



Machine Learning Classifiers on Predicting the Survival of Pediatric Hematologic Transplant Patients

Bryan S. Subingsubing¹, Danielle Cyrele D. Azarraga¹, Marvic Gabriel Ruiz¹, Ma Sheila A. Magboo¹[0000-0002-6221-7892] and Vincent Peter C. Magboo*¹[0000-0001-8301-9775]

¹ University of the Philippines Manila, Manila, Philippines

bssubingsubing@up.edu.ph, ddazarraga@up.edu.ph, mruiz@up.edu.ph, mamagboo@up.edu.ph, vcmagboo@up.edu.ph

Abstract. Predicting the survival outcomes of pediatric hematologic transplant patients remains to be a formidable task for clinicians. This study aims to assess the capability to predict survival post bone marrow transplantation of machine learning classifiers namely: random forest, decision trees, logistic regression, Naïve Bayes, support vector machine and AdaBoost. Several model configurations representing various methodological enhancements as to feature selection using mutual information, synthetic minority oversampling technique to address class imbalance and hyperparameter tuning were analyzed. The results showed random forest model consistently besting other models in all model configurations generating an accuracy of 97.37%, 93.75% recall, 100% precision, and 96.77% F1-score. The study underscores the importance of these methodological enhancements to further strengthen the predictive performance of machine learning models in pediatric patients post bone marrow transplantation. The results of this research study represent a significant advancement in leveraging machine learning techniques to upgrade the care for pediatric patients with hematologic malignancies.

Keywords: Pediatric hematologic transplant survivorship, mutual information, SMOTE, hyperparameter tuning

1 INTRODUCTION

Bone marrow transplantation, sometimes called Hematopoietic stem cell transplantation (HSCT) is one of the treatment armamentariums prescribed by physicians for various hematologic disorders such as acute myeloid leukemia, relapsed or refractory acute lymphoid leukemia, relapsed and refractory chronic myeloid leukemia, myelodysplastic syndrome, lymphomas as well as non-malignant diseases in the pediatric age group [1][2]. In children with relapsed/refractory acute leukemia, it remains to be the only curative option [3]. Through the years, there has been significant improvement in HCT techniques coupled with cellular therapy and the corresponding supportive care, paving the way for newer indications of HCT with beneficial clinical prognosis leading to improved survivorship [4]. Nonetheless, significant risks and

complications of HSCT persist to be a challenge brought about by infections, toxicity induced by chemotherapy and radiation therapy, and graft versus host disease predisposing patients to significant morbidity and mortality [1]. As such, predicting the outcomes of HSCT treatment, more particularly the survivorship, remains to be a formidable task. It is in this context where machine learning technologies can be useful adjunctive tools to the health professionals in predicting the survival of HSCT, assist clinicians in the necessary preparatory steps before commencement of HSCT, and potentially contributing to the development of an automated diagnostic system with high accuracy with reduced computation time. The use of machine learning methodologies has paved the way for the development of multiple clinical decision support applications generating actionable insights utilizing huge volume of patient clinical data [5].

Very few studies have been made to ascertain the usefulness of machine learning to predict survival among pediatric patients undergoing HSCT. To address this research gap, this study leverages machine learning algorithms to predict the outcomes of pediatric HSCT patients using a comprehensive dataset that includes various clinical and demographic features. This study aims to compare the performance of six classifiers—Decision Trees (DT), Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF) and Adaptive Boosting (AdaBoost) in predicting the survival of pediatric patients who underwent HSCT. Additionally, the study also aims to assess the added utility of employing various methodological enhancements such as feature selection via mutual information, synthetic minority oversampling technique (SMOTE) to address class imbalance and hyperparameter tuning in the predictive capability of the various models.

The significant contribution of this research is by providing clinicians valuable insights in the development of accurate and robust machine learning models that will be guide them in the decision-making in the clinical setting, tailor-fit treatment options ultimately leading to improved survival outcomes of our pediatric patients and thus enhancing their quality of life.

This study is organized in the following format: an introduction section is followed by Literature Review specifying the relevant research studies assessing machine learning approaches to predict HSCT survival. The Materials and Methods section specifies the dataset description, the preprocessing steps applied to the dataset, methodological enhancements and the machine learning models employed in the study. The presentation and discussion of the results are then highlighted in the next section. Finally, the study culminated with a synopsis of the study and further recommendations for future research study.

2 LITERATURE REVIEW

Artificial intelligence techniques have been applied in various clinical prediction models for assessing survival in many clinical scenarios. Alawneh and Hasasneh analyzed several machine learning classifiers to predict survival after bone marrow transplantation with RF, AdaBoost and Gradient Boosting generating the best accuracy

rate of 97.37% [6]. Authors concluded that the use of a feature selection approach based on feature importance plot as well as hyperparameter tuning have significantly improved the prediction performance. In the study by Chadaga et al., authors utilized machine learning with explainable artificial intelligence techniques such as Shapley additive values (SHAP), local interpretable model-agnostic explanations (LIME) to predict survival rates of pediatric patients who underwent HSCT [7]. The best performing model yielded an 89% accuracy, 88% recall, 88% F1-score and 92% area under the curve (AUC). Authors also concluded that the most important features to predict survival were relapse, donor age, recipient age, and platelet recovery time. In [8], researchers investigated several machine learning classifiers such as logistic regression, multilayer perceptron, random forest and Bayes network to predict the survival of bone marrow transplant patients with multilayer perceptron and Bayes network generating 94.1% accuracy. The prediction performance between a full dataset and a reduced dataset obtained using a correlation feature subset method was likewise compared with the reduced dataset yielding better prediction capability. Ratul et al., analyzed several supervised machine learning classifiers to predict the survival of patients who underwent HSCT [9]. The Chi-square feature selection method as well as hyperparameter optimization using the grid search cross-validation technique were also employed for improved prediction which resulted in 94.73% accuracy. Authors concluded that this approach was a noteworthy contribution to the development of machine learning-based healthcare prediction systems more prominently in scenarios where resources are very limited.

