***Dissertation Title***

*Machine Learning for Predictive Analytics in Hospital Patient Care: A Data-Driven Approach to Early Disease Detection and Resource Optimization*

**Final Thesis**

In Partial Fulfillment

of the Requirements for the Degree of

Master in Computer Science

|  |  |  |
| --- | --- | --- |
| Student Name | : | Roshan Shrestha |
| Student ID | : | u2732965 |
| Supervisor | : |  |

Abstract

Apply the font of Times New Roman to the paragraphs of the abstract using font size of 12. An abstract is usually one to three paragraphs long with a length of 150 to 350 words.

**Example:**

The prediction of human diseases precisely is still an uphill battle task for better and timely treatment. A multidisciplinary diabetic disease is a life-threatening disease all over the world. It attacks different vital parts of the human body, like Neuropathy, Retinopathy, Nephropathy, and ultimately Heart. A smart healthcare recommendation system predicts and recommends diabetic disease accurately using optimal machine learning models with the data fusion technique on healthcare datasets. Various machine learning models and methods have been proposed in the recent past to predict diabetes disease. Still, these systems cannot handle the massive number of multi-features datasets on diabetes disease properly. A smart healthcare recommendation system is proposed for diabetes disease based on deep machine learning and data fusion perspectives. Using data fusion, we can eliminate the irrelevant burden of system computational capabilities and increase the proposed system’s performance to predict and recommend this life-threatening disease more accurately. Finally, the ensemble machine learning model is trained for diabetes prediction. This intelligent recommendation system is evaluated based on a well-known diabetes dataset, and its performance is compared with the most recent developments from the literature. The proposed system achieved 99.6% accuracy, which is higher compared to the existing deep machine-learning methods, therefore, our proposed system is better for multidisciplinary diabetes disease prediction and recommendation. Our proposed system’s improved disease diagnosis performance advocates for its employment in the automated diagnostic and recommendation systems for diabetic patients

**Keywords**: A maximum; of 6 words, or phrases; separated with semi-colons; no punctuation; at the end; use the font of Arial and the size of 11.

Smart diabetes recommendation system; deep ensemble learning; data fusion

Acknowledgements

Use font: Times New Roman; font size: 12; line spacing: 1.5

Contents (This is an example)

[Abstract ii](#_Toc128143288)

[Acknowledgements iii](#_Toc128143289)

[Contents (This is an example) iv](#_Toc128143290)

[List of Tables v](#_Toc128143291)

[List of Figures vi](#_Toc128143292)

[List of Acronyms vii](#_Toc128143293)

[Chapter 1 Introduction (font: Times New Roman; font size: 20) 1](#_Toc128143294)

[1.1 Background(font: Times New Roman; font size 16) 1](#_Toc128143295)

[1.2 Problem Statement 1](#_Toc128143296)

[1.3 Research Question and Objectives 2](#_Toc128143297)

[1.4 Expected outcomes 3](#_Toc128143298)

[Chapter 2 4](#_Toc128143299)

[Literature Review/Related Work 4](#_Toc128143300)

[2.1 Comprehensive Overview of the Existing Literature 4](#_Toc128143301)

[2.2 Critical Analysis of Existing Studies 5](#_Toc128143302)

[Chapter 3 Methodology 7](#_Toc128143303)

[3.1 Data Collection and Preprocessing 8](#_Toc128143304)

[3.2 ML/AI Model Development 9](#_Toc128143305)

[3.3 Evaluation of the Proposed System 9](#_Toc128143306)

[3.4 Others (according to your selected topic) 9](#_Toc128143307)

[Chapter 4 10](#_Toc128143308)

[Experimental Results 10](#_Toc128143309)

[4.1 Experimental Setup 10](#_Toc128143310)

[4.2 Dataset Description 10](#_Toc128143311)

[4.3 Results 11](#_Toc128143312)

[4.4 Comparison with Baseline Methods 11](#_Toc128143313)

