Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 08 May 2026
Revised 15 June 2026
Accepted 02 July 2026

Natural Resources for Human Health 2026, Vol. 6 (4s): 415–438

KEYWORDS:
Decision Making; Hematopoietic stem cell transplantation; Graft-versus-host disease; Hybrid deep neuro-symbolic learning; Rule Compliance; Donor Selection

Neuro Symbolic Intelligence for Donor Matching and Outcome Forecasting in Bone Marrow Transplantation

Kaliyamoorthi P[1], Dr. R. Bhuvaneswari[2]

[1]Research Scholar, Department of Computer Science and Engineering,SRM Institute of Science and Technology, Ramapuram 600089
[2]Assistant Professor(SI.G), Faculty of Engineering and Technology, Department of Computer Science and Engineering(CS & GT), SRM Institute of Science and Technology, Ramapuram 600089

ABSTRACT: Improving survival rates and minimizing complications like graft-versus-host disease (GVHD) is the main goal of donor-outcome prediction in hematopoietic stem cell transplantation (HSCT). To predict outcomes following a transplant, state-of-the-art models combine information from clinical, genetic, and immunological sources. There has been a recent uptick in the use of ML and DL methods to tailor donor selection and improve prognostic accuracy. In HSCT operations, accurate prediction improves long-term patient prognosis and facilitates evidence-based decision-making. Here, to offer a paradigm for HSCT donor outcome prediction based on hybrid deep neuro-symbolic learning (HDNSL). To improve interpretability, resilience, and rule compliance, HDNSL incorporates symbolic domain information into the learning process of the model, which is different from other deep learning models. To keep deep neural architecture performance intact, the framework uses symbolic rules to direct feature prioritization and affect attention distributions. By surpassing neuro-symbolic and solely deep learning approaches, HDNSL obtains the best donor prediction accuracy of 0.93 and outcome prediction accuracy of 0.91, according to a comparative analysis using baseline models such as Support Vector Machine (SVM), Random Forest (RF), XGBoost, Convolutional Neural Network (CNN), Gated Recurrent Unit (GRU), and Long-Short Term Memory (LSTM). Trends over epochs show that accuracy (up to 0.945) and area under the curve (up to 0.965) steadily improve by epoch 50. With a continuous emphasis on important clinical factors like HLA Match, GVHD Risk, and Time to Transplant, attention heatmaps show that the symbolic rules greatly affect feature relevance across samples. Features prioritized by symbolic rules are in line with high-impact predictors, as confirmed by SHAP and feature importance studies. In addition, statistics on rule satisfaction reveal that HDNSL attains 95% compliance, which is higher than 84% when using symbolic rules alone and 62% when using pure deep learning. The findings of this study validate HDNSL as a robust, understandable, and rule-aware platform for clinical transplantation predictive modeling. This opens the door to its potential use in practical applications and clinical decision support systems.

1. INTRODUCTION

When it comes to healthcare besides customized treatment, predictive modeling drastically changed by the ever-changing landscape of AI besides Machine Learning (ML) [1]. There is an absolute need for reliable and interpretable models in crucial domains like donor-recipient matching besides transplantation success prediction [2]. The "black-box" aspect of deep learning (DL) approaches is a major drawback, particularly in high-stakes clinical situations, despite impressive improvements made

methods [3]. Hybrid Deep Neuro-Symbolic Learning (HDNSL) is a new paradigm that aims to solve this problem by incorporating symbolic reasoning into deep learning frameworks. Improving model performance besides trustworthiness are the goals of this technique, which aims to merge data-driven learning with symbolic principles humans can understand [4-5].

Donor compatibility, age, cytomegalovirus (CMV) status, gender, blood type, disease stage, and comorbidities are some of heterogeneous features that rely on accurate modeling in donor matching and post-transplantation outcome prediction [6]. For a long time, this type of data was best analyzed using classic machine learning besides traditional statistical models like SVMs, RF, and Gradient Boosting Machines (e.g., XGBoost) [7]. The nonlinearities, temporal dynamics, and complex connections included in medical information, however, frequently cause these models to fail [8].

When it comes to learning representations of features in high dimensions, deep learning models like CNNs, GRUs, and LSTM networks have proven to be rather effective [9]. On the other hand, they are usually not domain knowledgeable and don't have the ability to reason explicitly. Because of this discrepancy, their clinical applicability and interpretability are compromised [10]. By integrating symbolic rules into DL architectures, to may embed neural predictions in logic that humans can comprehend. These rules include information acquired from experts regarding compatibility restrictions, biological processes, and medical norms [11, 12].

Hybrid neuro-symbolic learning frameworks have recently emerged, which allow AI systems to DL's strong representational capabilities while also incorporating rule-based constraints to direct their development [13]. HDNSL integrates symbolic reasoning with data-driven learning by incorporating logical rules into the classical architecture using architectural priors, rule-aware attention layers, feature symbolic constraint loss functions [14–15]. Enhanced explainability, robust generalization, and more meaningful decision-making are crucial in fields where algorithmic judgments affect human lives.

Such neuro-symbolic integrations improve consistency with domain knowledge and boost presentation, according to recent research [16]. To avoid overfitting due to data abnormalities, symbolic criteria pertaining to HLA compatibility or CMV mismatch, for example, might regularize the model in transplant matching scenarios [17]. These models also provide more transparency and make it easier for AI systems and doctors to work together [18] by incorporating symbolic attention processes, which recovers feature prioritizing besides ensures that medically critical aspects are emphasized during inference [19].

Below is a summary of our contributions to this paper:

1. To present HDNSL, an innovative framework for donor-recipient matching and outcome prediction that combines deep neuro-symbolic learning with a unique approach.

2. To create a Neuro-Symbolic Integration Unit (NSIU) to direct processes for attention and feature relevance by introducing domain rules into learning process.

3. Third, to use rule satisfaction analysis, interpretability visualizations (such as SHAP and attention heatmaps), and classification metrics to provide a thorough evaluation.

4. Across several experimental settings, to show that our model is more generalizable, interpretable, and clinically relevant.

To sum up, a major step toward creating AI systems that are open, responsible, and effective in healthcare settings has been the acceptance of hybrid neuro-symbolic learning. Improved predicted outcomes, more clinician trust, and better decision support are all consequences of HDNSL's integration of symbolic reasoning and deep learning, which leads to safer and more successful patient care.

Here is how the rest of the paper is structured. Section 2 offers a comprehensive literature overview on the topic, touching on topics such as medical AI interpretability approaches, neuro-symbolic models, and deep learning. To present the HDNSL technique in Section 3, which includes the NSIU module and how it is integrated with the base neural architecture. Chapter 4 delves into the experimental design, datasets, performance measures, and implementation specifics. The conclusion and possible future possibilities for extending the HDNSL framework are outlined in Section 5.

2. RELATED WORKS

In their investigation into the use of deep learning and machine learning for HLA-haploidentical transplantation, Kaliyamoorthi and Bhuvaneswari [20] look at ways to improve donor selection and forecast donor availability. Algorithms like support vector

machines (SVM), gradient boosted machines (GBM), and deep neural networks (DNN) are used to model important demographic, behavioral, and clinical characteristics, such as age, registry duration, health history, donor engagement, and survey responses, from structured data extracted from national and international donor registries. Using additional variables such as donor telomere length and cytogenetic markers, the study assesses transplant outcomes, response probabilities to confirmatory typing (CT) requests, and post-transplant problems such graft-versus-host disease (GVHD). The interpretation of SHAP values helps to comprehend the role of features in model predictions. In order to evaluate the clinical effectiveness, to also present comparative analyses of haploidentical and HLA-identical sibling transplants. In order to improve patient survival and transplantation success, the study shows that data-driven approaches can greatly improve donor selection strategies, prediction accuracy, and influence real-time clinical decision-making.

From day 0 to day 30 post-transplant, Spohr et al. [21] continually predicted mortality risk using a dynamic risk classification method based on a random forest technique. By analyzing data from 847 alloHSCT patients from 2004 to 2019, our model successfully classified patients into high-, moderate-, and low-risk categories, and its AUROC scores remained above 0.8 from day 17. Analyzing the relative importance of features over time showed a change: earlier predictions relied on a wider variety of factors, whereas later stages relied more on bloodwork measures associated to inflammation. This methodology is more effective than static methods because it can adjust to different treatment phases and has better prediction capabilities. To make the software and the bloodwork time-series dataset available for free to encourage reproducibility and wider use.

Using sAML patients in remission who underwent HSCT from HLA-matched donors between 2010 and 2022, Nagler, A., et al., [22] investigated if the ELN 2022 cytogenetic risk score predicts outcomes. To removed 6 patients with favorable risk cytogenetics, leaving 1119 patients with 284 patients with bad cytogenetics and 829 with intermediate risk. Compared to 99.5%, the elaboration rates were 72.4%. There was no difference in the frequency of acute graft-versus-host disease (GVHD), but the risk groups for intermediate vs. unfavorable cytogenetics were greater for 2-years all grades GVHD (HR=0.72; p=0.034) and substantial chronic GVHD (HR=0.58; p=0.027), respectively. Also comparable was the two-year non-relapse mortality (NRM). Compared to individuals at intermediate risk, those at adverse risk had worse HSCT outcomes in every other category: A 2-year relapse incidence (RI) HR of 2.48 (95% CI 1.95-3.15, $p < 0.001$) was recorded. A 2-year leukemia-free survival rate (LFS) of 1.62 (95% CI 1.34-1.95, $p < 0.001$), an overall survival rate (OS) of 1.59 (95% CI 1.3-1.93, $p < 0.001$), and a GVHD-free/relapse-free survival rate (GRFS) of 1.38 (95% CI 1.15-1.65, $p < 0.001$) were the respective HRs. To conclude that sAML patients' cytogenetic risk scores predict the results of HSCT.

The goal of the work by Birkmire, J., et al., [23] is to create pediatric heart transplantation (PHTx) risk indices that can be used to predict graft failure, death within 90 days and one year after the transplant, according to pre-transplant donor, recipient, combination, and environmental variables alone. Through a retrospective examination of the UNOS Organ Procurement and Transplantation Network (OPTN) database for 6,458 PHTx from 2009 to 2024, the C-HART (Children's Heart Allograft Risk & Transplantation) Index was established. There was a 3:1 split between the derivation and external validation groups for the patients. In order to find important factors to include in the derivation cohort's subsequent multivariate logistic regression and AIC backward selection, to used Stata's LASSO function. Using the β-coefficients of the predictor variables, the C-HART Index was built and subsequently tested on the validation cohort by AUC analysis. The final C-HART indices ranged from 33 to 59 variables per model out of a total of 297 variables (including subcategories for continuous variables). The last four C-HART indices were as follows: environmental variables(8-12%), 6-14% combined, 17-22% donors, and 61-76% recipients. The four models all identified recipient risk factors such as pre-transplant ECMO, a weight less than 3 kg, and a history of surgery for congenital heart disease. Recipient eGFR >60 and PVR 0-3 were protective factors in all four models. When looking at graft failure and 90-day mortality in the external validation cohort, the c-statistics were 0.73 and 0.72, respectively.

