Survival Prediction for Patients Undergoing HSCT Using Machine Learning Techniques

Rishabh Hanselia[1], Dilip Kumar Choubey[1,*], Ashutosh Mishra[2], Ratnesh Kumar Dubey[1], and Kanchan Bala[3]

ARTICLEINFO

Article history:
Received: 12 September 2023
Revised: 8 March 2024
Accepted: 5 October 2024
Online: 31 October 2025

Keywords:
HSCT
HCT
Allo-HSCT
EBMT
Survival Prediction
Machine Learning

ABSTRACT

Multipotent hematopoietic stem cells are transplanted in HSCT to proliferate within patients and generate healthy blood cells, typically derived from the umbilical cord, peripheral blood, or bone marrow. This procedure is common among patients with particular bone marrow disorders, such as leukemia or multiple myeloma. Despite being a life-saving intervention, HSCT carries inherent risks. Thus, predicting patient survivability and the influencing factors becomes paramount. In this study, the researchers employed a dataset from HSCT patients to thoroughly assess the predictive capacities of diverse ML models for patient survivability. The investigation also delved into prior research on patient survivability in HSCT. By evaluating and contrasting the performance of different ML models, the study objective is to identify the optimal technique, thereby aiding medical professionals in informed decision-making for their patients. The outcomes highlighted the Random Forest Classifier as the top-performing algorithms, which obtained an accuracy of 94.74% and an F1 score of 0.944.

1. INTRODUCTION

Undifferentiated cells, commonly referred to as stem cells, possess the remarkable ability to develop into distinct specialized cell types within the body. These versatile cells can be sourced from various origins, such as bone marrow, peripheral blood, or umbilical cord blood. In different scenarios, a patient's own stem cells may be used for Autologous SCT, while Allogeneic SCT involve cells from a donor, and syngeneic stem cell transplants utilize cells from an identical twin.

Hematopoietic Stem Cell Transplantation (HSCT), often called a bone marrow transplant or HPSCT, is a procedure wherein individuals with impaired or deficient bone marrow receive healthy HSC. Hemoglobinopathies, immune deficiency syndromes, and other illnesses benefit from this therapy strategy. It can either bolster bone marrow function or facilitate the replacement of malfunctioning cells, thereby either eliminating malignancy or generating functional replacements.

Usually, chemotherapy or radiation is used to suppress the recipient's immune system prior to transplantation. Notably, graft-versus-host disease and infections are two serious side effects of allogeneic HSCT. HSCT is only utilized for patients with terminal conditions because of the hazards involved. Over time, its application has grown beyond the treatment of cancer to include autoimmune diseases and hereditary skeletal dysplasias, especially diseases like mucopolysaccharidosis and malignant infantile osteopetrosis. Improved post-procedure survival rates are the cause of this increased use.

1.1 Motivation

While mortality rates following transplantation have seen a decline in latest years, they still remain noteworthy. A number of risk scores and indexes are now available to help doctors make decisions. These include the EBMT risk score and the Hematopoietic Cell Transplant Comorbidity Index (HCT-CI). However, these methods rely on conventional statistical approaches and exhibit less than optimal predictive accuracies. Given the burgeoning medical data landscape and the need for improved predictions, there is an increasing demand for more cutting-edge models. As demonstrated in [1], ML and various DM approaches hold promise in meeting this demand.

1.2 Contribution

This research focuses on an in-depth exploration of current literature concerning the prediction of patient survival during HSCT. The primary objective is to assess the predictive efficacy of ML in anticipating the survival outcomes of patients undergoing HSCT. This study intends to establish the foundation for upcoming research initiatives

[1]*Department of Computer Science and Engineering, Indian Institute of Information Technology Bhagalpur, Bihar, India.*
[2]*Department of Computer Science & Engineering, Thapar Institute of Engineering & Technology, Patiala, Punjab, India.*
[3]*Department of Computer Science & Engineering, Gaya College of Engineering, Gaya, DSTTE, Bihar.*
**Corresponding author:* Dilip Kumar Choubey; Email: dkchoubey.cse@iiitbh.ac.in.

by exploring previous studies, offering insightful information on the possibilities application of ML to enhance clinical decision-making regarding patient survival in the context of HSCT. The results of this manuscript could potentially help significantly to the advancement of medical knowledge and could serve as a foundation for further investigations in this critical area.

The structure of the manuscript is as follows: Related Work is presented in Section 2, the Proposed Approach described in Section 3, Section 4 showcases the Experimental Results, and Section 5 is dedicated to Experiment Insights, Discussion, and Future Directions, the same has been showcased in Fig 1.

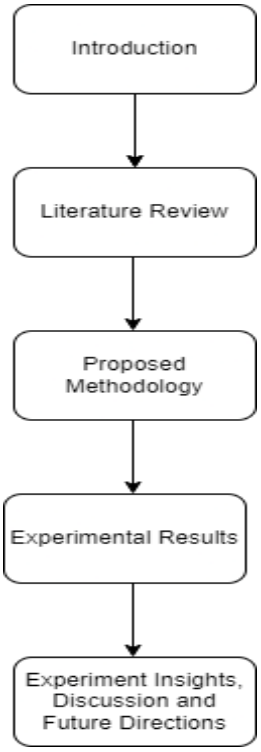

Fig. 1. Organizational flow diagram.