In the study of Zhou et al., a risk-prediction algorithm to predict short- and long-term survival after allo-HCT in pediatric patients was developed using Naïve Bayes machine learning by analyzing clinical measurements performed before and after allo-HCT [10]. Authors stressed that unlike traditional prognostic tools utilizing fixed baseline demographic, clinical and laboratory variables for risk assessment and survival prediction, their method incorporated dynamic variability in the clinical and longitudinal laboratory data to supplement the predictive capability which has significantly improved mortality prediction. In [11], authors developed a model based on XGBoost to predict survival after HSCT utilizing the European Society for Blood and Marrow Transplant registry data generating a significantly higher AUC at 73% at 1 year compared to a simple linear Cox model. The effects of the main predictors of the model included diagnosis, performance score, patient age and donor type and were readily visualized using an explainable artificial intelligence method via SHAP ML. Choi et al., developed a gradient boosting machine learning with 10-fold cross-validation predictive model for assessing 5-year post-transplantation survival after allogeneic HSCT in patients with hematologic malignancies [12]. The model generated a 78.8% mean AUC. Authors concluded that the model aided clinicians for decision-making on issues involving patient and donor candidates, and other transplantation-related resources. Xiang et al., analyzed various machine learning classifiers (DT, RF, SVM, AdaBoost, Gradient Boosting and linear regression) utilizing demographic and baseline laboratory data from Granulocyte colony-stimulating factor (G-CSF) – mobilized donors, to predict patients who may have less than optimal stem cell yields after a single leukapheresis session (early donor screening stage) [13]. The

best classifier generated a significantly higher AUC at 89% as compared to the traditional statistical model. In [14], authors developed a random forest – based predictive model for leukemic relapse utilizing pre-transplant and post-transplant patient variables with emphasis on specifically lineage-specific chimerism. This is coupled with a LIME interpretable machine learning tool. The predictive model generated the following performance metrics: 85% accuracy, 85% sensitivity, 89% specificity for acute lymphocytic leukemia, and 81% accuracy, 75% sensitivity, and 100% specificity for acute myeloid leukemia in predicting relapses within 24 months after HSCT.

In summary, most of the published works emphasized the prognostic capacity of machine learning classifiers to assess survival of patients post HSCT. These studies collectively represent a significant advancement in leveraging machine learning techniques to enhance care for pediatric patients with hematologic malignancies undergoing HSCT. Moreover, only a handful of these studies appraised the contributory impact of methodological enhancements as to the application of a feature selection approach, addressing presence of data imbalance and utility of hyperparameter tuning to amplify prognostic capability of machine learning models.

3 MATERIALS AND METHODS

The study adhered to the machine learning pipeline outlined in Figure 1, analyzing a total of eight model configurations representing various combinations of methodological enhancements. The dataset was subjected to data cleaning and preprocessing steps. The configurations varied based on the application of mutual information-based feature selection (MIFS), addressing imbalance with SMOTE, and applying hyperparameter tuning to the various models.

3.1 Dataset Description

The study utilized the "Bone Marrow Transplant: Children" dataset from the University of California Irvine Machine Learning Repository [15]. This dataset comprises 187 instances and 36 features, encompassing both numerical and categorical data, and includes missing values. It features data on pediatric patients with various conditions, including cancerous ailments such as acute myelogenous leukemia and acute lymphoblastic leukemia, as well as nonmalignant diseases like Fanconi anemia and aplastic anemia. The target variable, 'survival status', is represented by '0' for alive (55%) and '1' for dead (45%). All patients in the dataset underwent unmanipulated allogeneic unrelated donor hematopoietic stem cell transplantation.

