

**FORECASTING CARDIOVASCULAR DISEASES OF
ADULT PATIENTS IN EMERGENCY DEPARTMENT
USING AI MODELS**

BY
OBASI, CHIGOZIE FRANCIS

A Thesis Presented to the
**COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA**

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE
In
**BUSINESS STUDIES, CONCENTRATION IN MANAGEMENT
INFORMATION SYSTEM**

MAY 2025

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **OBASI, CHIGOZIE FRANCIS** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN BUSINESS STUDIES, CONCENTRATION IN MANAGEMENT INFORMATION SYSTEM.**



Dr. Mohammad Saad Al-Ahmadi
Department Chairman



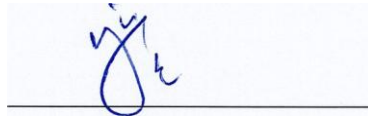
Dr. Suliman S. Al-Homidan
Dean of Graduate Studies

May 6, 2025

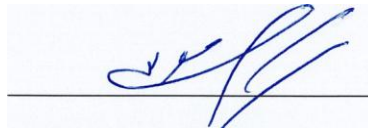
Date



Dr. Rasha Alahmad
(Advisor)



Dr. Faleh Althiyabi
(Member)



Dr. Mousa Al-Bashrawi
(Member)

© Obasi, Chigozie Francis
2025

ALL RIGHT RESERVED

Dedication

This work is dedicated to the members of my family, whose steadfast support and unwavering love have been the root of my success. To my wife and my mother, who have always had faith in my abilities and on days when I had self-doubt, and to my brothers and sister whose moral encouragement and kind words made the challenges lighter. Their continued presence in my life contributed significantly to molding me in areas I cannot fully express.

To the faculties and mentors, I want to say a big thank you for their unmeasurable wisdom and guidance throughout this process. Your insights have been very important in my academic growth, and I am most grateful for your belief in my potential.

Finally, I dedicate this work to my friends, who have been a constant source of positivity, motivation, and joy. You made the challenging periods easier to handle, and your support was unforgettable.

This thesis is a testament to all of you.

ACKNOWLEDGMENTS

I would like to express my profound appreciation to my supervisor, Dr. Rasha Alahmad whose guidance, patience and unwavering assistance, advice and support made this research project possible. Your immeasurable feedback, insight, and constant encouragement have improved my personal and academic growth and development throughout this phase of my developmental experience. I am most thankful for the opportunity to work under your tutelage.

I am also immensely grateful to the members of my thesis committee, Dr. Mousa Al-Bashrawi and Dr. Faleh Althiyabi for their time, constructive additions and feedback which have greatly improved the quality of this thesis work. Your insight has been important in reshaping my research.

I would like to extend my gratitude to the government of Saudi Arabia for providing me with the immense scholarship to study at the prestigious King Fahd University of Petroleum and Minerals. Additionally, I am most thankful to the management and staff of KFUPM for providing the right environment and resources that will enable the success of the project. Finally, I recognize the contributions of the Dean, faculties and all the members of KFUPM Business School for providing necessary resources and access to labs for the running and execution of the data mining processes.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	ix
ABSTRACT	xi
ملخص الرسالة	xiii
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 LITERATURE REVIEW	5
2.1 Cardiovascular Diseases	5
2.2 Forecasting Modeling in Healthcare Systems Emergency Department.....	7
2.2.1 Time Series Approach	7
2.2.2 Regression Methods Approach.....	8
2.2.3 Time Series Regression Approach.....	9
2.2.4 Stacking Ensemble Model (SEM)	10
CHAPTER 3 METHODOLOGY	12
3.1 Forecasting Models.....	13
3.2 Study Design, Data Source and Data Description	15
3.3 Data Preparation and Preprocessing	16
3.4 Data Analysis.....	17
3.5 Forecasting Model Development	18
CHAPTER 4 RESULTS.....	21
4.1 Descriptive Analysis	21
4.2 Monthly CVD Time Series Overview	24
4.3 Forecasting Performance	26
4.4 Visualization of Actual data vs Forecasted data for each Models	28
CHAPTER 5 CONCLUSION AND PRACTICAL IMPLICATIONS.....	32
CHAPTER 6 LIMITATIONS AND FUTURE WORK	34
REFERENCES.....	36
APPENDIXES.....	49
Appendix 1: CVD Keywords	49
Appendix 2: Text Normalization	50
VITAE.....	51

LIST OF TABLES

Table 1 Forecasting Algorithms Used: Time Series vs. Regression-Based Models.....	15
Table 2 Most Frequently Reported Cardiovascular Diseases	21
Table 3 AI Forecasting Models Performance	27

LIST OF FIGURES

Figure 1 Methodology framework	13
Figure 2 Data preparation	17
Figure 3 Data Analysis.....	18
Figure 4 StackingRegressor	19
Figure 5 Forecasting models development	20
Figure 6 Distribution of CVD Diagnoses by Gender.....	22
Figure 7 Distribution of Patients by Severity Level	23
Figure 8 Average Hospital Stay Duration by Severity Level	24
Figure 9 CVD Admission Count Trend	25
Figure 10 Train/Test Splitting of Data	26
Figure 11 forecast models (a).....	30
Figure 12 forecast models (b)	31

LIST OF ABBREVIATIONS

ACS	:	Acute Coronary Syndrome
AHA	:	American Heart Association
AI	:	Artificial Intelligence
ANOVA	:	Analysis of Variance
ARIMA	:	Auto-Regressive Integrated Moving Average Regression
ARMA	:	Auto-Regressive Moving Average Regression
CVD	:	Cardiovascular Disease
DL	:	Deep Learning
ED	:	Emergency Department
EHR	:	Electronic Health Record
ES	:	Exponential Smoothing
ESC	:	European Cardiology Society
GBDT	:	Gradient Boosting Decision Tree
GDP	:	Gross Domestic Product
IEEE	:	Institute Electrical and Electronics Engineers
IHCA	:	In-hospital Cardiac Arrest
KSA	:	Kingdom of Saudi Arabia
LightGBM	:	Light Gradient Boosting Machine
LR	:	Linear Regression
LSTM	:	Long Short-Term Memory
MAE	:	Mean Absolute Error

MAPE	:	Mean Absolute Percentage Error
ML	:	Machine Learning
MLR	:	Multiple Linear Regression
PDPL	:	Personal Data Protection Law
RF	:	Random Forest
RMSE	:	Root Mean Squared Error
SARIMA	:	Seasonal Auto-Regressive Integrated Moving Average
SARIMAX	:	Seasonal Auto-Regressive Integrated Moving Average + Exogenous Variables
SEM	:	Stacking Ensemble Model
SMOTE	:	Synthetic Minority Oversampling
STLM	:	Seasonal Decomposition
StructTS	:	Structural Time Series Model
SVR	:	Support Vector Regression
WHO	:	World Health Organization
XGBoost	:	Extreme Gradient Boosting

ABSTRACT

Full Name : Obasi, Chigozie Francis
Thesis Title : Forecasting Cardiovascular Diseases of Adult Patients in Emergency
Department using AI Models
Major Field : Business Studies, Management Information System Track
Date of Degree : May, 2025

Background: The impact of CVDs on the economy of Saudi Arabia and mortality of Saudis continue to show worrying trends. In 2016, the CVDs effect on the KSA economy was estimated to be \$3.5 billion and is expected to increase to \$9.8 billion in 2035. It has been reported that CVDs account for approximately 28% of all deaths caused by the major non-communicable diseases in KSA, resulting in a direct and indirect economic burden of around SAR 52 billion according to the Ministry of Health. Furthermore, in KSA, over 30% of adults are at risk of CVD. In view of this, there is a need to be involved in scholarly work to find better techniques in handling CVD cases, especially in the health systems emergency department (ED) where patients get their first contact with medical practitioners at the hospital. However, overcrowding at ED is a major global challenge. In Saudi Arabia, there is limited study on the percentage of CVD patient within the midst of overcrowded patients in ED, top CVD risk factors and accurate forecast of CVD patient volume in ED.

Objectives: This study aimed at analyzing the collected data in the Emergency Cases at Jubail Health Services Program from 2019-2023 to identify percentage of CVD cases amongst the overcrowded medical cases, find the top CVD risk factors in Jubail and forecast the volume of CVD patients for the next 12 months at the Jubail health services.

Methods: This retrospective study involved collecting data from the Saudi Open Data Portal. Specifically, the data downloaded contains information relating to “emergency cases in Jubail Health Services Program” from 2019 to 2023 dataset. Python programming language was used in integrating the different years record into one merged dataset. On the merged dataset, python regex functionality was used to filter out CVD cases based on domain-specific keywords such as "chest pain," "hypertension," "heart failure, etc. seen on the “main diagnosis” column based on CVD expert opinion. Thereafter, text normalization and monthly aggregate volume count were applied on the filtered dataset to create the

overall database used to answer the proposed research questions. Descriptive and forecasting analysis was done using python. Forecast models include statistical time series models (ARMA, ARIMA, and SARIMA) and machine learning models (Lasso, Ridge, Random Forest, Support Vector Regression, LightGBM, XGBoost and Stacking ensemble regression). Models' performance was measured using MAPE and RMSE.

Results: Out of 418,835 emergency visits, 9,968 were related to cardiovascular diseases, accounting for 2.38% of total visits from 2019 to 2023. Among the CVD cases, the top five CVD risk factors are Chest Pain, Hypertension, Stroke, Heart Failure, Cardiac Arrest. Additionally, the performance of forecasting models to predict monthly CVD patients were evaluated with MAPE and RMSE. With MAPE of 7.92% and RMSE of 18.96 XGBoost emerged as the best-performing model in terms of MAPE and RMSE followed closely by StackingEnsembleRegressor, Random Forest and ARIMA. Using RMSE, and except for ARIMA, most of the machine learning models outperform the statistical time series models of ARMA and SARIMA.

Conclusion: This study provides a data-driven analysis of cardiovascular disease cases in Saudi Arabian emergency departments and forecasts monthly future cases using multiple predictive models.

ملخص الرسالة

الاسم الكامل: أوباسي، تشيغوزي فرانسيس

عنوان الرسالة: التنبؤ بأمراض القلب والأوعية الدموية لدى المرضى البالغين في قسم الطوارئ باستخدام نماذج الذكاء الاصطناعي

التخصص: دراسات الأعمال، مسار نظام المعلومات الإدارية

تاريخ الدرجة العلمية: مايو 2025

الخلفية: لا يزال تأثير أمراض القلب والأوعية الدموية على اقتصاد المملكة العربية السعودية ومعدل وفيات السعوديين يُظهر اتجاهات مثيرة للقلق. في عام 2016، قُدِّر تأثير أمراض القلب والأوعية الدموية على اقتصاد المملكة العربية السعودية بنحو 3.5 مليار دولار أمريكي، ومن المتوقع أن يرتفع إلى 9.8 مليار دولار أمريكي في عام 2035. وقد أفادت التقارير أن أمراض القلب والأوعية الدموية تُمثل ما يقرب من 28% من جميع الوفيات الناجمة عن الأمراض غير المعدية الرئيسية في المملكة العربية السعودية، مما ينتج عنه عبء اقتصادي مباشر وغير مباشر يبلغ حوالي 52 مليار ريال سعودي وفقاً لوزارة الصحة. علاوة على ذلك، فإن أكثر من 30% من البالغين في المملكة العربية السعودية معرضون لخطر الإصابة بأمراض القلب والأوعية الدموية. وفي ضوء ذلك، هناك حاجة إلى المشاركة في العمل الأكاديمي لإيجاد تقنيات أفضل في التعامل مع حالات أمراض القلب والأوعية الدموية، وخاصة في قسم الطوارئ في حيث يحصل المرضى على أول اتصال لهم مع الممارسين الطبيين في المستشفى. ومع ذلك، (ED) أنظمة الصحة فإن الاكتظاظ في قسم الطوارئ يمثل تحدياً عالمياً كبيراً. في المملكة العربية السعودية، توجد دراسة محدودة حول نسبة مرضى أمراض القلب والأوعية الدموية ضمن المرضى المكتظين في قسم الطوارئ، وعوامل الخطر الرئيسية لأمراض القلب والأوعية الدموية، والتنبؤ الدقيق بحجم مرضى أمراض القلب والأوعية الدموية في قسم الطوارئ.

الأهداف: هدفت هذه الدراسة إلى تحليل البيانات التي تم جمعها في حالات الطوارئ في برنامج خدمات الجنبيل الصحية للفترة من 2019 إلى 2023 لتحديد نسبة حالات أمراض القلب والأوعية الدموية بين الحالات الطبية المكتظة، والعثور على عوامل الخطر الرئيسية لأمراض القلب والأوعية الدموية في الجنبيل، والتنبؤ بحجم مرضى أمراض القلب والأوعية الدموية. الدورية للأشهر الـ 12 المقبلة في خدمات الجنبيل الصحية.

المنهجية: تضمنت هذه الدراسة الاستراتيجية جمع البيانات من بوابة البيانات المفتوحة السعودية. على وجه التحديد، تحتوي البيانات التي تم تنزيلها على معلومات تتعلق بـ "حالات الطوارئ في برنامج خدمات الجنبيل الصحية" من مجموعة بيانات من 2019 إلى 2023. تم استخدام لغة برمجة بايثون في دمج سجلات السنوات المختلفة في مجموعة بيانات واحدة مدمجة. في مجموعة البيانات المدمجة، استُخدمت وظيفة التعبيرات العادية في بايثون لتصفية حالات أمراض القلب والأوعية الدموية بناءً على كلمات رئيسية خاصة بالمجال، مثل "ألم الصدر" و"ارتفاع ضغط الدم"

و"قصور القلب"، وما إلى ذلك، والتي تظهر في عمود "التشخيص الرئيسي" بناءً على رأي خبير أمراض القلب والأوعية الدموية. بعد ذلك، طبقت معايير النص وحساب الحجم الإجمالي الشهري على مجموعة البيانات المفكّكة لإنشاء قاعدة البيانات الشاملة المستخدمة للإجابة على أسئلة البحث المقترحة. أُجري التحليل الوصفي والتنبؤ باستخدام ونماذج (SARIMA و ARIMA و ARMA) بايثون. تتضمن نماذج التنبؤ نماذج السلاسل الزمنية الإحصائية LightGBM و Support Vector Regression و Random Forest و Ridge و Lasso) التعلم الآلي و RMSE و MAPE تم قياس أداء النماذج باستخدام (Stacking ensemble Regression و XGBoost). النتائج: من بين 418,835 زيارة طوارئ، كانت 9,968 منها مرتبطة بأمراض القلب والأوعية الدموية، وهو ما يمثل 2.38% من إجمالي الزيارات من عام 2019 إلى 2023. من بين حالات أمراض القلب والأوعية الدموية، كانت عوامل الخطر الخمسة الرئيسية لأمراض القلب والأوعية الدموية هي ألم الصدر، وارتفاع ضغط الدم، والسكتة الدماغية، وقصور القلب، والسكتة القلبية. بالإضافة إلى ذلك، تم تقييم أداء نماذج التنبؤ للتنبؤ بمعدلات مرضى أمراض بنسبة RMSE بنسبة 7.92% و MAPE مع RMSE و MAPE القلب والأوعية الدموية الشهرية باستخدام ، يليه RMSE و MAPE كأفضل نموذج أداء من حيث XGBoost 18.96، برز ، وباستثناء RMSE باستخدام ARIMA و Random Forest و StackingEnsembleRegressor ، تتفوق معظم نماذج التعلم الآلي على نماذج السلاسل الزمنية الإحصائية لـ ARIMA و SARIMA و ARMA.

