Reproducibility Checklist

This paper:

- Includes a conceptual outline and/or pseudocode description of AI methods introduced ([Yes]).
- Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results ([Yes]).
- Provides well marked pedagogical references for less-familiare readers to gain background necessary to replicate the paper ([Yes]).

Does this paper make theoretical contributions? ([No])

Does this paper rely on one or more datasets? ([Yes]).
If yes, please complete the list below.

- A motivation is given for why the experiments are conducted on the selected datasets ([Yes]).
- All novel datasets introduced in this paper are included in a data appendix ([Yes]).
- All novel datasets introduced in this paper will be made publicly available upon publication of the paper with a license that allows free usage for research purposes ([Yes]).
- All datasets drawn from the existing literature (potentially including authors' own previously published work) are accompanied by appropriate citations ([Yes]).
- All datasets drawn from the existing literature (potentially including authors' own previously published work) are publicly available ([Yes]).
- All datasets that are not publicly available are described in detail, with explanation why publicly available alternatives are not scientifically satisficing ([Yes]).

Does this paper include computational experiments? ([Yes]).
If yes, please complete the list below.

- Any code required for pre-processing data is included in the appendix ([Yes]).
- All source code required for conducting and analyzing the experiments is included in a code appendix ([Yes]).
- All source code required for conducting and analyzing the experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes ([Yes]).
- All source code implementing new methods have comments detailing the implementation, with references to the paper where each step comes from ([Yes]).
- If an algorithm depends on randomness, then the method used for setting seeds is described in a way sufficient to allow replication of results([Yes]).
- This paper specifies the computing infrastructure used for running experiments (hardware and software), including GPU/CPU models; amount of memory; operating system; names and versions of relevant software libraries and frameworks ([Yes]).
- This paper formally describes evaluation metrics used and explains the motivation for choosing these metrics ([Yes]).
- This paper states the number of algorithm runs used to compute each reported result ([Yes]).
- Analysis of experiments goes beyond single-dimensional summaries of performance (e.g., average; median) to include measures of variation, confidence, or other distributional information ([Yes]).
- The significance of any improvement or decrease in performance is judged using appropriate statistical tests (e.g., Wilcoxon signed-rank) ([Yes]).
- This paper lists all final (hyper-)parameters used for each model/algorithm in the paper's experiments ([Yes]).
- This paper states the number and range of values tried per (hyper-) parameter during development of the paper, along with the criterion used for selecting the final parameter setting ([Yes]).