Appendix. Cover letter

Dear Editors,

Thank you for giving us the opportunity to resubmit our manuscript! We appreciate the insightful comments provided by the reviewers, as it has helped us improve the quality of our paper. In this revised version, we have made the following main changes:

- In the Related Work section (Section 3), we have added a new summary table (Table 1) to visually compare existing NeSy-related studies with ours. We also provide an overview of the key differences between this extended version and our previous NeSy 2024 paper, and have incorporated the paper recommended by the reviewer to enrich the related work section.
- We reported standard deviations over multiple runs in our result tables. Additionally, we provide more results in the Appendix, with a particular focus on recall metrics.
- We added legends and improved the caption for Figures 3-10 to make them more understandable.
- In the Appendix we added an overview of the dataset features.
- We added supplementary material [9], where all approaches are illustrated using the boxology notation from [50]
- We corrected all the typos.

We hope these revisions address all the reviewers' comments, improve the manuscript's clarity and make it suitable for publication. Below, we have included a detailed response letter where we address each of the reviewers' comments individually.

The Author Team

Appendix. Response letter

Dear Reviewers,

We sincerely appreciate the time and effort you have invested in reviewing our manuscript. Your thorough and insightful feedback has been invaluable in refining our work, strengthening its clarity, and improving its overall quality.

In this response letter, we have reproduced the reviewers' comments as received, explicitly numbering each comment raised. We then provide a detailed response to each comment and highlight the corresponding revisions made in the manuscript where applicable.

We hope that the revisions and improvements we have incorporated make our submission suitable for acceptance.

The Author Team

Review 1:

The paper explores the enhancement of machine learning (ML) predictions in data-scarce environments through semantic-based data augmentation leveraging knowledge graphs (KGs). It enriches tabular datasets with various KG-derived embeddings and evaluates their impact on the predictive performance of ML models (e.g., KNN, SVM, XGBoost, Neural Networks) across different embedding techniques and augmentation strategies. The methodology is applied to binary classification tasks for heart disease and chronic kidney disease using public datasets. The findings demonstrate notable improvements, particularly when distance-based KG features are incorporated, with XGBoost and Neural Networks showing the most significant gains.

[9] https://semsys.ai.wu.ac.at/data-augmentation/home.html.

The paper fits well within the scope of the Neuro AI journal. It is well-written and clear. Therefore, I recommend accepting the paper with minor revisions. Below are some suggested improvements.

Presentation:

R1C1: Include a summary of the main changes in this extended version compared to the NeSy 2024 paper.
Answer*: We have included a summary of the main changes of this paper compared to our NeSy 2024 paper in the last subsection of the Related Work section. This summary highlights the additional approaches we propose, the formalization of the approaches, the expanded methodology with two more KG embedding algorithms, and the more comprehensive evaluation applied to heart and chronic kidney disease prediction.*

R1C2: Add a summary table in the related works section to visualize the relationship between the current work and NeSy-related studies.
Answer*: We have included Table 1 in the related works section, which provides an overview of NeSy-related studies. This table outlines the types of semantic knowledge used, the domains or tasks covered, the ML models applied, the incorporation of KGEs, and the integration methods employed. This addition helps to visually summarize the relationship between our work and existing NeSy-related studies, making it easier to compare different approaches.*

R1C3: Position tables summarizing the results in the relevant sections of the main text.
Answer*: As recommended, we have moved the tables summarizing the results to the relevant sections of the main text to improve readability.*

R1C4: Move all algorithms to an appendix for better readability.
Answer*: We considered moving all algorithms to an appendix but found that it compromised readability due to frequent back-and-forth references.*

R1C5: Utilize Figure 14 in the main text to reference all approaches instead of having separate figures.
Answer*: Similar to our response to R1C4, we considered referencing all approaches in Figure 14 instead of using separate figures but found that it compromised readability.*

Experiments and Results:

R1C6: Section 6.2: While the paper reports an averaged performance across three embedding dimensions to ensure robustness, it is common practice to also average the results of multiple runs for each experiment and report the standard deviation.
Answer*: In response to the reviewer's suggestion, we have included the standard deviations to capture variability between runs for Tables 5 and 7 in the main text, as well as for Tables 11 and 12 in the appendix, based on experiments conducted across three embedding dimensions.*

R1C7: Table 2: Provide details on how the hyperparameters for each embedding method were selected.
Answer*: As suggested by the reviewer, we have added details on the selection of hyperparameters in Table 3 (previously Table 2). Specifically, we explain that the embedding dimensions ([64, 128, 100]) were selected to provide a range of vector sizes that are large enough to capture meaningful patterns but small enough to maintain computational efficiency. For Node2Vec and RDF2Vec, the walk length and the number of walks per node were adapted to the size and complexity of each ontology. For smaller ontologies, shorter walks and fewer iterations, while larger or more complex ontologies required slightly longer walks.*

R1C8: Section 7: Clarify how the impact of KGs was computed. Since the ontologies are used to build the KGs and subsequently implement the described approaches, specify whether the reported value represents an average across all these approaches.

***Answer**: We have clarified that the reported results represent the average accuracy and F2 score across all evaluated approaches implemented with each ontology.*

Minor comments:

R1C9:

- p10, line 44: approachess -> approaches
- p10, line 51: c_j for each target class is computed. How is the centroid computed?
- p11, line 19: and no classes -> and noDisease classes.
- p11 in Alg3, $\vec{v_i}$ is not defined
- p12, line 43: In this approach, referred to as EmbedClusterAugTab -> In this approach, referred to as ClusterAugTab
- p12, line 45: Algorithm 6 -> Algorithm 5
- p17, line 33: we only used on the third approach -> we only used the third approach
- p17, line 39: Detailed descriptions -> I would rather say 'An overview of these models is provided in Section 2.'
- p19, line 34: in the other hand -> on the other hand

***Answer**: We have addressed all minor comments by correcting the reported typos, clarifying how the centroids for the target classes are computed, and defining $\vec{v_i}$ in Algorithm 3.*

Review 2:

[Paper Summary]

The paper explores the integration of Knowledge Graphs into Machine Learning pipelines to address challenges in data-scarce domains. The authors hypothesize that enriching training datasets with semantic information derived from KGs can improve ML models' predictive capabilities. The paper explores three primary research objectives: identifying optimal methods for integrating KG-derived features into ML pipelines; analyzing the impact of various KG embedding techniques on model performance; and comparing the effectiveness of ML algorithms when augmented with KG information. To address the objective the authors tested five sub-hypotheses across eight approaches and conducted experiments on binary classification tasks focused on predicting heart and chronic kidney diseases. The results indicate substantial improvements when models are augmented with distance features derived from KG embeddings.

[Review]

I reviewed this paper for NeSy 2024, and I am pleased to note that it has significantly improved since then. The authors have addressed almost all my previous comments, resulting in a much stronger and more refined submission. The methodology and results are now presented more clearly, and the contributions are well-articulated. Additionally, the paper has been expanded considerably with new content. The paper in its current form is well-prepared and demonstrates substantial progress. However, I still have a few minor comments that could further enhance the quality and clarity of the work:

R2C1: The paper briefly mentions the types of features used in the datasets but does not provide detailed specifications. It would be beneficial to clarify whether the features are categorical, continuous, or integers. Including a table in the appendix that lists each feature along with its type and any relevant characteristics would enhance clarity and reproducibility.

***Answer**: As suggested by the reviewer, we have added Tables 9 and 10 in the Appendix A to provide more details for the features in the used datasets.*

R2C2: It is unclear whether literals are considered part of the entity set E. If literals are not included in E, the graph should be formally represented KG =(E,L,R',Tr) where L denotes the set of literals. Additionally, the paper should explain how literals are mapped into the embedding space, as their representation may influence the model's effectiveness.

Answer*: We appreciate the reviewer's insightful comment. In our approach, literals are* ***not*** *explicitly included in the entity set E, therefore, we have updated the formal representation of the KG to $KG = (E, L, R', Tr)$, where L represents literals, which are currently not included in the embedding space. This is because the embedding models we use do not handle literals unless they are explicitly converted into entities or nodes. We acknowledge that the representation of literals may influence the model's effectiveness, and as part of future work, we plan to explore methods such as transforming literals into entity nodes or using other models that support literals (such as TransEA, LiteralE).*

R2C3: The notation for the embedding function requires more consistency. Initially, it is defined as $\phi : E \cup R \rightarrow \mathbb{R}^d$ mapping entities and relations into a d-dimensional space. Later, the notation shifts to $\phi : P \rightarrow \mathbb{R}^d$, where $P \subseteq E$.

