DAYANANDA SAGAR UNIVERSITY
SCHOOL OF ENGINEERING
HAROHALLI, KANAKAPURA ROAD – 562112

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
(DATA SCIENCE)

FODS PROJECT REPORT

ON

"DRUG DOSE OPTIMIZATION USING MATLAB SIMULATIONS"

2024-2025

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING (DATA SCIENCE)

Submitted by

Shalini S- ENG23DS0032

Bhumika Damodar Moger- ENG23DS0006

Pavithra s- ENG23DS0023

Under The Supervision of:

Prof. Sindhu A

Assistant Professor

Department of CSE (Data Science), DSU

DAYANANDA SAGAR UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(DATA SCIENCE)

SCHOOL OF ENGINEERING

HAROHALLI, KANAKAPURA ROAD – 562112

CERTIFICATE

It is certified that the mini project work entitled "Drug Dose Optimization Using MATLAB Simulations" has been carried out at *Dayananda Sagar University*, Bangalore, by *Shalini S- ENG23DS0032, Bhumika Damodar Moger- ENG23DS0006, Pavithra S- ENG23DS0023*, Bonafide student of fourth Semester, B-Tech in partial fulfilment for the award of degree in *Bachelor of Technology in Computer Science & Engineering (Data Science)* during academic year *2024-25*. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in departmental library.

The project report has been approved as it satisfies the academic requirements in respect of project work for the said degree.

Signature of the Guide **Signature of the chairperson**

ACKNOWLEDGEMENT

A project's successful completion offers a sense of satisfaction, but it is never finished without expressing gratitude to everyone who contributed to its accomplishment. We would like to convey our sincere gratitude to our esteemed university, Dayananda Sagar University, for offering the first-rate facilities.

We are especially thankful to our Chairperson, Dr. Shaila S G, for providing necessary departmental facilities, moral support and encouragement. The largest measure of our acknowledgment is reserved for Prof. Sindhu A. whose guidance and support madeit possible to complete the project work in a timely manner.

We would want to thank everyone who has assisted us in successfully completing this project work, both directly and indirectly. The staff has provided us with a great deal of direction and cooperation.

Shalini S- ENG23DS0032
Bhumika Damodar Moger- ENG23DS0006
Pavithra S- ENG23DS0023

DECLARATION

We hereby declare that the project entitled **"Drug Dose Optimization Using MATLAB Simulations"** submitted to Dayananda Sagar University, Bengaluru, is a bona fide record of the work carried out by me under the guidance of Prof. Sindhu A., Assistant Professor in the Dayananda Sagar University School of Engineering's Department of Computer Science and Engineering (Data Science). This work is submitted toward the partial fulfillment of the requirements for the award of a Bachelor of Technology in Computer Science and Engineering (Data Science).

Shalini S- ENG23DS0032

Bhumika Damodar Moger- ENG23DS0006

Pavithra S- ENG23DS0023

ABSTRACT

Drug dose optimisation is a critical component of precision medicine, aiming to maximize therapeutic efficacy while minimizing adverse effects. This process involves tailoring drug regimens based on individual patient characteristics such as age, weight, genetic profile organ function, and co-existing conditions.

Advances in pharmacokinetics, pharmacodynamics, and pharmacogenomics have significantly enhanced the ability to predict optimal dosing strategies. Furthermore, the integration of clinical decision support tools and machine learning algorithms has facilitated more accurate, real-time adjustments to dosing regimens. This report explores the principles, methodologies, and technologies employed in drug dose optimisation, with a focus on improving patient outcomes, reducing drug-related toxicity, and promoting cost-effective healthcare.

Drug dose optimisation is a critical component of precision medicine, aiming to maximize therapeutic efficacy while minimizing the risk of adverse drug reactions. It focuses on tailoring drug regimens to individual patient characteristics such as age, weight, genetic profile, organ function, and the presence of co-existing conditions. Traditional "one-size-fits-all" dosing strategies often overlook these inter-individual differences, potentially leading to under-dosing (resulting in therapeutic failure) or over-dosing (leading to toxicity). As such, personalized dosing is emerging as a key objective in modern pharmacological practice and clinical therapeutics.

