A Review Paper on Classification of Stem Cell Transplant to Identify the High Survival Rate

Ms. Jayshree S. Raju[1]
Department of I.T.
DMIETR, Wardha
Maharastra,India
jayshree90.raju@gmail.com

Mr. Prafulla L. Mehar[2]
Department of I.T.
DMIETR, Wardha
Maharastra,India
prafulla.mehar@gmail.com

Ms. Dipali P. Pethe[3]
Department of I.T.
GNIET, Nagpur
Maharastra,,India
dipali.pethe@gmail.com

***Abstract*:** A patient undergoing hematopoietic stem cell transplant faces various risk factors and has become the standard of care for congenital or acquired disorders of the hematopoietic system or with chemo-sensitive, radiosensitive or immunosensitive malignancies. Analyzing and classifying the data from past transplant can enhance the understanding of the factors leading to highest survival rates among the patients. Over the last few decades there has been tremendous use of technology in this field. Stem cell transplant remains a dangerous procedure as it requires significant infrastructure and a network of specialists from all fields of medicine. In this paper, we are using a classification algorithm known as Support Vector Machine to classify the patients who have undergone stem cell transplant with high odds of survival. We are also keeping track of information about the donors within the family and outside the family which has a direct impact in the prioritization of resources. Classification of this information is useful to create the need for a global perspective for all cell, tissue, and organ transplants and to reveal statistical structure with potential implications in evidence-based prioritization of resources. Machine-learning techniques proved useful in analyzing the correct data from various datasets as this techniques were previously been considered too complex to analyze.

Keywords- *Stem cell transplant, SVM*

*****

I. Introduction

Hematopoietic stem cell transplant is a procedure of transferring the hematopoietic stem cells from one individual to another or the return of previously harvested cells to the same individual after testing and manipulating of the cells and/or the recipient. It is the transplantation of multipotent hematopoietic stem cells, usually derived from bone marrow, peripheral blood, or umbilical cord blood. There are three types of stem cell transplant that differ in the source of the donor cells.

Autologous

Donor and recipient is the same individual. Treatment for blood and marrow cancers can include large amounts of chemotherapy and radiation. These types of therapy can damage the marrow, resulting in very low blood cell production. Autologous transplant is used to restore the damaged marrow. During remission; stem cells are harvested after intensive therapy when very few cancerous cells remain. The stem cells are frozen for later use. This type of transplant will not result in rejection or graft versus host disease.

Syngenic

Donor and recipient are identical twins. Donor cells are not rejected and recipient tissue is not attacked by donor cells (no GVHD).

Allogenic

Donor and recipient are two different individuals. There are two types of allogenic transplant; related allogenic and unrelated allogenic. Related allogenic is usually a sibling. Unrelated allogenic is usually a closely matched individual from a donor list; this is often referred to as a matched unrelated donor or MUD transplant.

The procedure injects or infuses the healthy stem cells in the body which have been damaged or destroyed during the treatment of certain type of cancer such as leukemia, lymphoma, and neuroblastoma. During the stem transplant procedure the patients are prone to various risk factors and complications such as infection and graft-versus-host disease. The hematopoietic stem cells can be transplanted in two ways: autologous transplant, in which the patient receives his or her own stem cells, and allogeneic transplant, in which stem cells are donated by another person. The stem cell transplant is done in stages so much of the information regarding the patient treatment can be stored and later retrieved to find a particular patient with high odds of survival. Analyzing and classifying this type of data can provide a serious outcome in the prioritization of resources and in donor matching. Various data mining techniques or algorithms efficiently operate on challenging datasets to provide better results

The review presented here tackles typical challenges faced during the analysis of real-time hospital records including missing data, incomplete data, and contradictory information in the records. The support vector machine algorithm has been considered as a good classifier because of its high generalization performance. The aim of

SVM is to find the best classification function to distinguish between members of the two classes in the training data. The data obtained after analysis can be useful to predict the patients survival based on the pretransplant attributes. Often there are occasional discrepancies as the dataset is generated by gathering information from multiple sources which are subject to data entry errors.

