Review of Bone Marrow Transplantation Predictive Model

[1] Reena Rani   [2] Dr. Dinesh Kumar
[1] M.Tech. CSE Scholar [2] Professor
Department of Computer Science & Engineering
BRCM College of Engineering & Technology, Bahal, Bhiwani, India
[1]Reenasharma5544@gmail.com
[2] hodcse@brcm.edu.in

Abstract: Organ transplantation is a life-saving procedure that has transformed the field of medicine. Despite significant advancements in transplantation techniques and immunosuppressive therapies, challenges still remain, including organ availability and the risk of rejection. In this review paper, we provide a comprehensive overview of the Bone Marrow Transplantation Predictive Model, a novel predictive model that aims to enhance the accuracy of predicting transplant outcomes. We summarize the existing literature on transplantation predictive models, highlight the unique features and strengths of the Bone Marrow model, and discuss its potential implications for improving transplant success rates and patient outcomes.

Keywords: Bone Marrow, predictive model, organ transplantation, transplant outcomes, clinical factors.

I. INTRODUCTION

Bone marrow, a vital component of the human body, serves as the primary site for hematopoiesis and houses a rich source of multipotent stem cells. These specialized cells, known as bone marrow stem cells (BMSCs), have garnered significant attention in the field of regenerative medicine due to their remarkable ability to differentiate into various cell types and promote tissue repair. The exploration of BMSCs has unfolded a promising avenue for treating a wide range of diseases and injuries, including hematological disorders, autoimmune conditions, and musculoskeletal injuries.

Understanding the timeline of bone marrow development is crucial in comprehending the complexity and significance of BMSCs. During embryogenesis, hematopoiesis occurs in the yolk sac, followed by the fetal liver and spleen. However, it is in the bone marrow where the production of blood cells and the presence of BMSCs become prominent throughout postnatal life. This timeline underscores the critical role that bone marrow plays in the continuous replenishment of the body's blood cells and the regenerative potential of BMSCs.

The function of BMSCs extends beyond hematopoiesis, as these cells possess unique properties that make them attractive for therapeutic applications. BMSCs not only have the ability to self-renew but also exhibit immunomodulatory effects and

secrete various growth factors, cytokines, and extracellular vesicles that promote tissue regeneration and modulate the immune response. Harnessing the therapeutic potential of BMSCs requires a comprehensive understanding of their behavior and interactions within the complex biological systems.

Machine learning approaches, such as the Bone Marrow Transplantation Predictive Model, have emerged as powerful tools to analyze and predict the behavior of BMSCs. By leveraging large datasets encompassing molecular profiles, clinical parameters, and transplantation outcomes, these models aim to uncover patterns and correlations that can enhance our understanding of BMSC function and optimize therapeutic strategies. The integration of machine learning techniques with the vast biological knowledge about BMSCs has the potential to revolutionize the field of regenerative medicine and improve patient outcomes.

Machine learning, a subfield of artificial intelligence, encompasses various techniques and algorithms that enable computers to learn from data and make predictions or decisions without explicit programming. In the context of your research paper, discussing the types of machine learning can provide a comprehensive understanding of the methodologies employed in the Bone Marrow Transplantation Predictive Model. Here are some key types of machine learning:
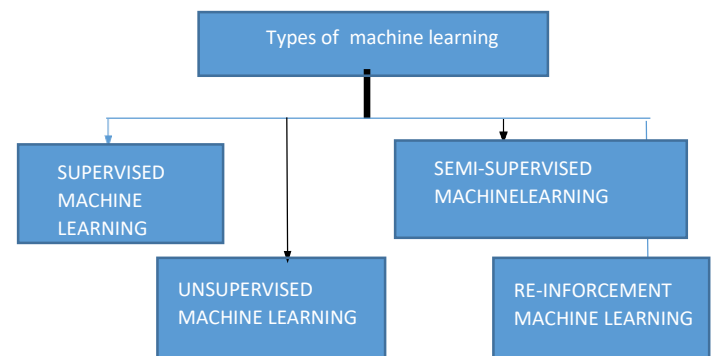
Supervised Learning: This type of machine learning involves training a model using labeled data, where the input variables (features) are mapped to known output variables (labels or target

variables). The model learns the underlying patterns in the data and can then make predictions on new, unseen data. Examples of supervised learning algorithms include linear regression, decision trees, support vector machines, and neural networks.

