

Dear Author,

Thank you for the opportunity to review your manuscript titled " Application of Graph Neural Networks for Modeling Stem Cell Donor-Recipient Compatibility in the Detection and Classification of Leukemia." Your work addresses a highly relevant topic. Below is my detailed feedback:

The study has a promising topic but suffers from critical flaws: unsubstantiated claims, lack of reproducibility, poor writing/informal language

1. The methodology is fundamentally unsound. The use of the 1000 Genomes Project dataset is inappropriate for leukaemia detection and classification, as it comprises genomic data from healthy individuals across populations, lacking any labels for disease status, transplant outcomes, or compatibility metrics.
2. The claimed contributions - improved accuracy over traditional ML models and interpretability - are not sufficiently substantiated, as the methodology relies on inappropriate data (*1000 Genomes Project, which consists of healthy individuals without leukaemia or transplant outcomes*).
3. The authors simulate donor-recipient pairs by imputing missing values and normalizing features but fail to explain how ground truth labels for "compatibility", "relapse", or "leukaemia subtypes" are derived, rendering the high reported accuracies (97.68% - 99.74% for detection; 98.76% - 99.4% for classification) implausible and likely artefactual.
4. Graph construction (nodes as individuals, edges as compatibility scores) and models (GAT for attention, DGNN for dynamics) are described at a high level with equations, but lack implementation details (no code repository, hyperparameter tuning specifics beyond basics like Adam optimizer and dropout). Experiments on "patients" (Tables 4, 8) contradict the dataset description, suggesting fabrication or confusion.
5. Results are not reproducible, the data processing (PCA retaining 95% variance, stratified splitting) and evaluations (AUC-ROC, F1-score) are not supported by verifiable evidence. Confusion matrices and accuracy claims on larger "datasets" (190/200 true positives) appear arbitrary without raw data or cross-validation details. Fluorescence microscopy images (Figures 7 - 9) seem unrelated to the GNN model and do not validate computational predictions. Interpretations overstate clinical relevance without biological validation.
6. Abstract/Introduction (Lines 1–144): Overstates novelty; clarify how 1000 Genomes data enables leukaemia classification. Remove repetitive "advanced research" phrasing.
7. Methods (Lines 223 - 410): Provide code/pseudocode for graph construction. Explain label generation for supervised training. Fluorescence methods (implied in results) belong here if relevant.
8. Results (Lines 428 - 536): High accuracies need benchmarking code/data. Figures 7 - 9 seem irrelevant, remove or justify. Tables 8 - 13 mix simulated "patient" data with 1000 Genomes; resolve inconsistency.

9. Discussion/Conclusions (Lines 550 - 670): Acknowledge data limitations (no real HSCT outcomes). Suggest real datasets (e.g. from bone marrow registries) for future work
10. Specify whether any external clinical datasets (e.g., transplant registries) were used alongside 1000 Genomes. If not, clarify how relapse/classification labels were generated.
11. For figures 7 - 9, provide details on the origin of the microscopy images. If experimental, describe protocols; if illustrative, clearly state so.
12. Results lack statistical tests to support claims of superiority over SVMs, CNNs, and RNNs
13. The 1000 Genomes Project provides SNP and partial HLA data but does not include actual transplant outcomes. The manuscript mentions relapse and transplant results, but it is unclear how these were inferred or simulated. This raises reproducibility concerns
14. The integration of confocal microscopy images as "biological validation" is underexplained. It is unclear whether these are original experiments or illustrative references.
15. Expand on limitations of using synthetic/augmented datasets and discuss transferability to real-world patient cohorts.
16. Tables lack references
17. Tables 8 -13 are very detailed, but some redundancy could be reduced to improve readability.
18. Language is informal, major revision is needed to use academically sound terms
19. Review of related work needs to highlight the novelty of the authors' specific contribution compared to existing transplant-focused ML models