[Chapter 5 Conclusion and Future Work 12](#_Toc128143314)

[5.1 Conclusion 12](#_Toc128143315)

[5.2 Future Work 12](#_Toc128143316)

[References 13](#_Toc128143317)

[Appendix A. Title of Appendix 14](#_Toc128143318)

[Appendix B. Title of Appendix 15](#_Toc128143319)

List of Tables

[Table 2.1 Critical analysis/ Summary of the existing studies 6](#_Toc128143320)

List of Figures

[Figure 3.1 Architecture of the proposed system 8](#_Toc128143321)

List of Acronyms

Term Initial components of the term (examples are below)

FEC Forward Error Correction

FET Field Effect Transistor

Please insert one term per table row; this will ensure appropriate spacing and alignment between each term and its components. It will also allow you to sort the terms alphabetically.

# Introduction (font: Times New Roman; font size: 20)

This section presents the background of the study, the problem statement, research objectives, and significance of the study. The section introduces the main focus of the dissertation and sets the context for the reader. It should have more or less following sub sections

Background of the study, Problem statement, Research objectives, Significance of the study

**Below are some examples provided based on the selected example title.**

## Background (font: Times New Roman; font size 16)

(Font: Times New Roman; font size: 12, Line spacing 1.5, double sided printing preferred)

This section contains problems description, motivation, aims and objective.

The healthcare industry is undergoing a digital transformation, powered by the integration of artificial intelligence (AI), machine learning (ML), and data-driven analytics. Among the most critical applications of these technologies is the early detection of diseases and the optimization of healthcare resources. With the increasing volume of patient data being collected from electronic health records (EHRs), hospitals now have an unprecedented opportunity to derive insights that can enhance clinical decision-making and improve patient outcomes.

Early disease detection is not merely a matter of clinical interest—it has direct implications for patient survival, cost savings, and healthcare system efficiency. Detecting life-threatening conditions such as cancer, cardiac arrest, or stroke in their early stages allows healthcare professionals to intervene promptly, potentially saving lives and avoiding the need for intensive, prolonged treatment. However, early detection remains a major challenge, especially in settings with limited resources or delayed access to specialist care.

In parallel, healthcare institutions are constantly facing challenges related to resource allocation. Hospital beds, medical personnel, and diagnostic facilities must be distributed in a way that meets patient demands without overwhelming the system. Predictive analytics can play a significant role here. By forecasting patient admission trends and clustering patients based on risk and resource usage, hospitals can anticipate future loads and make informed logistical decisions.

This dissertation leverages machine learning and predictive analytics to address two core challenges in hospital care: (1) the early detection of high-risk patients and (2) the optimization of hospital resources through clustering and forecasting. By using real-world clinical data, this research demonstrates the potential of machine learning not only as a classification tool but also as a system-level optimizer in hospital environments.

## Problem Statement

Hospitals today generate vast amounts of patient data through daily operations. Despite this wealth of information, many clinical decisions continue to be reactive rather than predictive. Critical conditions often go undiagnosed until symptoms become severe, and hospitals frequently experience resource shortages due to an inability to predict patient inflow or stratify patient risk efficiently.

The core problem addressed in this dissertation is the limited use of machine learning tools in converting hospital data into actionable insights for **early disease detection** and **hospital resource optimization**. Traditional rule-based systems and manual triage often fail to detect critical illnesses in time or allocate resources appropriately, resulting in delayed interventions and increased mortality or costs.

This project aims to fill this gap by developing and evaluating machine learning models that can:

* Predict whether a patient is in a critical condition using routinely available hospital features.
* Cluster patients into risk groups based on their clinical and demographic data.
* Forecast hospital admissions to support better resource planning.

Note: use references to make more convincing problem statement

1.2 AIM and Objective

## Research Question and Objectives

This section introduces the motivation/research questions, and objective of the project.

