IMPROVING CARDIAC CARE DELIVERY USING PREDICTIVE AND PRESCRIPTIVE ANALYTICS

A Dissertation

Presented to

The Faculty of Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Industrial Engineering

By

Haya Salah

Dr. Sharan Srinivas, Dissertation Supervisor

May 2022

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

IMPROVING CARDIAC CARE DELIVERY USING PREDICTIVE AND PRESCRIPTIVE ANALYTICS

Presented by Haya Salah, a candidate for the degree of Doctor of Philosophy in Industrial Engineering, and hereby certify that in their opinion it is worthy of acceptance.

_____

Professor Sharan Srinivas

_____

Professor Luis Occena

_____

Professor Kangwon Seo

_____

Professor Lincoln Sheets

_____

Professor Sangdo Choi

Acknowledgment

I would like to thank my faculty adviser, Dr. Sharan Srinivas for his support, guidance and expertise. I would like to thank my husband, parents, brothers and sisters for their encouragement, love and support through my professional development.

Table of Contents

List of Figures

List of Tables

Abstract

Endured by the growing prevalence of cardiovascular diseases (CVDs), the demand for cardiac care services has increased. On the other hand, the supply of cardiologists is expected to be insufficient to meet this growing demand. Considering this imbalance between demand and capacity, cardiology clinics strive to improve their services. Thus, this dissertation proposes two approaches for improving cardiac care services through the application of predictive and prescriptive analytics. The first approach develops an efficient appointment system (AS) that allocates patients' demand during the clinic session effectively to improve resource utilization and patient satisfaction. The proposed AS also addresses the problem of clinical uncertainties, such as patient no-shows and service-time variability, which adversely impact AS efficiency by developing a predict-then-schedule framework. In the predict step, patient-specific no-show risk and service duration are estimated using machine learning (ML) models. The schedule step determines the appointment time and interval for each patient using a sequential AS that leverages the ML predictions. In addition, four new ML-enabled sequencing rules are proposed. The proposed approach and sequencing rules are validated using real clinical data. Besides, the effectiveness of integrating ML-based uncertainty predictions into the AS design is also evaluated for 32 different clinic environments. Results indicate that an AS design adopting the predict-then-schedule approach always dominates the conventional system and could improve the efficiency by 60%. The new sequencing rules can improve the AS performance by up to 40% when compared to the existing policies. Finally, several managerial insights on sequencing and overbooking are also provided. On the other hand,

the second approach develops an ML-based model to predict the long-term CVD risk that can aid in the early detection of CVD. Unlike the existing tools for CVD risk assessment which are only applicable to adults and use cross-sectional data. This research provides the first long-term ML-based CVD risk prediction model among adolescents based on a longitudinal dataset. Our results indicated the capability of ML models to accurately predict the long-term risk of CVD among adolescents. In addition, the most significant factors for predicting CVD risk among adolescents are identified. Furthermore, the proposed model can be used to identify individuals who are at high risk of developing CVD early in life and provide them with the necessary guidance and preventive treatment, which improves the quality of life, lower healthcare costs, and reduces the demand for cardiac care services later.

Chapter 1. Introduction

Cardiovascular diseases (CVDs) affect more than half of the U.S. adult population [1], accounting for an average of 40.5 million visits between 2014 and 2016 [2–4]. Almost 40% of the U.S. population is expected to have CVD by 2030, with direct medical costs approaching $1 trillion [5]. Despite these statistics, mortality related to CVD declined by 38% between 2003 and 2013. This could be attributed to the risk factor modification, advanced treatments, and technologies [6]. However, decreasing mortality is not the same as decreasing CVD prevalence. By 2030, the occurrence of CVD is projected to increase by 4.5% [5]. While some progress has been made, cardiovascular disease remains the leading cause of death in the United States, and the demand for cardiac services has increased significantly. This increasing trend in CVD has been attributed to insufficient awareness of CVD, limited planning for early prevention interventions, and the increasing dominance of CVD risk factors, such as obesity, diabetes, blood pressure, and elevated blood cholesterol [7]. As a result of this growing demand , cardiac care services are struggling to provide timely access, patients are experiencing long wait times, and doctors are overburdened with clinical responsibilities [8]. Thus, this study proposes two methods to improve the delivery of cardiology clinics: the first approach focuses on improving resource utilization and patient experience through the design of an appointment system (AS). An AS can smooth patient demand by matching appointment requests to available capacity, thereby effectively utilizing resources such as physicians, nurses, and medical equipment. From the patient's perspective, the AS has the potential to reduce their waiting time at the clinic through proper planning and allocation of resources [9]. On the other

hand, the second approach tries to reduce the demand for cardiology clinics by developing a clinical decision support tool that can aid in the early detection of CVD risk. Since around 20-40% of heart attacks occur in individuals who have never been diagnosed with CVD [10], such a tool can identify individuals who are at high risk of developing CVD later in life and provide them with the necessary guidance and preventive treatment. Which in turn reduce the demand for cardiac care services later.

1.1 Designing an effective AS for cardiology clinic

Cardiology clinics should have an efficient appointment system (AS) to schedule patients and deliver a quality of care within their constrained resources. An Appointment system is designed to allocate patient demand during the clinic session such that the physicians' time is effectively utilized, and patients are satisfied. Therefore, designing an AS for a cardiology clinic can help improve the clinics delivery and resource utilization.

1.1.1 Appointment system (AS) in healthcare

Several studies have been carried out to increase the efficiency of AS in healthcare settings since the pioneering work of Baily's in 1952 [11]. The ultimate goal is to increase the productivity of resources and synchronously provide a high quality of care [12]. From a doctor's perspective, it is imperative to reduce idle time and clinic overtime. On the other hand, long waiting times can't be endured by patients. Thus, a well-designed appointment

system will try to balance patients' and physicians' preferences, such that patients are satisfied (less waiting time), and physicians are effectively utilized (less doctor's idle time and overtime). Outpatient clinics use AS to distribute their workload over time. For a particular day, the clinic session is divided into smaller appointment durations called (slots). Each patient is assigned to a slot/slots depending on three decisions adapted by the clinic AS, and these include (i) allocation decisions, (ii) sequencing decisions, (iii) and no-show adjustments.

1.1.1.1 Allocation decisions

Allocation decisions assign patients' appointments depending on two characteristics: block size and block interval. The block size determines the number of patients be scheduled in each slot; patients can be scheduled individually, in groups of constant size, or variable block size. In addition to block size, block intervals are used to identify the appointment duration (number of blocks) assigned for each patient. These also can be constant or variable. Usually, appointment intervals are set equal to some value, such as the mean consultation times for patients. Some of the most common allocation decisions are:

i. *Single block*: assign all patients to a single block at the start of the clinic session. For example, all patients for the morning session are assigned the same appointment start time, and they will be served on a first call, first-appointment basis.

ii. *Individual block/Fixed interval*: allocates each patient to a specific appointment slot with constant appointment duration (i.e., the mean of consultations durations)

iii. *Multiple block/Fixed interval*: schedule *m* number of patients to each slot with constant appointment intervals.

iv. *Variable bock/Fixed interval*: schedules patients in different block sizes with appointment durations kept constant.

v. *Individual block/Variable interval*: schedules patients individually by varying the appointment intervals.

1.1.1.2 Sequencing decisions

Sequencing decisions are used to define the order in which patients are scheduled to the appointment slot. When no distinct group is used to classify patients, they are scheduled on the first call, first-appointment basis. However, if patients are classified based on certain characteristics, then they can be sequenced based on these characteristics. Several characteristics were used to categorize patients in the literature, such as return/new patients, high/low variation in service time, and the type of procedure. [13,14].

1.1.1.3 No-show adjustments decisions

Even though many administrative policies effectively mitigate the effect of no-shows, such as sending automated text messages, reminders, e-mails, and phone calls, it is not completely possible to eliminate no-shows [15]. Therefore, the AS should be adjusted to reduce their harmful effect. Previous work has investigated two approaches to adjust for no-shows. White et al. [16] consider the overbooking policy where extra patients are scheduled to accommodate the average number of no-shows. The results showed that adding extra patients can improve the system performance. Fetter et al. [17] suggested adjusting the appointment intervals according to the expected number of no-shows. On the other hand, Cox et al. [18] introduced an approach that sets the appointment intervals equal to the revised consultation length based on the expected probability of no-show. Thus, to compensate for no-shows, one can choose either overbooking or adjusting the appointment intervals.

1.1.2 Challenges that undermine the effectiveness of AS in specialty clinics

Outpatient specialty clinics, such as cardiac and cancer care, use an appointment system to manage access to their services. However, many uncertainties affect the efficiency of appointment systems, such as demand uncertainty, variation in consultation length, and patient no-shows.

i. *Patient No-shows and late cancellations*: An important cause of uncertainty in AS is patient no-shows, when a patient fails to show up for an appointment without a prior notice. This affects the clinic adversely since the allocated consultation length can be totally wasted. Notably, a clinic loses an average of $200 in revenue for each unused time slot, and the U.S. healthcare system is projected to lose $150 billion annually due to no-shows [19] . To accommodate for no-shows, clinics tend to schedule more patients than their capacity. The majority of clinics assume patients are homogenous and have the same probability of no-shows. Therefore, they double-book by an even percentage equal to the average probability of no-show in the clinic.

ii. *Variation in consultation length*: High variation in consultation length in specialty clinics affects the scheduler's ability to allocate provider time appropriately, thereby posing a significant challenge for effective scheduling [20]. Underestimating consultation duration can lead to physician burnout, misdiagnosis, and patient dissatisfaction [8,21]. On the other hand, consultation durations that are longer than required lead to idle resources and fewer patient visits, which, in turn, affects the clinic's revenue [22]. The clinic's settings also consider as another challenge for effective allocation of service times and scheduling. In primary care, provider service times tend to be divided into equal-length slots as most of the patients require service times that can be completed within a fixed time interval [20,23]. On the other hand, allocating service times in specialty clinics are more complicated as service times vary depending on patient diagnosis and other

characteristics. Thus, service times can't be divided into fixed intervals [20]. Therefore, a reasonable estimation of a service provider's time with a patient can lead to better scheduling, improving timely access to care, resource utilization, quality of care, and patient satisfaction.

iii. *Sequential appointment requests*: sequential arrivals of appointment requests pose a significant challenge in sequencing the patients. Prior works have demonstrated sequencing to have a significant impact on patient waiting time and server idle time [24]. To handle the sequential call-in requests, a first-call, first-appointment (FCFA) policy is typically adopted in the case of a homogeneous patient population; otherwise, patient classification is commonly used for sequencing [25]. For instance, the patients visiting the clinic for the first time are scheduled in the morning, while patients returning for a follow-up visit are scheduled in the afternoon session.

1.2 Developing ML-based model for early detection of CVD risk

While improving the AS design by taking into account patient-specific characteristics can enable better management of growing cases of CVD and limited physician capacity, it does not provide an opportunity for reducing CVD-related hospital visits. In other words, it is not beneficial in reducing the demand for cardiac care. According to the Centers for Disease Control and Prevention (CDC), more than 200,000 deaths from heart disease and

stroke each year are preventable. However, the primary challenge is that the treatment and intervention strategies used for CVD are initiated late due to several reasons, such as lack of awareness, symptoms, motivation, or misconceptions. Although CVD detection appears later in life, the risk factors leading to CVD begin in childhood as young as 3 years old, develop in early adulthood, and manifest into clinical disease at later stages [26,27]. Compelling research and empirical studies have identified several childhood/adolescent risk factors that have been associated with CVD in adulthood. These include different adolescent behaviors and characteristics, such as unhealthy diet, smoking, physical inactivity, obesity, blood pressure, and lipids [28–34]. Thus, CVD risk assessments among adolescents can facilitate early intervention and primordial prevention. However, a clinical decision support tool for long-term CVD risk prediction among adolescents is not available. Existing CVD risk assessment tools are based on measuring common risk factors (i.e., age, hypertension, cholesterol, smoking, and diabetes) and predicting events over 10 years or a lifetime.

Existing CVD risk assessment tools are based on measuring traditional risk factors (i.e., age, hypertension, cholesterol, smoking, and diabetes) and predicting events over 10 years or a lifetime. In other words, current CVD risk prediction models do not consider behavioral and lifestyle factors as inputs, but instead focus on factors such as age, gender, race, cholesterol, blood pressure, smoking status, and presence of diabetes to estimate the 10-year or 30-year risk of heart disease or stroke [31,35,36]. Moreover, most of these models are based on cross-sectional data. Further, these risk calculators are only suitable for adults. Therefore, these models are unable to predict the long-term risk of CVD in its

early stages and help the development of preventive interventions. In addition, all standard CVD risk assessment tools are based on the assumption that each risk factor is related to the CVD events in a linear fashion [37]. However, such an assumption may oversimplify the complex relationships between the CVD outcome and a large number of risk factors with non-linear interactions. Thus, other approaches that consider multiple risk factors and determine complex relationships between risk factors and outcomes need to be explored.

Machine learning (ML) offers an alternative approach to standard prediction modeling that may address current limitations. It has the potential to predict an outcome measure with high accuracy based on its relationship with a set of input variables. Besides, it can learn the complex and non-linear interactions between the predictors and an outcome, and identify latent variables, which are unlikely to be observed but might be inferred from other variables [38]. The recent growth in computational power and data accessibility has paved the way for the use of ML algorithms in many healthcare applications. In particular, ML techniques have shown promising results in prior research on disease prediction or risk calculation [39–41].