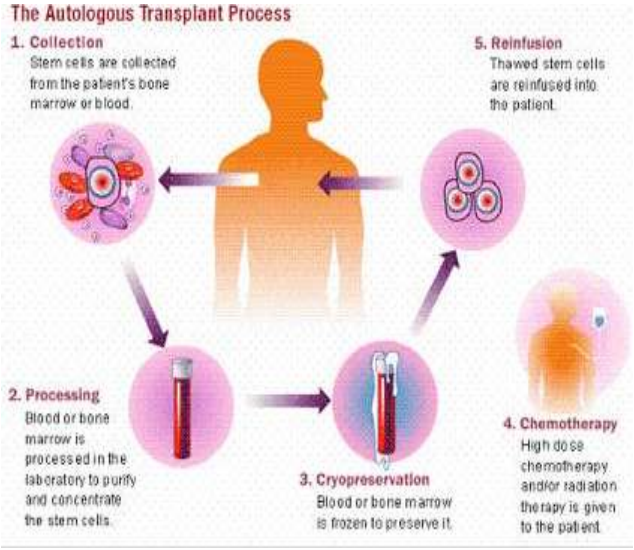

Fig: Autologous Transplant Process

This paper is organized as follows. Section II, address the survey of existing methods. In Section III, proposed work has been described. The concluding remarks are in Section IV.

II. Background

Most of the previous techniques focused on the prediction of graft-versus-host disease and much information is not published in the medical terminology exploring the use of data mining techniques in classifying the stem cell transplant records to identify matching donors. Most of the available techniques analyze the data to correctly find a donor so the cases of graft-versus-host disease can be reduced. In [1] which gives the information about the need for creating awareness about stem cell transplant over worldwide perspective. The information about the stem cell transplant records are often merged from different resources so the data is not available for analysis. The ten-fold cross validation technique is used basically to evaluate the performance of various classification techniques or algorithms to predict the results related to research [2].One of the difficult tasks in KDD process is the selection of correct classification algorithm. This work gives an idea to a selection of better data mining technique for solving a particular problem. The selection of classification algorithm is important because they provide a solution to a problem with effective results which can be useful in analysis.

The data mining techniques or the algorithms are parameter based and there is a need to set various parameters for the algorithm to analyze the better results. If the parameters are not set properly then the algorithm will to find true patterns or it can define new patterns which does not exist. The data mining techniques or algorithms should have minimum parameters set and should generate true patterns [4]. In [5] which gives information about the complications during the stem cell transplant and provides a solution to reduce the infections. In allogeneic transplant, where the stem cells are taken from another person there is a risk that the patient may get infected. This can lead to graft-versus-host disease which causes the patient to suffer more after the stem cell transplant. In one study [6], which tells about the decrease of human involvement in examining data mining methods and gives information about gene profile expressions. The gene profile expressions if used require extensive manual analysis and potential costs, which is not suitable for clinical applications. In [7] which gives information about support vector machine algorithm. SVM algorithm has been developed as a tool for classification and regression in noisy and complex domains. The key features of support vector machine are the generalization theory and the kernel functions. The SVM algorithm considers very few parameters.

III Proposed System

The proposed system will identify the chances of survival by identifying the factors which affects the BMT Transplantation. Up till now only few factors has been identified. This work will input the number of factors and after applying the data mining and machine learning techniques it will identify the important factors. For this purpose, Support Vector Machines (SVM) is used. Support Vector Machines (SVM) has recently gained prominence in the field of machine learning and pattern classification. The factors which are considered to be important are the overall health, age, medical condition and donor matching.

The Overall Health is considered because this factor covers most of the important information about the patients. This factor will provide the health parameters such as weight, height, blood pressure, fluid levels etc. Age of the Patient is also considered as a factor because transplants are performed on various age groups. Patients over the age of 50 are more likely to have complications; including graft-versus-host disease (GVHD).Medical Condition and Likelihood of Response factor is important because ongoing research and advances have increased the success of stem cell transplantation. The procedure may be difficult because the patient should respond to the treatment. Stem cell transplant can be viewed as pre-transplant and post-transplant procedure as monitoring the patient through the treatment is

must. Availability of Donor Cell is also considered one of the factors because if there are no available matching donor cells, then a transplant is not possible. When the patient is donating their own cells (autologous transplants) a sufficient amount of healthy cells must be available. For procedures in which the patient is NOT the donor (allogenic transplants), a close HLA match (discussed below) is necessary. A close HLA match is defined as at least 5 out of 6 HLA markers matching for adults getting donor stem cells and 4 out of 6 for children getting stem cells from cord blood. One of the most important factors is the graft-versus-host disease which is mostly considered after the stem cell transplant procedure. This factor is influential as this causes the patient to suffer mostly during the treatment. This factor can be helpful to choose the donor perfectly.