2. RELATED WORK

Jahan Ratul et al. [1] đã sử dụng tối ưu tham số và nhiều kỹ thuật học máy để dự đoán tỉ lệ tử vong ở trẻ nhận ghép tế bào gốc tạo máu (HSCT), đồng thời tiến hành phân tích so sánh hiệu quả các thuật toán được dùng. Karami et al. [2] đã ứng dụng một loạt các thuật toán học máy nhằm nhận diện các yếu tố ảnh hưởng đến tỷ lệ sống sót của bệnh nhân mắc bệnh bạch cầu myeloid cấp tính (AML). Pan et al. [3] tập trung vào ứng dụng thuật toán học máy để dự đoán tái phát ở trẻ bị bạch cầu lympho cấp (ALL), một yếu tố quan trọng tác động đến tỷ lệ sống còn của bệnh nhân. Iwasaki et al. [4] sử dụng ba thuật toán học máy khác nhau (Dynamic-DeepHit, Random Survival Forest và AdaBoost) để ước đoán tỷ lệ tử vong tổng thể và không tái phát sau ghép tế bào gốc đồng loại lần đầu tiên từ 1986 đến 2016 ở bệnh nhân các bệnh huyết học. Goswami et al. [5] phát triển hệ thống hai giai đoạn với ba yếu tố để phân loại bệnh nhân đa u tủy đang điều trị ghép tự thân vào nhóm nguy cơ tái phát cao hoặc thấp. Okamura et al. [6] xây dựng ứng dụng web tương tác dùng mô hình Random Survival Forest nhằm cung cấp dự báo tiên lượng cá nhân cho người nhận ghép tế bào gốc đồng loại. Taati et al. [7] áp dụng khai phá dữ liệu và học máy để xác định bệnh nhân ghép tủy xương có khả năng sống sót cao hơn. Choi et al. [8] dùng năm mô hình học máy khác nhau để dự đoán sống sót dài hạn của bệnh nhân ung thư huyết học sau ghép đồng loại, trong đó mô hình Gradient Boosting (GB) cho kết quả tốt nhất.

Iwasaki et al. [9] áp dụng học tập kết hợp (ensemble learning) tích hợp nhiều thuật toán để dự đoán thời gian sống còn không có GVHD hoặc tái phát. Eisenberg et al. [10] dùng phương pháp Gradient Boosting kết hợp dữ liệu nền và CMV để tiên đoán tử vong theo thời gian trong ghép tế bào gốc tạo máu. Shouval et al. đã đề xuất thuật toán cây quyết định (DT) nhằm tạo mô hình dễ hiểu cho dự đoán tử vong ghép đồng loại HSCT, cho thấy các mô hình học máy vượt trội so với các điểm số và chỉ số truyền thống [11]. Nghiên cứu tiếp theo của nhóm này sử dụng năm thuật toán học máy đa dạng cùng dữ liệu EBMT để phân tích các yếu tố ảnh hưởng đến tử vong liên quan HSCT [12] và phát triển mô hình dự đoán tử vong sau 100 ngày ghép [13]. Nazha et al. [14] đề xuất mô hình cá nhân hóa dự báo ở nhiều mốc thời gian sau ghép tế bào gốc tạo máu. Ngoài ra, Jangir et al. [15] ứng dụng mạng CNN liên kết chức năng cho phân loại bệnh tiểu đường. Các nghiên cứu của Choubey và cộng sự trình bày hệ thống lai phân loại tiểu đường [16], triển khai nhiều thuật toán phân loại khác nhau [17], phân tích so sánh sử dụng và không sử dụng các phương pháp phân loại [18], đánh giá hiệu quả phân loại bằng PCA và PSO [19], cũng như sử dụng mạng nơ-ron RBF và thuật toán GA_RBF cho phân loại tiểu đường [26]. Lấy cảm hứng từ các công trình trên, ý tưởng nghiên cứu dự đoán tử vong liên quan HSCT được hình thành. Các nhà nghiên cứu khác đã bàn luận về ứng dụng tính toán mềm trong dự báo ung thư vú [20], phân tích bạch cầu bằng học máy và khai phá dữ liệu [21], và phân tích dự báo các kết quả liên quan gan bằng các kỹ thuật tính toán mềm [22]. Choubey và Paul [23] cùng Srivastava et al. [25] thực hiện phân tích so sánh học máy (ML), học sâu (DL) và tính toán mềm cho bệnh tiểu đường và tim mạch, tập trung vào hiệu suất, lợi ích và thách thức của các phương pháp này. Trong bối cảnh nghiên cứu đó, đề tài dự đoán tử vong liên quan ghép tế bào gốc tạo máu được khởi xướng.

Different decision trees, like the J48 tree, the random tree, the REP tree, the SOM, logistic regression, and naïve Bayes, were used and compared to find dengue early on [31]. Decision tree classifiers have been utilized to perform the classification of dengue disease [31]. Machine learning, big data, and IoT have been analyzed, compared, and implemented in the healthcare sector [32]. Several classification techniques, namely Naïve Bayes, MLP, Voted Perceptron, and J48, were utilized for two data sets (sick and

breast cancer), and their performance was assessed and compared with existing methods [33]. CNN+LSTM techniques were proposed for detecting hate speech and offensive language, outperforming other existing approaches [34]. A computational classifier on protein data, where Random Forest outperforms existing techniques, was also proposed [35]. Additionally, ML, data mining, and soft computing have been thoroughly reviewed for the classification of dengue and ovarian cancer [36, 37].

3. PROPOSED APPROACH

This study involved analysis of data of pediatric patients undergoing HSCT. The data was then prepared to be able to be fed to ML algorithms. Performance metrics of the applied ML have been compared to find the optimum model for survivability prediction. All the experiments have been performed in Jupyter Notebook with the use of Python 3.9.

There are three sub-sections in this proposed methodology: section 3.1 deals with dataset descriptions, section 3.2 deals with experimental details, including information about the applied ML algorithms, and section 3.3 discusses the workflow of the experiments.

3.1 Dataset

The data set describes pediatric individuals with numerous hematologic diseases: malignant disorders and nonmalignant instances. The unaltered allogeneic unrelated donor HSCT was performed on all patients.