الخلاصة: تقدم هذه الدراسة تحليلاً قائماً على البيانات لحالات أمراض القلب والأوعية الدموية في أقسام الطوارئ في المملكة العربية السعودية، وتتنبأ بالحالات المستقبلية الشهرية باستخدام نماذج تنبؤية متعددة.

CHAPTER 1

INTRODUCTION

Cardiovascular Disease (CVD) continues to be the leading cause of death worldwide as highlighted in a recent report by the American Heart Association (AHA), which stated that CVD accounts for 37% of deaths among people under 70 years of age (Tsao et al., 2023). In the United States, AHA report showed that in 2020, the prevalence of CVD in people from 20 years and above was 127.9 million with about 380 deaths per 100,000 people (Tsao et al., 2023). Also, a report from the European Cardiology Society (ESC) indicated that 26 million people globally suffer from heart disease, 3.6 million are diagnosed every year with more than half of this number expected to die within 2 years (Haq et al., 2018). CVD includes a wide range of conditions such as acute coronary syndrome (ACS), Ischemic heart, valvular, hypertension, heart failure, all of which pose significant challenges for early detection and effective management. The World Health Organization estimated an increasing trend of CVD cases.

In almost all cases of managing CVD, emergency department (ED) is the healthcare facility where patients begin their medical treatments by providing almost 24 hours health services to all patients arriving via walk-in or through an ambulance (Sudarshan et al., 2021). As the first medical station every patient visits, it is prone to congestion. The American College of Emergency Physicians noted that ED congestion is a scenario in which ED needs are more than the current resources for patient care (Sudarshan et al., 2021). It has been noted that ED about fifty percent of EDs is more than the ideal patient to nurse ratio for normal ED beds and patient to nurse ratio for urgent care beds (Rose & Skariah, 2013;

Schneider et al., 2003). Congestion in ED is a worrying situation with many disadvantages. The extensive overcrowding challenge of emergency departments worldwide in healthcare systems brings with it massive negative outcomes which include increasing patient count, death and morbidity rates, and an unoptimized, slow operational capabilities to give important medical services to sick individuals bedeviled with health emergencies (Deng et al., 2024). Past studies have further elaborated that overcrowding leads to high morbidity (Edmondson et al., 2013), mortality (Jo et al., 2014), treatment delays (Hong et al., 2013), increase length of stay at healthcare facilities (Derose et al., 2014), increase cost (Sun et al., 2013), readmission (McCusker et al., 2014) and reduction in patient satisfaction (Tekwani et al., 2013). Previous studies have assessed the factors and effect of ED overcrowding (Darraj et al., 2023). Effect of ED congestion, poor healthcare workers management, has led to increasing possibilities of In-hospital Cardiac Arrest (IHCA) in the ED. IHCA is the unexpected discontinuation of the pumping of the heart in a medical facility (Deng et al., 2024). Data from the American Heart Association's Resuscitation Registry showed that IHCA in inpatient adults is roughly 10 situations per 1000 bed-days equivalent to 290,000 patients annually of which 10% happens in the ED with surviving to discharge if adults IHCA patients put at only 25% to 40% (Holmberg et al., 2019).

Various scholars have proposed varying solutions to resolving overcrowding in ED. These suggestions include input-throughput-output framework (Asplin et al., 2003), upscaling resources, demand handling operation research (Hoot & Aronsky, 2008), speedy evaluation areas (Bullard et al., 2012) and diagnostic decision divisions (Roberts et al., 2010). Apart from these different proposals, developing an automated prediction of future patient arrival quantity is valuable and important in enhancing improvements in planning

and handling of ED operations and invariably solve the issues of ED congestion. Hence, forecasts of CVD patients' arrivals in time intervals like weekly or monthly can facilitate staff management, resource allocation and purchase. The ability to predict patient volume enables healthcare providers to make data-driven decisions about allocating human and material resources. This capability can enhance operational efficiency, reduce costs, and alleviate the challenges associated with overcrowding (Ho et al., 2019). Additionally, many healthcare experts have noted that patient volume and complexity inhibit participation in share decision-making - a framework that fosters collaboration between health professionals and patients (Himmelfarb et al., 2023). Therefore, there is a need to forecast CVD patients' volume. Additionally, despite advances in medical tests, diagnosing CVD remains costly and agonizing for patients, emphasizing the need for more accessible and less invasive diagnostic tools (Rani et al., 2021).

Additionally, Proper analysis of medical data is part of the wider concept called healthcare analytics. Clinical data analytics involves collecting, analyzing and appraising data from different data sources such as electronic health, cost records, patients survey, social media and many others) to assist healthcare firms deliver better quality care, reduce medical cost and improve patient experience (Fawagreh & Gaber, 2020). Furthermore, healthcare firms aim to improve their service; hence, they concentrate on sustainable delivery healthcare operations and optimizing the medical activities they offer (Eschenbrenner & Brenden, 2022). By analyzing huge amounts of medical data, healthcare organizations can get more clarity on medical treatment results and apply learnings to personalized health care plans (Eschenbrenner & Brenden, 2022). Malik et al., (2018) noted that healthcare organizations are utilizing data mining techniques to diagnose and create healthcare plans. Therefore, by

analyzing CVD medical records and identifying results such as the most common CVD among patients, this will help hospitals optimize their medical resources and improve patient outcomes.

To address these health challenges, this paper aims to answer the following questions (1) What is the percentage of CVD patients within the midst of overcrowded patients in ED? (2) What are the top CVD risk factors? and (3) What is the monthly forecast of CVD patient volume in ED? To find answers to these questions, I retrospectively reviewed and analyzed downloaded data from the Saudi Open Data Portal. Specifically, the data downloaded contains data related to “emergency cases in Jubail Health Services Program” from 2019 to 2023 dataset. I employed data mining methods in identifying the percentage of adult cardiac patients in the ED, identifying the top CVD diagnoses leading to ED visits on CVD diseases. Additionally, I developed an AI model to predict the top CVD diagnoses and forecast the number of CVD patient admission.

CHAPTER 2

LITERATURE REVIEW

2.1 Cardiovascular Diseases

Cardiovascular diseases (CVDs) are defined by the World Health Organization (WHO, 2021), as a range of anomalies impacting the heart and blood vessels. According to the American Heart Association, CVDs include coronary heart diseases, cerebrovascular diseases, peripheral arterial diseases, rheumatic heart diseases, congenital heart diseases, deep vein thrombosis, pulmonary embolism and many others (Tsao et al., 2023). The development of CVDs is driven mainly by hereditary, environmental, and lifestyle choices. The key risk factors of CVDs include lack of exercises, eating habits and dietary habits, age, depression, stress and gender (Tsao et al., 2023).

WHO reported that in 2019 CVDs led to 17.9 million deaths, representing an estimated 32% of all global deaths, and 38% of the 17 million non-communicable diseases related deaths for people under 70 were due to CVDs (WHO, 2021). Roth et al., (2020) identified lack of exercise, smoking, aging populations and limitations in accessing healthcare as some of the factors contributing to this medical burden. CVDs continue to pose significant costs to many countries. CVDs and its related prevalent treatments cost \$23 billion in France, about 0.9% of its 2020 gross domestic product (GDP) (Rachas et al., 2022). In 2018, over \$400 billion equivalent to almost 2% of its GDP were spent on CVD expenses in the United States (Benjamin et al., 2018), While the 2009 economic effect of CVDs in Russia was about \$30 billion equivalent to 2.8% of its GDP (Kontsevaya et al., 2013). Despite having one of the best healthcare systems in the world, annually Australia spent

more than \$14 billion on CVD (Figtree et al., 2024). By 2030, it is estimated that the cost of handling CVD worldwide will be in excess of \$1 trillion yearly due to increasing medical needs and elimination of economic productivity (Alhabib et al., 2020). In the Kingdom of Saudi Arabia (KSA), medical treatment burden is estimated to cover more than 8% of the government total budget, equivalent to about \$56 billion; specifically (Tash & Al-Bawardy, 2023). Overall, the impact of CVDs cut across various countries including developed and developing countries.

The Gulf countries, mainly KSA, have experienced a massive increase in CVDs over the past few decades. Increased structural and economic developments, changes in lifestyles have contributed to the high number of obesity, diabetes and hypertension which are all high risk factors for CVD; hence, CVDs such as coronary artery disease, heart failure have been increasing in the Gulf region (Gulf Heart Association, 2022). Over the past decade, KSA has undertaken a major transformation of its healthcare system aimed at achieving easy access to medical functions, improving healthcare system standards and proficiency, and advocating prevention activities against health risks (Tyrovolas et al., 2020). Nonetheless, the impact of CVDs on the economy of KSA and mortality of Saudis continue to show worrying trends. In 2016, the CVDs effect on the KSA economy was estimated to be \$3.5 billion and is expected to increase to \$9.8 billion in 2035 (Tash & Al-Bawardy, 2023). It has been reported that CVDs account for approximately 28% of all deaths caused by the major non-communicable diseases in KSA, resulting in a direct and indirect economic burden of around SAR 52 billion (Ministry of Health KSA et al., 2023). Furthermore, in KSA, over 30% of adults are at risk of CVD diagnosis. Additionally, 35% of Saudi citizens have cholesterol levels above recommended value. This a concerning

statistic compared to other countries (National Health Center & Saudi Health Council, 2022). Overall, this means that, on average, people in KSA are at risk of developing CVDs a decade earlier than those in the western countries.

2.2 Forecasting Modeling in Healthcare Systems Emergency Department

The rapid advancement of AI and its subsets, such as Machine Learning (ML) and Deep Learning (DL), has empowered researchers and AI specialists to develop innovative solutions to improve healthcare quality. In their review paper, Jiang et al., (2023) identified different AI forecasting techniques that have been employed in the emergency department. These methods include the time series, regression, time series regression and deep learning methods. Jiang et al., (2023) stated that it remains unclear which method is superior.

2.2.1 Time Series Approach

Time series approaches, in particular, have been employed in making forecasts in healthcare (Marcilio et al., 2013; (Sorjamaa et al., 2007). For instance, various time series algorithms have been applied to predict the daily number of admitted patients in emergency departments, including average models, and Univariate Seasonal Auto-Regressive Integrated Moving Average (SARIMA). Multivariate SARIMA gave the favorable result with a mean absolute percentage error (MAPE) of 7.4% (Kam et al., 2010). Zhu et al., (2017) employed a model that integrates seasonal regression, ARIMA, and other time-series models in generating forecasts of the number of inpatients discharged daily. Furthermore, single seasonal ARIMA and other time-series models were employed to predict the monthly and daily quantity of patients admitted (Zhou et al., 2018). In forecast ED arrivals which were grouped based on treatment, age, admission status and triage, Mai

et al., (2015) used vector ARIMA model to create a forecasting model. Also, ARMA was used to develop a forecasting model by Afilal et al., (2016). Also, Vollmer et al.,(2021) compared time series models of ARIMA, STLM(seasonal decomposition), Structural Time Series Model (StructTS), Exponential smoothing (ES) to forecast ED demands with machine learning models, in which the linear time series showed better performance than the machine learning models. Cheng et al., (2021) used SARIMAX to estimate ED occupancy, yielding a 60% enhancement in MSE for short-term forecasts.

However, it has been reported that time series models have limitations in their generalization ability to handle convoluted nonlinear patterns in multiple features (Hu et al., 2020). Additionally, it is difficult to automate with ARIMA and SARIMA as they need to be retrained when new data are added in the training data series which then affect their practicability (Marcilio et al., 2013). Also, time series algorithms use a recursive prediction strategy (Tran et al., 2009), known to have issues like accumulated inaccuracy and lack details (Sorjamaa et al., 2007).

2.2.2 Regression Methods Approach

According to Shah & Mishra (2024), recent years have seen application of machine learning models to forecast important medical outcomes and improve health activities by giving upscaled capabilities to assess past data, find patterns and create reliable predictions. These strategies are now becoming continuously frequently used in forecasting healthcare systems bed occupancy especially in locations where demand is high like the ED and during global health crises (Vollmer et al., 2021). Jiang et al., (2023) posited that regression is a statistical technique for measuring the cause-and-effect connection among a target variable and at least one predictor variable. They noted that estimated results can

be understood by the relationship between independent variables. Examples of regression-based algorithms used for forecasting in emergency department include generalized linear model, poisson regression and multiple linear regression (MLR) (Jiang et al., 2023). Various studies have used MLR Ho et al., (2019), Zhang et al., (2022), and Boyle et al., (2012). McCarthy et al., (2008), and Marcilio et al., (2013) employed poisson regression method while Yousefi et al., (2019), and Tello et al., (2022) used support vector regression (SVR). Additionally, a binary logistic regression model was developed to forecast patients arrivals in varying timelines with the forecasting model getting the best achievement when fed with the timing of 30 min, 8 hours and 12 hours (Chase et al., 2012).

2.2.3 Time Series Regression Approach

This method integrates the characteristics of both regression and time series methods as It incorporates past trends and causal relationships in one algorithm (Jiang et al., 2023). Fralick et al., (2021) fed the output of a generalized additive model in a random forest model to forecast ED volumes. Also, Marcilio et al., (2013) used a generalized estimating equation that used upcoming patient visit day as the autocorrelation dependence term. Additionally, an integration of linear regression, SARIMAX, and multi-layer perceptron was used to estimate everyday arrivals and occupancy amount, improving healthcare system organizational long term planning (Whitt & Zhang, 2019).

2.2.4 Stacking Ensemble Model (SEM)

While individual machine learning algorithms have been studied and shown to give high performance, Hu et al., (2020) noted that across different application cases, no one individual machine learning method can surpass others and that adding multiple machine learning models may significantly give better performance when compared to the individual model. (Hu et al., 2020; Wolpert, 1992). The ensemble algorithm combines multiple models, each with different strengths and weaknesses, to create a more robust model. Ensemble methods achieve this by using techniques such as bagging, boosting or voting (Ghasemieh et al., 2023). The SEM methodology has been widely used in scientific research and has shown to give improved prediction than each of an individual model. In predicting CVD daily number of hospital admissions. Hu et al (2020) utilized a gradient boosting decision tree (GBDT), support vector regression (SVR), random forest (RF), linear regression (LR) and , extreme gradient boosting (XGBoost) algorithms as the base learners in a stacking ensemble model with RF as the meta learner, the stacking model reduced the MAPE, RMSE and MAE by 6.3%, 7.4% and 6.3% respectively and the R2 improving by 1.7%. Atallah & Al-Mousa, (2019) developed a boosting ensemble model that used and compared the performance of stochastic gradient descent, k-nearest neighbors, random forest, and logistic regression models individually and together as ensemble model with the majority voting method been used to determine the final prediction. With an accuracy of 90%, the ensemble model outperformed the individual models. Also, Lakshmanarao et al., (2021) employed a feature selection technique such as Analysis of Variance (ANOVA) F-value and mutual information to determine useful variables which were used in the model training and applied sampling techniques like

random oversampling, synthetic minority oversampling (SMOTE) and adaptive Synthetic Sampling. The classifier algorithms used were logistic regression, k-nearest neighbors, support vector machine, decision trees, naïve bayes, random forest, and adaboost and the ensemble techniques of stacking and voting. The stacking ensemble model gave an accuracy of 99% on the Kaggle dataset using random oversampling, while the SMOTE gave a 93% accuracy. Furthermore, using a dataset of 1190 occurrences and 14 variables from the Institute of Electrical and Electronics Engineers (IEEE) database, an ensemble model comprising of Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting, AdaBoost and RF with LightGBM been used as the meta learner gave an accuracy of 98% (Deneshkumar et al., 2023).

