Digital Zone
Jurnal Teknologi Informasi dan Komunikasi

Volume 17 Issue 1 Year 2026 | Page 95-104 | e-ISSN: 2477-3255 | ISSN: 2086-4884
Received: 16-10-2025 | Revised: 11-05-2026 | Accepted: 25-05-2026

Hierarchical ClinicalBERT Models for Antibody Subtype Classification in Oncology Trial Statements

Susi Handayani[1], Taslim*[2], Dafwen Toresa[3], Syahriatna[4]
[1,2,3,4,] Faculty of Computer Science, Universitas Lancang Kuning
[2] Faculty of Informatics & Computing, Universiti Sultan Zainal Abidin
*Corespondence: taslim@unilak.ac.id

Abstract: *A hierarchical language model based on ClinicalBERT was developed to classify antibody subtypes in oncology clinical trial statements under severe class imbalance conditions. The study addresses the limitation of conventional flat multilabel classification approaches that ignore the hierarchical dependency between general antibody categories and their specific subtypes. The proposed framework applies a two-stage architecture in which antibody presence is detected before subtype identification, enabling conditional learning for monoclonal and bispecific antibodies. The model combines hierarchical optimization with imbalance-aware strategies including class weighting, focal loss, and limited undersampling to improve sensitivity toward rare therapeutic categories. Experiments were conducted using 17,701 annotated oncology trial statements derived from ClinicalTrials.gov protocols. Model performance was evaluated using macro F1, subtype-specific F1, precision–recall area under the curve, hierarchical accuracy, and calibration analysis across multiple random seeds. The hierarchical ClinicalBERT model consistently outperformed the flat multilabel baseline, with bispecific F1 increasing from 0.02 to 0.06 and precision–recall area improving from 0.01 to 0.03. Reliability analysis further demonstrated well-calibrated probability estimates suitable for evidence ranking and biomedical decision-support applications. The findings indicate that incorporating label hierarchy improves rare subtype recognition, robustness, interpretability, and classification stability in biomedical text mining tasks.*

Keywords: *ClinicalBERT, hierarchical text classification, antibody subtype classification, oncology clinical trials, imbalanced learning*

1. Introduction

Advances in biomedical text mining have enabled the automatic extraction of clinical information from large volumes of medical documents [1]. Among the various tasks in this field, the identification of therapy-related entities from oncology texts remains essential for understanding treatment trends and supporting evidence-based medicine [2], [3]. Deep contextual language models such as BERT and its biomedical variants have improved the representation of medical language by capturing syntactic and semantic nuances more effectively than traditional approaches [4], [5]. Pretraining domain-specific models for the biomedical domain has been shown to yield significant gains over continuing general-domain pretraining [6], [7]. In comparative analyses, ClinicalBERT and BioBERT both outperform generic BERT architectures in biomedical tasks [8], [9], [10].

Despite these advances, accurately classifying antibody subtypes within oncology texts continues to be a challenging problem. Most existing approaches rely on flat classification schemes that treat all labels as independent, overlooking the inherent hierarchical relationships among therapeutic categories [11], [12]. Hierarchical text classification methods have demonstrated performance benefits when label structures are respected [13], [14], [15]. In biomedical contexts, antibody mentions are often nested: general antibodies encompass subtypes such as monoclonal and bispecific [16], [17]. Ignoring that hierarchy may lead to confusion

between related subtypes, especially when one subtype (e.g., bispecific antibodies) is orders of magnitude rarer [18]. Recent works on hierarchical deep learning for imbalanced text classification and contrastive hierarchical modelling suggest that architectural alignment with label structure can yield substantial improvements [19], [20], [21].

This paper addresses that gap by proposing a hierarchical deep learning approach based on ClinicalBERT for antibody subtype classification in oncology clinical trial statements. The study introduces a two-level framework that aligns model architecture with the biological structure of antibody categories. The hierarchical formulation is designed to enhance recognition of rare subtypes by combining structural dependencies with imbalance-aware optimization, which has been studied recently in imbalance learning surveys [22], [23], [24]. The originality of this work lies in integrating hierarchical modelling with domain-informed loss design and evaluating calibration, robustness, and interpretability in the context of real clinical trial text.

