



Children Bone Marrow Transplant

Machine Learning Project



Children Bone Marrow Transplant

Machine Learning Project

Pediatric Bone Marrow Transplantation

- Essential treatment for children with serious blood disorders.
- Involves transplanting healthy stem cells to replace damaged or diseased marrow.
- Multiple factors influence the success of transplants.
- Predicting outcomes like transplant success and patient survival is complex.



Key Focus Areas for Predicting BMT Outcomes

- **Donor-Recipient Compatibility:**

Examine how age, blood type, and HLA matching impact transplant success.

- **Disease Characteristics:**

Analyze the effect of the type of blood disorder on survival rates and relapse.

- **Patient Demographics and Health Status:**

Assesses the influence of pre-transplant health conditions on outcomes.

- **Treatment Protocols:**

Evaluates the effectiveness of stem cell sources and conditioning regimens in improving patient survival.



01

Dataset Description

Initial insights



Origin and Collection: UCI Machine Learning Repository.



**Number of
entries**

187 rows.



**Number of
attributes**

37 columns.

Variables



Pre-Transplant Variables

Demographic Information: Recipient gender, age, and body mass.

Genetic Markers: Donor and recipient blood groups, Rh factor, HLA compatibility, cytomegalovirus status, gender compatibility.

Treatment Specifics: Stem cell source, risk group, disease type, second transplant after relapse.



Post-Transplant Variables

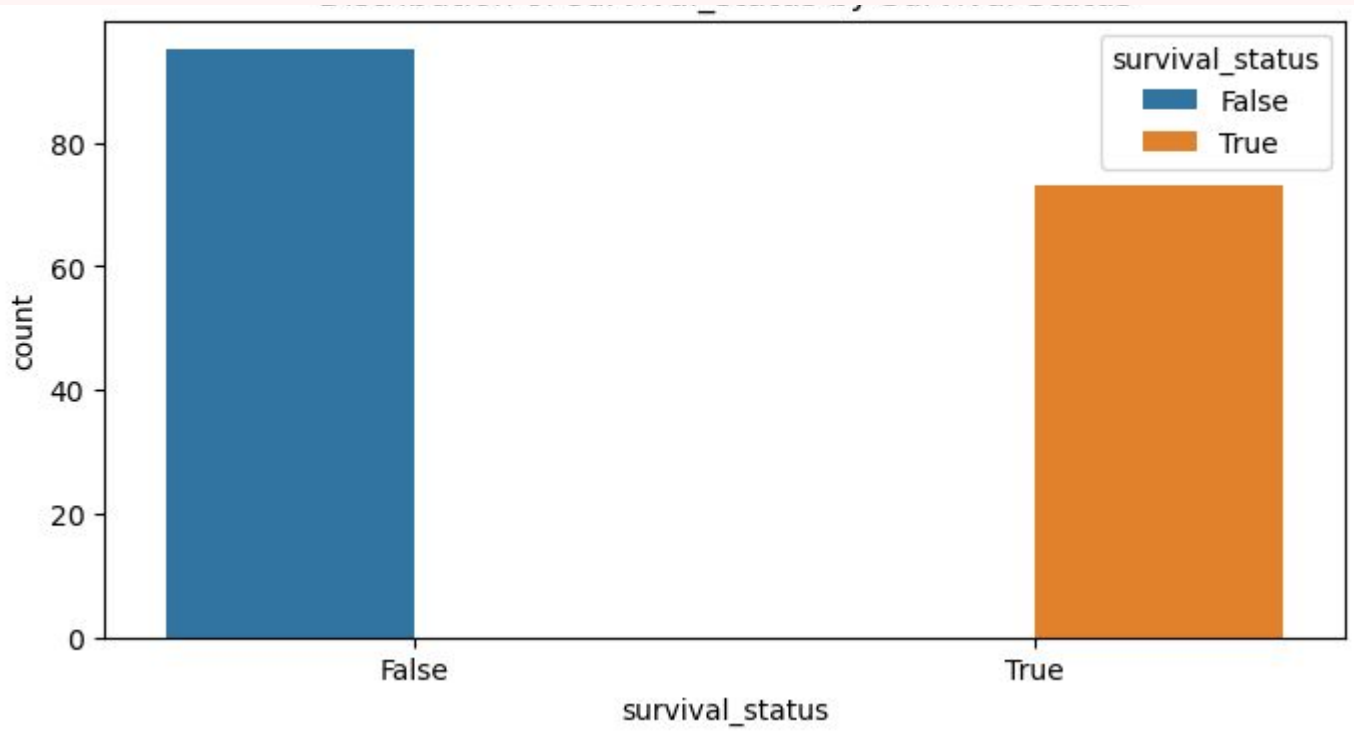
Recovery Metrics: Severity of acute and chronic GvHD, time for neutrophil and platelet recovery, cell dose measurements, time to acute GvHD development.

Outcomes: Disease recurrence, survival time, survival status

Summary Statistics

	Age (Recipient/Donor)		Cell Measurements	Survival Time	Survival Status
Min	1	18	0.79	6 days	Surviving: 45.5%
Max	20	56	57.78	3364 days	Dead: 54.5%
Mean	8	35		939 days	balanced dataset
Significance	considerable variation in outcomes.				

Summary Statistics





02

Methodology

Prediction Approaches

- **Basis:** the nature of the available features and the goal of the study.
- **Goal:** creating both accurate and unbiased models.



Comprehensive Process Analysis Approach

evaluates the entire BMT process by utilizing all available features, both pre-transplant and post-transplant.



Predictive Modeling Approach

predict the survival status of patients using only pre-transplant features.

Prediction Approaches



Comprehensive Process Analysis Approach

- **Accuracy:** Post-transplant features capture critical outcomes.
- **Pattern Analysis:** Understand variable interactions throughout the whole process.
- **Insights:** Identify key survival factors.
- **Effectiveness:** Offer comprehensive insights into BMT.



Predictive Modeling Approach

- **Realistic Assessment:** Realistic survival prospects pre-procedure.
- **Avoid Bias:** Exclude post-transplant features.
- **Preemptive:** Use pre-transplant info for early decisions.
- **Utility:** Crucial for early planning, valuable for clinicians.



03

**Data
Preprocessing**

Handling Missing Values

Recipientgender	0
Stemcellsource	0
Donorage	0
Donorage35	0
IIIV	0
Gendermatch	0
DonorABO	0
RecipientABO	1
RecipientRh	2
ABOmatch	1
CMVstatus	16
DonorCMV	2
RecipientCMV	14
Disease	0
Riskgroup	0
Txpostrelapse	0
Diseasegroup	0
HLAmatch	0
HLAmismatch	0
Antigen	1

Antigen	1
Alel	1
HLAgrI	0
Recipientage	0
Recipientage10	0
Recipientageint	0
Relapse	0
aGvHDIIIIV	0
extcGvHD	31
CD34kgx10d6	0
CD3dCD34	5
CD3dkgx10d8	5
Rbodymass	2
ANCRecovery	0
PLTRecovery	0
time_to_aGvHD_III_IV	0
survival_time	0
survival_status	0

Handling Missing Values



K-Nearest Neighbors (KNN) Imputation

- For continuous values (**CD3dCD34**, **CD34kgx10d6**, and **CD3dkgx10d8**).
- Suitable for clinical data where patients with similar vital indicators receive similar dosages.



Age-Specific Imputation

- For features related to age (**Body mass**).
- Especially important in pediatric patients.



Dropping Rows

- For rows with missing values in critical columns (**RecipientABO**, **RecipientRh**, **ABOMatch**, **Antigen**, **Alel**, **extcGvHD** and **CMVstatus**)



CMVstatus Custom Imputation

- For related columns (**DonorCMV** and **RecipientCMV**).
- The **CMVstatus** indicates donor and recipient compatibility regarding infection.

Other Preprocessing



Normalization/Scaling

Normalized using **StandardScaler** to have a mean of 0 and a standard deviation of 1.



One-Hot Encoding

Applied to categorical column “**Disease**” to transform them into a format suitable for machine learning algorithms.

