Survival Prediction of Children after Bone Marrow Transplant Using Machine Learning Algorithms

Hussam Alawneh
Department of Natural Engineering and Technology Sciences, Arab American University, Palestine
h.alawneh3@student.aaup.edu

Ahmad Hasasneh
Department of Natural Engineering and Technology Sciences, Arab American University, Palestine
Ahmad.Hasasneh@aaup.edu

Abstract: *Bone marrow is the source of many blood-related diseases, such as blood cancers, and Bone Marrow Transplantation (BMT), also known as Hematopoietic Stem Cell Transplantation (HSCT), is a life-saving surgical procedure. However, this treatment is associated with a high risk of mortality. Predicting survival after BMT is therefore essential for effective and accurate treatment. BMT is considered a treatment-related mortality due to several primary causes of death such as infections, toxicity, and Graft-versus-Host Disease (GvHD) that occur after treatment. In addition, several risk factors affect the success of BMT and long-term survival after treatment. Therefore, there is a need for a prediction system based on machine learning techniques that can predict whether the patient will survive after BMT or not, which will definitely help the physicians to make the right decisions before performing the surgery for the patient. In this paper, using a publicly available BMT dataset from the University of California, Irvine ML repository (UCI ML repository), different machine learning models were investigated to predict the survival status of children undergoing BMT treatment. In particular, Random Forest (RF), Bagging Classifier, Extreme Gradient Boost (XGBoost), Adaptive Boosting (AdaBoost), Decision Tree (DT), Gradient Boost (GB), and K-Nearest Neighbors (KNN) were trained on the given dataset. The dataset consists of 45 variables after applying a series of preprocessing steps and removing the multicollinearity features based on the correlation heat map. Then, a feature engineering and modelling step was applied to identify the most significant features, followed by the use of machine learning models to simplify the overall classification process. It's important to note that the most important features obtained by DT and those obtained by GB were the most suitable for training the Bagging classifier and the KNN model, respectively. In addition to that, hyper-parameters optimization using Grid Search Cross-Validation (GSCV) was applied to both approaches to improve the accuracy of the survival prediction. RF, AdaBoost, GB, and Bagging techniques have achieved the best accuracy of 97.37%.*

Keywords: *Bone marrow transplant, hematopoietic stem cell, random forest, bagging, adaboost, gradient boost, decision tree, XGBoost.*

Received December 21, 2023; accepted March 7, 2024
https://doi.org/10.34028/iajit/21/3/4

1. Introduction

Bone marrow is the elastic, delicate, fatty tissue found inside most skeletal structures that is responsible for producing Red Blood Cells (RBCs) and is important for curing some diseases [31, 34, 35]. However, Bone Marrow Transplantation (BMT) is a common treatment for some types of cancer, such as leukaemia and lymphoma, and thalassaemia, among other diseases [39, 40], in which the malfunctioning or damaged bone marrow is replaced with healthy bone marrow [13, 40]. Although BMT is a life-saving treatment, it also carries potentially life-threatening risks [40]. According to the American Cancer Society, leukaemia, also called blood cancer, caused approximately 595,690 deaths in the US in 2016 [32], and approximately 399,967 people will be diagnosed with leukaemia in 2019 [21]. Leukaemia, lymphoma and other diseases such as thalassaemia are typically treated with BMT, which can be a potentially life-threatening risk [40]. More than 50,000 haematopoietic stem cell transplants are performed worldwide each year [16]. Therefore, the need for an automated system to accurately determine the survival of BMT patients makes it a hot topic for research in the medical field. In addition, leukaemia is the most common cancer in Palestinian children, with a rate of 2.6 per 100,000 children [24]. BMT is a life-threatening procedure, and healthcare providers can't predict survival [34]. In addition, the major causes of death in children following BMT remain unclear [34]. Therefore, early prediction of BMT survival will support clinician decision making and guide them towards alternative treatments. As machine learning techniques can automate various tasks in BMT by extracting patterns from data and building predictive models [18], they will be very useful in predicting patient survival after BMT treatment.

In this research, we aim to introduce an intelligent prediction system to assist physicians in predicting patient survival following BMT treatment. The development of this intelligent prediction system involved the investigation of several machine learning techniques, as well as the study and identification of the key features that contribute to improving the accuracy of prediction rates.

The rest of the paper is organized as follows. The related work on survival prediction of children after BMT treatment is discussed and presented in section 2. Section 3 describes the proposed model, including a description of the dataset, feature engineering and modelling, followed by a brief description of the different machine learning algorithms that have been trained and tuned to perform the prediction process. The results obtained from the different machine learning techniques are presented in section 4 and discussed in section 5. Finally, section 6 presents the conclusions and outlines future directions for this research.

2. Literature Review

Machine learning algorithms have been used to build predictive models to automate various tasks in the field of BMT, including identification of the matching donor for an Hematopoietic Stem Cell Transplantation (HSCT) recipient, modelling of Graft-versus-Host Disease (GvHD) risk stratification, and identification of biomarkers for early diagnosis of post-HSCT complications [18]. Its application in BMT is not limited to these tasks, but the use of supervised ML techniques has shown that it is possible to predict the survival status of children after BMT treatment.

Ratul *et al*. [34] proposed a set of machine learning algorithms including Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbors (KNN), Gradient Boost (GB), Adaptive Boosting (AdaBoost) and Extreme Gradient Boost (XGBoost) to predict the survival of children undergoing HSCT. These algorithms were applied to a synthetic dataset obtained from the University of California, Irvine ML repository (UCI ML repository) consisting of 59 features and then reduced to the 11 most important features using the Chi-square feature selection method. In addition, the Grid Search Cross-Validation (GSCV) was used to optimize the hyperparameters and improve the accuracy and was applied to the original and reduced features. The model achieved an accuracy of 94.73% for both the original and reduced features, which required less time and resources to predict child survival after BMT. Using the same dataset, Chadaga *et al*. [7] used machine learning and eXplainable Artificial Intelligence (XAI) techniques to predict survival in children undergoing HSCT. The XAI used to make the models more accurate, interpretable and understandable included Shapley additive scores, local interpretable model agnostic explanations, ELI5 and QLattice. Harris hawks optimization, salp swarm optimization and mutual information feature selection techniques were also used to select the most important features. The best accuracy of 89% was achieved by the custom stacked model. While Sapra *et al*. [35] achieve slightly lower accuracy than Ratul *et al*. [34] by using LR, Multi-Layer Perceptron (MLP), RF and Bayes network to predict the survival of bone marrow transplant patients. These algorithms are applied to the full features and to the reduced features, which are reduced using the correlation feature subset method. The highest accuracy achieved is 94.1% using MLP and Bayes network. An improvement was then provided by Gourisaria *et al*. [15] who used two feature selection methods, namely fuzzy discernibility matrix and principal component analysis, to improve prediction accuracy. The survival status of patients after BMT was predicted using the AdaBoost machine learning classifier. The best result was achieved when AdaBoost was trained on the features selected by the fuzzy discernibility matrix with an accuracy of 95.23%. In addition, AdaBoost with fuzzy discernibility matrix has the shortest training and testing times. Other studies have linked prediction to survival time, such as the study by Choi *et al*. [8], which used a GB, RF, deep neural network, LR and AdaBoost machine learning algorithms to predict 5-year survival after allogeneic HSCT transplantation. This model was applied to data collected from 1470 patients with haematological malignancies who underwent allogeneic HCT at Asan Medical Center, Seoul, South Korea. The model achieved the best mean Area Under the receiver operating characteristic Curve (AUC) of 0.788 using a gradient-boosting machine learning algorithm with 10-fold cross-validation. Shouval *et al*. [36] present the Alternating Decision Tree (ADT) algorithm with 10-fold cross-validation applied to predict all-cause mortality at 100 days after allogeneic HSCT in patients with acute leukaemia. It was applied to a dataset consisting of 29685 patients and 20 features describing recipient, donor and procedure characteristics. The AUC value of the ADT model is 0.702. In addition, the secondary objectives of this study were to estimate Non-Relapse Mortality (NRM), leukaemia-free survival and overall survival at 2 years. Shouval *et al*. [37] delve deeper into NRM by using an in silico approach based on various machine learning algorithms including naive Bayes, ADT, LR, MLP, RF and AdaBoost to predict NRM 100 days after allogeneic HSCT. In this study, the authors developed thousands of models based on different conditions including specific subpopulations, increasing sample size and increasing number of variables. A dataset of 25,923 adults with 22 variables for acute leukaemia patients from the European Society for Blood and Marrow Transplantation (EBMT) registry was analyzed. The maximum predictive AUC achieved using LR and MLP ranged from 0.65-0.67 when using 6 variables. These were selected by ranking all variables based on their contribution to each algorithm, then applying the model to the top 1 variable, followed by increasing the number of variables one at a time.