TABLE OF CONTENTS

INTRODUCTION

Drug dose optimisation is a critical aspect of modern healthcare that aims to improve the safety and effectiveness of pharmacological treatments by personalizing drug regimens based on individual patient characteristics. Traditional drug dosing strategies, often based on population averages, fail to account for significant inter-patient variability in factors such as age, sex, genetics, body weight, organ function, and comorbidities. As a result, the concept of individualizing drug doses to better match the needs of each patient has become a central focus in the field of pharmacology and clinical medicine.

The primary goal of drug dose optimisation is to achieve the maximum therapeutic benefit while minimizing the risk of adverse effects and toxicity. This project focuses on estimating an optimal drug dosage based on multiple patient health parameters such as age, temperature, medical condition, blood pressure, and sugar level. Using MATLAB, we analyze data-driven insights and rule-based logic to provide safe and customized dosage recommendations.

The core objective of drug dose optimisation is to maintain drug concentrations within the therapeutic window—a concentration range where the drug is most effective without being toxic. This requires a strong understanding of pharmacokinetics (how the body affects the drug) and pharmacodynamics (how the drug affects the body). By analyzing these factors, one can better predict how a drug will behave in different individuals and make informed adjustments to dosing schedules.

This project centers around the development of a computational model for dose estimation using MATLAB, based on a rule-based decision system. Patient health indicators such as **age**, **body temperature**, **medical condition**, **blood pressure**, and **blood sugar levels** are used to inform and adjust drug dosage recommendations. The methodology combines clinical logic with data analysis, enabling a semi-automated approach to personalized dosing.

OBJECTIVE AND SCOPE OF WORK

OBJECTIVE

The aims of this project are grounded in the application of data-driven techniques to improve drug dose estimation and personalization. A fundamental component of the methodology involves the initial cleaning and preprocessing of patient health data to ensure accuracy, consistency, and usability. This includes handling missing values, removing duplicates, and standardizing the dataset to prepare it for analysis.

Following data cleaning, the next phase focuses on normalizing the patient metrics to ensure that variables with different units and scales are brought to a common range. This step is essential to facilitate meaningful comparisons across features and to support robust visualizations as part of exploratory data analysis (EDA). Visual tools such as histograms, scatter plots, and correlation matrices are employed to uncover underlying trends, outliers, and relationships among health parameters such as age, temperature, blood pressure, and sugar level.

SCOPE OF WORK

This project serves multiple purposes, including data analysis and clinical decision support by providing interpretable dosage recommendations based on patient data. It is also designed for educational use, demonstrating how rule-based systems combined with basic machine learning techniques like PCA can enhance healthcare analytics. Additionally, the framework is well-suited for small to medium clinical settings seeking a quick, transparent, and easy-to-implement method for personalized drug dosage **estimation.**

DESCRIPTION OF WORK

The project is designed to estimate patient-specific drug dosages based on health-related input variables using MATLAB. It involves a structured sequence of steps that encompass data loading, preprocessing, analysis, modeling, and visualization.

Initially, the dataset is loaded from a CSV file and subjected to cleaning operations to remove incomplete or missing entries. Once the data integrity is ensured, normalization is applied to key numerical variables .

Exploratory Data Analysis (EDA) is then performed to understand the distribution, spread, and variability of the data using descriptive statistics, histograms, and boxplots

Following this, Principal Component Analysis (PCA) is employed to reduce the dimensionality of the dataset and identify patterns or clusters among patients based on age and temperature.

A rule-based algorithm is developed for each patient based on clinical logic. The estimated dosage is then categorized into three levels:Low, Moderate, and High, for better interpretability

To address the high dimensionality and complexity of the data, Principal Component Analysis (PCA) is conducted to reduce the number of variables while preserving most of the variance in the data. PCA helps to uncover hidden structures and relationships within the patient population by projecting the data into a lower-dimensional space. This facilitates the identification of patient clusters or patterns related to critical features such as age and temperature, which may correlate with dosage requirements.