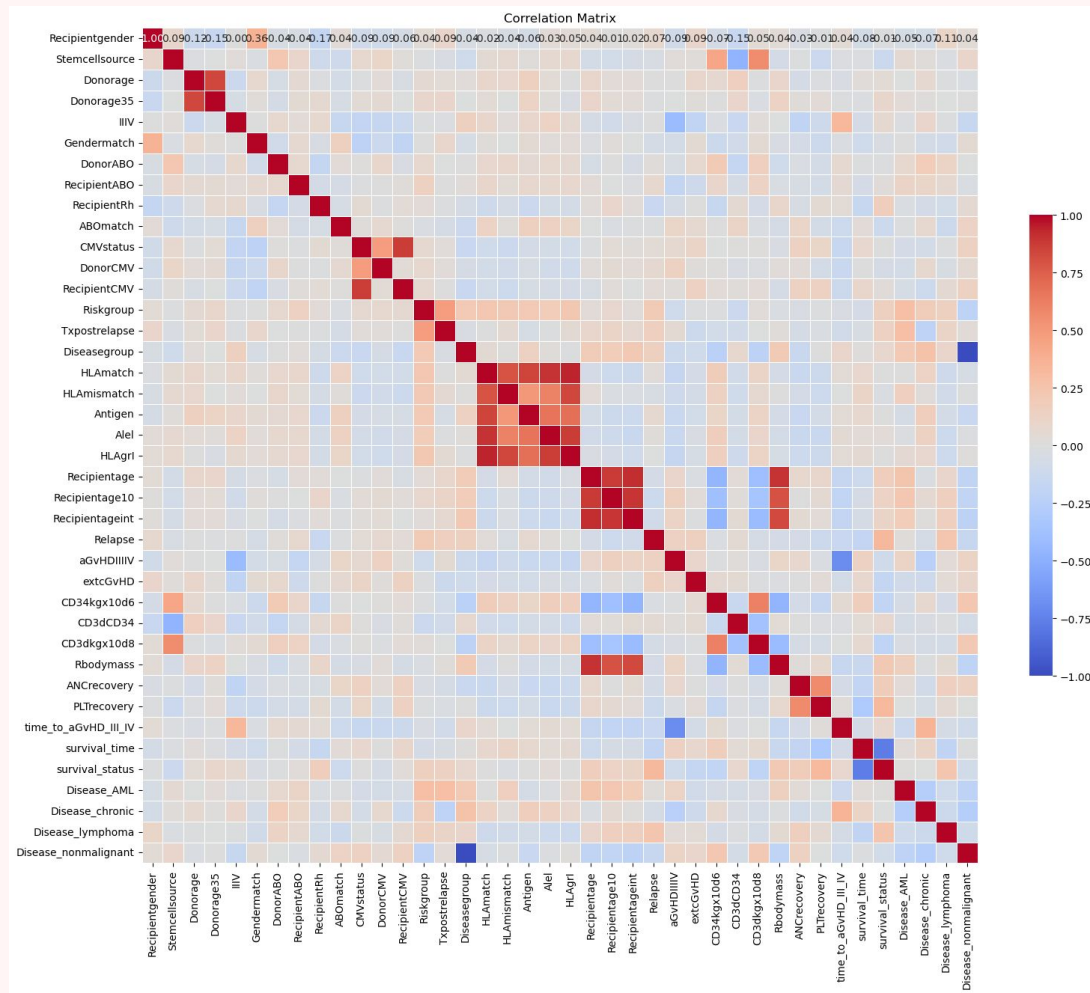
Disease_AML
Disease_chronic
Disease_lymphoma
Disease_nonmalignant



Conversion of columns

Converted the target column to boolean “**survival_status**”.

0 - didn't survive.
1 - survived.





04

**Machine
Learning Models**

Machine Learning Models



Decision Trees

Understands variable hierarchical importance (donor age, recipient mass, HLA matching) for predicting survival.



Random Forests

Enhances accuracy by combining trees; identifies complex interactions (cell dose, recovery metrics).



K-Nearest Neighbors

Predicts outcomes based on similarity of patients' features (vital signs, cell counts).



Naive Bayes

Makes probabilistic predictions for high-dimensional data using independent features (disease type, CMV status).



Support Vector Machines

Distinguishes survivors by maximizing class margin in high-dimensional spaces.



Artificial Neural Networks

Captures complex nonlinear relationships (genetic markers, post-transplant complications).

Prediction Approaches



Comprehensive Process Analysis Approach

(1st Approach)

Using pre-transplant and post-transplant features.



Predictive Modeling Approach

(2nd Approach)

Using pre-transplant features only.



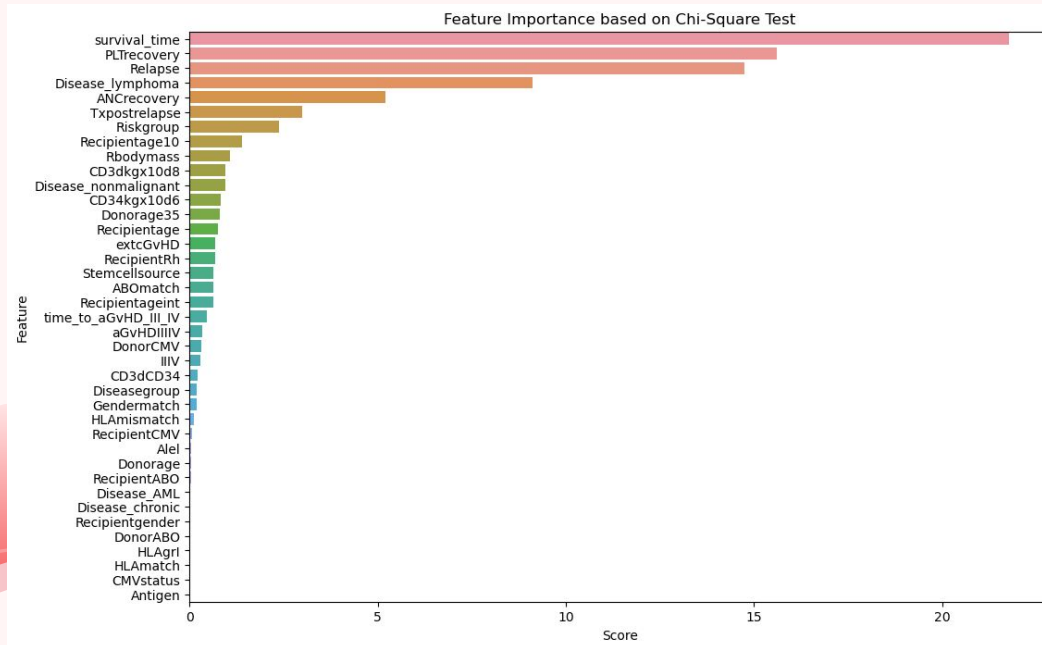
05

**Modeling and Results
(1st Approach)**

1

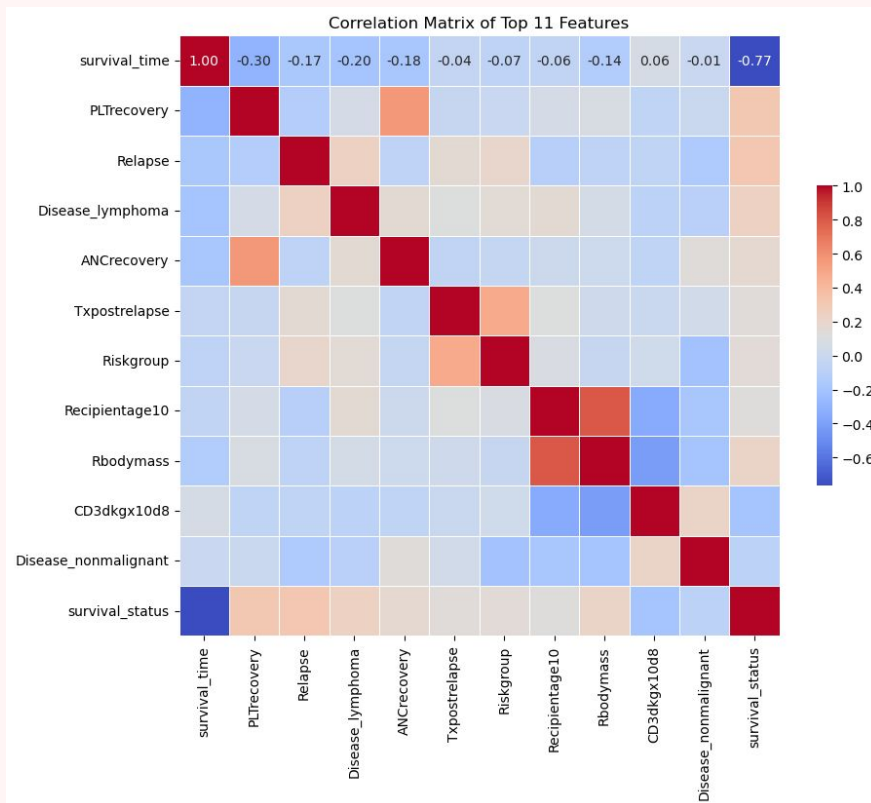
Feature Selection

- We applied the **SelectKBest** method using the **Chi-Square** test to extract the top 11 features.