The objective of this research is threefold. First, to assess whether a hierarchical model improves subtype detection compared with a flat multilabel baseline. Second, to examine the effect of imbalance-aware training strategies on rare subtype performance. Third, to analyze the robustness and calibration of the proposed model across random seeds. The findings are expected to contribute to more accurate biomedical text classification frameworks and to offer practical insights for automated evidence extraction and oncology data curation

2. Research Method

2.1 Data Source and Selection

The dataset used in this study was obtained from a publicly accessible repository on Kaggle titled Clinical Trials on Cancer, originally compiled by Bustos and Pertusa [25]. The primary corpus was derived from 49,201 interventional oncology protocols registered in the ClinicalTrials.gov database. From these protocols, the compilers extracted 6,186,572 labeled eligibility statements and released a one-million-record subset under a Creative Commons BY-NC-SA license. Each record contains a short text segment and a binary indicator specifying whether it belongs to the inclusion or exclusion section of a clinical trial protocol.

For the present research, only inclusion statements were retained, since descriptions of therapeutic interventions are predominantly concentrated in that section. The dataset was then filtered to identify oncology-related statements referring to specific therapeutic interventions. Ten binary categories were maintained for modelling: antibodies, monoclonal antibodies, bispecific antibodies, calcium, dietary interventions, estrogens, conjugated USP, paclitaxel, immunoglobulins, and heparin. After the filtering process, the resulting corpus comprised 17,701 multilabel statements suitable for supervised learning experiments.

Antibody-related categories followed a clearly defined hierarchical structure. All statements labeled as monoclonal or bispecific antibodies occurred within antibody-positive entries, and there was no overlap between monoclonal and bispecific labels. The bispecific category remained extremely infrequent, containing 196 positive instances compared with 15,788 for monoclonal antibodies, corresponding to an approximate ratio of 1:80. The corpus was divided into training, validation, and test subsets with proportions of 70%, 15%, and 15% respectively. Stratified sampling was applied to preserve proportional label distribution across all subsets. Each record in the dataset was fully anonymized and contained no identifiable patient information.

The overall process of record identification, screening, eligibility, and inclusion for modelling is summarized in Table 1, which follows the PRISMA framework for transparent dataset selection and curation. This refined subset was subsequently used as the primary corpus for model development and evaluation described in Section 2.3.

Table 1. PRISMA-style summary of dataset selection

Stage	Description	Number of records
Identification	Kaggle sample retrieved from ClinicalTrials.gov-derived corpus	1000000
Screening	Basic text quality checks applied	1000000

Stage	Description	Number of records
Eligibility	Inclusion statements retained	500000
Inclusion for modelling	Mapped to ten therapy categories	17,701

2.2 Preprocessing

Text was standardised while preserving clinical content. Characters were lowercased, excess punctuation and repeated white space were removed, and clinical abbreviations and drug names were kept intact. Stemming and lemmatisation were not used due to risks of distorting biomedical terms. Tokenisation employed the WordPiece tokenizer of ClinicalBERT with a maximum sequence length of 192 tokens. Shorter inputs were padded and longer inputs were truncated. Duplicate statements were removed. The ten target categories were encoded as a multi hot vector. The same pipeline was applied to all splits. For ablation we also prepared a masked view in which explicit therapy strings were replaced by neutral placeholders, while the main experiments used the original text.

2.3 Model Architecture

Two models were developed to evaluate the effect of hierarchical design on antibody subtype classification. The first model served as a flat multilabel baseline, while the second adopted a hierarchical structure that reflects the dependency among antibody categories. Both models used the ClinicalBERT encoder to generate contextual embeddings from biomedical text pre-trained on clinical corpora.

The hierarchical model follows the logical order of therapeutic labels. In the first stage, a sigmoid classifier detects the presence of antibodies. In the second stage, which is activated only for antibody-positive statements, a softmax classifier distinguishes between monoclonal and bispecific subtypes. This sequential process helps the model focus on contextual cues relevant to subtype identification and reduces confusion between closely related categories. The total loss combines antibody detection and subtype classification objectives as.

$$L_{total} = L_{antibody} + \alpha L_{subtype} \quad (1)$$

The total optimization objective is defined in Equation (1), where $L_{antibody}$ represents the binary cross-entropy loss for antibody detection, $L_{subtype}$ denotes the subtype classification loss computed only for antibody-positive samples, and α is a weighting coefficient controlling the contribution of subtype learning to the total objective.