Building on these analyses, a rule-based algorithm is developed to estimate drug dosages tailored to each patient's unique profile. This algorithm incorporates established clinical guidelines and pharmacological knowledge, using conditional statements to adjust dosages based on key health metrics and risk factors. Dosage recommendations are then categorized into three levels—Low, Moderate, and High—to simplify interpretation and support clinical decision-making.

.

METHODOLOGY

This project uses a structured, rule-based approach to estimate drug dosages for patients based on health parameters. First, patient data is imported and cleaned by removing missing entries. Numerical features like age and temperature are normalized to ensure consistency. Exploratory Data Analysis (EDA) is performed using statistics and visualizations to understand data distribution. Principal Component Analysis (PCA) is then applied to reduce dimensionality and highlight patterns.

A rule-based logic model is developed to adjust a base dosage according to the patient's age, temperature, medical condition, blood pressure, and sugar level. The resulting dosages are categorized into Low, Moderate, or High levels. Visualization techniques are used to explore relationships and patterns in the data. Finally, an interactive MATLAB module allows users to input new patient details and receive personalized dosage predictions.

The core of the project is the development of a rule-based logic model that calculates an optimal drug dosage for each patient. This model adjusts a base dosage by incorporating clinical decision rules derived from patient-specific factors such as age, body temperature, diagnosed medical conditions, blood pressure, and blood sugar levels. These factors are used to increase or decrease the dosage according to clinically relevant thresholds, ensuring that recommendations are both safe and effective. The final dosage outputs are grouped into three categories—Low, Moderate, and High—to facilitate clear and actionable guidance for clinicians.

Visualization techniques are extensively utilized to explore and communicate the relationships between patient variables and dosage levels. Graphical representations such as bar charts, heatmaps, and pie charts illustrate dosage distributions and risk profiles across different patient subgroups, making it easier to interpret complex data patterns.

SOURCE CODE

```
data = readtable('drug1.csv');

data = rmmissing(data);

fprintf('Cleaned Data Rows: %d\n', height(data));

disp(head(data, 5));

%% Step 2: Normalize Age and Temperature

norm_data.age = normalize(data.age);

norm_data.temperature = normalize(data.temperature);

fprintf('\n--- Data Normalization ---\n');

disp(table(data.age(1:5), norm_data.age(1:5), 'VariableNames', {'Original Age', 'Normalized Age'}));

disp(table(data.temperature(1:5), norm_data.temperature(1:5), 'VariableNames', {'Original Temp', 'Normalized Temp'}));

%% Step 3: Data Visualization and EDA

fprintf('\n--- Data Visualization and EDA ---\n');

fprintf('Age - Mean: %.2f, Std: %.2f\n', mean(data.age), std(data.age));

fprintf('Temperature - Mean: %.2f, Std: %.2f\n', mean(data.temperature), std(data.temperature));

figure;

subplot(1,2,1); histogram(data.age); title('Age Distribution'); xlabel('Age'); ylabel('Count');

subplot(1,2,2); histogram(data.temperature); title('Temperature Distribution'); xlabel('Temperature'); ylabel('Count');

figure;

subplot(1,2,1); boxplot(data.age); title('Boxplot - Age');

subplot(1,2,2); boxplot(data.temperature); title('Boxplot - Temperature');

%% Step 4: PCA (CPA)

fprintf('\n--- Principal Component Analysis (PCA) ---\n');

X = [norm_data.age, norm_data.temperature];
```

```
[coeff, score, latent, ~, explained] = pca(X);
disp('PCA Coefficients:');
disp(coeff);
disp('Explained Variance (%):');
disp(explained);
for i = 1:height(data)
    age = data.age(i);
    temp = data.temperature(i);
    condition = string(data.condition(i));
    bp = string(data.bp(i));
    sugar = string(data.sugar(i));
    dose = 200;
    if age > 60, dose = dose + 50;
    elseif age < 30, dose = dose - 30; end
    if temp > 99, dose = dose + 20; end
    if condition == "Depression"
        dose = dose + 25;
    elseif condition == "Diabetes"
        dose = dose + 40;
    elseif condition == "Hypertension"
        dose = dose + 35;
    end
    if bp == "high", dose = dose + 15;
    elseif bp == "low", dose = dose - 10; end
    if bp == "high", dose = dose + 15;
    elseif bp == "low", dose = dose - 10; end
      if sugar == "high", dose = dose + 10;
    elseif sugar == "low", dose = dose - 10; end
```

```
    dosages(i) = dose;
end
%% Step 7: Dosage Visualizations
figure;
scatter(data.age, data.estimated_dosage, 60, 'filled');
xlabel('Age'); ylabel('Estimated Dosage'); title('Dosage vs Age'); grid on;
%% Step 8: PCA Plot with Labels
figure;
gscatter(score(:,1), score(:,2), data.condition);
xlabel('PC1'); ylabel('PC2'); title('PCA Grouped by Condition'); grid on;
```