In developing machine learning models, careful actions are done to identify important features and their correct combination to be used in the creation of accurate prediction models (Kausar et al., 2016). AI models depend heavily on training and test data consistency; feature selection techniques assist to get the data ready in a state to make a reliable forecast. After relevant features are selected, prediction algorithms can be employed to forecast disease likelihood (Ghosh et al., 2021).

CHAPTER 3

METHODOLOGY

This section outlines the research methodology employed in this study including data collection, data preparation, data analysis, forecasted model development, and model performance assessment (Figure 1). It also provides summaries of the algorithms used, and the dataset description. Having reviewed literature to identify different forecasting algorithms being employed, the first step was to further select and review these algorithms that have been utilized in the forecasting literature. Next, data preparations were done to get the data ready for analysis and model development. Thereafter, the identified models were then trained and tested on the selected dataset. Model performance measurement and comparison was done to identify the best performing forecasting model.

The paper aims to address three key questions: (1) What are the top CVD diagnoses? (2) What is the percentage of CVD in the emergency department and (3) What is the forecasting monthly cardiovascular disease diagnoses in the Saudi Arabia Healthcare system using real-world secondary data sources from the Saudi Open Data Portal.

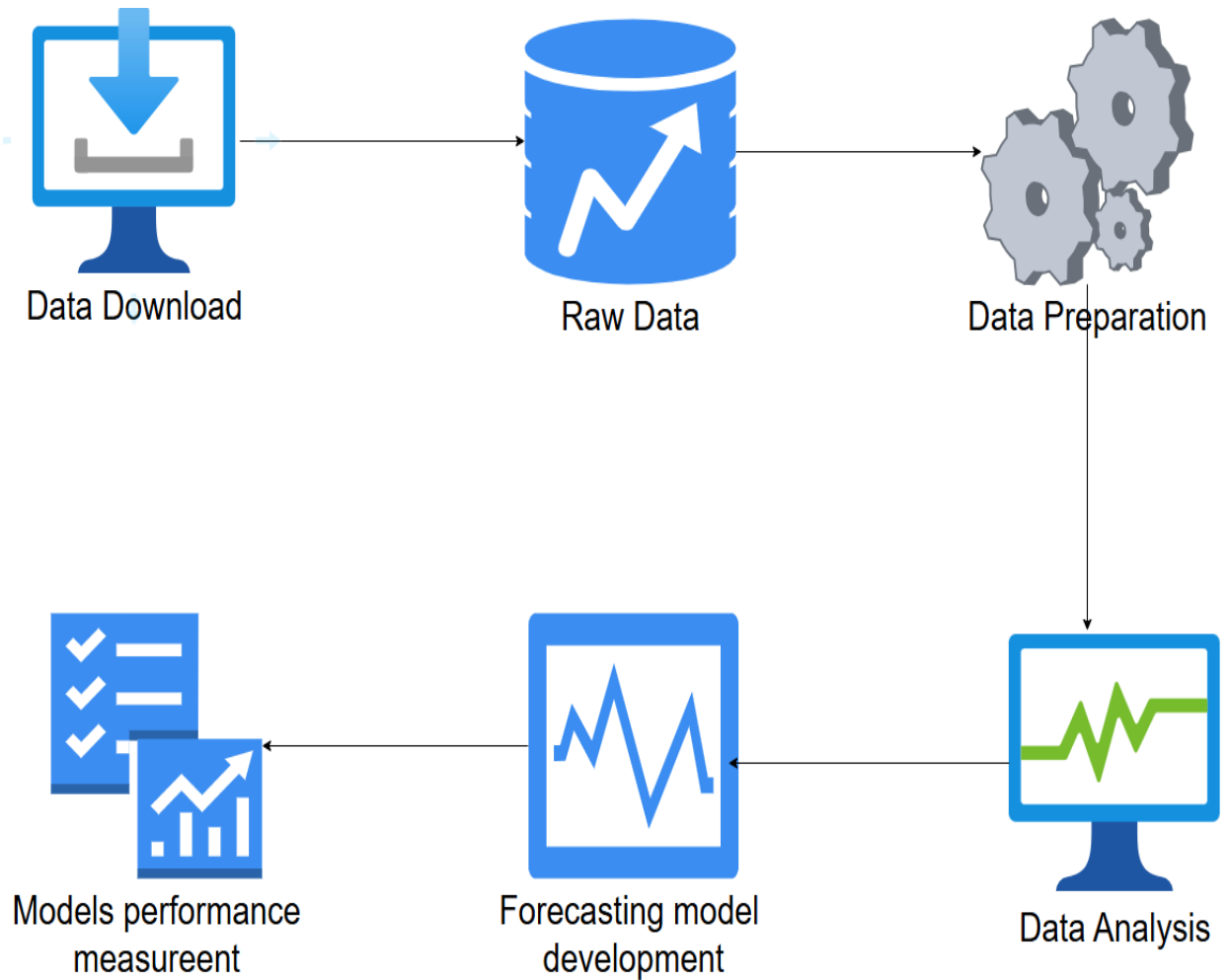


Figure 1 Methodology framework

3.1 Forecasting Models

This section discusses the algorithms used to develop forecasting models. Comparison is made between the performance of statistical time series models and machine learning regression models (Table 1).

We used two varying types of models for our estimation: time series and machine learning models (Makridakis et al., 2020). A distinct time series is a chain of data elements in sequential order separated into consistent time spans; And the main premise of both kinds

of algorithms is the values are nearer to one another which indicates close behavioral pattern and that there exists an association between data points noted at the same point within the time cycle (Vollmer et al., 2021). Time series models employ event sequence and defined time span to make deduction and sever the time series into varying linear elements like residual, seasonality and trend. The residual is postulated to have a certain correlative pattern that is mostly modelled via autoregressive random sequence or exponential decay from which estimated data are forecasted based on previous figures (Makridakis et al., 2020). On the other hand, machine learning models define a given function class like trees or smooth curves with enough capacity to learn complicated formulas (Vollmer et al., 2021). Both methods were applied to maximize their individual strengths. For example, Hu et al., (2020) stated that statistical time series models are unable to handle convoluted nonlinear data as it assumes that the data is stationary and may not perform well with irregularities in data. Regression models can handle non-stationary data. Additionally, current study in our field increasingly adopts hybrid approaches. By following this, our study remains methodologically relevant and contributes meaningfully to ongoing discussions.

Table 1 Forecasting Algorithms Used: Time Series vs. Regression-Based Models

Time Series Models	Regression Models
Auto-regressive Moving Average (ARMA)	Linear Regression
Auto-regressive Integrated Moving Average (ARIMA)	Random Forest
SARIMA	Support Vector Machine
	LightGBM
	Ridge
	XGBoost
	StackingEnsembleRegressor

3.2 Study Design, Data Source and Data Description

This work utilized a retrospective observational design in executing this research. Real-world secondary data employed in this research was downloaded from the Saudi Open Data Portal. Specifically, the data downloaded contains data related to “emergency cases in Jubail Health Services Program” from 2019 to 2023 dataset (Royal Commission for Jubail and Yanbu, 2019, 2020, 2021, 2022, 2023). Across the different years, each patient visit record contains variables such as visit and discharge timestamps, hospital visits, gender, severity level, main diagnosis, Waiting Time (Minutes), and Length of Stay (Minutes). This data was utilized as it contains information relating to medical emergency healthcare cases within the Saudi Arabia context. Additionally, the Jubail dataset is the available Saudi Arabia real-world secondary data on the government Saudi Open Data Portal.

3.3 Data Preparation and Preprocessing

The data preparation and preprocessing involve activities which include integrating the 2019 to 2023 datasets into a combined dataset using a unique identifier generated by combining extracted year and rownum columns into a single column on each year's data. A likely CVD dataset was created from the merged dataset. On the merged dataset Main Diagnosis column, using python regex, the column was filtered to extract CVD domain-specific keywords such as "chest pain," "hypertension," "heart failure," "myocardial infarction," and "stroke." were categorized as likely CVD cases. This keyword insight was obtained based on expert opinion. (See appendix 1). Additional text normalization was performed to standardize synonymous diagnoses (See Appendix 2). To develop the forecast model, I aggregated the total number of monthly CVD cases from the likely CVD dataset created earlier. Monthly aggregates were computed to construct a time series reflecting CVD cases over 60 months. The resulting data was split into training (January 2019 to December 2022) and testing (January to December 2023) sets.

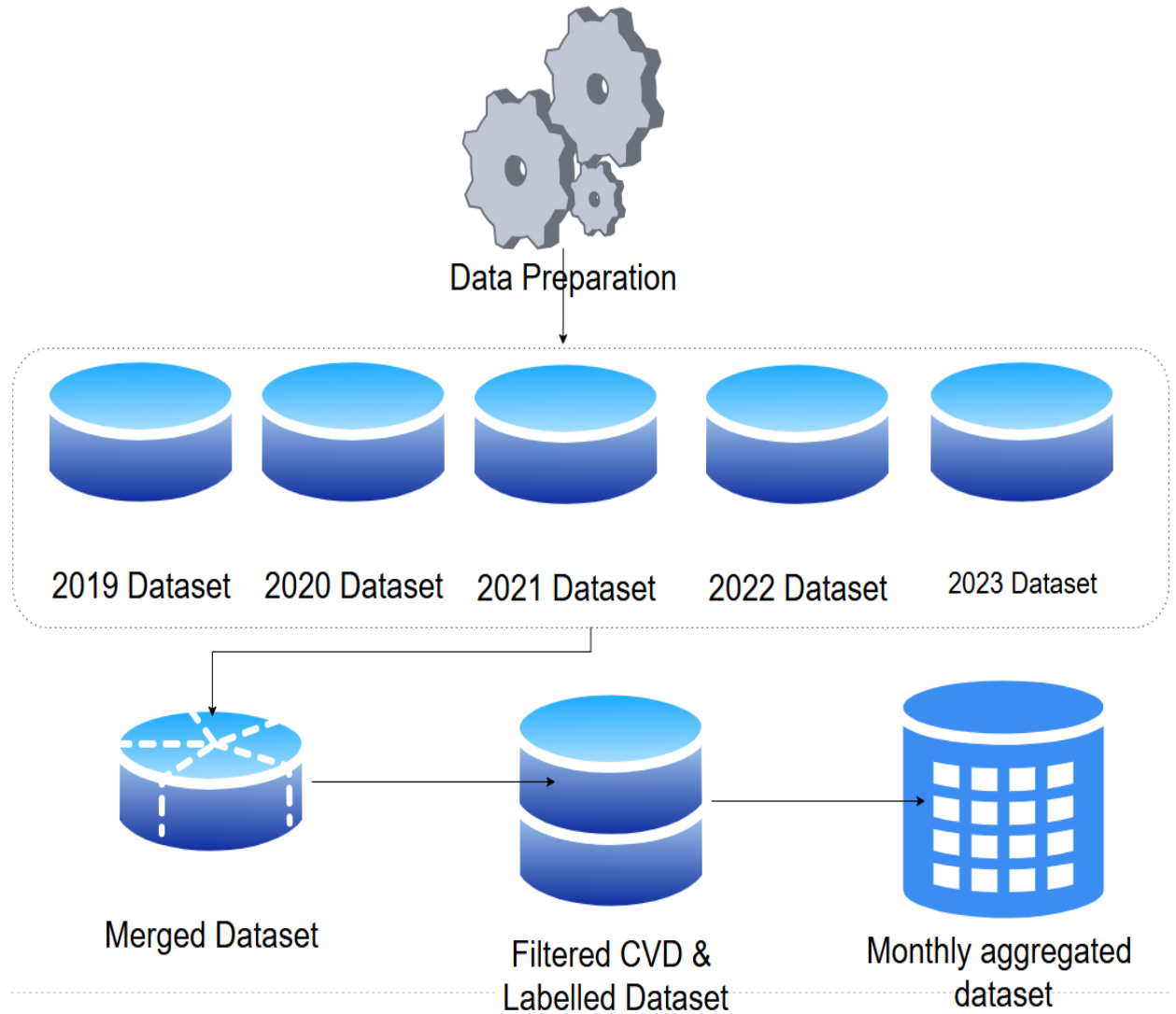


Figure 2 Data preparation

3.4 Data Analysis

To answer the first two research questions, descriptive data analysis was done using python to get insights into the percentage of CVD present in the overall emergence crowded visits, the top CVD risk factors, Additional analysis to derive the male to female CVD patients ratio, the number of patients with varying severity level of CVD and the impact of CVD severity on patients length of stay at the ED was assessed.

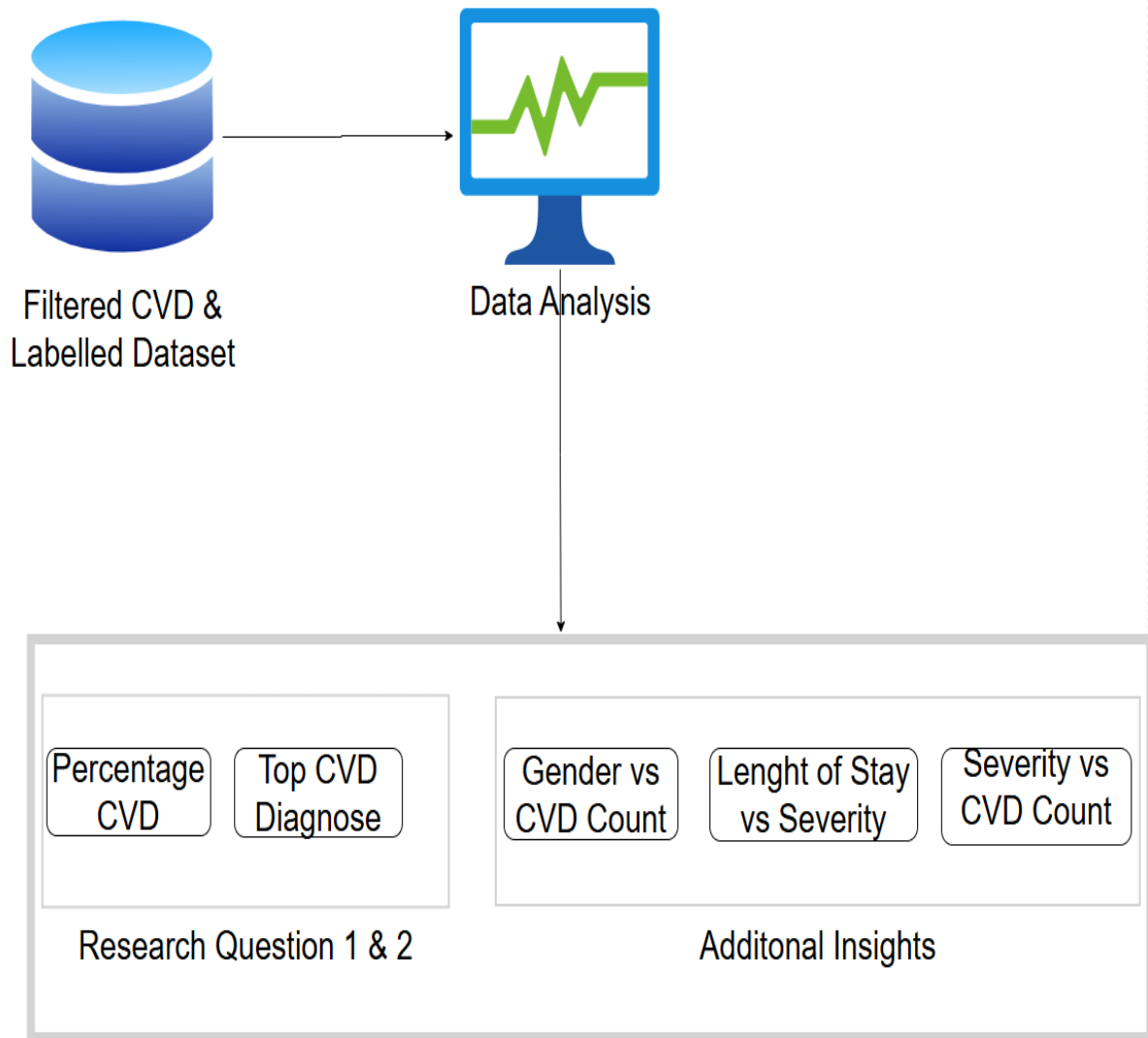


Figure 3 Data Analysis

3.5 Forecasting Model Development

Forecasting Models and Implementation To forecast CVD diagnoses, monthly aggregates of CVD cases were computed based on the patient arrival timestamp, creating a time series from January 2019 to December 2023. A comparative evaluation of classical time series models and machine learning approaches was conducted. The following models were implemented:

- **Statistical Models:** ARMA, ARIMA, and SARIMA were implemented using the statsmodels library. These models were evaluated for their ability to handle stationarity, trend, and seasonality in the monthly CVD time series.
- **Machine Learning Models:** Recursive multi-step forecasting was carried out using skforecast (Amat Rodrigo & Escobar Ortiz, 2023), combining lag features and algorithms like Lasso Regression, Ridge Regression, Random Forest, Support Vector Regression, LightGBM, XGBoost and Stacking ensemble regression.