Support Vector Machine Algorithm

Support vector machine algorithm is a good classifier and has been effective method for regression, classification and general pattern recognition. The aim of SVM is to find the best classification function to distinguish between members of the two classes in the training data. The SVM algorithm considers two cases: the data is linearly separable and the data is linearly inseperable. A technical measure of how clearly a hyper plane separates data is its margin. This is the distance of the hyper plane to the closest point in the dataset; a large margin means that the hyper plane very clearly separates the data. In the below figure there are several hyper plane (shown by light black line) but the one in the dark is selected as it gives the best generalization. The hyper plane is decided by some conditions. The SVM finds the maximum marginal hyper plane and selects the hyper plane which gives best classification results.

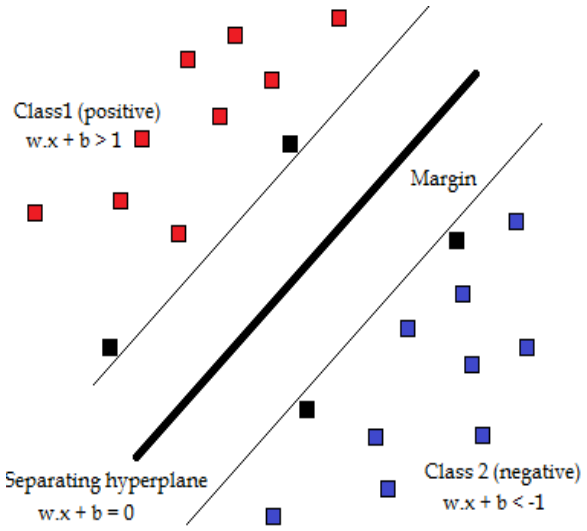

Fig: SVM

IV. Conclusion

This review paper provides information about the stem cell transplant and analyzing and identifying the patients with high odds of survival. Here, we are using the SVM Algorithm to classify the patients with high rate of survival in order to prioritize the resources and in selecting a donor. Analysis of clinical data regarding the stem cell transplant from past transplant could enhance the understanding of the factors influencing success. Sometimes during analysis correct amount of information is not available or the available information does not generate correct patterns. The information obtained can be useful to identify individuals with high success rate in survival. The factors are important in order to decide the success rate of survival among the patients. In future, more classification algorithms can be used and can evaluate the results. This review paper can be taken as a reference for any other diseases where analysis is important concept.

REFERENCES

[1] A. Gratwohl, H. Baldomero, M. Aljurf, M. C. Pasquini, L. F. Bouzas, A. Yoshimi, J. Szer, J. Lipton, A. Schwendener, M. Gratwohl, K. Frauendorfer, D. Niederwieser, M. Horowitz, and Y. Kodera, "Hematopoietic stem cell transplantation: A global perspective," *J. Amer. Med. Assoc.*,vol. 303, no. 16, pp. 1617–1624, 2010.

[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", IEEE journal, vol.18,no.1, 2014.

[3] K. Gibert, M. Sanchez-Marre, and V. Codina, "Choosing the right data mining technique: classification of methods and intelligent recommendation,"presented at the Biennial Meet. Int. Environmental Modelling Software Society, Int. Congr. Environmental Modelling Software, Ottawa,ON, Canada, 2010.

[4] E. Keogh, S. Lonardi, and C. A. Ratanamahatana, "Towards parameterfree data mining," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, pp. 206–215,2009

[5] M. Tomblyn, T. Chiller, H. Einsele, R. Gress, K. Sepkowitz, "Guidelines for Preventing Infectious Complications among Hematopoietic Cell Transplantation Recipients: A Global Perspective",ASBMT, 2009

[6] C. Baron, R. Somogyi, L. D. Greller, V. Rineau, P. Wilkinson *et al.*, "Prediction of graft-versus-host disease in humans by donor gene-expression profiling," *PLoS Med.*, vol. 4, no. 3, pp. 69–83, 2007.

[7] R.kumar, R.Verma," Classification Algorithms for Data Mining:A Survey",IJIET,vol no.1,2012.