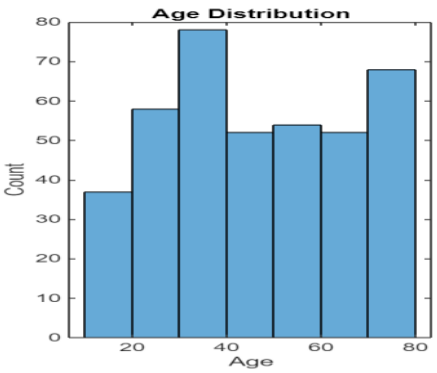
RESULT

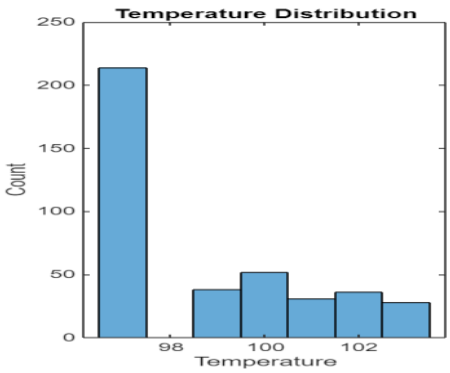
```
--- Data Cleaning ---
Cleaned Data Rows: 399

--- Data Normalization ---
```

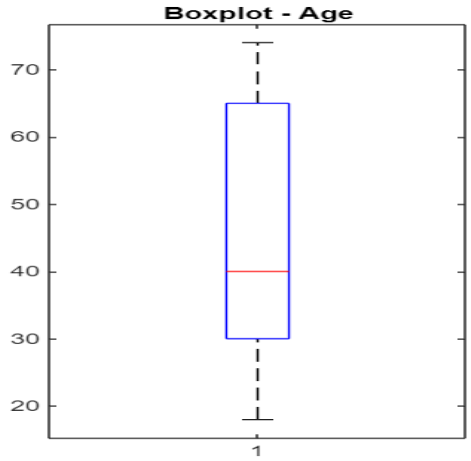
Original Age	Normalized Age
30	-0.77307
22	-1.1875
56	0.5739
35	-0.51404
30	-0.77307

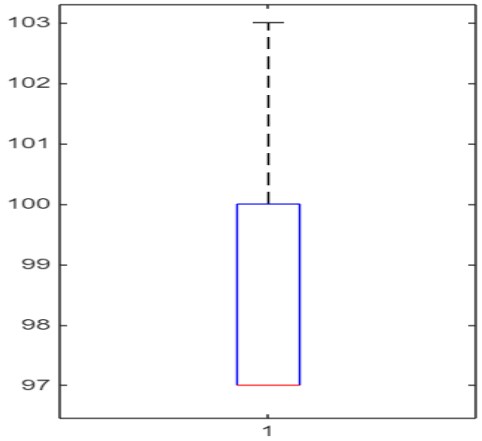
Original Temp	Normalized Temp
97	-0.83388
97	-0.83388
103	2.0018
100	0.58396
97	-0.83388

```
--- Data Visualization and EDA –

Age - Mean: 44.92, Std: 19.30
Temperature - Mean: 98.76, Std: 2.12
```