The study by Gandelman *et al*. [14] doesn’t predict survival status like the above studies, but stratifies survival after HSCT using a DT algorithm that identifies a number of risk groups (low, intermediate and high risk). Based on these groups, the model provides that the high- and intermediate-risk groups had significantly shorter overall survival than those in the low-risk group, with a hazard ratio of 2.79 and 95% confidence interval of 1.58-4.91 for the high- and intermediate-risk groups, and a hazard ratio of 1.78 and 95% confidence interval of 1.06-3.01 for the low-risk group. The data set was collected from 339 patients with incident chronic GvHD. Furthermore, Taati *et al*. [40] used a data mining technique by applying collaborative filtering techniques to process the highly sparse observations with 22.3% missing values and parameter tuning via Bayesian optimization. To evaluate the performance of the proposed model, different classification algorithms (RF, LR and Support Vector Machine (SVM)) with tenfold cross-validation were used and a dataset of 1751 observations and 120 features was collected from Shariati Hospital. The model achieves an AUC of 0.69 in predicting survival status, which is a modest level of accuracy. On the other hand, the model can identify the surgeries with the highest chance of success with high accuracy.

Despite the existence of these related research works, several challenges remain. For example, some related works such as [7, 8, 36, 40] have achieved low accuracies, while others such as [15, 34] have achieved high prediction rates, but they have not used the optimal feature selection method that deals with the multicollinearity problem between the independent features, nor have they used a model-based feature selection that will be more insightful for the model used by considering them as significant. Furthermore, some researchers as Shouval *et al*. [37] have suggested using sophisticated prediction algorithms to improve the prediction rates. In addition, not all existing works have tuned the model parameters, see for example [14, 35]. In this research work, the ultimate goal of the proposed model is to address these challenges by investigating different feature selection methods to identify the optimal and most important features that simplify the prediction process and lead to the best prediction rate. Several Machin Learning (ML) techniques and hyperparameter tuning procedures have been investigated to achieve the best performance.

3. Methods and Materials

Feature selection is an important step in improving prediction tasks by extracting the most significant features. It reduces dimensionality and improves model accuracy and interpretability. Selecting the most relevant features leads to improved accuracy, efficiency and robustness. Therefore, the ultimate goal of this proposed work is to investigate different feature selection methods to identify and extract the most important features. Therefore, as shown in Figure 1, the workflow of the proposed model includes five main phases, including:

- Data exploration.
- Data pre-processing.
- Investigation of machine learning models on the full-feature dataset, followed by hyperparameter tuning for these models.
- Feature selection is based on the important variable plot generated by each model trained on the best parameter set.
- Investigate machine learning models on the reduced features dataset, followed by hyperparameter tuning for these models.

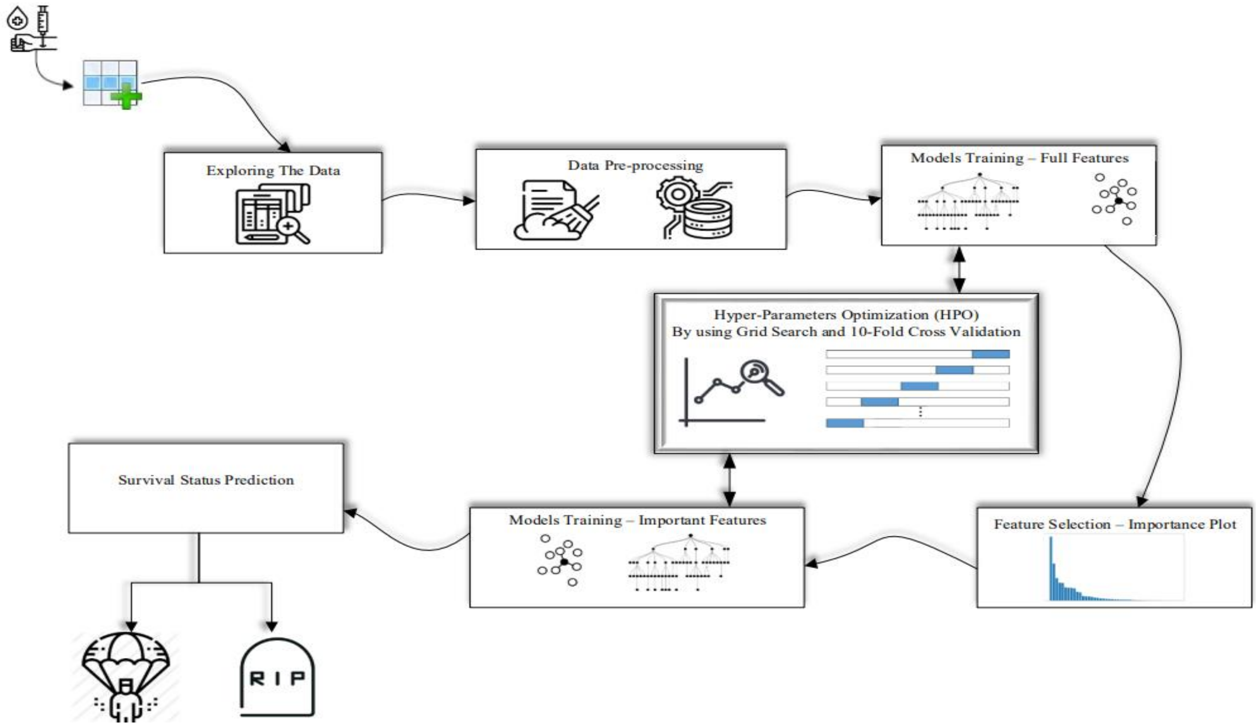

Figure 1. The workflow of the proposed model for the survival prediction of children after bone marrow transplant.

3.1. Dataset Description

The dataset used in this study was retrieved from the UCI ML repository [41]. It consists of 187 observations and 37 attributes describing paediatric patients with a variety of haematological disorders, including malignant and non-malignant cases [34, 41]. All paediatric patients in this dataset undergo unmodified allogeneic unrelated donor haematopoietic stem cell transplantation [34]. The majority of the characteristics are categorical and they also include numeric and Boolean characteristics. A detailed description of the dataset features can also be found in [41].

3.2. Dataset Pre-Processing

The dataset contains many missing values stored as "?", categorical variables such as the recipient_gender feature, and multicollinearity between independent features such as the multicollinearity between hla_match_7/10 and hla_group_1_mismatched features. Therefore, a number of pre-processing steps were performed to address these issues, resulting in a more robust and accurate prediction model. The following pre-processing steps were carried out on the raw data.

- Missing value treatment: missing values were imputed with the mean value for numerical characteristics and the most frequent value for categorical characteristics.
- Categorical variables encoding: dummies were created for each categorical variable and then the extra dummies were removed.
- Multicollinearity: based on the correlation heat map, all independent variables with a correlation greater than 0.7 were removed.
- Data scaling: the standard scaler is used to normalize each input feature before starting to train a model.

After applying these pre-processing steps, the generated dataset consists of 44 features and one target variable. In addition, all problems in the data, such as missing values, categories and multicollinearity, have been dealt with, so the data is now clean and ready to be fed into any machine learning model.

The following correlation heatmap in Figure 2 shows the correlation between the independent and dependent variables, and there's no collinearity between the variables.