	Feature	Score
34	survival_time	21.771535
32	PLT_recovery	15.610214
24	Relapse	14.735178
37	Disease_lymphoma	9.109589
31	ANC_recovery	5.204131
14	Txpostrelapse	3.002626
13	Riskgroup	2.384081
22	Recipientage10	1.395818
30	Rbodymass	1.081810
29	CD3dkgx10d8	0.958306
38	Disease_nonmalignant	0.949551

Feature Selection



2

Initial Modelling

Model	Accuracy	Precision	Recall	F1	ROC_AUC	Train Accuracy	Test Accuracy
DT	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RF	0.970588	0.9375	1.0	0.967742	1.0	1.0	0.970588
KNN	0.647059	0.666667	0.4	0.5	0.64386	0.843284	0.647059
NB	0.705882	0.857143	0.4	0.545455	0.926316	0.679104	0.705882
SVM	0.794118	0.9	0.6	0.72	0.905263	0.992537	0.794118
ANN	0.941176	0.933333	0.933333	0.933333	0.978947	1.0	0.941176

DT: Achieved perfect scores, indicating overfitting.

RF: High accuracy, precision, recall, and ROC AUC, making it reliable.

ANN: Robust performance with balanced precision, recall, and F1 scores.

SVM: Moderate performance with high precision but lower recall.

NB: Higher precision but lower recall, indicating trade-offs.

KNN: Lowest performance, highlighting limitations.

Our Candidates: Random Forest (RF) and Artificial Neural Network.

3

Repeated Cross Validation

(Reselecting Features for more accurate results) - RF MODEL

```
Average cross-validation score: 0.928974358974359 ± 0.04398920367747346
      precision    recall  f1-score   support

     0         1.00      1.00      1.00        19
     1         1.00      1.00      1.00        15

 accuracy          1.00          1.00          1.00        34
 macro avg         1.00          1.00          1.00        34
weighted avg         1.00          1.00          1.00        34
```

The repeated cross-validation scores indicate that the model is **performing well** on average.

The classification report on the test set shows perfect precision, recall, and F1-scores for both classes, which could be a **sign of overfitting**, especially given the small dataset.

3

Repeated Cross Validation

(Reselecting Features for more accurate results)

	Feature	Score
34	survival_time	21.771535
32	PLT_recovery	15.610214
24	Relapse	14.735178
37	Disease_lymphoma	9.109589
31	ANC_recovery	5.204131
14	Txpostrelapse	3.002626
13	Riskgroup	2.384081
22	Recipientage10	1.395818
30	Rbodymass	1.081810
29	CD3dkgx10d8	0.958306
38	Disease_nonmalignant	0.949551

**Best 11 Features Before Cross
Validation**

	Feature	Importance
34	survival_time	0.379839
26	extcGvHD	0.076298
24	Relapse	0.062799
27	CD34kgx10d6	0.048523
29	CD3dkgx10d8	0.044298
30	Rbodymass	0.043647
28	CD3dCD34	0.040714
21	Recipientage	0.038742
31	ANC_recovery	0.038190
32	PLT_recovery	0.036636
2	Donorage	0.026022

**Best 11 Features After Cross
Validation**

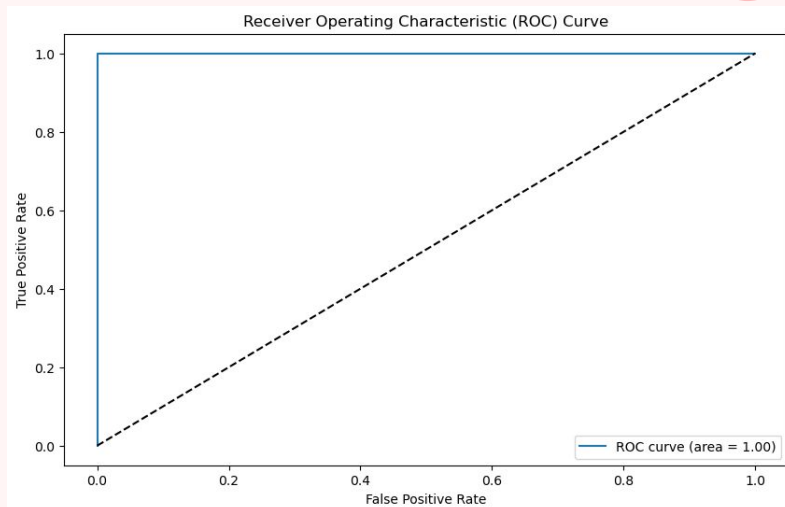
3

Repeated Cross Validation

(Using top 7 features)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	19
1	1.00	1.00	1.00	15
accuracy			1.00	34
macro avg	1.00	1.00	1.00	34
weighted avg	1.00	1.00	1.00	34

Cross-validated AUC scores: [0.99444444 0.97777778 0.97777778 1.
Average AUC score: 0.9755 ± 0.0257
Test ROC AUC score: 1.0000



- The results show a **perfect performance** of the RF model.
- The cross-validated AUC scores are also very high, indicating **strong predictive performance** across different folds. The ROC curve further confirms the model's robustness, achieving an area under the curve (AUC) of 1.0 on the test set.

4

Models Evaluation

(Using Cross Validation on All Models)

Model	Accuracy
DT	1.000000
RF	0.970588
KNN	0.647059
NB	0.705882
SVM	0.794118
ANN	0.941176

Before Cross Validation

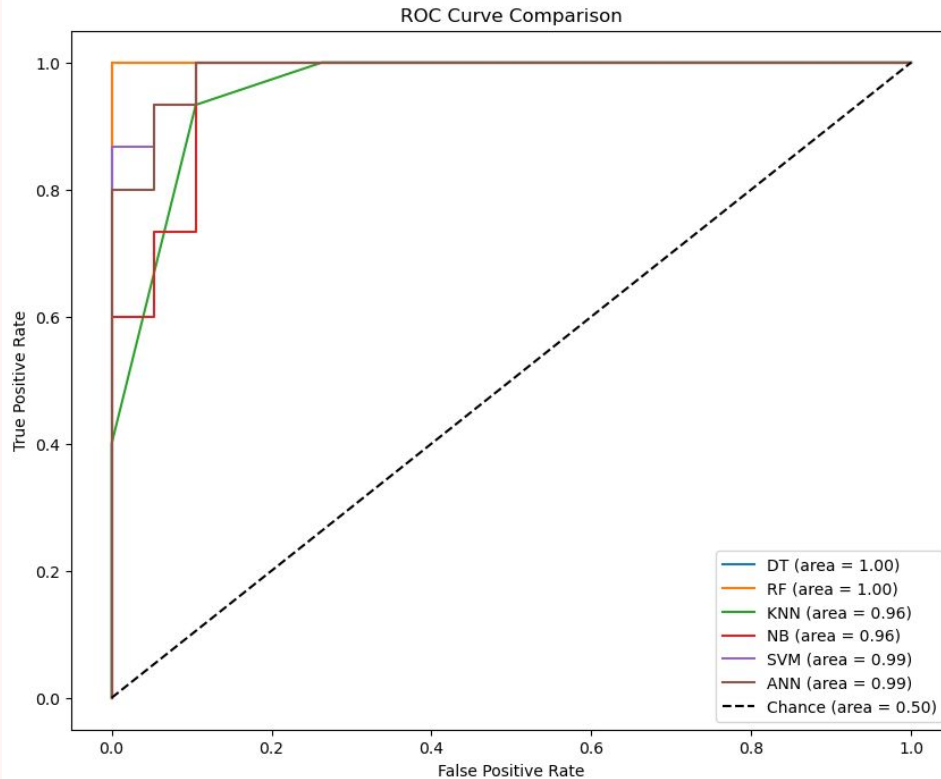
DT accuracy: 0.8931 ± 0.0479
RF accuracy: 0.9356 ± 0.0407
KNN accuracy: 0.8274 ± 0.0750
NB accuracy: 0.8938 ± 0.0572
SVM accuracy: 0.9004 ± 0.0573
ANN accuracy: 0.9110 ± 0.0498

After Cross Validation

- **RF model** stands out as the most accurate and stable model.
- The **ANN** also shows promising results with high accuracy and reasonable stability.
- **KNN model** appears to be the least reliable with the lowest accuracy and the highest variability.