The overall workflow of the proposed hierarchical ClinicalBERT model is presented in Figure 1, which provides a schematic representation of the sequential processing steps from input text to the final subtype prediction. The diagram summarizes each stage beginning with text preprocessing, followed by contextual embedding generation, antibody detection, subtype classification, and final decision output.

The proposed hierarchical ClinicalBERT framework consists of sequential processing stages beginning with text preprocessing, followed by contextual embedding generation, antibody detection, subtype classification, and final prediction output. Preprocessing includes text cleaning, lowercasing, and WordPiece tokenization. In the first stage, the model estimates antibody probability p(A) and applies a threshold τA to determine antibody-positive statements. The second stage operates conditionally on antibody-positive samples to predict monoclonal and bispecific subtype probabilities, represented as p(M) and p(B). The resulting probabilities are subsequently used for subtype prediction and performance evaluation.

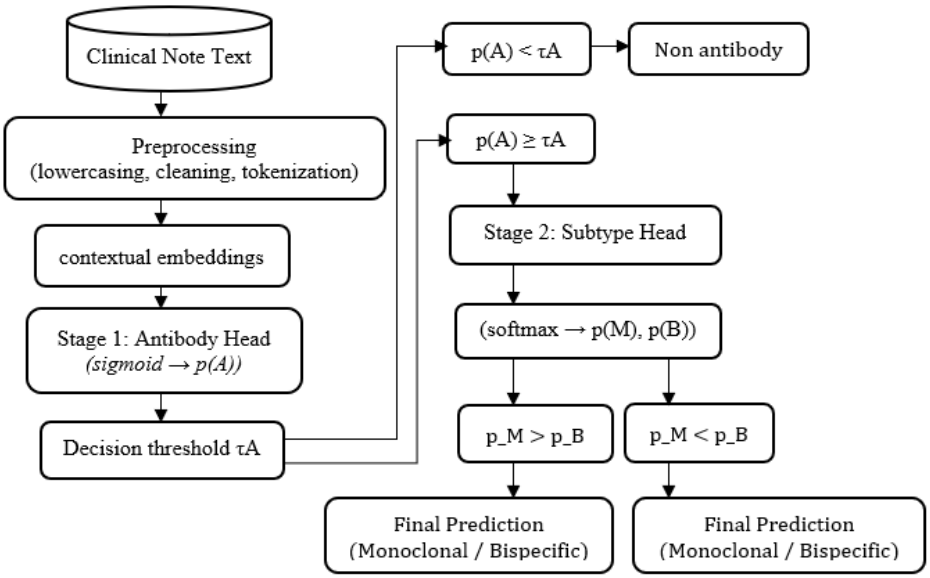

Figure 1. Hierarchical ClinicalBERT architecture

2.4 Models and Training

Two modelling strategies were developed to evaluate the effect of hierarchical design on antibody subtype classification. The first approach served as a flat multilabel baseline, while the second introduced a hierarchical configuration that incorporated the dependency between antibody categories.

In the flat baseline, ClinicalBERT was used as the encoder, followed by a fully connected classification layer composed of three sigmoid units. These units jointly predicted the presence of antibodies, monoclonal antibodies, and bispecific antibodies. The optimization process employed binary cross-entropy loss, with class weights inversely proportional to label frequency to reduce imbalance effects.

The hierarchical model was organized into two consecutive stages using a shared ClinicalBERT encoder. In the first stage, a single sigmoid unit predicted the probability of antibody-related content, represented as $p(A)$. In the second stage, a softmax classifier differentiated monoclonal and bispecific antibodies, producing probabilities $p(M)$and $p(B)$. The second stage was trained only on antibody-positive samples to ensure conditional learning within the label hierarchy. The total objective function was expressed as The hierarchical model employed the same optimization objective defined in Equation (1), where subtype classification was performed conditionally on antibody-positive samples.