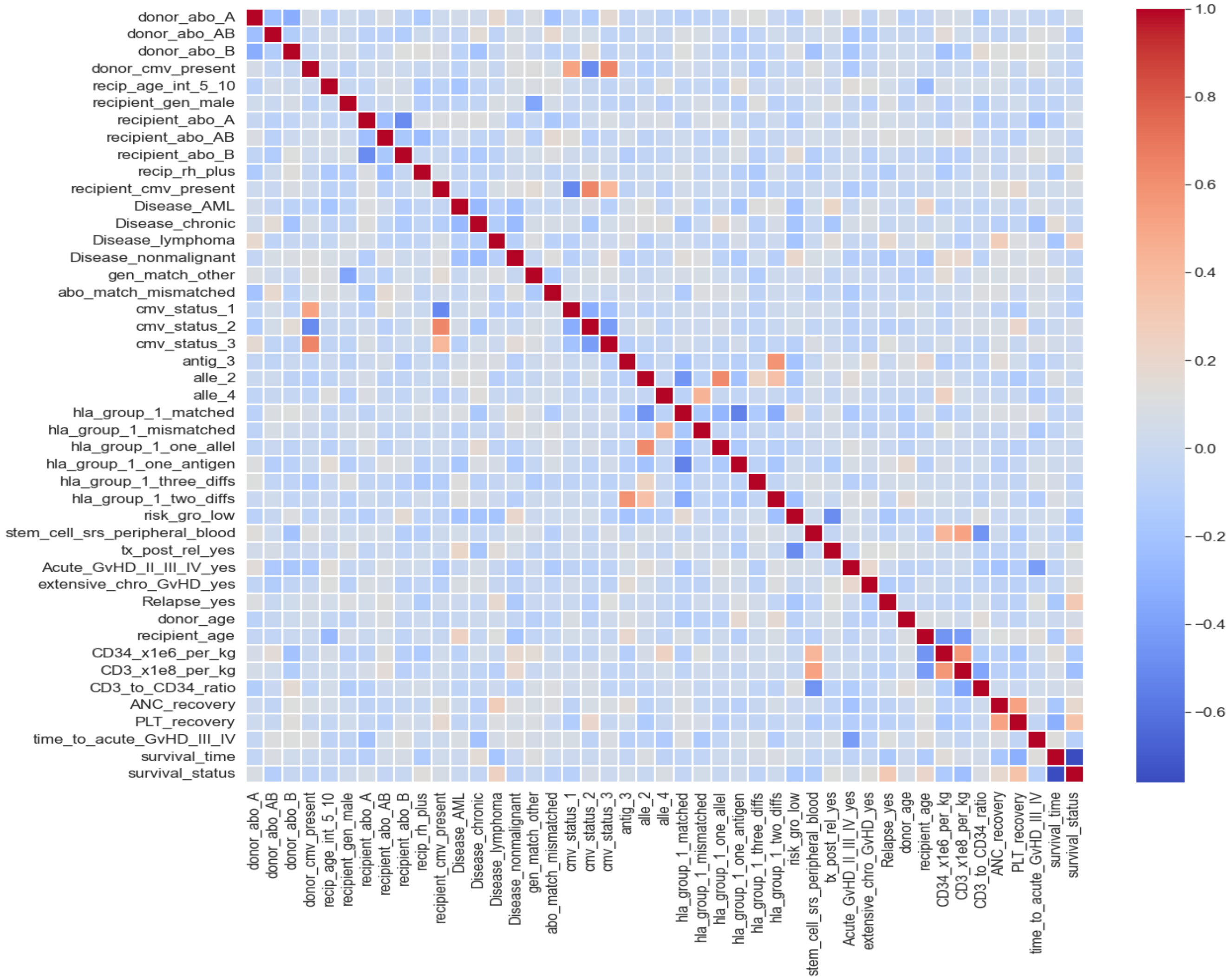

Figure 2. Correlation heat map.

3.3. Feature Extraction and Modelling

Identifying and extracting the most significant features that contribute more to improving the linear separation in subsequent classification or prediction tasks is the main objective of this research. More specifically, in this research work, several feature selection methods have been used by training each model on the full feature set and then recursively eliminating the features with lower scores until the best set of features that obtained the best score for the evaluation metrics is obtained. In addition, the features extracted by one model are used to improve the same model in the next experiments, as the model considers these features as insights for itself.

Each model uses a different way of calculating the importance of the features, The RF depends on the mean decrease in impurity (mean decrease in Gini) for node splitting by finding the total decrease in impurity of the split overall DT [17] to determine the contribution of each feature in constructing the RF model [33]. The features with higher mean decrease impurity factor value represent higher importance than other features [20]. Similar to RF, the mean decrease in impurity for the DT is calculated [25] for a single DT. In contrast, XGBoost has three different feature importance scores which are weight, gain and cover [42]. The weight score, which is used in this research, represents the number of times a feature is used for DT splitting, while the gain, which is calculated by averaging the error reduction of a feature used for DT splitting, and the cover, which is the average number of samples affected by the splitting [42]. In AdaBoost, the importance of a particular feature is found by calculating the total reduction in the criterion brought about by that feature [38]. Specifically, the more frequent feature used in the weak learners to correctly classify the samples will have a higher importance, while the feature importance in GB is calculated by summing the improvement in the loss function due to splitting on that feature over all trees, and the features that lead to larger reductions in the loss function will be more important.

3.4. Machine Learning Prediction and Tuning

In this paper, seven ML models were experimented with, including KNN, RF, Bagging, GB, XGBoost, AdaBoost and DT. The tree-based models can handle numerical and categorical variables [22], which makes them suitable for the heterogeneous BMT dataset. More specifically, DT classifiers can capture complex interactions between features [11], while RF and Bagging are an ensemble of DTs, which reduces overfitting from these DTs [6, 27] and improves prediction accuracy.GB, XGBoost and AdaBoost are aimed at improving the weak learner to produce a strong classifier [5]. In addition, they have a good ability to handle complex relationships between features, which can provide accurate predictions of patient survival status after BMT. On the other hand, K-NN is effective for small data [29], which makes it suitable for the task of predicting the survival status of children undergoing BMT.

- **RF**: the RF is a tree-based ML model of an ensemble of DTs to minimise variance by averaging the DTs. RF uses bootstrapping for sampling by selecting a subset of observations and features. Feature selection is done using the rule of thumb [11]. A DT is then run on each sample and the Gini index is performed to determine the final class for each tree [11] as in Equation (1).

$$Gini = 1 - \sum_{i} {p_i}^2 \tag{1}$$

- **Bagging**: bagging is a tree-based ML model that uses the bootstrap aggregation method to reduce variance by aggregating the results of repeated sampling with replacement (unlike RF sampling, which is done only for the observations and selects all features) [11]. The aggregation technique described in Equation (2) is used to compute the mode for classification and to average the predictions for regression tasks to obtain a model with low variance and bias.

$$\widehat{f_{avg}}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(x) \tag{2}$$

- **GB**: GB [11, 12] is an ensemble boosting classification algorithm that iteratively combines weak learners to perform a strong classifier. It does this by focusing on the problematic observations and correcting them in the next iteration. The gradient descent optimisation procedure is used to minimise the loss as it depends on the loss function.
- **XGBoost**: XGBoost [4] is a highly scalable DT ensemble based on gradient boosting. It minimises a loss function tailored for DTs and uses the hyperparameters including multiple leaves and regularisation to control the complexity of the tree. To reduce overfitting and increase the speed of the training phase, randomisation techniques have been implemented in XGBoost, including subsampling and column subsampling. XGBoost handles sparsity in the data and incorporates interesting features such as monotonic and feature interaction constraints. Furthermore, the training speed was increased by implementing several optimised split-finding methods, such as a compressed column-based structure and selective evaluation of candidate splits using percentiles.
- **AdaBoost**: AdaBoost [28] is considered to be the first successful boosting algorithm. It is one of the boosting algorithms that combine multiple week learners to produce a strong classifier. Similar to GB, AdaBoost trains the base classifier iteratively, but unlike GB, it adjusts the weights of misclassified samples at each iteration. AdaBoost gives more attention to difficult instances and improves its ability

to handle complex classification problems by assigning higher weights to misclassified samples and lower weights to correctly classified samples.

- **DT**: DT [10, 11] is one of the most widely used supervised learning algorithms, which recursively splits the data based on set impurity criteria to create a tree-like model. The shape of the tree gives the name of the DT for this method, where the leaves represent the class names and the branches represent the features. DTs are easy to understand, visualise and interpret. They can handle both numerical and categorical features, making them suitable for a wide range of classification tasks. Disadvantages of DTs include the problem of overfitting due to high variance and the difficulty of classifying multiple output classes.
- **KNN**: KNN [9, 11] is a non-parametric machine learning algorithm that classifies instances based on the class of their nearest neighbours. Given a new instance, KNN determines the K nearest neighbours in the training data set and assigns the majority class label among these neighbours to the new instance. KNN is also called memory-based classification because all training data is in memory at runtime.