1.3 Motivation

A practical AS should be designed to improve the operations of cardiology clinics such that patient satisfaction and resource utilization are achieved. In addition, developing tools that facilitate the early detection of CVD is needed to decrease the prevalence of CVD. Given the above, the following serve as a motivation for this research:

1.3.1 Motivation for designing an effective AS for cardiology clinic

i. The increased demand for cardiac care services due to increased CVD prevalence.
ii. The increasing costs of cardiac care could be controlled by improving resource utilization through an effective AS.
iii. The challenges of AS, such as no-shows, variation in consultation length and Sequential appointment requests.
iv. The need for a better way for estimating consultation length and no-show rate.
v. The availability of historical data and electronic medical records (EMR) that could be used to design a data-driven AS that incorporates patients' characteristics.
vi. The dearth of studies that integrate ML-based predictions into the AS design.

1.3.2 Motivation for developing ML-based model for early detection of CVD risk

i. The prevalence of CVD affects more than 50% of the U.S. population, and the projection for its rate to increase to 4.5% in the coming decade.
ii. The increasing costs of cardiac care could be controlled by an early detection of CVD.
iii. The increased prevalence of CVD due to insufficient awareness of CVD, limited planning for early prevention interventions, and the increasing dominance of CVD risk factors.
iv. The need for long-term CVD prediction tools which assist in the early detection of CVD.

v. The lack of CVD risk prediction tools among adolescents.

vi. The need to address the influence of adolescents' risk factors in developing CVD later in life.

vii. The dearth of ML-based models for long-term CVD prediction among the available CVD assessment tools.

viii. The need to develop explainable ML models to enable stakeholders to understand and trust ML solutions.

1.4 Research objectives

The following serve as key objectives for this research:

i. Developing a data-driven ML approach to automate the prediction of consultation length and no-shows.

ii. Identifying critical variables for predicting no-shows and consultation length.

iii. Assessing the effectiveness of adopting the ML-based predictions for allocating the consultation duration during schedule planning.

iv. Developing different sequencing rules for patient scheduling by leveraging no-show and consultation length predictions.

v. Evaluating the proposed appointment rules based on the average total cost, which is a combination of patient waiting time, doctor's idle time, and overtime.

vi. Investigating the impact of the ML-integrated AS design on several clinic environments.

vii. Examine whether a comprehensive set of adolescent factors can be used as inputs to predict the long-term CVD risk score with higher accuracy.

viii. Identify the relative importance of each risk-factor in predicting the CVD risk category.

ix. Assess whether ML models can be converted into more transparent and explainable solutions.

1.5 Organization of the proposal

The remainder of the proposal is organized as follows. A literature review about previous work related to consultation length predictions, design of appointment system, and long-term CVD prediction and the gap in the current work is presented in Chapter 2. The data-driven approach to predict no-shows and consultation length is presented in Chapter 3. The development of ML-enabled appointment system, different appointment and sequencing rules for patient scheduling is detailed in Chapter 4. The development of ML-based model for long-term CVD prediction is described in Chapter 5.

Chapter 2. Literature Review

In this chapter, we review previous work related to allocation decisions, sequencing decisions, and no-show adjustment decisions. Additionally, we review notable literature related to predictive analytics in healthcare, consultation length prediction, and long-term CVD prediction.

2.1 Allocation decisions

The allocation decisions (also referred to as appointment rules) specify the number of patients to be scheduled in each appointment slot (i.e., block size) and the appointment duration. Depending on the allocation decision, an appointment slot can accommodate one or more patients, while appointment intervals can be fixed across all blocks or varied. One of the earliest AS designs was the single-block system, where all patients are scheduled for the same appointment time [42]. However, such an allocation scheme will lead to long waiting times for patients and an excessive burden on physicians. Thus, most of the previous work recommended shifting from a single-block rule to an individual block fixed interval (IBFI) setting, where the clinic session is divided into n blocks of equal intervals, and one patient is scheduled in each block [11,43,44]. Several variants of the IBFI setting were also investigated in the literature. For instance, one of the popular IBFI variants is the 2ATBEG rule, in which two patients are allocated to the first block, and one patient is scheduled to each of the remaining blocks. Since the 2ATBEG rule assumes identical service time distribution for all patients, the appointment interval was set to be equal to the clinic's average service duration [11]. Likewise, Ho and Lau [45] proposed the 4ATBEG

rule and conducted a comparative analysis of 50 allocation rules under different clinic environments. The authors found 2ATBEG and 4ATBEG to be robust as they yielded lower total cost (weighted sum of patient waiting time and doctor idle time) across all test instances.

Previous studies have also investigated allocation decisions that assign fixed appointment intervals with fixed block sizes as well as a more complex system with varying block sizes [17,18,46]. Fries et al. [46] introduced the variable block fixed interval allocation system and showed that it could reduce the weighted cost by up to 40% as opposed to a multi-block assignment. More recently, Choi and Banerjee [47] proposed several variants of variable block rules and found them to provide near-optimal solutions. However, only a few research works have considered variable interval allocation rules [24,48,49]. Ho et al. [48] conducted a simulation study to investigate the performance of different variable-interval rules against the traditional rules. They have concluded that rules with larger intervals at the end of the session yielded the best performance. Several analytical studies also observed that allocation rules that gradually increase the appointment interval up to the middle of the session and then modestly decrease it till the session end time ("Dome" pattern) to be well-performing in most cases, especially when the service times are independently and identically distributed [50,51].

2.2 Sequencing decisions

Sequencing decisions determine the appointment time for each patient. Most prior works consider a static scheduling problem (i.e., all appointment requests are known apriori) and use mathematical programming models or heuristic algorithms to determine the best sequence [13,24,52]. In the case of the sequential scheduling problem, several works classified the patient population into distinct classes (e.g., in terms of service time or patient status) and used them to sequence the appointment requests [42]. Some of the classes used to sequence patients in the literature include new/return patients, variability in the service time (i.e., low/high variation), and type of procedure [18,53,54]. For instance, Klassen et al. [13] proposed to sequence patients based on their service time (i.e., low and high) and evaluated different scheduling rules. They concluded that scheduling patients with low service time standard deviation from the beginning of clinic session could consistently yield lower total cost in different clinic environments. Similarly, Cayirli et al. [42] used patient status (whether new or return) to sequence the patients. They showed that sequencing new patients from the beginning perform well for multi-block fixed interval rules, while sequencing return patients from the beginning are best suited for individual block fixed interval allocation system. Erdogan et al. [55] developed a stochastic mixed-integer programming model to determine the optimal sequence and found FCFA to be a good sequencing rule if the service time distribution and waiting time cost of all patients are assumed to be the same.

2.3 No-show adjustment decisions

The previous work on no-show adjustment can be grouped into two overbooking strategies: (i) schedule one patient per slot but shorten the consultation duration proportional to the expected no-show probability and (ii) schedule two or more patients for the same slot (e.g., double-booking) without altering the estimated appointment interval. One of the earliest studies on appointment shortening was by LaGanga and Lawrence [56], where a utility function was developed to achieve a trade-off between the benefits and risks of overbooking. Their analytical study concluded that it is beneficial to overbook by compressing the appointment interval if the clinic experiences high demand, low service time variability and high no-show rates. Subsequently, LaGanga and Lawrence [57] conducted an extensive simulation study to evaluate several scheduling rules that employ both the overbooking strategies (shortening and double-booking). They concluded the appointment shortening to be suitable for lowering patient waiting time, while double-booking can reduce provider overtime. Cayirli and Yang [12] proposed a universal appointment rule by considering patient classification and adjusted for no-shows by shortening the appointment interval. Their simulation study demonstrated the benefit of such an approach for different clinic environments. With regard to the second overbooking strategy (i.e., two or more patients in the same slot), several works considered patient-specific no-show risk to determine the best slots for overbooking [49,58–61]. Muthuraman et al. [58] classified patients according to their probability of no-show and scheduled them sequentially based on a stochastic overbooking policy as well as a myopic scheduling rule. Srinivas and Ravindran [49] double-booked only patients with low and high no-show risk

to the same slot and compared two policies – double-booking the first available slot and evenly distributing the overbooked slots throughout the schedule. The authors found evenly distributed double-booked slots to reduce provider overtime and patient waiting time, while the other policy lowered doctor idle time. Srinivas and Ravindran [59] designed an AS by considering a double-booking strategy to compensate for no-shows. Their experimental results suggested double-booking at the beginning for clinics with high no-shows. They also found the optimal number of slots to double-book to be dependent on no-show rate, patient waiting time cost and doctor idle time cost.

2.4 Data analytics in healthcare

Over the years, the healthcare industry has generated a large amount of data, resulting from record-keeping, compliance and regulatory requirements, and patient care [62]. Driven by the necessity to improve healthcare delivery while reducing costs, this large amount of data has the potential to support a wide range of healthcare functions such as decision support, disease surveillance, and prevention [63,64]. Therefore, the topic of healthcare data has been investigated recently due to the potential of big data analytics in improving care, saving lives, and reducing costs. Analytics in healthcare is characterized by three domains, namely, descriptive, predictive, and perspective. Descriptive analytics has been investigated in healthcare to analyze patient behavior and gather insights from data summary [65]. For instance, Abdel-Basset et al. [66] proposed a framework based on data obtained from body sensors and patients' mobile phones regarding heart failure symptoms; this data then was used by clinicians to categorize patients into different groups

depending on their symptoms; afterward, different techniques were used to recognize, observe, and control heart failure disease with minimal costs. The application of predictive analytics in healthcare abounds; some of these applications are predicting inpatients' length of stay [67]. Predicting the likelihood of diabetes at early stages using a set of different attributes (i.e., age, BMI, and blood pressure) [68] . Predicting the risk of CVD based on patient lifestyle [69]. Despite the wide range of descriptive and predictive analytics in healthcare, researchers are still investigating the domain of perspective analytics in healthcare. According to Raghupathi et al. [70], healthcare applications of prescriptive analytics include personalized medicine and evidence-based medicine. Some studies have explored data-driven approaches for improving AS performance. For instance Srinivas [49] has suggested different scheduling rules that integrate patient-specific no-shows. It was found that his proposed scheduling rules outperformed the existing rules across all clinic settings. Bentayeb et al. predicted the radiotherapy treatment duration using ML algorithms, then using the predicted durations, new scheduling rules were proposed. Based on the proposed rules, more patients were served, patients' waiting times, and technologists over time were significantly improved [71].

2.5 Consultation length and no-show prediction

Realizing the importance of consultation length, a substantial number of existing works have focused on examining the factors influencing the consultation length [21,72–77]. The impact of patient characteristics was of particular interest in many studies [8,21,72,77]. Some works found older adults, females, and first-time patients to be associated with more

extended physician office visits [8,21,72,77], while a meta-analysis indicated no significant difference in consultation length for male and female patients [73]. Prior research also examined the variables associated with the care provider and the clinic [21,72,77]. Nevertheless, these works may not be suitable or effective for estimating the consultation length of individual patients. To overcome the drawbacks, a limited number of recent studies have sought to predict the duration of physician visits using machine learning (ML) approaches [22,78,79]. Strahl [80] estimated the consultation duration of an outpatient clinic using linear regression and found the patient's gender, appointment time, and surgery history to be significant predictors. They showed a regression-based estimation of consultation length to achieve superior performance over the traditional approach of allocating fixed durations. In another study, Agnes [23] predicted the consultation length for a digital primary care clinic using the doctor's average service duration and other pertinent information extracted from patient-reported symptom form. Recently, Bentayeb et al. [71] predicted the radiotherapy treatment duration based on four factors – cancer category, care plan, patient status, and appointment room. They tested four ML algorithms (linear regression, neural networks, multivariate adaptive regression splines, regression trees), and found the tree-based model to perform the best.

Nonetheless, the literature pertaining to ML-based models for predicting consultation length had several limitations. First, prior prediction models postulate that all patients would adhere to their appointments and do not consider the possibility of no-shows (i.e., zero service time) when predicting consultation duration. However, in practice, consultation length is a semicontinuous variable. It is a mixture of zeros (if patients miss

their appointment) and continuously distributed positive values (if patients attend their appointment). Thus, predicting such semicontinuous data solely based on a continuous ML model is not desirable [81]. Second, they only consider a limited number of risk factors (typically varying from 3 – 4) that were primarily related to the patient characteristic. As a result, many potential risk factors that can improve the prediction performance of the model, such as appointment type, may have been overlooked. Third, raw data obtained from electronic medical records (EMR) were typically used as risk factors in prior research, and little emphasis was given on transforming them into derived variables (feature engineering), which could better represent the problem and improve prediction accuracy [82,83]. Finally, none of the existing work quantifies or assesses the impact of adopting ML-based consultation length prediction on appointment systems' efficiency.

2.6 Long-term CVD prediction

While Sections 2.1-2.5 reviewed the prior research pertaining to the key decisions on designing an AS and approaches for handling clinical uncertainties, this section focuses on reviewing studies pertaining to long-term CVD prediction. The association between childhood/adolescent risk factors and the development of CVD in adulthood has been investigated extensively in the literature. Several prior works examined the impact of a single risk factor on CVD [32,84,85]. Tirosh et al. [32] conducted a prospective study in which the BMI of 37,674 young men were tracked to investigate the association between adolescent BMI and adulthood diabetes as well as CHD. The results showed that higher BMI in adolescence is a significant predictor for both diabetes and CHD in adulthood.

Ferreira et al. [84] observed the carotid stiffness measures for 373 subjects between 13 to 36 years old and concluded that subjects who have stiffer carotid arteries at the age of 36 years were characterized by higher levels of blood pressure between the ages of 13 to 36 years. Besides BMI and hypertension, lipids and their distribution were found to determinants of CVD. Considering its significance, Ferreira et al. [85] assessed the relationship between estimates of body fats and its distribution during adolescent (13 -16 years) and the arterial stiffness at age of 36. It was concluded that the adolescent truncal fat accumulation was correlated with the increased arterial stiffness at the age of 36.