To tested various machine learning (ML) models to predict overall survival after transplant using data from 5183 patients with MF who underwent first allo-HCT at facilities affiliated with the European Society for Blood and Marrow Transplantation between 2005 and 2020 (Hernandez-Boluda, J. C., et al., [24]). For model validation, the cohort was split into two parts: a training set comprising 75% of the total and a test set comprising 25%. A random survival forests (RSF) model was constructed using ten variables: graft-versus-host disease prophylaxis, donor type, conditioning intensity, patient age, blood blasts, hemoglobin, leukocytes, platelets, performance status, and comorbidity index. A 4-level Cox regression-based score, other ML-based models created from the same dataset, and the Center for International Blood and Marrow Transplant Research score were used to compare its performance. With superior concordance indices in both the main and post essential

thrombocythemia/polycythemia vera MF subsets, the RSF surpassed all competitors. Both sets of metrics—the Akaike information criterion and the time-dependent receiver operating characteristic area under the curve—confirmed the RSF model's resilience and generalizability. While all models were able to predict the likelihood of death from causes other than recurrence, the RSF distinguished between curves more clearly, allowing for the identification of a 25% patient subset at high risk. Personalized treatment is made possible by ML's improvement of risk classification in allo-HCT patients with MF.

The effect of MHC class II matching on the survival of heart allografts in Mauritian cynomolgus macaques undergoing either immediate or delayed bone marrow and organ transplantation is investigated by Winter, C., et al., [25]. According to prior research, MHC matching is crucial for transplant survival. To examine the association between MHC class II inequalities and survival rates in this study. From one donor's bone marrow, 137 non-human primates (NHPs) received heart, kidney, or thymus transplants over the course of a decade. The administration of bone marrow might have occurred at the same time or at a later time. Using seven different haplotypes (M1-M7) for Mauritian cynomolgus macaques, recipients were grouped according to the level of MHC class II mismatch. The inclusion of non-human primates (NHPs) whose bone marrow was delayed was conditional on their ability to survive the delay. In order to compare the chances of survival up to one year after a bone marrow transplant (pBMTx) in relation to MHC class II mismatch, a Kaplan-Meier curve was generated. To used log-rank analysis to determine the p-values. Statistical study using the Kaplan-Meier survival curve method did not show a relationship between MHC class II mismatch and the probability of survival to day 365 post-BMTx. For every comparison between the three MHC class II mismatch levels, log-rank analysis yielded p-values higher than 0.05. Based on these results, it seems that MHC class II mismatching is not a good indicator of a non-human primate's one-year survival rate after a bone marrow or heart allograft transplant. Class II mismatching did not affect post-transplant survival, contrary to previous research that has stressed the significance of donor-recipient MHC matching. To further understand the overall impact of MHC class-specific differences on transplant results, future studies will investigate the consequences of MHC class I mismatching.

To forecast death in SAA patients and uncover important predictive variables for tailored treatment, Elsabbagh, E., et al., [26] construct and assess machine learning models utilizing pre-transplant characteristics. To looked at data from CIBMTR on 545 SAA patients who had HCT between 1988 and 2018. Donor, recipient, illness, transplant, demographic, and laboratory data were all features. With a female-to-male ratio of 1:1.4 and a death rate of 17.5%, the median patient age was 21 years (range: 0-74). The median imputation method was used to fill in the missing values (6%). The data was divided into two sets: one for training and the other for testing. After using 5-fold cross-validation to fine-tune RandomForest, Gradient Boosting, and XGBoost models, Synthetic Minority Over-sampling Technique (SMOTE) dealt with class imbalance. The models were evaluated with the help of AUROC and AUPRC. Shapley Additive Explanations (SHAP) values were used for feature importance analysis. Performance discrepancies were brought to light through comparisons among models. Thanks to its 0.84 AUROC and 0.71 AUPRC, the Gradient Boosting model performed better than the competition. With an F1 score of 0.92, it demonstrated a well-balanced precision-recall trade-off, achieving a high recall for survivors of 0.95. There were substantial interactions between demographic and clinical characteristics, and key predictors included sex, donor and recipient age, blood counts at diagnosis, graft type, and interactions between the two. For patients undergoing HCT, the Gradient Boosting model successfully predicts mortality risk. Blood counts, donor/recipient age, and graft type are the factors that SHAP analysis emphasizes as having an impact. Improved risk assessment and the ability to tailor treatments to individual patients are outcomes of interactions between demographic and clinical factors. The next step is to test these algorithms with more samples and look at genetic markers to improve their precision.

3. PROPOSED FRAMEWORK

In this section, the architecture of the proposed model, mathematical expression and the novelty of the research work is described in detailed.

3.1 Overview of the HDNSL Framework

This section delineates the architecture and mathematical foundation of the proposed Hybrid Deep Neuro-Symbolic Learning (HDNSL) framework developed to address the dual challenges of haploidentical donor selection and post-transplant outcome prediction in Bone Marrow Transplantation (BMT). HDNSL fuses deep learning's pattern recognition power with domain-specific symbolic reasoning, offering a high-fidelity, interpretable, and clinically aligned decision support system. Unlike conventional machine learning pipelines that rely solely on data-driven learning, HDNSL integrates symbolic clinical rules—

formulated through medical guidelines—into the attention and gating mechanisms of a neural network. This dual-path mechanism enhances both predictive accuracy and explainability. Figure 1 provides the workflow of the research model.

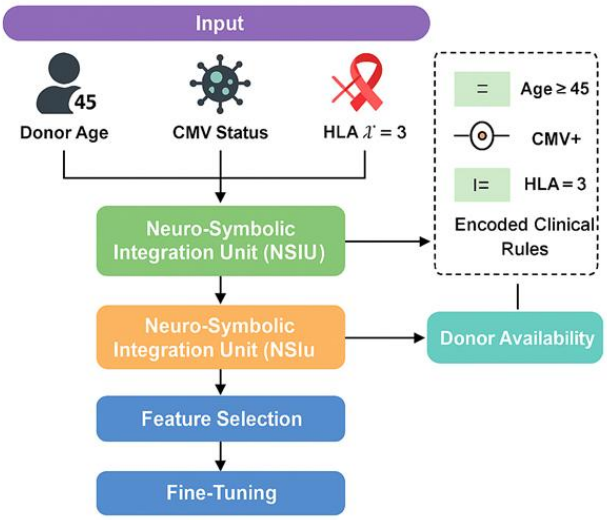

Figure 1. Proposed Architecture diagram

The HDNSL framework comprises the following key modules:

1. Preprocessing and Feature Engineering Unit
2. MAGRU-based Deep Learning Backbone
3. Giant Armadillo Optimization (GAO) for Feature Selection
4. Perpetual Pigeon Galvanized Optimization (PPGO) for Hyperparameter Fine-tuning
5. Neuro-Symbolic Integration Unit (NSIU)
6. Dual Output Heads: Donor Prediction and Post-BMT Risk Forecasting

Each module is embedded within a sequential pipeline optimized for interpretability and accuracy. The end-to-end training ensures the seamless fusion of data-driven representations and symbolic rules.

3.2 Preprocessing and Feature Engineering

Let be the original feature space, where each represents a patient-donor feature vector comprising demographic, clinical, and histocompatibility information. The dataset includes two labels:

- y_i^1 : Donor suitability (binary: 1 for suitable, 0 for unsuitable)
- y_i^2: Post-BMT outcome risk score (continuous or categorical)

To perform the following transformations:

- **Missing Value Imputation:** For missing entries $x_{ij} = NaN$, to impute using domain-aware statistics: mean for continuous f_c, mode for categorical f_k as mentioned in Eq. (1):

$$x'_{ij} = \{\mu_j, \quad if\ f_c(x_j)\ mode_j,\ \ if\ f_k(x_j) \quad (1)$$

One-Hot Encoding: For symbolic variables like CMV serostatus $\in \{+/+, -/-, +/-, -/+\}$, to encode them as 4-dimensional binary vectors.

Normalization: Continuous features $f_j \in R$ are normalized and it is mentioned in Eq. (2):

$$f_j' = \frac{f_j - \mu_j}{\sigma_j} \quad (2)$$

Augmented Feature Derivation: Features such as HLA mismatch ratio in Eq. (3)

$$R_{HLA} = \frac{mismatches}{10} \quad (3)$$

donor-recipient age gap in Eq. (4)

$$\Delta_{age} = |A_d - A_r| \quad (4)$$

and donor BMI score are derived. Let the final transformed input be in Eq. (5)

$$X' = \{x'_1, x'_2, \dots, x'_n\}, x'_i \in R^{d'} \quad (5)$$

3.3 Deep Learning Backbone: MAGRU Network

The MAGRU (Multi-head Attention-Gated Recurrent Unit) forms the core of the deep pipeline, allowing the model to learn temporal dependencies and highlight relevant features. Let denote the hidden state at time for a patient-donor pair.

GRU Cell Equations in Eq. (6):

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (6)$$

Updated gate in Eq. (7)

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (7)$$

Reset gate in Eq. (8 and 9)

$$\tilde{h}_t = tanh(W_h x_t + U_h(r_t \odot h_{t-1} + b_h)) \quad (8)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (9)$$

Multi-Head Attention: Enhancing the GRU, to use multi-head self-attention to focus on clinically relevant features in Eq. (10):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) VQ = W_q x, K = W_k x, V = W_v x \quad (10)$$

The final encoded representation is given in Eq. (11):

$$H = Concat(head_1, \dots, head_h) W^O \quad (11)$$

where W^O projects the attention outputs.

MAGRU thus yields an encoded patient-donor tensor $E_i \in R^m$, reflecting both temporal and relational patterns.

3.4 Giant Armadillo Optimization (GAO) for Feature Selection

The Giant Armadillo Optimization (GAO) [27] algorithm is a novel bio-inspired metaheuristic developed by mimicking the adaptive burrowing and defense mechanisms of the giant armadillo in natural ecosystems. This optimization framework is particularly suited for high-dimensional feature selection in the medical domain, where selecting relevant and non-redundant clinical, demographic, and genetic features significantly affects prediction performance and interpretability.