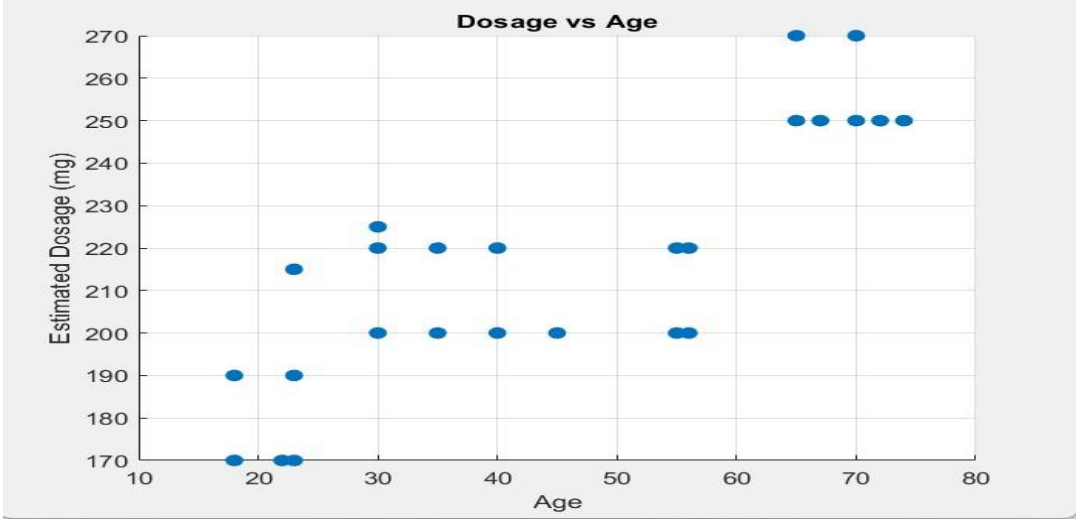


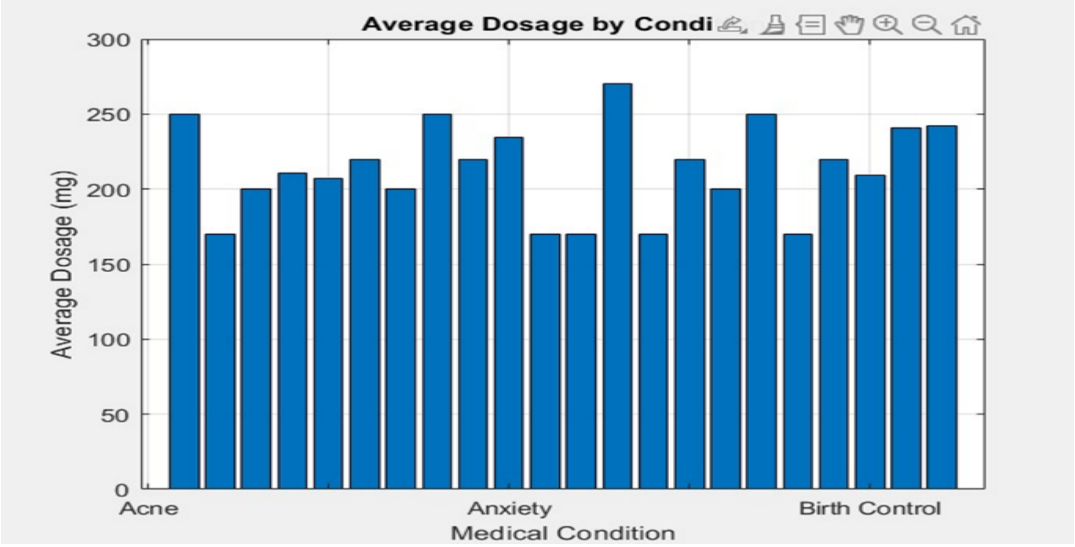





```
\n--- Summary Statistics ---
           drug: [1×1 struct]
         dosage: [1×1 struct]
      condition: [1×1 struct]
         review: [1×1 struct]
         rating: [1×1 struct]
    usefulCount: [1×1 struct]
             bp: [1×1 struct]
          sugar: [1×1 struct]
    temperature: [1×1 struct]
            age: [1×1 struct]
    Sideeffects: [1×1 struct]

--- Principal Component Analysis (PCA) ---
PCA Coefficients:
   -0.7071    0.7071
    0.7071    0.7071

Explained Variance (%):
   63.6673
   36.3327

--- Dosage Calculation ---


--- Estimate Dosage for New Patient ---
Estimated Dosage: 315.0 mg (High Dose)
```

```
Dosage Statistics by Condition:
Condition                    | Mean (mg)  | Median (mg)  | Std Dev (mg)
-------------------------------------------------------------
Acne                         |     250.00 |       250.00 |         0.00
Alcohol Dependence           |     170.00 |       170.00 |         0.00
Anxiety                      |     200.00 |       200.00 |         0.00
Bipolar Disorde              |     210.87 |       200.00 |        41.99
Birth Control                |     207.00 |       200.00 |        21.23
Bowel Preparation            |     220.00 |       220.00 |         0.00
Constipation, Drug Induced   |     200.00 |       200.00 |         0.00
Cough and Nasal Congestion   |     250.00 |       250.00 |         0.00
Depression                   |     219.67 |       215.00 |         5.16
Diabetes, Type 2             |     234.62 |       250.00 |        23.53
Emergency Contraception      |     170.00 |       170.00 |         0.00
Hyperhidrosis                |     170.00 |       170.00 |         0.00
Insomnia                     |     270.00 |       270.00 |         0.00
Lymphocytic Colitis          |     170.00 |       170.00 |         0.00
Major Depressive Disorde     |     220.00 |       220.00 |         0.00
Narcolepsy                   |     200.00 |       200.00 |         0.00
Pain                         |     250.00 |       250.00 |         0.00
Panic Disorde                |     170.00 |       170.00 |         0.00
Rosacea                      |     220.00 |       220.00 |         0.00