Other studies have investigated adolescent lifestyle (i.e., smoking, diet, and physical activity) and its association with CVD in adulthood [30,31,86,87]. Van De Laar et al. [86] studied the interrelation between smoking in adolescence (age 15 years), and the arterial stiffness in young adults (23 years). Their results showed that smoking in adolescence was associated with higher arterial stiffness in adulthood. Mikkilä et al. [30] showed that a dietary pattern characterized by high consumption of rye, potatoes, butter, sausages, milk, and coffee was positively correlated with developing subclinical atherosclerosis among men. Adolescent physical activity, another important lifestyle factor, has been investigated by prior studies to examine its association with adulthood vascular measures. According to the Young Finns Study [31] and the European Youth Heart Study [87], higher physical activity levels in adolescents are associated with lower carotid intimal-medical thickness in adulthood. On the other hand, mental health-related factors, such as stress and depression, were found to be associated with poor health outcomes, including CVD [88–90]. In addition to the traditional CVD risk factors (i.e., hypertension, lipids, and diabetes),

social determinants of health, which can be represented by socioeconomic status (SES), have a significant impact on CVD development [91]. According to [92–95], four factors of SES have revealed an association with CVD in high-income countries and these include income level, educational attainment, employment status, and environmental factors.

Prior research also investigated the relationship between multiple risk factors such as biomarkers and lifestyle factors in childhood and the development of CVD later in life [31,35,36]. For instance, Raitakari et al. [35] found that the development of atherosclerosis in adulthood was significantly associated with the childhood CVD risk factors such as triglycerides levels, body mass index, systolic blood pressure, and smoking. Similarly [31] investigated if high LDL-cholesterol, low HDL-cholesterol, high blood pressure, obesity, diabetes, smoking, low physical activity, infrequent fruit consumption in childhood were associated with a 6-year change in carotid intima-media thickness (IMT) in young. They found that children with these attributes have a significant association with increased IMT, which indicates a higher atherosclerosis progression rate in adulthood. Another study conducted by [36] examined multiple CVD risk factors (i.e., gender, age, serum lipoprotein concentrations, smoking, hypertension, obesity, and hyperglycemia) in adolescents and their association with the IMT in adulthood. These factors are found to be associated with the increased IMT, indicating a higher CVD risk for young adults.

Although most of the previous research seeks to find an association between adolescent risk factors and adulthood CVD risk, some researchers developed multivariable prediction algorithms intended to assist clinicians in CVD risk assessment. The importance of these

algorithms is emphasized in their incorporation into the treatment recommendations for adults [96–99]. For instance, Assmann et al. [96] constructed a simple scoring model for calculating the risk score of acute coronary events based on the 10-year follow-up for around 5000 men aged 35 to 65 years old. First, they developed a cox proportional hazards model using eight risk variables, including – age, LDL cholesterol, smoking, HDL cholesterol, systolic blood pressure (SBP), family history of premature myocardial infarction, diabetes mellitus, and triglycerides. Then a simple scoring system was created based on the coefficients of the cox model. Their system was able to predict the coronary events with an area under the receiver-operating characteristics curve (AUROC) of 82.4%. Ridker et al. [97] developed the Reynolds risk score for men, and considered C-reactive protein and family history as independent risk factors. Their results showed that including these two variables improves the risk prediction significantly. Another 10-year scoring system is developed by the SCORE project to estimate the risk score of fatal cardiovascular disease in Europe. The system is developed based on 12 European cohort studies and the risk is calculated based on the Weibull model using three independent risk variables (i.e., smoking, cholesterol, stylistic blood pressure) [98]. Unlike the previous CVD risk prediction models, Pencina et al. [99] developed a 30-year risk prediction model for CVD based on the Framingham heart study. They used a modified Cox model that allows adjustment for competing risk of non-CVD death and utilizes standard risk factors that can be collected during a physician's office visit (i.e., sex, age, SBP, smoking).

Most of the previous risk prediction algorithms for CVD used a limited number of risks factors and assumed a linear relationship between the CVD events and input predictors.

Thus, some studies have been conducted to assess whether recent techniques such as machine learning (ML) can improve the prediction accuracy of CVD [100–102]. Kakadiaris et al. [100] developed a ML risk calculator based on support vector machines (SVM) and compared it to the American College of Cardiology/American Heart Association (ACC/AHA) risk calculator. The ML-based tool provided better predictive performance than the ACC/AHA calculator yielding higher sensitivity, specificity, and AUCROC. In another study, four ML algorithms, including random forest, logistic regression, gradient boosting machines, and neural networks, were compared to the ACC/AHA risk calculator to predict the first cardiovascular event over 10-years. As opposed to the established risk calculator, ML algorithms improved the prediction by accurately predicting more CVD and non-CVD events [101]. Kim et al. conducted a comparative analysis of 10 different machine learning algorithms to predict CVD using the National Health Insurance Service – Health Screening dataset. Their results showed that ML models outperformed previously proposed prediction models and improved the CVD prediction. Moreover, their comparative analysis revealed that extreme gradient boosting machines and random forest achieved the highest accuracy and AUC of 81.2% and 81.1%, respectively [102].

2.7 Research gap and contributions

Based on the analysis of the previous work, different gaps in the literature have been identified relating to the design of AS, consultation length prediction, and long-term CVD prediction.

2.7.1 Gaps in the literature

The previous work related to the design of AS has several limitations. First, the majority of studies assumed homogenous patients, and therefore, they consider all patients to have the same probability of no-show and same consultation length. However, these two can differ according to the patients' characteristics and the clinic's environment. Given the previous work, there is a lack of using data-driven approaches in the design of AS. Finally, the vast majority of the previous studies sequenced patients without considering the patient-specific characteristics such as predicted service times of patients. The literature pertaining to ML-based models for predicting consultation length had several limitations. First, prior prediction models postulate that all patients would adhere to their appointments and do not consider the possibility of no-shows (i.e., zero service time) when predicting consultation duration. However, in practice, consultation length is a semicontinuous variable. It is a mixture of zeros (if patients miss their appointment) and continuously distributed positive values (if patients attend their appointment). Thus, predicting such semicontinuous data solely based on a continuous ML model is not desirable [81]. Second, they only consider a limited number of risk factors (typically varying from 3 – 4) that were mostly related to the patient characteristic. As a result, many potential risk factors that can improve the prediction performance of the model, such as appointment type, may have been overlooked. Third, raw data obtained from electronic medical records (EMR) were typically used as risk factors in prior research, and little emphasis was given on transforming them into derived variables (feature engineering), which could better represent the problem and improve prediction accuracy [82,83]. Finally, none of the

existing work quantifies or assesses the impact of adopting ML-based consultation length prediction on appointment systems' efficiency.

With regard to early detection of CVD, existing association studies and prediction models have several limitations. First, most previous research focuses on the impact of a single risk factor (such as gender, total cholesterol, HDL cholesterol, systolic blood pressure, and smoking) on cardiovascular disease, thereby providing limited scope for risk assessment among adolescents. Second, studies that consider the impact of more than one risk factor (such as Framingham risk-score model and ASCVD Risk Estimator Plus) on CVD use different forms of regression or multivariate analysis and assume the risk factors are related to CVD in a linear pattern. As a result, the complex synergistic interaction of risk factors is not recognized. Third, almost all the existing prediction models use factors such as age, gender, race, cholesterol, blood pressure, smoking status, and presence of diabetes to estimate the 10-year or 30-year risk of heart disease or stroke and don't consider other behavioral and lifestyle's factors as inputs to their models. Most importantly, all existing risk prediction models are applicable only for adults above the age of 30 years and are not suitable for determining the long-term impact of unhealthy behavior in the earlier adolescent years. Finally, very little academic research is devoted to the development of a predictive model that is able to categorize adolescents as high or low risk of CVD in adulthood. This research aims to overcome the above-mentioned limitations in the literature. Specifically, the ML-based CVD risk predictor developed in this research applies to the adolescent population. Moreover, the proposed risk scoring method is unconstrained, assuming all possible forms of relatedness of risk factors and incidence of CVD risk.

2.8 Research Contributions

This research has several contributions related to the design of AS, consultation length prediction, and long-term CVD prediction.

2.8.1 Contributions related to consultation length prediction

This study deviated from the literature in the following manner:

i. This is the first study that applied data-driven approaches to predict consultation length in a cardiology clinic to the best of our knowledge

ii. Considering the heterogeneity of patients, patient-specific no-shows and consultation length are predicted using ML algorithms.

iii. The length of office visits for cardiac care was predicted using an ML-based two-part modeling approach, where a classification model was employed to predict no-shows (i.e., categorize the consultation length as zero or positive), and a ML regression algorithm was developed for estimating consultation lengths of patients who are predicted to come for the appointment.

iv. A comprehensive list of 16 risk factors that include both raw and feature engineered data relating to the patient, appointment, and provider were considered. Specifically, this is the first study on ML-based consultation length prediction to include information relating to the appointment as well as other unique risk factors such as the patient-specific historical average of service time and provider's experience.

v. Deep learning and ensemble ML models (such as random forests and gradient boosting) were developed as preceding works on consultation length prediction have not employed them despite their popularity.

vi. Comparative analysis of multiple ML algorithms was considered, and the most prominent model-specific variables for predicting no-shows and consultation duration were identified.

vii. The impact of leveraging ML-based predictions for AS design was studied.

2.8.2 Contributions related to ML-enabled AS design

This research differs from previous works and contributes in the following ways:

i. This is the first work to investigate the impact of integrating ML-based prediction of no-show and consultation length into existing AS configurations.

ii. A predict-then-schedule framework that capitalizes on patient-specific predictions of clinical uncertainties to design an AS is proposed. The proposed framework designs an AS by jointly considering the three decisions (allocation, sequencing, and no-show adjustment) along with patient classification.

iii. Four new online sequencing rules based on the ML-enabled prediction of consultation length are developed.

iv. Clinical data obtained from a cardiology clinic is used for ML-algorithm development, simulation of the scheduling process and validation of the proposed approach.

v. The impact of ML algorithm's predictive accuracy on the performance of the AS is investigated.

2.8.3 Contributions related to long-term CVD risk prediction

This research overcome the beforementioned limitations in the literature and innovates the field in five ways:

i. This is the first study to provide long-term CVD risk ML prediction model among adolescents based on a longitudinal dataset to best of our knowledge.

ii. The proposed risk calculation applies to the adolescent population while currently developed CVD risk calculators only apply to the adult population.

iii. While the statistical expectations of the currently used CVD risk calculator limit the model's prediction ability, the proposed risk scoring method is unconstrained, assuming all possible forms of relatedness of risk factors and incidence of CVD risk.

iv. This study identifies the influence and relative importance of each adolescent risk factor in predicting the adulthood CVD risk score, which, in turn, can highlight the importance of unidentified factors until now. This would provide a new opportunity for future research.

v. This study develops non-parametric machine learning (ML) models, which tend to identify relationships previously masked through the use of stochastic models [37,103].

Chapter 3. Consultation Length and No-show Prediction for Improving Appointment Scheduling Efficiency at a Cardiology Clinic: A Data Analytics Approach

The observed consultation length at specialty clinics, such as cardiology care, is represented by two underlying groups - one with zero service time due to patient no-shows, and the other characterized by positive values with high variance. This inconstancy affects the scheduler's ability to accurately estimate consultation length, which, in turn, hinders effective utilization of the clinic's resources and timely access to care. The objectives of this study were to: (i) predict the consultation length by accounting for its semicontinuous nature (i.e., zero in case of no-shows and positive otherwise), using machine learning (ML) algorithms, (ii) identify important features for predicting no-shows and non-zero consultation length, and (iii) assess the impact of integrating the ML-based prediction with the appointment scheduling system.

3.1 Methodology

To predict the consultation length using an ML-based two-part modeling approach, we adopt the CRoss-Industry Standard Process for Data Mining (CRISP-DM) methodology – an industry-proven framework that provides a structured approach for planning, developing, evaluating and executing data analytics tasks [104,105]. An overview of the CRISP-DM methodology is shown in Fig. 1, and the following subsections describe its main phases.