This research is guided by the following core question:

*How can machine learning models be employed to enhance early disease detection and improve hospital resource optimization using patient data?*

From this question, the following objectives are derived:

1. **To preprocess and engineer meaningful features from hospital patient data** to improve the performance of machine learning models.
2. **To build a supervised learning model (Random Forest)** capable of classifying patients into critical and non-critical conditions with high accuracy.
3. **To implement clustering algorithms (K-Means, Agglomerative, DBSCAN)** to group patients based on risk and analyze group patterns.
4. **To evaluate and compare clustering models** using internal validation metrics such as Silhouette Score and Davies-Bouldin Index.
5. **To develop a hospital admission forecasting model** using time series analysis (ARIMA) to predict daily patient inflow.
6. **To ensure explainability** of the predictive model using SHAP (SHapley Additive exPlanations) to support clinical trust.
7. **To evaluate the effectiveness of the proposed framework** using metrics such as accuracy, precision, recall, AUC-ROC for classification and Silhouette/DB Index for clustering.

**For example:**

**Research question:** Can a smart healthcare recommendation system based on deep ensemble learning improve healthcare outcomes for multidisciplinary diabetes patients?

**Objectives:**

1. Develop a smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning.
2. Evaluate the effectiveness of the smart healthcare recommendation system in improving healthcare outcomes for multidisciplinary diabetes patients.
3. Identify any barriers and limitations to the adoption of the smart healthcare recommendation system in clinical practice.

## Expected outcomes

This section contains the expected outcomes. It should align with research objectives to understand how objectives turn into outcomes.

The expected outcomes of this dissertation are twofold:

1. **An integrated machine learning framework** capable of identifying high-risk patients using structured hospital data with interpretable outputs. The classifier is expected to achieve an accuracy above 60%, with SHAP visualizations to explain key features like emergency status, billing amount, and length of stay.
2. **A clustering and forecasting pipeline** that supports hospital resource optimization. The clustering analysis is expected to provide meaningful patient segmentation (e.g., low, medium, and high risk) using K-Means as the most effective method. Additionally, time-series forecasting should provide 30-day predictions for hospital admissions, enabling better preparedness.

By achieving these goals, the dissertation will contribute both technically and clinically, demonstrating that a data-driven ML approach can significantly aid in the early detection of critical illnesses and help hospital administrators optimize resources based on predictive insights.

For example:

The results of this research will provide insight into the effectiveness of a smart healthcare recommendation system based on deep ensemble learning in improving healthcare outcomes for multidisciplinary diabetes patients.

The findings will also identify any barriers and limitations to the adoption of the system in clinical practice, and provide recommendations for addressing these barriers.

This research will be useful for healthcare professionals, researchers, and policymakers who are interested in improving healthcare outcomes for diabetes patients through the use of AI and ML.

# 

# Literature Review/Related Work

Note: You can have literature review in chapter 1, just like the original template. But I like to have a separate chapter.

While working on the literature review, it is important to note the methodologies and datasets used in the selected papers and evaluate their effectiveness. It would also be useful to identify any limitations in the selected papers.

To organize your literature review, you may consider structuring it based on different ML algorithms or applications. I suggest selecting at least 30 recent papers (published within the last five years) from reputable journals such as Springer, Elsevier, ScienceDirect, etc.

To get started, you can select a recent paper that you feel is closely related to your problem. While reading that paper, you will come across references to other relevant papers. As you continue reading, a structure will emerge based on either type of ML algorithms or related applications.

I would like to provide you with some guidance on organizing your literature review. While reading the papers, you will get an idea of how people organize their literature. Therefore, your literature review should cover the following topics:

## Comprehensive Overview of the Existing Literature

This section should provide a thorough overview of the existing studies and their challenges. After discussing the existing studies, you should introduce your proposed system.

For each paper, you include in your literature review, ensure you provide a brief summary of the study, including the problem, methodology, results, and limitations. You should also evaluate the relevance and significance of the paper to your research problem.

