Predicting Death Events From Clinical Data Using Statistical Control Charts

Muhammad Hamza Motiwala
Faculty of Computer Science
GIK Institute of Engineering Sciences and Technology
u2022380@giki.edu.pk

Muhammad Adeel
Faculty of Computer Science
GIK Institute of Engineering Sciences and Technology
u2022331@giki.edu.pk

Ahmed Bilal Nazim
Faculty of Computer Science
GIK Institute of Engineering Sciences and Technology
u2022064@giki.edu.pk

***Abstract*— This paper applies statistical techniques, including covariance matrices, control charts (CUSUM, EWMA, Shewhart), and dimensionality reduction(single value decomposition) to clinical data to predict mortality risk in heart disease patients. By examining relationships between key variables such as serum creatinine, ejection fraction, and survival time, the study identifies predictors for mortality and provides tools for enhanced patient monitoring.**

Keywords—Linear Algebra, Control Charts, Process Monitoring, Dimensionality Reduction, Covariance Matrix, Mortality Prediction, Clinical Data Analysis.

I. Introduction

Heart failure is a significant global health issue, and early identification of patients at high risk of mortality can significantly improve clinical outcomes. The objective is to use statistical and dimensionality reduction techniques to identify patterns in clinical data related to heart failure patients and to understand the relationship between these pattern and patient survival. The analysis is based on clinical data that includes variables like age, ejection fraction, serum creatinine, and survival outcomes (death event). The focus is to explore these relationships and leverage control charts and dimensionality reduction to enhance the predication of the patient outcomes.

II. Methodology

This study involved four primary phases:

A. Data Cleaning:

Handling missing values and standardizing data.

B. Covariance Analysis:

Utilizing covariance matrices to explore linear relationships between clinical variables.

C. Control Charts:

Employing control charts(CUSUM, EWMA, Shewhart) to track trends and variability in key health indicators.

D. Dimensionality Reduction:

Applying Singluar Value Decompostion (SVD) to reduce the dimensionality of data to enhave visualization.

The Following control charts were used:

I. **CUCUM for Platelets count** variation over time
II. **EWMA for Serum Creatinine** (kidney health) over time
III. **Shewhart for Ejection Fraction** (heart pumping efficiency)

Dimensionality reducation was used to project high-dimensional data into two principal components for easier interpretation of trends and patterns, particularly in seperating survival and non-survival groups

III. Viusal and Statistics

I. Covariance matrix

The Covariance matrix (fig. 1) illustrates the linear relationship between clinical variables, highlighting string association with mortality.

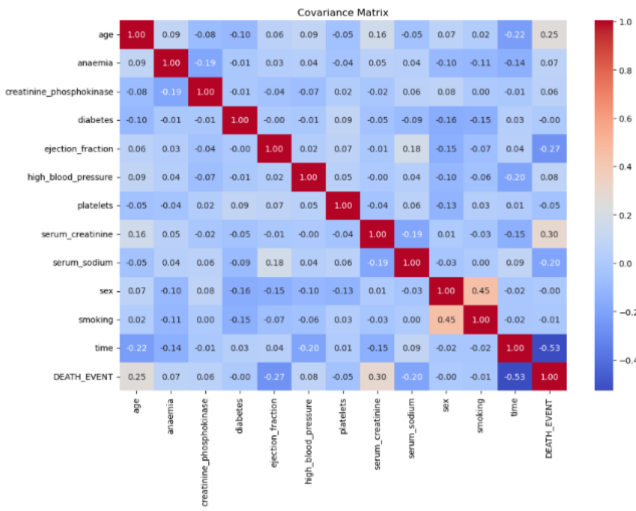

Figure 1

Significant findings include:

- **Serum Creatinine and Mortality**: A positive correlation (0.30) suggests that increased serum creatinine levels are a significant predictor of mortality.

- **Fraction and Death Events**: A negative correlation (-0.27) indicates that lower ejection fraction is associated with poor survival outcomes.

- **Time vs. Mortality**: A strong negative correlation (-0.53) shows that shorter survival time correlates with increased mortality risk.

II. CUSUM Chart for Platelets

The CUSUM chart (Fig. 2) reveals the platelets' variability over time, demonstrating trends above and below the expected range.

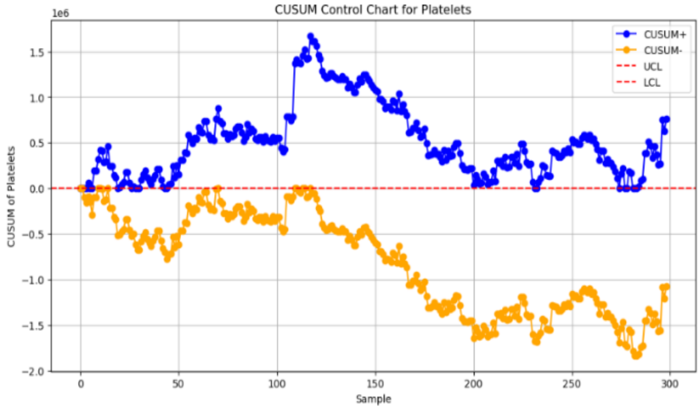

Figure 2

This helps identify potential health issues as trends diverge from normal values.