Unsupervised Learning: Unsupervised learning deals with unlabeled data, where the model's objective is to discover patterns, relationships, or structures within the data without any predefined output labels. Common unsupervised learning techniques include clustering algorithms like k-means clustering, hierarchical clustering, and dimensionality reduction techniques like principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE).

Semi-Supervised Learning: This type of learning lies between supervised and unsupervised learning. It utilizes both labeled and unlabeled data during the training process. Typically, the labeled data is limited, while the unlabeled data is abundant. Semi-supervised learning algorithms aim to leverage the unlabeled data to improve the model's performance and generalization.

Reinforcement Learning: Reinforcement learning involves training an agent to interact with an environment and learn from the feedback received in the form of rewards or penalties. The agent learns to take actions that maximize cumulative rewards over time. This type of learning is often used in scenarios where an agent learns to play games, control robotic systems, or optimize decision-making processes.

Deep Learning: Deep learning is a subset of machine learning that focuses on the development and training of deep neural networks, which are neural networks with multiple hidden layers. Deep learning has demonstrated remarkable success in various domains, such as image recognition, natural language processing, and speech recognition. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are examples of deep learning architectures.

Transfer Learning: Transfer learning involves leveraging knowledge learned from one task or domain to improve learning and performance in another related task or domain. By transferring pre-trained models or knowledge, transfer learning can significantly reduce the need for large labeled datasets and improve learning efficiency.

Ensemble Learning: Ensemble learning combines multiple individual models (often referred to as base learners) to make predictions or decisions. The ensemble model aggregates the predictions from the base models, which often leads to better overall performance. Popular ensemble learning methods include random forests, gradient boosting, and stacking.

It is important to note that the Bone Marrow Transplantation Predictive Model may utilize specific types of machine learning algorithms or a combination of different approaches. Understanding these types of machine learning provides a foundation for discussing the methodologies employed in the model and their implications for transplantation predictions.

Machine learning algorithms:

Machine learning algorithms have revolutionized the field of data analysis and prediction. Among the various algorithms, linear regression is a widely used technique for predicting a continuous output variable based on input variables. It assumes a linear relationship between the inputs and the output. On the other hand, logistic regression is specifically designed for binary classification problems, estimating the probability of an instance belonging to a particular class.

Decision trees are popular for their ability to partition data into hierarchical structures based on attribute values, allowing for efficient classification or regression. Random forests take this concept further by combining multiple decision trees to make predictions. Each tree is constructed independently using a random subset of the training data, and the final prediction is obtained by aggregating the predictions of all the trees.

Support Vector Machines (SVM) are powerful algorithms used for both classification and regression tasks. They construct hyperplanes that maximize the margin between different classes or regression points. Bayes naive, based on Bayes' theorem, is a simple yet effective probabilistic classifier that assumes feature independence, simplifying probability calculations.

KNN, or nearest neighbors, is a versatile algorithm used for classification and regression. It predicts values based on the majority vote or average of the values of the k nearest neighbors in the training set. Neural systems, inspired by the human brain, consist of interconnected nodes organized in layers. They excel in learning complex patterns and relationships in data.

Inclination helping machines, or gradient boosting machines, sequentially combine weak prediction models to create a strong predictive model. Clustering techniques, on the other hand, are unsupervised learning algorithms used to identify groups or clusters within a dataset. K-means clustering partitions data into a predetermined number of clusters, while hierarchical clustering builds a hierarchy of clusters based on similarity measures. DBSCAN groups together data points that are close to each other, separating outliers or noise points. Gaussian Mixture Models (GMM) assume the data is generated from a mixture of Gaussian distributions and can be used for clustering and density estimation tasks.

Overall, these algorithms represent a diverse set of tools for analyzing and interpreting data, enabling us to gain valuable insights and make accurate predictions in various domains.

Feature selection:

Feature selection plays a crucial role in machine learning and data analysis. It refers to the process of selecting a subset of relevant features from a larger set of variables or attributes in a dataset. The need for feature selection arises due to several reasons.

Firstly, feature selection helps to improve model performance and generalization. Including irrelevant or redundant features in the model can lead to overfitting, where the model becomes too specialized to the training data and performs poorly on unseen data. By selecting only the most informative and relevant features, we can reduce the complexity of the model and enhance its ability to generalize to new instances.