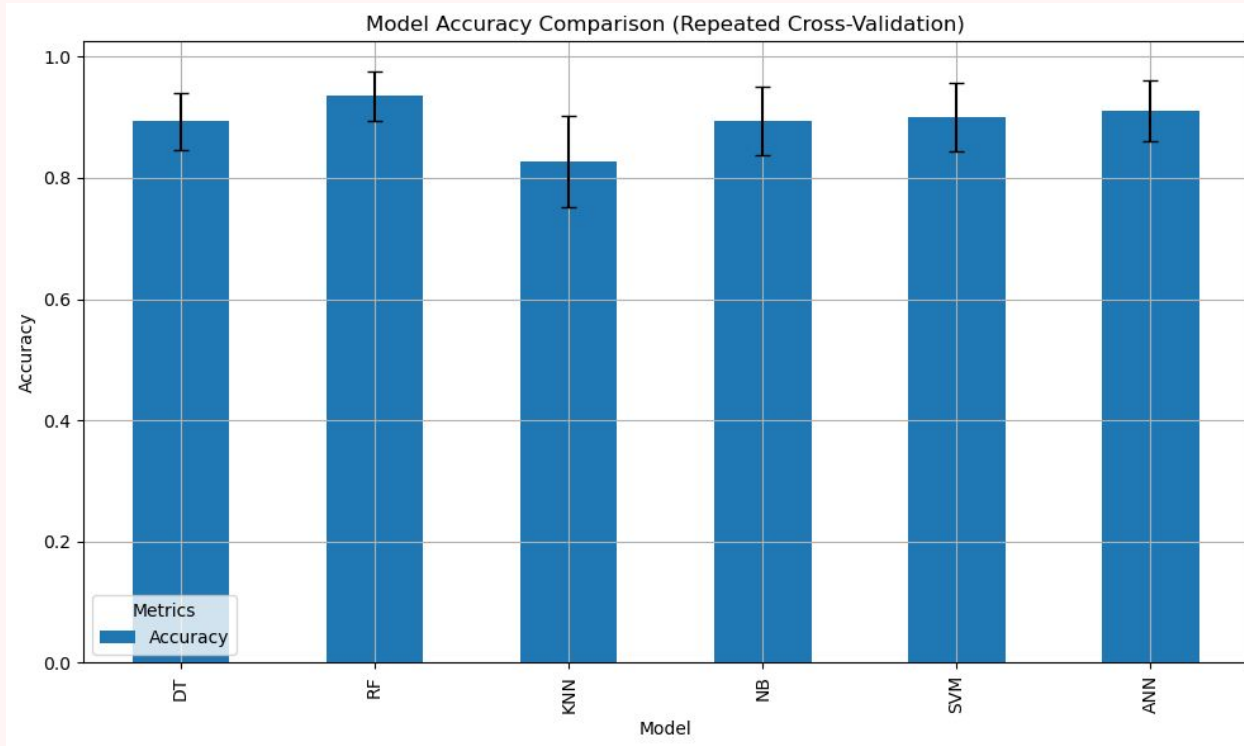
4

Model Evaluation



4

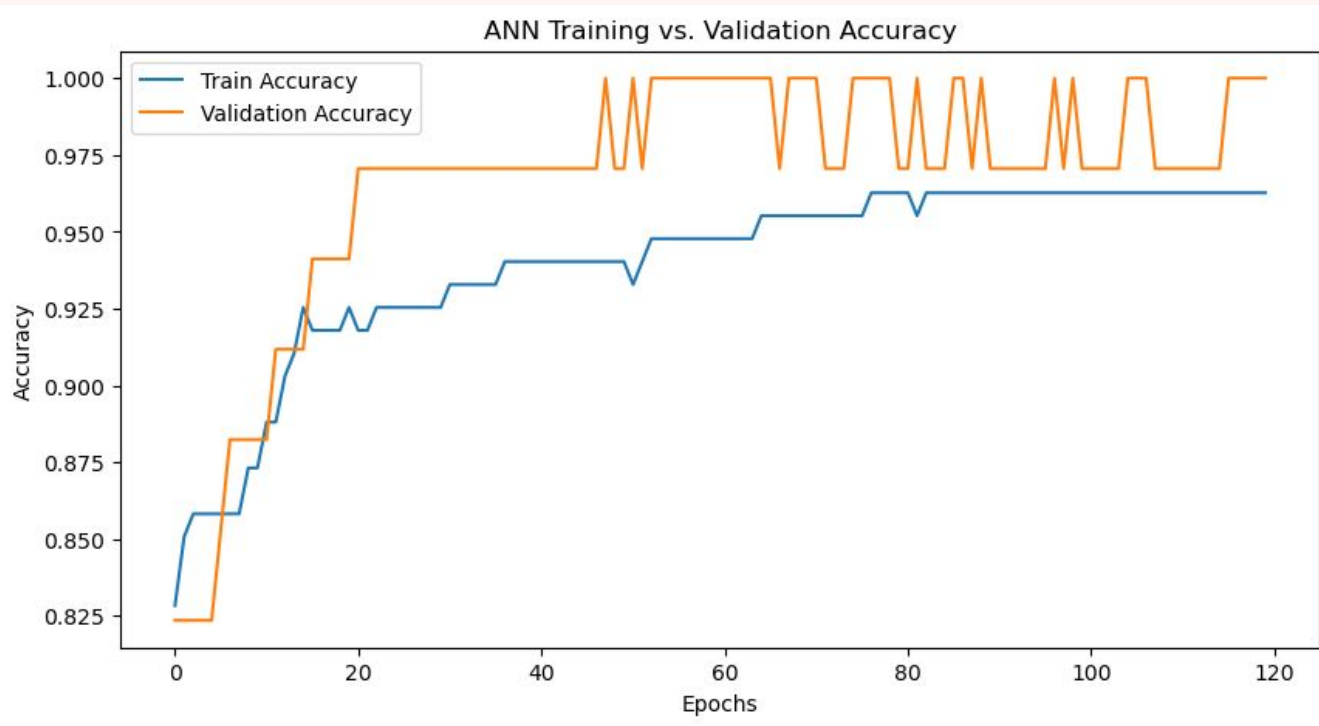
Model Evaluation



4

Model Evaluation

(ANN Training vs. Validation Accuracy)



5

Model Selection

Model	Precision (Class 0/1)	Recall (Class 0/1)	F1-Score (Class 0/1)	Accuracy	Repeated Cross-Validation Accuracy
RF	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.0	0.9319 ± 0.0446
ANN	0.86 / 0.92	0.95 / 0.80	0.90 / 0.86	0.88	0.9086 ± 0.0594
SVM	0.75 / 0.90	0.95 / 0.60	0.84 / 0.72	0.79	0.9004 ± 0.0573
NB	0.67 / 0.86	0.95 / 0.40	0.78 / 0.55	0.71	0.8938 ± 0.0572
KNN	0.64 / 0.67	0.84 / 0.40	0.73 / 0.50	0.65	0.8274 ± 0.0750
DT	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.0	0.8954 ± 0.0477

Random Forest (RF) shows the best balance of precision and recall across both classes and the highest repeated cross-validation accuracy.

Artificial Neural Network (ANN) performs well with balanced precision and recall, making it a strong candidate.

Support Vector Machine (SVM) and **Naive Bayes (NB)** provide high precision and recall but slightly lower than RF and ANN.

K-Nearest Neighbors (KNN) has the lowest performance with the highest variability, indicating less consistent performance.