I have provided an example below: Please note that this is just an example to help you understand how to structure your literature review.

**Overview of existing studies/Literature Review/related work:**

Numerous studies have focused on developing machine learning (ML) classifier and artificial intelligence (AI) assistance for the prediction of diabetes disease. With the help of artificial intelligence, we can easily collect healthcare data. After collecting the big data from the healthcare center, we can easily predict human diseases, including multidisciplinary diabetes diseases.

In the realm of diabetes disease detection, various machine learning models have been employed to achieve accurate results. For instance, N. Sneha and T. Gangil [ref] adopted the feature selection technique with SVM classifier for early detection of diabetes disease, and their results were compared with random forest (RF), naive Bayes (NB), decision tree (DT), and KNN classification models. The highest accuracy achieved with SVM was 77.73%. The feature selection technique was used to reduce computational requirements and improve accuracy.

In early detection of diabetes disease Aminah et al. [ref], used k-nearest neighbor (KNN) classifier model was used and the result was compared with that of the support vector machine (SVM) model achieving 85.6% accuracy. The results comparison was made with another machine learning classifier called SVM to authenticate the work.

And so on.

At the end of the literature review you should discuss overall finding and highlight any limitations.

## Critical Analysis of Existing Studies

This section should include a critical analysis of the existing studies. You can identify parameters on which to compare and contrast the studies. For instance, you can use a table with various parameters and compare them with each other and with your proposed system.

**Critical analysis of existing studies**

In this section you should select few parameters for example dataset used, Algorithm applied, accuracy achieved, limitation etc. After discussing each evaluation parameter a table can be provided to give a quick overview of existing studies including your own. Below is an example for the selected paper.

Table 2.1 Critical analysis/ Summary of the existing studies

A picture containing text, screenshot, font, number

Description automatically generated

# Methodology

This chapter describes the proposed system's architecture, data collection, preprocessing, feature extraction and selection, model development, performance evaluation metrics, and implementation details.

It is recommended to provide an end-to-end model pipeline in a block diagram with all the necessary components shown in the Figure 3.1.

* Data collection and preprocessing methods
* Feature extraction
* Machine learning/AL algorithms used for model development
* Performance evaluation metrics and methods

After the figure, you should explain with reference each of the component of the proposed methodology.

A picture containing text, diagram, plan, sketch

Description automatically generated

Figure 3.1 Architecture of the proposed system

## Data Collection and Preprocessing

Describe the data collection and preprocessing methods

Example:

**Data collection:** Patient data will be collected from multiple sources, including electronic health records, medical imaging, and patient-generated data such as wearable devices and mobile apps.

If you are using open source dataset, describe the dataset to understand what feature the dataset has and how they are going to be ingest by the model

**Data preprocessing:** The collected data will be preprocessed to ensure consistency and compatibility across different data sources.

Note: you should use references to explain data collection and preprocessing techniques

## ML/AI Model Development

Describe ML/AI/DL models you are using to develop the proposed system.

For example, for the selected title, A deep ensemble learning model will be developed to integrate the preprocessed data from multiple sources and provide personalized healthcare recommendations for each patient.

Here it should have an end-to-end pipeline from the model input to generate desired output, and describe with the help of related reference the internal functionality of the proposed models.

## Evaluation of the Proposed System

In this section describe how you are going to evaluate your system.

Define the evaluation metrics used to assess the performance of the AI-based system, such as accuracy, sensitivity, specificity, and F1 score. Use relevant references, equation definition etc. to justify why you are selecting the specific evaluation metrics for the performance measurements.

Example: The smart healthcare recommendation system will be evaluated using a randomized controlled trial with multidisciplinary diabetes patients. The primary outcome measure will be healthcare outcomes, including glycemic control, blood pressure, and quality of life. Secondary outcome measures will include patient satisfaction and healthcare resource utilization.

