Hierarchy-Aware Dense Retrieval for Greek Cardiology Clinical Coding at ELCardioCC 2026

BIOASQ at CLEF 2026

Fernando Gallego-Donoso[1,*], Salvador Lima-López[1], Eulàlia Farré-Maduell[1], Lorenzo Pratesi[2] and Martin Krallinger[1]

[1]Barcelona Supercomputing Center, Plaça Eusebi Güell, 1-3, 08034, Barcelona, Spain

Abstract

Clinical coding in low-resource languages remains limited by the scarcity of annotated corpora, language-specific terminologies, and robust normalization systems. This paper describes our participation in the ELCardioCC shared task, where Greek cardiology discharge letters must be processed for Named Entity Recognition (NER) and Entity Linking (EL). We approach the task as a two-stage normalization pipeline: first, clinical mentions are detected with a Greek-BERT sequence labeling model; next, detected spans are linked to standardized concepts using Greek-ClinLinker, a retrieval-based EL model adapted to Greek clinical text. To support semantic normalization beyond lexical matching, we built a Greek SNOMED CT-derived terminology using automatic translation and exploiting parent and grandparent relations during representation learning. We also explore data augmentation with ELCardioCC 2025 and translated CodiEsp mention–concept pairs. Across five official submissions, our best configuration achieved an F1 score of 0.7617, with balanced recall and precision. The results suggest that hierarchy-aware dense retrieval is a promising strategy for Greek clinical EL, while also showing that span detection, data selection, and terminology-specific adaptation remain key factors for further improvement.

Keywords

Clinical coding, Named entity recognition, Entity linking, Biomedical normalization, Greek clinical text, Cardiology, Dense retrieval, SNOMED CT

1. Introduction

Cardiovascular diseases remain a major global health burden and are frequently documented in clinical narratives through diagnoses, symptoms, procedures, risk factors, comorbidities, and treatment-related information [1]. Although some of this information is represented in structured fields, a substantial proportion remains embedded in the free text of Electronic Health Records (EHRs), particularly in discharge letters and longitudinal clinical documentation. As a result, clinically relevant evidence may not be readily available for structured coding, cohort construction, clinical research, or downstream decision-support applications.

Natural Language Processing (NLP) methods can help bridge this gap by transforming unstructured clinical narratives into standardized and machine-readable information. In particular, Named Entity Recognition (NER) can identify clinically relevant textual spans, whereas Entity Linking (EL) can associate these mentions with controlled terminology concepts. This two-stage formulation is especially relevant in cardiology, where the same condition may be expressed through abbreviations, lexical variants, descriptive formulations, or partially overlapping mentions.

The International Classification of Diseases (ICD) is widely used for reporting, administrative coding, and epidemiological purposes. In contrast, SNOMED CT provides a more detailed and polyhierarchical representation of clinical concepts, including clinical findings, disorders, procedures, and contextual

CLEF 2026 Working Notes, 21 – 24 September 2026, Jena, Germany

*Corresponding author.

†These authors contributed equally.

fernando.gallego@bsc.es (F. Gallego-Donoso); salvador.limalopez@bsc.es (S. Lima-López); eulalia.farre@bsc.es (E. Farré-Maduell); lorenzopratesi3@gmail.com (L. Pratesi); martin.krallinger@bsc.es (M. Krallinger)

0000-0001-8053-5707 (F. Gallego-Donoso); 0000-0002-7384-1877 (S. Lima-López); 0000-0002-9116-1592 (E. Farré-Maduell); 0000-0002-2646-8782 (M. Krallinger)

information. This richer semantic structure can support representation learning beyond direct lexical matching, even when the final task requires predictions in an ICD-oriented coding space.

These challenges are particularly relevant for Greek clinical text, where annotated resources and terminology coverage remain limited. In this work, we address the ELCardioCC 2026 [2] shared task as a two-stage NER and EL problem for Greek cardiology discharge letters, combining Greek-BERT mention detection with hierarchy-aware dense retrieval for concept normalization.