Decision Tree (DT) performs perfectly in terms of precision, recall, and F1-score but might indicate overfitting.

5

Model Selection

Model	Precision (Class 0/1)	Recall (Class 0/1)	F1-Score (Class 0/1)	Accuracy	Repeated Cross-Validation Accuracy
RF	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.0	0.9319 \pm 0.0446
ANN	0.86 / 0.92	0.95 / 0.80	0.90 / 0.86	0.88	0.9086 \pm 0.0594
SVM	0.75 / 0.90	0.95 / 0.60	0.84 / 0.72	0.79	0.9004 \pm 0.0573
NB	0.67 / 0.86	0.95 / 0.40	0.78 / 0.55	0.71	0.8938 \pm 0.0572
KNN	0.64 / 0.67	0.84 / 0.40	0.73 / 0.50	0.65	0.8274 \pm 0.0750
DT	1.00 / 1.00	1.00 / 1.00	1.00 / 1.00	1.0	0.8954 \pm 0.0477

Best Model: Random Forest (RF), based on:

- The analysis of precision and recall.
- It provides the highest balanced F1-scores and robust performance across both classes.

making it the most reliable model for predicting survival status in this context. The high cross-validated AUC scores and consistent performance further validate its effectiveness.

5

Conclusion

Our goal was evaluating Bone Marrow Transplant survival prediction with various ML models: DT, RF, KNN, NB, SVM, ANN.

DT: Perfect scores (overfitting).

RF: High accuracy and stability; top performer.

ANN: Also demonstrated high accuracy and stability.

SVM: High precision but lower recall.

NB: Moderate performance.

KNN: Lowest performance among models tested.

This approach highlighted the importance of using **comprehensive features**, **tuning hyperparameters**, and **managing class imbalance** for reliable predictive performance in medical applications.

Random Forest Model was identified as the most reliable model for analysing survival status with both pre- and post- transplant features.



05

**Modeling and Results
(2nd Approach)**

1

Feature Exclusion

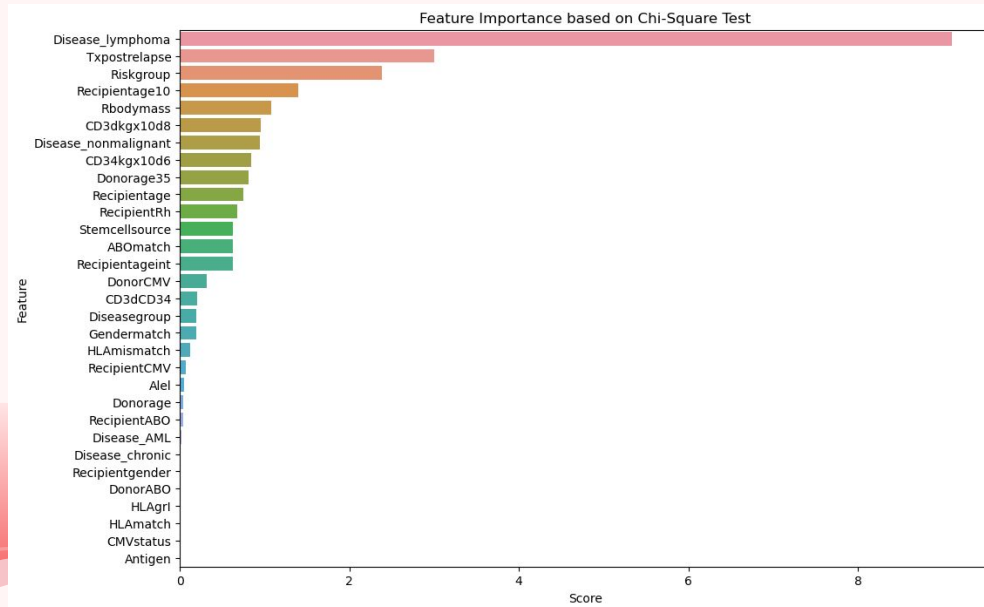
(Excluding Post-Transplant Features)

- **IIIIV:** Post-transplant complication.
- **GvHDIIIIV:** Severity of acute Graft-versus-Host Disease (II, III, or IV).
- **ANCrecovery:** Time to absolute neutrophil count recovery.
- **PLTrecovery:** Time to platelet recovery.
- **Survival_time:** Time to survive post-transplant.
- **Relapse:** Post-transplant relapse status.
- **extcGvHD:** Development of extensive chronic graft-versus-host disease
- **time_to_aGvHD_III_IV:** Time to develop an acute Graft-versus-Host Disease.

2

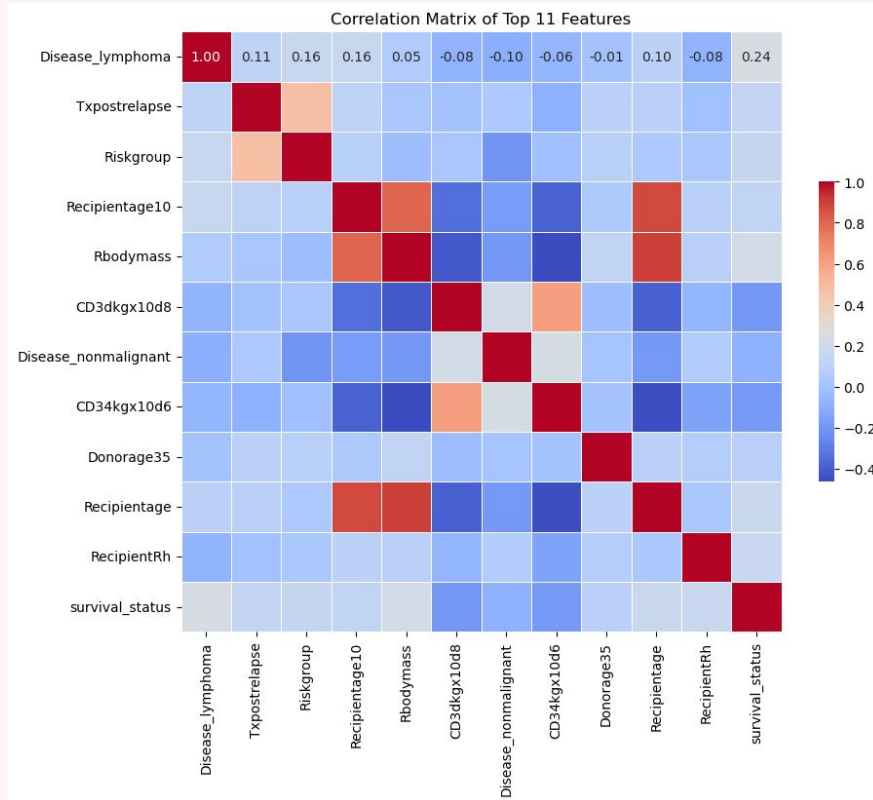
Feature Selection

- We applied the **SelectKBest** method using the **Chi-Square** test to extract the top 11 features.



	Feature	Score
29	Disease_lymphoma	9.109589
13	Txpostrelapse	3.002626
12	Riskgroup	2.384081
21	Recipientage10	1.395818
26	Rbodymass	1.081810
25	CD3dkgx10d8	0.958306
30	Disease_nonmalignant	0.949551
23	CD34kgx10d6	0.841769
3	Donorage35	0.813180
20	Recipientage	0.747882
7	RecipientRh	0.673072

Feature Selection



3

Initial Modelling

(Using top 7 features)

Model	Precision	Recall	F1	Train Accuracy	Test Accuracy
DT	0.58	0.58	0.58	0.5154	0.5882
RF	0.5	0.5	0.5	0.5672	0.5
KNN	0.39	0.39	0.38	0.6721	0.3824
NB	0.81	0.6	0.55	0.5969	0.6471
SVM	0.44	0.44	0.44	0.6652	0.4412
ANN	0.44	0.44	0.44	0.7892	0.665

- Disease_lymphoma
- Riskgroup
- Txpostrelapse
- Recipientage10
- Rbodymass
- CD3dkgx10d8
- Disease_nonmalignant

Using the **top 7 features**, the results indicated potential overfitting due to the small dataset size (~180 samples)

4

Feature Reduction

(Reselecting Features for more accurate results)