Secondly, feature selection helps to reduce computational complexity. Large datasets with a high number of features can significantly increase the computational requirements for training and testing models. By eliminating irrelevant or redundant features, we can reduce the dimensionality of the dataset and streamline the computational processes, making them more efficient and manageable.

Thirdly, feature selection aids in improving model interpretability. When dealing with a large number of features, it becomes challenging to understand the underlying patterns and relationships captured by the model. By selecting a smaller subset of features, we can simplify the model and make it more interpretable, allowing us to gain insights and extract meaningful information from the data.

Furthermore, feature selection can contribute to data preprocessing and data cleaning. It helps in identifying and removing features that are highly correlated, noisy, or contain missing values. By eliminating such problematic features, we can enhance the quality and reliability of the data, leading to more accurate and robust models.

In addition, feature selection can facilitate domain knowledge integration. Domain experts often have valuable insights into which features are more relevant for a specific problem. By involving them in the feature selection process, we can leverage their expertise to identify and prioritize the most informative features, leading to improved model performance and decision-making.

Overall, feature selection is crucial for optimizing model performance, reducing computational complexity, improving interpretability, enhancing data quality, and incorporating domain knowledge. It allows us to focus on the most relevant aspects of the data, leading to more effective and efficient machine learning models.

Literature Survey:

Gourisaria et al. [2022] proposed in their research paper the prediction of the survival status of patients using machine learning algorithms with the help of feature selection techniques. they used two feature selection processes which are Fuzzy Discernibility Matrix and Principal Component Analysis. they gained better results using FDM as compared to the other feature selection techniques [1].

Taati et al. [2014] examined on Bone Marrow transplant risk factors. They collected the record of the transplant procedure from 1751 patients. They used collaborative filtering techniques to collect records with missing values [2].

Bosewell et al. [2015] worked on human leukocyte antigens. They proposed that bone marrow transplant is the fundamental source of human leukocyte antigen type match for patients and donors. They proposed a web-based virtual data integration system to improve the human leukocyte antigen matching process for donors [3].

Odeh et al. [2018] worked on specimen tracking or cancer centers. They tried to explain the lab specimen states and identify them in the specimen tracking system without taking or thinking about the lab business process carefully for a long time. For this one stakeholders and the accounting department are responsible to charge for the services. Not thinking for a long time it creates a gate between the current identified states and the lab's process. This paper informed about the work of specimen tracking from lab business processes. It also studied the cell therapy and Bone marrow transplant lab at a cancer center in Jordan. Also, it contributes to enhancing the quality of records in lab data and cancer center data [4].

Ramalhete et al. [2019] proposed that the methods that evaluate Lymphocyte function play an important role in patients with bone marrow transplants, organ transplantation, and therapy. He worked with an infrared spectroscopy-based method[5].

Gresh et al. [2002] worked on the creation of an application that can monitor and analyze the patient record. This application is used to visualize and understand the record of the patient. They can be understood as a huge collection of patient records data of bone marrow transplants in Israel. This application was based on an information visualization toolkit, opal developed at IBM Watson Research Center. This tool kit is very flexible to use. This application is very good for information visualization techniques. This application is very good to deal with unfamiliar data. This application is very good to deal with records and creates a useful, flexible, and iterative information environment [6].

Liu et al. [2017] proposed the first deep-reinforcement learning framework to find the optimal Dynamic Treatment Regimes from medical records. This framework is very flexible for state spaces to model real life in a disease. This framework is more flexible to use in heterogeneous disease propagation according to doctors' and patients' decisions and recommendations. The supervised learning step also belongs to this learning framework to estimate the actions and long- term value function. They implemented this framework on a recorded dataset from the Center for International bone marrow

transplantation research. They found the result with promising accuracy [7].

Chen et al. [2010] proposed that vasculature tumors. This is a type of tumor that can occur anywhere on the skin or in the skin tissues. In this paper, to assisting improving tumor therapy, the features of tumor vascular patterns between the two vasculature pathways are investigated. They proposed the potential of high-frequency ultrasound 3 D Doppler images for assisting the pathways of the tumor vasculatures in recurrent and primary tumors. In this process, the Murine tumor is transplanted into non-irradiated and pre-irradiated tissues to simulate the primary and recurrent tumor models. After the 7th day, taking inoculation, Doppler mode images are captured to reconstruct 3-D tumor vasculature images. Recurrent tumors show the blood vessels were wider than the primary tumor. This study is about the feature of patterns in tumor vasculature and improving the curability of recurrent tumors [8].