2. Related Work

Automatic clinical coding is often framed as large-scale multi-label text classification, where a clinical document is mapped directly to one or more standardized codes. Early neural approaches were mainly evaluated on MIMIC-II and MIMIC-III, English critical-care benchmarks annotated with ICD-9 diagnosis and procedure codes [3]. In this setting, convolutional and recurrent encoders were used to model discharge summaries, while label-wise attention became a common mechanism to learn code-specific document representations and to identify textual evidence for each prediction [4]. Later work improved these architectures with deeper convolutional encoders, multiple filter sizes, residual connections, and recurrent label-aware representations to better handle long documents, variable-length clinical evidence, and imbalanced code distributions [5, 6]. When pretrained language models became the dominant encoders in NLP, they were also explored for clinical coding. However, this transition is not straightforward: clinical notes often exceed the input length of standard transformers, the number of possible codes remains large and highly imbalanced, and the language of clinical documentation differs substantially from general pretraining corpora [7].

Beyond English, clinical coding has been addressed in Spanish, Brazilian, French, and Dutch clinical documents, including cardiology discharge letters [8, 9, 10]. Following this line, shared tasks such as CodiEsp [11] and ELCardioCC 2025 [12] have provided comparable benchmarks for clinical coding.

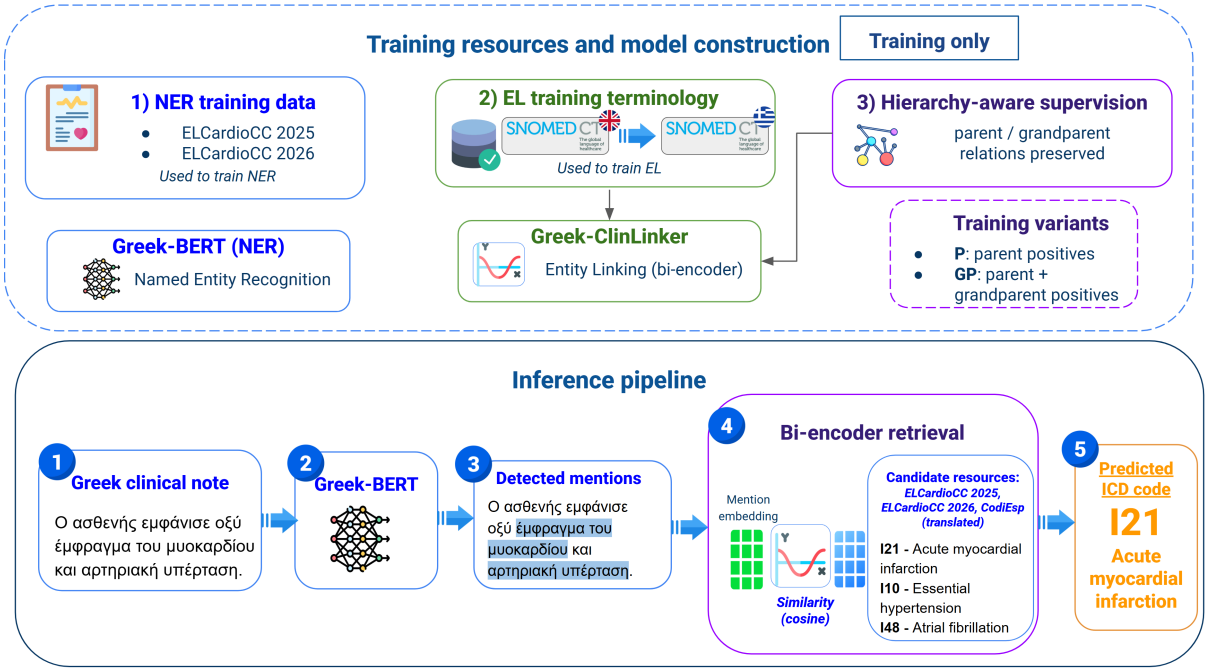
In this paper, we address the ELCardioCC 2026 task as a NER and EL problem. Our system first detects clinical mentions in Greek cardiology discharge letters and then links them to standardized concepts using a retrieval-based EL approach inspired by previous biomedical normalization methods [13, 14].