Stacking ensemble regression is developed by combining the forecasting properties of Lasso, Ridge, Random Forest, Support Vector Regression, LightGBM and XGBoost. XGBoost served as the final estimator, giving it better performance in terms of MAPE and RMSE.

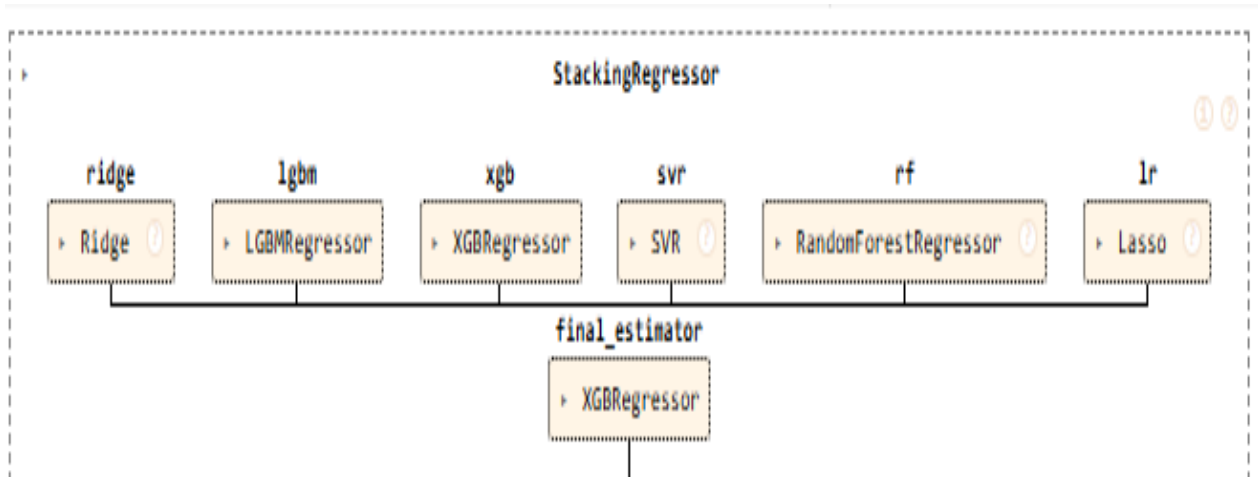


Figure 4 StackingRegressor

- The dataset was separated into two parts: training and testing. Data from the years 2019 to 2022 representing 48 months of historical data were used for the training. While data for the year 2023 was used to ascertain (test) the forecast models performance. Each model used a lag of 12 months, with the final 12 months held out as the test set and performance was evaluated using MAPE and Root Mean Squared Error (RMSE).

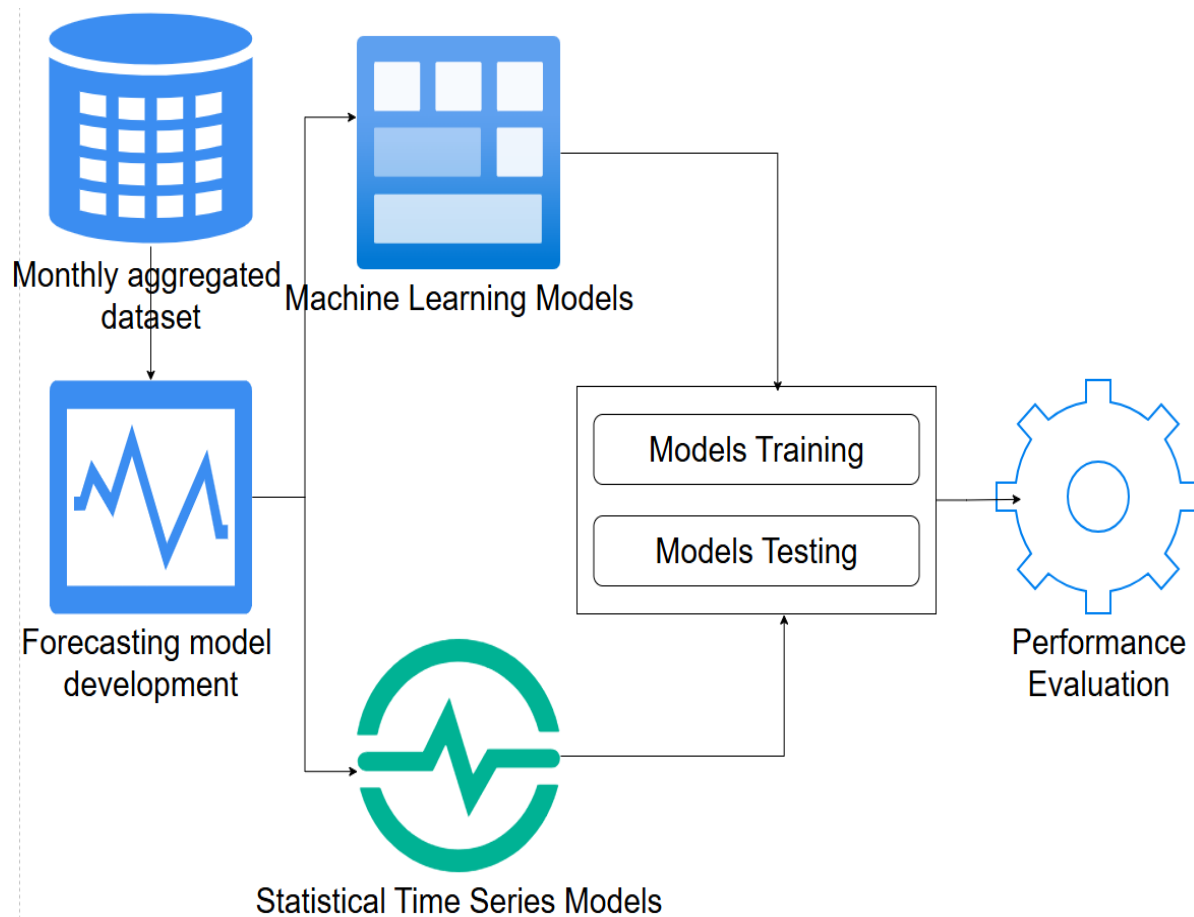


Figure 5 Forecasting models development

CHAPTER 4

RESULTS

4.1 Descriptive Analysis

Out of 418,835 emergency visits, 9,968 were related to cardiovascular diseases, accounting for 2.38% of total visits. The top diagnoses included are chest pain, hypertension, stroke, heart failure, cardiac arrest (Table 2)

Table 2 Most Frequently Reported Cardiovascular Diseases

Top Cardiovascular Diseases	Number of Cases
Chest Pain	8,702 cases
Hypertension	362 cases
Stroke	362 cases
Heart Failure	77 cases
Cardiac Arrest	55 cases

Additionally, from the dataset, there are more male patients with CVD diagnosis than female patients (figure 6).

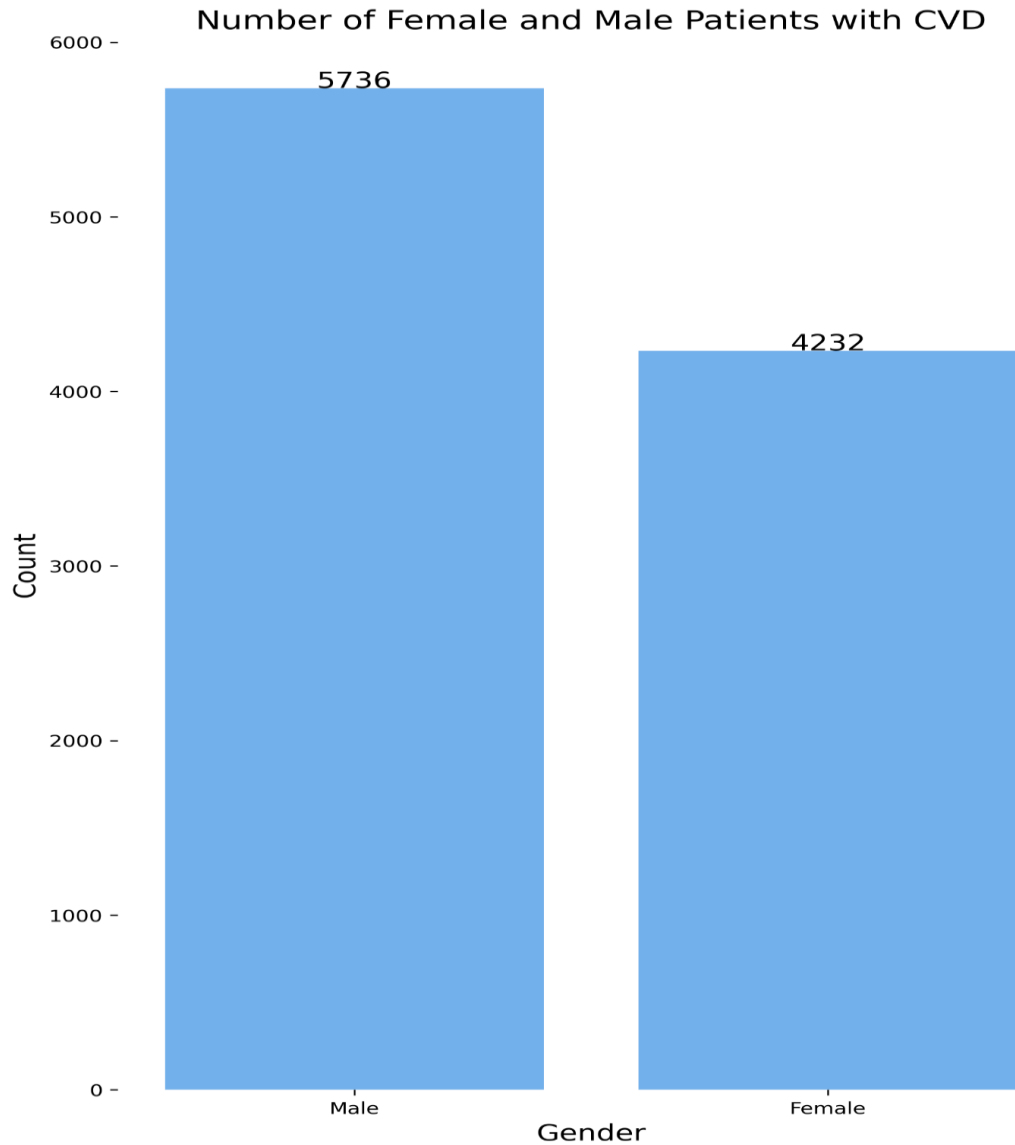


Figure 6 Distribution of CVD Diagnoses by Gender

Furthermore, we defined CVD severity as the level to which CVD illness has progressed with its resultant effect on medical resource usage, presence of varying medical coexisting health challenges on the patient and the risk of patient's death (Zghebi et al., 2021). Based on the data, higher levels indicate a high severity degree for the recorded patient. Figure 7 indicates that the severity level of most CVD patients at the emergency department was at level III followed by level IV and level II revealing patients suffered from mild to serious CVD cases.

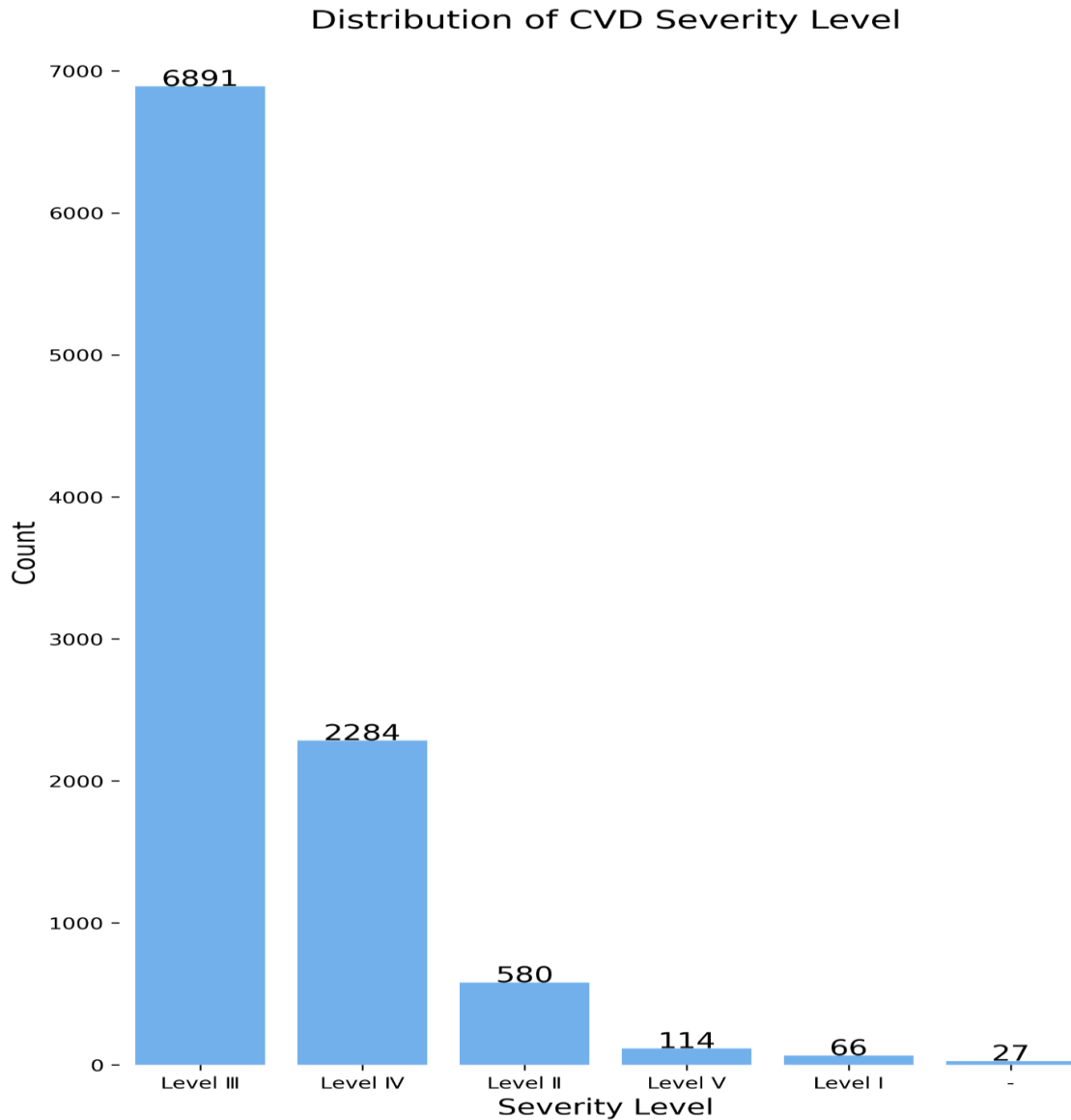


Figure 7 Distribution of Patients by Severity Level

From details on Figure 8, It could also be seen that severity level II and level III had the highest average length of stay, while the highest level IV and V had the lowest average length of stay. This may be because patients who displayed severe CVD symptoms may have been moved quickly out of the emergency care to specialist care for focus disease treatment operations.

