Neuroblastoma
Chemotherapy Regimen	Type of Drugs, Dosage, Frequency	Particular Chemotherapy Regimen (e.g., high-dosage)
Chemotherapeutic Agents	Mechanism of Action, Side Effects	Doxorubicin
Neutropenia	Severity (Grades), Duration	Instance of Neutropenia (e.g., severity [1, 4])
Neutrophil Count Recovery	Time to Recovery, Recovery Level	A specific recovery treatment
Deep Learning Model	Algorithm, Accuracy, Predictive Performance	Google's TFT
Mathematical Models	Type (e.g., PK-PD), Parameters	PK-PD model
Clinician	Experience Level, Specialization, Number of Patients Handled	Specific Clinician
Infections	Type, Severity	Bacterial, Funghi, Virus
Medical Datasets	Size, Diversity, Representativeness	EHR, CDW at Samsung Medical Center
Guidelines checklist	Compliance Level, Scope	MI-CLAIM, CONSORT-AI
Criterion of severe neutropenia	Threshold Value, Associated Risks	Specific Criterion (e.g., Specific Patient Case, severe)
Loss function	Type, Formula, Objective, Computation Complexity, Differentiability	sMAPE

Table 4.1: Attributes and Instances for the classes

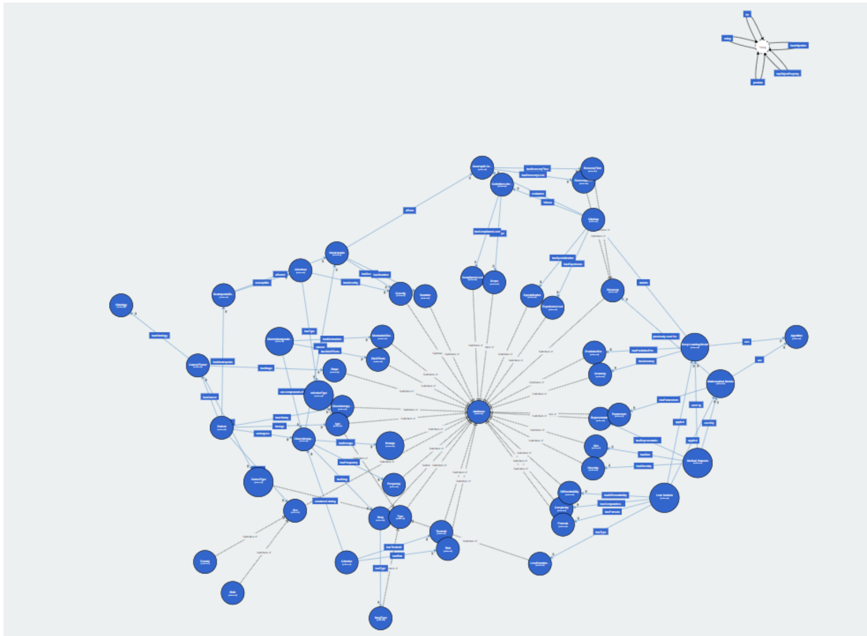

Figure 5.1: Ontology Overview

5.1 Classes

This section of the ontology focuses on defining the core classes that represent the key entities. Each class is designed to encapsulate a distinct aspect of the process, ranging from patient demographics to the specific activation layer of a neural networks. The attributes of each class provide detailed characterizations, offering insights into the complex interactions and dependencies within the treatment landscape. The following descriptions offer an overview of some primary classes, highlighting their attributes and roles within the ontology structure.

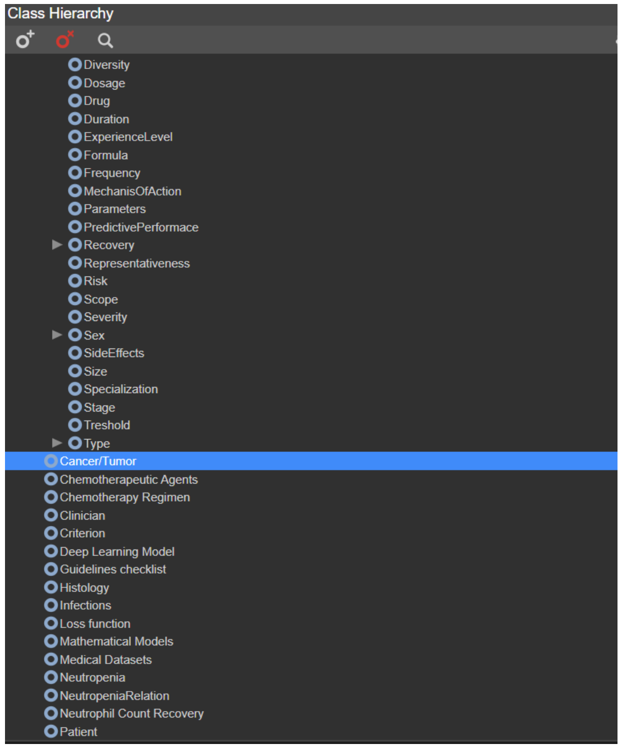

Figure 5.2: Classes

5.1.1 Patients

- **Attributes:** Age, Sex, Neutrophil Count.
- **Description:** Represents individuals undergoing chemotherapy, with attributes detailing their demographics and relevant medical indicators.