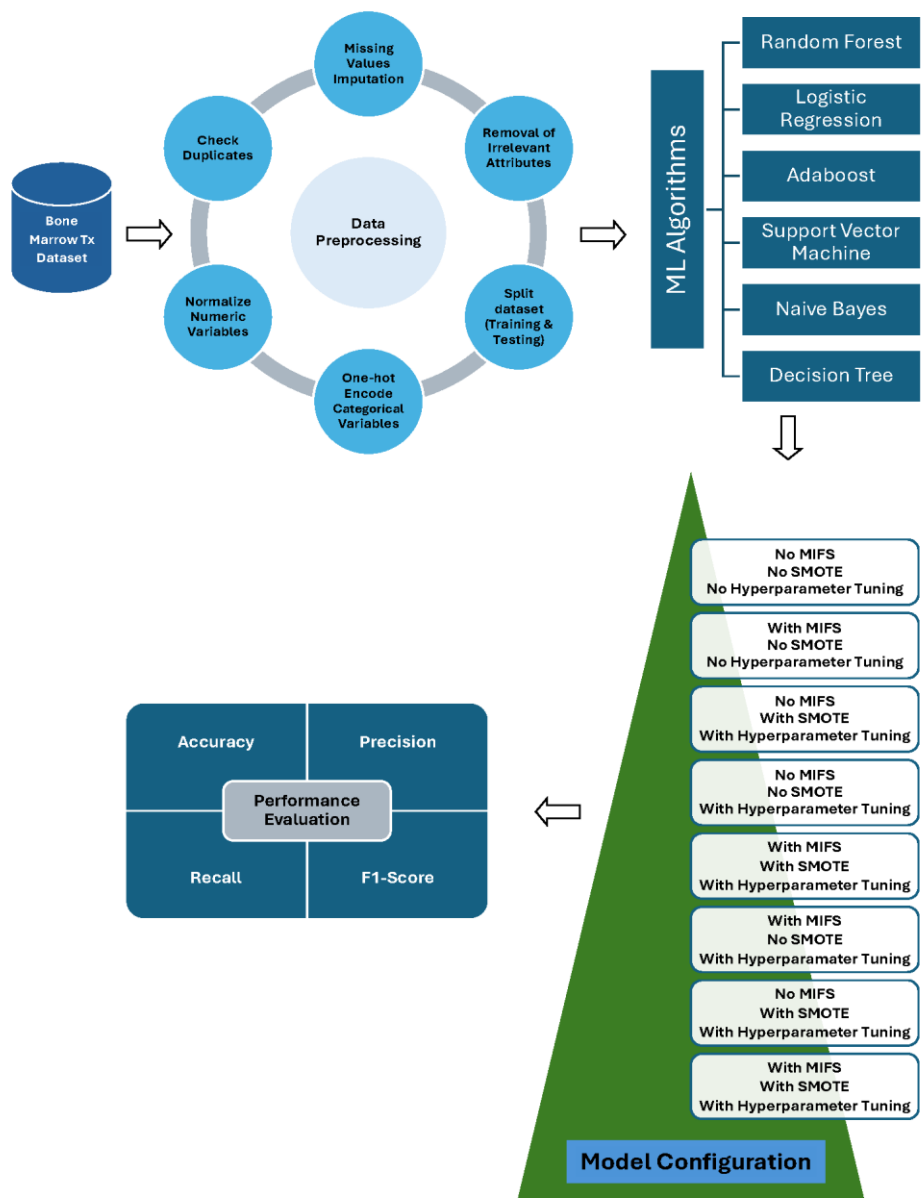


Fig. 1. Machine Learning Pipeline for Predicting Survival of Pediatric Hematologic Transplant Patients

3.2 Pre-processing Steps

To adequately prepare the dataset for machine learning processes, it underwent thorough data cleaning during the preprocessing phase. Missing values were imputed, duplicate rows were checked, irrelevant and redundant attributes were removed, and categorical variables were encoded using one-hot encoding to prevent the tendency of the model to misinterpreting the data as having some sort of numerical order. Numerical predictors with missing values were imputed using their median, as the mean was avoided due to the significant influence of outliers. Missing categorical variables were replaced with their respective modes. Additionally, the data was standardized using the StandardScaler library to ensure uniformity across the dataset.

3.3 Feature Selection and Class Imbalance Handling

In a full dataset, irrelevant and redundant features as well as noise are also included by the models in predicting survival status after HSCT. To assess whether reducing the number of features could enhance the performance of various machine learning models, four out of the eight model configurations employed mutual information-based feature selection (MIFS), as indicated in Table 1. MIFS is a filter-based feature selection method which showcases its capacity to measure linear and non-linear association between the target and independent attributes [16]. It is also a popular feature selection approach as it possesses the maximum relevance and minimum redundancy criteria of feature selection. [16]. In this study, the top 20 features were selected based on their mutual information scores. Additionally, the dataset exhibited a mild imbalance in the target feature, 'survival status', with 55% of cases classified as '0' (alive) and 45% as '1' (dead). In the presence of class imbalance, machine learning models usually misclassify the minority class because of the fewer number of samples. To address this mild imbalance, a data level method using SMOTE was utilized. SMOTE is an oversampling method which generates synthetic minority class to increase its number leading to a balanced distribution using an iterative search and selection method [17]. Through this approach, the significance of the minority class samples is enhanced in assessing the decision criterion of the machine learning models.

3.4 Hyperparameter Tuning

Hyperparameter tuning was employed to optimize the model parameters for achieving the best possible results. This process involved outlining potential values for each parameter and conducting a grid search using GridSearchCV method to evaluate the performance of models across each combination of parameters. The parameters from the best performing model were then used for training and testing the machine learning models. The parameter grid varied across models: Random Forest (RF) utilized settings such as the number of trees in the forest, maximum tree depth, and minimum samples required at node and leaf levels; Logistic Regression (LR) adjusted factors like the inverse of regularization strength, penalization norm, and optimization algorithm; AdaBoost set the number of boosting stages and learning rate; Support Vector

Machines (SVM) configured the regularization constant (C), kernel type, and gamma value; Naive Bayes (NB) applied variance smoothing; and Decision Trees (DT) managed maximum tree depth and minimum samples required at node and leaf levels. These adjustments were critical in refining the models to deliver optimal performance metrics.

3.5 Machine Learning Models and Evaluation

This study employed six supervised machine learning models—Random Forest (RF), Logistic Regression (LR), AdaBoost, Support Vector Machines (SVM), Naive Bayes (NB), and Decision Trees (DT)—to predict the survival outcomes of pediatric hematologic transplant patients using the provided dataset features. Given the tabular nature of the dataset, which includes a limited number of instances and lacks image data, these traditional ML models were deemed adequate. The analysis included eight model configurations, as detailed in Table 1, resulting in a total of 48 different machine learning model evaluations. The dataset was partitioned into an 80% training set and a 20% testing set. Python 3.8 and its machine learning libraries, including scikit-learn, pandas, Matplotlib, seaborn, and NumPy, were utilized for this analysis. The models’ performances were assessed using metrics such as accuracy, recall, precision, and F1-score.

Table 1. Model configurations in this study.