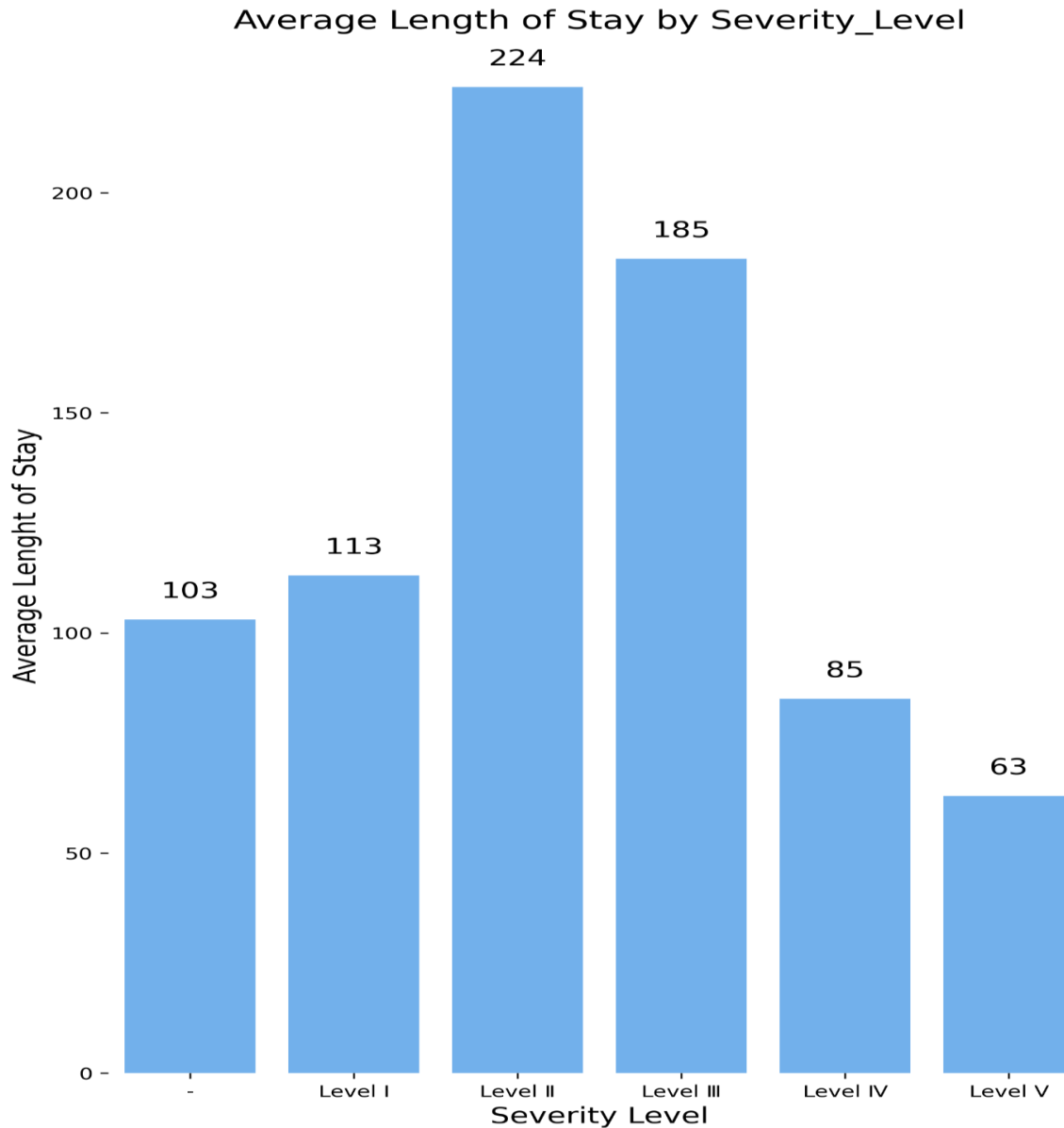


Figure 8 Average Hospital Stay Duration by Severity Level

4.2 Monthly CVD Time Series Overview

Details seen on figure 9 indicate that the peak admission count of CVD patients at the Jubail emergence healthcare system between 2019 and 2023. There was a steep downward trend in 2020. This is mostly due to the COVID-19 pandemic that impacted physical medical care as efforts were made to curtail its spread. Additionally, figure 10 displayed the visual of the splitting of data into train and test data.

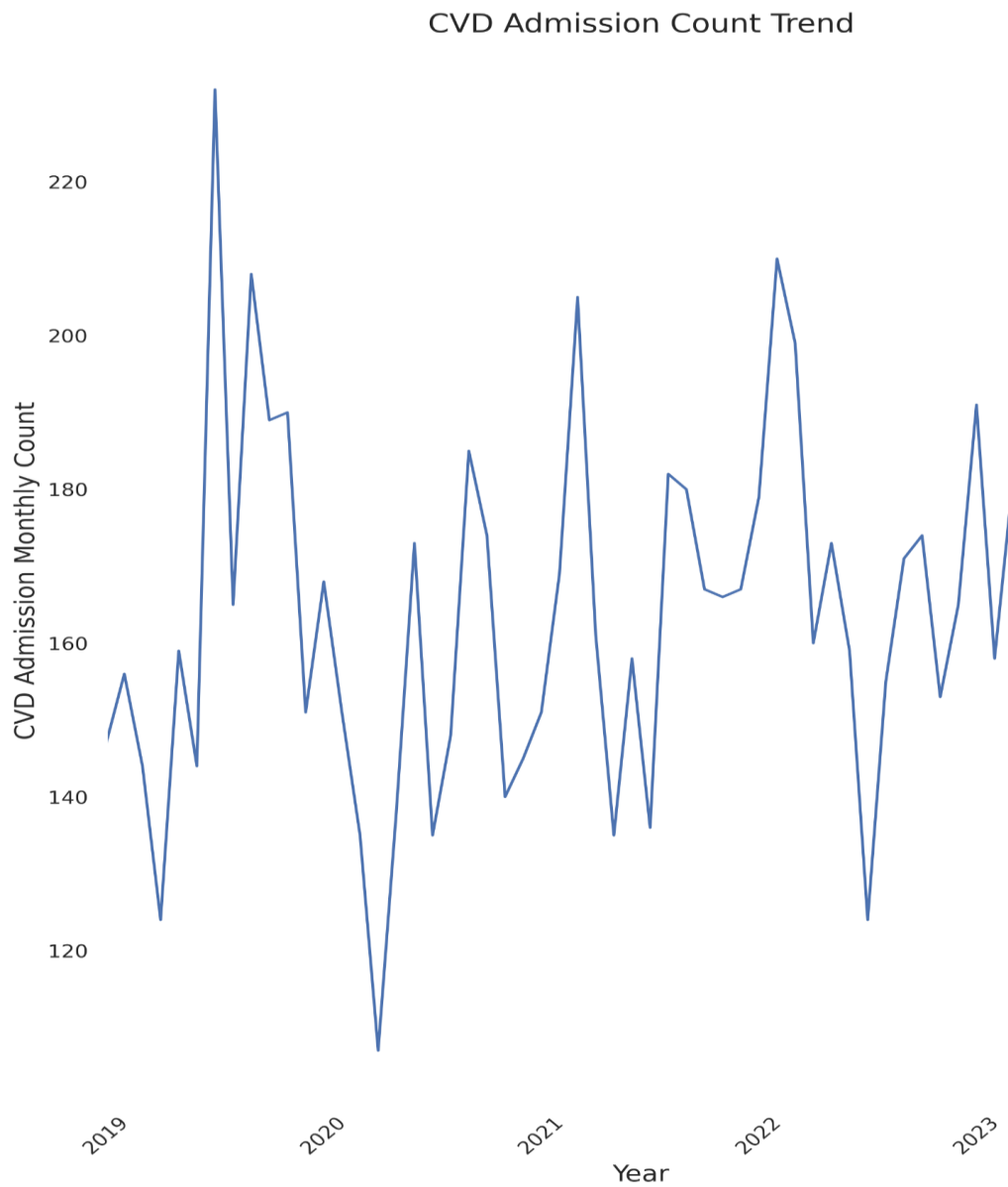


Figure 9 CVD Admission Count Trend

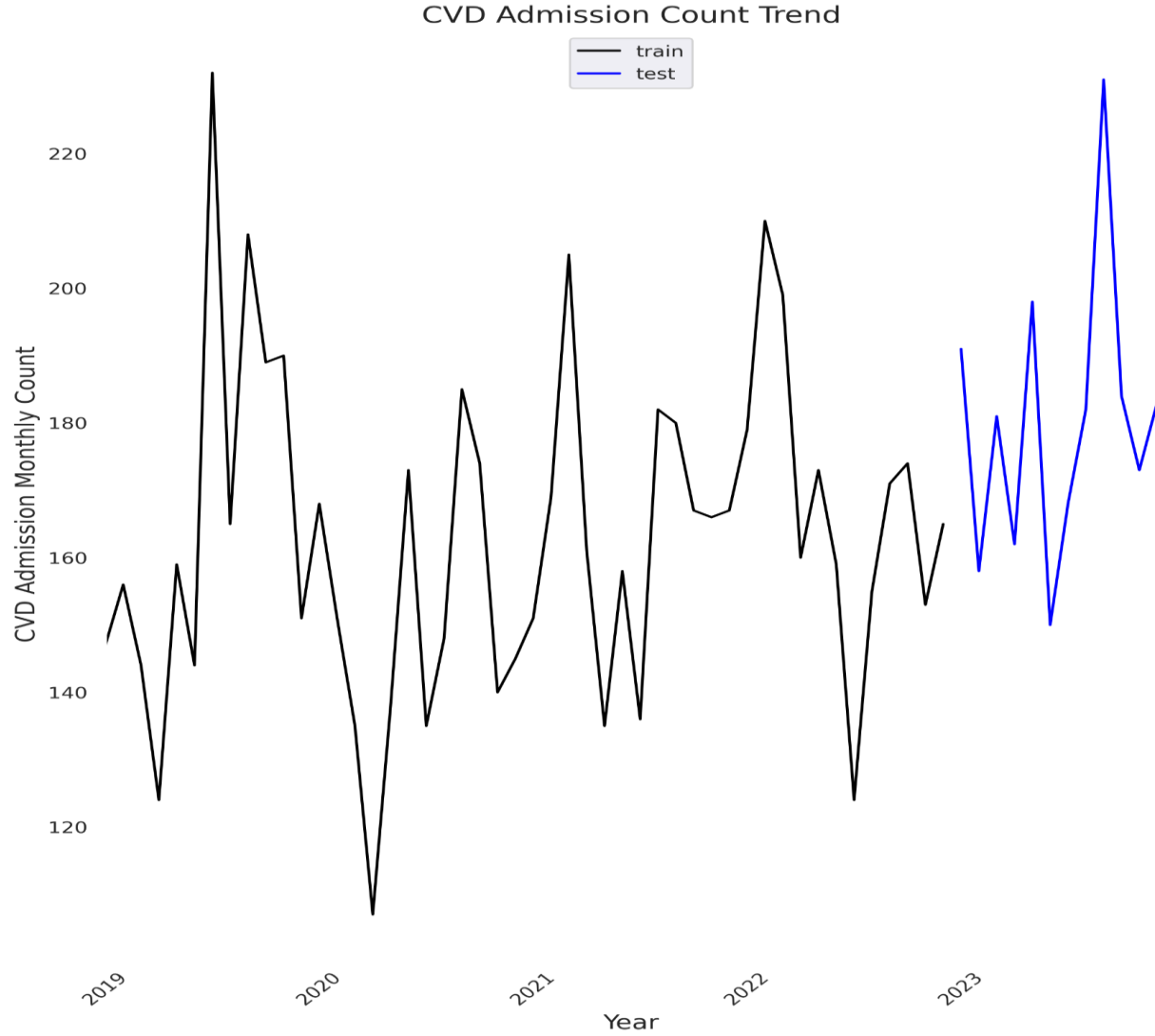


Figure 10 Train/Test Splitting of Data

4.3 Forecasting Performance

We used two performance metrics to measure the performance of the forecast models, root Mean Squared Error (RMSE), and the Mean Absolute Percentage Error (MAPE) (Cihan, 2024; Fan et al., 2022) (Formula 1 and 2). Where n indicates the number of observations, x_i and y_i denotes actual and forecasted values respectively.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}$$

Formula (1)

$$\text{Formula (2)} \quad \text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \times 100$$

RMSE indicates the deviation between the forecasted and real values and it is the standard deviation of predicted errors while MAPE represents the mean amount of percentage of forecasted errors (Cihan, 2024). These metrics reflect a high level of model accuracy when their values are low, as this suggests more accurate and reliable predictions.

The comparative results of all forecasting models on the test set (last 12 months) are summarized in Table 3. As shown in the table, XGBoost emerged as the best-performing model in terms of MAPE and RMSE followed closely by StackingEnsembleRegressor, Random Forest and ARIMA. Using RMSE, and except for ARIMA, most of the machine learning models outperform the statistical time series models of ARMA and SARIMA.

Table 3 AI Forecasting Models Performance

Model	MAPE	RMSE
ARMA	11.45%	28.59
ARIMA	10.93%	22.22
SARIMA	28.39%	58.44
Lasso	11.24%	26.04
Ridge	11.25%	26.03
RF	10.73%	24.7
SVR	12.65%	28.57
LightGBM	11.04%	27.87
XGBoost	7.92%	18.96
StackingEnsembleRegressor	10.56%	25.78

4.4 Visualization of Actual data vs Forecasted data for each Models

Figure 11a and 11b highlights the comparison between the actual and forecasted values of each model. From figure 11a, it could be seen that the linear regression model shows good performance for low-volatility periods but found difficulty in adapting to non-linear spikes. It under forecast peaks and overestimates troughs. Random forest model forecasted values relatively follow the overall trend of rise and dipping of the test values but with identifying deviations in magnitude. For high admission months like May (198) and September (231) the model greatly under forecast the actual values. However, the forecast for April (actual: 162, forecast: 169) and October (actual: 184, forecast: 162) is close to test values. Overall, the model underestimates actual figures mostly for periods with sharp spikes. Its strengths lie in capturing the general trends, seasonality and forecast well for stable months. Support vector regression model follows the general trend directions for months with limited variation, follows the seasonal wave of rise and fall of the data, but misses months of extreme fluctuations. It performs reasonably well in April, June and November. It does not estimate well with months of spikes and does not generalize well to sharp upwards trends. The LightGBM model forecasts a single figure for the twelve months. It ignores the fluctuations and offers no variation in its forecast. It did not learn time-based changes in the data.