The dataset was collected by Faculty of Silesian University of Technology from the period spanning 2000 to 2008, a total of 187 individuals in the pediatric and adolescent age group underwent un-manipulated allogeneic Umbilical Cord Blood HSCT. Among them, 112 were male, and 75 were female. The cohort included patients with both malignant (n = 155) and nonmalignant disorders (n = 32). The median age of these individuals at the time of transplantation stood at 9.6 years, with a range spanning from 0.6 to 20.2 years. This comprehensive dataset forms the basis for our examination of outcomes and contributes to the understanding of un-manipulated allogeneic UD HSCT in children and adolescents during the specified timeframe. The dataset comprises of 187 instances and 37 attributes [29].

3.2 Machine Learning Models

The dataset was used to train 10 ML models namely, Logistic Regression, DT, Random Forest, XGBoost, KNN, Naïve Bayes Bernoulli, SVM, Perceptron, MLP and Naïve Bayes Gaussian.

3.2.1 Logistic Regression

Logistic regression is one statistical method for binary classification problems. It uses the logistic function to model the relationship between a set of independent variables and a binary result. It forecasts the likelihood that an event will occur, which is subsequently converted into a binary outcome, in contrast to linear regression. It is frequently used to forecast outcomes like the occurrence of diseases or loan defaults in industries like medical and finance. By estimating the coefficients for every variable, logistic regression calculates how those coefficients affect the event's log-odds. It is comprehensible, simple to implement, and a fundamental technique in statistical analysis and ML. In order to diagnose diabetes, the same logistic regression methodology was employed in [18].

3.2.2 Decision Tree Classifier

Both classification and regression tasks make use of the Decision Tree (DT) Classifier. Each internal node represents a characteristic, each branch represents a decision rule, and each leaf node represents an outcome class. It traverses from the root to a leaf node to make decisions by recursively dividing the data according to features. Both qualitative and numerical data can be handled with decision trees, which are interpretable. Because of their simplicity, adaptability, and capacity to manage non-linear patterns, they are utilized in numerous sectors, such as healthcare and finance, and they record intricate relationships. Diabetes has been diagnosed using the similar decision tree classification approach in [18, 27].

3.2.3 Random Forest Classifier

One flexible ensemble learning technique in machine learning is the Random Forest Classifier. For better prediction performance, it develops various DT during training and aggregates their outputs. To improve generalization and lessen overfitting, each tree is trained on a random selection of features and data. It classifies input data by combining the choices made by individual trees using a voting method. Random Forest is appropriate for classification and regression applications, robust, and capable of managing intricate interactions. Because it can manage noisy data and produce accurate predictions, it performs well in a variety of fields, including bioinformatics and finance. In [1], the same random forest classification methodology was employed.

3.2.4 XGBoost

XGBoost is an influential and popular gradient boosting algorithm in ML. It improves the accuracy of predictions by sequentially building an ensemble of weak decision trees. XGBoost intelligently addresses bias-variance trade-offs through regularization techniques and employs gradient optimization to minimize prediction errors. It handles missing data, supports various objective functions, and can be utilized for classification, regression, and ranking tasks. XGBoost's efficiency, scalability, and feature importance analysis make it a preferred choice in competitions and real-world applications, spanning areas like finance, healthcare,

and recommendation systems. The same used algorithm of XGBoost has been used in [1].

3.2.5 KNN

KNN is an instance-based, non-parametric ML algorithm. It makes predictions by locating the 'k' closest data points to a given input, based on a defined distance metric. KNN is versatile, handling both classification and regression tasks, and adapts well to various data types. It captures local patterns and doesn't assume specific underlying distributions. However, its performance can be influenced by the 'k' selection and sensitive to noisy data. KNN finds applications in recommendation systems, anomaly detection, and pattern recognition, offering simplicity and effectiveness in tasks where local context plays a vital role. The same used algorithm of KNN has been used in [18] for the diagnosis of diabetes.

3.2.6 Naïve Bayes Bernoulli

A probabilistic classification technique used in machine learning is called Naive Bayes (NBs) Bernoulli. It's particularly suited for binary feature data, where each feature can take on only two values (usually 0 or 1). It considers that attributes are conditionally independent, simplifying calculations. It calculates the likelihood of a class based on the presence or absence of particular features by using Bayes' theorem. When word occurrences are employed as features in text classification and spam filtering, Naive Bayes Bernoulli frequently works well despite its "naive" assumption of independence. It is a useful tool in many different applications due to its ease of use, speed, and efficiency while handling high-dimensional data. The same used algorithm of NBs Bernoulli has been used in [17], [18] for the diagnosis of diabetes.

3.2.7 SVM

The SVM is a powerful supervised learning method for both regression and classification. Finding a decision surface that best divides data into discrete classes while optimizing the margin between them is SVM's main goal. By converting data into a higher-dimensional space, SVM makes it possible to find intricate associations. To establish the decision boundary, it chooses the most significant data points, called as support vectors. SVM uses kernel functions, such as polynomial or radial basis functions, to handle both linear and non-linear data. Because of its adaptability, resilience to overfitting, and capacity to manage high-dimensional data, SVM is a major force in a variety of domains, including biology, finance, and image recognition. The similar SVM method has been utilized to diagnose diabetes in [24, 34].

3.2.8 Perceptron

A basic neural network model for binary classification tasks is the perceptron. It comprises input nodes, weighted connections, and an activation function that produces an output. During training, it adjusts weights based on misclassified samples, seeking convergence towards accurate classifications. While simple and efficient, the Perceptron is limited to linearly separable data. It played a pivotal role in the development of Neural Networks, paving the way for more complex architectures. Though less capable of handling intricate patterns than modern networks, the Perceptron remains an essential concept, contributing to the evolution of DL and AI. The same used algorithm of Perceptron has been used in [28].