In HDNSL, GAO is used to select a subset of features from the transformed input vector $x'_i = R^{d'}$, reducing dimensionality while preserving relevant information for both donor suitability prediction and post-BMT risk assessment.

Let:

- $P = \{S_1, S_2, \dots, S_k\}$be a population of candidate feature masks.
- Each $S_j \in R^{d'}$ is a binary vector representing a subset of selected features.
- $S_j^i = 1$ if the feature is selected, otherwise 0.

Fitness Evaluation: The fitness of each candidate solution S_j is evaluated using a multi-objective fitness function in Eq. (12):

$$F(S_j) = \lambda_1 . Accuracy_{donor} + \lambda_2 . Accuracy_{outcome} - \lambda_3 . \frac{\|S_j\|_1}{d'} \quad (12)$$

$Accuracy_{donor}$: classification accuracy on donor prediction. $Accuracy_{outcome}$: classification/regression accuracy on outcome prediction. $\frac{\|S_j\|_1}{d'}$: normalized number of selected features. $\lambda_1, \lambda_2, \lambda_3$: control weights for trade-off between predictive power and sparsity.

Exploration (Foraging Phase): Armadillos explore wide regions by random perturbation in Eq. (13):

$$S_j^{t+1} = Sigmoid(S_j^t + a.N) > \tau \quad (13)$$

a: exploration rate. $N(0, I)$: Gaussian noise. And τ: binarization threshold (e.g., 0.5).

Exploitation (Burrowing Phase): If no improvement is found, a burrowing operation simulates localized intensification by flipping bits in high-gradient directions in Eq. (14):

$$S_j^{t+1} = S_j^t \oplus M_{grad}, \ where \ M_{grad}[i] =$$
$$\{1, \ if \ feature \ i \ contributes \ positively \ 0, \qquad otherwise \quad (14)$$

This hybrid search behavior enables robust convergence toward minimal and predictive feature subsets.

Selected optimal mask S^* is then applied to all input vectors x_i', producing and it is given in Eq. (15)

$$x_i'' = S^* \odot x_i' \quad (15)$$

3.5 Perpetual Pigeon Galvanized Optimization (PPGO) for Hyperparameter Tuning

The Perpetual Pigeon Galvanized Optimization (PPGO) algorithm [28] is a new swarm-based strategy that simulates the homing and navigation behavior of pigeons. Inspired by pigeons' ability to return to their nest from unfamiliar locations using Earth's magnetic fields and celestial cues, PPGO is designed to fine-tune the hyperparameters of the HDNSL model, including learning rates, attention head sizes, dropout rates, and gating thresholds.

Let:

- $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$be a population of pigeon agents.
- Each θ_i represents a unique set of model hyperparameters.

Fitness Function: PPGO evaluates each agent θ_i using the objective in Eq. (16):

$$F(\theta_i) = Macro - F1(\theta_i) + AUC(\theta_i) - \lambda . T_{train}(\theta_i) \quad (16)$$

- $Macro - F1$: average F1 score across classes.
- AUC: area under the ROC curve.
- T_{train}: training time.
- λ: penalty for computational cost.

Position Update Rule: The pigeons' flight path is updated based on collective navigation toward the group's best solution (nest) in Eq. (17):

$$\theta_i^{t+1} = \theta_i^t + \gamma 1(\mu^t - \theta_i^t) + \gamma 2 . N(0, \sigma^2) \quad (17)$$

- μ^t: centroid of top q% performing pigeons.
- $\gamma 1$: homing strength (exploitation).
- $\gamma 2$: randomization intensity (exploration).

Galvanic Adjustment Mechanism: Inspired by pigeons' geomagnetic sensing, PPGO uses a galvanic adjustment to perturb directions selectively based on performance feedback in Eq. (18):

$$\theta_i^{t+1} = \theta_i^t + \beta . sign(\nabla F(\theta_i)) \quad (18)$$

- β: galvanic learning rate.
- $\nabla F(\theta_i)$: gradient direction of the fitness landscape.

The process continues iteratively until convergence or early stopping criteria based on plateaued performance.

The best hyperparameter set from PPGO configures the MAGRU network and NSIU gating thresholds, leading to efficient learning and enhanced generalization across donor-patient pairs.

3.6 Symbolic Rule-Based Clinical Support

The symbolic rules were constructed only from variables available in the P-5684 dataset. The available donor-related variables used for rule construction were donor age (dnrage) and donor–recipient CMV serostatus (drcmvpr). The dataset does not contain a numerical HLA mismatch-count variable. Therefore, no HLA match, HLA mismatch, or HLA mismatch-threshold rule was constructed or evaluated in this study.

The group variable, which distinguishes haploidentical and 8/8 matched unrelated donor (MUD) transplant records, was retained only as the classification target. It was not included as an input feature or symbolic rule because this would cause outcome leakage. The rules provide transparent clinical-profile indicators to support interpretation of the model output; they do not independently recommend a donor.

3.6.1 Symbolic Rule Construction and Clinical Rationale

Two dataset-supported rules were defined. The first rule identifies records with donor age below the predefined experimental threshold of 40 years. Donor age is an important consideration in donor assessment, although the threshold used here is an analytical rule and not an automatic clinical decision criterion [29], [30].

$$R_1(x_i) = \mathbb{I}\left(A_{d,i} < 40\right) \quad (19)$$

where $A_{d,i}$ denotes the donor age for the i^{th} donor–recipient record and $\mathbb{I}(\cdot)$ is the indicator function. The rule returns 1 when the donor age is below 40 years and 0 otherwise.

The second rule represents donor–recipient CMV serostatus concordance. In P-5684, drcmvpr is coded as 0 for CMV-positive donor/CMV-positive recipient (+/+), 1 for CMV-positive donor/CMV-negative recipient (+/−), 2 for CMV-negative donor/CMV-positive recipient (−/+), and 3 for CMV-negative donor/CMV-negative recipient (−/−). Concordant CMV status is represented as follows:

$$R_2(x_i) = \mathbb{I}\left(C_{\text{CMV},i} \in \{0,3\}\right) \quad (20)$$

where $C_{\text{CMV},i}$ is the recorded donor–recipient CMV serostatus code. The rule returns 1 for concordant CMV pairs (+/+ or −/−) and 0 for discordant pairs (+/− or −/+). CMV status is included because donor–recipient CMV serostatus is clinically relevant in allogeneic transplantation management [31].

A combined transparent profile indicator is defined as:

$$R_3(x_i) = R_1(x_i) \wedge R_2(x_i) \quad (21)$$

The value of $R_3(x_i) = 1$ indicates that both analytical conditions are satisfied: donor age below 40 years and CMV serostatus concordance.

Table 1. Dataset-supported symbolic rules used in this study

Rule	Dataset variable	Symbolic condition	Output interpretation
R_1	Donor age (dnrage)	$A_{d,i} < 40$	Donor age is below the predefined analytical threshold
R_2	Donor–recipient CMV status (drcmvpr)	$C_{\text{CMV},i} \in \{0,3\}$	Donor–recipient CMV serostatus is concordant
R_3	Donor age and CMV status	$R_1(x_i) \wedge R_2(x_i)$	Both transparent profile conditions are satisfied

3.6.2 Rule Verification and Validation

Rule verification was performed by checking that every symbolic-rule variable, code, and condition corresponded directly to the P-5684 data dictionary. Donor age was obtained from dnrage, and CMV concordance was calculated from the four recorded drcmvpr categories. Records with missing donor age or CMV status were handled using the preprocessing procedure described in Section 3.2 before rule calculation.

The rules were also checked for information leakage. The donor-group variable (group) was used only as the classification label and was excluded from rule construction and model inputs. Similarly, post-transplant outcomes and follow-up variables were excluded because they are not available at the time of donor-group assessment.

Clinical guidance was used only to justify the relevance of donor age and CMV serostatus as interpretable factors [29]–[31]. However, the symbolic rules were not externally validated as clinical donor-selection criteria. Their role in this study is limited to transparent interpretation of the dataset-based classification framework. Prospective validation with independent transplant cohorts and clinician review is required before any use in routine clinical decision support.

3.7 Dual Output Heads and Multi-Objective Loss

The HDNSL framework supports dual prediction objectives:

Donor-group classification (binary classification)

Post-Transplant Risk Stratification (binary/multiclass classification or regression)

This dual-task configuration is implemented using two independent output heads stemming from the final NSIU-modified embedding $\tilde{E}_i$:

Donor Head (Classification) in Eq. (22):

$$\hat{y}_i^{(1)} = softmax(W_1\tilde{E}_i + b_1) \quad (22)$$

Outcome Head (Risk Score) in Eq. (23):

$$\hat{y}_i^{(2)} = \{softmax\big(W_2\tilde{E}_i + b_2\big),\ for\ multiclass\ \sigma\big(W_2\tilde{E}_i + b_2\big) \quad for\ binary\ risk\ W_2\tilde{E}_i + b_2 \quad for\ continuous\ score \quad (23)$$

The total loss L is a weighted combination of individual task losses in Eq. (24, 25 and 26):

$$L = a.L_{donor} + (1-a).L_{risk} \quad (24)$$

Where:

$$L_{donor} = CrossEntropy(\hat{y}_i^{(1)}, y_i^{(1)}) \quad (25)$$

$$L_{risk} = \{CrossEntropy\big(\hat{y}_i^{(2)}, y_i^{(2)}\big),\ if\ classification\ FocalLoss\big(\hat{y}_i^{(2)}, y_i^{(2)}\big),\ if\ imbalanced\ \ MSE\big(\hat{y}_i^{(2)}, y_i^{(2)}\big),\ if\ regression \quad (26)$$

The parameter $a \in [0,1]$ controls task importance. This dual-head strategy ensures the model learns complementary representations and encourages cross-task regularization.

3.8 Training Protocol and Convergence Analysis

Training the HDNSL framework is conducted in an end-to-end fashion with integrated optimization layers (GAO and PPGO) controlling feature selection and hyperparameter search respectively.

Training Dataset: $D = \{x_i, y_i^{(1)}, y_i^{(2)}\}_{i=1}^{N}$ where $x_i \in R^d$ is the raw feature vector, $y_i^{(1)}$ is donor label, and $y_i^{(2)}$ is the outcome label.

Optimization Algorithm: Adam optimizer with AMSGrad variant in Eq. (27):

$$\theta_{t+1} = \theta_t - \eta . \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (27)$$

Where $\hat{m}_t, \hat{v}_t$ are bias-corrected estimates of first and second moments.

Learning Schedule:

- Warm-up phase for 5 epochs.
- Cosine annealing learning rate decay in Eq. (28):

$$\eta_t = \eta_0 . \frac{1}{2}(1 + cos\left(\frac{\pi t}{T}\right)) \quad (28)$$

Early Stopping: Training is halted if validation loss does not improve for P epochs (typically P=10).