Qualitative interviews will be conducted with healthcare professionals and patients to identify any barriers and limitations to the adoption of the smart healthcare recommendation system in clinical practice.

## Others (according to your selected topic)

This section contains major results achieved through the project.

# 

# Experimental Results

Note: You can have results as a lst section of the methodology, just like the original template. But I like to have a separate chapter.

Results section should provide a clear and objective presentation of your study's findings, using appropriate statistical analysis/ML/Dl and visualization techniques. The Results section is where you provide the evidence that supports your research claims and ultimately helps to answer your research questions.

It can have more or less following sub-sections

## Experimental Setup

Describe software/hardware setup for conducting experiments

## Dataset Description

Describe the dataset used to train and test the AI- system, including the data collection method and data preprocessing steps.

For example:

The proposed SHRS-M3DP model is simulated with two different diabetes disease datasets: Hospital Frankfurt Germany diabetes dataset [ref] and Pima Indians diabetes dataset [ref]. The Hospital Frankfurt Germany diabetes dataset consists of 2000 cases with eight features. The Pima Indians diabetes dataset consists of 768 patients with eight features. The fused features dataset for an experiment was made with the combination of features of both datasets. Both datasets’ cases with some missing values are managed with proposed filtering, and normalization techniques were discussed earlier in chapter 4. The combined, fused dataset features are shown in Table 4.1. The fused features attributes, measuring units, and their ranges are also mentioned.

## Results

Describe the steps taken to evaluate the performance of the AI-based models, including how the system was trained and how the performance was evaluated.

Present the results of the experiments, including the performance metrics and its comparison with the baseline methods, such as traditional machine learning techniques or other existing systems.

Present your findings using tables, graphs or figures. Start with the most important results and present them in a logical order that follows the research questions. Be sure to label all tables, graphs and figures and to provide detailed explanation.

Discuss any limitations of your study and their potential impact on your results. Be honest about the limitations of your study, but avoid undermining the significance of your findings.

## Comparison with Baseline Methods

If necessary provide a separate section for comparing your main findings with baseline methods, such as traditional machine learning techniques or other existing systems.

# Conclusion and Future Work

## Conclusion

This section contains conclusions drawn about the subject.

Summarize the main findings of your study, emphasizing their significance and implications. Be sure to tie your conclusions back to the research questions or hypotheses that guided your study.

## Future Work

This section contains limitations on the conclusions and specifies the future work.

Identify specific areas in which further research could build on your work. These might include unanswered questions, new approaches, or potential applications of your findings.

Provide a rationale for why each area of future research is important and how it could advance the field.

References

The references will follow an IEEE style (or any other standard referencing format such as Harvard, APA, among others).

**Books**

1. J. K. Author, “Title of chapter in the book,” in Title of His Published Book, xth ed.

City of Publisher, Country if not USA: Abbrev. of Publisher, year, ch. x, sec. x, pp.

xxx–xxx.

**Journals**

1. J. K. Author, “Name of paper,” *Abbrev. Title of Periodical*, vol. x, no. x, pp. xxx-xxx, Abbrev. of Month, year.

**Conference Technical Articles**

1. J. K. Author, “Title of paper,” in *Unabbreviated Name of Conf.*, City of Conf., Abbrev. State (if given), year, pp.xxx-xxx.

**Online Sources**

1. J. K. Author, (year, Month. day). *Title of web page* [Online]. Available: URL [http://www.../..../.../doc/text.](http://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=126387&org=NSF&from=news)

Appendix A. Title of Appendix

**Appendix Heading 1**

Text of the appendix goes here

**Appendix Heading 2**

Text of the appendix goes here

**Appendix Table and Figure Captions**

In appendices, table and figure caption labels and numbers are typed in manually (e.g., Table A1, Table A2, etc.). These do not get generated into the lists that appear after the Table of Contents.

Appendix B. Title of Appendix

Text of the appendix goes here if there is only a single heading.