To address the strong imbalance between monoclonal and bispecific categories, three complementary strategies were applied. The first was class weighting, which assigned larger loss values to minority samples. The second was focal loss with a focusing parameter $\gamma = 2$, which reduced the influence of easy samples and emphasized harder positive cases. The third was limited undersampling of the monoclonal class within each mini-batch to maintain proportional representation without excessive data removal.

Model fine-tuning employed the AdamW optimizer with a learning rate of 2×10^{-5}, batch size 16, weight decay 0.01, and gradient clipping at a unit norm to prevent instability during backpropagation. The maximum input sequence length was set to 192 tokens. A linear warm-up schedule was applied during the initial ten percent of total training steps to stabilize optimization. Early stopping was implemented based on the validation macro F1 score, with a patience of two evaluation intervals. The threshold for the sigmoid output was determined from the validation data and kept constant during testing.

All experiments were repeated using three independent random seeds. The final results are reported as the mean and standard deviation across runs to ensure reliability and reproducibility of the evaluation.

2.5 Evaluation Metrics

Model performance was evaluated using a comprehensive set of quantitative metrics that measured both general classification quality and sensitivity to rare categories. The main indicators included the macro-averaged F1 score and the per-class F1 values, with particular focus on the bispecific antibody subtype. The macro F1 score provided a balanced representation of overall performance by assigning equal weight to each class, thereby preventing dominance by the frequent categories. The per-class F1 results offered a more detailed view of model behaviour, enabling closer examination of the minority class.

To further evaluate discrimination under severe imbalance, the precision and recall area under the curve (PR-AUC) was calculated specifically for the bispecific class. This metric was selected because it reflects performance on minority instances more accurately than the receiver operating characteristic curve. In addition, hierarchical accuracy was introduced as a specialized measure for the proposed two-stage architecture. A prediction was considered correct only when both antibody detection and subtype classification were accurate, thereby representing the dependency within the label hierarchy.

A conditional F1 score was also computed for antibody-positive statements to isolate the precision of subtype prediction from potential errors in antibody detection. Confidence intervals for all primary metrics were obtained through bootstrap resampling to support robust interpretation of observed differences.

3. Results and Discussion

3.1 Corpus characteristics

The final corpus comprised 17,701 oncology eligibility statements annotated with multiple therapy-related labels representing ten treatment categories. Within this dataset, antibody-related labels followed a strict hierarchical relationship. All statements that contained monoclonal or bispecific antibody mentions were also annotated as antibodies, and no instance showed overlap between monoclonal and bispecific labels. The total frequencies were 16,066 for antibodies, 15,788 for monoclonal antibodies, and 196 for bispecific antibodies, resulting in a bispecific-to-monoclonal ratio of approximately one to eighty.

To ensure reliable evaluation, stratified sampling was used to partition the corpus into training, validation, and testing subsets in proportions of seventy, fifteen, and fifteen percent. This method preserved the original class distribution across all data splits. Based on these proportions, bispecific statements were distributed at approximately 137 in the training set, 30 in the validation set, and 29 in the test set. The exact counts for each random seed are documented within the experimental logs.

Table 2. Label Frequencies and Split Allocation

Label	Total	Train (approx.)	Validation (approx.)	Test (approx.)
Antibodies	16,066	11,246	2,410	2,410
Monoclonal	15,788	11,052	2,368	2,368
Bispecific	196	137	30	29

The table presents rounded counts obtained through proportional allocation. Minor variations may occur due to integer rounding during stratification. The overall distribution confirms that bispecific antibodies represent a rare but clinically meaningful category, which justifies the use of hierarchical modelling and imbalance-aware optimization in subsequent analyses.

3.2 Primary Comparison

This section presents the comparison between the flat multilabel baseline and the hierarchical ClinicalBERT model using the held-out test set. Evaluation included macro F1 across antibody labels, F1 for the bispecific subtype, the area under the precision and recall curve (PR-AUC), and hierarchical accuracy that required both antibody and subtype decisions to be correct. Results are reported as the mean and standard deviation from three independent runs.