Getting these ML algorithms to perform their best hyperparameter optimisation is a critical step [2], since these parameters define the architecture of ML models and fine-tuning them will have a critical impact on the performance of an ideal model [34]. Grid search is the most popular method for learning the hyperparameter configuration by extensively exploring all combinations of hyperparameters fed into the grid configuration [3]. For this task, the GridSearchCV method is used in addition to grid search to perform cross-validation by training the model and optimising the hyperparameters using the validation data [34]. By performing hyperparameter optimisation with GSCV, they systematically explore different hyperparameter combinations and select those that perform best for predicting survival after BMT. This process helps to ensure that the generated model is well tuned and capable of making accurate predictions in a real-world setting.

After the training phase and to evaluate the ability of the proposed model to predict the survival of children after BMT, several evaluation measures were considered. Accuracy is the most widely used metric across all systems and is the ratio of correctly predicted instances to the total number of observations [23]. Accuracy is a useful metric for the task of model evaluation, but it may not provide a complete picture of model performance [30]. Therefore, other measures such as Area Under the Receiver Operating Characteristic Curve (AUC-ROC), precision, sensitivity, F1 score and confusion matrix [43] are proposed to evaluate the model performance more accurately. Table 1 presents mathematical equations for some of these evaluation measures used in this study, based on the number of False Positives (FP), True Positives (TP), False Negatives (FN), and True Negatives (TN), [26] which are commonly used to evaluate the performance of classification models.

Table 1. Evaluation measures.

Evaluation measure	Formula
Accuracy	(TP+TN)/(TP+TN + FP+FN)
Precision	TP/(TP+FP)
Sensitivity	TP/(TP+FN)
F1-score	2×(precision×recall)/(precision+recall)

4. Results

4.1. Features Extraction and Selection

After fitting the RF, AdaBoost, DT, GB and XGBoost with the best set of hyperparameters, an important variable plot is generated for each model. The Bagging classifier was applied to the features extracted from the DT important variable plot, and the KNN was applied to the features extracted from the GB, as they achieve the best accuracy based on these features. As shown in Table 2, the number of most important features is 10 for RF and DT, while it is 11 for AdaBoost, 9 for XGBoost, and 6 for GB, indicating that the number of important features varies depending on the model.

Table 2. Important features based on importance plot.

AdaBoost	RF	DT and Bagging based on DT	XGBoost	GB and KNN based on GB
survival_time: 0.2399	**survival_time**: 0.312	**survival_time**: 0.7959	**survival_time**: 359.0	**survival_time**: 0.7939
CD34_x1e6_per_kg: 0.139	PLT_recovery:0.1167	**Relapse_yes**: 0.0872	**donor_age**: 325.0	**Relapse_yes**: 0.0836
CD3_x1e8_per_kg: 0.0843	**Relapse_yes**: 0.0747	**donor_age** : 0.0342	**CD34_x1e6_per_kg**: 258	**CD34_x1e6_per_kg**: 0.0358
donor_age: 0.0675	CD3_x1e8_per_kg: 0.0731	recipient_abo_B: 0.0339	CD3_x1e8_per_kg: 255	recip_rh_plus: 0.0279
recipient_age: 0.0666	**CD34_x1e6_per_kg**: 0.0695	**CD34_x1e6_per_kg**: 0.0226	**recipient_age**: 155	risk_gro_low: 0.0168
ANC_recovery: 0.0495	ANC_recovery: 0.0527	recipient_gen_male: 0.0113	ANC_recovery: 110	ANC_recovery: 0.0144
PLT_recovery: 0.0486	**recipient_age**: 0.0516	**recipient_age**: 0.0086	Acute_GvHD_2_3_4_yes: 106	
cmv_status_1: 0.048	CD3_to_CD34_ratio: 0.0494	extensive_chro_GvHD_yes: 0.0063	risk_gro_low: 96	
Relapse_yes: 0.0458	**donor_age**: 0.0289	recipient_cmv_present: 0.0	**Relapse_yes**: 89	
CD3_to_CD34_ratio: 0.0332	Disease_lymphoma: 0.0248	abo_match_mismatched: 0.0		
recipient_cmv_present: 0.0305				

As shown in Table 2, the important features of these models have some similarities, five features are important for AdaBoost, RF, DT, Bagging and XGBoost which are (survival_time, CD34_x1e6_per_kg, donor_age, recipient_age and relapse_yes). While GB and KNN models share them with the first three features. It can also be observed that survival time is the most important feature for all machine learning models.

Table 3. RF's most important 20 features.

Feature number	Feature name
43	survival_time: 0.312
41	PLT_recovery: 0.1167
34	Relapse_yes: 0.0747
38	CD3_x1e8_per_kg: 0.0731
37	CD34_x1e6_per_kg: 0.0695
40	ANC_recovery: 0.0527
36	recipient_age: 0.0516
39	CD3_to_CD34_ratio: 0.0494
35	donor_age: 0.0289
13	Disease_lymphoma: 0.0248
42	time_to_acute_GvHD_III_IV: 0.0166
1	donor_abo_AB: 0.0141
10	recipient_cmv_present: 0.0115
31	tx_post_rel_yes: 0.0102
17	cmv_status_1: 0.01
9	recip_rh_plus: 0.0083
30	stem_cell_srs_peripheral_blood: 0.0072
29	risk_gro_low: 0.007
18	cmv_status_2: 0.0059
12	Disease_chronic: 0.0055
33	extensive_chro_GvHD_yes: 0.0047
25	hla_group_1_one_allel: 0.0041
28	hla_group_1_two_diffs: 0.004

Since RF achieves stable results in all experiments, whereas Bagging, AdaBoost, DT, XGBoost, KNN and GB results were not constant in all experiments. Therefore, Table 3 shows the feature number and name for the top 20 features in descending order based on the RF importance plot.

4.2. Experiment 1: Survival Prediction with all Features and Default Hyperparameters

In this experiment, RF, XGBoost, Bagging, AdaBoost, DT, GB and KNN were applied to a full feature dataset without removing any features. Table 4 shows the classification report for each model, and Table 5 shows a comparison between the accuracy and the area under the ROC curve for each model in this study and from the study in [34].

Table 4. Experiment 1 classification reports.

Model	Classification report			
	Target	Precision	Recall	F1-Score
RF	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
Bagging	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
XGBoost	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
AdaBoost	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
GB	0	0.92	1	0.96
	1	1	0.88	0.93
DT	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
KNN	0	0.68	0.86	0.76
	1	0.70	0.44	0.54

Table 5. Models accuracy.

Model	This study		Study in [34]	
	Accuracy	ROC_AUC	Accuracy	ROC_AUC
RF	**0.9737**	**0.9688**	0.9210	0.9229
Bagging	0.9474	0.9460	-	-
XGBoost	0.9474	0.9460	0.8947	0.8991
AdaBoost	**0.9737**	**0.9688**	**0.9473**	**0.9467**
GB	0.9474	0.9375	**0.9473**	**0. 9467**
DT	0.9474	0.9460	**0.9473**	**0.9523**
KNN	0.6842	0.6506	0.6052	0.5756

RF and AdaBoost have achieved the best accuracy of 97.37% and area under the curve of 96.88%, as shown in Table 5 and Figure 3, outperforming the results in [34] and other models. This indicates the suitability of RF and AdaBoost for the categorical data. In addition, XGBoost, GB, DT and KNN have also provided accuracy improvements over the same models in [34], giving further evidence of the importance of preprocessing in model improvement. The Receiver Operating Characteristics (ROC) curve for each model is shown in Figure 3.

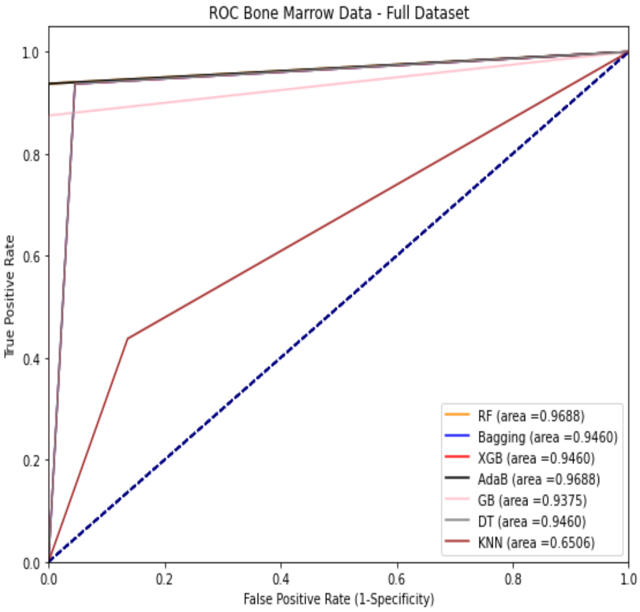

Figure 3. ROC curve -all dataset features.