Additionally, from figure 11b, the ARMA model predicts stable CVD admission when the admission increases. However, the ARMA model is limited in that it fails for non-stationary time series and does not capture seasonality. ARIMA model forecasted a steady increase in the number of CVD patients monthly. It did not consider the steep fluctuations when compared with the actual values for 2023 data. Also, the SARIMA model forecasted

nearly opposite values when compared with the actual values for 2023 data. That is, it increases when the actual values decrease from the month of July. Additionally, ridge model forecasted values move in the same direction of the actual values, though not perfectly. It shows a good response to the ups and downs but underrepresents sharp peaks. XGBoost forecasted values closely follow actual values in many months. It captures both local dips and peaks and responds reasonably to changes. However, it underestimates the sharp spike in September. StackingEnsembleRegressor forecasts showed great stability and smoothness, however, it found difficulty in handling extreme fluctuations and sudden spikes in admission quantity.

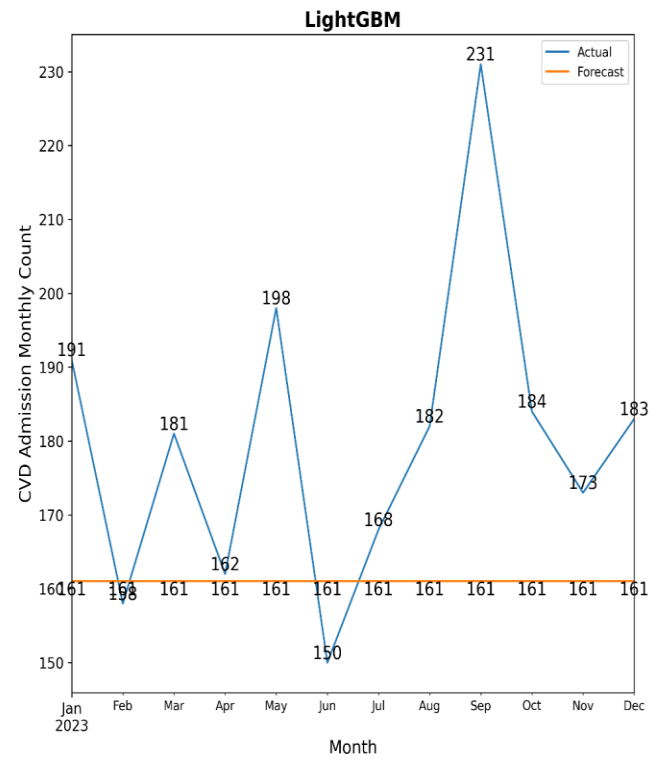
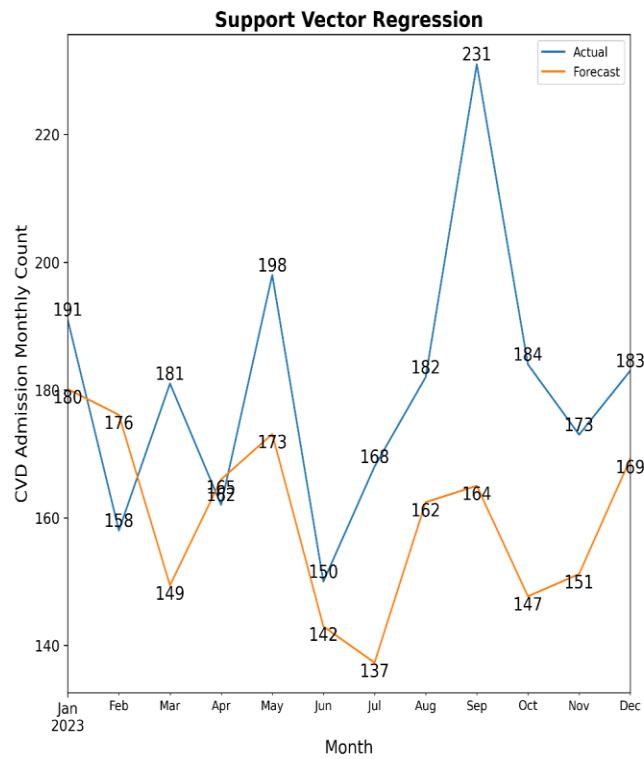
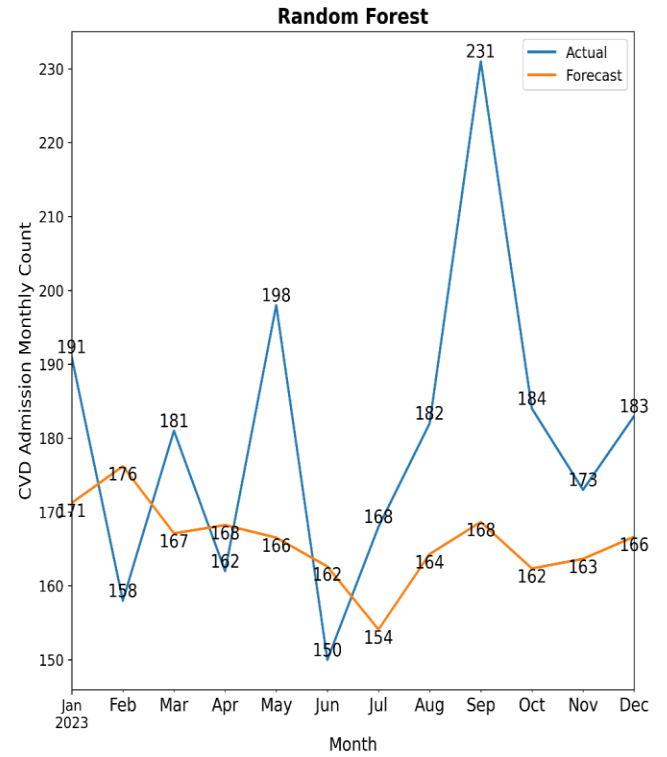
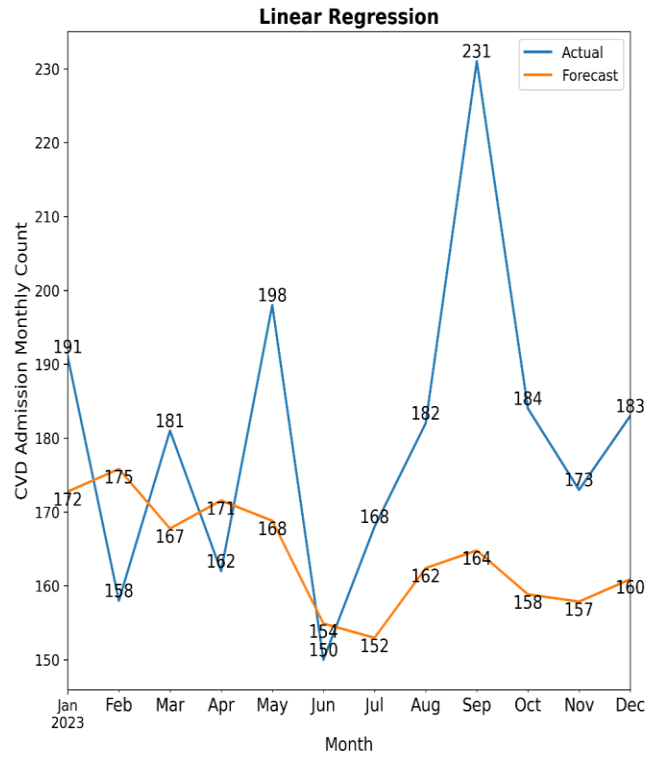


Figure 11 forecast models (a)

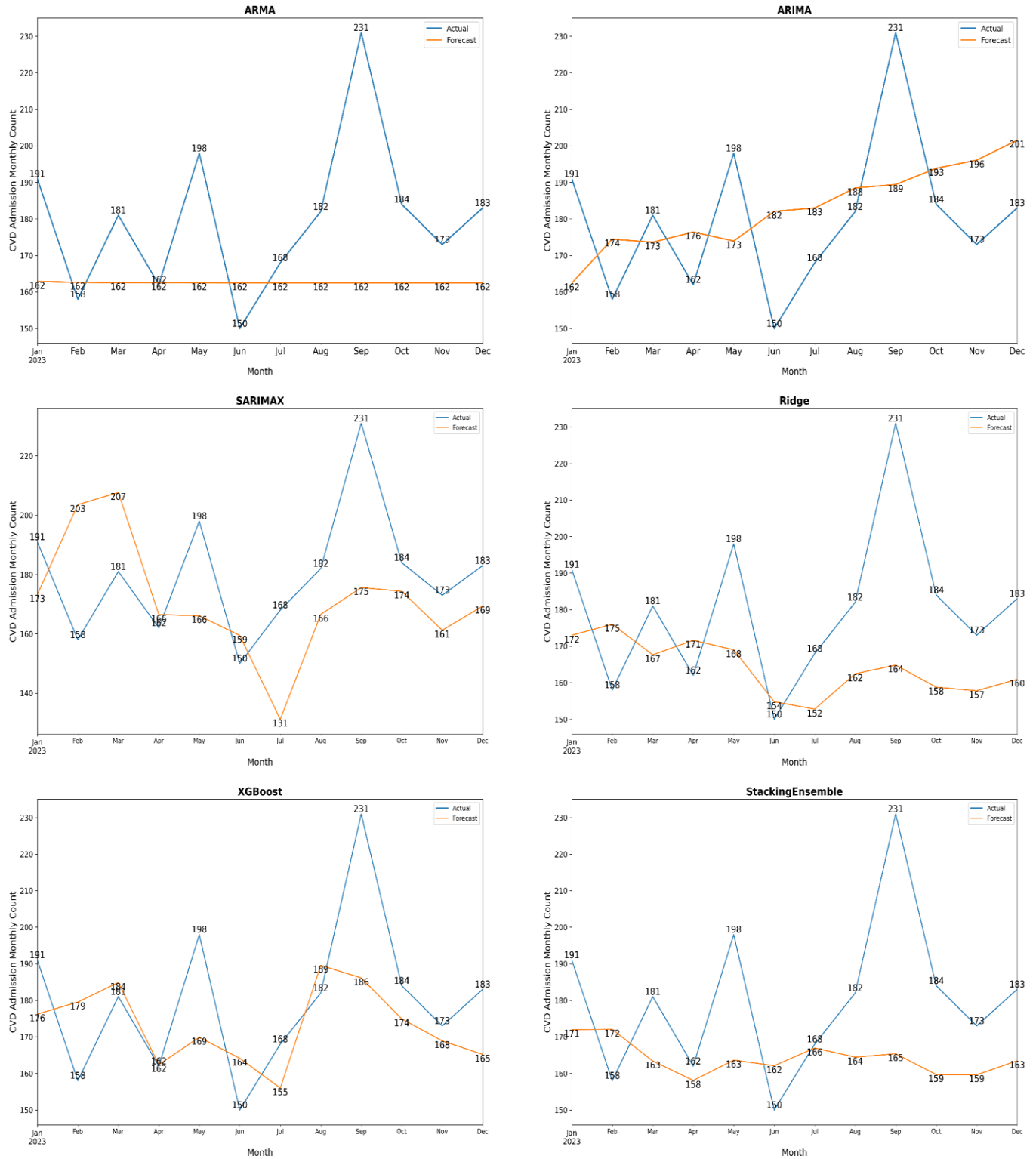


Figure 12 forecast models (b)

CHAPTER 5

CONCLUSION AND PRACTICAL IMPLICATIONS

This study provided a data-driven examination of cardiovascular disease presentations in Saudi Arabia emergency departments and forecasted monthly future cases using multiple predictive models. CVD diagnoses accounted for approximately 2.38% of cases from 2019–2023 of all adult ED visits. Among various diagnoses, chest pain was overwhelmingly dominant, aligning with clinical expectations of CVD symptoms.

Forecasting evaluations highlighted the superiority of machine learning methods (XGBoost , Random Forest, StackingEnsembleRegressor) over statistical time series models. While ARIMA performed comparably well, SARIMA underperformed due to its inability to generalize over out-of-sample data. The results are consistent with findings in previous research (e.g., Fan et al., 2022; Ho et al., 2019), reinforcing that nonlinear and ensemble models yield better performance in healthcare forecasting scenarios.

By proffering a technology approach to the challenges of CVDs, the study aligns with the goal of enhancing quality health delivery through digital innovation as seen in the healthcare transformation strategy of Saudi Arabia vision 2030 (Alasiri & Mohammed, 2022; Alghamdi & Alashban, 2024). Finally, by utilizing analytical and forecasting methodology in the research process, the study aligns with KSA National Biotechnology Strategy of integrating data analysis into healthcare for more precise precautionary diagnosis of CVDs and execution of remedies.

The practical implication of this research is that with reliable monthly forecasts of CVD patients, healthcare centers can effectively apportion their human and material capabilities to ensure that

they have optimum resources to handle monthly CVD patients. Additionally, private and public officials can make use of the insights from this study to focus on CVD medical challenges in the region. Finally, the methodology utilized in this research can be upscaled to different hospitals and regions to better manage Saudi Arabia's CVD cases.

CHAPTER 6

LIMITATIONS AND FUTURE WORK

During this study, we relied on using text data in the main diagnosis column which may introduce bias. The seasonal and pandemic-related anomalies were not modelled independently, and diagnosis texts may be prone to variability and underreporting. Univariate forecasting was done with the monthly number CVD patients being the only variable employed in the model development. Future work can incorporate deep learning methods like Long Short-Term Memory (LSTM) as alternative forecasting models and compare their performance with statistical and machine learning time series forecasting models. Also, forecasting models can be trained with multi variables to enhance model accuracy. And the integration of AI agents may improve CVD data analysis and forecast model performance. Additionally, this study used a sample size of 60 months which may limit the models' statistical power to identify important associations which could have impact on the generalizability of the findings from the models. For future research, incorporating larger datasets, potentially through multi-center collaborations or national health databases, would enhance the robustness of the predictive model and allow for more precise estimation of risk factors. The current models were trained with retrospective data, utilizing historical clinical presentations and symptoms. Although this data comes with important insights, it comes with limitations in predicting future events. Future studies should aim for prospective data collection to develop AI Algorithms that can more accurately predict CVD admission forecasts based on real-time clinical parameters, and more diverse populations to enhance the model's applicability across different demographics. Also, future work should Incorporate a broader range of clinical and lifestyle variables, such as diet, physical activity, and socioeconomic factors, to improve the model's predictive accuracy and undertake external validation of the model

to assess its performance in different clinical settings and populations within Saudi Arabia. This highlights the importance of conducting large-scale studies on CVD cases within Saudi Arabia. Also, accessing holistic and unstained medical records in Saudi Arabia comes with huge challenges because of tough data protection regulations and the absence of standardized, interoperable electronic health record (EHR) systems. The Personal Data Protection Law (PDPL) enforces stringent requirements on the gathering, processing and distribution of health data, limiting its availability for research purposes. Finally, the fragmentation of EHR systems across different healthcare providers hinders the aggregation of data necessary for large-scale studies.