Model Configuration	Methodological Enhancements
1	No MIFS, No SMOTE, No Hyperparameter Tuning
2	With MIFS, No SMOTE, No Hyperparameter Tuning
3	No MIFS, With SMOTE, No Hyperparameter Tuning
4	No MIFS, No SMOTE, With Hyperparameter Tuning
5	With MIFS, With SMOTE, No Hyperparameter Tuning
6	With MIFS, No SMOTE, With Hyperparameter Tuning
7	No MIFS, With SMOTE, With Hyperparameter Tuning
8	With MIFS, With SMOTE, With Hyperparameter Tuning

4 RESULTS AND DISCUSSION

The initial or base model configuration utilized the full dataset without any methodological enhancements. It did not involve feature selection, did not address the mild class imbalance, and did not include any hyperparameter tuning. Under these conditions, the RF model achieved the highest performance with a 97.37% accuracy, 93.75% recall, 100% precision, and a 96.77% F1-score, as shown in Table 2. AdaBoost and DT also generated superior prediction performance, both achieving 92.11% accuracy. Both LR and SVM displayed much lower predictive capability while the NB model exhibited very poor performance, with only 60.53% accuracy and a notably

unsatisfactory 21.05% F1-score. These highlight the challenges of managing the dataset without the added methodological enhancements. These results served as baseline values for assessing comparative performance of the methodological enhancements involving feature selection, class imbalance and hyperparameter tuning were implemented with the goal of amplifying predictive capabilities.

Table 2. Performance Metrics of the Base Model Configuration

Model Configuration	Model	Accuracy	Recall	Precision	F1-score
No MIFS No SMOTE No Hyperparameter Tuning	DT	92.11	93.75	88.24	90.91
	LR	76.32	68.75	73.33	70.97
	NB	60.53	12.50	66.67	21.05
	SVM	73.68	56.25	75.00	64.29
	RF	97.37	93.75	100.00	96.77
	AdaBoost	92.11	87.50	93.33	90.32

Fig. 2 details the performance metrics in the second model configuration, which assess the contributory effect of one methodological enhancement, more particularly MIFS. Using MIFS, twenty features were selected based on the mutual information score, including *survival_time*, *PLT_recovery*, *Antigen_b'2'*, *HLA_g_I_b'3'*, *Relapse*, *HLA_g_I_b'2'*, *Txpostrelapse*, *Riskgroup*, *Stemcellsource*, *Disease_b'lymphoma'*, *CD34kgx10d8*, *HLAmatch_b'1'*, *Diseasegroup*, *Recipientage10*, *DonorABO_b'1'*, *IIIV*, *CD34kgx10d6*, *Donorage35*, *DonorCMV*, and *Rbodymass*. The mutual information scores are shown in the treemap in Fig. 3 with the *survival_time* (time of observation (if alive) or time to event (if dead) in days) *PLT_recovery* (time to platelet recovery defined as platelet count >50000/mm3) and *Antigen_b'2'* (in how many antigens there is difference between the donor and the recipient) being classified as the most important attributes. As shown in Fig. 2, the best performing models included LR, RF, and AdaBoost with a 94.74% accuracy. As compared to the baseline model configuration, the implementation of a feature selection using MIFS significantly enhanced the performance of LR across all metrics, more prominently from an accuracy of 76.32% (baseline) to 94.74%. In contrast, RF and DT experienced a modest decline in its performance (97.37% accuracy in baseline down to 94.74% for RF and 92.11% accuracy in baseline down to 89.47% for DT) while AdaBoost showed slight improvement in all its metrics. Generally, there was no significant improvement in the overall predictive performance for both SVM and NB models.

In configuration 3 highlighting the effect of SMOTE in the predictive capability of the models, RF bested all other models with the highest accuracy of 97.37% as illustrated in Fig. 4. This is followed very closely by AdaBoost and DT with a 94.74% and 92.11% accuracy rates, respectively. However, in comparison with the baseline model configuration, there was no improvement in the metrics for RF and DT while AdaBoost exhibited slight improvements in its metrics. Similarly, LR also generated slight increases in its metrics suggesting that LR and AdaBoost benefited with the application of SMOTE in addressing class imbalance. SVM and NB generated much

lower predictive performance in configuration 3 as well as having no significant change in their metrics when compared to the baseline configuration.

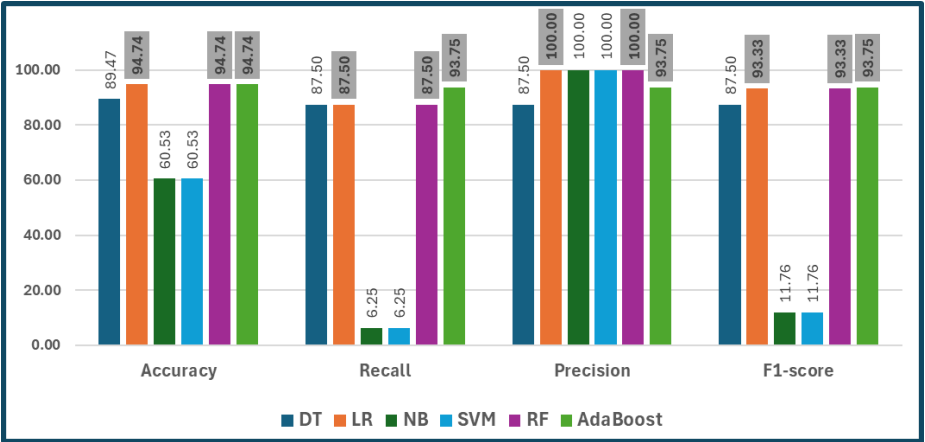


Fig. 2. Performance Metrics of Model Configuration 2 Assessing Mutual Information-based Feature Selection