4.3. Experiment 2: Survival Prediction with all Features and Hyperparameters Optimization

In this experiment, RF, XGBoost, Bagging, AdaBoost, DT, GB, KNN with Grid Search and 10-fold cross-validation were applied to the processed dataset to optimize the hyper-parameters and improve the model. Table 6 shows the classification report for each model and Table 7 shows the accuracy and area under the ROC curve for each model compared to the results of the study in [34].

After optimizing the hyperparameters, RF, Bagging and GB achieve the best accuracy and area under the curve as shown in Table 7 and Figure 4. The ROC curve for all models with GSCV is shown in Figure 4.

Table 6. Experiment 2 classification reports.

Model	**Classification report**			
	Target	**Precision**	**Recall**	**F1-Score**
RF	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
Bagging	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
XGBoost	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
AdaBoost	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
GB	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
DT	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
KNN	0	0.69	0.91	0.78
	1	0.78	0.44	0.56

Table 7. Models accuracy with GSCV.

Model	**This study**		**Study in [34]**	
	Accuracy	**ROC_AUC**	**Accuracy**	**ROC_AUC**
RF	**0.9737**	**0.9688**	0.9210	0.9229
Bagging	**0.9737**	**0.9688**	-	-
XGBoost	0.9474	0.9460	0.9210	0.9229
AdaBoost	0.9474	0.946	**0.9473**	**0.9467**
GB	**0.9737**	**0.9688**	**0.9473**	**0.9467**
DT	0.9474	0.946	**0.9473**	**0.9411**
KNN	0.7105	0.6733	0.6842	0.6638

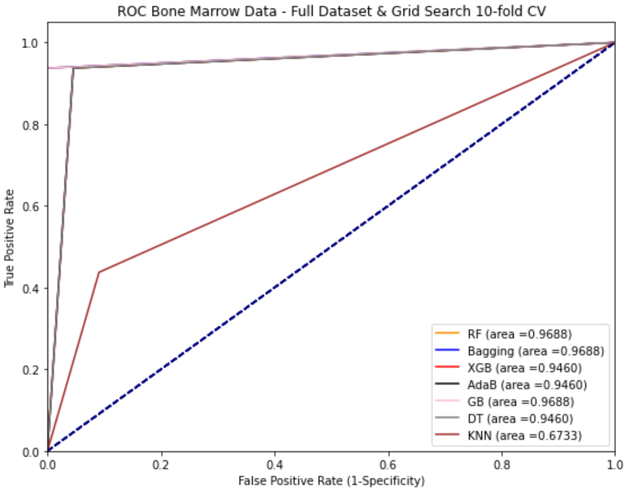

Figure 4. ROC curve - full features and GSCV.

RF, Bagging and GB have achieved the best accuracy of 97.37% and area under the curve of 96.88%, as shown in Table 7, outperforming the results in [34] and other models in this study. After hyper-parameter optimization, the accuracy of Bagging and GB is increased, which indicates the role of hyper-parameter optimization in building robust models. In addition, KNN has also provided accuracy improvements over the same models in [34] and in Experiment 1, further demonstrating the importance of fine tuning in model improvement.

4.4. Experiment 3: Survival Prediction with a Reduced-Features Dataset Based on Important Variable Plot

In this experiment, RF, XGBoost, Bagging, AdaBoost, DT, GB, and KNN were applied to the processed dataset after eliminating the insignificant features for each model. Table 8 shows the classification report for each model, and Table 9 shows the accuracy and area under the ROC curve for each model compared to the results of the study in [34].

After eliminating the insignificant features according to the importance plot the RF, Bagging classifier, and AdaBoost achieve the best accuracy and area under the curve as shown in Table 9 and Figure 5 compared to the results of the study in [34]. The ROC curve for all seven models after keeping the most important features for each model is shown in Figure 5.

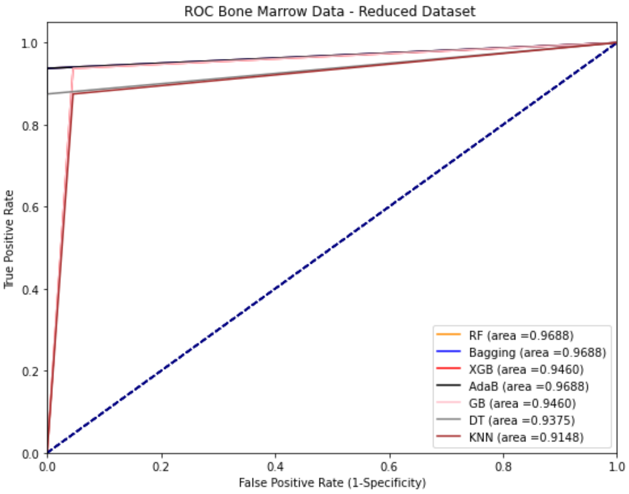

Figure 5. ROC curve-significant feature.

Table 8. Experiment 3 classification reports.

Model	**Classification report**			
	Target	**Precision**	**Recall**	**F1-Score**
RF	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
Bagging	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
XGBoost	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
AdaBoost	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
GB	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
DT	0	0.92	1	0.96
	1	1	0.88	0.93
KNN	0	0.91	0.95	0.93
	1	0.93	0.88	0.90

Table 9. Models accuracy-significant features.

Model	**This study**		**Study in** [34]	
	Accuracy	**ROC_AUC**	**Accuracy**	**ROC_AUC**
RF	**0.9737**	**0.9688**	0.8157	0.8053
Bagging	**0.9737**	**0.9688**	-	-
XGBoost	0.9474	0.9460	0.8157	0.8053
AdaBoost	**0.9737**	**0.9688**	0.7894	0.7815
GB	0.9474	0.946	0.8157	0.8053
DT	0.9474	0.9375	0.8157	0.8165
KNN	0.9211	0.9148	**0.9210**	**0.9229**

RF, Bagging and AdaBoost have achieved the best accuracy of 97.37% and area under the curve of 96.88% as shown in Table 9, outperforming the results in [34] and other models in this study. After maintaining the significant features, the accuracy of AdaBoost and KNN is increased, while RF, Bagging, XGBoost and GB have maintained their accuracy level. This indicates the role of dimension reduction by keeping the significant features in building robust models.

4.5. Experiment 4: Survival Prediction with a Reduced-Features Dataset and Hyperparameter Optimization

In this experiment, grid search and 10-fold cross-validation with RF, XGBoost, Bagging, AdaBoost, DT, GB and KNN were applied to the processed dataset after eliminating the insignificant features for each model. Table 10 shows the classification report for each model.

Table 10. Experiment 4 classification reports.

Model	Classification report			
	Target	Precision	Recall	F1-Score
RF	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
Bagging	**0**	**0.96**	**1**	**0.98**
	1	**1**	**0.94**	**0.97**
XGBoost	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
AdaBoost	0	0.92	1	0.96
	1	1	0.88	0.93
GB	0	0.95	0.95	0.95
	1	0.94	0.94	0.94
DT	0	0.92	1	0.96
	1	1	0.88	0.93
KNN	0	0.91	0.91	0.91
	1	0.88	0.88	0.88

Table 11 shows the accuracy and the area under the ROC curve for each model after applying it to the most important features and optimizing the hyperparameters, compared to the results of the study in [34]. The ROC curve for all models with GSCV after keeping the most important features for each model is shown in Figure 6. RF and Bagging have the best results as shown in Table 11 and Figure 6. Moreover, Table 12 summarises the optimal value of the hyperparameters for each model when applying the model to the full and reduced feature datasets.