REFERENCES

- Afilal, M., Yalaoui, F., Dugardin, F., Amodeo, L., Laplanche, D., & Blua, P. (2016). Forecasting the Emergency Department Patients Flow. *Journal of Medical Systems*, 40(7), 175.
<https://doi.org/10.1007/s10916-016-0527-0>
- Alasiri, A. A., & Mohammed, V. (2022). Healthcare Transformation in Saudi Arabia: An Overview Since the Launch of Vision 2030. *Health Services Insights*, 15, 11786329221121214. <https://doi.org/10.1177/11786329221121214>
- Alghamdi, S. A., & Alashban, Y. (2024). Medical science students' attitudes and perceptions of artificial intelligence in healthcare: A national study conducted in Saudi Arabia. *Journal of Radiation Research and Applied Sciences*, 17(1), 100815.
<https://doi.org/10.1016/j.jrras.2023.100815>
- Alhabib, K. F., Gamra, H., Almahmeed, W., Hammoudeh, A., Benkheddah, S., Al Jarallah, M., Al-Motarreb, A., Alquraishi, M., Sobhy, M., Yousif, M. G., Alkindi, F., Fellat, N., Amin, M. I., Ali, M., Al Saleh, A., Ullah, A., & Zannad, F. (2020). Acute myocardial infarction and acute heart failure in the Middle East and North Africa: Study design and pilot phase study results from the PEACE MENA registry. *PLOS ONE*, 15(7), e0236292.
<https://doi.org/10.1371/journal.pone.0236292>
- Amat Rodrigo, J., & Escobar Ortiz, J. (2023). Skforecast (Version v0.10.1dev) [Computer software]. Zenodo. <https://doi.org/10.5281/ZENODO.8382788>
- Asplin, B. R., Magid, D. J., Rhodes, K. V., Solberg, L. I., Lurie, N., & Camargo, C. A. (2003). A conceptual model of emergency department crowding. *Annals of Emergency Medicine*, 42(2), 173–180. <https://doi.org/10.1067/mem.2003.302>
- Atallah, R., & Al-Mousa, A. (2019). Heart Disease Detection Using Machine Learning Majority

- Voting Ensemble Method. *2019 2nd International Conference on New Trends in Computing Sciences (ICTCS)*, 1–6. <https://doi.org/10.1109/ICTCS.2019.8923053>
- Benjamin, E. J., Virani, S. S., Callaway, C. W., Chamberlain, A. M., Chang, A. R., Cheng, S., Chiuve, S. E., Cushman, M., Delling, F. N., Deo, R., De Ferranti, S. D., Ferguson, J. F., Fornage, M., Gillespie, C., Isasi, C. R., Jiménez, M. C., Jordan, L. C., Judd, S. E., Lackland, D., ... Muntner, P. (2018). Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association. *Circulation*, *137*(12). <https://doi.org/10.1161/CIR.0000000000000558>
- Boyle, J., Jessup, M., Crilly, J., Green, D., Lind, J., Wallis, M., Miller, P., & Fitzgerald, G. (2012). Predicting emergency department admissions. *Emergency Medicine Journal*, *29*(5), 358–365. <https://doi.org/10.1136/emj.2010.103531>
- Bullard, M. J., Villa-Roel, C., Guo, X., Holroyd, B. R., Innes, G., Schull, M. J., Vandermeer, B., Ospina, M., & Rowe, B. H. (2012). The role of a rapid assessment zone/pod on reducing overcrowding in emergency departments: A systematic review. *Emergency Medicine Journal*, *29*(5), 372–378. <https://doi.org/10.1136/emj.2010.103598>
- Chase, V. J., Cohn, A. E. M., Peterson, T. A., & Lavieri, M. S. (2012). Predicting Emergency Department Volume Using Forecasting Methods to Create a “Surge Response” for Noncrisis Events. *Academic Emergency Medicine*, *19*(5), 569–576. <https://doi.org/10.1111/j.1553-2712.2012.01359.x>
- Cheng, Q., Argon, N. T., Evans, C. S., Liu, Y., Platts-Mills, T. F., & Ziya, S. (2021). Forecasting emergency department hourly occupancy using time series analysis. *The American Journal of Emergency Medicine*, *48*, 177–182. <https://doi.org/10.1016/j.ajem.2021.04.075>

- Cihan, P. (2024). Comparative Performance Analysis of Deep Learning, Classical, and Hybrid Time Series Models in Ecological Footprint Forecasting. *Applied Sciences*, 14(4), 1479. <https://doi.org/10.3390/app14041479>
- Darraj, A., Hudays, A., Hazazi, A., Hobani, A., & Alghamdi, A. (2023). The Association between Emergency Department Overcrowding and Delay in Treatment: A Systematic Review. *Healthcare*, 11(3), 385. <https://doi.org/10.3390/healthcare11030385>
- Deneshkumar, V., Jithu, G., & Jebitha, R. (2023). Ensemble Machine Learning Algorithms for Predicting Cardiovascular Disease. In S. Roy, D. Sinwar, N. Dey, T. Perumal, & J. M. R. S. Tavares (Eds.), *Innovations in Computational Intelligence and Computer Vision* (Vol. 680, pp. 427–436). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-2602-2_32
- Deng, Y.-X., Wang, J.-Y., Ko, C.-H., Huang, C.-H., Tsai, C.-L., & Fu, L.-C. (2024). Deep learning-based Emergency Department In-hospital Cardiac Arrest Score (Deep EDICAS) for early prediction of cardiac arrest and cardiopulmonary resuscitation in the emergency department. *BioData Mining*, 17(1), 52. <https://doi.org/10.1186/s13040-024-00407-8>
- Derose, S. F., Gabayan, G. Z., Chiu, V. Y., Yiu, S. C., & Sun, B. C. (2014). Emergency Department Crowding Predicts Admission Length-of-Stay But Not Mortality in a Large Health System. *Medical Care*, 52(7), 602–611. <https://doi.org/10.1097/MLR.0000000000000141>
- Edmondson, D., Shimbo, D., Ye, S., Wyer, P., & Davidson, K. W. (2013). The Association of Emergency Department Crowding During Treatment for Acute Coronary Syndrome With Subsequent Posttraumatic Stress Disorder Symptoms. *JAMA Internal Medicine*, 173(6), 472. <https://doi.org/10.1001/jamainternmed.2013.2536>

- Eschenbrenner, B., & Brenden, R. (2022). Deriving Value from Big Data Analytics in Healthcare: A Value-focused Thinking Approach. *AIS Transactions on Human-Computer Interaction*, 14(3), 289–313. <https://doi.org/10.17705/1thci.00170>
- Fan, B., Peng, J., Guo, H., Gu, H., Xu, K., & Wu, T. (2022). Accurate Forecasting of Emergency Department Arrivals With Internet Search Index and Machine Learning Models: Model Development and Performance Evaluation. *JMIR Medical Informatics*, 10(7), e34504. <https://doi.org/10.2196/34504>
- Fawagreh, K., & Gaber, M. M. (2020). Resource-efficient fast prediction in healthcare data analytics: A pruned Random Forest regression approach. *Computing*, 102(5), 1187–1198. <https://doi.org/10.1007/s00607-019-00785-6>
- Figtree, G. A., Doyle, K., Nedkoff, L., Cadilhac, D. A., Kovacic, J., & For the Cardiovascular National Health Leaders Research Forum Writing Group. (2024). The national Cardiovascular Health Leadership Research Forum: A new data-driven model placing research at the centre of improving patient outcomes. *Medical Journal of Australia*, 221(9), 452–456. <https://doi.org/10.5694/mja2.52482>
- Fralick, M., Murray, J., & Mamdani, M. (2021). Predicting emergency department volumes: A multicenter prospective study. *The American Journal of Emergency Medicine*, 46, 695–697. <https://doi.org/10.1016/j.ajem.2020.10.047>
- Ghasemieh, A., Lloyed, A., Bahrami, P., Vajar, P., & Kashef, R. (2023). A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients. *Decision Analytics Journal*, 7, 100242. <https://doi.org/10.1016/j.dajour.2023.100242>
- Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. M. J. M., Ignatious, E., Shultana, S.,

- Beeravolu, A. R., & De Boer, F. (2021). Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques. *IEEE Access*, 9, 19304–19326.
<https://doi.org/10.1109/ACCESS.2021.3053759>
- Gulf Heart Association. (2022). Regional insights into cardiovascular disease and risk factors in the Gulf Cooperation Council countries. *GHA Journal of Cardiology*, 45(3), 112–118.
- Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms. *Mobile Information Systems*, 2018, 1–21. <https://doi.org/10.1155/2018/3860146>
- Himmelfarb, D., Cheryl R., Beckie, T. M., Allen, L. A., Commodore-Mensah, Y., Davidson, P. M., Lin, G., Lutz, B., & Spatz, E. S. (2023). Shared Decision-Making and Cardiovascular Health: A Scientific Statement From the American Heart Association. *Circulation*, 148(11), 912–931. <https://doi.org/10.1161/CIR.0000000000001162>
- Ho, A. F. W., To, B. Z. Y. S., Koh, J. M., & Cheong, K. H. (2019). Forecasting Hospital Emergency Department Patient Volume Using Internet Search Data. *IEEE Access*, 7, 93387–93395. <https://doi.org/10.1109/ACCESS.2019.2928122>
- Holmberg, M. J., Wiberg, S., Ross, C. E., Kleinman, M., Hoeyer-Nielsen, A. K., Donnino, M. W., Andersen, L. W., & For the American Heart Association’s Get With The Guidelines-Resuscitation Investigators. (2019). Trends in Survival After Pediatric In-Hospital Cardiac Arrest in the United States. *Circulation*, 140(17), 1398–1408.
<https://doi.org/10.1161/CIRCULATIONAHA.119.041667>
- Hong, K. J., Shin, S. D., Song, K. J., Cha, W. C., & Cho, J. S. (2013). Association between ED crowding and delay in resuscitation effort. *The American Journal of Emergency*

- Medicine*, 31(3), 509–515. <https://doi.org/10.1016/j.ajem.2012.09.029>
- Hoot, N. R., & Aronsky, D. (2008). Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Annals of Emergency Medicine*, 52(2), 126-136.e1. <https://doi.org/10.1016/j.annemergmed.2008.03.014>
- Hu, Z., Qiu, H., Su, Z., Shen, M., & Chen, Z. (2020). A Stacking Ensemble Model to Predict Daily Number of Hospital Admissions for Cardiovascular Diseases. *IEEE Access*, 8, 138719–138729. <https://doi.org/10.1109/ACCESS.2020.3012143>
- Jiang, S., Liu, Q., & Ding, B. (2023). A systematic review of the modelling of patient arrivals in emergency departments. *Quantitative Imaging in Medicine and Surgery*, 13(3), 1957–1971. <https://doi.org/10.21037/qims-22-268>
- Jo, S., Jin, Y. H., Lee, J. B., Jeong, T., Yoon, J., & Park, B. (2014). Emergency Department Occupancy Ratio is Associated With Increased Early Mortality. *The Journal of Emergency Medicine*, 46(2), 241–249. <https://doi.org/10.1016/j.jemermed.2013.05.026>
- Kam, H. J., Sung, J. O., & Park, R. W. (2010). Prediction of Daily Patient Numbers for a Regional Emergency Medical Center using Time Series Analysis. *Healthcare Informatics Research*, 16(3), 158. <https://doi.org/10.4258/hir.2010.16.3.158>
- Kausar, N., Palaniappan, S., Samir, B. B., Abdullah, A., & Dey, N. (2016). Systematic Analysis of Applied Data Mining Based Optimization Algorithms in Clinical Attribute Extraction and Classification for Diagnosis of Cardiac Patients. In A.-E. Hassanien, C. Grosan, & M. Fahmy Tolba (Eds.), *Applications of Intelligent Optimization in Biology and Medicine* (Vol. 96, pp. 217–231). Springer International Publishing. https://doi.org/10.1007/978-3-319-21212-8_9
- Kontsevaya, A., Kalinina, A., & Oganov, R. (2013). Economic Burden of Cardiovascular

- Diseases in the Russian Federation. *Value in Health Regional Issues*, 2(2), 199–204.
<https://doi.org/10.1016/j.vhri.2013.06.010>
- Lakshmanarao, A., Srisaila, A., & Kiran, T. S. R. (2021). Heart Disease Prediction using Feature Selection and Ensemble Learning Techniques. *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 994–998. <https://doi.org/10.1109/ICICV50876.2021.9388482>
- Mai, Q., Aboagye-Sarfo, P., Sanfilippo, F. M., Preen, D. B., & Fatovich, D. M. (2015). Predicting the number of emergency department presentations in Western Australia: A population-based time series analysis. *Emergency Medicine Australasia*, 27(1), 16–21.
<https://doi.org/10.1111/1742-6723.12344>
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74.
<https://doi.org/10.1016/j.ijforecast.2019.04.014>
- Malik, M. M., Abdallah, S., & Ala'raj, M. (2018). Data mining and predictive analytics applications for the delivery of healthcare services: A systematic literature review. *Annals of Operations Research*, 270(1–2), 287–312. <https://doi.org/10.1007/s10479-016-2393-z>
- Marcilio, I., Hajat, S., & Gouveia, N. (2013). Forecasting Daily Emergency Department Visits Using Calendar Variables and Ambient Temperature Readings. *Academic Emergency Medicine*, 20(8), 769–777. <https://doi.org/10.1111/acem.12182>
- McCarthy, M. L., Zeger, S. L., Ding, R., Aronsky, D., Hoot, N. R., & Kelen, G. D. (2008). The Challenge of Predicting Demand for Emergency Department Services. *Academic Emergency Medicine*, 15(4), 337–346. <https://doi.org/10.1111/j.1553-2712.2008.00083.x>
- McCusker, J., Vadeboncoeur, A., Lévesque, J., Ciampi, A., & Belzile, E. (2014). Increases in

- Emergency Department Occupancy Are Associated With Adverse 30-day Outcomes. *Academic Emergency Medicine*, 21(10), 1092–1100. <https://doi.org/10.1111/acem.12480>
- Ministry of Health KSA, United Nations Development Programme, World Health Organization, & Secretariat of the UN Inter-Agency Task Force on NCDs. (2023). THE CASE FOR INVESTMENT IN PREVENTION AND CONTROL OF NON-COMMUNICABLE DISEASES IN THE KINGDOM OF SAUDI ARABIA THE CASE FOR INVESTMENT IN PREVENTION AND CONTROL OF NON-COMMUNICABLE DISEASES IN SAUDI ARABIA.
- National Health Center, & Saudi Health Council. (2022). Cardiovascular Disease: A Public Health Priority.
- Rachas, A., Gastaldi-Ménager, C., Denis, P., Barthélémy, P., Constantinou, P., Drouin, J., Lastier, D., Lesuffleur, T., Mette, C., Nicolas, M., Pestel, L., Rivière, S., Tajahmady, A., Gissot, C., & Fagot-Campagna, A. (2022). The Economic Burden of Disease in France From the National Health Insurance Perspective: The Healthcare Expenditures and Conditions Mapping Used to Prepare the French Social Security Funding Act and the Public Health Act. *Medical Care*, 60(9), 655–664. <https://doi.org/10.1097/MLR.0000000000001745>
- Rani, P., Kumar, R., Ahmed, N. M. O. S., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments*, 7(3), 263–275. <https://doi.org/10.1007/s40860-021-00133-6>
- Roberts, M. V., Baird, W., Kerr, P., & O'Reilly, S. (2010). Can an Emergency Department-based Clinical Decision Unit successfully utilize alternatives to emergency hospitalization?: *European Journal of Emergency Medicine*, 17(2), 89–96. <https://doi.org/10.1097/MEJ.0b013e32832f05bf>