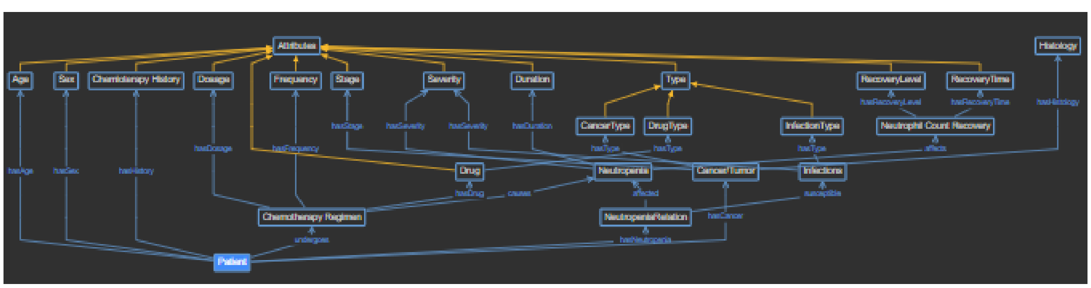

Figure 5.3: Patients

5.1.2 Cancer/Tumor

- **Attributes:** Type, Stage, Genetic Mutations.
- **Description:** Encompasses various types of cancers and tumors, classified by their characteristics and progression stages.

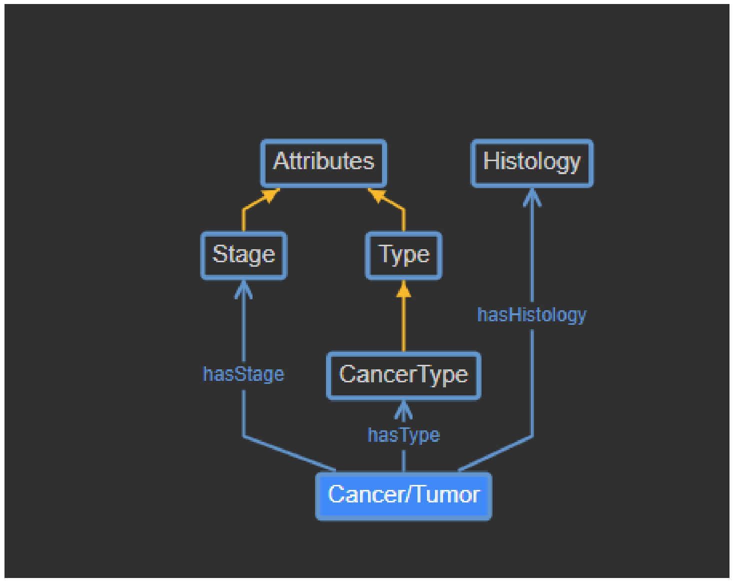

Figure 5.4: Cancer/Tumor

5.1.3 Deep Learning Model

- **Attributes:** Accuracy, Predictive Performance, Type.
- **Description:** Represents sophisticated computational models used in healthcare analytics, particularly in oncology. These models are characterized by their ability to accurately predict clinical outcomes, such as patient response to treatments or disease progression, based on a variety of input data. Key attributes include the model's accuracy in making predictions and its overall predictive performance, which are critical for ensuring reliable and effective patient care.