3. Material & Methods

In this paper we present a two-stage methodology for the ELCardioCC shared task, targeting NER and EL in Greek cardiology reports. The overall design follows the structure of the task itself: clinical mentions are first identified in free-text discharge summaries and are then mapped to normalized clinical concepts. In this setting, the NER stage provides the textual spans to be normalized, whereas the EL stage is responsible for resolving each span to the corresponding concept representation.

Figure 1 summarizes the complete workflow. We first describe the data sources used throughout the study, including the ELCardioCC 2025 release, the new ELCardioCC 2026 data, and the external CodiEsp [11] corpus used for augmentation (Section 3.1). Next, we introduce the mention detection component, which is formulated as a transformer-based sequence labeling problem over Greek clinical text (Section 3.2). Building on the detected mentions, we finally describe the normalization component, which constitutes the main methodological contribution of this work (Section 3.3). The proposed EL strategy is based on Greek-ClinLinker, a semantic retrieval framework in which mention and concept representations are projected into a shared embedding space. To support normalization beyond surface-level lexical matching, we built a Greek SNOMED-CT-derived terminology through automatic translation and incorporated hierarchical concept information during training. This design allows the model to exploit both textual similarity and ontology-driven semantic structure when ranking candidate concepts.

Figure 1: Overview of the proposed two-stage workflow. Greek-BERT detects mentions, while Greek-ClinLinker retrieves the best ICD code from ELCardioCC 2025/2026 and translated CodiEsp candidate resources.



3.1. Data distribution

The experimental setting combines the official ELCardioCC data with two additional sources used to increase the amount and diversity of training examples. Table 1 reports the main corpus-level statistics for each split, including the number of documents, the amount of text, the number of annotated mentions, and the number of unique normalized codes. Parenthetical values complement these global counts with document-level information: the number of reports without gold-standard annotations in the *Documents* column, the average number of words per report in the *Words* column, and the average number of annotated mentions per report in the *Mentions* column.

Table 1
Dataset statistics across splits.

Split	Documents (w/o anns)	Words (avg. words/doc.)	Mentions (avg. mentions/doc.)	Unique codes
Train 2025	1,000 (0)	297,109 (297.11)	10,168 (10.17)	144
Train 2026	2,500 (1,007)	702,038 (280.82)	11,979 (4.79)	85
Test	500 (500)	131,104 (262.21)	—	—
CodiEsp	1,000 (0)	396,988 (396.99)	18,435 (18.44)	1,853

The official ELCardioCC 2026 training release constitutes the primary corpus of the current shared task. It contains 2,500 Greek cardiology reports and 702,038 words, with an average document length of 280.82 words. However, only part of this collection includes annotations: 1,007 documents do not contain gold-standard mentions. As a result, the 2026 training split provides 11,979 annotated mentions, corresponding to 4.79 mentions per document on average. This configuration makes the 2026 release the central in-domain resource for the task, while also reflecting a relatively sparse annotation density at the document level.

To complement the official 2026 training material with additional in-domain supervision, we also used the ELCardioCC 2025 release from the previous edition of the task. This split contains 1,000 fully annotated Greek cardiology reports, comprising 297,109 words and 10,168 mention-level annotations. Although it is smaller than the 2026 training release in terms of both documents and total word count,

it is considerably denser in annotations, with 10.17 mentions per document on average. This makes the 2025 release especially useful for training the span detection component, since all documents contain gold-standard annotations and the average number of mentions per report is more than twice that of the 2026 split. Because it shares the same language, clinical domain, and annotation setting as the current task, we used it together with the 2026 training data for the NER stage, and also as additional in-domain supervision for the EL component.

The second additional resource was CodiEsp, which was incorporated as an external augmentation corpus for the EL stage. As shown in Table 1, CodiEsp contains 1,000 clinical cases and 396,988 words, with an average document length of 396.99 words. It also provides 18,435 annotated mentions, corresponding to 18.44 mentions per document, and a substantially larger concept inventory than the ELCardioCC splits, with 1,853 unique non-granular codes. This high annotation density and broader label space make CodiEsp useful for increasing the lexical and conceptual variability of mention–code pairs during semantic retrieval training. However, since CodiEsp is originally written in Spanish and does not follow the same Greek NER setting as ELCardioCC, it was not used to train the NER model. Instead, its annotated mentions were automatically translated into Greek and incorporated into the candidate resources used by the EL component at inference time.