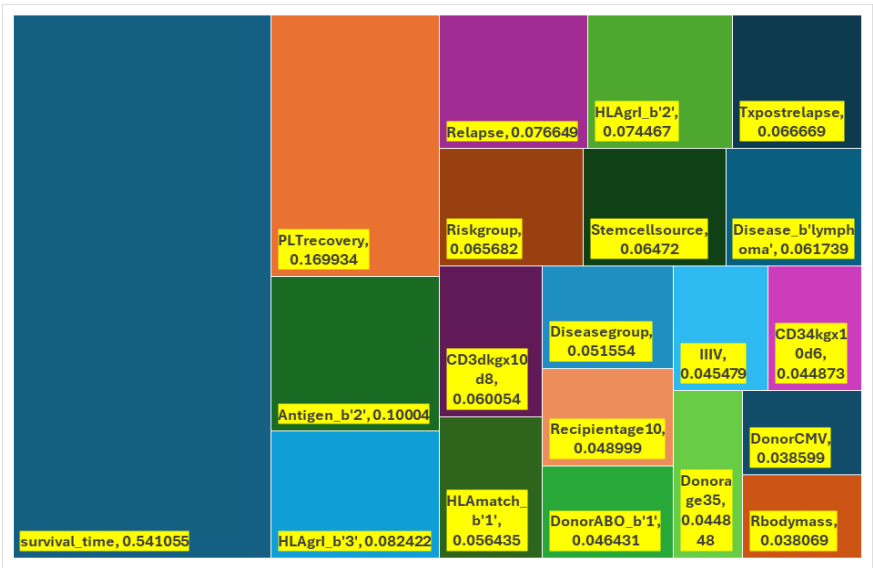


Fig. 3. Top twenty attributes based on Mutual Information Scores

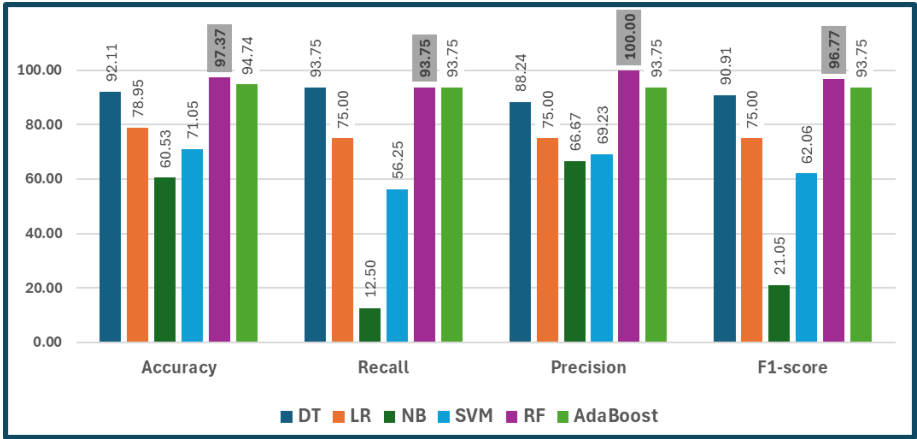


Fig. 4. Performance Metrics of Model Configuration 3 Assessing SMOTE

RF, AdaBoost and LR obtained the highest predictive performance with a 94.74% accuracy rate in configuration 4 (hyperparameter tuning) as seen in Fig. 5. Likewise, DT also performed well with 92.11% accuracy. SVM and NB consistently generated very low predictive performance like the earlier configurations (baseline, 2 and 3). In comparison with the baseline configuration, LR had the most significant improvement in its performance in the 4th configuration with a marked increase in all its metrics, more prominently in its accuracy from 76.32% to 94.74%. On the other hand, AdaBoost had a slight increase while RF had a slight decrease in their accuracy rates. There was no improvement seen for DT, SVM and NB in the 4th configuration. The significant gains for LR highlight the importance of tuning to achieve optimal model performance.

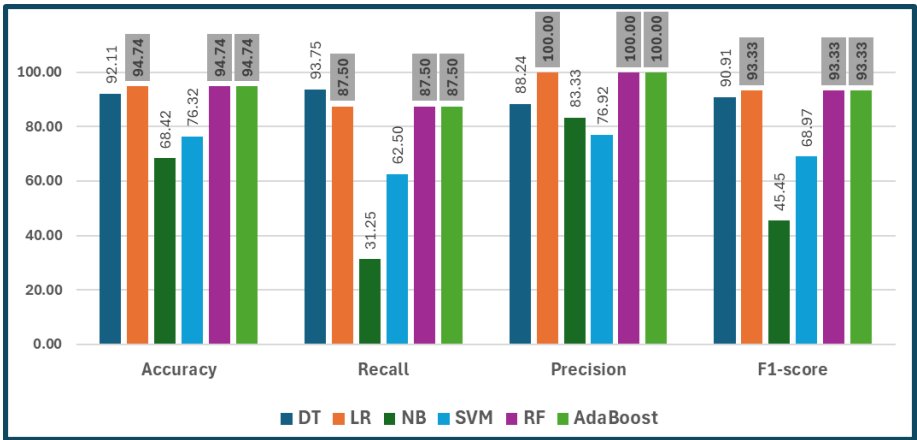


Fig. 5. Performance Metrics of Model Configuration 4 Assessing Hyperparameter Tuning

Fig. 6 presents the metrics for the pairwise combination of methodological enhancements applied to the dataset, more particularly configuration 5 (MIFS and

SMOTE). In this configuration, RF bested all other models in its predictive capability with the highest values across all metrics. Similarly, LR, AdaBoost and DT have good performance across all metrics. SVM yielded lower predictive capacity while NB was underperforming with very low metrics. In comparison with the baseline configuration, noticeably LR had the marked improvement showing significant increases in all metrics. SVM likewise yielded moderate increase in all its metrics when compared with the baseline configuration. In contrast, there was not much upgrade in the metric values for the other models (DT, NB, RF, AdaBoost). As such, both LR and SVM benefitted from both feature selection via mutual information and the balanced dataset provided by SMOTE.