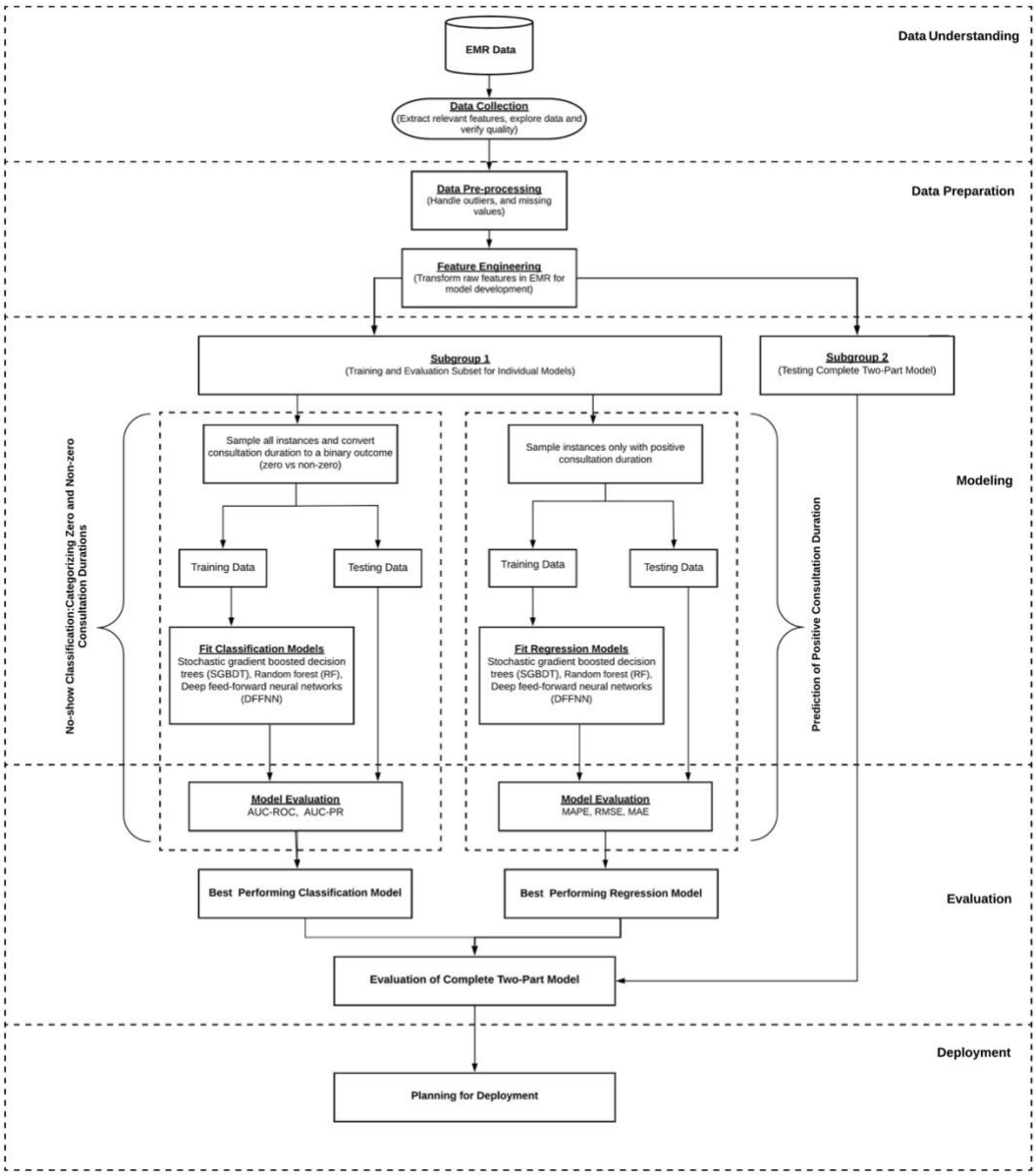

Figure 1: Overview of CRISP-DM framework for predicting consultation length

3.1.1 Business and data understanding

The clinic's goal was to utilize the resources effectively without compromising on care quality by scheduling patients for the optimum consultation duration. In this research, we

used two years of EMR containing 25,523 patient visits to the cardiology clinic. Each appointment booking in the EMR included information regarding the following three categories.

- Patient Characteristics – Provides information on the patient's age, gender, race, marital status, insurance type, and Zone Improvement Plan (ZIP) code.
- Appointment Information – Provides a detailed description of events (action) for each patient call and visit, such as booked, canceled, no-show, checked-in, and checked-out. Also, the EMR provides the associated timestamp (mm/dd/yyyy hh:mm:ss format) for each action. This information is imperative to retrieve various information such as scheduled consultation duration, appointment time, and observed consultation duration.
- Resource Description – Indicates the doctor assigned for each appointment.

3.1.2 Data pre-processing and feature engineering

Typical of any real-world data, the EMR also contained incorrect information, missing values, outliers, and redundant features, which, in turn, reduced the data quality. As a result, the data had to be pre-processed before being used in the predictive models. Incorrect information (such as negative values for patient age) and missing values accounted for 1,456 cases, out of which 1,117 instances were rectified by obtaining the updated information from the clinic. The remaining 339 unrecovered missing values were imputed by chained equations [106], where the distribution of unobserved value was estimated based on the observed values. In particular, if there were M independent variables, then the variable (e.g., x_1) with missing values was regressed on the other independent variables,

(x_2, x_3, …, x_M), by considering only the observed values, and subsequently, the missing values in x_1 were obtained using the predictions from the fitted model. The procedure was repeated for each variable containing one or more missing values to obtain a complete dataset. Also, the data included 434 outliers, where the consultation duration of patients who came for the appointment was either unreasonably long (1200 minutes) or short (< 2 minutes). This was mainly due to the data entry error by the clinical staff responsible for entering the timestamps for check-in and check-out. Thus, these outliers were treated as missing values and subsequently imputed. Upon pre-processing, the consultation length was zero in 4,594 instances (i.e., ~18% no-shows and late cancellations) and positive for the remaining 20,929 cases. The non-zero visit length varied from 15 to 122 minutes with a mean of 50 minutes and a standard deviation (SD) of 15.62 minutes.

Upon dealing with data inconsistencies, we leveraged available data to establish 16 predictors pertinent to the patient, appointment, and doctor, as shown in Table 1. Four features were directly obtained from EMR, while the remaining predictors were derived using existing information. In particular, the patient status was derived from the historical appointment visits and categorized as new or return. The neighborhood socioeconomic status (nSES) was estimated based on seven characteristics associated with the patient's ZIP code - unemployment rate, proportion below poverty level, median value of dwelling, median household income, percentage of households with mean of ≥1 person per room, high school graduation rate, and 4-year college graduation rate. The census information collected by the American Community Service (ACS) [107] was leveraged to extract ZIP code-related data and the SES index, a weighted combination of the seven neighborhood

characteristics, conceived by the Agency for Healthcare Research and Quality (AHRQ) was used to calculate nSES [108]. The nSES ranges from 0 to 100, where a higher number reflects a more significant neighborhood deprivation.

Table 1: Description of features extracted after pre-processing

Predictors	Value	Type	Source
Patient Characteristic			
• Gender	Male or Female	Categorical	Raw
• Age	1 – 90 years	Continuous	Raw
• Marital Status	Single Married Divorced Separated Widowed	Categorical	Raw
• Insurance Group	Medicare Medicaid Private	Categorical	Raw
• Patient Status	New or Return	Categorical	Derived
• Neighborhood Socioeconomic Status	0 – 100	Continuous	Derived
Appointment Information			
• Appointment Month	Jan, Feb, …, Dec	Categorical	Derived
• Appointment Day of Week	Mon, Tue, …., Fri	Categorical	Derived
• Appointment Time Category	Morning or Afternoon	Categorical	Derived
• Appointment Delay	0 – 48 days	Categorical	Derived
• Appointment type	Cardiology Cardiothoracic Surgery Heart/Vascular Intervention Vascular Surgery	Categorical	Derived
• Prior no-show rate	0% – 100%	Continuous	Derived
• Service Time History	15 – 122 minutes	Continuous	Derived
Doctor Information			
• Gender	Male Female	Categorical	Derived
• Race	White Non-White	Categorical	Derived
• Experience	8 – 55 years	Continuous	Derived

The appointment information, namely, month, day, category, and delay (the difference between appointment date and booking request date) were derived from the timestamp data. Similarly, the appointment type (i.e., the reason for visit) was extracted from the action description notes stored in the EMR and categorized as cardiology, cardiothoracic surgery, heart & vascular intervention, and vascular surgery. The history of zero service time rate (or prior no-show rate) for a patient was calculated as the percentage of appointments for which that patient did not show up during the last one year. Likewise, the positive service time history for a patient was determined by averaging all previously observed visit duration for that patient in the past year. With regards to the provider's characteristics, all three attributes, namely, gender, race, and experience, were extracted from public online sources based on the resource description provided in the EMR. Subsequently, the categorical variables with more than two levels were encoded as separate feature vectors with one-hot encoding [109].

The dependent variable for the classification part was a binary variable representing the consultation duration as zero (equivalently no-show) or non-zero. The outcome variable for the regression part was the positive consultation length – a continuous value calculated as the difference between the patients' check-out time and check-in time. The outcome variables were derived using action descriptions and the corresponding timestamps in EMR.

3.1.3 Machine learning-based two-part predictive modeling and evaluation

The problem of predicting the no-show risk (i.e., categorizing consultation length as zero or non-zero) was modeled as a supervised classification problem, and the aspect of predicting the non-zero visit length was treated as a supervised regression problem. Recent research had demonstrated the capability of random forests, stochastic gradient boosted trees, and deep neural networks to accurately predict both binary and high-variance continuous variables in healthcare domain such as disease risk [28,29], length of inpatient stay [30] and mortality risk [31]. Therefore, we employed random forest classifier (RFC), stochastic gradient boosted classification trees (SGBCT), and deep neural network classifier (DNNC) for no-show classification (i.e., categorizing consultation length as zero or positive). Likewise, we considered random forest regressor (RFR), stochastic gradient boosted regression trees (SGBRT), and deep neural network regressor (DNNR) for predicting the non-zero consultation length.

Similar to the procedure adopted by Povak et al. [110], the ML models for classification and regression were developed separately. A stratified random sampling was performed to divide the data into two parts – 75% was used for training and evaluating both classification and regression models (Subgroup 1), and the remaining 25% was held-out for assessing the complete two-part model (Subgroup 2). There were 19,142 samples (out of which 15,702 instances have non-zero consultation duration) in Subgroup 1 and 6,381 records (5,227 samples with > 0 visit length) in Subgroup 2. The no-show classification model was trained using 75% of the randomly sampled data from Subgroup 1 and tested on the remaining samples. Note that the service time values were converted to a dichotomous

variable (zero or non-zero outcome) for no-show modeling and evaluation. Similarly, a 75/25 random split of samples containing the positive consultation durations from Subgroup 1 is utilized for training and testing regression models. Besides, the important features for predicting the outcome were established by randomly permuting the value of the predictor and evaluating its impact on prediction error [111]. To eschew overfitting (learning noise) in the learning phase and optimize the parameters of ML models, a 10-fold cross-validation procedure was performed on the training data [112,113]. A brief description of the three ML techniques is provided in the following subsections.

3.1.3.1 Random forest regressor (RFR)

Random forest algorithm is an ensemble of uncorrelated decision trees, where each tree is trained based on two concepts – bagging and random feature selection [114]. The bagging technique samples a subset of the training data with replacement and uses it to train a decision tree, while the random feature selection technique only considers a subset of predictors for splitting at every node of a decision tree [115]. The combination of these two techniques reduces the risk of multiple decision trees learning the same information from the training set. Each individual tree (*t*) is grown as follows [116]:

- From *N* samples in the training dataset, *n* (<< N) samples are drawn randomly with replacement to construct a regression tree (f_t)
- At each node of a regression tree, $m << M$ variables are drawn randomly and the sum of squared errors (SSE) for splitting the dataset into two regions (R_1 and R_2) on every possible value for the chosen *m* variables is computed (Equation 1). The

variable/value combination that yields the lowest SSE is chosen for splitting the node into two sub-nodes.

$$SSE = \sum_{i \epsilon R_1} (x_i - \bar{x}_1)^2 + \sum_{i \epsilon R_2} (x_i - \bar{x}_2)^2 \quad (1)$$

where x_i is the value of predictor *i* and $\bar{x}_1, \bar{x}_2$ are the means of right and left sides of the possible splits respectively. This procedure is repeated for all the sub-nodes until a stopping criterion is reached (e.g., maximum depth is reached)

Upon independently training *T* regression trees in parallel, the final prediction is obtained by aggregating the individual predictions the *T* trees. The performance of RF is affected by the value of *T* and *m*.

$$\hat{Y} = \frac{1}{T} \sum_{t=1}^{T} \hat{Y}_t \quad (2)$$

3.1.3.2 Stochastic gradient boosted classification trees (SGBRT)

Stochastic Gradient boosted decision trees (SGBDT) also use a multitude of regression trees for learning the mapping function between predictors and outcome. However, unlike RF, it builds the regression trees in a stage-wise fashion, where subsequent tree seeks to boost the performance over previous tree's prediction by emphasizing on samples with high SSE. Basically, the algorithm for boosted regression trees is an iterative process that follows the following steps [117]:

- Initialize model with a constant value $f_0(x) = \bar{y}$
- For each boosted model *t*
    - Compute the residual of previous model as $h_t(x) = y - f_{t-1}(x)$
    - Fit a regression tree ($g_t(x)$) on the residuals of the previous tree by using a subset of the training data drawn without replacement
    - Compute the gradient descent step size ($\rho_t(x)$)
    - Combine the trained trees to obtain a boosted model, $f_t(x)$, which has lower SSE than $f_{t-1}(x)$. Also, to prevent overfitting a shrinkage parameter ($J \in (0,1]$) is used to control the learning rate. Therefore, the combined model is represented as $f_t(x) = f_{t-1}(x) + J \times \rho_t(x) \times g_t(x)$

The procedure is repeated until the reduction in SSE is insignificant. The performance of the SGBDT depends on the following hyperparameters – number of trees to construct (*T*), learning rate (*J*), and the minimum number of samples in terminal node (*C*) of tree to control its depth.

3.1.3.3 Deep neural network regressor (DNNR)

Deep neural network regressor (DNNR) is a type of deep learning model, where several nodes arranged as linear stack of three different layer types, input, hidden and output, are used to best approximate the mapping function [118]. The nodes in adjacent layers are interconnected by weighted edges. Unlike a simple neural network with one hidden layer, a DNNR has many hidden layers to facilitate the learning of complex functions and better prediction [119]. The input layer nodes receive information from the *M* predictors and

transmit the weighted input to the nodes in the adjacent hidden layer, which then transform it using activation function, to account for non-linear relationships, before feeding that information to the subsequent hidden or output layer. This process, known as forward pass, continues until the information from the last hidden layer is processed and transformed by the output layer nodes to predict the consultation length. The training data is fed into the DNNR in small batches, where a batch contains one or more training records for learning purposes. The connection weights between nodes i and j (w_{ij}) of DNNR are initialized randomly and progressively altered Equation (3) for every successive batch (b) based on the difference between the actual and predicted output (E) obtained from a previous batch (b – 1) and network's learning rate (η) - referred to as backward pass.

$$w_{ij}(b) = w_{ij}(b-1) + \eta \frac{\partial E}{\partial w_{ij}} \quad (3)$$

Moreover, a DNNR may require several epochs, number of passes through the entire training data, where each epoch has one or more batches before approximating the mapping function. The learning rate (η), batch size (b), number of epochs (P), number of hidden layers (H) and the number of nodes in hidden layer k (n_k) affect the performance of the DNNR.