Finally, the official test set contains 500 Greek cardiology reports and 131,104 words, with an average length of 262.21 words per document. Since gold-standard mentions and normalized codes were not released for this split, all test documents are reported as documents without annotations in Table 1, and the mention and code columns are left undefined. The test set was therefore used only for final prediction and evaluation, and not for model training.

This data configuration assigns a specific role to each resource. The ELCardioCC 2026 release defines the official in-domain training scenario, and the ELCardioCC 2025 release contributes additional Greek annotations that are directly compatible with the current task. The translated CodiEsp data are instead used as auxiliary supervision for retrieval-based EL, increasing the diversity of mention–concept associations without introducing out-of-domain text into the NER training stage.

3.2. Mention Detection with Greek-BERT

Given the annotated corpora described in Section 3.1, we formulated the NER stage as a supervised token classification problem. Each clinical report was represented as a sequence of tokens, and the model was trained to assign an IOB2 label to each token in order to recover the boundaries of clinical mentions. In this scheme, the beginning of a mention is marked with a `B` tag, subsequent tokens belonging to the same mention are marked with `I`, and tokens outside annotated spans are assigned the `O` label. The predicted spans are then passed to the EL component as mention candidates for normalization.

For the NER component, we fine-tuned a transformer encoder based on Greek-BERT[1] [15]. This model was selected because it slightly outperformed XLM-RoBERTa [16] in the previous edition of the task [17]. Although ensemble strategies combining multiple multilingual and monolingual backbones are often competitive for sequence labeling tasks, we restricted our experiments to a single backbone due to computational constraints. This design keeps the mention detection component lightweight and reproducible, while allowing the main methodological effort of this work to focus on the EL stage.

The NER model was trained using the annotated ELCardioCC 2025 and 2026 training data. Both releases provide in-domain Greek cardiology reports with mention-level annotations, adapting them for supervised span detection. In contrast, the translated CodiEsp data were not used for NER training, since they originate from a different language and annotation setting and were incorporated only as auxiliary supervision for the EL component. The Greek-BERT model was fine-tuned for 8 epochs with a learning rate of 2×10^{-5} and a batch size of 32; no validation split was used, as the main methodological focus of this work was on retrieval-based EL and automatic terminology translation.

[1] https://huggingface.co/nlpaueb/bert-base-greek-uncased-v1

3.3. Hierarchy-Aware Entity Linking with Greek-ClinLinker

Building on the mention spans produced by the NER component, we addressed EL as a semantic retrieval problem. Instead of treating normalization as a closed-set classification task over the codes observed during training, each detected mention was encoded and compared against a set of candidate clinical concepts represented in the same embedding space. This formulation is particularly suitable for biomedical normalization, where the number of possible concepts is large, surface forms are highly variable, and unseen concepts may appear at inference time. Our approach follows the retrieval-based line of work explored in previous Spanish biomedical EL studies, where ClinLinker bi-encoders were trained to align mentions and ontology concepts through contrastive learning [13].

The main difficulty in adapting this strategy to ELCardioCC is the lack of Greek clinical ontologies with a level of granularity and hierarchical structure comparable to SNOMED-CT. Although the official task annotations are provided using ICD-style codes, our objective was to exploit a richer conceptual space during representation learning. SNOMED-CT provides a fine-grained biomedical terminology with explicit parent–child relations between concepts, which can be used to inject ontological structure into the training process. However, SNOMED-CT is not directly available in Greek with sufficient lexical coverage for this task. To address this limitation, we started from the English SNOMED-CT terminology and automatically translated concept names into Greek. The resulting resource preserves the original concept identifiers and hierarchical relations, while providing Greek textual forms that can be encoded by the retrieval model.

Following this terminology construction step, we trained two Greek-ClinLinker variants. The first model, Greek-ClinLinker-P, follows a pair-based bi-encoder formulation in which clinical mentions and concept descriptions are independently encoded. During training, positive pairs are constructed from mention–concept associations, and the model learns to place the mention representation close to the representation of its corresponding concept. Retrieval is then performed by computing vector similarity between the encoded mention and the encoded candidate concepts. This model constitutes the basic ontology-aware semantic retrieval setting used in our work.