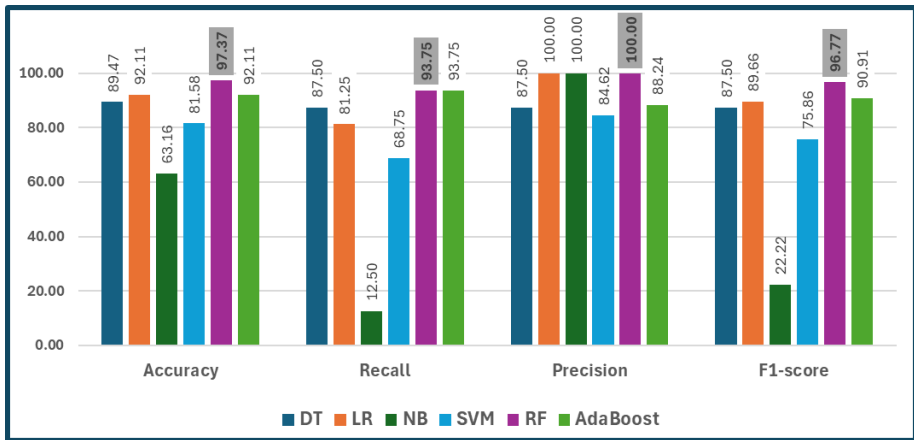


Fig. 6. Performance Metrics of Model Configuration 5 Assessing MIFS + SMOTE

The performance metrics of model configuration 6 assessing MIFS and hyperparameter tuning are presented in Fig. 7. RF and AdaBoost generated the highest predictive ability across all metrics. DT, LR and SVM also yielded good predictive performance while NB had the lowest performance. In comparison with the baseline configuration, a substantial improvement is evident in LR showing a marked increase in all its metrics suggesting significant benefits from feature selection and hyperparameter tuning. Likewise, SVM also had a slight improvement in its metrics. The rest of the models did not show any noticeable difference in their predictive capability.

The predictive performance of model configuration 7 highlighting the impact of SMOTE and hyperparameter tuning is seen in Fig. 8. Again, RF demonstrated its excellent predictive ability as it obtained the highest values across all metrics. This is closely followed by AdaBoost, LR and DT with very good metric values. Consistently, NB was an underperforming model as it generated poor accuracy, recall and F1-score. In comparison with the baseline model configuration, LR, NB, and SVM posted significant gains in their predictive performance. This highlighted the advantageous effects of combining SMOTE with hyperparameter tuning for these models. The rest of the model did not show any noticeable improvements in their performance.

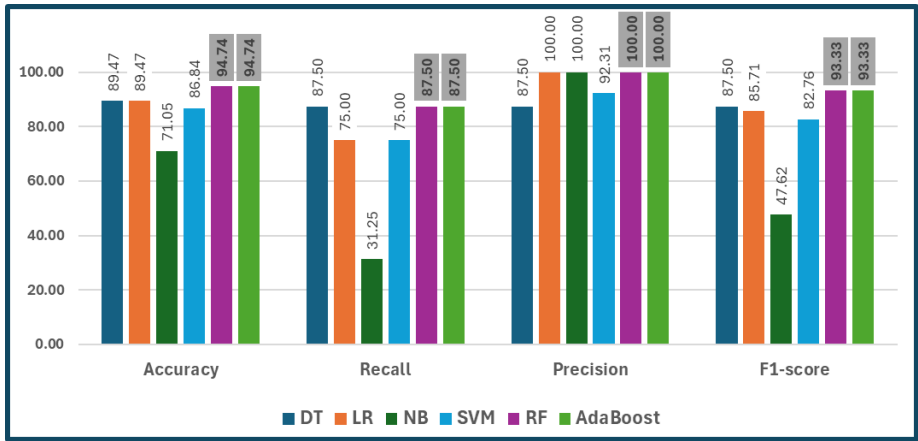


Fig. 7. Performance Metrics of Model Configuration 6 Assessing MIFS + Hyperparameter Tuning

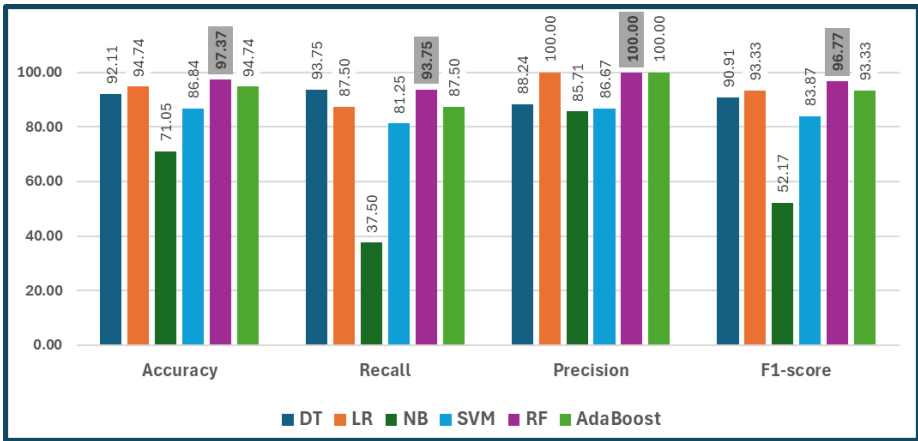


Fig. 8. Performance Metrics of Model Configuration 7 Assessing SMOTE + Hyperparameter Tuning