Model	Best Feature Set	Train Accuracy	Test Accuracy
DT	('Disease_lymphoma', 'Riskgroup', 'Txpostrelapse')	0.5969	0.6471
RF	('Disease_lymphoma', 'Riskgroup', 'Txpostrelapse')	0.6199	0.6471
KNN	('Disease_lymphoma', 'Riskgroup', 'CD3dkgx10d8', 'Disease_nonmalignant')	0.5977	0.6765
NB	('Disease_lymphoma', 'Riskgroup', 'Txpostrelapse')	0.5969	0.6471
SVM	('Disease_lymphoma', 'Riskgroup', 'Txpostrelapse', 'CD3dkgx10d8')	0.598	0.6765
ANN	('Disease_lymphoma', 'Txpostrelapse', 'Rbodymass')	0.6941	0.7017

Best Feature Combination: ['Disease_lymphoma', 'Txpostrelapse', 'Riskgroup', 'Recipientage10']

4

Feature Reduction

(Training with reduced 4 best models)

Model	Precision	Recall	F1	Train Accuracy	Test Accuracy
DT	0.54	0.54	0.53	0.59	0.5294
RF	0.54	0.54	0.53	0.5974	0.5294
KNN	0.67	0.61	0.6	0.5903	0.6471
NB	0.81	0.6	0.55	0.5969	0.6471
SVM	0.54	0.54	0.53	0.59	0.5294
ANN	0.54	0.54	0.53	0.6586	0.59

Benefits: KNN and SVM improved with fewer features.

Drawbacks: DT, RF, and ANN performed worse with fewer features.

Stability: NB's performance remained stable with fewer features.

Generalization Improvement: Reduced features helped mitigate overfitting in KNN and SVM.

Overfitting Addressed: The primary goal of reducing features was to address overfitting.

5

The hyperparameter tuning

(SVM - KNN)

Model	Best Parameters	CV Score	Precision	Recall	F1-Score	Test Accuracy
SVM	{'kernel': 'linear', 'gamma': 1, 'C': 100}	0.6123	0.58	0.58	0.58	0.5789
KNN	{'weights': 'distance', 'n_neighbors': 7, 'metric': 'euclidean'}	0.6248	0.61	0.61	0.6	0.6053

RF Performance: Highest CV score (0.6514) and test accuracy (0.6316).

SVM Performance: Used a Linear kernel, CV score (0.6123), test accuracy (0.5789).

KNN Performance: Distance weighting, seven neighbors, CV score (0.6248), test accuracy (0.6053).

Generalization: RF showed best generalization and balanced metrics.

Overfitting: KNN showed signs of overfitting.

Conclusion: RF is the most reliable model for predicting survival status.

5

The hyperparameter tuning

(Artificial Neural Network + cross validation)

Classification Report for ANN:

	precision	recall	f1-score	support
0	0.60	0.47	0.53	19
1	0.47	0.60	0.53	15
accuracy			0.53	34
macro avg	0.54	0.54	0.53	34
weighted avg	0.54	0.53	0.53	34

Average Training Accuracy for ANN: 0.7463

Testing Accuracy for ANN: 0.6946

Test Set Performance: Balanced precision and recall, overall accuracy of 0.53.

Overfitting: Moderate overfitting observed.

Learning Patterns: High training accuracy indicates good learning on training data.

Generalization: Struggles to generalize to unseen data.

Improvement Needed: Further tuning or additional features needed to improve generalization and reduce overfitting.

5

The hyperparameter tuning

(Random Forest)

```
Fitting 3 folds for each of 100 candidates, totalling 300 fits  
Best parameters found for RF: {'n_estimators': 100, 'min_samples_split':  
2, 'min_samples_leaf': 2, 'max_features': 'sqrt', 'max_depth': 20}  
Best cross-validation score for RF: 0.6513725490196078
```

Classification Report for RF (Test):

	precision	recall	f1-score	support
0	0.60	0.67	0.63	18
1	0.67	0.60	0.63	20
accuracy			0.63	38
macro avg	0.63	0.63	0.63	38
weighted avg	0.64	0.63	0.63	38

Test Accuracy for RF: 0.6316

Tuning Method: Randomized Search
CV.

CV Score: 0.651.

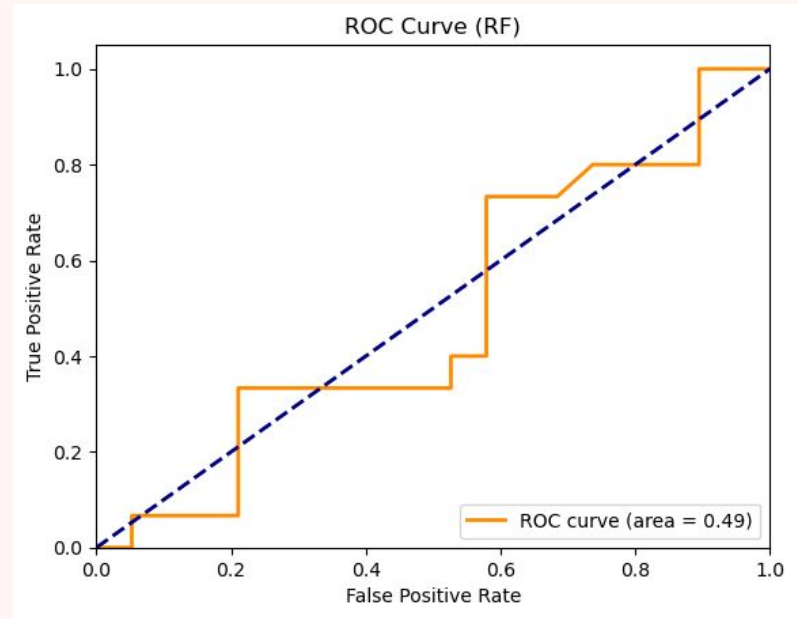
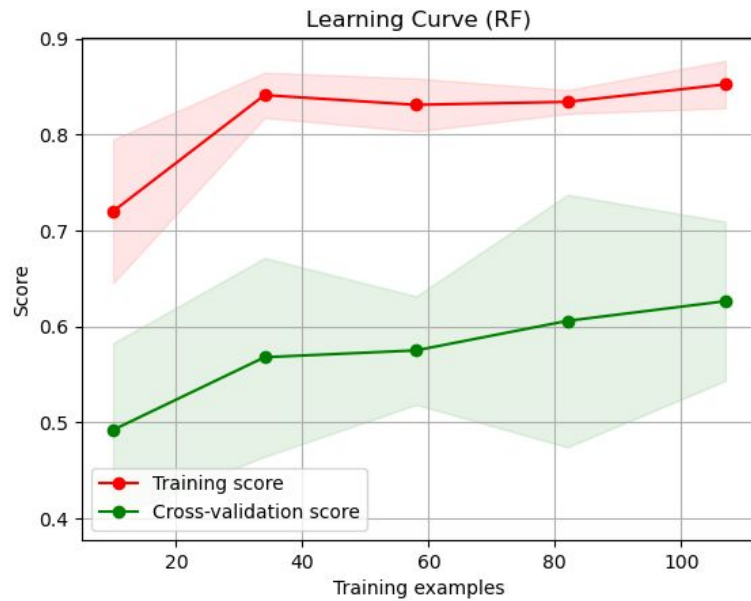
Test Accuracy: 0.63.