The second model, Greek-ClinLinker-GP, extends this formulation by incorporating hierarchical information from the translated SNOMED-CT-derived terminology. In addition to direct mention–concept pairs, the training signal is enriched with concepts that are hierarchically related to the target concept. More specifically, we exploit local ontological relations by considering parent concepts, and, in the extended variant, higher-level ancestors, as additional semantically related positives during training. This design follows the intuition that clinically related concepts should occupy nearby regions of the embedding space, while still requiring the model to discriminate between fine-grained alternatives. The use of hierarchical information is therefore intended to improve candidate retrieval by moving beyond purely lexical or synonym-based alignment. Both Greek-ClinLinker variants were fine-tuned for one epoch using the multi-similarity loss adopted in SapBERT [13], with a learning rate of 2×10^{-5}.

To avoid relying exclusively on the limited number of Greek ELCardioCC annotations, we also incorporated the translated CodiEsp mentions described in Section 3.1. These examples were not used for span detection, but they provided additional mention–concept pairs for training the EL retrieval model. In practice, the translated CodiEsp data increased the lexical diversity of clinical mentions and exposed the model to a broader set of normalized concepts. This augmentation is especially relevant for a similarity-based EL approach, since the quality of the learned embedding space depends on the diversity and coverage of the mention–concept associations observed during training.

At inference time, each mention predicted by the NER system is encoded and compared with the concept representations stored in the candidate terminology. Candidate concepts are ranked according to their semantic similarity to the mention, and the top-ranked concept is selected as the normalized prediction. This inference scheme provides a common retrieval framework in which different sources of supervision and matching signals can be incorporated without modifying the preceding NER component. Therefore, the final ELCardioCC submissions were designed as a set of progressively enriched configurations of the same semantic retrieval pipeline.

The first two submissions focused on the effect of hierarchical enrichment. The first run used Greek-

ClinLinker-P, based on parent-level relations in the translated SNOMED-CT-derived concept graph. The second run evaluated Greek-ClinLinker-GP, which extends this formulation by also incorporating grandparent-level relations during training. These two configurations allowed us to compare whether local hierarchical information was sufficient for the task or whether a deeper expansion of the ontology could provide additional benefit.

Since the number of official submissions was limited to five, the remaining runs were built around the configuration that we considered most promising according to our previous Spanish EL experiments [14, 18]. In those studies, parent-level enrichment generally provided a more selective and stable training signal than deeper parent-plus-grandparent expansion. We therefore used the parent-based Greek-ClinLinker configuration as the basis for the subsequent submissions and progressively added external augmentation and lexical matching components.

The third submission, `Greek_ClinLinker-GP_greek_bert_DA_codiesp`, incorporated data augmentation from the automatically translated CodiEsp corpus. This run was intended to evaluate whether additional mention–concept pairs from an external clinical corpus, once translated into Greek, could improve the semantic structure of the retrieval space. The fourth submission, `Greek_ClinLinker-P_greek_bert_DA_codiesp_enriched_sm`, further complemented the augmented dense retriever with exact string matching. This lexical component was introduced to recover high-confidence cases in which the detected mention directly matched a concept label or synonym in the candidate terminology.

Finally, the fifth submission, `Greek_ClinLinker-P_greek_bert_DA_codiesp_enriched_all`, added fuzzy matching on top of the previous configuration. This last step was designed to capture near-exact variants, including minor spelling differences, inflectional variation, and orthographic inconsistencies that are frequent in clinical text. Overall, the submitted systems follow a gradual enrichment strategy: starting from dense hierarchy-aware retrieval, then incorporating translated external supervision, and finally combining semantic retrieval with exact and approximate lexical matching signals.

4. Results

Table 2 reports the official ELCardioCC results for the five submitted runs. The best-performing configuration was Greek_ClinLinker-P_greek_bert_DA_codiesp, which obtained an F1 score of 0.7617, with a recall of 0.7915 and a precision of 0.7340. The closely related Greek_ClinLinker-GP_greek_bert_DA_codiesp run obtained an F1 score of 0.7557. Within this comparison, the parent-based configuration slightly outperformed the parent-plus-grandparent variant, suggesting that a more local hierarchy-aware signal was preferable in the submitted setting.