Yali et al. [2011] studied the treatment of Myocardial infarction using stem cell transplantation from an umbilical cord source. This study explores the feasibility of the treatment of stem cells by using a new source umbilical cord. The body that has Mesenchymal stem cells, is transplanted by using the human umbilical cord. After 8 weeks, take out the heart of the rat model to check the

transplantation of mesenchymal stem cells if the result of this is Myosin heavy chain it means stem cells are alive or transplantation of mesenchymal stem cells is able to survive. This research discovered a new source for seed cells because of the limited application of mesenchymal stem cells coming from bone marrow [9].

Parekkadan et al. [2009] proposed methods to derive or manipulate targets and prepare the stem cells for medical use. These methods will help humans to derive and test human embryonic stem cells, analyzing bone marrow stem cells' function. In this study, an explanation of many types of methods to assist the image of stem cell transplant, and cryopreserve stem cells, using microscale techniques [10].

Nishio et al. [2010] examined the study of wound healing. This is the major problem faced by elderly patients. Normal wound healing is the category of an inflammatory reaction means that is spread by bacteria into the skin tissue. These causes are swelling and redness on the skin. This inflammatory reaction is categorized by sequential infiltration of neutrophils, macrophages, and lymphocytes. this study examined the effects of monocyte lineage cells on wound healing. Here they transplanted bone marrow cells from GFP mice to C57BL/6 Mice. They found the rate of healing wound repair was improved by bone marrow transplantation. Finally old bone marrow cells also improved wound repair by bone marrow transplant. They check the rate of old bone marrow to new bone marrow. The rate of wound repair in old bone marrow was lower than in the transplant of young bone marrow. The results show that wound healing is controlled by monocyte lineage

cells[11].

Xie et al. [2008] studied bone segmental bone defects by marrow stromal cell transplant with demineralized bone matrix. This study examined the construction of engineered tissue in a large animal model. In this study, marrow stromal cells were seeded or transplanted with the help of a demineralized bone matrix for bone regeneration of large segmentation bone defects [12].

Result and Discussion:

This section illustrates the outcomes of our use of six machine learning methods to perform some metrics, such as correlation coefficient, coefficient determinant, mean absolute error, and root mean squared error.

The correlation coefficient measures the strength and direction of the linear relationship between two variables, providing insights into their association.

Table:1 Correlation Coefficient

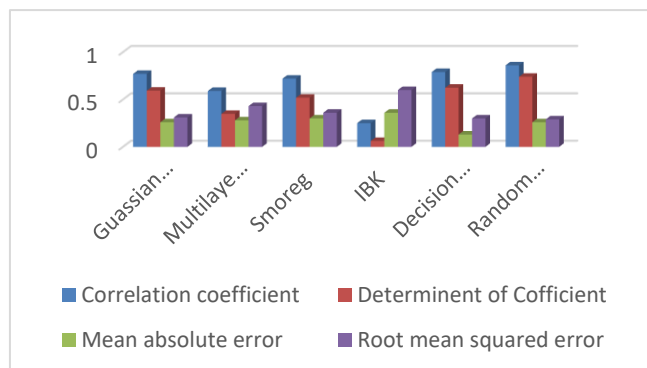| Algorithm	Correlation coefficient	Determinent of Coefficient	Mean absolute error	Root mean squared error
Guassian Process	0.77	0.59	0.26	0.31
Multilayer Perceptron	0.59	0.35	0.28	0.43
Smoreg	0.72	0.52	0.3	0.36
IBK	0.25	0.06	0.36	0.6
Decision Table	0.79	0.62	0.13	0.3
Random Forest	0.86	0.74	0.26	0.29

Figure:1 Correlation Coefficient

We can examine various correlation coefficient numbers in the graph.

Coefficients represent the slope or the rate of change of the predictor variables. They indicate the impact or contribution of

each variable on the outcome variable.

