MI-Checklist

Study design (Part 1)	Completed: page number	Notes if not completed
The clinical problem in which the model will be employed is clearly detailed in the paper.	Page 1	---
The research question is clearly stated.	Page 1	—
The characteristics of the cohorts (training and test sets) are detailed in the text.	—	No training/test split — unsupervised single-cell clustering on a public dataset.
The cohorts (training and test sets) are shown to be representative of real-world clinical settings.	—	Not applicable — dataset is experimental, not clinical.
The state-of-the-art solution used as a baseline for comparison has been identified and detailed.	Pages 1-2	----
Data and optimization (Parts 2, 3)	**Completed: page number**	**Notes if not completed**
The origin of the data is described and the original format is detailed in the paper.	Page 1	—
Transformations of the data before it is applied to the proposed model are described.	Page 1-2	—
The independence between training and test sets has been proven in the paper.	—	Not applicable — unsupervised clustering; no model training/testing split.
Details on the models that were evaluated and the code developed to select the best model are provided.	Pages 1–6 — models introduced (1–2), baseline and variant evaluations (2–6), best model justified (6); code provided (page 2).	---
Is the input data type structured or unstructured?	Structured	---
Model performance (Part 4)	**Completed: page number**	**Notes if not completed**
The primary metric selected to evaluate algorithm performance (e.g., AUC, F-score, etc.), including the justification for selection, has been clearly stated.	Pages 2 & 6	---
The primary metric selected to evaluate the clinical utility of the model (e.g., PPV, NNT, etc.), including the justification for selection, has been clearly stated.	—	Not a clinical model — utility measured through ROC marker recovery, not predictive value.

The performance comparison between baseline and proposed model is presented with the appropriate statistical significance.	Pages 2–6; baseline (Table 1) vs variants (Tables 3A–4B) compared quantitatively using Silhouette, ARI, and NMI.	---
Model examination (Part 5)	**Completed: page number**	**Notes if not completed**
Examination technique 1a	---	Not required by rubric — structured single-cell data; interpretability performed via marker-gene enrichment and ROC gene-set scoring instead of coefficient or sensitivity analysis
Examination technique 2a	---	Not required by rubric — unsupervised structured data; biological interpretability achieved through ROC gene-set scoring and visualization (Figures 2A–2B) rather than sensitivity or saliency analysis
A discussion of the relevance of the examination results with respect to model/algorithm performance is presented.	Pages 2–6	Discussion not required by rubric — addressed implicitly through performance interpretation on Pages 2–6 (e.g., Silhouette/ARI/NMI changes and ROC marker enrichment across methods)
A discussion of the feasibility and significance of model interpretability at the case level if examination methods are uninterpretable is presented.	---	Not required by rubric — model is unsupervised and interpretable through marker-level analysis; case-level interpretability (individual prediction explanations) not applicable.
A discussion of the reliability and robustness of the model as the underlying data distribution shifts is included.	Page 6	

Reproducibility (Part 6): choose appropriate tier of transparency	**Notes**
☑ **Tier 1:** complete sharing of the code	Full public GitHub repository and Google Colab notebook with executable code provided.
☐ Tier 2: allow a third party to evaluate the code for accuracy/fairness; share the results of this evaluation	—
☐ Tier 3: release of a virtual machine (binary) for running the code on new data without sharing its details	—
☐ Tier 4: no sharing	—