Fig. 9 details the performance metrics of the 8th model configuration showing all three enhancement methods—MIFS, SMOTE, and hyperparameter tuning. RF maintained its optimum performance like that of the base model configuration. It yielded 97.37% accuracy, 100% precision, 93.75% recall, and a 96.77% F1- score. In this configuration, LR and SVM had a substantial increment in their metric values which strongly suggests the importance of combining methodological enhancements. Likewise, NB posted a modest increase in its performance while DT and AdaBoost did not show any noticeable changes in its predictive capacity.

Given the mild imbalance in the dataset, accuracy continued to be a reliable metric for evaluating how well the models performed. The highest accuracy achieved by RF at 97.37% suggests the model’s effectiveness in identifying both survivors and

non-survivors. The consistent excellent predictive performance of RF in all model configurations evidently shows the suitability of RF in survival prediction of pediatric patients who underwent HSCT. The high recall rate (sensitivity) at 93.75% indicates the model's capability to correctly predict those patients who will survive after HSCT. This shall result in fewer false negatives or fewer missed cases of survivors. Additionally, the excellent precision indicates the ability of the model to make fewer false positive predictions.

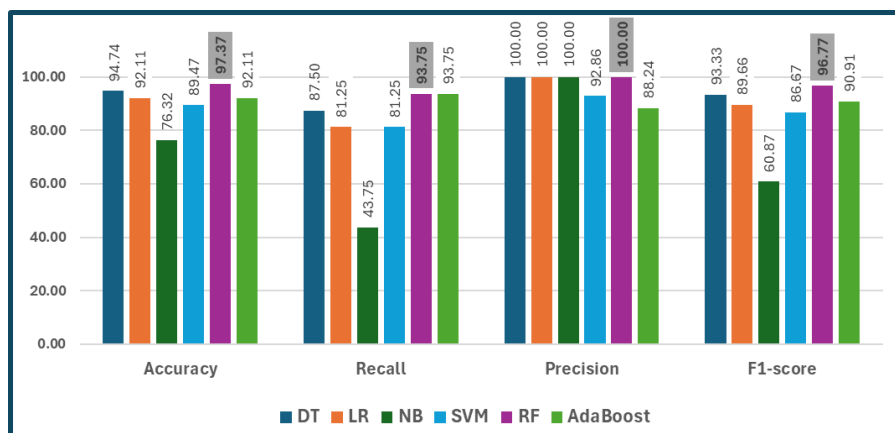


Fig. 9. Performance Metrics of Model Configuration 8 (All Enhancement Methods)

Similar with RF, the performance of AdaBoost and DT remained relatively stable, consistently demonstrating very good predictive metrics. This robustness indicates less sensitivity to changes in feature selection, class balancing, or hyperparameter tuning. On the other hand, LR generally performed better with the addition of methodological enhancements, more particularly with MIFS and hyperparameter tuning, potentially offering greater robustness and interpretability for future applications in predicting the survival of pediatric hematologic transplant patients. NB improved notably with the inclusion of hyperparameter tuning in the model configurations though its overall performance remained much lower compared to other models, specifying its limitations in managing complex datasets even with enhancements. SVM also saw modest increases in the predictive metrics when two or more methodological enhancements were applied. Thus, the study underscores the importance of these methodological enhancements to further improve the predictive performance of various models in pediatric patients who underwent HSCT. The results of this research study represent a significant advancement in leveraging machine learning techniques to upgrade the care for pediatric patients with hematologic malignancies who underwent HSCT. The potential for creating robust and efficient prognostic systems with minimal computation time ultimately contributes to tailored treatment strategies and improved patient outcomes. All of these shall provide the directions for a brighter future for precision medicine in pediatric hematology.

Limitations of this research study included the use of a smaller dataset coming from a single institution. There is a need to externally validate the resulting model in larger datasets with the incorporation of more relevant clinical and laboratory parameters. Furthermore, other machine learning algorithms and feature extraction approaches need to be fully appraised to determine their contributory impact on strengthening prognostic capability of machine learning classifiers.

5 CONCLUSIONS AND RECOMMENDATIONS

Several machine learning classifiers on a publicly available bone marrow transplant in children dataset were investigated in this study. The predictive performance of these models was analyzed through various combinations of methodological enhancements particularly feature selection via mutual information, synthetic minority oversampling technique to address imbalance, and hyperparameter tuning. The top performing prediction model was achieved by random forest which performed robustly across all model configurations with enhancements. The excellent predictive metrics of random forest underscore its reliability and effectiveness in predicting survival outcomes of pediatric hematologic bone marrow transplant patients. The incorporation of methodological enhancement methods suggests a synergistic effect in balancing dataset characteristics while optimizing model parameters through comprehensive feature selection. The integration of hyperparameter tuning as a standard practice is an added advantage to significantly enhance the performance of several models. To ensure the reliability and generalizability of these findings, it would be beneficial to validate these models on larger external datasets. The results of this research study represent a significant advancement in leveraging machine learning techniques to upgrade the care for pediatric patients with hematologic malignancies who underwent bone marrow transplantation. The potential for creating robust and efficient prognostic systems with minimal computation time ultimately contributes to tailored treatment strategies and improved patient outcomes.