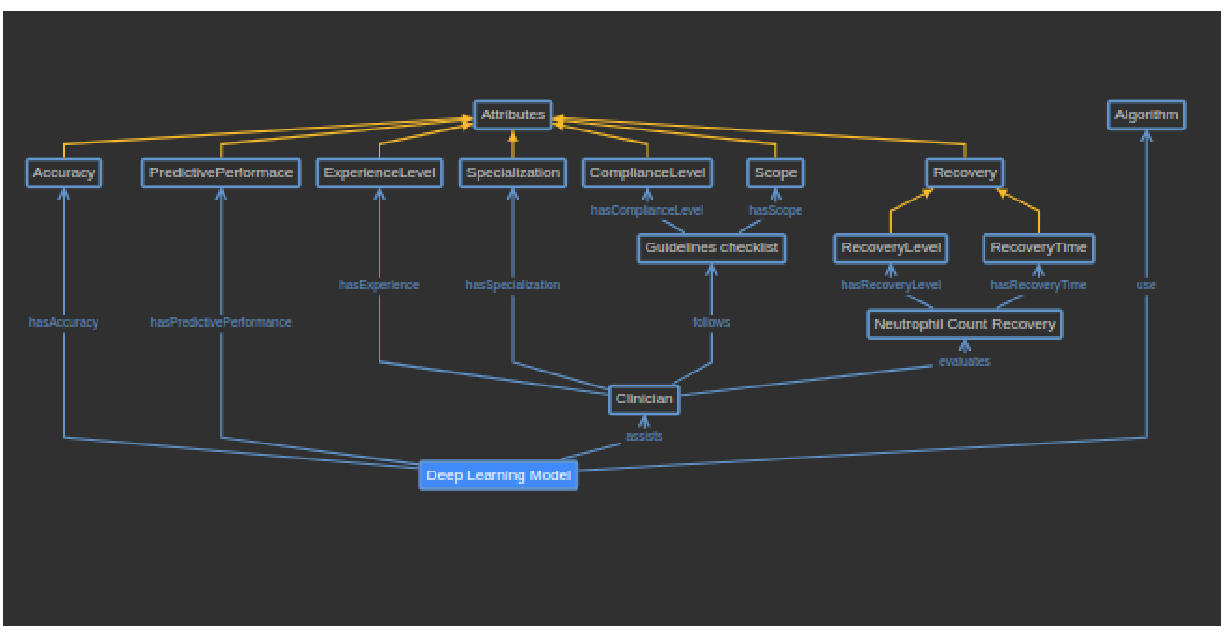

Figure 5.5: Criterion

6 Properties Description

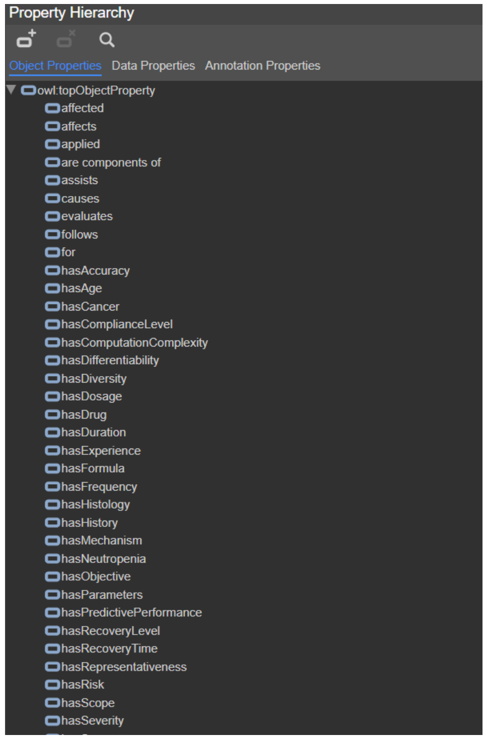

Figure 6.1: Properties

In this ontology, properties play a crucial role in defining the relationships and characteristics of the various classes. They serve as connectors that link classes to each other, thereby forming a coherent and interrelated structure. These properties range from simple attributes that describe class features to complex relationships that illustrate interactions between different classes. Here's an example of some properties:

1. **hasRecoveryTime**: This property relates to the duration it takes for a patient's recovery, in the context of chemotherapy or another medical treatment.

2. **hasPredictivePerformance**: This property is associated with the performance of predictive models in relation to medical outcomes or treatment efficacy.

3. **hasDrug**: A property that links patients or medical conditions to specific drugs used in their treatment, indicating a relationship between a medical condition and a pharmaceutical agent.

7 Individuals Description

In the development of our ontology, we have incorporated several individuals directly from the referenced academic paper to enhance its relevance and applicability. Examples of such individuals include "TFT" for DeepLearning Models, "PK-PD model" representing Mathematical Models, and "Doxorubicin", a chemotherapeutic agent. These specific instances enrich the ontology with real-world data, facilitating more accurate modeling and analysis within the scope of chemotherapy treatment for cancer patients.