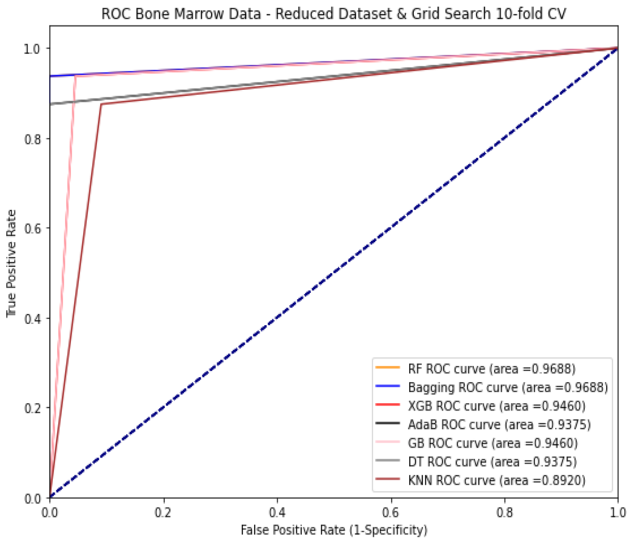

Figure 6. ROC-significant features and GSCV.

Table 11. Models accuracy-significant features and GSCV.

Model	This study		Study in [34]	
	Accuracy	ROC_AUC	Accuracy	ROC_AUC
RF	**0.9737**	**0.9688**	0.8421	0.8347
Bagging	**0.9737**	**0.9688**	-	-
XGBoost	0.9474	0.9460	0.8157	0.8053
AdaBoost	0.9474	0.9375	0.8157	0.8053
GB	0.9474	0.946	0.8157	0.8053
DT	0.9474	0.9375	**0.9473**	**0.9467**
KNN	0.8947	0.892	0.8947	0.8879

Table 12. Hyper parameters optimal value.

ML Classifier	Parameter	Full-features dataset	Reduced-features dataset
RF	Number of estimators	250	25
	Min samples split	10	50
	Min samples leaf	1	2
	Max depth	3	5
Bagging	Number of estimators	200	100
	Max samples	0.25	0.3
	Max features	0.5	0.5
	Bootstrap features	True	False
	Bootstrap	True	False
	Base estimator max depth	3	3
XGBoost	Number of estimators	1500	1000
	Max depth	2	3
	Learning rate	0.05	0.05
AdaBoost	Number of estimators	50	50
	Learning rate	0.1	0.1
GB	Number of estimators	1000	200
	Min samples split	2	2
	Min samples leaf	2	1
	Max depth	5	5
	Learning rate	0.1	0.1
DT	Max features	0.5	default value (non)
	Min samples split	2	2
	Min samples leaf	2	1
	Max depth	15	1
	Criterion	Gini	Gini
	Splitter	Best	Best
KNN	Number of neighbors	7	7
	Weights	uniform	Distance
	algorithm	ball_tree	ball_tree

The importance of training each model on the appropriate features for it is clearly shown in Table 11. As the models in this study maintain their accuracy score, while the accuracy score in [34] decreased for the same models. All models in [34] were trained on significant features extracted using the Chi-square test, which may not be appropriate for them.

5. Discussion

In this paper, seven machine learning algorithms were proposed for predicting the survival status of children undergoing BMT during 4 experiments for each model. These experiments started by applying each model to all features, then finding the most important features based on the best parameter set and reapplying the models to the reduced features dataset after using GSCV with 10-fold for hyperparameter optimisation for model improvement when applied to the full features and important features. The RF, AdaBoost, GB and Bagging results were the best with an accuracy of 97.37% and the RF results were similar and stable throughout the experiments.

The RF achieves an accuracy of 97.37% and precision (0.98), recall (0.97), f1-score (0.97) and ROC_AUC (0.97) for all experiments and Bagging achieves the same results in the second, third and fourth experiments, AdaBoost achieves the same results in the first and third experiments when using all features and the 11 most important features and GB achieves the same results only in the second experiment when using all features with the GSCV. While XGBoost, DT and KNN achieve a lower score for accuracy and AUC with the best score for accuracy being 94.74% for XGBoost and DT and 92.11% for KNN. The stability of the results of the RF before and after the elimination of the insignificant features shows the suitability of the RF in the task of survival prediction and its ability to extract the most important features for this type of data.

The interpretation of the high accuracy obtained is that the developed models are applied to good pre-processed data, using a scientific method for the treatment of the missing value based on [34]. What is more, the stability and improvement of the results after the model-based feature extraction step by using the important variable plot generated from the trained model with the best set of parameters, feature extraction made by selecting the most important 15 features and removing feature by feature until the best results are obtained with the lowest number of features. For the experiments with Hyperparameter Optimization (HPO), the results in Tables 7 and 11 show the improvement of the results for the GB and KNN models and the saving of the same results for the other models after using the significant features. Fine-tuning for each model also plays an important role in improving the results, which is clear from the results of Bagging, GB and KNN in Table 5, which are improved in Table 7 after tuning the model parameters.

The RF model produces a high AUC across all trials, which gives us an indication of the high performance of this model. We can see that the curve rises dramatically near the y-axis (top left), which means that the true positive rate is increasing and the false positive rate is almost zero, reflecting the high accuracy.

The Bagging classifier and AdaBoost achieve their highest scores when using the DT as the base estimator with the best set of parameters generated from the GSCV. Also, the most significant variables for the Bagging model were extracted based on the importance variable plot of the DT, as it obtained its highest score based on them.

The previous studies developed different ML models for the task of predicting the survival of patients who received three BMTs in 2022. The study in [34] used the same dataset for this paper and applied different ML on the full-features dataset of 59 features after pre-processing and the reduced-features dataset with HPO to achieve an accuracy of 94.73%. The main difference is that the study in [34] used the full-features dataset with a multicollinearity between the independent variables and an ambiguous step of selecting the features of the full dataset without determining the selection strategies and then using the Chi-square test to determine the most significant features, while this study focused on treating the multicollinearity by removing the highly correlated independent variables based on a heat map to generate a dataset of 44 features. And the use of model-based feature extraction, which is done by using the important variable plot for 7 machine learning models and then selecting the most important features. In addition, fine-tuning for each model by selecting the best set of parameters based on the pre-processed dataset to achieve an accuracy of 97.37% on reduced and full feature datasets.

The models in this study achieve a real improvement in accuracy for predicting survival in children post-HSCT, increasing the prediction accuracy for all models and across all experiments over the results achieved by [34]. Table 13 shows the difference between the accuracy achieved by the models in this study and the results achieved by [34].

Table 13. The accuracy improvements for all models in this study over the study in [34].

Model	**Experiment 1**	**Experiment 2**	**Experiment 3**	**Experiment 4**
RF	5.27%	5.27%	15.80%	13.16%
Bagging	-	-	-	-
XGBoost	5.27%	2.64%	13.17%	13.17%
AdaBoost	2.64%	0.01%	18.43%	13.17%
GB	0.01%	2.64%	13.17%	13.17%
DT	0.01%	0.01%	13.17%	0.01%
KNN	7.90%	2.63%	0.01%	0.00%

The RF model developed in this study used the most important features from the RF importance plot and it achieves the best result when using the most important 10 features, while the study in [34] used the most important 11 features based on the Chi-Square test. Table 14 shows a comparison between the important features, it is clear that out of the 10 features, there are 9 features shared between the two studies with a difference in the degree of importance.

Table 14. Features comparison.

RF	**The Chi-square test for the study in [34]**
survival_time	**PLT_recovery**
PLT_recovery	**ANC_recovery**
Relapse_yes	time_to_acute_GvHD_III_IV
CD3_x1e8_per_kg	**survival_time**
CD34_x1e6_per_kg	recipient_body_mass
ANC_recovery	**CD34_x1e6_per_kg**
recipient_age	**CD3_x1e8_per_kg**
CD3_to_CD34_ratio	**CD3_to_CD34_ratio**
donor_age	**recipient_age**
Disease_lymphoma	**relapse_yes**
RF	**disease_lymphoma**

6. Conclusions

BMT is a treatment for many types of cancer, and at the same time there is a risk of patients dying after the treatment [34, 35]. Therefore, the aim of this work was to use seven ML models to predict the outcome of this treatment before it is performed. This will provide the medical sector with real insight into the outcome of treatment and guide them in finding alternatives when BMT has a high risk. We also aimed to find the best set of features for each model by using the importance plot as a feature selection technique. It shows the most important features in determining the BMT outcome from the view of each trained model on the full feature dataset, with five features shared by all models except GB and KNN, which share only three features with other models.

This work shows that the DT-based ML models are suitable for this classification problem, with a slight advantage for the RF. Also, the selection of the most important features has a real contribution to improve the performance of the models, as shown by obtaining the same accuracy after eliminating insignificant and multicollinearity and keeping the most important features out of 59 features.