3.2.9 MLP

MLP is a versatile ANN architecture. It is made up of interconnected nodes in input, hidden, and output layers. Each node applies weights and an activation function to incoming data, enabling the network to capture complex relationships. MLP can handle non-linear patterns and is trained using techniques like backpropagation, adjusting weights to minimize errors. Its capacity to model intricate data makes it appropriate for several tasks, from classification and regression to image recognition. Despite its effectiveness, MLP may suffer from overfitting and requires careful tuning. It remains a cornerstone in Deep Learning, inspiring advanced Neural Network architectures and advancements in Machine Learning. The same used algorithm of MLP has been used in [17, 35] for the diagnosis of diabetes.

3.2.10 Naïve Bayes Gaussian

Naive Bayes Gaussian is a probabilistic classification algorithm that assumes features to follow a Gaussian/normal distribution. It's well-suited for continuous data, where attributes are assumed to have a bell-shaped curve. Using Bayes' theorem, it calculates the likelihood of a class given feature values, facilitating predictions. Despite its "naive" assumption of feature independence, Naive Bayes Gaussian remains efficient and can handle high-dimensional datasets. It's particularly useful in natural language processing and sentiment analysis tasks. By modeling continuous data distributions, Naive Bayes Gaussian provides a simple yet effective way to categorize data points into classes based on their attribute values. The same used algorithm of Naïve Bayes Gaussian has been used in [17], [18] for the diagnosis of diabetes.

3.3 Workflow

The dataset was first imported and analyzed using python-based libraries like NumPy, matplotlib, seaborn. After this the work has been splinted into 3 experiments.

3.3.1 Experiment I

This subset of the work performed by the authors involves the use of unaltered data to train the ML models. The

missing data has not been considered by the authors in this experiment by dropping them.

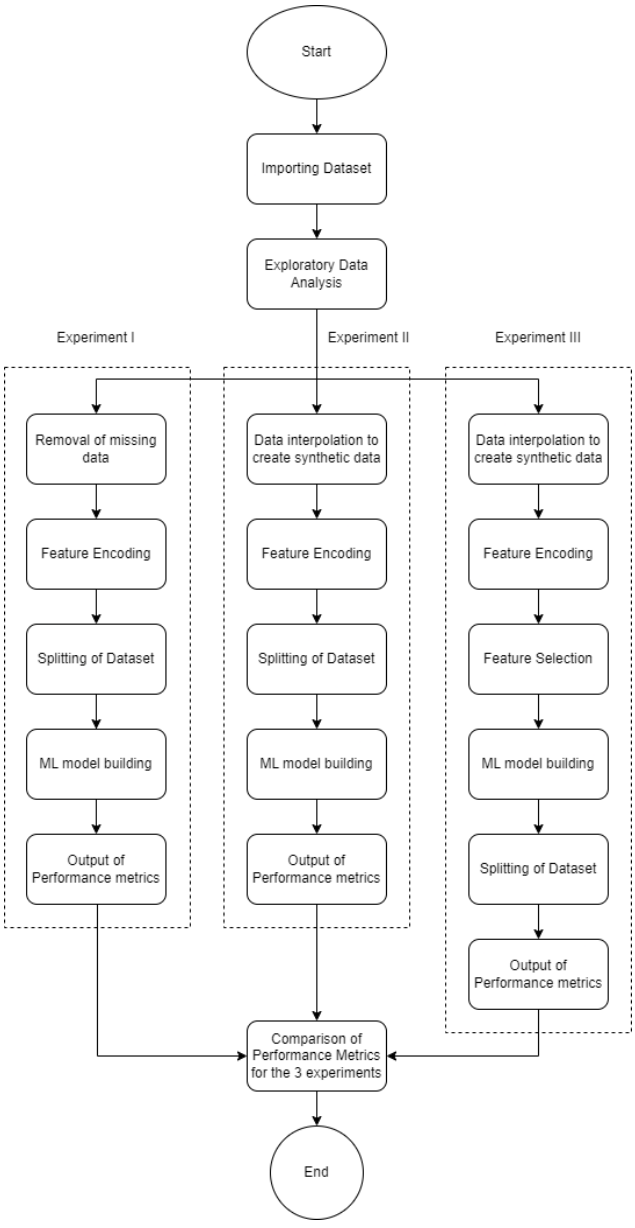

Fig. 2. Working approach of the proposed methodology.

3.3.2 Experiment II

In this experiment synthetic data has been created by the authors via the use of various data interpolation techniques to compensate for the missing data.

The working approach of the proposed methodology has been graphically represented in the Fig. 2.

3.3.3 Experiment III

The last experiment involved the use of attribute selection in addition to the use of synthetic data created in Experiment II. Chi square test was selected as the feature selection technique by the authors.

3.3.4 Chi Square Test

A statistical technique for examining the relationship between categorical variables in a dataset is the Chi-Square Test. It assesses whether observed frequencies differ significantly from expected frequencies, helping determine if variables are independent or related. It involves calculating a test statistic by comparing observed and expected counts and then comparing it to a critical value from the Chi-Square distribution. The test is majorly applied in several fields, such as social sciences, genetics, and market research, to uncover relationships or dependencies between categorical variables and provide insights into the underlying data structure.

$$x^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

where, x^2 denotes/represents the Chi-Square test statistic, O_i represents the observed frequency and E_i represents the expected frequency.

4. EXPERIMENTAL RESULTS

The results of the proposed methodology are as follows:

4.1 Experiment I

Among all the employed Machine Learning models, the Decision Tree Classifier exhibited the most favorable results on the original dataset. Notably, it achieved the highest accuracy of 93.10% and an F1 Score of 0.8.

The performance metrics for Experiment I have been presented in Table 1, while their graphical representation is depicted in Fig. 3.