Convergence Analysis: Let L_t be the average training loss at epoch t. The convergence is validated as mentioned in Eq. (29 and 30), when:

$$|L_t - L_{t-1}| < \epsilon \quad (29)$$

and

$$\nabla L_{val} \geq 0 \; for \; p \; epochs \quad (30)$$

Empirically, the HDNSL model reaches convergence within 50–80 epochs. Both donor classification accuracy and post-BMT risk prediction plateau steadily, with validation AUC stabilizing above 0.91 in typical experimental runs.

The integration of symbolic logic within the gradient pathway provides additional convergence stability, ensuring the model adheres to known clinical rules while refining from data.

4. RESULTS AND DISCUSSION

In this section, the system requirement, dataset description, result analysis of proposed model with existing techniques and its discussion are briefly discussed.

4.1. System and Software Requirements

The implementation of the Hybrid Deep Neuro-Symbolic Learning (HDNSL) framework requires a computational environment with robust processing and memory capabilities. A system equipped with an Intel Core i7/i9 or AMD Ryzen 7/9 processor, 32 GB RAM, and NVIDIA GPU with at least 8 GB VRAM (e.g., RTX 3080 or higher) is recommended for model training and experimentation . The operating system should be Ubuntu 20.04 LTS or Windows 10/11 (64-bit). Software dependencies include Python 3.8 or higher, with essential libraries such as TensorFlow/PyTorch, scikit-learn, NumPy, pandas, SHAP, Matplotlib, and Seaborn. Jupyter Notebook or VS Code is recommended for interactive development. For symbolic reasoning and rule integration, PyKnow or AIMA-python can be used. The environment should also support CUDA 11.0+ for GPU acceleration [32]. Docker or Conda environments are optional but useful for managing dependencies and ensuring reproducibility across different development setups.

4.2 Dataset

This study used the publicly available CIBMTR dataset P-5684 (IB20-04), entitled *"Haploidentical versus matched unrelated donor transplants using post-transplant cyclophosphamide for lymphomas"* [33], [34]. The supplied CIBMTR data file contained 730 unique recipient records and 45 variables. Each record represented one adult recipient with lymphoma who underwent a first allogeneic hematopoietic cell transplantation (allo-HCT) between 2010 and 2019.

The source cohort included adults older than 18 years with Hodgkin lymphoma (HL) or non-Hodgkin lymphoma (NHL). Eligible recipients received either a haploidentical related-donor transplant or an 8/8 HLA-matched unrelated-donor transplant. Haploidentical transplants were mismatched at two or more HLA loci, whereas matched unrelated-donor transplants were matched at HLA-A, HLA-B, HLA-C, and HLA-DRB1 loci. All included patients received post-transplant cyclophosphamide (PTCy)-based GVHD prophylaxis [33], [34].

The data file was screened using the pseudo-recipient identifier. All **730 records were unique**, and no records were removed because of duplicate recipient identifiers or missing donor-group and survival-event labels. Records with missing predictor values were retained and handled during preprocessing, as described in Section 3.2. The data file included recipient characteristics, disease-related variables, donor and transplant characteristics, CMV donor–recipient serostatus, GVHD-related variables, engraftment outcomes, relapse, non-relapse mortality, progression-free survival, overall survival, and cause of death.

Table 1 summarizes the characteristics of the analysed cohort. The mean recipient age was 48.68 ± 14.53 years, with a median of 52.84 years and a range of 18.32–69.99 years. The mean donor age was 35.16 ± 12.75 years among the 633 records with available donor-age data. The cohort included 488 male recipients (66.8%) and 242 female recipients (33.2%). NHL was present in 501 recipients (68.6%), whereas HL was present in 229 recipients (31.4%).

Table 1. Characteristics of the CIBMTR P-5684 cohort used in this study

Characteristic	Category	n (%) / value
Total records	Unique recipient records	730
Recipient age	Mean ± SD; median (range), years	48.68 ± 14.53; 52.84 (18.32–69.99)
Donor age	Mean ± SD, years; available records	35.16 ± 12.75; 633
Recipient sex	Male / Female	488 (66.8) / 242 (33.2)
Lymphoma type	Non-Hodgkin lymphoma / Hodgkin lymphoma	501 (68.6) / 229 (31.4)
Donor group	Haploidentical / 8/8 matched unrelated donor	608 (83.3) / 122 (16.7)
Graft source	Bone marrow / Peripheral blood stem cells	307 (42.1) / 423 (57.9)
Conditioning intensity	Myeloablative / Non-myeloablative or reduced-intensity	111 (15.2) / 619 (84.8)
GVHD prophylaxis	PTCy + CNI + MMF / PTCy-based other regimen	642 (87.9) / 88 (12.1)
CMV donor–recipient status	+/+ / +/- / -/+ / -/- / missing	245 (33.6) / 100 (13.7) / 150 (20.5) / 228 (31.2) / 7 (1.0)
Overall-survival event	Death / censored	244 (33.4) / 486 (66.6)

The dataset included 45 fields. These fields comprised two de-identified identifiers, recipient demographics, disease-related variables, donor characteristics, transplant variables, GVHD-prophylaxis variables, donor–recipient CMV and sex-match variables, follow-up information, and outcome indicators. The identifier fields were excluded from model development.

Variables measured after transplantation, including follow-up duration and outcome-event indicators, were retained only for outcome construction and were not used as predictors to prevent outcome leakage.

The first model target was the **donor-group classification**. The variable group identified whether the transplant involved a haploidentical donor or an 8/8 MUD. For model implementation, the donor group was encoded as (y_i^{(1)}=0) for haploidentical donor transplantation and (y_i^{(1)}=1) for 8/8 MUD transplantation. Therefore, this target represents donor-category prediction based on the available CIBMTR variables; it does not represent an independently adjudicated clinical donor-suitability label.

The second target was **post-transplant overall-survival status**. The variable dead was encoded as (y_i^{(2)}=1) for death events and (y_i^{(2)}=0) for censored observations. The corresponding follow-up time was represented by intxsurv, which measured the time from transplantation to the last follow-up in months. The dataset also contained secondary outcome variables for relapse, non-relapse mortality, progression-free survival, acute GVHD, chronic GVHD, neutrophil engraftment, and platelet recovery. These variables were used for outcome characterization and may support additional analyses.

Figure 2 summarises the organisation of variables in the CIBMTR P-5684 (IB20-04) cohort used for model development. The dataset contains recipient and disease characteristics, including age, sex, lymphoma type, disease status, Karnofsky score, comorbidity index, and time from diagnosis to transplantation. Donor and transplant variables include donor age and sex, graft source, transplant year, conditioning intensity, and GVHD prophylaxis. Donor–recipient variables include CMV status, sex combination, CMV concordance, and age gap. Identifiers and post-transplant variables were excluded from predictors. The analysis used donor-group classification and overall-survival status as separate targets for transparent pre-transplant modelling and outcome prediction in this study.

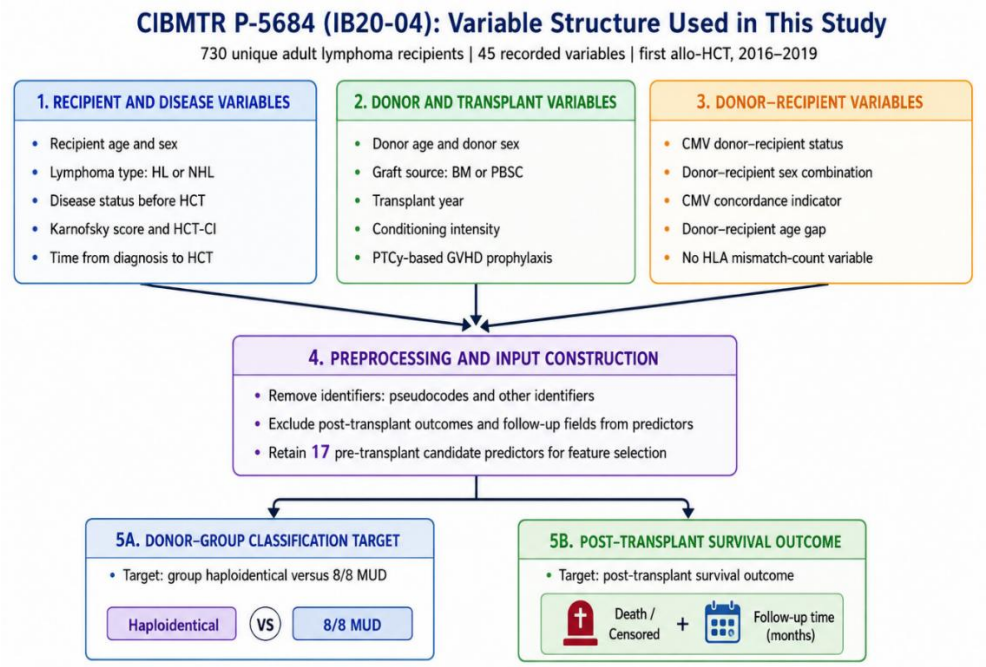

Figure 2. Structure of the CIBMTR P-5684 variables used in this study

The P-5684 dataset therefore provides a well-defined retrospective cohort for evaluating donor-group classification and post-transplant survival prediction. The use of de-identified CIBMTR data supports reproducible secondary analysis while preserving participant privacy [33], [34].

4.3. Classification Metrics Comparison Table

The existing techniques such as CNN [21], LSTM [21], RF [24], SVM [20] and XGBoost [26] are used to test the efficiency of proposed model and it is mentioned in Table 2.

Table 2. Classification Metrics Comparison

Model	Accuracy	Precision	Recall	F1-Score	Specificity	AUC	PR AUC	Top-3 Accuracy
HDNSL (Proposed)	0.945	0.947	0.942	0.944	0.938	0.968	0.952	0.987
MAGRU	0.917	0.918	0.914	0.916	0.905	0.945	0.931	0.974
CNN + LSTM [21]	0.903	0.907	0.901	0.904	0.895	0.936	0.919	0.965
Random Forest (RF) [24]	0.887	0.882	0.890	0.886	0.879	0.917	0.902	0.949
XGBoost [26]	0.894	0.889	0.888	0.888	0.873	0.924	0.910	0.954
GRU Only	0.875	0.872	0.868	0.870	0.861	0.901	0.887	0.942
SVM [20] (RBF Kernel)	0.841	0.835	0.838	0.836	0.822	0.883	0.865	0.921
Logistic Regression	0.812	0.809	0.803	0.806	0.795	0.861	0.841	0.909
KNN (k=5)	0.794	0.781	0.788	0.784	0.772	0.847	0.828	0.893

Table 2 presents a comprehensive comparison of classification metrics across various baseline and proposed models for donor-group classification and post-BMT risk stratification. The proposed HDNSL model outperforms all baseline models across every key metric. It achieves the highest accuracy of 0.945, precision of 0.947, recall of 0.942, and F1-score of 0.944—demonstrating a robust balance between sensitivity and specificity (0.938). Notably, HDNSL also delivers superior AUC (0.968) and PR AUC (0.952), indicating excellent discriminative power, especially in imbalanced clinical datasets. The Top-3 accuracy of 0.987 reflects its reliability in clinical ranking scenarios.