This study makes a real contribution to improving the prediction performance of the survival status of children after BMT by using feature selection and fine hyperparameter tuning methods. This model achieves an accuracy of 97.37%, which is higher than the accuracy achieved by Ratul *et al*. [34].

Since the strategy used in preprocessing the data by handling missing values and multicollinearity, and using the importance plot for feature selection shows real improvement in the survival status prediction task, our next step is to apply this model to a larger dataset with a larger number of observations. We will also attempt to interpret the model predictions using explicable AI techniques. In addition, new deep learning models and explicable artificial intelligence techniques [1, 19, 43] will be further explored in this research.

References

[1] Al-Fayoumi M., Abu Al-Haija Q., Armoush R., and Amareen C., “XAI-PDF: A Robust Framework for Malicious PDF Detection Leveraging SHAP-Based Feature Engineering,” *The International Arab Journal of Information Technology*, vol. 21, no. 1, pp. 128-146, 2024. https://doi.org/10.34028/iajit/21/1/12

[2] Anelli V., Di Noia T., Di Sciascio E., Pomo C., and Ragone A., “On the Discriminative Power of Hyper-Parameters in Cross-Validation and how to Choose them,” *in Proceedings of the 13th ACM Conference on Recommender Systems*, Copenhagen, pp. 447-451, 2019. https://doi.org/10.1145/3298689.3347010

[3] Belete D. and Huchaiah M., “Grid Search in Hyperparameter Optimization of Machine Learning Models for Prediction of HIV/AIDS Test Results,” *International Journal of Computers and Applications*, vol. 44, no. 9, pp. 875-886, 2021. https://doi.org/10.1080/1206212X.2021.1974663

[4] Bentéjac C., Csörgő A., and Martínez-Muñoz G., “A Comparative Analysis of Gradient Boosting Algorithms,” *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937-1967, 2021. https://doi.org/10.1007/s10462-020-09896-5

[5] Bloehdom S. and Hotho A., “Text Classification by Boosting Weak Learners Based on Terms and Concepts,” *in Proceedings of the 4th IEEE International Conference on Data Mining*, Brighton, pp. 331-334, 2004. https://ieeexplore.ieee.org/document/1410303

[6] Boruah A., Biswas S., and Bandyopadhyay S., “Transparent Rule Generator Random Forest: An Interpretable Random Forest,” *Evolving Systems*, vol. 14, no. 1, pp. 69-83, 2023. DOI:10.1007/s12530-022-09434-4

[7] Chadaga K., Prabhu S., Sampathila N., and Chadaga R., “A Machine Learning and Explainable Artificial Intelligence Approach for Predicting the Efficacy of Hematopoietic Stem Cell Transplant in Pediatric Patients,” *Healthcare Analytics*, vol. 3, pp. 1-15, 2023. https://doi.org/10.1016/j.health.2023.100170

[8] Choi E., Jun T., Park H., Lee J., Lee K., and Kim Y., “Predicting Long-Term Survival after Allogeneic Hematopoietic Cell Transplantation in Patients with Hematologic Malignancies: Machine Learning-based Model Development and Validation,” *JMIR Medical Informatics*, vol. 10, no. 3, pp. 1-9, 2022. https://medinform.jmir.org/2022/3/e32313

[9] Cunningham P. and Delany S., “K-Nearest Neighbour Classifiers-A Tutorial,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1-25, 2021. https://doi.org/10.1145/3459665

[10] Dabiri H., Farhangi V., Moradi M., Zadehmohamad M., and Karakouzian M., “Applications of Decision Tree and Random Forest as Tree-based Machine Learning Techniques for Analyzing the Ultimate Strain of Spliced and Non-Spliced Reinforcement Bars,” *Applied Sciences*, vol. 12, no. 10, pp. 1-13, 2022. https://doi.org/10.3390/app12104851

[11] Dangeti P., *Statistics for Machine Learning*, Packt Publishing, 2017. https://books.google.jo/books?id=f2rpswEACAAJ

[12] Dasari K. and Devarakonda N., “Detection of DDoS Attacks Using Machine Learning Classification Algorithms,” *International Journal of Computer Network and Information Security*, vol. 14, no. 6, pp. 89-97, 2022. DOI: https://doi.org/10.5815/ijcnis.2022.06.07

[13] De la Morena M. and Gatti R., “A History of Bone Marrow Transplantation,” *Hematology/Oncology Clinics*, vol. 25, no. 1, pp. 1-15, 2011. DOI:10.1016/J.HOC.2010.11.001

[14] Gandelman J., Byrne M., Mistry A., Polikowsky H., and Diggins K., “Machine Learning Reveals

Chronic Graft-Versus-Host Disease Phenotypes and Stratifies Survival after Stem Cell Transplant for Hematologic Malignancies," *Haematologica*, vol. 104, no. 1, pp. 189, 2019. https://doi.org/10.3324/haematol.2018.193441

[15] Gourisaria M., Patel A., Chatterjee R., and Sahoo B., "Predicting the Survival Status of Patient after Bone Marrow Transplant Using Fuzzy Discernibility Matrix," *in Proceedings of the OPJU International Technology Conference on Emerging Technologies for Sustainable Development*, Raigarh, pp. 1-6, 2023. DOI:10.1109/OTCON56053.2023.10114043

[16] Gratwohl A., Baldomero H., Aljurf M., Pasquini M., and Bouzas L., "Hematopoietic Stem Cell Transplantation: A Global Perspective," *JAMA*, vol. 303, no. 16, pp. 1617-1624, 2010. DOI:10.1001/JAMA.2010.491

[17] Guo L., Wang W., Xie X., Wang S., and Zhang Y., "Machine Learning-based Models for genomic Predicting Neoadjuvant Chemotherapeutic Sensitivity in Cervical Cancer," *Biomedicine and Pharmacotherapy*, vol. 159, pp. 1-7, 2023. https://doi.org/10.1016/j.biopha.2023.114256

[18] Gupta V., Braun T., Chowdhury M., Tewari M., and Choi S., "A Systematic Review of Machine Learning Techniques in Hematopoietic Stem Cell Transplantation," *Sensors*, vol. 20, no. 21, pp. 1-19, 2020. https://doi.org/10.3390/s20216100

[19] Hasasneh A., Hijazi H., Abu Talib M., Afadar Y., Bou Nassif A., and Nasir Q., "Wearable Devices and Explainable Unsupervised Learning for COVID-19 Detection and Monitoring," *Diagnostics*, vol. 13, no. 19, pp. 1-21, 2023. https://doi.org/10.3390/diagnostics13193071

[20] Hassanijalilian O., Igathinathane C., Bajwa S., and Nowatzki J., "Rating Iron Deficiency in Soybean Using Image Processing and Decision-Tree Based Models," *Remote Sensing*, vol. 12, no. 24, pp. 1-24, 2020. https://doi.org/10.3390/rs12244143

[21] Hegde R., Prasad K., Hebbar H., Singh B., and Sandhya I., "Automated Decision Support System for Detection of Leukemia from Peripheral Blood Smear Images," *Journal of Digital Imaging*, vol. 33, no. 2, pp. 361-374, 2019. DOI:10.1007/S10278-019-00288-Y

[22] Jiang D., Lin W., and Raghavan N., "A Novel Framework for Semiconductor Manufacturing Final Test Yield Classification Using Machine Learning Techniques," *IEEE Access*, vol. 8, pp. 197885-197895, 2020. https://ieeexplore.ieee.org/document/9244159

[23] Jun W. and Liyuan Z., "Brain Tumor Classification Based on Attention Guided Deep Learning Model," *International Journal of Computational Intelligence Systems*, vol. 15, no. 1, pp. 1-9, 2022. https://link.springer.com/article/10.1007/s44196-022-00090-9

[24] Kharroubi A. and Seir R., *Cancer Care in Countries and Societies in Transition*, Springer, 2016. https://doi.org/10.1007/978-3-319-22912-6_6

[25] Lee S., Lee C., Mun K., and Kim D., "Decision Tree Algorithm Considering Distances between Classes," *IEEE Access*, vol. 10, pp. 69750-69756, 2022. DOI: 10.1109/ACCESS.2022.3187172

[26] Li J. and Wang R., "An Anomaly Detection Method for Weighted Data Based on Feature Association Analysis," *The International Arab Journal of Information Technology*, vol. 21, no. 1, pp. 117-127, 2024. https://doi.org/10.34028/iajit/21/1/11