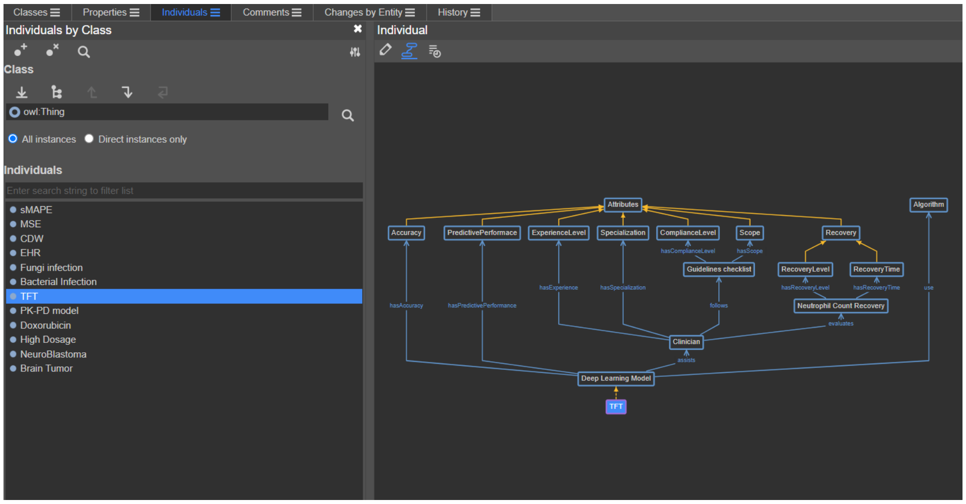

Figure 7.1: Individuals

8 Ontology Extension

After creating our ontology based on the paper, we decided to extend it by reusing two ontologies found at `https://bioportal.bioontology.org/`.

- `Artificial Intelligence Ontology` Joachimiak [2023] models classes and relationships describing deep learning networks, their component layers and activation functions, as well as potential biases.
- `National Cancer Institute Thesaurus` NCI [2023], a vocabulary for clinical care, translational and basic research, and public information and administrative activities.

We chose these two ontologies because combined with the one we created starting from the paper they can help create a broader link between artificial intelligence tools and the fight against cancer, designing and testing new relationships as the paper does.

Metrics

Axiom	19.621
Logical axiom count	1.816
Declaration axioms count	1.573
Class count	1.468
Object property count	148
Data property count	0
Individual count	12
Annotation Property count	155

Figure 8.1: Numerical Data of the Entire Ontology

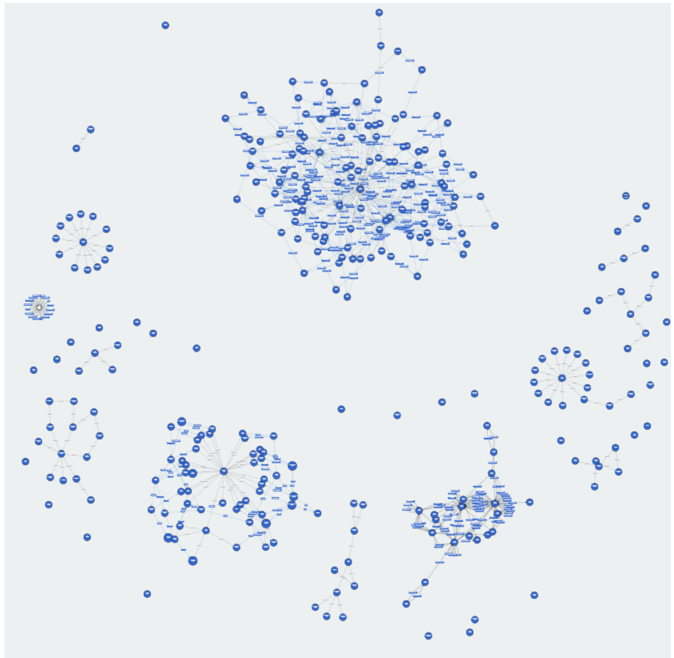

Figure 8.2: Graphical Representation of the Entire Ontology

Annex

Please see the attached copy of the original article used for this assignment. This document provides foundational information and insights that were critical for the development of the ontology.
Deep-learning-based personalized prediction of absolute neutrophil count recovery and comparison with clinicians for validation

Contributions

Each member of the group actively supervised the entire process of carrying out the assignment, dividing the work equally.

References

H Choo, SY Yoo, S Moon, and et al. Deep-learning-based personalized prediction of absolute neutrophil count recovery and comparison with clinicians for validation. *J Biomed Inform*, 137:104268, 2023. doi: 10.1016/j.jbi.2022.104268.

Marcin Pawel Joachimiak. Artificial intelligence ontology, 2023. URL `https://bioportal.bioontology.org/ontologies/AIO`.

Bryan Lim, Nicolas Loeff, Sercan Arik, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. 2021.

NCI. National cancer institute thesaurus, 2023. URL `https://bioportal.bioontology.org/ontologies/NCIT`.