Table 1. Performance measures for Experiment I

ML Algorithm	Accuracy	Precision	Recall	F1 Score
Logistic Regression	89.66%	1.000	0.5	0.667
Decision Tree Classifier	93.10%	1.000	0.667	0.8
Random Forest Classifier	89.66%	1.000	0.5	0.667
XGBoost	89.66%	1.000	0.5	0.667
KNN	86.21%	0.625	0.833	0.714
Naïve Bayes Bernoulli	82.76%	0.571	0.667	0.615
SVM	89.66%	0.714	0.833	0.769
Perceptron	79.31%	0.000	0	0
MLP	89.66%	0.800	0.667	0.727
Naïve Bayes Gaussian	27.59%	0.220	1	0.364

Fig. 3. depict the performance measures for Experiment I.

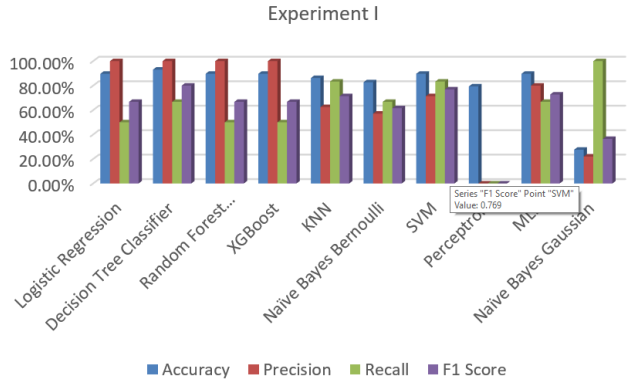

Fig. 3. Performance measures for Experiment I.

4.2 Experiment II

In the case of synthetic data, the Random Forest Classifier and XGBoost demonstrated superior performance, obtained an accuracy of 94.74% and an F1 Score of 0.944.

The performance measures for Experiment II are succinctly summarized in Table 2, while their visual depiction is showcased through Fig. 4.

Table 2. Performance measures for Experiment II

ML Algorithm	Accuracy	Precision	Recall	F1 Score
Logistic Regression	89.47%	0.895	0.895	0.895
Decision Tree Classifier	92.11%	0.9	0.947	0.923
Random Forest Classifier	94.74%	1	0.895	0.944
XGBoost	94.74%	1	0.895	0.944
KNN	89.47%	0.895	0.895	0.895
Naïve Bayes Bernoulli	68.42%	0.684	0.684	0.684
SVM	86.84%	0.937	0.789	0.857
Perceptron	73.66%	0.68	0.895	0.772
MLP	92.11%	0.944	0.895	0.919
Naïve Bayes Gaussian	60.53%	1	0.211	0.348

4.3 Experiment III

Even with reduced attributes after the addition of feature selection techniques to the synthetic data same levels of performance was observed in Experiment III that was seen in Experiment II. Highest accuracy was 94.74% and F1 Score of 0.944 was achieved by random forest classifier and xgboost.

Fig. 4. depict the performance measures for Experiment II.

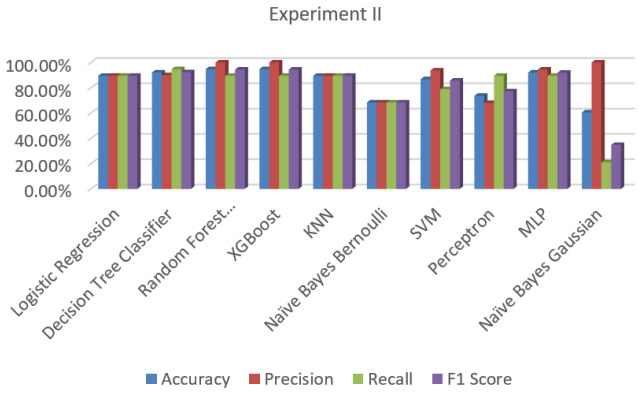

Fig. 4. Performance measures for Experiment II.

The outcomes of the implemented ML algorithms for Experiment III are succinctly presented in a tabular format within Table 3, while their graphical representation is illustrated in Fig. 5.

Fig. 5. depict the performance measures for Experiment III.

Table 3. Performance measure for Experiment III

ML Algorithm	Accuracy	Precision	Recall	F1 Score
Logistic Regression	81.58%	0.875	0.737	0.799
Decision Tree Classifier	92.11%	0.944	0.895	0.919
Random Forest Classifier	94.74%	1	0.895	0.944
XGBoost	94.74%	1	0.895	0.944
KNN	89.47%	0.895	0.895	0.895
Naïve Bayes Bernoulli	50.00%	0	0	0
SVM	89.47%	0.895	0.895	0.895
Perceptron	73.68%	0.68	0.895	0.773
MLP	92.11%	0.944	0.895	0.919
Naïve Bayes Gaussian	50.00%	0.5	1	0.667

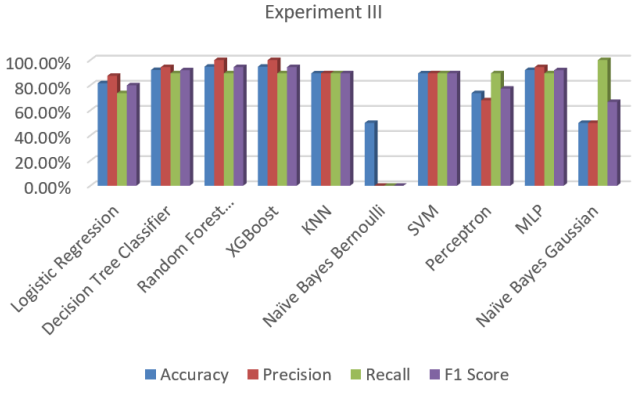

Fig. 5. Performance measures for Experiment III.