3.1.3.4 Hyper parameters tuning

The hyperparameters of the three ML models were tuned using the grid search technique as it was well-suited for models with few parameters and easy-to-implement [120]. The hyperparameter space screened for each algorithm was based on the commonly used values

in the literature [121–126]. The number of trees (T) considered for RFC, RFR, SGBCT, and SGBRT were (2^5, 2^6, 2^7, 2^8). Besides, the number of variables randomly sampled at each candidate split (m) in the RFC and RFR algorithms were varied from 5 to 15. For SGBCT and SGBRT, the learning rate (J) explored was between 0.025 to 0.1, and the minimum number of observations in a trees' terminal node (C) was varied from 2 to 8. For DNNC and DNNR, the number of hidden layers (H) examined were (1, 2, 3, 4, 5), while the number of nodes in each hidden layer (n_k) was varied from 1 to 16. Moreover, the values of learning rate (η), number of epochs (P) and batch size (b) searched were (0.01, 0.05, 0.1), (10, 20, 30, 40, 50), and (32, 64 ,..., 512), respectively. Besides the hyperparameters, the cut-off (γ) value, which converts the predicted class probability of no-show classification model to a label, was varied from 0.4 to 0.8 in increments of 0.01.

Once the ML models for no-show classification and consultation length prediction were trained and evaluated on their respective datasets, the overall prediction performance of the complete two-part model (best-performing model from each part) was assessed on a separate held-out dataset (Subgroup 2). Note that the outcome variable for testing the complete two-part model was semicontinuous and had both zero and non-zero consultation durations. Hence, when evaluating a record using the complete two-part model, if the classification algorithm predicted the no-show probability to be below a cut-off value (γ), only then the regression model was leveraged to predict the visit length of that patient; else the consultation duration was set to zero.

3.1.3.5 Evaluation metrics

The classification models were compared based on two measures, namely, area under the receiver operating characteristic curve (AUC-ROC) and area under the precision-recall curve (AUC-PR). The AUC-ROC is a single-measure for evaluating the overall discriminative performance of the ML model [127]. Besides, AUC-ROC had been consistently used in prior research dealing with no-show predictions [59]. On the other hand, AUC-PR is considered to be a robust metric for evaluating models dealing with class imbalances [128]. The value for both these metrics ranges from 0 to 1, and a higher score is preferred. In addition, the best cut-off value (γ) was chosen based on the harmonic mean of precision and recall (i.e., F_1 score) and the misclassification rate (MCR) associated with the best γ was also considered to assess classification model performance.

The ML models for non-zero consultation length prediction were evaluated based on three commonly used evaluation metrics in the literature – mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) [129]. Since these three are error measures, lower values are preferred. Finally, the complete two-part model that predicts the consultation length in Subgroup 2 was evaluated based on MAE and RMSE. Note that MAPE was not employed owing to the semicontinuous distribution of the outcome that would lead to undefined or infinite values for instances involving zero consultation duration. Besides facilitating the selection of the best performing model, the AUC-PR and RMSE measures were also used to determine the best combination of hyperparameters for classification and regression models, respectively.

3.1.4 Deployment of ML approach

The predictive model in this study could be integrated into the clinical decision support system to automate the task of estimating the no-show risk and consultation duration for the scheduler. Fig. 2 provides the schematic for deployment and integration.

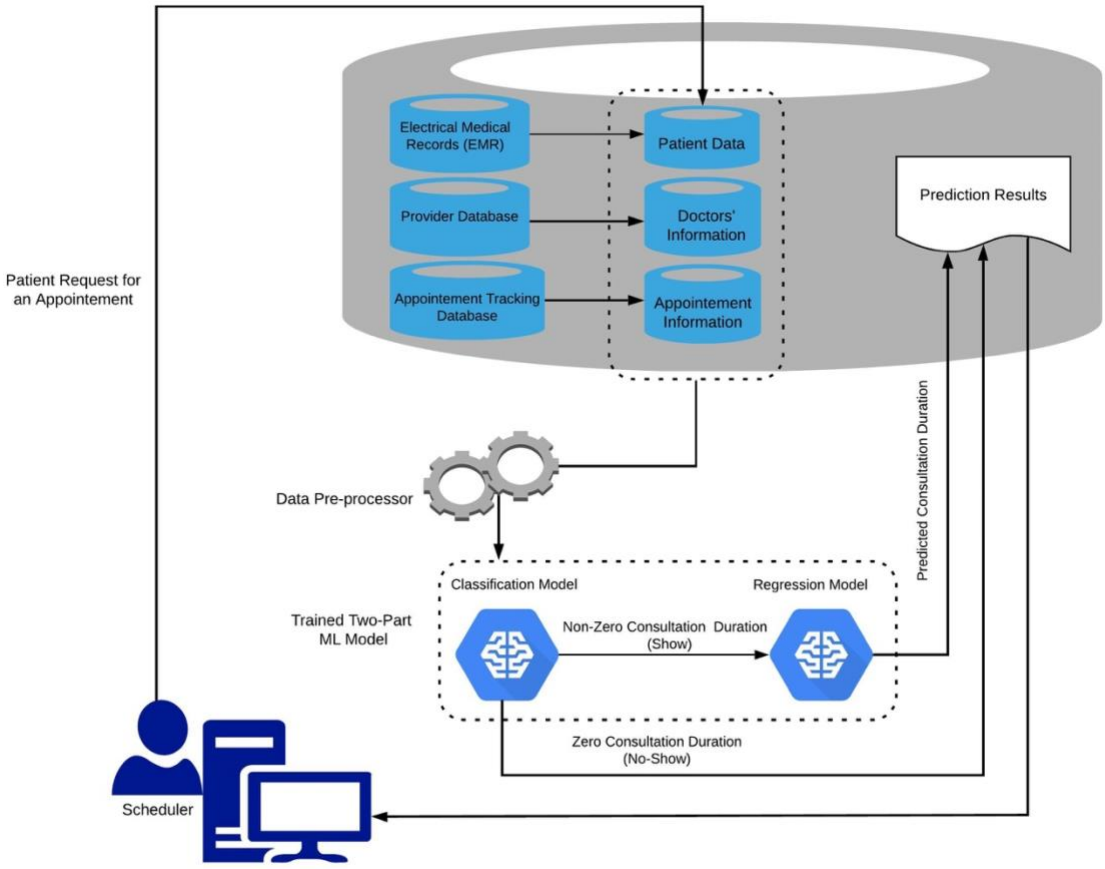

Figure 2: Deployment of two-part prediction approach for scheduling patients

When a patient calls for an appointment, the best performing ML algorithm will leverage the patient, doctor, and appointment information from the electronic database to estimate the patient-specific consultation duration in real-time. Subsequently, the scheduler would

be able to use this information to allocate the consultation duration while scheduling the patient. Furthermore, the integrated system would routinely update its model to learn from new data.

3.1.5 Impact of integrating ML predictions for automating scheduling decisions

As discussed earlier, allocating the right consultation duration is necessary to optimize the provider's schedule. Besides, the effectiveness of a schedule has a direct impact on the provider's productivity and patient satisfaction [130]. To assess the potential impact of adopting the ML-based predictions for allocating consultation durations at the time of scheduling patients, a comparison with the clinic's existing approach was conducted using simulation modeling. Specifically, the procedure to estimate the consultation length in the current approach depends on the patient status (new vs. return) and reason for the visit. The scheduler follows the clinic's guideline to allocate the consultation duration for new patients, which varies depending on the appointment type (e.g., cardiology, vascular surgery). However, in the case of an established patient, the provider indicates the expected consultation length of the return visit in their notes, which is then used by the scheduler to allocate the specified duration. Moreover, the clinic does not adopt a pre-defined template for making scheduling decisions. It determines the appointment time for each patient, depending on his/her preference and appointment availability. To compensate for no-shows, the clinic accepts more appointments than its capacity and randomly overbooks some time slots based on the clinic's average no-show rate.

In this research, the existing scheduling approach (i.e., sequence of patients scheduled for appointments and their allocated consultation length) was reconstructed by retrospective retracing of the EMR data. Two simulation models were developed – (i) the first model mimicked the existing schedule based on the information extracted from EMR, and (ii) the second model represented the proposed approach, where the reconstructed sequence was adopted, but the prediction from the best performing classification and regression ML algorithms was leveraged for determining the overbooking decision and consultation duration, respectively. Therefore, if a patient was predicted to be a no-show, then that patient's appointment is scheduled to overlap with previously booked patients in the reconstructed sequence; otherwise, a patient is not overbooked. On the other hand, the consultation duration allocated for a patient is equal to the visit length estimated from the best performing ML regression model rounded to the nearest multiple of five. The simulation models of the two methods were run, and their performances were compared based on critical schedule outcome measures, namely, patient waiting time and doctor idle time.

3.2 Results

The CRISP-DM procedure for predicting consultation length using the two-part ML-based approach was implemented in R statistical software on a computer configured with Intel Core i7 3.4 GHz processor, macOS Sierra operating system, and 32 GB RAM. The training and testing data for classification models contained 14,356 and 4,786 samples, respectively. Out of these, 2,588 training instances and 814 testing records had zero

consultation duration (i.e., patient no-show). On the other hand, training data for regression models consisted of 11,776 instances (with mean consultation length of 49.3 minutes and SD of 16.1 minutes), and the corresponding testing set included 3,926 instances (with mean visit length of 50.2 minutes and SD of 14.7 minutes). The best set of hyperparameters for classification and regression models are summarized in Tables 2(a) and 2(b), respectively. Besides, the best cut-off values for RFC, SGBCT and DNNC were 0.58, 0.61 and 0.66, respectively.

Table 2: Best hyperparameters for classification and regression models

(a) Best hyperparameter values for classification models from grid-search

Classifier	**Hyperparameter**	**Values**
RFC	T	64
	m	7
SGBCT	T	64
	C	4
	J	.05
DNNC	H	2
	n_k	5
	P	20
	η	.01
	b	32

(b) Best hyperparameter values for regression models from grid-search

Regressor	**Hyperparameter**	**Values**
RFR	T	128
	m	12
SGBRT	T	128
	C	6
	J	.025
DNNR	H	4
	n_k	10
	P	50
	η	.01
	b	32

3.2.1 Predictive performance of classification and regression models

The 10-fold cross-validation performance of different no-show classification and non-zero consultation length prediction models is illustrated in Fig. 3(a) and Fig. 3(b), respectively. The evaluation metrics indicate a good discriminative capability of zero and non-zero service times of all the classification models. In particular, SGBCT yielded the best average values for all the classification evaluation metrics under consideration. It achieved 2% better AUC-ROC, 4% higher AUC-PR, and 7% lower MCR than RFC. Similarly, when compared to DNNC, SGBCT showed 4%, 9%, and 23% improvement in AUC-ROC, AUC-PR, and MCR values, respectively. Among the ML models tested for non-zero consultation length prediction, the results demonstrated the DNNR to be the best performing model with respect to all error measures. It dominated RF by engendering 18%, 19% and 16% lower RMSE, MAE and MAPE, respectively. Likewise, it systematically outperformed SGBRT with 10%, 9%, and 7% reduction in RMSE, MAE, and MAPE, respectively.

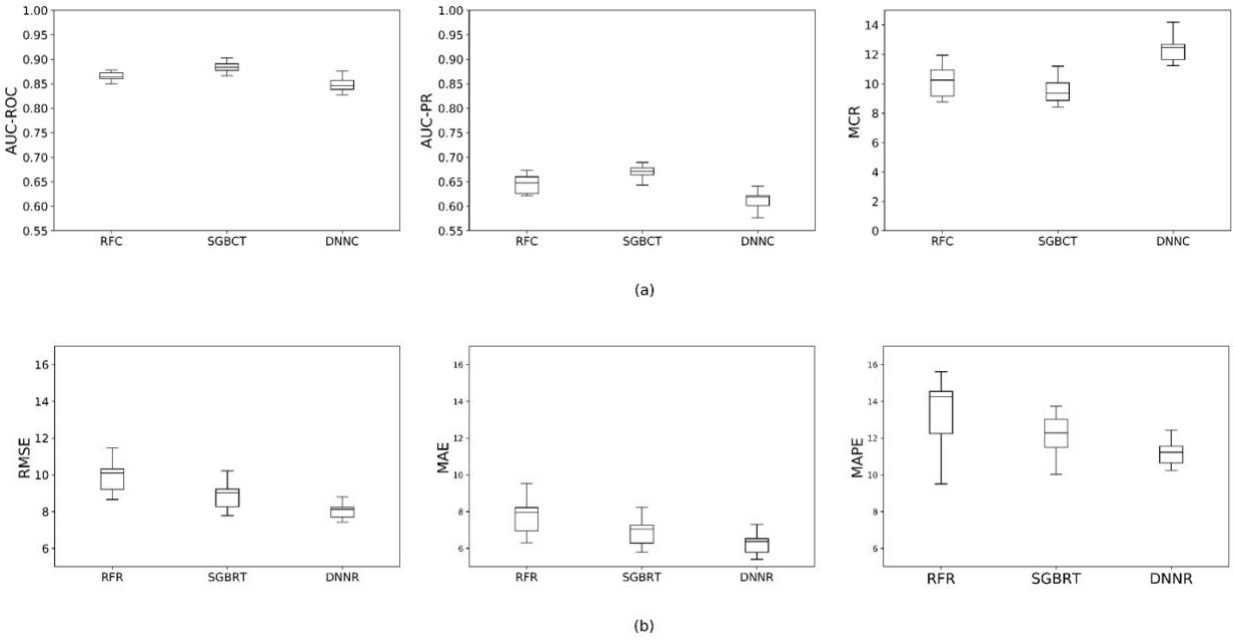

Figure 3: Cross-validation performance of ML algorithms on (a) no-show classification and (b) non-zero consultation length prediction

Table 3 presents the performance of ML models for no-show classification and consultation length prediction on the testing dataset, and confirms our previous findings based on cross-validation. In other words, the performance of classification and regression models on testing data were comparable to the cross-validation results, thereby suggesting the generalization capability of the ML models considered. Besides, SGBCT and DNNR consistently outperformed the other two algorithms for no-show classification and non-zero consultation length prediction, respectively.