Performance: Balanced precision and
recall for both classes

6

Model Evaluation

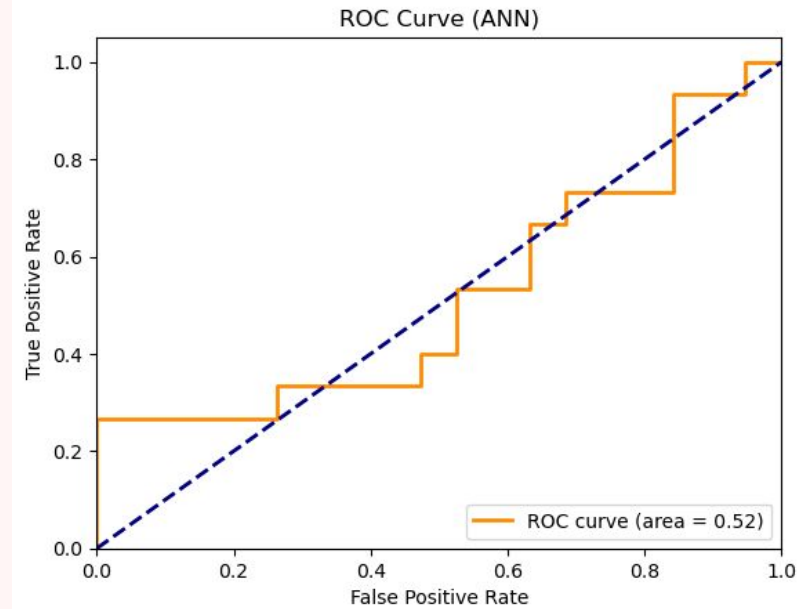
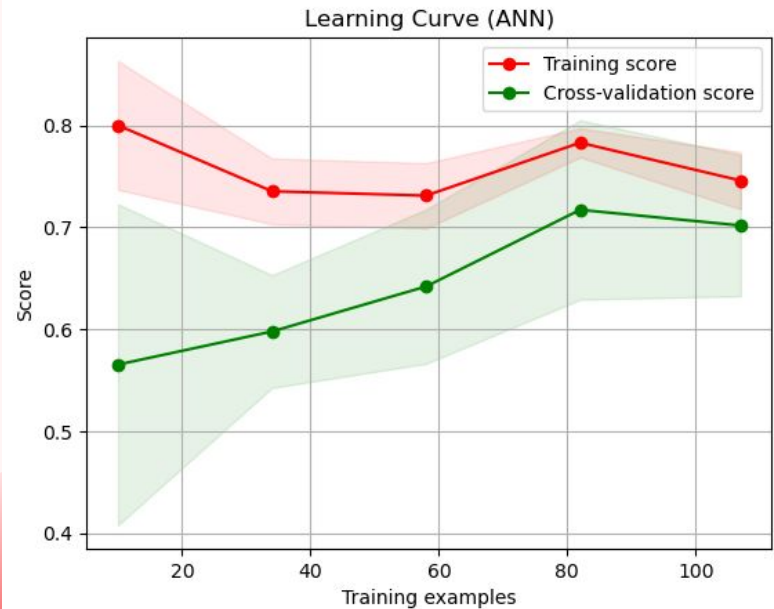
(Random Forest)



6

Model Evaluation

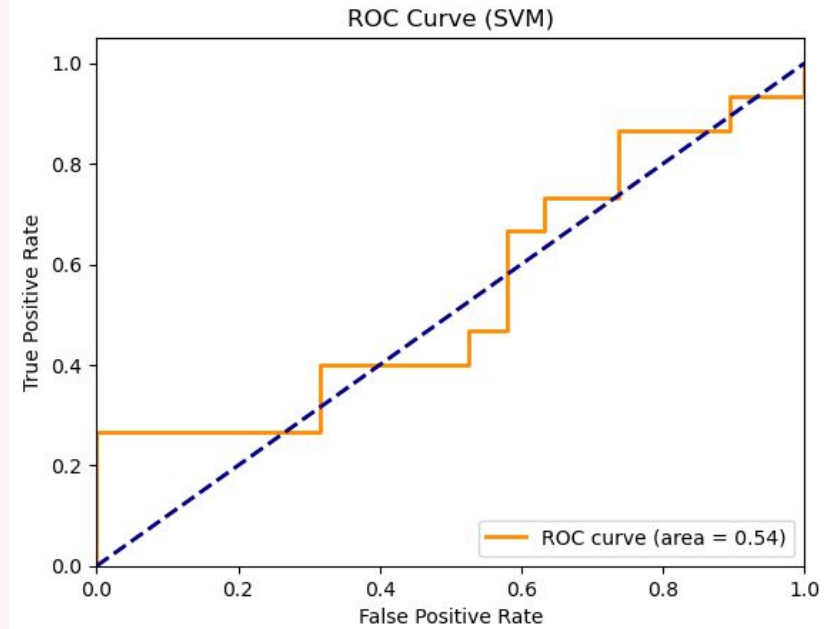
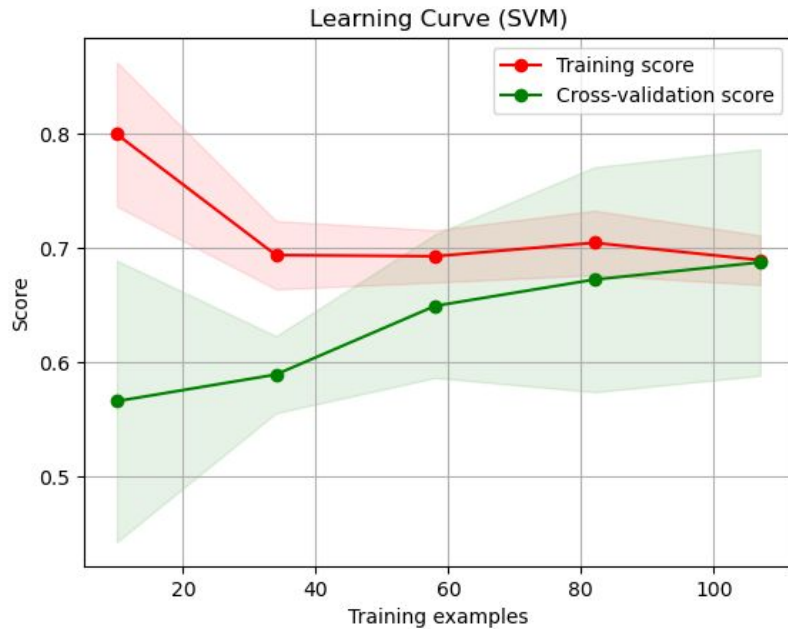
(Artificial Neural Network)



6

Model Evaluation

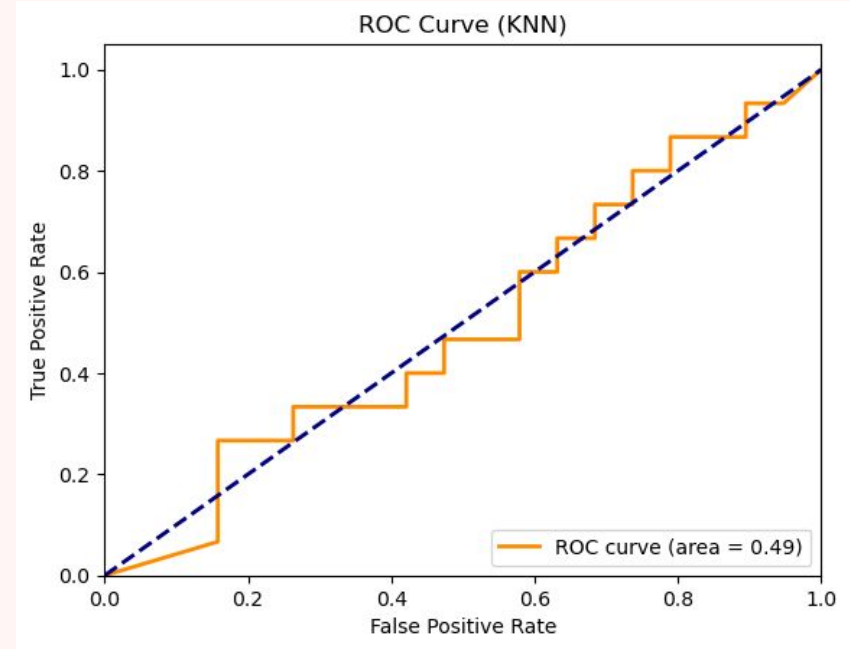
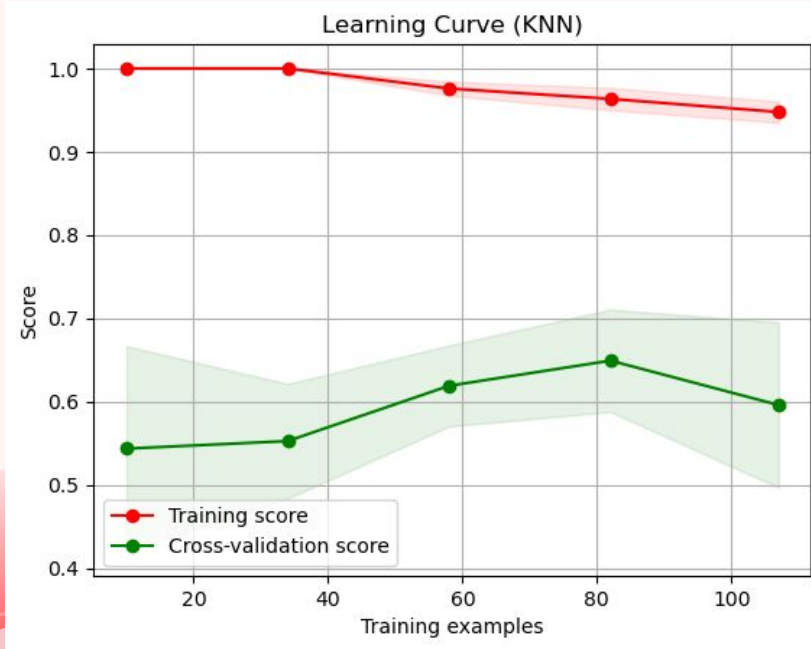
(Support Vector Machine)



