# Predicting Blood Transfusions for Coronary Artery Bypass Graft Patients using Deep Neural Networks and Synthetic Data

**Hsiao-Tien Tsai1 • Jichong Wu1 • Puneet Gupta2 • Eric R Heinz2 • Amir Jafari1**

## Abstract

Sample\_text.

**Keywords** Deep Neural Network • Synthetic Data Generation • Blood Transfusion • Coronary Artery Bypass Graft

✉Hsiao-Tien Tsai  
[jennytsai@gwu.edu](mailto:jennytsai@gwu.edu)

Jichong Wu  
[jwu36@gwu.edu](mailto:jwu36@gwu.edu)

Puneet Gupta MD  
[gupta14@gwu.edu](mailto:gupta14@gwu.edu)

Eric R Heinz MD PhD  
[gupta14@gwu.edu](mailto:gupta14@gwu.edu)

Amir Jafari PhD  
[ajafari@gwu.edu](mailto:ajafari@gwu.edu)

1 Data Science, George Washington University, DC, United States

2 Department of Anesthesiology and Critical Care Medicine, George Washington University School of Medicine and Health Sciences, DC, United States

## Introduction

Coronary Artery Bypass Graft (CABG) is a common cardiac surgery but continues to have many associated risks, including major bleeding which might need blood transfusion. Previous research has shown that blood transfusion during CABG surgery is associated with an increased risk for mortality after surgery [1]. Specially, post-operative blood transfusion after CABG is associated with higher odds of readmission and heart failure within 30-days [2].

To lower the risk of mortality after surgery, there is a need to develop models that preoperatively predict which patients will need an intra-operative or post-operative blood transfusion. This will not only help to improve patient selection and patient education, but also physician preoperative awareness and perioperative guidelines for CABG patients. Therefore, the goal of this research is to explore different modern data analysis techniques and find the models that can best make predictions, including synthetic data generation and deep neural networks.

## Related Work

Research have been conducted to investigate factors that can help to predict major bleeding [3] and the need for red blood cell transfusion after cardiac surgery [4]. In one of the studies [5] that is most relevant to the current research, the researchers employed machine learning models to predict perioperative allogeneic blood transfusion for cardiac patients. The best model (Random Forest) showed good performance (RUC ranged from .76 - .86), however, the study has several limitations. For example, the data was from a single adult cardiac surgery center in Austria with a relatively small sample size (N = 3,782), thus the results may not be generalizable to other samples with different demographics or nationalities. Moreover, the studies only predicted allogeneic blood transfusion (i.e., transfusion of more than 10 units of packed red blood cells (pRBC)), while blood transfusion regardless of volume has been associated with many known risks. Lastly, the study only tested the basic machine learning models (e.g., tree-based models), and it is likely that the performance can be significantly improved using more advanced techniques and deep neutral networks.

To address this research gap, the current project will use the national medical database in the U.S. with a large sample size of over 13,500 data points. Additionally, we will predict the need for blood transfusion regardless of volume. Lastly, we will experiment with various approaches in order to optimize the performance, including synthetic data generation and deep neural networks.

## Solution and Methodology

### Data Source and Data Preprocessing

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

### Exploratory Data Analysis

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

### Analysis Strategy

Sample\_text.

Sample\_text.

Sample\_text.

## Results and Discussion

### Fully Connected Neural Networks (FNN)

Sample\_text.

Sample\_text.

#### Original data

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

#### Synthetic Data - Realtabformer

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

#### Synthetic Data – Data Synthesizer

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

### Convolutional Neural Networks (CNN)

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.

## Conclusion

Sample\_text.

Sample\_text.

Sample\_text.

Sample\_text.



**Fig. 1** A green box

Figure 1, Table 1

**Table 1** An empty table

|  |  |  |
| --- | --- | --- |
| 1 | 2 | 3 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## References

(Style – springer-Vancouver)

1. Chin WW (2010) How to Write Up and Report PLS Analyses. Handb Partial Least Squares 655–690