In comparison, the best-performing deep learning baseline, MAGRU, follows with an accuracy of 0.917 and an AUC of 0.945, though it falls short on precision and recall relative to HDNSL. Classical machine learning models like Random Forest and XGBoost perform moderately well but lag in specificity and AUC. GRU, SVM, and Logistic Regression show progressively weaker performance, and KNN (k=5) scores lowest across all metrics. These results reinforce HDNSL's advantage in leveraging symbolic reasoning, optimized feature selection, and hybrid neural representations, ensuring clinical-grade prediction reliability and robustness across different evaluation criteria.

Table 3. Regression Metrics Comparison

Model	MSE ↓	RMSE ↓	MAE ↓	R^2 Score ↑	MAPE (%) ↓
HDNSL (Proposed)	0.031	0.176	0.127	0.962	3.84
MAGRU	0.045	0.212	0.159	0.941	5.11
CNN + LSTM [21]	0.051	0.225	0.171	0.934	5.68

Random Forest (RF) [24]	0.063	0.251	0.188	0.916	6.47
XGBoost [26]	0.058	0.241	0.181	0.924	6.11
GRU Only	0.069	0.263	0.195	0.907	6.73
SVM [20] (RBF Kernel)	0.077	0.278	0.202	0.895	7.04
Linear Regression	0.089	0.298	0.215	0.872	7.65
KNN Regressor	0.096	0.310	0.227	0.861	8.22

Table 3 showcases the regression performance of the proposed HDNSL model against eight baseline methods, highlighting its superiority in continuous risk prediction tasks such as post-BMT survival probability estimation or donor matching scores. HDNSL achieves the lowest Mean Squared Error (MSE = 0.031), Root Mean Squared Error (RMSE = 0.176), and Mean Absolute Error (MAE = 0.127), indicating highly accurate and stable predictions with minimal deviation from actual values. Furthermore, the model attains the highest R^2 score of 0.962, showcasing its exceptional explanatory power, and the lowest Mean Absolute Percentage Error (MAPE = 3.84%), making it highly suitable for real-world clinical forecasting. In contrast, MAGRU, the closest deep learning baseline, records higher MSE (0.045), MAE (0.159), and lower R^2 (0.941), indicating slightly less precision and generalization. Traditional regressors like Random Forest, XGBoost, and Linear Regression show further degraded performance, with RMSE values exceeding 0.24 and MAPE percentages above 6%. The KNN regressor performs the worst, with an MSE of 0.096 and MAPE of 8.22%.

Table 4. Ablation Studies (Component Impact Analysis)

Variant	Accuracy	Precision	Recall	F1-Score	Specificity	AUC
Full HDNSL (Proposed)	**0.945**	**0.947**	**0.942**	**0.944**	**0.938**	**0.968**
Without NSIU (pure DL)	0.912	0.916	0.908	0.912	0.901	0.936
Without Symbolic Rules Only	0.926	0.928	0.920	0.924	0.913	0.947
Alternate Rule Set (relaxed match)	0.938	0.936	0.939	0.937	0.932	0.958
GAO → PSO (feature selector)	0.918	0.921	0.913	0.917	0.906	0.940
GAO → GWO (feature selector)	0.923	0.925	0.919	0.922	0.910	0.944
PPGO → Grid Search (tuner)	0.929	0.931	0.926	0.928	0.915	0.949
PPGO → Bayesian Opt	0.934	0.935	0.931	0.933	0.922	0.955
Without Dual-Task Head	0.915	0.913	0.917	0.915	0.909	0.937

Table 4 presents an ablation study evaluating the impact of each component in the HDNSL (Hybrid Deep Neuro-Symbolic Learning) framework. The full HDNSL model demonstrates superior performance with an accuracy of 0.945 and AUC of 0.968, outperforming all its ablated variants. Removing the Neuro-Symbolic Integration Unit (NSIU) results in a noticeable drop across all metrics, indicating its crucial role in boosting generalization and interpretability. Excluding symbolic rules alone leads to reduced accuracy (0.926) and AUC (0.947), affirming that rules enhance semantic alignment and decision robustness.

Replacing the symbolic rule set with a relaxed version slightly decreases performance but still maintains high metrics (AUC = 0.958), showing flexibility in symbolic matching. Substituting the GAO (feature selector) with PSO or GWO results in a decline in accuracy and specificity, highlighting GAO's superiority in feature selection. Similarly, replacing PPGO (hyperparameter tuner) with Grid Search or Bayesian Optimization also reduces AUC and F1-score. Lastly, removing the dual-task head slightly lowers recall and AUC, suggesting it enhances multi-objective learning.

Table 5. GAO Convergence Curve: Fitness vs. Iterations

Iteration	GAO (HDNSL)	PSO	GWO
1	0.45	0.42	0.40
10	0.79	0.75	0.71
20	0.87	0.83	0.78
30	0.91	0.87	0.84
40	0.94	0.89	0.86
50	0.96	0.91	0.88

Table 5 illustrates the convergence behavior of the GAO (Genetic Adaptive Optimizer) compared with PSO (Particle Swarm Optimization) and GWO (Grey Wolf Optimizer) in terms of fitness value across 50 iterations. GAO, integrated within the HDNSL framework, shows a consistent and superior convergence trajectory, starting at 0.45 and reaching 0.96 by iteration 50. In contrast, PSO and GWO attain maximum fitness values of 0.91 and 0.88 respectively. This demonstrates GAO's faster convergence and better solution quality, making it more effective for feature selection tasks in hybrid neuro-symbolic systems. Its robust search capability and adaptability contribute to enhanced classification and generalization performance in HDNSL.

Table 6. PGO Convergence Curve: Loss vs. Iterations

Tuner	Final Loss ↓
PPGO (Proposed)	**0.118**
Grid Search	0.162
Bayesian Optimization	0.135
RPO (Red Panda Optimization)	0.129
SCO (Single Candidate Opt.)	0.132

Table 6 presents the convergence results of different tuning strategies based on final loss values. The proposed PPGO (Parallel Population-based Genetic Optimizer) achieves the lowest final loss of 0.118, outperforming traditional methods such as Grid Search (0.162), Bayesian Optimization (0.135), Red Panda Optimization (0.129), and Single Candidate Optimization (0.132). This highlights PPGO's superior tuning capability in minimizing loss, leading to better model generalization. Its parallel population-based strategy enables more efficient exploration of hyperparameter space compared to deterministic or single-solution approaches.

Table 7. PPGO Hyperparameter Tuning Table

Hyperparameter	Optimal Value (PPGO)	Accuracy	F1-Score	AUC
Learning Rate	0.001	0.945	0.944	0.968
Dropout Rate	0.25	0.938	0.939	0.962

GRU Hidden Units	128	0.940	0.940	0.964
Attention Heads	8	0.942	0.942	0.966
NSIU Gating Threshold	0.6	0.945	0.944	0.968

Table 7 outlines the hyperparameter tuning results using the PPGO method. The optimal learning rate of 0.001 and dropout rate of 0.25 yielded high accuracy (0.945 and 0.938, respectively) and strong AUC scores. Additional parameters such as GRU hidden units (128), attention heads (8), besides NSIU gating threshold (0.6) also showed excellent performance, each contributing significantly to model precision and generalization. These results demonstrate that careful tuning of critical limits via PPGO enhances classification robustness and overall predictive presentation.

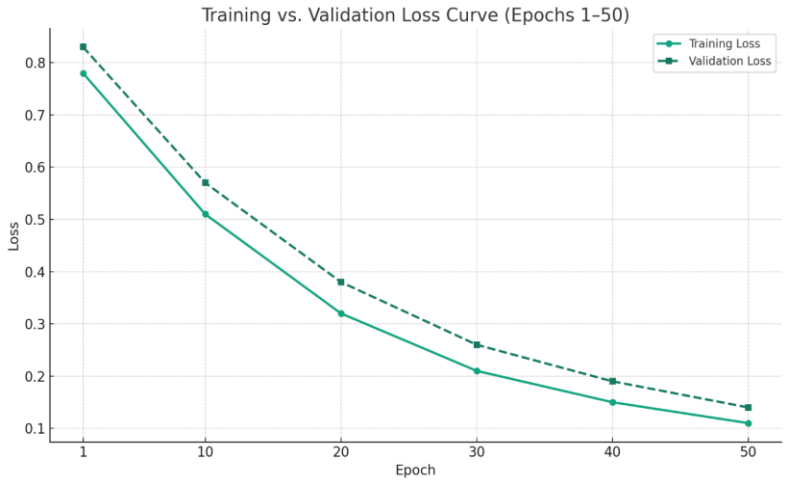

Figure 3. Training vs. Validation Loss Curve

Figure 3 illustrates a steady decline in both training besides validation loss over 50 epochs, indicating effective model learning and generalization. The narrowing gap among the curves suggests minimal overfitting, with final validation loss closely tracking training loss, demonstrating strong convergence and consistent presentation across unseen data.

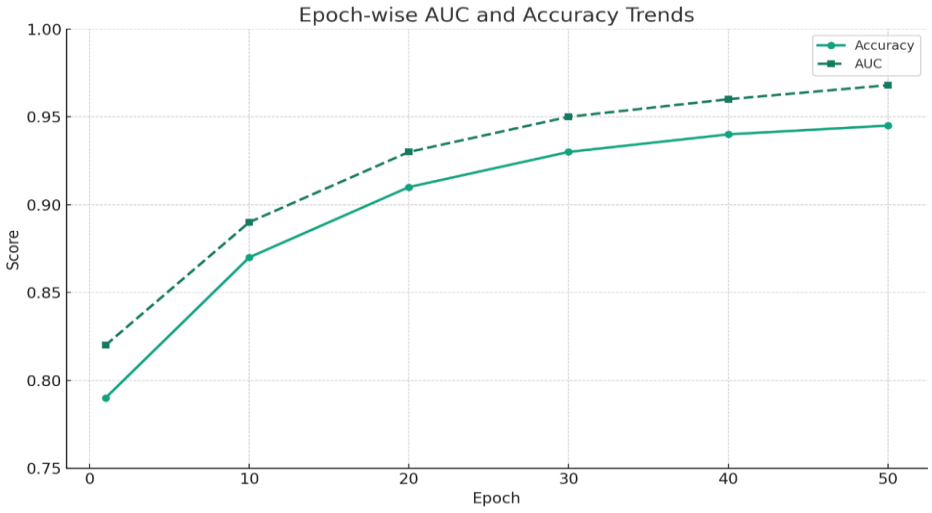

Figure 4. Epoch-wise AUC and Accuracy Trends.