4.4 Comparison among the experiments

Here the performance of the ML algorithm across all the three experiments have been put into consideration.

Accuracy of the ML models for the three experiments have been represented in tabular in Table 4 and shown in Fig. 9.

Table 4. Comparison of accuracy

ML Algorithm	Experiment I	Experiment II	Experiment III
Logistic Regression	89.66%	89.47%	81.58%
Decision Tree Classifier	93.10%	92.11%	92.11%
Random Forest Classifier	89.66%	94.74%	94.74%
XGBoost	89.66%	94.74%	94.74%
KNN	86.21%	89.47%	89.47%
Naïve Bayes Bernoulli	82.76%	68.42%	50.00%
SVM	89.66%	86.84%	89.47%
Perceptron	79.31%	73.66%	73.68%
MLP	89.66%	92.11%	92.11%
Naïve Bayes Gaussian	27.59%	60.53%	50.00%

Fig. 6. depict the accuracy comparison of the particular algorithm.

Fig. 6. Comparison of accuracy.

The precision scores of the Machine Learning models across the three experiments are succinctly showcased in a tabular manner within Table 5. Furthermore, their visual representation is vividly depicted in Fig. 7.

Fig. 7. depict the precision comparison of the particular algorithm.

The recall values for the Machine Learning models in the three experiments have been concisely displayed in tabular format within Table 6. Correspondingly, their graphical depiction is presented in Fig. 8.

Table 5. Comparison of precision

ML Algorithm	Experiment I	Experiment II	Experiment III
Logistic Regression	1.000	0.895	0.875
Decision Tree Classifier	1.000	0.9	0.944
Random Forest Classifier	1.000	1	1
XGBoost	1.000	1	1
KNN	0.625	0.895	0.895
Naïve Bayes Bernoulli	0.571	0.684	0
SVM	0.714	0.937	0.895
Perceptron	0.000	0.68	0.68
MLP	0.800	0.944	0.944
Naïve Bayes Gaussian	0.220	1	0.5

Fig. 7. Comparison of precision.

Fig. 8. depict the recall comparison of the particular algorithm.

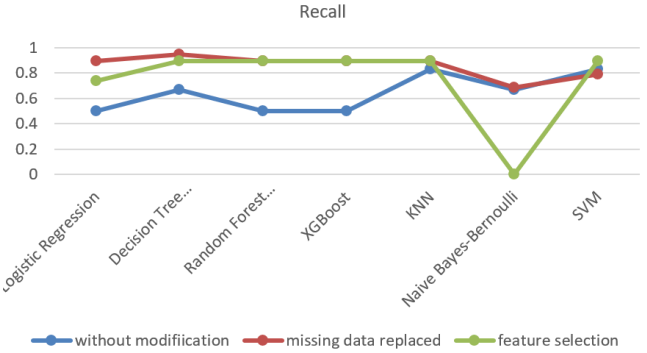

Fig. 8. Comparison of recall.

Table 6. Comparison of recall

ML Algorithm	Experiment I	Experiment II	Experiment III
Logistic Regression	0.5	0.895	0.737
Decision Tree Classifier	0.667	0.947	0.895
Random Forest Classifier	0.5	0.895	0.895
XGBoost	0.5	0.895	0.895
KNN	0.833	0.895	0.895
Naïve Bayes Bernoulli	0.667	0.684	0
SVM	0.833	0.789	0.895
Perceptron	0	0.895	0.895
MLP	0.667	0.895	0.895
Naïve Bayes Gaussian	1	0.211	1

F1 Score of the ML algorithm for the three experiments have been represented in tabular in Table 7 and shown in Fig. 9.

Table 7. Comparison of F1 Score

ML Algorithm	Experiment I	Experiment II	Experiment III
Logistic Regression	0.667	0.895	0.799
Decision Tree Classifier	0.8	0.923	0.919
Random Forest Classifier	0.667	0.944	0.944
XGBoost	0.667	0.944	0.944
KNN	0.714	0.895	0.895
Naïve Bayes Bernoulli	0.615	0.684	0
SVM	0.769	0.857	0.895
Perceptron	0	0.772	0.773
MLP	0.727	0.919	0.919
Naïve Bayes Gaussian	0.364	0.348	0.667

Fig. 9. depict the F1 score comparison of the particular algorithm.

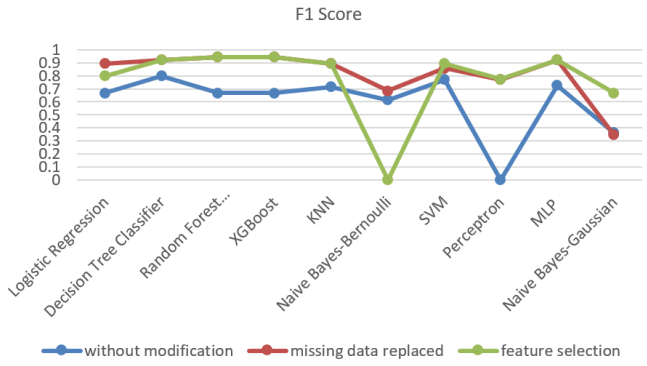

Fig. 9. Comparison of F1 score.

5. EXPERIMENT INSIGHTS, DISCUSSION AND FUTURE DIRECTIONS

The initial dataset encompassed 187 instances and 37 attributes, and this attribute count was subsequently elevated to 58 through feature encoding techniques aimed at data preparation for ML models. In Experiment I, the instance count was reduced to 142 due to the removal of missing data. In Experiment III, the authors employed chi-square for feature selection, impressively condensing attributes to a mere 6 while retaining optimal performance. Notably, the Random Forest Classifier and XGBoost exhibited superior performance, achieving a remarkable prediction accuracy of 94.74% and an F1 Score of 0.944. It was noted that ensemble-based models performed significantly higher than the other models.