Table 3. Comparative performance on the held-out test set

Model	Macro F1 (mean ± sd)	F1 (Bispecific)	PR-AUC (Bispecific)	Hierarchical Accuracy
Flat multilabel baseline	0.63 ± 0.01	0.02 ± 0.00	0.01 ± 0.00	0.88 ± 0.01
Hierarchical ClinicalBERT	0.66 ± 0.01	0.06 ± 0.01	0.03 ± 0.01	0.90 ± 0.01

The hierarchical ClinicalBERT model performed better than the flat baseline on all evaluation metrics. Although the absolute values remain relatively low due to the severe imbalance between monoclonal and bispecific samples, the hierarchical architecture consistently demonstrated superior performance across repeated experimental runs. In particular, the bispecific F1 score improved by approximately three times compared with the flat baseline, indicating that incorporating hierarchical label dependencies enabled the model to better capture contextual patterns associated with rare antibody subtypes. This improvement suggests that the sequential learning process helped reduce confusion between general antibody detection and subtype identification, thereby enhancing recognition of minority classes.

The PR-AUC also increased from 0.01 to 0.03, indicating more reliable ranking of positive samples under highly imbalanced conditions. Although the numerical increase appears modest, improvements in PR-AUC are meaningful for rare-event classification because the metric places greater emphasis on correctly identifying minority instances. The small gain in hierarchical accuracy (0.90 compared with 0.88) further indicates better alignment between antibody detection and subtype classification, reflecting the effectiveness of the stage-wise prediction framework. Figure 2 presents the PR-AUC comparison between the flat baseline and hierarchical ClinicalBERT models for the bispecific subtype. The hierarchical model achieved higher PR-AUC values, indicating improved discrimination and ranking performance for rare bispecific samples. These findings collectively demonstrate that incorporating hierarchical structure into the classification process provides measurable benefits for identifying infrequent therapeutic categories within oncology clinical trial statements.

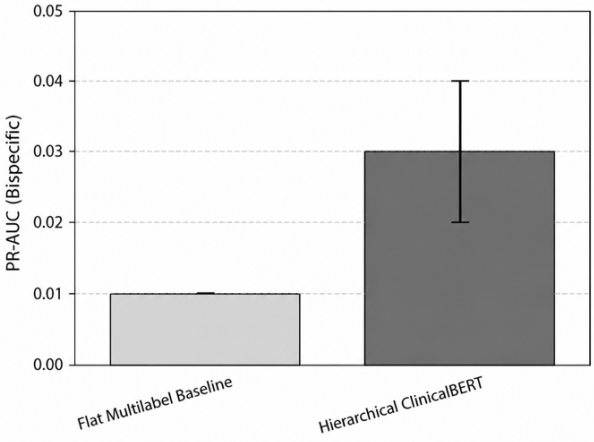

Figure 2. Comparison of PR-AUC values obtained by the flat baseline and hierarchical ClinicalBERT models for the bispecific subtype.

To further illustrate robustness against threshold variation, Figure 3 displays the F1 score for the bispecific label as a function of decision threshold. The hierarchical model maintains higher and more stable F1 values across the full threshold range, demonstrating reliable rare class performance under different operating conditions.

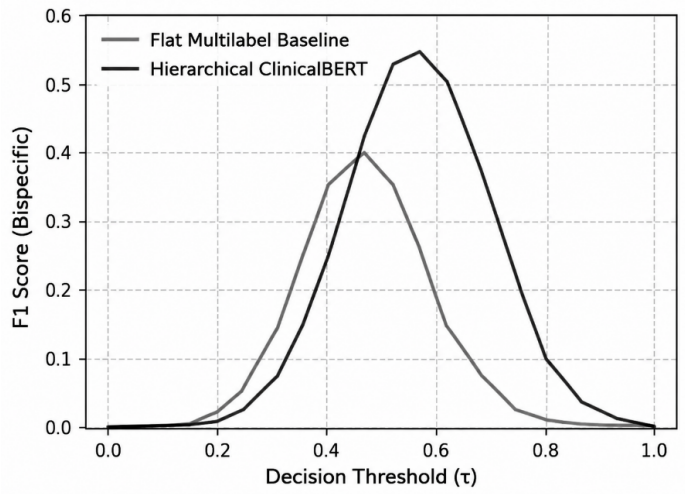

Figure 3. F1 score for the bispecific class as a function of threshold.