Table:2 Comparing Coefficient

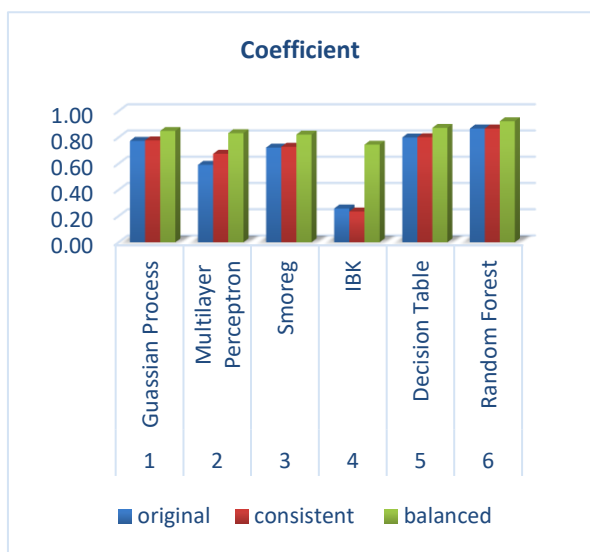| Algorithm	original	consistent	balanced
Guassian Process	0.77	0.78	0.85
Multilayer Perceptron	0.59	0.68	0.83
Smoreg	0.72	0.73	0.82
IBK	0.26	0.24	0.74
Decision Table	0.80	0.80	0.87
Random Forest	0.87	0.87	0.92

Figure:2 Comparing Coefficient

We can examine various Coefficient numbers in the graph.

Table:3 Accuracy

Algorithm	Original	Consistent	Balanced
Guassian Process	73.87	74.2	79.21
Multilayer Perceptron	71.99	75.02	85.09
Smoreg	69.53	74.25	80.2
IBK	63.64	70.25	86.91
Decision Table	86.57	87.24	90.86
Random Forest	73.93	79.85	84.72

The simplest metric to assess the accuracy of a forecast is Mean Absolute Error. The MAE, or mean of the absolute errors, is precisely what its name implies. The absolute error is the difference among the projected value, the real value, and the real value, presented as a complete and comprehensive number.

Conclusion:

We prepare our model to support bone marrow transplantation. A decent model's performance success rate is a crucial component. Numerous researchers put forth their well-known bone marrow transplant-related studies, methodologies, technologies, and models. As a result, we read out the number of earlier articles written by various researchers. After evaluating their performance, we must ptimise for the bone marrow transplant problem. Here, we used six machine learning algorithms to investigate several parameters, including mean absolute error, accuracy, etc.
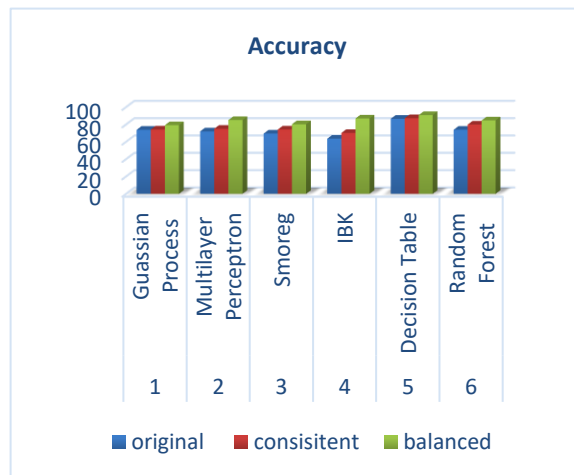

Figure:3 Accuracy

In this figure there is a representation if accuracy according to different algorithms

REFERENCES

1.   M. K. Gourisaria, A. V. Patel, R. Chatterjee and B. Sahoo, "Predicting the Survival Status of Patient after Bone Marrow Transplant Using Fuzzy Discernibility Matrix," 2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON), Raigarh, Chhattisgarh, India, 2023, pp. 1-6, doi: 10.1109/OTCON56053.2023.10114043.

2.   B. Taati, J. Snoek, D. Aleman and A. Ghavamzadeh, "Data Mining in Bone Marrow Transplant Records to Identify Patients With High Odds of Survival," in IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 1, pp. 21-27, Jan. 2014, doi: 10.1109/JBHI.2013.2274733.