- Rose, R., & Skariah, E. N. (2013). Active power management of hybrid fuel cell, photovoltaic unit, and supercapacitor power conversion system in a microgrid. *2013 International Conference on Renewable Energy and Sustainable Energy (ICRESE)*, 200–206.
<https://doi.org/10.1109/ICRESE.2013.6927815>
- Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., ... Fuster, V. (2020). Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019. *Journal of the American College of Cardiology*, 76(25), 2982–3021.
<https://doi.org/10.1016/j.jacc.2020.11.010>
- Royal Commission for Jubail and Yanbu. (2019). *Emergency Cases in Jubail Health Services Program in 2019* [Dataset]. <https://open.data.gov.sa/en/datasets/view/d294debb-e98d-48bd-a15d-88b9cd127307>
- Royal Commission for Jubail and Yanbu. (2020). *Emergency Cases in Jubail Health Services Program in 2020* [Dataset]. <https://open.data.gov.sa/en/datasets/view/648bf9a5-daf9-49f4-b857-20b426cd9102>
- Royal Commission for Jubail and Yanbu. (2021). *Emergency Cases in Jubail Health Services Program in 2021* [Dataset]. <https://open.data.gov.sa/en/datasets/view/968a1ee6-79d0-4a31-97e8-c550ef686f80>
- Royal Commission for Jubail and Yanbu. (2022). *Emergency Cases in Jubail Health Services Program in 2022* [Dataset]. <https://open.data.gov.sa/en/datasets/view/50667990-6231-45c8-ba9a-9342d235d8f2>
- Royal Commission for Jubail and Yanbu. (2023). *Emergency Cases in Jubail Health Services*

- Program in 2023* [Dataset]. <https://open.data.gov.sa/en/datasets/view/3377a024-c833-4c2c-bc10-33ee2d417387>
- Schneider, S. M., Gallery, M. E., Schafermeyer, R., & Zwemer, F. L. (2003). Emergency department crowding: A point in time. *Annals of Emergency Medicine*, 42(2), 167–172. <https://doi.org/10.1067/mem.2003.258>
- Shah, I. A., & Mishra, S. (2024). Artificial intelligence in advancing occupational health and safety: An encapsulation of developments. *Journal of Occupational Health*, 66(1), uiad017. <https://doi.org/10.1093/joccuh/uiad017>
- Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y., & Lendasse, A. (2007). Methodology for long-term prediction of time series. *Neurocomputing*, 70(16–18), 2861–2869. <https://doi.org/10.1016/j.neucom.2006.06.015>
- Sudarshan, V. K., Brabrand, M., Range, T. M., & Wiil, U. K. (2021). Performance evaluation of Emergency Department patient arrivals forecasting models by including meteorological and calendar information: A comparative study. *Computers in Biology and Medicine*, 135, 104541. <https://doi.org/10.1016/j.combiomed.2021.104541>
- Sun, B. C., Hsia, R. Y., Weiss, R. E., Zingmond, D., Liang, L.-J., Han, W., McCreath, H., & Asch, S. M. (2013). Effect of Emergency Department Crowding on Outcomes of Admitted Patients. *Annals of Emergency Medicine*, 61(6), 605-611.e6. <https://doi.org/10.1016/j.annemergmed.2012.10.026>
- Tash, A. A., & Al-Bawardy, R. F. (2023). Cardiovascular Disease in Saudi Arabia: Facts and the way forward. *Journal of the Saudi Heart Association*, 35(2), 148–162. <https://doi.org/10.37616/2212-5043.1336>
- Tekwani, K., Kerem, Y., Mistry, C., Sayger, B., & Kulstad, E. (2013). Emergency Department

- Crowding is Associated with Reduced Satisfaction Scores in Patients Discharged from the Emergency Department. *Western Journal of Emergency Medicine*, 14(1), 11–15.
<https://doi.org/10.5811/westjem.2011.11.11456>
- Tello, M., Reich, E. S., Puckey, J., Maff, R., Garcia-Arce, A., Bhattacharya, B. S., & Feijoo, F. (2022). Machine learning based forecast for the prediction of inpatient bed demand. *BMC Medical Informatics and Decision Making*, 22(1), 55. <https://doi.org/10.1186/s12911-022-01787-9>
- Tran, V. T., Yang, B.-S., & Tan, A. C. C. (2009). Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems. *Expert Systems with Applications*, 36(5), 9378–9387. <https://doi.org/10.1016/j.eswa.2009.01.007>
- Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Anderson, C. A. M., Arora, P., Avery, C. L., Baker-Smith, C. M., Beaton, A. Z., Boehme, A. K., Buxton, A. E., Commodore-Mensah, Y., Elkind, M. S. V., Evenson, K. R., Eze-Nliam, C., Fugar, S., Generoso, G., Heard, D. G., Hiremath, S., Ho, J. E., ... on behalf of the American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. (2023). Heart Disease and Stroke Statistics—2023 Update: A Report From the American Heart Association. *Circulation*, 147(8). <https://doi.org/10.1161/CIR.0000000000001123>
- Tyrovolas, S., El Bcheraoui, C., Alghnam, S. A., Alhabib, K. F., Almadi, M. A. H., Al-Raddadi, R. M., Bedi, N., El Tantawi, M., Krish, V. S., Memish, Z. A., Mohammad, Y., Molassiotis, A., Panagiotakos, D., Salam, N., Sobaih, B. H., & Mokdad, A. H. (2020). The burden of disease in Saudi Arabia 1990–2017: Results from the Global Burden of Disease Study 2017. *The Lancet Planetary Health*, 4(5), e195–e208.
[https://doi.org/10.1016/S2542-5196\(20\)30075-9](https://doi.org/10.1016/S2542-5196(20)30075-9)

- Vollmer, M. A. C., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., Klaber, R., Cooke, G., Flaxman, S., & Bhatt, S. (2021). A unified machine learning approach to time series forecasting applied to demand at emergency departments. *BMC Emergency Medicine*, 21(1), 9. <https://doi.org/10.1186/s12873-020-00395-y>
- Whitt, W., & Zhang, X. (2019). Forecasting arrivals and occupancy levels in an emergency department. *Operations Research for Health Care*, 21, 1–18. <https://doi.org/10.1016/j.orhc.2019.01.002>
- WHO. (2021, June 11). *Cardiovascular diseases (CVDs)*. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- Yousefi, M., Yousefi, M., Fathi, M., & Fogliatto, F. S. (2019). Patient visit forecasting in an emergency department using a deep neural network approach. *Kybernetes*, 49(9), 2335–2348. <https://doi.org/10.1108/K-10-2018-0520>
- Zghebi, S. S., Mamas, M. A., Ashcroft, D. M., Rutter, M. K., VanMarwijk, H., Salisbury, C., Mallen, C. D., Chew-Graham, C. A., Qureshi, N., Weng, S. F., Holt, T., Buchan, I., Peek, N., Giles, S., Reeves, D., & Kontopantelis, E. (2021). Assessing the severity of cardiovascular disease in 213 088 patients with coronary heart disease: A retrospective cohort study. *Open Heart*, 8(1), e001498. <https://doi.org/10.1136/openhrt-2020-001498>
- Zhang, Y., Zhang, J., Tao, M., Shu, J., & Zhu, D. (2022). Forecasting patient arrivals at emergency department using calendar and meteorological information. *Applied Intelligence*, 52(10), 11232–11243. <https://doi.org/10.1007/s10489-021-03085-9>
- Zhou, L., Zhao, P., Wu, D., Cheng, C., & Huang, H. (2018). Time series model for forecasting

the number of new admission inpatients. *BMC Medical Informatics and Decision Making*, 18(1), 39. <https://doi.org/10.1186/s12911-018-0616-8>

Zhu, T., Luo, L., Zhang, X., Shi, Y., & Shen, W. (2017). Time-Series Approaches for Forecasting the Number of Hospital Daily Discharged Inpatients. *IEEE Journal of Biomedical and Health Informatics*, 21(2), 515–526.
<https://doi.org/10.1109/JBHI.2015.2511820>

APPENDIXES

Appendix 1: CVD Keywords

- chest
- stroke
- angina
- ischemia
- fibrillation
- bradycardia
- tachycardia
- arteriosclerosis
- thrombosis
- hypotension
- cardiomyopathy
- pericarditis
- endocarditis
- arrhythmia
- myocardial
- cardiac
- heart
- hypertension
- cardiomyopathy
- coronary
- aortic
- ventricular
- atrial
- mitral
- tricuspid

- pulmonary

Appendix 2: Text Normalization

- Any Main Diagnosis row value containing keyword “chest pain” in them was replaced with “Chest pain”
- Main Diagnosis row value containing the keyword “hypertension” in them was replaced with “Hypertension”
- Main Diagnosis row value containing the keyword “heart failure” in them was replaced with “Heart failure”
- Main Diagnosis row value containing the keyword “stroke” in them was replaced with “Stroke”

VITAE

Name OBASI, CHIGOZIE FRANCIS

Nationality NIGERIAN

Date of Birth January 8, 1993

Email obasichigoziefrancis@gmail.com

Address 13, Alhaji Kaleijaye Street off Bajualaye Road, Somolu, P.O Box 102216
Lagos, Nigeria.

EDUCATION

- King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia June 2025

Master of Science in Business Studies, Management Information System Track (CGPA 3.9/4.0).

Thesis: Forecasting Cardiovascular Diseases of adult patients in health care systems emergency department: A data mining approach.

- Federal University of Technology, Owerri (FUTO), Nigeria.
October 2014

Bachelor of Agricultural Technology, Crop Science Technology (CGPA: 4.11/5.00).

WORKING PAPERS

Obasi, C.F, Alahmad, R. & Althiyabi, F. Forecasting Cardiovascular Diseases of Adult Patients in Healthcare Systems Emergency Department: A Data Mining Approach (will be submitted soon)

Abbasi, G.A, Obasi, C.F, Albahhar, A. A, Ebrima, J & Abdullah, M. Assessing Artificial Intelligence Adoption in Healthcare workplace in Saudi Arabia using Behavioral Reasoning Theory (will be submitted soon)

Alahmad, R, Obasi, C.F, & Javed, I. Impact of Artificial Intelligence Inclusion on User Satisfaction and Interactional Justice (will be submitted soon)

AWARDS, GRANT AND RECOGNITIONS

- Co-PI with Ghazanfar Ali Abbasi on the project titled “Application of Artificial Intelligence tools in hospitals across the Kingdom of Saudi Arabia.” This project is funded by KFUPM (\$13000).
- **Shortlisted for the best master thesis at KFUPM forum in 2024.**

- Recipient of the best graduating student award in Crop Science and Technology from the Federal University of Technology, Owerri, Imo state, Nigeria in 2014.

RESEARCH EXPERIENCE

Graduate Student

King Fahd University of Petroleum and Minerals, Saudi Arabia
June 2025

August 2023 –

- Conducted extensive field research on the impact of enablers & disablers on Students' Switching Intention from generative Chatbots using the push and pull theory.
- Co-PI to fully funded research on the application of AI tools in hospitals within the Kingdom of Saudi Arabia.

Undergraduate Student

Federal University of Technology, Owerri, Nigeria.
2009 – October 2014

November

- Executed field research on the effect of Cassava varieties and Neem leaves on termite incidence and severity on Cassava (*manihot esculenta*) in Owerri, Imo state, Nigeria. The experiment spanned a total of 16 weeks and was conducted using a 3 X 3 Factorial in a randomized complete block design (RCBD) with three replications. The treatment combination of Variety NR80-83 and Neem Leaves 20.0tonnes/ha significantly reduced termite incidence especially from 12WAP where termite incidence was suppressed up to zero level.

RESEARCH CONFERENCE

- International Conference on Smart Mobility and Logistics Ecosystems (SMiLE)
September 2024
- Saudi Conference on Information Systems (SaudiCIS).
November 2024

King Fahd University of Petroleum and Minerals, Saudi Arabia

- Presented my team's research paper on the impact of enablers & disablers on Students' Switching Intention from generative artificial intelligent chatbots using the push and pull theory.

UNIVERSITY SERVICE AND ENGAGEMENT

Graduate Student

King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia
2024 – June 2025

September

- Nominated student representative, responsible use of artificial intelligence tools taskforce at KFUPM

- 2024/2025 academic year strategic planning and steering committee member, KFUPM Business School.
- Nominated student participant, Green Campus Initiative at KFUPM.

PROFESSIONAL WORK EXPERIENCE

Payment Administrator (Finance), The Heineken Company June 2021 – September 2023

- Explored the lean six sigma project management framework to resolve finance control issues relating to missing goods in transit, wrong purchase orders creation, goods receipts valued at an estimated sum of \$100,000.
- Employed agile project management methodology in identifying and ensuring that a system payment control gap is resolved, thereby avoiding duplicating vendors payment, saving a total \$8,035,714 every week.
- Analyzed and prepared management reports such as goods in transit, ledger reclassification, and balance sheet account review, and trade debt using Microsoft excel and ensured data correctness, completeness during month end, quarter end and the year-end close process.

Purchase to Pay Administrator (Finance), The Heineken Company November 2019 – May 2021

- Improved the position of intercompany vendor reconciliation account by reconciling and clearing all aged items over 720 days and reducing the number of aged items over 360 days by thirty percent through in-depth analyses of supplier statements, collaboration with stakeholders, following up on missing invoices and any unclear transactions.
- Improved the reconciliation position of customs duty account as for the first in over 4 years, all customs duty items were cleared within 16 months on the Job by processing custom duties invoices on time, clearing of all imports transactions valued at over \$4,000,000.
- Improved vendors invoice posting issues by 12% by complying with SAP 3-way payable matching rules and process automation.

TEACHING EXPERIENCE

Teaching Assistant, College of Education, Warri, Delta State. May 2015 – April 2016

- Developed and implemented agricultural teaching techniques for 60 students.
- Continuously reviewed class achievements in test, exams and in-class questions, and adjusting training activities for improved assimilation of biology and agricultural materials.
- Organized assessment activities such as quizzes, tests and examination in line with the school's curriculum requirement.

- Created agricultural and biological training content that focused on efficient management of cassava, maize and cowpea for improved crop health

COMMUNITY AND VOLUNTERING SERVICE

- **Education Project Co-lead, Catholic Youth Organization of Nigeria, St. Matthew's, Lagos, Nigeria:** Trained about 50 youths on career development, delivered food and material resources to one orphanage home with about 30 children between ages 3 to 11 and over 200 homeless individuals in the community.
- **Yellow Cassava Project Assistant, Forward Africa (NGO), Owerri, Imo state, Nigeria:** A Bill and Melinda Gates Funded Project. Distributed about two tons of Vitamin A enriched cassava stems to over 500 farmers in Imo state, Nigeria. Trained over 100 farmers on the best cassava cultivation technique and on the application of the right quantity of pesticides and fertilizers.

TECHNICAL SKILLS

- Programming Languages: Python, SQL, R.
- Software: PowerBI , SPSS, SMARTPLS, Microsoft Excel .