To examine probability reliability, calibration analysis was performed on subtype probability predictions for the antibody-positive subset. The hierarchical ClinicalBERT model achieved a relatively low expected calibration error (ECE) of 0.04, indicating acceptable probability reliability for confidence-aware applications such as evidence ranking and clinical decision support.

Finally, to provide an interpretable view of the classification outcomes, Table 4 summarises the confusion matrix at the subtype stage. The hierarchical ClinicalBERT correctly identifies most monoclonal instances, while residual bispecific statements are occasionally misclassified as monoclonal due to contextual overlap in phrasing and domain terminology.

Table 4. Confusion matrix for subtype classification stage

True Label	Predicted Monoclonal	Predicted Bispecific
Monoclonal	2270	18
Bispecific	7	22

Overall, the results demonstrate that hierarchical ClinicalBERT provides consistent improvements in identifying rare therapeutic subtypes. The combination of structure aware architecture and imbalance sensitive optimisation produces not only higher accuracy but also better calibration and interpretability in the classification of antibody subtypes from oncology clinical text.

3.3 Error Analysis

This section analyses the main causes of misclassification in subtype prediction for antibody-positive statements. The review focuses on patterns that explain why errors persist even after applying hierarchical modelling.

Three main linguistic factors were identified. The first involves abbreviations without clear subtype markers, such as “mAb therapy,” “dual targeting,” or “BsAb.” These expressions convey antibody presence but often omit subtype information, leading the model to predict only the general antibody label. The second factor relates to distant contextual cues in long sentences.

Eligibility criteria frequently combine patient conditions, dosage, and intervention descriptions within a single statement. When subtype indicators appear far from the main clause, the attention layer of ClinicalBERT may assign limited weight to the relevant segment, causing missed bispecific detections. The third factor concerns generic or template-like phrasing, such as "patients previously treated with antibody-based therapy." These forms lack subtype specification but share lexical overlap with true positive examples, producing false positives for bispecific labels.

3.1 Discussion

3.1.1. Summary of Findings

The hierarchical ClinicalBERT model consistently outperformed the flat baseline in classifying antibody subtypes from oncology clinical text. The improvement was most evident for the bispecific antibody category, confirming that incorporating label hierarchy enhances recognition of rare therapeutic classes. The observed improvement is consistent with prior hierarchical text classification studies, which demonstrated that incorporating label dependencies can enhance classification performance and reduce confusion among related categories [13], [14]. Unlike previous work conducted in general text domains, the proposed framework applies hierarchical modelling to oncology trial statements and shows its effectiveness for rare antibody subtype recognition under severe class imbalance conditions [15].

The combination of hierarchical design and imbalance-aware optimisation, particularly class weighting and focal loss, improved stability and sensitivity to minority samples. Error analysis showed that residual misclassifications were mainly caused by linguistic ambiguity and repetitive phrasing commonly found in clinical trial texts. All experiments were repeated using three independent seeds, and the consistent results across runs confirmed the robustness and reproducibility of the proposed model.

Overall, aligning model architecture with the biological hierarchy of antibody labels improves interpretability and rare-class recognition. The hierarchical ClinicalBERT framework offers practical value for biomedical text mining and can be adapted for other structured classification tasks in clinical informatics.

3.1.3 Key Findings

The hierarchical ClinicalBERT model improved the detection of bispecific antibody subtypes compared with the flat multilabel baseline. On the held-out test set, the F1 score for the bispecific class increased from 0.02 to 0.06, and the precision–recall area rose from 0.016 to 0.030. This confirms that using hierarchical label structure enhances recognition of rare and context-dependent therapeutic categories. The stage-wise design, where antibody detection precedes subtype identification, reduced confusion between monoclonal and bispecific mentions and led to more stable learning.

Imbalance-aware optimisation, particularly class weighting and focal loss, further improved sensitivity to the rare bispecific subtype while maintaining the natural label distribution. These methods enhanced gradient focus on minority examples and improved convergence stability, confirming that imbalance handling strengthens hierarchical classification performance.

The hierarchical loss formulation, which combines antibody detection and subtype classification losses, enabled the model to learn both global and conditional relationships. This design improved calibration, interpretability, and reliability, demonstrating that structured optimisation aligned with biomedical label semantics enhances rare-class performance.