3.   D. Boswell, B. Keltch, N. Roberts, X. Qian and C. Bayrak, "Match-Me:

A web based Human leukocyte antigen matching environment," 2015 World Congress on Information Technology and Computer Applications (WCITCA), Hammamet, Tunisia, 2015, pp. 1-4, doi: 10.1109/WCITCA.2015.7367040.

4.   Y. Odeh, D. Tbaishat, S. Sharma, F. Aburub, A. Tbakhi and M. Odeh, "Using Business Process Models to Inform Deriving State Transition Diagrams: Applied to BMT Lab in a Cancer Organization," 2018 1st International Conference on Cancer Care Informatics (CCI), Amman, Jordan, 2018, pp. 180-187, doi: 10.1109/CANCERCARE.2018.8618238.

5.   L. Ramalhete and C. R. C. Calado, "Assessing the molecular fingerprint of T lymphocyte activation," 2019 IEEE 6th Portuguese Meeting on Bioengineering (ENBENG), Lisbon, Portugal, 2019, pp. 1-3, doi: 10.1109/ENBENG.2019.8692471.

6.   D. L. Gresh, D. A. Rabenhorst, A. Shabo and S. Slavin, "PRIMA: A case study of using information visualization techniques for patient record analysis," IEEE Visualization, 2002. VIS 2002., Boston, MA, USA, 2002, pp. 509-512, doi: 10.1109/VISUAL.2002.1183817.

7.   Y. Liu, B. Logan, N. Liu, Z. Xu, J. Tang and Y. Wang, "Deep Reinforcement Learning for Dynamic Treatment Regimes on Medical Registry Data," 2017 IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, USA, 2017, pp. 380-385, doi: 10.1109/ICHI.2017.45.

8.   J. -J. Chen, Y. -H. Lin, C. -S. Chiang, J. -H. Hong and C. -K. Yeh, "Characterization of tumor vasculature derived from angiogenesis and vasculogenesis by high-frequency three-dimensional Doppler ultrasound," 2010 IEEE International Ultrasonics Symposium, San Diego, CA, USA, 2010, pp. 2319-2322, doi: 10.1109/ULTSYM.2010.5935993.

9.   Z. Yali and X. Wei, "The experimental study on treatment of myocardial infarction using mesenchymal stem cell transplantation from umbilical cord source," Proceedings 2011 International Conference on Human Health and Biomedical Engineering, Jilin, China, 2011, pp. 303-306, doi: 10.1109/HHBE.2011.6027959.

10.   Biju Parekkadan; Martin Yarmush, Methods in Bioengineering: Stem Cell Bioengineering , Artech, 2009.

11.   N. Nishio, S. Ito, Y. Okawa and K. -i. Isobe, "The wound repair is control by monocyte linage cells," 2010 International Symposium on Micro-NanoMechatronics and Human Science, Nagoya, Japan, 2010, pp. 278-282, doi: 10.1109/MHS.2010.5669544.

12.   Z. Xie, F. Luo, J. -Z. Xu and S. -w. Dong, "Repaire of Large Segmental Bone Defects by Marrow Stromal Cells Transplant with Demineralized Bone Matrix," 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, Shanghai, China, 2008, pp. 1042-1045, doi: 10.1109/ICBBE.2008.255.

13.   E. C. Weiss, P. Anastasiadis, C. Hildebrandt, E. Gorjup and R. M. Lemor, "9D-3 Characterization of Adipogenic, Chondrogenic and Osteogenic Differentiation with Time-Resolved Acoustic Microscopy," 2007 IEEE Ultrasonics Symposium Proceedings, New York, NY, USA, 2007, pp. 809- 812, doi: 10.1109/ULTSYM.2007.207.

14.   E Donnall Thomas, Rainer Storb, Reginald A Clift, Alexander Fefer, F Leonard Johnson, Paul E Neiman, Kenneth G Lerner, Harold Glucksberg, C Dean Buckner New England Journal of Medicine 292 (17), 895-902, 1975

15.   F. Aburub, Y. Odeh, D. Tbaishat, S. Sharma, A. Tbakhi and M. Odeh, "Deriving Informational Needs from Process Models: the Case of Cell Therapy and Bone Marrow Transplant Workflow in a Cancer Center Organization," 2018 1st International Conference on Cancer Care Informatics (CCI), Amman, Jordan, 2018, pp. 110-118, doi: 10.1109/CANCERCARE.2018.8618228.