The study involved an extensive review of existing research and implementations focusing on predicting patients' survival post HSCT [30] [38]. The authors meticulously examined various techniques employed in these studies, conducted performance comparisons, provided insights into motivations, and highlighted the contributions made to the field of survival prediction. The authors' work demonstrated superior performance compared to the research outlined in [1]. Despite surpassing the performance claimed in [1], the models in this study obtained similar results with reduced resource utilization, achieved through a decrease in the number of required features. Furthermore, the authors introduced prospective Machine Learning models not covered in [1], adding novel approaches to predicting post-HSCT survival.

Traditionally, medical decisions concerning HPSCT risk were collaboratively made by medical experts, sometimes aided by statistical models that furnished risk scores or indexes for informed decision-making. However, the body of research showcased the enhanced predictive capabilities of Machine Learning models while retaining interpretability. Leveraging ML-based risk prediction methods not only delivers comparable or even superior

predictive accuracy but also mitigates data, resource, and time requirements. Moreover, these methods enable the deployment of interactive tools to provide patient-specific prognostic insights. Notably, certain studies even demonstrated the potential for tailored strategies in selecting optimal patients or donors. Additionally, the accessibility and user-friendliness of ML-based approaches further bolster their appeal.

While the ML algorithms advanced by the authors and their contemporaries outperformed conventional statistical methods, avenues for refinement persist, as elucidated in the comprehensive comparative study presented in the literature review section. In the context of this study, potential enhancements involve employing a larger dataset enriched with more extensive medical data. Although the MLP Neural Network model demonstrated commendable performance, its effectiveness could potentially be augmented through larger dataset application for improved outcomes. The incorporation of other DL models could also be explored to effectively manage the expanding dimensions and intricacies inherent in medical data.

REFERENCES

[1] Jahan Ratul, I., Habiba Wani, U., Muntasir Nishat, M., Al-Monsur, A., Ar-Rafi, A. M., Faisal, F., & Ridwan Kabir, M. (2022). 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. arXiv e-prints, arXiv-2201.

[2] Karami, K., Akbari, M., Moradi, M. T., Soleymani, B., & Fallahi, H. (2021). Survival prognostic factors in patients with acute myeloid leukemia using machine learning techniques. PloS one, 16(7), e0254976.

[3] Pan, L., Liu, G., Lin, F., Zhong, S., Xia, H., Sun, X., & Liang, H. (2017). Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. Scientific Reports, 7(1), 1-9.

[4] Iwasaki, M., Kanda, J., Arai, Y., Kondo, T., Ishikawa, T., Ueda, Y., & Takaori-Kondo, A. (2022). Establishment of a predictive model for GVHD-free, relapse-free survival after allogeneic HSCT using ensemble learning. Blood Advances, 6(8), 2618-2627.

[5] Goswami, C., Poonia, S., Kumar, L., & Sengupta, D. (2019). Staging system to predict the risk of relapse in multiple myeloma patients undergoing autologous stem cell transplantation. Frontiers in oncology, 9, 633.

[6] Okamura, H., Nakamae, M., Koh, S., Nanno, S., Nakashima, Y., Koh, H., & Nakamae, H. (2021). Interactive web application for plotting personalized prognosis prediction curves in allogeneic hematopoietic cell transplantation using machine learning. Transplantation, 105(5), 1090-1096.

[7] Taati, B., Snoek, J., Aleman, D., & Ghavamzadeh, A. (2013). Data mining in bone marrow transplant records to identify patients with high odds of survival. IEEE journal of biomedical and health informatics, 18(1), 21-27.

[8] Choi, E. J., Jun, T. J., Park, H. S., Lee, J. H., Lee, K. H., Kim, Y. H., & Lee, J. H. (2022). Predicting Long-term Survival After Allogeneic Hematopoietic Cell Transplantation in Patients with Hematologic Malignancies: Machine Learning–Based Model Development and Validation. JMIR Medical Informatics, 10(3), e32313.

[9] Iwasaki, M., Kanda, J., Arai, Y., Kondo, T., Ishikawa, T., Ueda, Y., & Takaori-Kondo, A. (2022). Establishment of a predictive model for GVHD-free, relapse-free survival after allogeneic HSCT using ensemble learning. Blood Advances, 6(8), 2618-2627.

[10] Eisenberg, L., XplOit consortium, Brossette, C., Rauch, J., Grandjean, A., Ottinger, H., ... & Turki, A. T. (2022). Time-dependent prediction of mortality and cytomegalovirus reactivation after allogeneic hematopoietic cell transplantation using machine learning. American Journal of Hematology, 97(10), 1309-1323.

[11] Shouval, R., Nagler, A., Labopin, M., & Unger, R. (2015). Interpretable boosted decision trees for prediction of mortality following allogeneic hematopoietic stem cell transplantation. J Data Mining Genomics Proteomics, 6(4), 2.

[12] Shouval, R., Labopin, M., Unger, R., Giebel, S., Ciceri, F., Schmid, C., & Nagler, A. (2016). 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, 11(3), e0150637.

[13] Shouval, R., Labopin, M., Bondi, O., Mishan Shamay, H., Shimoni, A., Ciceri, F., & Mohty, M. (2015). 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, 2015, vol. 33, num. 28, p. 3144-3151.

[14] Nazha, A., Hu, Z. H., Wang, T., Lindsley, R. C., Abdel-Azim, H., Aljurf, M., & Saber, W. (2020). A personalized prediction model for outcomes after allogeneic hematopoietic cell transplant in patients with myelodysplastic syndromes. Biology of blood and marrow transplantation, 26(11), 2139-2146.