The parent-based configuration without the additional translated CodiEsp candidate resource obtained an F1 score of 0.7510. Thus, the inclusion of the translated CodiEsp resource was associated with a 1.07-point increase in F1 in this specific end-to-end comparison. This result should be interpreted cautiously: it reflects the behaviour of the complete submitted configuration, including the candidate pool available at inference time, rather than an isolated estimate of the contribution of any single resource or training decision.

The two lexically enriched submissions obtained very high recall but very low precision. Both runs reached a recall of 0.9728, whereas precision decreased to 0.0570, resulting in an F1 score of 0.1077. This pattern indicates that the matching rules substantially increased the number of predicted spans, but generated redundant, overlapping, or overly narrow predictions under the official exact-match evaluation. Consequently, lexical matching did not improve the final end-to-end system in its current form.

The official evaluation was end-to-end and performed on a blind test set. Therefore, the reported scores jointly reflect the effects of span detection, span-boundary accuracy, representation learning, candidate-resource coverage, and ranking. Given the limit of five official submissions, the submitted runs should be interpreted as constrained system-level comparisons rather than as a complete factorial

Table 2
Official ELCardioCC results for the five submitted runs.

Submission	Recall	Precision	F1
`Greek_ClinLinker-P_greek_bert_DA_codiesp`	0.7915	0.7340	0.7617
`Greek_ClinLinker-GP_greek_bert_DA_codiesp`	0.7864	0.7274	0.7557
`Greek_ClinLinker-P_greek_bert`	0.7919	0.7142	0.7510
`Greek_ClinLinker-P_greek_bert_DA_codiesp_enriched_all`	0.9728	0.0570	0.1077
`Greek_ClinLinker-P_greek_bert_DA_codiesp_enriched_sm`	0.9728	0.0570	0.1077

ablation study. The results do not isolate the independent contribution of the NER component, translated SNOMED CT terminology, hierarchy-aware supervision, or individual candidate resources.

5. Discussion

The official results indicate that the proposed two-stage normalization pipeline is a promising approach for Greek cardiology clinical coding. The best submitted configuration achieved an F1 score of 0.7617, with a relatively balanced trade-off between recall and precision. This is encouraging in a low-resource setting, where Greek clinical terminologies with broad coverage and explicit semantic structure remain limited. The results therefore suggest that hierarchy-aware dense retrieval deserves further investigation for Greek clinical EL.

The submitted configurations also provide useful initial insights into the design of the EL component. The parent-based Greek-ClinLinker configuration outperformed the parent-plus-grandparent variant by 0.60 F1 points, suggesting that a more local hierarchy-aware signal may be preferable to a broader ancestor expansion in this setting. Likewise, the configuration including translated CodiEsp candidate resources achieved a modest improvement over its corresponding configuration without them. Although these comparisons do not isolate every modeling decision, they indicate promising directions for refining hierarchy-aware retrieval and candidate-resource design.

The official scores reflect the behaviour of the complete pipeline, combining NER quality, span boundaries, representation learning, candidate-resource coverage, and ranking. The shared-task setting focused development on a small number of complete configurations, evaluated on a blind test set and under a limit of five official submissions. This makes the current results particularly useful as an end-to-end assessment of the proposed approach, while also identifying several components that can be examined in greater detail in a subsequent stage of the work.

One important question concerns the relative contribution of NER and EL errors. Mention detection relied on a single Greek-BERT sequence-labeling model, without ensembles or an extensive hyperparameter search. Since EL operates directly on predicted spans, missed mentions, incomplete spans, and boundary mismatches can propagate to the final normalization output. A controlled comparison of EL performance using predicted and gold-standard spans would help determine whether the main opportunity for improvement lies in mention detection, concept representation, candidate coverage, or ranking.