- **CUSUM+ (Increasing Trend)**: Suggests elevated platelet counts.
- **CUSUM- (Decreasing Trend)**: Indicates decreasing platelet levels

III. EWMA Control Chart for Serum Creatinine

The EWMA chart (Fig. 3) tracks serum creatinine levels, with control limits indicating potential kidney dysfunction when values exceed the upper control limit (UCL).

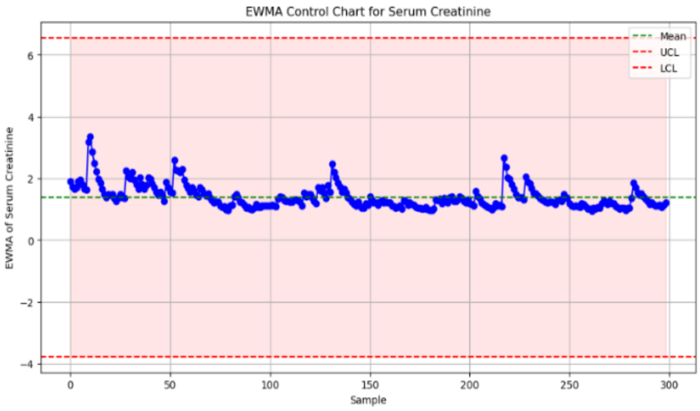

Figure 3

Initial Increase: Indicates periods of normal renal stress.

Post-Sample 50: Occasional fluctuations indicate possible kidney health concerns when values cross the UCL.

IV. Shewhart Control Chart for Ejection Fraction

The Shewhart diagram (Fig. 4) assesses the variation in heart pump efficiency (ejection fraction), with control limits ranging from 10 to 70.

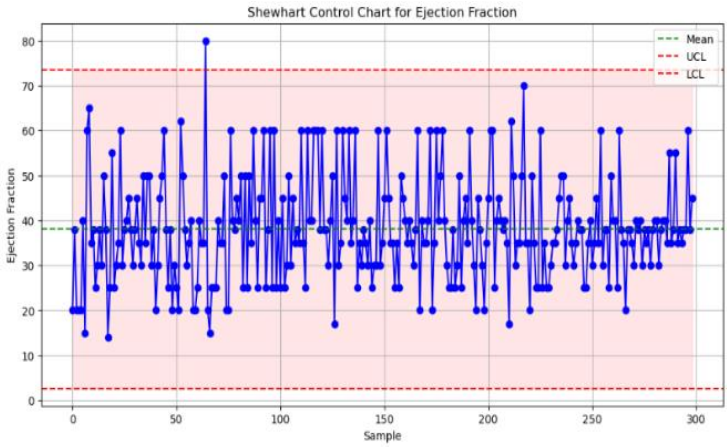

Figure 4

Data points near these limits signal potential issues with cardiac function.

- **Fluctuating Points**: Indicate inconsistent heart performance, which is critical for patient care.

V. Dimensionality Reduction Using SVD

SVD scatterplot (Fig. 5) reduces the data's dimensionality to two principal components (PC1, PC2), capturing over 75% of the data's variance.

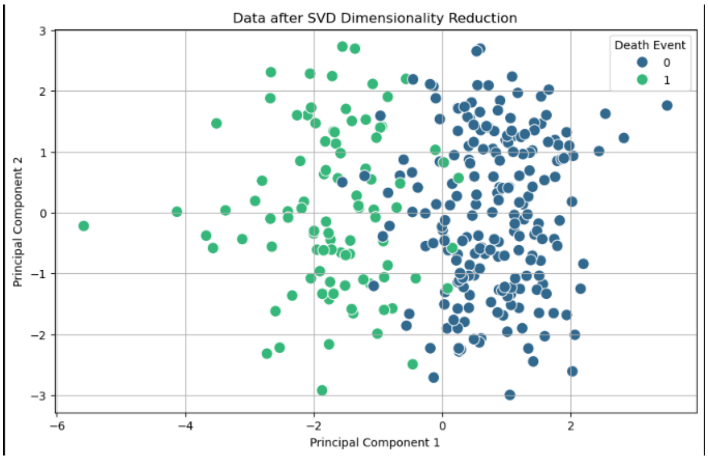

Figure 5

- **DEATH_EVENT = 1 (Green) and DEATH_EVENT = 0 (Blue)**: Clear separation between surviving and deceased patients based on clinical features.
- **Significant Partitioning**: The clustering along PC1 emphasizes the importance of key variables like serum creatinine, ejection fraction, and time for predicting mortality.

s

IV. Result and Discussion

V. Conclusion

This study demonstrates how statistical tools such as covariance analysis, control charts, and dimensionality reduction can provide valuable insights into clinical data for predicting mortality in heart failure patients. Key findings include:

- **Strong Predictors**: Serum creatinine, ejection fraction, and survival time are significant predictors of mortality.

- **Effective Monitoring**: Control charts are useful for tracking trends and deviations in key health markers.

- **Clear Separation**: SVD clearly distinguishes between survival and non-survival groups, further confirming the relevance of the selected predictors.

In the future, these techniques could be expanded to create predictive models using machine learning, improving automated mortality risk assessments for heart failure patients across larger datasets.

References

[1] Montgomery, D. C. (2009). Introduction to Statistical Quality Control. John Wiley & Sons.

[2] Jolliffe, I. T. (2011). Principal Component Analysis. Springer.

[3] Heart Failure Dataset Description. UCI Machine Learning Repository.

[4] Rousseeuw, P. J., & Leroy, A. M. (2003). Robust Regression and Outlier Detection. Wiley..