References

1. Agarwal, N., Rotz, S., Hanna, R.: Medical emergencies in pediatric blood & marrow transplant and cellular therapies. *Frontiers in Pediatrics* 11 (2023). <https://doi.org/10.3389/fped.2023.1075644>.
2. Fraint, E., Abdel-Azim, H., Bhatt, N., et al.: Evaluation of Children with Malignancies for Blood and Marrow Transplantation: A Report from the ASTCT Committee on Practice Guidelines. *Transplantation and Cellular Therapy* 29(5), 293-301 (2023). <https://doi.org/10.1016/j.jtct.2023.02.003>.
3. Garima, N., Subhashish, P., Vaibhav, C., et al.: Improving outcomes of hematopoietic stem cell transplant for leukemia in children and young adults in resource-limited setting. *Pediatric Hematology Oncology Journal* 9(2), 105-113 (2024). <https://doi.org/10.1016/j.phoj.2024.02.006>.
4. Rotz, S., Bhatt, N., Hamilton, B., et al.: International Recommendations for Screening and Preventative Practices for Long-Term Survivors of Transplantation and Cellular Therapy: A

- 2023 Update. *Transplantation and Cellular Therapy* 30(4), 349-385 (2024). <https://doi.org/10.1016/j.jtct.2023.12.001>.
5. Hong, N., Liu, C., Gao, J., et al.: State of the Art of Machine Learning-Enabled Clinical Decision Support in Intensive Care Units: Literature Review. *JMIR medical informatics* 10(3), e28781 (2022). <https://doi.org/10.2196/28781>.
6. Alawneh, H., Hasasneh, A.: Survival Prediction of Children after Bone Marrow Transplant Using Machine Learning Algorithms. *The International Arab Journal of Information Technology* 21(3), (2024). <https://doi.org/10.34028/iajit/21/3/4>.
7. Chadaga, K., Prabhu, S., Sampathila, N., Chadaga, N.: A machine learning and explainable artificial intelligence approach for predicting the efficacy of hematopoietic stem cell transplant in pediatric patients. *Healthcare Analytics* 3, 100170 (2023). <https://doi.org/10.1016/j.health.2023.100170>.
8. Sapra, V., Sapra, L., Bansal, Y., Chhabra, G., Tanwar, R.: Machine Learning Approach for Identifying Survival of Bone Marrow Transplant Patients. In: Singh, P.K., Kolekar, M.H., Tanwar, S., Wieruchoń, S.T., Bhatnagar, R.K. (eds) *Emerging Technologies for Computing, Communication and Smart Cities. Lecture Notes in Electrical Engineering*, vol 875 (2022). Springer, Singapore. https://doi.org/10.1007/978-981-19-0284-0_3.
9. Ratul, I.J., Wani, U.H., Nishat, M.M., et al.: Survival Prediction of Children Undergoing Hematopoietic Stem Cell Transplantation Using Different Machine Learning Classifiers by Performing Chi-Square Test and Hyperparameter Optimization: A Retrospective Analysis. *Computational and Mathematical Methods in Medicine*, 9391136 (2022). <https://doi.org/10.1155/2022/9391136>.
10. Zhou, Y., Smith, J., Keerthi, D., et al.: Longitudinal clinical data improve survival prediction after hematopoietic cell transplantation using machine learning. *Blood advances*, 8(3), 686–698 (2023). <https://doi.org/10.1182/bloodadvances.2023011752>.
11. von Asmuth, E. G. J., Neven, B., Albert, M. H., et al.: Predicting Patient Death after Allogeneic Stem Cell Transplantation for Inborn Errors Using Machine Learning (PREPAD): A European Society for Blood and Marrow Transplantation Inborn Errors Working Party Study. *Transplantation and cellular therapy*, 29(12), 775.e1–775.e8 (2023). <https://doi.org/10.1016/j.jtct.2023.09.007>.
12. Choi, E.J., Jun, T.J., Park, H.S., et al.: Predicting Long-term Survival After Allogeneic Hematopoietic Cell Transplantation in Patients With Hematologic Malignancies: Machine Learning-Based Model Development and Validation, *JMIR Med Inform* 10(3):e32313 (2022). <https://doi.org/10.2196/32313>.
13. Xiang, J., Shi, M., Fiala, M., et al.: Machine learning-based scoring models to predict hematopoietic stem cell mobilization in allogeneic donors. *Blood Adv* 6 (7): 1991–2000 (2022). <https://doi.org/10.1182/bloodadvances.2021005149>.
14. Shyr, D., Zhang, B.M., Saini, G., Brewer, S.C.: Exploring Pattern of Relapse in Pediatric Patients with Acute Lymphocytic Leukemia and Acute Myeloid Leukemia Undergoing Stem Cell Transplant Using Machine Learning Methods. *Journal of Clinical Medicine* 13(14):4021 (2024). <https://doi.org/10.3390/jcm13144021>.
15. Sikora, M., Wróbel, Ł., Gudyś, A.: Bone marrow transplant: children. *UCI Machine Learning Repository* (2020). <https://doi.org/10.24432/C5NP6Z>.
16. Lall, S., Sinha, D., Ghosh, A., Sengupta, D., Bandyopadhyay, S.: Stable feature selection using copula based mutual information. *Pattern Recognition* 112, 107697 (2021). <https://doi.org/10.1016/j.patcog.2020.107697>.
Raghuwanshi, B.S., Shukla, S.: SMOTE based class-specific extreme learning machine for imbalanced learning. *Knowledge-Based Systems* 187, 104814 (2020). <https://doi.org/10.1016/j.knosys.2019.06.022>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