6

Model Evaluation

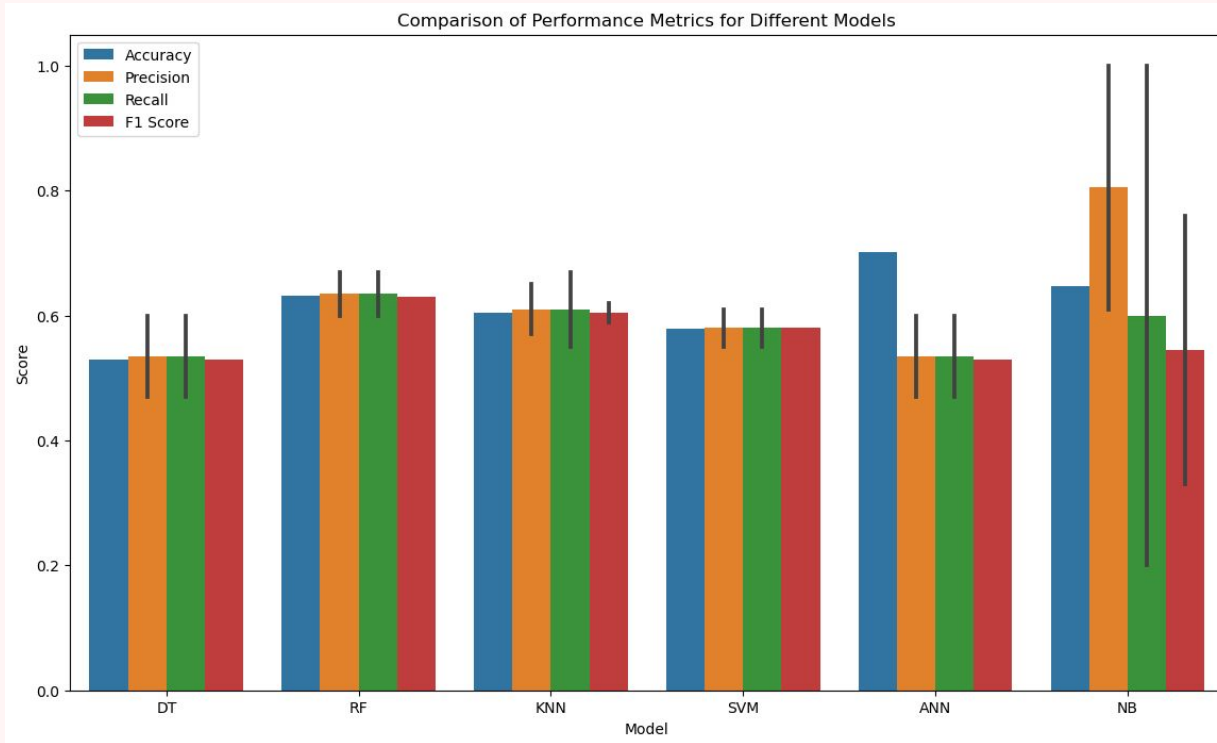
(K-Nearest Neighbors)



6

Model Evaluation

(Models Comparison)



7

Model Selection

Model	Precision (0/1)	Recall (0/1)	F1-Score (0/1)	Accuracy
DT	0.60 / 0.47	0.47 / 0.60	0.53 / 0.53	0.53
NB	0.61 / 1.00	1.00 / 0.20	0.76 / 0.33	0.65
RF	0.60 / 0.67	0.67 / 0.60	0.63 / 0.63	0.63
ANN	0.60 / 0.47	0.47 / 0.60	0.53 / 0.53	0.53
SVM	0.55 / 0.61	0.61 / 0.55	0.58 / 0.58	0.58
KNN	0.57 / 0.65	0.67 / 0.55	0.62 / 0.59	0.61

Support Vector Machine (SVM) shows a good balance of precision and recall across both classes, with F1-scores of 0.58 for both classes and stability indicated by the learning curve.

Random Forest (RF) also performs well with balanced precision and recall, but slightly lower stability as indicated by the learning curve.

K-Nearest Neighbors (KNN) has good metrics but less stability compared to SVM.

Artificial Neural Network (ANN) shows lower precision and recall metrics compared to RF, SVM, and KNN.

7

Model Selection

Model	Precision (0/1)	Recall (0/1)	F1-Score (0/1)	Accuracy
DT	0.60 / 0.47	0.47 / 0.60	0.53 / 0.53	0.53
NB	0.61 / 1.00	1.00 / 0.20	0.76 / 0.33	0.65
RF	0.60 / 0.67	0.67 / 0.60	0.63 / 0.63	0.63
ANN	0.60 / 0.47	0.47 / 0.60	0.53 / 0.53	0.53
SVM	0.55 / 0.61	0.61 / 0.55	0.58 / 0.58	0.58
KNN	0.57 / 0.65	0.67 / 0.55	0.62 / 0.59	0.61

Best Model: Support Vector Machine (SVM), based on:

- the analysis of precision and recall combined with learning curve stability.

It provides balanced F1-scores and robust performance across both classes, making it the most reliable model for predicting survival status in this context.

Conclusion

This approach aimed to predict survival status using various ML models, focusing on **handling class imbalance** and **optimizing performance**.

DT: Perfect scores (overfitting).

RF: High accuracy and stability; top performer.

ANN: Improved with added complexity, but struggled with a reduced feature set.

SVM: Best performer based on precision, recall, and stability; balanced precision and recall effectively.

NB: Moderate performance.

KNN: Lowest performance among models tested.

Initially, models included post-transplant features, resulting in high accuracy due to the richness of the data. Removing these features significantly dropped accuracy, highlighting their importance.

Despite challenges, the use of advanced techniques allowed for reasonable predictive performance. **The Support Vector Machine (SVM) model** was identified as the most reliable for this task, underscoring the need for comprehensive data and sophisticated models in medical predictions.



05

Conclusion

Conclusion

In this study, we aimed to predict **patient survival** using various machine learning models, including Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Machine, and Artificial Neural Networks.

Initially, we used a **full feature set**, including post-transplant data, then focused on **pre-transplant** features.

Models with the **full feature** set had high accuracy, but performance dropped when post-transplant features were removed.

Our methodical approach involved **feature selection**, **handling class imbalances**, and optimizing model performance through **hyperparameter tuning**.

Conclusion

Feature Selection and Reduction: Post-transplant features boosted the accuracy; removing them reduced performance. Feature reduction simplified models and reduced overfitting.

Class Imbalance Handling: Resampling balanced classes, helped improving metric performance.

Model Optimization: Hyperparameter tuning and cross-validation refined models. RF had the best precision-recall balance.

Model Performance: Despite fewer features and limited dataset, models performed well. ANN improved with increased complexity.

Conclusion

The insights gained from the project can guide future efforts in predictive modeling for bone marrow transplants and similar medical scenarios.

Future works could involve:

- **Data Collection:** Gather more comprehensive data, including larger sample sizes and additional features.
- **Advanced Techniques:** Explore ensemble methods and neural network architectures to improve accuracy and reliability.
- **Domain Knowledge:** Integrate domain expertise into feature selection and model tuning for robust and interpretable models.

The background features abstract, flowing shapes in various shades of pink and red, creating a soft, organic feel. These shapes are layered, with some appearing more prominent than others, and they frame the central text.

Thanks