Table 3: Performance assessment of classification and regression models on the testing dataset

No-show Classification				**Non-zero Consultation Length prediction**			
Classifier	**AUC-ROC**	**AUC-PR**	**MCR**	**Regressor**	**RMSE**	**MAE**	**MAPE**
RFC	0.83	0.61	13.1%	RFR	10.88	8.22	14.71
SGBCT	0.85	0.64	11.6%	SGBRT	9.65	7.10	13.26
DNNC	0.83	0.60	13.7%	DNNR	8.55	6.88	12.24

3.2.2 Testing complete two-part model on new data

Based on the initial analysis, the best performing classification and regression models were combined to establish the complete two-part model (i.e., SGBCT + DNNR). The performance of SGBCT+DNNR was tested on new data (Subgroup 2), which had a semicontinuous consultation duration with 1,154 zero instances and 5,227 positive samples (mean $\pm$ SD = 49.6 $\pm$ 15.8 minutes). To assess the efficacy of the complete two-part model, we also calculated the error measures for the clinic's current approach to visit length estimation (discussed in Section 2.6) by comparing the consultation duration allocated by the scheduler against the observed consultation length for all the instances in Subgroup 2. The results illustrated in Fig. 4 demonstrated the dominance of the ML-based two-part approach. As opposed to the current approach, the consultation length predictions from SGBCT+DNNR resulted in 50% and 52% lower RMSE and MAE, respectively.

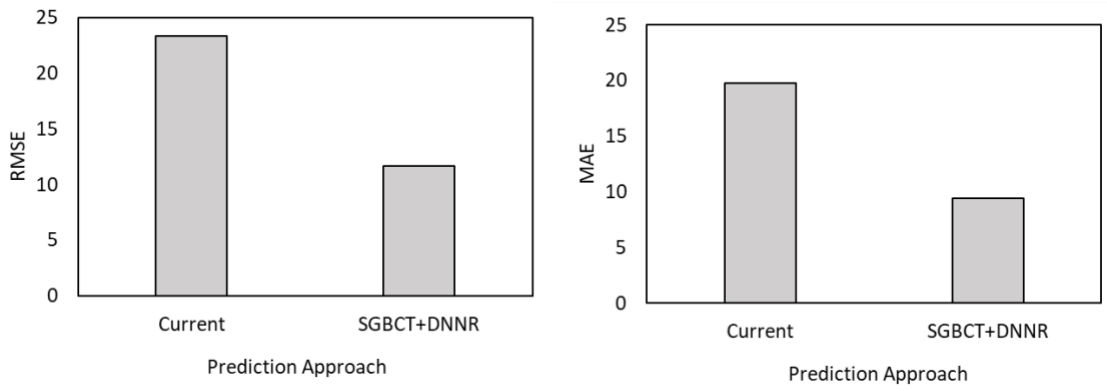

Figure 4: Performance assessment of the complete two-part model on testing dataset

3.2.3 Important variables for predicting no-show and consultation length

Fig. 5(a) illustrates the five most important variables for no-show classification (i.e., zero vs. non-zero consultation length). Appointment delay and prior no-show rate were consistently ranked as the two most important predictors by all three models. Besides, the patient's age was identified to be among the five most important variables by all no-show classification models, while nSES and appointment time category (morning vs. afternoon) were ranked by at least two of the three classification models. Interestingly, none of the attributes associated with the provider appeared as critical predictors for the problem of no-show classification.

Fig. 5(b) shows the top five features for predicting positive consultation length by the regression algorithms. Patient status (new vs. return) was identified as the most important variable by all the regression algorithms. In addition, the provider's years of working experience consistently emerged as the second most important feature. While the encoded appointment type feature representing cardiothoracic surgery, appointments was estimated

to be a critical predictor by the three models, its importance ranking varied depending on the ML algorithm. Patient's age and service time history were listed as essential predictors for two of the three regression models. Also, the patient's age was the only variable to be listed in both no-show classification and consultation length prediction as a prominent predictor. Certain features were considered to be critical only by a single ML algorithm. For instance, the encoded appointment type variable for vascular surgery was an essential predictor for SGBRT, while the doctor's gender was a vital predictor for RFR. Besides, regression algorithms included variables from all three levels in the list of the top five important variables.

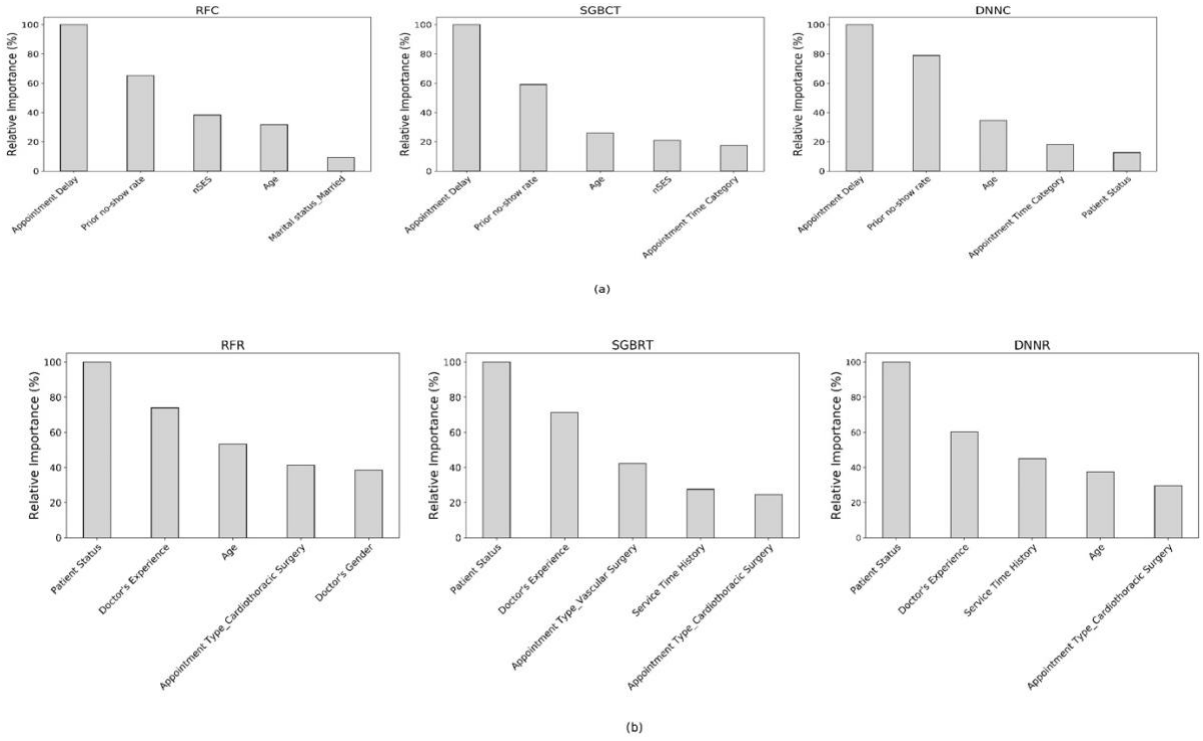

Figure 5: Top 5 algorithm-specific variables and its relative importance for (a) no-show classification and (b) non-zero consultation length prediction

3.2.4 Schedule performance of ML-enabled approach and existing approach

In order to assess the effectiveness of adopting the ML-based predictions for allocating the consultation length during schedule planning, the simulation models for the proposed and existing scheduling approaches (discussed in Section 2.6) were run for 100 clinical days by replicating the exact patient sequence observed in the EMR. The average waiting time per patient and average idle time per doctor across 100 clinical days for the two approaches are presented in Fig. 6(a) and 6(b), respectively. If the clinic adopts the prediction obtained from the best performing two-part model (SGBCT+DNNR) for determining the consultation length during patient scheduling, then it has the potential for a 56% improvement in patient waiting time and 52% reduction in provider idle time. A *t*-test with $\alpha = 0.05$ indicated the proposed approach to be significantly better than the current practice with respect to patient waiting time and physician idle time. In other words, ML-enabled consultation length determination would substantially improve patient satisfaction (due to the reduction in waiting time) and provider productivity (as a result of the cutback in idle time).

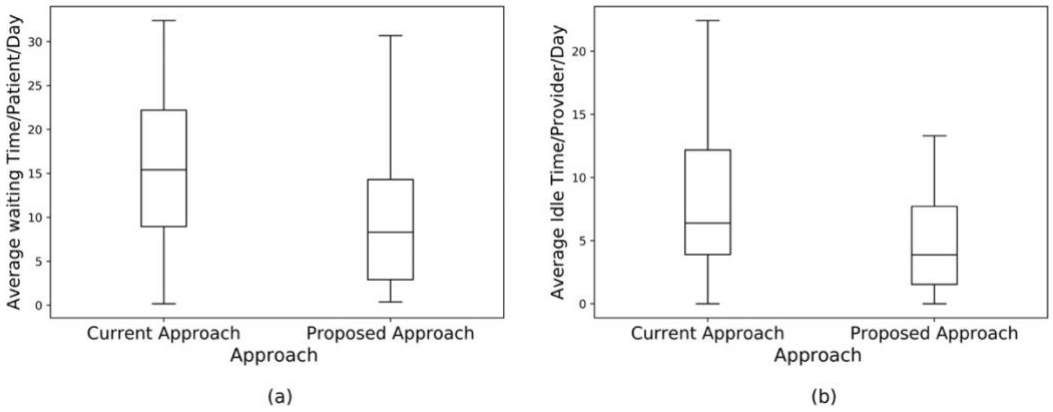

Figure 6: Comparison of schedule outcomes for the current and proposed scheduling approaches

3.3 Discussion

Our results demonstrated the superiority of the ML-based two-part approach in the estimation of semicontinuous consultation length. These results also added to the existing knowledge of the literature on consultation length prediction [22,78,79]. Moreover, the ML algorithms under consideration were scalable for deployment, as observed by prior research works [131]. This study also identified the variables that were most critical in predicting the consultation length. Consistent with previous research, we found appointment delay, prior no-show rate, and patient's age to be important features for predicting missed consultations [132,133]. Earlier studies had also established an association between low SES and a higher likelihood of no-shows [132,134], and our study results further substantiated this as nSES was found to be a critical predictor of appointment non-adherence. Likewise, the two most prominent variables identified for predicting the non-zero consultation length, namely, patient status, and doctor's experience, also

concurred with the findings of previous research [8,21,72,77]. Our experiment also divulged new critical predictors, such as appointment type and service time history. Besides, the study results provided new insights into the learning pattern of the ML algorithms. Specifically, the tree-based ensembles (RFC and SGBCT) were able to extract more information from features such as nSES as opposed to DNNC. Also, information pertaining to the doctor appeared to be inconsequential for no-show classification but played a pivotal role in predicting the non-zero consultation length.

The findings of this research had several clinical implications, especially for specialty clinics such as cardiac care, which experiences high provider burnout [135], depressed patients [136], and rising costs [137]. First, the proposed approach automates the consultation time management for clinics, which, in turn, would reduce the administrative burden for the scheduler. Besides, the integration of the prediction model would enable a digital healthcare system, an important transition necessary for transforming processes and achieving cost-effectiveness [138,139]. Second, the variables and methods adopted in this research could be promising for designing and improving the performance of the clinic's appointment system. Third, the risk of schedule underruns and overruns reduces significantly, thereby improving clinical workflow. Therefore, providers' schedule will be time-synchronized, and they will not be burdened with inadequate consultation length, a factor which has the prospect to mitigate burnout and medical errors [21,72,74]. It will also benefit the patients as they are likely to wait less and spend quality time with their provider. These aspects constitute factors that may lead to higher patient satisfaction, as suggested by prior studies [21,73,77]. Finally, the critical predictors identified in this study can be

leveraged to control the consultation duration. For instance, this study found patients from economically disadvantaged neighborhoods to miss their appointment. This could be due to restricted access to transport facilities for patients from areas with low SES [140,141]. To mitigate no-shows, clinics could offer transportation assistance, including local transit options and low-cost transport services. Similarly, this study found new patients to take longer consultations. This was primarily due to additional administrative pre-processing and medical history reporting required for new patients. To shorten the visit length, this information could be collected via online forms before the patient's visit.

Although this study had many merits, it also had two notable limitations. First, the approach was tested only for a single cardiology clinic and therefore limited the generalization of the study findings. Future research should repeat the study procedure across multiple clinics, specialties, and locations to support the findings of this study. Second, the fields available with the clinic under consideration were used to build the prediction model. Therefore, it is plausible to have missed certain relevant predictors, such as medical reports and prior test results. The creation of extended databases and standardization of the fields collected across clinics would further enhance the capability of these ML algorithms.