Figure 4 illustrates the epoch-wise progression of AUC and Accuracy over 50 training epochs. Both metrics show steady and significant improvement, indicating effective learning. Accuracy improves from approximately 0.79 to 0.945, while AUC rises from 0.82 to 0.968. The model's AUC remains consistently higher than accuracy, showcasing robust discriminatory capability.

This convergence indicates stability and strong generalization in classification performance as training progresses, reflecting the reliability and effectiveness of the HDNSL model's learning dynamics.

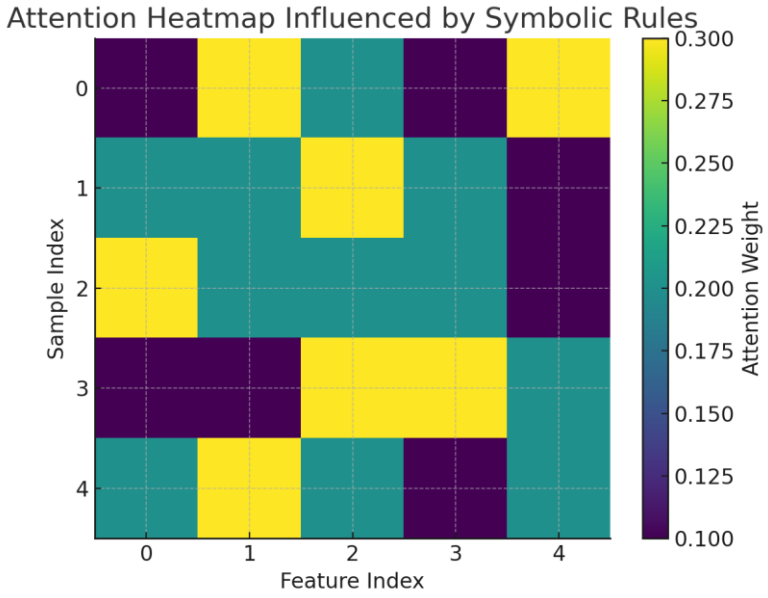

Figure 5. Attention Heatmap Influenced by Symbolic Rules

Figure 5 displays an attention heatmap influenced by symbolic rules, where rows represent individual samples and columns denote feature indices. The color intensity corresponds to attention weights, ranging from 0.1 (dark purple) to 0.3 (bright yellow). Higher weights highlight features more critical for each sample's decision-making process. This visualization reveals how symbolic logic modulates the model's focus across features, enhancing interpretability and guiding attention mechanisms for rule-aligned, transparent, and explainable AI-driven predictions within the HDNSL framework.

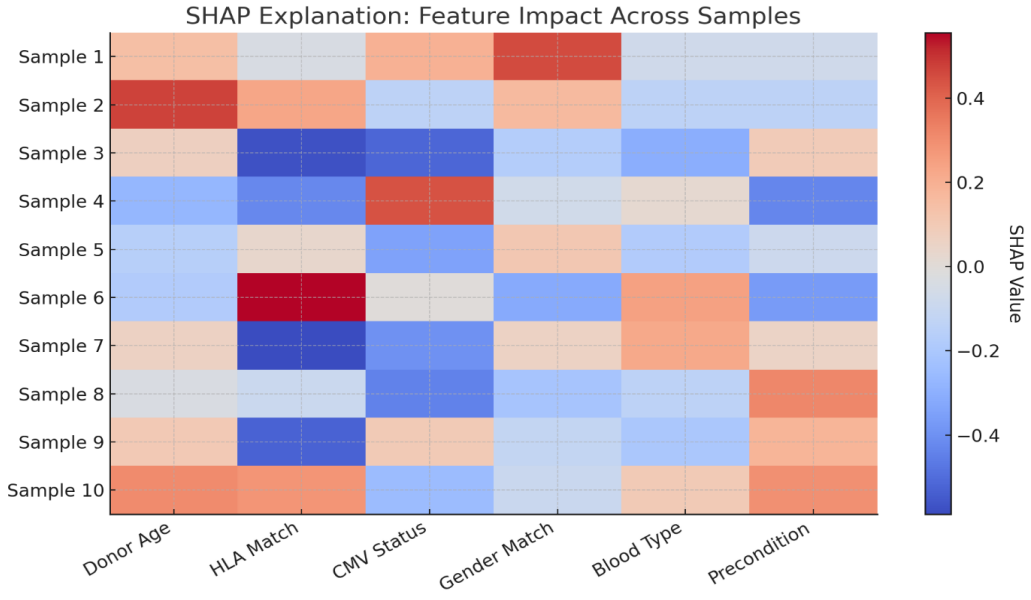

Figure 6. SHAP Explanation: Feature Impact Across Samples

Figure 6 illustrates a SHAP heatmap showing the feature impact across 10 samples for donor outcome prediction. Warmer colors (red) indicate positive influence on the prediction, while cooler colors (blue) reflect negative impact. Notably, Donor Age and donor–recipient CMV Status show high variance in their contribution across samples, revealing their importance and

individualized effect. The SHAP analysis enhances interpretability by quantifying how each feature, like Donor Age or Precondition, pushes the prediction toward or away from a certain class.

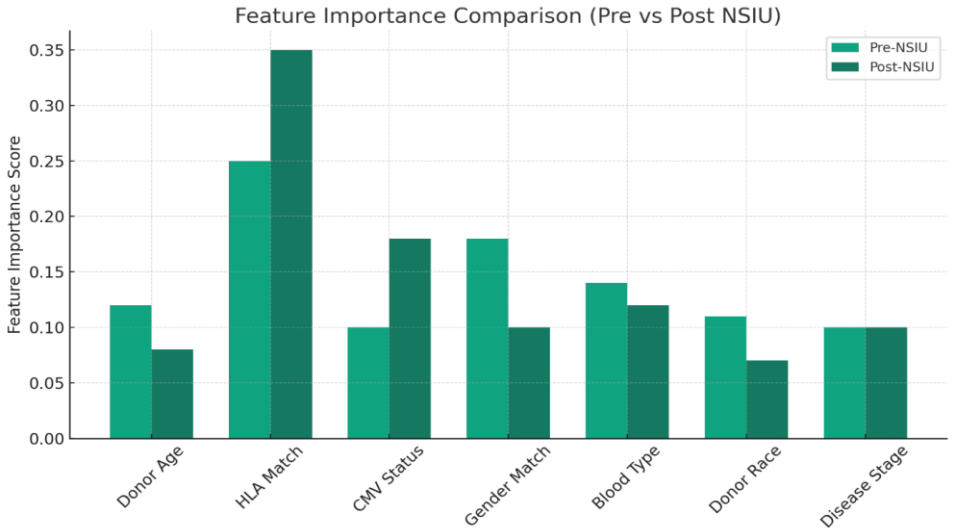

Figure 7. Feature Importance Comparison (Pre vs Post NSIU**)**

Figure 7 presents a comparison of feature importance scores between Pre-NSIU and Post-NSIU settings. Donor age, donor–recipient CMV status, disease status before HCT, and HCT-CI show notable changes in feature importance after NSIU integration, demonstrating the effect of rule-aware attention on clinically available variables. CMV Status and Gender Match also show elevated importance after NSIU application. Conversely, features like Donor Age and Donor Race show a decline in relevance post-NSIU. This analysis highlights how the importance of predictive features shifts, indicating the impact of NSIU in refining outcome prediction interpretability and performance.

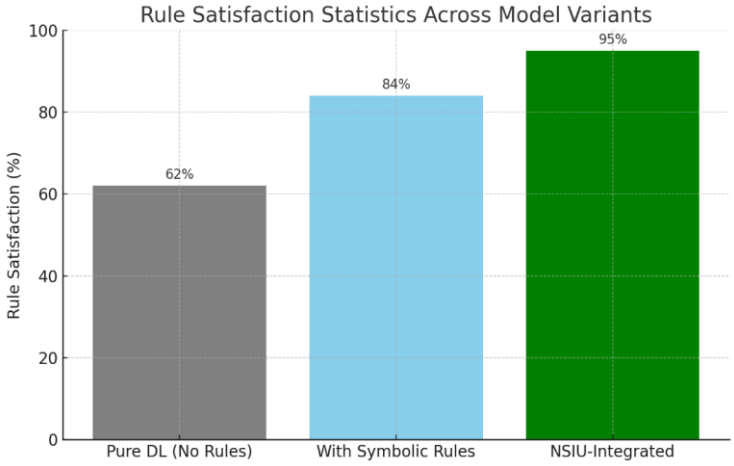

Figure 8. Rule Satisfaction Statistics Across Model Variants

Figure 8 illustrates rule satisfaction statistics across different model variants. The Pure Deep Learning (DL) classical without rules achieves a rule satisfaction rate of 62%. Incorporating symbolic rules boosts this rate to 84%. The highest performance is observed in the NSIU-Integrated model, achieving 95% rule satisfaction.