A further aspect to explore concerns the terminology resources used during training and inference. Greek-ClinLinker was trained with SNOMED CT concept labels automatically translated from English to Greek, while preserving the original concept identifiers and hierarchical relations. Parent and grandparent relations were used as additional positive signals during representation learning. In contrast, SNOMED CT was not used as the output space of the submitted system. At inference time, detected mentions were compared against candidate representations derived from ELCardioCC 2025, ELCardioCC 2026, and translated CodiEsp resources, and the final output was an ICD-oriented code. Thus, SNOMED CT provides auxiliary lexical and hierarchical supervision during training, whereas the candidate resources available at inference time define the final coding space.

The translated SNOMED CT terminology was generated using a proprietary translation service developed by TRANSLATED. We adopted this approach because SNOMED CT offers a highly detailed

and polyhierarchical clinical terminology, providing semantic relations that are not available in the Greek resources used in this task. As no Greek SNOMED CT release with comparable coverage and hierarchical structure was available, translating the English terminology allowed us to use its lexical and relational information as auxiliary supervision during training. Details regarding the translation system's architecture, training data, and decoding configuration cannot be disclosed. The resulting resource should therefore be understood as an auxiliary experimental terminology rather than as a clinically validated Greek SNOMED CT release.

The results also highlight opportunities to adapt ClinLinker more closely to an ICD-oriented candidate space. ClinLinker was originally developed for broader biomedical normalization settings, whereas ELCardioCC focuses on a more restricted set of cardiology-related diagnoses and conditions. Future work could therefore explore terminology-specific candidate construction, negative sampling, similarity calibration, and hierarchy-aware objectives. This may be especially relevant for disease normalization, which has proven more challenging than the normalization of procedures or symptoms in previous ClinLinker experiments.

Finally, the lexically enriched submissions provide a clear indication of the importance of span selection and post-processing in exact-match evaluations. Exact and fuzzy matching substantially increased recall, but redundant, overlapping, or overly narrow predictions severely affected precision. Rather than ruling out lexical matching, this result suggests that it should be combined with span consolidation, overlap resolution, and confidence-based filtering before being integrated with dense retrieval.

6. Conclusions

In this work, we presented a two-stage approach for the ELCardioCC shared task, addressing Greek cardiology clinical coding as a NER and EL problem. The system first detects clinical mentions in discharge letters and then normalizes them through a retrieval-based Greek-ClinLinker model enriched with hierarchy-aware concept representations. Across the five official submissions, the best configuration achieved an F1 score of 0.7617, indicating that dense retrieval combined with in-domain mention detection is a promising direction for Greek clinical EL. The approach uses translated SNOMED CT as auxiliary hierarchical supervision during representation learning, while final predictions are retrieved from the ICD-oriented candidate resources derived from ELCardioCC 2025, ELCardioCC 2026, and translated CodiEsp.

The submitted runs also identify clear opportunities to extend this work. In particular, the relative contribution of NER quality, hierarchy-aware supervision, candidate-resource composition, and lexical matching remains to be studied in a more controlled setting. Future work will focus on error analysis and targeted ablations, including alternative NER configurations, evaluation with gold-standard spans, span consolidation strategies, and terminology-specific adaptations of the EL model. These analyses will help identify the main bottlenecks and refine the most promising components of the proposed framework for cardiology-related disease coding.

Acknowledgments

This work was funded by the European projects DataTools4Heart (Grant Agreement No. 101057849) and AI4HF (Grant Agreement No. 101080430). Fernando Gallego-Donoso and Salvador Lima-López fellowship within the "Generación D" initiative, Red.es, Ministerio para la Transformación Digital y de la Función Pública, for talent attraction (C005/24-ED CV1). Funded by the European Union NextGenerationEU funds, through PRTR.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools in the preparation of this manuscript.

References

[1] World Health Organization, Cardiovascular diseases (cvds), 2025. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), accessed: 2026-05-28.

[2] D. Dimitriadis, V. Patsiou, E. Stoikopoulou, A. Toumpas, A. Bekiaridou, A. Samaras, G. Giannakoulas, G. Tsoumakas, Overview of ElCardioCC Task on Clinical Coding in Cardiology at BioASQ 2026, in: E. Sánchez Salido, A. Barrón-Cedeño, A. García Seco de Herrera, S. MacAvaney, J. M. Struß (Eds.), CLEF 2026 Working Notes, 2026.