3.4 Conclusions

Despite the potential for automating routine clinical works using ML algorithms, its adoption in the healthcare domain is not prevalent. In this study, we adopted the CRISP-DM methodology to build an ML-based two-part model for predicting the consultation

length while considering the possibility of no-shows. A supervised classification model categorized the consultation duration as zero or non-zero (equivalently patient no-show or attendance), and a supervised regression algorithm predicted the consultation length for instances classified as non-zero. The proposed approach was validated using the data obtained from the EMR system of a cardiology clinic, thereby demonstrating its feasibility. Besides, several clinical insights were obtained from our proof-of-concept study, including the identification of critical predictors, impact on appointment system design, and potential benefits to patients as well as providers. The two-part ML-based technique considered in this research could be extended to other outpatient clinics to facilitate better planning and potentially improve the health outcomes. Also, future work can leverage patient-specific no-show and consultation length predictions to develop dynamic scheduling rules that optimize the clinic's operational performance.

Chapter 4. Predict, Then Schedule: Prescriptive Analytics Approach for Designing Machine Learning-enabled Appointment System

As discussed in chapter 1, several uncertainties can pose a challenge for designing an effective appointment system (AS), which can lead to lower patient satisfaction and ineffective resource utilization. One of the leading causes of uncertainty in AS is patient no-shows – which arise when patients fail to show for their appointments without earlier notice. This affects the clinic adversely since the allocated consultation length can be wasted. To compensate for no-shows, clinics tend to double-book more patients than their capacity [9]. The majority of clinics assume patients are homogenous and have the same probability of no-shows. Therefore, they double-book by an even percentage equal to the average probability of no-show in the clinic [142]. Another significant source of uncertainty in AS is variation in consultation length, which affects the ability of the scheduler to allocate provider time accurately. Scheduling shorter appointment times than needed overburden their physicians with more patients and increase the average waiting time for patients. Furthermore, such practice may also hurt the quality of care provided. On the contrary, appointment durations that are longer than required lead to doctor idle time and fewer patient visits. Therefore, this chapter leverages the predictions of service time and no-show risk obtained in chapter 3 to develop different appointment systems for patient scheduling.

In particular, this chapter considers an outpatient clinic that serves patients on an appointment basis during a clinic session. The daily operating hours (or session duration)

are divided into smaller time intervals called blocks or slots. The patient population served by this clinic can be categorized into two distinct classes (or type) depending on the patient attributes such as nature of visit or reason for visit. Besides, the characteristics for each patient class, such as average consultation duration and no-show rate varies. The patient calls for an appointment arrive sequentially, and the scheduler must decide whether to accept or reject each request before the end of the call. In addition, the scheduler must also determine the appointment start time and appointment duration for each accepted request. Therefore, from the clinic's perspective, each patient has an associated start time and end time for a particular appointment. However, the actual appointment times may differ from the expected durations. Patients are assumed to be punctual to their appointments. However, some patients may miss their appointments. To compensate for no-shows, the clinic either double book (by scheduling two patients to the same slot) or adjust appointment intervals according to the no-show rate. On the day of the appointment, the patients are seen in the order in which they are scheduled and arriving patients should not "jump the queue" or preempt patients that are already being served. All scheduled patients should be served, and the clinic will work overtime to serve patients who are still waiting for treatment at the end of a clinic session.

This chapter aims to integrate patient-specific uncertainty estimates into the AS design and assess its impact on AS efficiency (measured as the weighted sum of patient waiting time, doctor idle time, and overtime). Specifically, the patients are scheduled as they sequentially call the clinic for appointments based on their risk of no-show and estimated consultation duration. To this end, this chapter proposes a predict-then-schedule framework, a prescriptive analytics approach for designing the AS. In the predict step, patient-specific

no-show risk and service duration are estimated using machine learning (ML) models. The schedule step determines each patient's appointment time and interval by integrating the ML predictions with the three decisions (allocation, sequencing, and overbooking) pertaining to the AS design. This chapter also proposes new sequencing rules that use the predicted appointment duration and patient classification for scheduling. In particular, the key research questions addressed in this chapter are as follows.

- How to use the ML-based patient-specific prediction of clinical uncertainty to design an AS? In other words, how can these predictions be used for decisions pertaining to appointment duration allocation and overbooking?
- Can patient-specific prediction be used for effectively sequencing the patients?
- What is the impact of integrating the ML predictions of clinical uncertainties into AS design?

4.1 Research methodology

An overview of the predict-then-schedule research framework is shown in Figure 1. The "predict" phase employs ML algorithms to estimate patient-specific consultation lengths and their associated risk of no-show based on clinical data obtained from the electronic medical records (EMR). In the "schedule" phase, the predictions obtained from ML algorithms are integrated into three decisions associated with the AS and evaluated to assess its performance.

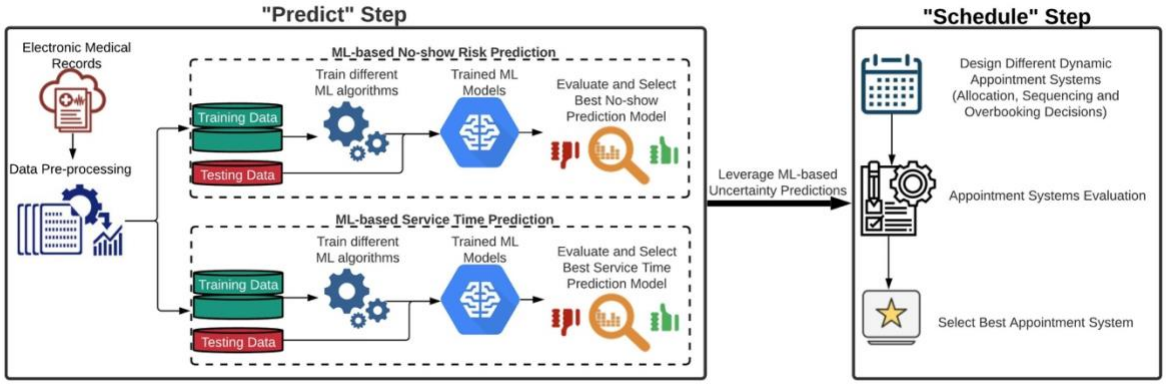

Figure 7: Overview of the predict-then schedule framework

4.1.1 Predicting patient-specific no-show and service duration

4.1.1.1 Data collection and feature selection

The data used in this research were obtained from the cardiology clinic at the University of Missouri Hospital. Two years of EMR containing a total of 30,000 patient visits are extracted for developing the ML algorithm. We identified 16 predictors that fall into four categories, namely, patient demographics, appointment detail, medical history, and patient behavior, as shown in Table 5. Some of the raw EMR variables, such as age and gender, are directly used as predictors in the ML algorithm. On the other hand, certain features were derived based on information extracted from the EMR. For instance, the timestamp data (hh:mm:ss mm/dd/yyyy) on the appointment date is transformed into three features – month, weekday and visit hour. Besides the predictors, the outcome variables, namely, appointment status (arrived or no-show) and appointment duration, are also obtained from the EMR.

Table 4: The description of features extracted from the EMR

Feature Category	**Feature**	**Description**	**Type**
Patient Demographic	Sex	Sex of the patient	Categorical
	Age	Age of the patient (in years)	Continuous
	Marital status	Classification of patient's legal marital status	Categorical
	Employment status	Patient's employment status at time of the	Categorical
	Insurance group	Primary payer for patient's medical claims	Categorical
	Primary language	Patient's preferred language of communication	Categorical
	Traveling distance	Traveling distance (in miles) from patient's residence	Continuous
Appointment Detail	Month	Month of appointment	Categorical
	Weekday	Day of the week on which the patient visits the clinic	Categorical
	Visit Hour	Clinical hour during which the appointment is	Categorical
	Care Delivery	Nature of patient's engagement with the care	Categorical
	Service Type	Classification of service provided during the visit	Categorical
Medical History	Patient status	Classification of patient's status based on their visit	Categorical
	Chronic Condition	Total number of chronic conditions diagnosed for the	Continuous
Patient Behavior	Service time history	Rolling average of patient's consultation duration	Continuous
	No-show history	Total number of appointments missed by the patient	Continuous

4.1.1.2 ML algorithm development and evaluation

The goal of a predictive algorithm is to best approximate the mapping function (*F*) from a set of *M* input features**,** $\mathbf{X} = (x_1, x_2, \ldots\ldots x_M)$, and an outcome variable (Y) based on *N* training samples [143]. Typically, each ML algorithm will adopt a unique approach to learn the mapping function between the predictors and an outcome variable [144]. Consequently, this will affect the performance and training time of the ML models. Hence, it is difficult to determine an appropriate ML model for a given dataset. To overcome this issue, prior research evaluates multiple ML algorithms and identifies the best-performing model by a comparative analysis [49,145] . In particular, random forests, boosted trees, and deep feed-forward neural networks have proven to perform well in predicting uncertainties related to outpatient scheduling, such as no-shows [143], patients' punctuality [59] , and consultation length [145]. Besides, these ML algorithms are well-suited for predicting continuous values (e.g., appointment duration) and discrete categories (e.g., no-show classification). Therefore, in this study, we implemented extreme gradient boosting (XGB), random forest (RF), and deep feed-forward neural networks (DNN) to predict both no-show classification and appointment duration. The learning approaches of the three ML algorithms are follows:

- Random Forest (RF): It is an ensemble of uncorrelated decision trees, where each tree is trained based on two concepts – bagging and random feature selection [144]. The bagging technique samples a subset of the training data with replacement and uses it to train a decision tree, while the random feature selection technique only considers a subset of predictors for splitting at every node of a decision tree. The combination of

these two techniques reduces the risk of multiple decision trees learning the same information from the training. Upon independently training multiple decision trees in parallel, the final prediction is obtained by aggregating the individual predictions of these trees (in the case of continuous outcome variable) and the majority vote of all trees (for the categorical outcome variable).

- Extreme Gradient Boosting (XGB): It also uses a multitude of decision trees for learning the mapping function between predictors and outcome [146] . However, unlike RF, it builds the decision trees in a stage-wise fashion, where the subsequent tree seeks to boost the performance over the previous tree's prediction by emphasizing on samples with high loss function (sum of squared errors in the case of prediction or the cross-entropy in case of classification). In particular, XGB uses second-order derivatives of the loss function, and advanced regularization techniques to prevent overfitting.

- Deep Feedforward Neural Networks (DNN): It is a type of deep learning model, where several nodes arranged as a linear stack of three different layer types, input, hidden, and output, are used to best approximate the mapping function. The nodes in adjacent layers are interconnected by weighted edges. Unlike a simple neural network with one hidden layer, a DNN has many hidden layers to facilitate the learning of complex functions and better prediction. The input layer nodes receive information from the predictors and transmit the weighted input to the nodes in the adjacent hidden layer, which then transforms it using an activation function, to account for non-linear relationships, before feeding that information to the subsequent hidden or output layer. This process,

known as forward pass, continues until the information from the last hidden layer is processed and transformed by the output layer nodes to make a prediction/classification. The training data is fed into the DNN in small batches, where a batch contains one or more training records for learning purposes. The connection weights between nodes of DNN are initialized randomly and progressively altered for every successive batch based on the difference between the actual and predicted output - referred to as backward pass. The forward and backward passes are repeated for several iterations to minimize the loss function.

During the predictive modeling phase, the data is randomly split into training and testing sets. The ML algorithm establishes the mapping function between the outcome variable and the predictors using the training data, while the trained model's ability to predict the outcome variable is evaluated using testing data. Moreover, to avoid overfitting (learning noise) in the training phase, a *k*-fold cross-validation technique is performed [147,148].

To evaluate the ML algorithm's performance of predicting consultation duration, we employed two commonly used error measures in the literature [149] – root mean square error (RMSE) and mean absolute percentage error (MAPE). Lower values are preferred for both measures since they indicate lower prediction error. Similarly, based on the review of literature, the following two metrics are used to evaluate the no-show classifiers [150] - area under the receiver operating characteristic curve (AUC) and misclassification rate (MCR). The values of these metrics range from 0-1, where higher AUC and lower MCR values are preferred. After training and evaluating the ML algorithms, the predictions from

the best performing no-show classification and appointment duration prediction models are leveraged to develop the AS.

4.1.2 Appointment System (AS) Design

In this research, the AS is designed by considering the key decision factors – allocation, sequencing, and no-show adjustment (i.e., no-show estimation and overbooking policy). The levels considered for each decision factor are described in the following subsections.

4.1.2.1 Allocation decisions

The components associated with the allocation decisions, namely, block size and appointment interval estimation, are determined based on the review of relevant literature [9,42]. Two block sizes are considered, namely, individual (i.e., one patient per slot) and variable (i.e., varying number of patients per slot). The block size of the AS is directly influenced by the overbooking policy proposed in this research, as discussed in Section 3.2.4. With regard to appointment interval estimation, we consider two strategies. The first strategy is akin to the approaches adopted in the literature and practice, where the appointment interval allocated to a patient is equal to the average service time of that patient's class (represented as AD^{μ}) [12,24,25]. The second strategy is the proposed approach of using the ML-based predictions, where patient-specific service time duration obtained from the best performing ML algorithm is used to allocate the appointment duration (denoted as AD^{ML}). In other words, we consider the heterogeneity present within each patient class by allowing service time values to be patient-specific. In both cases,

appointment intervals are considered variable since they vary depending on the patient class or ML predictions. Therefore, variable intervals are used to allocate patients to their appointment slots.