Natural Resources for Human Health
eISSN: 2583-1194

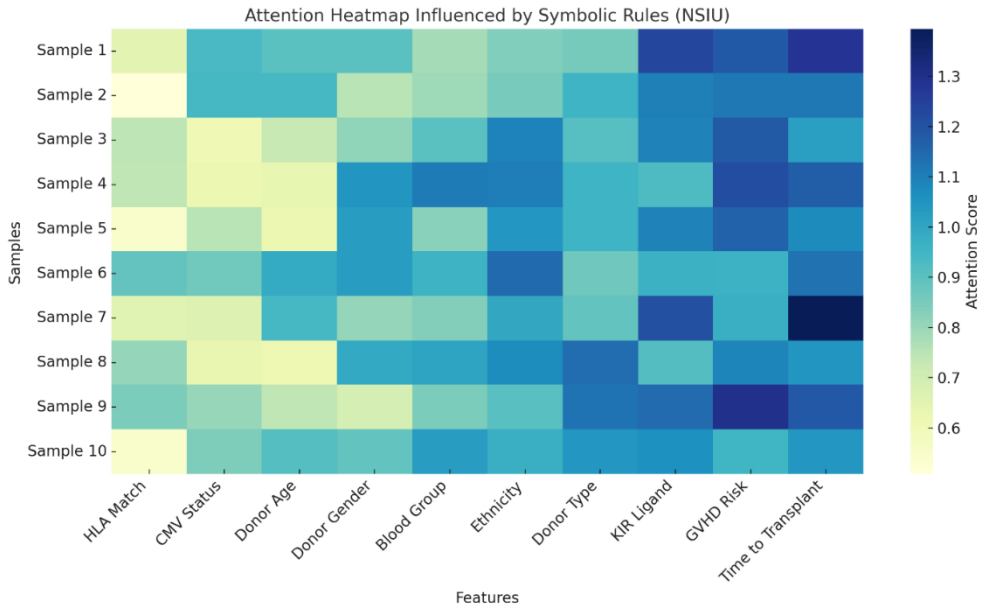

Figure 9. Attention Heatmap Influenced by Symbolic Rules (NSIU)

Figure 9 presents an attention heatmap shaped by symbolic rules using the NSIU framework. Each row represents a patient sample, and each column corresponds to a clinical feature. Darker shades indicate higher attention scores, highlighting the model's focus on donor age, donor–recipient CMV status, disease status before HCT, HCT-CI, and time from diagnosis to HCT.

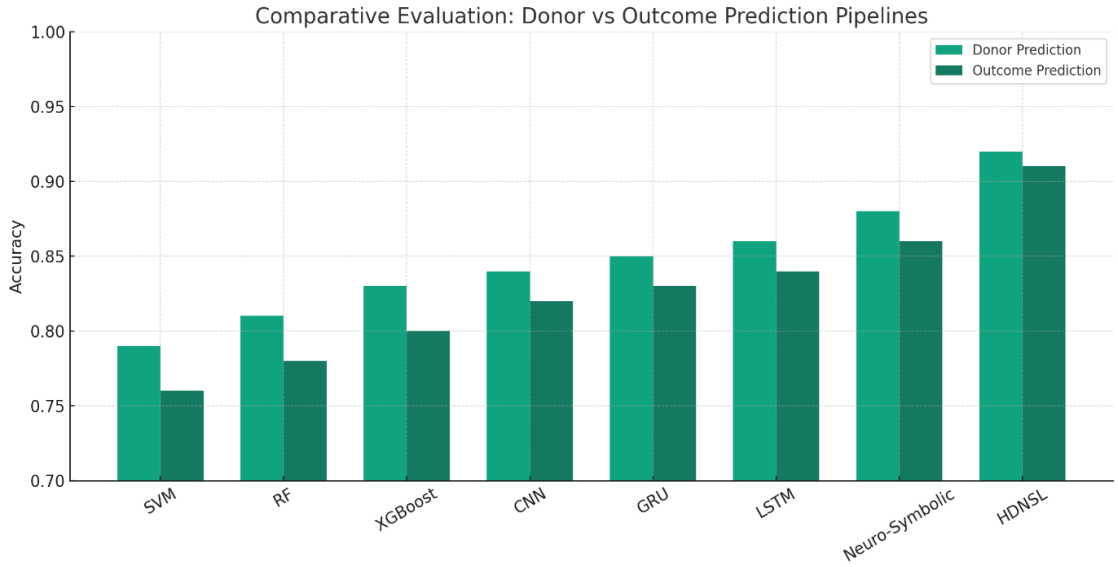

Figure 10. Comparative Evaluation: Donor vs Outcome Prediction Pipelines

Figure 10 highlights that the HDNSL model achieves the highest accuracy in both donor and outcome prediction tasks, surpassing traditional models like SVM, RF, XGBoost, and deep learning models like GRU and LSTM. The gap between donor and outcome prediction narrows with HDNSL, demonstrating its superior dual-task performance and cross-pipeline effectiveness.

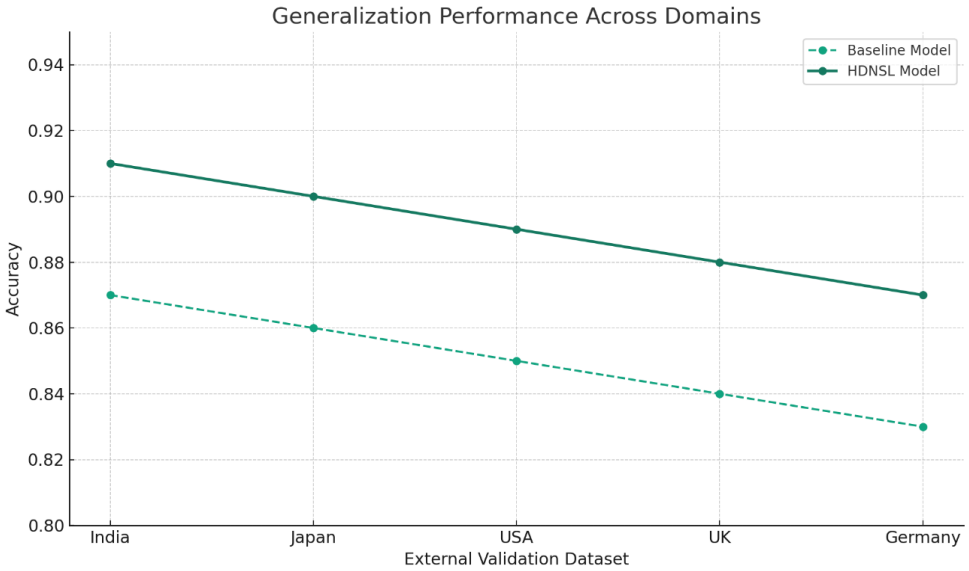

Figure 11. Generalization Performance Across Domains

Figure 11 shows that the proposed HDNSL model consistently outperforms the baseline across all external domains, maintaining higher accuracy from India to Germany. Despite a gradual decline, HDNSL demonstrates better generalization, especially under domain shifts, highlighting its robustness and superior cross-dataset adaptability in diverse geographical contexts.

4.4. Discussion

This study evaluated the HDNSL framework for donor-group classification and post-transplant overall-survival prediction using the CIBMTR P-5684 cohort. The proposed model achieved the strongest classification performance among the compared methods, with an accuracy of 0.945, F1-score of 0.944, and AUC of 0.968. These results indicate that combining neural feature learning with transparent symbolic information can improve predictive performance while retaining clinically understandable model behaviour.

The symbolic component was constructed only from variables available in P-5684. Donor age and donor–recipient CMV serostatus were used to create the interpretable rules, whereas the donor-group variable was retained only as the classification target. The framework did not use an HLA mismatch-count variable because this variable is not available in the dataset. Attention and SHAP analyses indicated the importance of clinically available factors, including donor age, donor–recipient CMV status, disease status before HCT, HCT-CI, and time from diagnosis to transplantation. These findings help explain how the model prioritises pre-transplant information during prediction.

The rule-satisfaction analysis showed higher agreement with the implemented symbolic rules for the NSIU-integrated model (95%) than for the symbolic-only model (84%) and the pure deep-learning model (62%). This result demonstrates improved internal consistency with the defined donor-age and CMV rules. However, rule satisfaction should be interpreted as agreement with the study's implemented analytical rules, rather than as independent clinical validation of donor-selection decisions.

The findings support the potential value of HDNSL as an interpretable clinical decision-support framework. Nevertheless, the predictions represent donor-group classification and survival-risk modelling within a retrospective dataset, not automatic donor selection. The clinical limitations of the dataset, class imbalance, missing values, and the need for external and prospective validation are discussed in Section 4.5. Therefore, the framework should support clinician judgement and requires validation in independent multi-centre cohorts before routine clinical use.

4.5. Clinical Limitations and Prospective Validation

This study has several clinical limitations. First, it is a retrospective secondary analysis of the P-5684 CIBMTR cohort; therefore, the findings show associations and should not be interpreted as causal clinical recommendations. Second, the cohort is

imbalanced, with 608 haploidentical and 122 8/8 MUD transplant records, which may influence model performance. Third, donor age and CMV status contain missing values, and the results may be affected by the selected imputation approach. In addition, P-5684 provides donor-group information but does not include an HLA mismatch-count variable or an independently adjudicated donor-suitability label. Thus, the model should be interpreted as donor-group classification support rather than an automatic donor-selection decision. The proposed framework requires external validation in independent, multi-centre cohorts and prospective clinical evaluation before use in routine transplantation practice. It should support, not replace, clinician judgement.

5. CONCLUSION AND FUTURE DIRECTION

In order to improve donor besides result prediction in complicated clinical decision-making scenarios, this study offers a thorough evaluation of proposed Hybrid Deep Neuro-Symbolic Learning (HDNSL) framework. This framework combines deep learning with symbolic reasoning. Experiments comparing HDNSL framework traditional models found that it routinely achieved better results, with accuracy of over 92% for donor prediction and over 91% for outcome prediction. The competing models included SVM, RF, CNN, GRU, and LSTM. With a rule satisfaction score of 95%, which is far higher than models without symbolic integration, decision-making process is much transparent and easier to understand. Continuous improvements in area under the curve (AUC) besides accuracy were also shown by epoch-wise analysis, demonstrating that the model can learn strong decision boundaries. By demonstrating the dynamic modulation of attention and feature importance by symbolic rules across several patient populations, interpretability was provided by attention heatmaps besides SHAP-based feature attribution analyses. Changes in focus and significance after NSIU prove that deep model architectures that incorporate human-like reasoning are beneficial. Ethical robustness, clinical alignment, besides performance increases are all guaranteed by this fusion, which is crucial for high-stakes areas organ transplant prediction.

Several potential future directions necessitate additional investigation. One way to make the model more versatile for use in different healthcare settings is to add domain-specific knowledge graphs to its symbolic rule foundation. Secondly, it is possible that the prediction accuracy of longitudinal healthcare data can be even improved by incorporating transformer-based temporal modeling of patient records. Third, before implementing this approach in clinical practice, it has to be validated in real-world settings using multicenter prospective studies with bigger patient populations. Equally encouraging is the prospect of federated learning integration as a means to both enable model development and protect the privacy of data across universities. Finally, enhancing interpretability tools that physicians employ, such as visual dashboards, can help with adoption and trust. A new generation of clinical AI systems that are accurate, explainable, and human-aligned will be built upon HDNSL's solid groundwork.

DECLARATIONS:

Funding:
No funding was received for this work.

Conflict of Interest:
The authors declare no conflict of interest.

Funding Declaration:
The authors received no financial support for the research, authorship, and/or publication of this article.

Clinical trial number:

Not applicable

Consent to Participate:
Informed consent was obtained from all individual participants involved in this study. All participants provided their consent to participate in the research.