Urinary Tract Infection      |     209.23 |       200.00 |        10.38
Vaginal Yeast Infection      |     240.45 |       250.00 |        14.30
Weight Loss                  |     241.92 |       250.00 |        13.57
-------------------------------------------------------------
```





CONCLUSION

In this project, a rule-based MATLAB model was successfully developed to estimate personalized drug dosages using patient-specific parameters such as age, temperature, medical condition, blood pressure, and sugar level. Through systematic data cleaning, normalization, exploratory analysis, and visualization, meaningful insights were derived from the dataset. Principal Component Analysis further enhanced understanding of underlying patterns. The model provides clear dosage categories and allows real-time predictions, making it a useful tool for preliminary clinical decision support and healthcare analytics.

Overall, the approach demonstrates how simple rule-based systems can mimic basic clinical reasoning. With further refinement and more data, the model can be expanded for broader medical applications**.**

While this initial model focuses on a defined set of parameters and rule-based logic, it lays a strong foundation for future enhancements. Incorporating larger and more diverse datasets, integrating pharmacogenomic information, and adopting machine learning techniques could further refine dosage predictions and expand the model's applicability to a wider range of therapeutic areas. Additionally, the system could be integrated into electronic health records (EHRs) to facilitate seamless clinical workflow integration.

Overall, this project highlights the promising role of computational modeling in healthcare analytics and precision medicine. By bridging clinical expertise with data-driven methods, it contributes to the ongoing effort to improve drug safety, efficacy, and personalized patient care.

REFERENCES

1. MathWorks. (2023). *MATLAB Documentation*. Retrieved from https://www.mathworks.com/help/

2. Durairaj, M., & Ranjani, V. (2013). Data mining applications in healthcare sector: A study. *International Journal of Scientific & Technology Research*, 2(10), 29–35.

3. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments.