Dear Editor and Reviewers,

We would like to thank you for your valuable feedback, which has enabled us to further improve the original version of the manuscript. We have made changes to reflect the proposed suggestions. Below is a list of the received feedback, along with a description of how we have addressed it. In addition, the entire manuscript has been proofread again.

Sincerely,

Giulia Murtas, Veselka Boeva and Elena Tsiporkova

Neuro-LENS: a neuro-symbolic framework integrating incomplete background knowledge and deep learning- Changes applied to address reviewers' comments

Reviewer 1

1. The second scenario is missing the main details about the symbolic elements, which have been limited to abstract categories without providing their semantics. This makes difficult to understand the complexity of the scenario. I encourage authors to describe what these categories are representing in this domain.

   Unfortunately, as the dataset is anonymized, it is not possible to explicitly state what these categories represent. Nevertheless, a few examples have been added in the text to make the use case clearer to the reader.

2. The paper requires more revision to fix typos.
   The whole text was revised in order to correct any residual typos.

3. Figures need to be improved (especially Figures 6, 7, 8).
   All graphics that were difficult to read were improved

Reviewer 2

1. Suggested wording changes:

a. Prognostic health monitoring with vehicle failure prediction → Prognostic health monitoring in vehicle failure prediction
b. The dataset comprises of 161 images → The dataset comprises 161 images
c. The current section → This section
d. The amount of time they encompass → The timespan they cover
e. Subsequently (second block in Fig. 3) → Subsequently (the second block in Fig. 3)
f. A few examples of these attributes can be seen in the first column of 6 → of Table 6
g. The time windows corresponding to the class labels defined in 5 are also shown → in Table 5

All wording suggestions were implemented.

2. Missing citations:
    a. Provide citations for the PETS2006 and AVS2007 datasets in section 5.1.1
    b. Provide citation for LSTM neural network

The citations were added to the text.

3. Fig. 6 (Fig. 4 in the revised version) should be closer to where it is first referenced in the text.

The figure was placed right below the first time it is mentioned in the text.

4. Transform the numbered list in section 5.1.2 into plain text with subsections, as the items are too long for a numbered list.

The list was rewritten as plain text, where each item is a separated paragraph titled with the name of the corresponding block.

5. “Next, the decision rules exploited by the rule-classifier are defined.” It is worth starting a new stage here.

The text was split into two new paragraphs, corresponding to the two grey blocks in Fig. 4: “Rule-based classifier” and “Score-based classifier.”

6. Table 2: keep the same precision in all numbers.

The integers were modified to have the same precision as the other numbers in the table.

7. The dataset names in Fig. 7 are unnaturally bulky.

The font size of the plot titles decreased.

8. Table formatting (gridlines) should be made uniform throughout the entire paper.

All tables now have consistent gridline formatting.

9. “In the data, each vehicle is described by 8 specifications.” (Section 5.2.1) A term like “feature” or “attribute” would be more fitting here.

The term “specifications” was substituted by “technical features” all throughout the text. The term “specification” is only retained in “specification data”, from the name of the file as presented in the original paper of the Scania dataset.

10. Enlarge the plots in Fig. 8.

The plots were enlarged.

11. Briefly explain what “forward filling” means. (Section 5.2.3)

A brief explanation of what forward filling means in the context of missing data imputation was added to the text.

12. Correct the sentence “All evaluated models exhibit a very high accuracy (over 95 is highly imbalanced, a high accuracy is not very indicative.” In section 5.2.3, Symbolic-to-neural chaining.

The sentence was corrected, the mistake was due to a few words accidentally being commented in the text.

13. "…uses them to derive meaningful attributes (or instances)" in section 5.1.2, Image characterization paragraph. Stick to the same term throughout the paper.

The text was modified to always use the term "attributes."

14. "A "none of known" class is also created for ambiguous images which do not satisfy either decision rule" is a repetition, as it was already stated a few lines above.

The repetition was removed from the text.

15. The captions should be placed at the top of the table in Table 4 and 5.

The captions were moved on top of the two tables.

16. Add technical details of the used LSTM-autoencoder: how was the history of the trucks presented to the network? Tokenized? What was the dimensionality of these tokens? What was the working dimensionality of the LSTM network? How was the model trained?

Technical details of the LSTM-autoencoder were added in the "Parallel neural-symbolic integration" paragraph in section 5.2.3.

Technical details of the LSTM model were added in the "Symbolic-to-neural chaining" paragraph of section 5.2.3.

17. Clarify which labelling is used for the SCANIA dataset: binary labelling (healthy and failing vehicles) or labelling with 5 decision classes? If only one is used, the other one should be completely discarded from the text; if both are used, they should be clearly named and referred to using the chosen names.

Both kinds of labelling are used in the paper: the binary labelling is always used within the symbolic components, as it is based on static properties of the trucks which do not depends on the time to failure; the multi-class labelling is used in the

final results of the symbolic-to-neural strategy. The fact that both are provided with the dataset is now specified in the dataset presentation in the 5.2.1 subsection, and a few sentences have been added throughout the whole 5.2 section to clearly state which labelling is being considered.

18. In the baseline comparison, the rule-based baseline model was manually built by the authors. Why not use an existing decision rule/tree inducer (e.g. C5.0 or C4.5) to induce such a rule, for the sake of greater objectivity? Also, the used rule should be shown.

A C5.0 decision tree was also used as a baseline and its performance was added to the comparison in the baseline section. Moreover, all the rules used are now explicitly stated in the form of equations.