Answer*: We have ensured consistency in the notation for the embedding function by explicitly using $\phi : P \cup R \rightarrow \mathbb{R}^d$ throughout the manuscript where applicable*

R2C4: The meaning of the yellow sections in the "augmented tabular data" block is not explained in Figures 3–10. Please provide a legend or annotation within the figures.

Answer*: We have added legends to all figures (Figures 3–10) and Figure 14 to clarify the meaning of the yellow sections in the "augmented tabular data" block. Additionally, we have updated the captions of the figures to explicitly describe the meaning of the yellow section.*

[Minor things]

R2C5: While the paper primarily focuses on accuracy and F2 scores, incorporating recall as an evaluation metric could offer valuable insights.

Answer*: As suggested by the reviewer, we have incorporated recall as an additional evaluation metric. In Appendix B, we have added tables that present the average recall along with standard deviations for different vector sizes, for different ML models and different embedding methods, comparing baseline tabular performance with KG-augmented data.*

R2C6: In the abstract and introduction, you mention a focus on accuracy and F2 score without explaining why. The explanation is provided later in section 4.2. Include a brief sentence in the introduction explaining this focus or reference where the explanation can be found.

Answer*: We have addressed this comment by adding a reference in the introduction to where the reason for choosing the selected metrics is explained.*

R2C7: Rescale the y-axis in your figures for clarity. For example, Fig 12 should have a y-axis range of 0.6 to 0.8. This will better illustrate performance differences.

Answer*: We considered rescaling the y-axis but decided to maintain the current scale for consistency with our previous publications and because altering the y-axis range can sometimes exaggerate differences and potentially mislead readers about the actual variability in performance. To avoid unintended misinterpretations, as highlighted by [11, 15], we have chosen to keep the original y-axis scale.*

R2C8: Typos:

– Page 3 line 42: R capital letter (R subset of CxC)
– H1.2 Analysis [line 14 page 19] missing capital letter

***Answer**: We have fixed the typos.*

Review 3:

This paper presents an impressive and innovative contribution to the field of machine learning, addressing the critical challenge of improving model performance in data-scarce or sensitive scenarios. The authors' hypothesis that semantic enrichment through knowledge graph (KG) integration can enhance predictive power is both compelling and highly relevant. The introduction of novel neuro-symbolic approaches and the systematic exploration of KG embedding techniques highlight the authors' dedication to advancing the state of the art. Their rigorous evaluation across multiple ML algorithms and KG embedding methods showcases the robustness of their approach. The focus on real-world applications, such as heart disease and chronic disease prediction, further emphasizes the practical significance of their work. The results are particularly noteworthy, demonstrating remarkable improvements in F2 scores, such as a dramatic boost in XGBoost performance for heart disease prediction. These findings convincingly illustrate the potential of KG-based augmentation to transform ML performance, especially in binary classification tasks. The clear, data-driven methodology and the emphasis on accuracy and F2 scores provide valuable insights for both researchers and practitioners. This reviewer emphasises that this paper is a significant step forward in the integration of symbolic reasoning with ML techniques, paving the way for more context-aware, robust, and effective predictive models. It is a must-read for anyone interested in enhancing ML performance through innovative data augmentation strategies. Summarizing, the work presented is interesting, relevant, important, well presented and also well written and fits well into this journal. For all these reasons, this reviewer argues for acceptance of this work and provides in the following just one minor suggestion for improvement to further enhance its usefulness to the potential reader:

R3C1: Page 5, last paragraph- In addition to the excellent work of Bhatt et al. (2020), a very new paper should be mentioned here, a related work that is very interesting for the reader: Kraisnikovic, C. 2025. Fine-tuning language model embeddings to reveal domain knowledge: An explainable artificial intelligence perspective on medical decision making. Engineering Applications of Artificial Intelligence, 139, 109561, doi:10.1016/j.engappai.2024.109561.

***Answer**: Thank you for your suggestion. We have incorporated the referenced paper into the Machine Learning Models in Disease Prediction paragraph of the Related Work section, as it aligns well with our discussion on leveraging embeddings for medical decision-making.*