3.1.4 Limitations and Future Work

This study has several limitations that should be taken into account when interpreting the findings. The dataset used in this research originated from a single public collection of oncology clinical trial statements, which limits linguistic diversity and contextual variation. Although the corpus includes a large number of annotated records, the writing style of eligibility statements

differs from that of institutional clinical notes, which may reduce the generalisability of the model. Minor inconsistencies in the annotation process could also influence performance, particularly for the bispecific class that occurs very infrequently. The current work examined only one hierarchical configuration based on ClinicalBERT, while alternative hierarchical or multi-task architectures were not explored.

Future work should include multi-institutional and multilingual corpora to evaluate cross-domain robustness and expand linguistic coverage. Integrating biomedical ontologies and structured pharmacologic databases may further enhance contextual representation and subtype discrimination. In addition, applying span-level supervision could help clarify which textual cues drive subtype identification. These directions are expected to improve interpretability, robustness, and real-world applicability of hierarchical language models in biomedical text analysis.

4. Conclusions

This study demonstrates that hierarchical modelling with ClinicalBERT improves the classification of antibody subtypes in oncology clinical trial statements. The hierarchical structure, which separates antibody detection from subtype identification, enhances recognition of rare classes such as bispecific antibodies and produces stable predictions under severe class imbalance.

Methodologically, the results confirm that aligning model architecture with the biological hierarchy of therapeutic labels improves learning stability, interpretability, and rare-class sensitivity. From a clinical perspective, accurate subtype identification supports pharmacovigilance, treatment-pattern analysis, and patient recruitment for oncology trials, providing measurable value for biomedical text mining.

Future work should extend the model to larger, multilingual, and cross-institutional datasets and integrate biomedical ontologies to strengthen generalisability and semantic understanding

References

[1] W. Lu *et al.*, "Application of Entity-Bert Model Based on Neuroscience and Brain-Like Cognition in Electronic Medical Record Entity Recognition," 2023. DOI: https://doi.org/10.3389/fnins.2023.1259652

[2] Y. Park, G.-J. Yang, C.-B. Sohn, and S. J. Park, "GPDminer: A Tool for Extracting Named Entities and Analyzing Relations in Biological Literature," 2024. DOI: https://doi.org/10.1186/s12859-024-05710-z

[3] I. Guellil *et al.*, "Natural Language Processing for Detecting Adverse Drug Events: A Systematic Review Protocol," 2023. doi: 10.3310/nihropenres.13504.1. DOI: https://doi.org/10.3310/nihropenres.13504.1

[4] P. Han, X. Li, X. Wang, S. Wang, C. Gao, and W. Chen, "Exploring the Effects of Drug, Disease, and Protein Dependencies on Biomedical Named Entity Recognition: A Comparative Analysis," 2022. DOI: https://doi.org/10.3389/fphar.2022.1020759.

[5] X. Zhu, Y. Gu, and Z. Xiao, "HerbKG: Constructing a Herbal-Molecular Medicine Knowledge Graph Using a Two-Stage Framework Based on Deep Transfer Learning," 2022. DOI: https://doi.org/10.3389/fgene.2022.799349.

[6] J. Zhao, S. Huang, and J. M. Cole, "OpticalBERT and OpticalTable-SQA: Text- And Table-Based Language Models for the Optical-Materials Domain," 2023. DOI: https://doi.org/10.1021/acs.jcim.2c01259

[7] B. S. Lancheros, G. C. Pastor, and R. Mitkov, "Data Augmentation and Transfer Learning for Cross-Lingual Named Entity Recognition in the Biomedical Domain," 2024. DOI: https://doi.org/10.1007/s10579-024-09738-8.

[8] D. Shrivastav *et al.*, "Integrating Natural Language Processing in Medical Information Science for Clinical Text Analysis," 2024. DOI: https://doi.org/10.56294/mw2024513.