[15] Jangir, S. K., Joshi, N., Kumar, M., Choubey, D. K., Singh, S., & Verma, M. (2021). Functional link convolutional neural network for the classification of diabetes mellitus. International Journal for Numerical Methods in Biomedical Engineering, 37(8), e3496.

[16] Choubey, D. K., & Paul, S. (2016). GA_MLP NN: a hybrid intelligent system for diabetes disease diagnosis. International Journal of Intelligent Systems and Applications, 8(1), 49-59.

[17] Choubey, D. K., Paul, S., Shandilya, S., & Dhandhania, V. K. (2020). Implementation and analysis of classification algorithms for diabetes. Current Medical Imaging, 16(4), 340-354.

[18] Choubey, D. K., Kumar, M., Shukla, V., Tripathi, S., & Dhandhania, V. K. (2020). Comparative analysis of classification methods with PCA and LDA for diabetes. Current diabetes reviews, 16(8), 833-850.

[19] Choubey, D. K., Kumar, P., Tripathi, S., & Kumar, S. (2020). Performance evaluation of classification methods with PCA and PSO for diabetes. Network Modeling Analysis in Health

Informatics and Bioinformatics, 9(1), 1-30.

[20] D. Sharma, P. Jain, and D. K. Choubey, "A Comparative Study of Computational Intelligence for Identification of Breast Cancer," in International Conference on Machine Learning, Image Processing, Network Security and Data Sciences, 2020, pp. 209–216.

[21] A. Parthvi, K. Rawal, and D. K. Choubey, "A Comparative study using Machine Learning and Data Mining Approach for Leukemia," Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020, pp. 672–677, 2020, doi: 10.1109/ICCSP48568.2020.9182142.

[22] S. Pahari and D. K. Choubey, "Analysis of Liver Disorder Using Classification Techniques: A Survey," Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020, pp. 1–4, 2020, doi: 10.1109/ic-ETITE47903.2020.300.

[23] Choubey, D. K., & Paul, S. (2016). Classification techniques for diagnosis of diabetes: a review. International Journal of Biomedical Engineering and Technology, 21(1), 15-39.

[24] Choubey, D. K., Tripathi, S., Kumar, P., Shukla, V., & Dhandhania, V. K. (2021). Classification of Diabetes by Kernel based SVM with PSO. *Recent Advances in Computer* Science and Communications (Formerly: Recent Patents on Computer Science), 14(4), 1242-1255.

[25] Srivastava, K., & Choubey, D. K. (2021). Soft Computing, Data Mining, and Machine Learning Approaches in Detection of Heart Disease: A Review. In Hybrid Intelligent Systems: 19th International Conference on Hybrid Intelligent Systems (HIS 2019) held in Bhopal, India, December 10-12, 2019 19 (pp. 165-175). Springer International Publishing.

[26] Choubey, D. K., & Paul, S. (2017). GA_RBF NN: a classification system for diabetes. International Journal of Biomedical Engineering and Technology, 23(1), 71-93.

[27] Choubey, D. K., Paul, S., Bala, K., Kumar, M., & Singh, U. P. (2019). Implementation of a hybrid classification method for diabetes. In Intelligent Innovations in Multimedia Data Engineering and Management (pp. 201-240). IGI Global.

[28] Sivanandam, S. N., & Deepa, S. N. (2013). Principles of Soft Computing. Retrieved from https://books.google.co.in/books?id=RisdlAEACAAJ.

[29] A. Gudys, *Bone marrow transplant children dataset*, Kaggle, 2020. Available: https://www.kaggle.com/datasets/adamgudys/bone-marrow-transplant-children. (Accessed: Sept. 2025).

[30] Hanselia, R., & Choubey, D. K. (2023, February). A Comparative Study for Prediction of Hematopoietic Stem Cell Transplantation-Related Mortality. In International Conference On Innovative Computing and Communication (pp. 641-652). Singapore: Springer Nature Singapore.

[31] Gambhir, S., Kumar, Y., Malik, S., Yadav, G., & Malik, A. (2019). Early diagnostics model for dengue disease using decision tree-based approaches. In Pre-Screening Systems for Early Disease Prediction, Detection, and Prevention (pp. 69-87). IGI Global.

[32] Kumar, Pradeep., Kumar, Yugal., and Tawhid, Mohammed A. (2021). Machine Learning, Big Data, and IoT for Medical Informatics. Academic Press.

[33] Kumar, Y., & Sahoo, G. (2012). Analysis of Bayes, neural network and tree classifier of classification technique in data mining using WEKA. Computer Science & Information Technology.

[34] Verma, A., Singh, A., Bihari, A., Tripathi, S., Agrawal, S., Pandey, S. K., & Verma, S. (2023). Identification of Hate Speech on Social Media using LSTM. GMSARN International Journal, 17, 468-474.

[35] Bhardwaj, S. K., Vishwakarma, S., Bihari, A., Tripathi, S., Agrawal, S., & Joshi, P. (2024). Protein Enzyme Sequence Class Prediction using Computational Model. GMSARN International Journal, 18, 62-70.

[36] Choubey, D. K., Choubey, A., Mahto, S., & Soni, V. (2024). Soft Computing Approaches for Ovarian Cancer: A Review. GMSARN International Journal, 18, 223-239.

[37] Choubey, D. K., Newton, R., Ojha, M. K., & Kumar, S. (2023, February). Soft Computing and Data Mining Techniques for Dengue Detection: A Review. In International Conference On Innovative Computing And Communication (pp. 295-309). Singapore: Springer Nature Singapore.

[38] Hanselia, R., Choubey, D. K., Bala, K., & Mishra, A. (2024). Transfer Learning for Multiclass Classification of Bone Marrow Cells. In Machine Learning and IoT Applications for Health Informatics (pp. 184-215). CRC Press.