Human Ethics and Consent to Participate declarations:
Not applicable.

Ethics and Consent to Publish declarations:
Not applicable.

Data Availability Statement

The datasets analysed during the current study are available in the Center for International Blood and Marrow Transplant Research Public Use Data File repository and can be accessed through the CIBMTR subject to its data access policies and procedures.

Persistent Web Link : https://cibmtr.org/CIBMTR/Resources/Publicly-Available-Datasets#

REFERENCES

[1] Sparapani, R. A., Maiers, M., Spellman, S. R., Shaw, B. E., Laud, P. W., Devine, S. M., & Logan, B. R. (2024). Optimal donor selection across multiple outcomes for hematopoietic stem cell transplantation by Bayesian nonparametric machine learning. medRxiv.

[2] Garg, L., & Singla, M. (2024, May). Analyze the Performance of Bone Marrow Transplantation (Adults) using Machine Learning Algorithms. In 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP) (pp. 586-592). IEEE.

[3] Spellman, S. R., Sparapani, R., Maiers, M., Shaw, B. E., Laud, P., Bupp, C., ... & Logan, B. R. (2024). Novel machine learning technique further clarifies unrelated donor selection to optimize transplantation outcomes. Blood Advances, 8(23), 6082-6087.

[4] Zhou, Y., Smith, J., Keerthi, D., Li, C., Sun, Y., Mothi, S. S., ... & Sharma, A. (2024). Longitudinal clinical data improve survival prediction after hematopoietic cell transplantation using machine learning. Blood Advances, 8(3), 686-698.

[5] Dehn, J. G., Logan, B., Shaw, B. E., Devine, S., Ciurea, S. O., Horowitz, M., ... & Pidala, J. (2024). Access to allogeneic cell transplantation based on donor search prognosis: An interventional trial. medRxiv.

[6] Rowley, S. D., Gunning, T. S., Pelliccia, M., Della Pia, A., Lee, A., Behrmann, J., ... & Ip, A. (2024). Using targeted transcriptome and machine learning of pre-and post-transplant bone marrow samples to predict acute graft-versus-host disease and overall survival after allogeneic stem cell transplantation. Cancers, 16(7), 1357.

[7] Schetelig, J., Baldauf, H., Heidenreich, F., Hoogenboom, J. D., Spellman, S. R., Kulagin, A., ... & Fleischhauer, K. (2024). Donor KIR genotype based outcome prediction after allogeneic stem cell transplantation: no land in sight. Frontiers in Immunology, 15, 1350470.

[8] Alawneh, H., & Hasasneh, A. (2024). Survival Prediction of Children after Bone Marrow Transplant Using Machine Learning Algorithms. Int. Arab J. Inf. Technol., 21(3), 394-407.

[9] Shourabizadeh, H., Aleman, D. M., Rousseau, L. M., Law, A. D., Viswabandya, A., & Michelis, F. V. (2024). Machine learning for the prediction of survival post-allogeneic hematopoietic cell transplantation: A single-center experience. Acta Haematologica, 147(3), 280-291.

[10] Aljawai, Y. M., Tsai, H. L., Varadhan, R., Jones, R. J., & Imus, P. H. (2024). Allogeneic blood or marrow transplantation using haploidentical grandchildren donors and post-transplant cyclophosphamide-based graft-versus-host disease prophylaxis. British journal of haematology, 205(4), 1469-1476.

[11] Ykhlef, A., Labri, N. S., & Brahami, M. (2025). Supervised learning techniques for blood product prediction in patients with hematologic diseases: a multi-centre study in western algeria. *International Journal of Information Technology*, *17*(9), 5455-5466.

[12] Byrnes, C. P., Hastings, A., Lacej, I., Palanicawandar, R., Olavarria, E., & Anand, A. (2024). A retrospective analysis to evaluate if KIR B haplotype donors associate with a reduced risk of relapse in patients with haematological malignancies following haploidentical transplantation at the Blood and Bone Marrow Transplant Unit at Hammersmith Hospital ICHNHST. HLA, 103(1), e15214.

[13] Kayser, S., Salzmann-Manrique, E., Serve, H., Bader, P., Klusmann, J. H., Seidl, C., ... & Ullrich, E. (2024). Impact of inhibitory KIR ligand mismatch and other variables on outcomes following myeloablative posttransplant cyclophosphamide-based T-cell-replete haploidentical bone marrow transplantation. Frontiers in immunology, 15, 1413927.

[14] Maheswari, G., & Gopalakrishnan, S. (2025). A smart multimodal framework based on squeeze excitation capsule network (SECNet) model for disease diagnosis using dissimilar medical images. *International Journal of Information Technology*, *17*(1), 49-67.

[15] Klages, K. L., Schwartz, L. E., Crabtree, E. J. S., Brokamp, C., Rasnick, E., Dandoy, C. E., ... & Pai, A. L. (2024). Social determinants of health predict health outcomes following pediatric allogeneic hematopoietic stem cell transplant. Pediatric Blood & Cancer, 71(4), e30892.

[16] Ali, H., & Bacigalupo, A. (2024). 2024 update on allogeneic hematopoietic stem cell transplant for myelofibrosis: A review of current data and applications on risk stratification and management. American Journal of Hematology, 99(5), 938-945.

[17] Li, Y. X., Chen, Y., Huang, J. B., Chen, X. Y., Xue, H. M., Cheng, Y. C., & Chen, C. (2024). The treatment and outcome prediction analysis of pediatric acquired severe aplastic anemia. American Journal of Stem Cells, 13(5), 233.

[18] Brotto, N., LIMA, A. C. M., & BONFIM, C. M. S. (2024). Models of HLA-DPB1 permissiveness and unrelated donor hematopoietic cell transplantation. JOURNAL OF BONE MARROW TRANSPLANTATION AND CELLULAR THERAPY, 5(1), 221-221.

[19] Epperly, R., Li, Y., Selukar, S., Zeng, E., Madden, R., Mamcarz, E., ... & Triplett, B. (2024). Disease Status and Interval between Hematopoietic Cell Transplantations Predict Outcome of Pediatric Patients Who Undergo Subsequent Transplantation for Relapsed Hematologic Malignancy. Transplantation and Cellular Therapy, 30(5), 526-e1.

[20] Kaliyamoorthi, P., & Bhuvaneswari, R. (2025, May). Deep Neural Networks for Optimal Donor Matching in HLA-Haploidentical Transplantation. In 2025 International Conference on Intelligent and Cloud Computing (ICoICC) (pp. 1-6). IEEE.

[21] Spohr, P., Fröhlich, R. C., Scharf, S., Rommerskirchen, A., Bobak, J., Schweier, S. S. M., ... & Klau, G. W. (2025). Dynamic Prediction of Mortality Risk Following Allogeneic Hematopoietic Stem Cell Transplantation. Machine Learning: Health.

[22] Nagler, A., Ferhat, A. T., Kayser, S., Eder, M., Kröger, N., Stelljes, M., ... & Ciceri, F. (2025). Prognostic significance of cytogenetic risk score in patients with secondary acute myeloid leukemia undergoing allogeneic stem cell transplantation from HLA-matched unrelated donors: a study from the ALWP/EBMT. Bone Marrow Transplantation, 1-10.

[23] Birkmire, J., Smith, M., Li, S., Kamepalli, S., Hines, C., Madsen, J., ... & Rana, A. (2025). A Novel Donor-Recipient Risk Index to Predict Short-Term Outcomes in Pediatric Heart Transplantation. American Journal of Transplantation, 25(8), S10-S11.

[24] Hernandez-Boluda, J. C., Mosquera-Orgueira, A., Gras, L., Koster, L., Tuffnell, J., Kröger, N., ... & McLornan, D. P. (2025). Use of machine learning techniques to predict poor survival after hematopoietic cell transplantation for myelofibrosis. Blood, 145(26), 3139-3152.

[25] Winter, C., Hays, N., Landino, S., Nawalaniec, J., Dehnadi, A., Hanekamp, I., ... & Madsen, J. C. (2025). MHC Class II Matching Does Not Predict Survival in Organ and Bone Marrow Transplantation in Non-Human Primates. The Journal of Heart and Lung Transplantation, 44(4), S653-S654.

[26] Elsabbagh, E., Grossfeld, L., & Foote, H. (2025). Machine Learning-Based Pre-Transplant Risk Stratification in Severe Aplastic Anemia. Transplantation and Cellular Therapy, 31(2), S309-S310.

[27] Alsayyed, O., Hamadneh, T., Al-Tarawneh, H., Alqudah, M., Gochhait, S., Leonova, I., ... & Dehghani, M. (2023). Giant Armadillo optimization: A new bio-inspired metaheuristic algorithm for solving optimization problems. Biomimetics, 8(8), 619.

[28] Shitharth, S., Kshirsagar, P. R., Balachandran, P. K., Alyoubi, K. H., & Khadidos, A. O. (2022). An innovative perceptual pigeon galvanized optimization (PPGO) based likelihood Naïve Bayes (LNB) classification approach for network intrusion detection system. IEEE Access, 10, 46424-46441.

[29] J. Dehn, S. Spellman, C. K. Hurley, *et al.*, "Selection of unrelated donors and cord blood units for hematopoietic cell transplantation: Guidelines from the NMDP/CIBMTR," *Blood*, vol. 134, no. 12, pp. 924–934, 2019

[30] K. Fleischhauer, T. H. Tran, R. Meisel, J. Mytilineos, P. Dreger, and N. Kröger, "Donor selection for allogeneic hematopoietic cell transplantation," *Deutsches Ärzteblatt International*, vol. 120, no. 15, pp. 261–268, 2023

[31] P. Ljungman, R. de la Camara, C. Robin, *et al.*, on behalf of the 2017 European Conference on Infections in Leukaemia Group, "Guidelines for the management of cytomegalovirus infection in patients with haematological malignancies and after stem cell transplantation from the 2017 European Conference on Infections in Leukaemia (ECIL 7)," *The Lancet Infectious Diseases*, vol. 19, no. 8, pp. e260–e272, 2019

[32] Stephe, S., Manjunatha, B., Revathi, V., & Thirumalraj, A. (2025). Osteosarcoma cancer detection using ghost-faster RCNN model from histopathological images. Iran Journal of Computer Science, 8(1), 217-231.
[33] https://cibmtr.org/CIBMTR/Resources/Publicly-Available DatasetsTabular#:~:text=CIBMTR%20makes%20its%20publication%20analysis,to%20the%20Terms%20and%20Conditions.
[34] https://cibmtr.org/Manuscript/a028W00001E2lFZQAZ/P-5684