[3] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R. G. Mark, Mimic-iii, a freely accessible critical care database, Scientific data 3 (2016) 1–9.

[4] J. Mullenbach, S. Wiegreffe, J. Duke, J. Sun, J. Eisenstein, Explainable prediction of medical codes from clinical text, in: Proceedings of the 2018 conference of the north American chapter of the association for computational linguistics: human language technologies, volume 1 (long papers), 2018, pp. 1101–1111.

[5] F. Li, H. Yu, Icd coding from clinical text using multi-filter residual convolutional neural network, in: proceedings of the AAAI conference on artificial intelligence, volume 34, 2020, pp. 8180–8187.

[6] T. Vu, D. Q. Nguyen, A. Nguyen, A label attention model for icd coding from clinical text, in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI'20, 2021.

[7] C.-W. Huang, S.-C. Tsai, Y.-N. Chen, Plm-icd: Automatic icd coding with pretrained language models, in: Proceedings of the 4th clinical natural language processing workshop, 2022, pp. 10–20.

[8] A. D. Reys, D. Silva, D. Severo, S. Pedro, M. M. de Sousa e Sá, G. A. Salgado, Predicting multiple icd-10 codes from brazilian-portuguese clinical notes, in: Brazilian Conference on Intelligent Systems, Springer, 2020, pp. 566–580.

[9] Y. Tchouka, J.-F. Couchot, D. Laiymani, P. Selles, A. Rahmani, Automatic icd-10 code association: A challenging task on french clinical texts, in: 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2023, pp. 91–96.

[10] A. Sammani, A. Bagheri, P. G. van der Heijden, A. S. Te Riele, A. F. Baas, C. Oosters, D. Oberski, F. W. Asselbergs, Automatic multilabel detection of icd10 codes in dutch cardiology discharge letters using neural networks, NPJ digital medicine 4 (2021) 37.

[11] A. Miranda-Escalada, A. Gonzalez-Agirre, J. Armengol-Estapé, M. Krallinger, Overview of automatic clinical coding: annotations, guidelines, and solutions for non-english clinical cases at codiesp track of CLEF eHealth 2020, in: Working Notes of Conference and Labs of the Evaluation (CLEF) Forum, CEUR Workshop Proceedings, 2020.

[12] D. Dimitriadis, V. Patsiou, E. Stoikopoulou, A. Toumpas, A. Kipouros, D. Papadopoulos, A. Bekiaridou, K. Barmpagiannos, A. Vasilopoulou, A. Barmpagiannos, et al., Overview of elcardiocc task on clinical coding in cardiology at bioasq 2025, in: CLEF, 2025.

[13] F. Liu, E. Shareghi, Z. Meng, M. Basaldella, N. Collier, Self-alignment pretraining for biomedical entity representations, in: Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies, 2021, pp. 4228–4238.

[14] F. Gallego, G. López-García, L. Gasco-Sánchez, M. Krallinger, F. J. Veredas, Clinlinker: Medical entity linking of clinical concept mentions in spanish, in: International Conference on Computational Science, Springer, 2024, pp. 266–280.

[15] J. Koutsikakis, I. Chalkidis, P. Malakasiotis, I. Androutsopoulos, Greek-bert: The greeks visiting sesame street, in: 11th Hellenic Conference on Artificial Intelligence, SETN 2020, Association for Computing Machinery, New York, NY, USA, 2020, p. 110–117. URL: https://doi.org/10.1145/3411408.3411440.

[16] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th annual meeting of the association for computational linguistics, 2020, pp. 8440–8451.

[17] B. Velichkov, A. Datseris, S. Vassileva, S. Boytcheva, Enigma@ elcardiocc: bridging ner and icd-10 entity linking-a hybrid method for greek clinical narratives, in: CLEF, 2025.
[18] F. Gallego, G. López-García, L. Gasco, M. Krallinger, F. J. Veredas, Clinlinker-kb: clinical entity linking in spanish with knowledge-graph enhanced biencoders, Available at SSRN 4939986 (2024).