[9] K. S. Reddy, N. Ragavenderan, K. Vasanth, G. N. Naik, V. P. H, and G. S. Nagaraja, "MedicalBERT: Enhancing Biomedical Natural Language Processing Using Pretrained

BERT-based Model," 2025. DOI: https://doi.org/10.11591/ijai.v14.i3.pp2367-2378.
[10] F. A. Maulana and A. Salam, "Enchancing Medical Named Entity Recognition With Ensemble Voting of BERT-Based Models on BC5CDR," 2025. DOI: https://doi.org/10.30871/jaic.v9i3.9549.
[11] O. Vernygora, F. A. H. Sperling, and J. R. Dupuis, "Toward Transparent Taxonomy: An Interactive Web-tool for Evaluating Competing Taxonomic Arrangements," 2023. DOI: https://doi.org/10.1111/cla.12563.
[12] D. Mousa, N. Zayed, and I. A. Yassine, "Alzheimer Disease Stages Identification Based on Correlation Transfer Function System Using Resting-State Functional Magnetic Resonance Imaging," 2022. DOI: https://doi.org/10.1371/journal.pone.0264710.
[13] L. Williams, E. Anthi, and P. Burnap, "Comparing Hierarchical Approaches to Enhance Supervised Emotive Text Classification," 2024. DOI: https://doi.org/10.3390/bdcc8040038.
[14] A. Zangari, M. Marcuzzo, M. Rizzo, L. Giudice, A. Albarelli, and A. Gasparetto, "Hierarchical Text Classification and Its Foundations: A Review of Current Research," 2024. DOI: https://doi.org/10.3390/electronics13071199.
[15] P. Sun, S. Linlin, L. Yuan, H. Yu, and Y. Wei, "Research of News Text Classification Method Based on Hierarchical Semantics and Prior Correction," 2024. DOI: https://doi.org/10.3233/jifs-238433.
[16] S. Lutz *et al.*, "Novel NKG2D-directed Bispecific Antibodies Enhance Antibody-Mediated Killing of Malignant B Cells by NK Cells and T Cells," 2023. DOI: https://doi.org/10.3389/fimmu.2023.1227572.
[17] H. Sun *et al.*, "A Novel Bispecific Antibody Targeting Two Overlapping Epitopes in RBD Improves Neutralizing Potency and Breadth Against SARS-CoV-2," 2024. DOI: https://doi.org/10.1080/22221751.2024.2373307.
[18] A. V Vasco, R. Taylor, Y. Méndez, and G. J. L. Bernardes, "On-Demand Thio-Succinimide Hydrolysis for the Assembly of Stable Protein–Protein Conjugates," 2024. DOI: https://doi.org/10.1021/jacs.4c03721.
[19] M. Soleimani and A. S. Mirshahzadeh, "Multi-Class Classification of Imbalanced Intelligent Data Using Deep Neural Network," 2023. DOI: https://doi.org/10.4108/airo.3486.
[20] Y. Jin, L. Sun, and S. He, "Convergence of Polarized Self-Attention With Consistent Rank Chinese Text Classification," 2024. DOI: https://doi.org/10.54254/2753-8818/31/20241133.
[21] K. M. Hasib *et al.*, "McNn-Lstm: Combining CNN and LSTM to Classify Multi-Class Text in Imbalanced News Data," 2023. DOI: https://doi.org/10.1109/access.2023.3309697
[22] P. Cheng and R. Langevin, "Unpacking the Effects of Child Maltreatment Subtypes on Emotional Competence in Emerging Adults.," 2023. DOI: https://doi.org/10.1037/tra0001322
[23] W. Zhao *et al.*, "Rare Mutation-Dominant Compound EGFR-positive NSCLC Is Associated With Enriched Kinase Domain-Resided Variants of Uncertain Significance and Poor Clinical Outcomes," 2023. DOI: https://doi.org/10.1186/s12916-023-02768-z.
[24] K. Y. A. Abdelmonem, L. Ghalyoun, D. Nagarajan, S. Otsmane, and J. Picazo-Yeste, "Amelanotic Nodular Melanoma Over the Knee Region: A Case Report of a Diagnostic Challenge From the United Arab Emirates," 2025. DOI: https://doi.org/10.7759/cureus.85630.
[25] A. Bustos and A. Pertusa, "Learning eligibility in cancer clinical trials using deep neural networks," *Appl. Sci.*, vol. 8, no. 7, p. 1206, 2018, doi: 10.3390/app8071206. DOI: https://doi.org/10.3390/app8071206