[27] Malikhah M., Sarno R., and Sabilla S., "Ensemble Learning for Optimizing Classification of Pork Adulteration in Beef Based on Electronic Nose Dataset," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 4, pp. 44-55, 2021. DOI: 10.22266/ijies2021.0831.05

[28] Mienye I. and Sun Y., "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, pp. 99129-99149, 2022. https://ieeexplore.ieee.org/document/9893798

[29] Mohanapriya M. and Lekha J., "Comparative Study between Decision Tree and KNN of Data Mining Classification Technique," *in Proceedings of the 2nd National Conference on Computational Intelligence*, Bangalore, pp. 1-8, 2018. DOI:10.1088/1742-6596/1142/1/012011

[30] Mustofa F., Safriandono A., Muslikh A., and Setiadi R., "Dataset and Feature Analysis for Diabetes Mellitus Classification Using Random Forest," *Journal of Computing Theories and Applications*, vol. 1, no. 1, pp. 41-49, 2023. https://doi.org/10.33633/jcta.v1i1.9190

[31] Navigating Cancer Care Cancer.Net, https://www.cancer.net/navigating-cancer-care/howcancer-treated/bone-marrowstem-cell-transplantation/what-bone-marrow-transplant-stem-cell-transplant, Last Visited, 2024.

[32] Raphael R. and Joy K., "Segmentation and Classification Techniques of Leukemia Using Image Processing: An Overview," *in Proceedings of the International Conference on Intelligent Sustainable Systems*, Palladam, pp. 378-384, 2019. DOI:10.1109/ISS1.2019.8907988

[33] Park S., Hamm S., and Kim J., "Performance Evaluation of the GIS-based Data-Mining Techniques Decision Tree, Random Forest, and Rotation Forest for Landslide Susceptibility Modeling," *Sustainability*, vol. 11, no. 20, pp. 1-20, 2019. https://doi.org/10.3390/su11205659

[34] Ratul I., Wani U., Nishat M., Al-Monsur A., Ar-Rafi A., Faisal F., and Kabir M., "Survival Prediction of Children Undergoing Hematopoietic

Stem Cell Transplantation Using Different Machine Learning Classifiers by Performing Chi-squared Test and Hyper-Parameter Optimization: A Retrospective Analysis," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1-14, 2022. https://doi.org/10.1155/2022/9391136

[35] Sapra V., Sapra L., Bansal Y., Chhabra G., and Tanwar R., "Machine Learning Approach for Identifying Survival of Bone Marrow Transplant Patients," *in Proceedings of the Emerging Technologies for Computing, Communication and Smart Cities,* Delhi, pp. 31-40, 2022. https://link.springer.com/chapter/10.1007/978-981-19-0284-0_3

[36] Shouval R., Labopin M., Bondi O., Mishan-Shamay H., Shimoni A., and Ciceri F., "Prediction of Allogeneic Hematopoietic Stem-Cell Transplantation Mortality 100 Days after Transplantation Using a Machine Learning Algorithm: A European Group for Blood and Marrow Transplantation Acute Leukemia Working Party Retrospective Data Mining Study," *Journal of Clinical Oncology*, vol. 33, no. 28, pp. 3144-3151, 2015. https://doi.org/10.1200/JCO.2014.59.1339

[37] Shouval R., Labopin M., Unger R., Giebel S., and Ciceri F., "Prediction of Hematopoietic Stem Cell Transplantation Related Mortality-Lessons Learned from the In-Silico Approach: A European Society for Blood and Marrow Transplantation Acute Leukemia Working Party Data Mining Study," *PLoS One*, vol. 11, no. 3, pp. 1-14, 2016. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150637

[38] Sklearn.ensemble.AdaBoostClassifier-scikit-learn 1.3.1 documentation, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html, Last Visited, 2024.

[39] Styczyński J., Tridello G., Koster L., and Iacobelli S., "Death after Hematopoietic Stem Cell Transplantation: Changes over Calendar Year Time, Infections and Associated Factors," *Bone Marrow Transplantation*, vol. 55, no. 1, pp. 126-136, 2019. DOI:10.1038/s41409-019-0624-z

[40] Taati B., Snoek J., Aleman D., and Ghavamzadeh A., "Data MIning in Bone Marrow Transplant Records to Identify Patients with High Odds of Survival," *IEEE Journal of mBiomed Health Informatics*, vol. 18, no. 1, pp. 21-27, 2014. DOI:10.1109/JBHI.2013.2274733

[41] UCI Machine Learning Repository: Bone Marrow Transplant: Children Data Set, https://archive.ics.uci.edu/ml/datasets/Bone+marrow+transplant%253A+children, Last Visited, 2024.

[42] Xia Y., Liu C., Li Y., and Liu N., "A Boosted Decision Tree Approach Using Bayesian Hyper-Parameter Optimization for Credit Scoring," *Expert Systems with Applications*, vol. 78, pp. 225-241, 2017. https://doi.org/10.1016/j.eswa.2017.02.017

[43] Zayed Y., Hasasneh A., and Tadj C., "Infant Cry Signal Diagnostic System Using Deep Learning and Fused Features," *Diagnostics*, vol. 13, no. 12, pp. 1-22, 2023. https://doi.org/10.3390/diagnostics13122107

Hussam Alawneh received a Bachelor's degree in Computer Systems Engineering from Arab American University in 2018. He has four years of experience in backend .NET development. Currently, he is completing his Master degree in Data Science and Business Analytics at Arab American University. Mr. Alawneh is honing his skills in Machine Learning, Deep Learning, and Data Analysis using Python. He has a good experience in C# and SQL, and exhibits a sharp proficiency in data preprocessing, visualization, and constructing predictive models with advanced techniques.

Ahmad Hasasneh received his B.Sc. degree in Computer Systems Engineering (CSE) from Palestine Polytechnic University (PPU) in 2005. Thereafter, in 2006, he obtained his M.Sc. degree in Computer Graphics and Programming from the University of Hull in the UK. Later, he served as a full-time lecturer in the Department of Computer Science at Hebron University for more than 3 years. In 2009, he was offered a full scholarship to join the doctoral program at Paris University, where he received his Ph.D. degree in Computer Science, majoring in Artificial Intelligence & Machine Learning, with distinction, in 2012. Dr. Hasasneh rejoined the Computer Science Department at Hebron University as an Assistant Professor for 2 years. From 2014 to 2015, he served as the Head of the Computer Department at Palestine Technical University Kadoorie (PTUK). From 2015 to 2021, he was an Assistant Professor and the Dean of the College of Engineering and Information Technology at Palestine Ahliya University (PAU). Since October 2021, he has been working as an Assistant Professor in the Department of Natural, Engineering, and Technology Sciences in the Faculty of Graduate Studies at the Arab American University (AAUP) and Head of Department of Data Science in the Faculty of Data Science at AAUP in the academic year of 2022/2023. He is currently heading the Department of Natural, Engineering, and Technology Sciences in the Faculty of Graduate Studies at AAUP. He was one of the key faculty members who established the Multimedia and Digital Production and Smart Systems Engineering programs at PAU. Dr. Hasasneh has held many positions at PAU related to curriculum development for IT programs, chairing the preparatory committee for organizing an international conference on technology and entrepreneurship and others.

Currently, Dr. Hasasneh has several collaborative research projects with international research centers, mainly focused on addressing medical problems based on new machine and deep learning models. In particular, within the Palestinian German Science Bridge (PGSB), a Ph.D. project entitled “Feature Extraction in Resting State Brain Activity Based on New Deep Learning Methods” is being carried out with the Juelich Research Centre in Germany. In addition, within the framework of the Palestinian-Quebec Science Bridge (PQSB), another research project entitled “Deep Learning-Based Cry Diagnostic System for the Early Identification of Infant Pathologies” is being implemented with the Canadian side (EST, University of Quebec, Montreal), with several results. Another international collaborative research project entitled “Wearable Devices and Interpretable Artificial Intelligence in Detecting and Classifying COVID-19” is currently being implemented with the University of Sharjah, UAE, and the University of Coimbra in Portugal. Dr. Hasasneh has published numerous international journals and conference papers, and his research interests include Machine Learning, Deep Learning, Robotics, Feature Extraction, Recognition, Robot Localization, Neurosciences, Image Processing, and Segmentation.