4.1.2.2 Sequencing rules

In this study, we consider two patient classes (Class A and Class B) for the following reasons: (i) the clinic under study categorizes patients based on their visit status as new or return for the purpose of scheduling, (ii) several prior works have used two patient classes and have reported them to be common in practice [12,24,25] (iii) considering many patient classes may hinder the practical/realistic application of the AS [42]. Class A patients are characterized by higher average service times than Class B patients. Such a classification scheme is chosen based on the empirical data observed in different specialties. For instance, patients who are visiting the clinic for the first time (new patients) can be regarded as Class A as they typically require longer consultations compared to patients returning to the clinic for follow-up appointments [151]. In this research, the following five sequencing rules that are commonly considered in the literature are evaluated [13,24]. Except for FCFA, the other rules are based on the two patient classes.

i. **FCFA**: represents the base experiment settings with no sequence. Patients are scheduled for an appointment based on a first-call, first-appointment basis, and this is similar to the clinic's existing approach.

ii. $\mathbf{A^{BG}}$: schedules Class A patients to the first unscheduled slot at the beginning of the session and Class B patients from the end (AAA. . . .BB).

iii. $\mathbf{B^{BG}}$: schedules Class B patients to the first available slot at the beginning of the session, while the patient class requiring longer service times are scheduled to the first feasible slot in the end (BBB. . . AA).

iv. $\mathbf{A^{BND}}$: schedules Class A patients at the beginning and end of the session, while class B patients are scheduled in the remaining positions (AAA...BB...A).

v. $\mathbf{B^{BND}}$: schedules Class B patients at the beginning and end of the session and class A patients in the remaining positions (BBB. . . AA...BB).

4.1.2.3 New Sequencing Rules based on ML Predictions

Some existing works have shown that sequencing patients with shorter and longer duration split across the clinic session to yield better resource utilization and patient waiting time [13,50,51]. However, these rules either rely on patient class characteristics (e.g., average and standard deviation of service time associated with a class) [50,51] or the experience of schedulers to determine the appointment time and duration for each patient [13]. In other words, existing sequencing rules do not consider patient-specific characteristics. Motivated by these findings and shortcomings in prior literature, this research proposes four sequencing rules that capitalize on ML-based patient-specific service time prediction to

sequentially schedule patients such that shorter and longer appointments are distributed across the session. Specifically, the proposed rules compare the patient's predicted service time with the average consultation duration corresponding to that patient's class to sequence them. Let D_i^A and D_j^B be the ML-based appointment duration prediction for patient i of Class A and patient j of Class B, respectively. If Class A and Class B patients have an average service time of μ^A and μ^B (where $\mu^A > \mu^B$), respectively, then the proposed sequencing rules can be described as follows.

i. $\mathbf{S_1^{ML}}$: schedule patient *i* of Class A in the beginning if $D_i^A < \mu^A$ and patient *j* of Class B in the beginning if $D_j^B > \mu^B$. Otherwise, the patient is scheduled from the end of the session.

ii. $\mathbf{S_2^{ML}}$**:** schedule the patient to the first available position in the beginning if $D_i^A > \mu^A$ or $D_j^B < \mu^B$; else schedule the patient to the first available position in the end.

iii. $\mathbf{S_3^{ML}}$: schedule the patients by alternating between the first available slot in the beginning and end of the session if $D_i^A > \mu^A$ or $D_j^B < \mu^B$. Otherwise, start scheduling from the middle of the clinic session while distributing the patient allocations between the two halves of the session. For example, if a clinic operates for 8 hours (i.e., 480 minutes) and an incoming appointment request has a predicted service time of $D_i^A < \mu^A$ or $D_j^B > \mu^B$, then the search for available slots starts in the middle (i.e., at 240 minutes), and then alternates between the left and right halves until the first feasible slot is found.

iv. $\mathbf{S_4^{ML}}$: schedules patients who have a predicted service time $D_i^A < \mu^A$ or $D_j^B > \mu^B$ at the beginning and end of the session; and other are patients scheduled from the middle of the clinic session. In other words, this sequencing rule is antithetical to S_3^{ML}.

4.1.2.4 No-show estimation and overbooking policies

We consider two approaches to estimate the no-show probability for each patient at the time of appointment request. In the first approach, the probability of a patient missing an appointment is considered homogeneous and is predicted to be equal to the average no-show rate of the clinic (denoted as NS^{μ}). For instance, if the clinic has an average no-show rate of 15%, then the no-show probability for each patient is set to 0.15. This is the most common approach considered in the literature [25,152] and is also similar to the practice adopted by the clinic under study. In the second case, the no-show probability is regarded as patient-specific (heterogeneous) and is estimated using the best performing no-show prediction algorithm (denoted as NS^{ML}). Thus, the second approach is a part of the proposed predict-then-schedule framework. The estimated no-show rates are then used for overbooking decisions.

To investigate the effect of the no-show adjustment on the clinic's performance, two overbooking policies are considered – double-booking and appointment duration shortening. Since the estimated no-show rates are used for both policies, we refer to these

policies as predictive double-booking (PD) and predictive shortening (PS). If $P(p_i)$ and $P(p_j)$ denote the probabilities of patients i and j attending the appointment, then the proposed PD policy schedules two patients in the same slot only when the following conditions are met: (i) the probability of at least one patient showing up for the appointment is greater than a threshold value β_1 (Equation 1) and (ii) the probability of both patients not showing up is greater than the threshold value β_2 (Equation 2). Moreover, a patient is double-booked only when a feasible empty slot is not available for that patient.

$$P(\text{at least one patient showing up }) = \left[P(p_i) + P\left(p_j\right) - P(p_i) * P\left(p_j\right)\right] > \beta_1 \quad (1)$$

$$P(\text{none of patients showing up }) = \left[(1 - P(p_i)) * (1 - P\left(p_j\right))\right] < \beta_2 \quad (2)$$

On the other hand, the PS policy compresses the appointment length proportional to the estimated no-show rate. In other words, the appointment duration for each patient is calculated as the product of that patient's estimated consultation length and no-show probability. Note that the overbooking policy has a direct impact on the block size. A PD policy allows more than one patient per slot, thereby resulting in a multi-block allocation system; whereas the PS policy restricts each slot to exactly one patient and can be regarded as an individual block system.

4.1.3 Performance measures

The key measures included in this study are patients' average waiting time (*PWT*), doctor's idle time per patient (*DIT*), and overtime per patient (*DOT*). Since we only consider punctual patients, the direct waiting time for a patient is defined as the positive difference between actual start time and appointment time (the same as the patient's arrival time). The doctor is considered idle if the actual service end time in an appointment slot is earlier than the expected end time. On the other hand, provider overtime occurs when the actual session completion time exceeds the scheduled end time. The total cost provides a trade-off among the key measures (*PWT*, *DIT*, *DOT*) since it allows the decision-makers to choose the importance of each primary measure based on the relative valuation given to the cost of patients' time (C^P) and doctor's time (C^D). Therefore, similar to prior works, this study also uses the total cost per patient (TC), a weighted sum of patient's and doctor's time, to evaluate the schedule efficiency [24,25]. The doctor's idle time and overtime are combined into one "doctor-related" measure, where the cost of doctor overtime is deemed to be higher than the idle time by π percent. Thus, the TC for an appointment system can be calculated as shown in Equation 3.

$$TC = C^P \times PWT + C^D([DIT + \pi \times DOT]) \quad (3)$$

4.1.4 Simulation of Appointment System

An AS design is composed of allocation, sequencing and no-show adjustment decisions, as discussed in Section 4.1.2. For instance, a clinic's AS design may adopt A^{BG} rule for sequencing, ML-based approach for appointment duration (AD^{ML}) and no-show estimation (NS^{ML}), and predictive double-booking (PD) as an overbooking strategy. Such an AS design is represented as A^{BG}-AD^{ML}-NS^{ML}-PD in this research, and depending on the levels described in Section 4.1.2, a total of 40 AS designs can be generated. A simulation model is developed to generate the schedule for a given AS design by considering a sequential patient call-in process. To simulate the scheduling process, the daily operating duration is partitioned into *k*-minute intervals called blocks (or slots). In addition, the following assumptions are used to model the scheduling process:

- Patients are seen in the order in which they are scheduled and arriving patients should not "jump the queue" or preempt patients already being served.

- All scheduled patients who show up must be served. The clinic will work overtime to serve patients who are still waiting for treatment at the end of a clinic session.

- Patients are punctual to their appointments if they show up.

- The estimated appointment duration using is rounded to the nearest k^{th} value, where *k* denotes the slot duration.

- A patient is assigned to a set of one or more continuous slots depending on their estimated appointment duration (either using AD^{μ} or AD^{ML} approach). For example, if a clinic divides the operating hours into 5-minute slots, and if the estimated service time for a patient is 20 minutes, then that patient is allocated 4 continuous slots.

The procedure for schedule generation is given by Algorithm 1. Upon obtaining the inputs necessary to design the AS, a set of patient calls ($p \in P$) along with different call-in sequences ($q \in \mathcal{Q}$) are randomly generated to simulate the sequential arrival of appointment requests (lines 1-3). The non-attendance probability and service time duration of patient p is determined (lines 6-7) based on the approach chosen for no-show (NS^{μ} or NS^{ML}) and appointment duration estimation (AD^{μ} or AD^{ML}). Subsequently, the number of slots required for patient p as well as the number of continuously available empty slots and single-booked slots are calculated (lines 8-9). Patient p is then either scheduled according to the chosen overbooking policy $\mathcal{O}$ and sequencing rule $\mathcal{R}$, or denied appointment for the day under consideration (lines 10-16). This procedure is repeated for all patient calls, $p \in P$, in sequence q in a sequential manner to construct the full schedule ψ_q (lines 5-18). To account for the impact of sequential appointment requests on the AS efficiency, the schedule generation procedure is replicated for $|\mathcal{Q}|$ different patient call-in sequences (lines 4-20).

Algorithm 1 Pseudocode for Schedule Construction

1: **Inputs**: sequencing rule ($\mathcal{R}$), no-show estimation approach ($\mathcal{N}$), overbooking policy ($\mathcal{O}$) and appointment duration estimation approach ($\mathcal{D}$), number of slots ($s \in \mathcal{S}$), slot duration (k)
2: Randomly generate a set of $p \in \mathcal{P}$ patients calls
3: Randomly generate $q \in \mathcal{Q}$ call-in sequences of the $|\mathcal{P}|$ patient calls
4: **for** each call-in sequence $q \in \mathcal{Q}$ **do**
5: **for** $p \in \mathcal{P}$ in sequence q **do**
6: Obtain no-show probability (η_p) for patient p based on $\mathcal{N}$
7: Obtain appointment duration for patient p (d_p) based on $\mathcal{D}$
8: Determine number of required slots for patient p as $r_p = \left\lfloor \frac{d_p}{k} \right\rceil$
9: Calculate number of consecutive empty (n_0) and single-booked slots (n_1)
10: **if** $n_0 \geq r_p$ **then**
11: Schedule patient p to r_p continuous slots according to selected sequencing rule $\mathcal{R}$
12: **else if** $\mathcal{O} == PD$ and $n_1 \geq r_p$ **then**
13: Double-book patient p to r_p continuous slots according to selected sequencing rule $\mathcal{R}$
14: **else**
15: Patient p cannot be scheduled for the day under consideration
16: **end if**
17: **end for**
18: Store schedule generated for sequence q (ψ_q)
19: **end for**
20: **return** $\psi_q, \forall q \in \mathcal{Q}$

The pseudocode for determining the AS efficiency based on the generated schedule is given by Algorithm 2. The key input parameters (line 1) for evaluating the AS are the schedule generated using Algorithm 1 for all call-in sequences, slot duration (k), total number of slots ($|S|$) and parameters related to TC (i.e., C^P, C^D, π). The total patients scheduled (P^{TOTAL}) as well as the appointment start and completion time (E_p^S and E_p^C) of each patient are obtained from the schedule generated using Algorithm 1 (lines 3-5). To simulate the patient arrival and consultation process, the actual no-show status (θ_p) and service time (τ_p) of each scheduled patient are determined (lines 6-7). The actual no-show status (arrived or missed appointment) can be obtained from the observed EMR data. Alternatively, similar to prior works, θ_p can be simulated using Bernoulli distribution based on the clinic's average no-show rate [49]. Likewise, τ_p can be retrieved from EMR or simulated based on a theoretical distribution [24,49]. In this research, we use both

approaches for obtaining θ_p and τ_p . The EMR data obtained from the cardiology clinic is used to extract the observed service time and no-show status and evaluate the generated schedule. In addition, to test the impact of the proposed approach on different clinic environments, the values of θ_p and τ_p are also simulated from a theoretical distribution. These values are then utilized to determine the actual service start time and service completion for each scheduled patient (lines 8-9). Subsequently, the performance measures for the generated schedule, ψ_q, are computed (lines 12-15), and this process is replicated for the $|\mathcal{Q}|$ schedules to obtain the TC associated with the AS design (lines 2-18).

Algorithm 2 Pseudocode for Schedule Evaluation

1: **Inputs:** Obtain $\psi_q, \forall q \in \mathcal{Q}$, k, $|\mathcal{S}|$, C^P, $C^D d$, and π
2: **for** each call-in sequence $q \in \mathcal{Q}$ **do**
3: Determine total number of patients scheduled in q (P^{TOTAL})
4: **for** each patient $p \in \psi_q$ **do**
5: Extract appointment start and end time for patient p (E_p^S and E_p^C)
6: Determine actual service time duration for patient p (τ_p)
7: Determine whether patient p actually missed the appointment ($\theta_p = 0$ for no-show and 1 otherwise)
8: Compute actual service start time for patient p, A_p^S. Note $A_p^S = 0$ for the first patient and completion time of previously scheduled patient ($A_{[p]-1}^C$) in all other cases.
9: Compute actual service completion time of patient p, $A_p^C = A_p^S + \tau_p \times \theta_p$
10: Compute waiting time for patient p, $WT_p = \max(0, A_p^S - E_p^S)$
11: **end for**
12: Calculate average patient waiting time for ψ_q, $PWT = \dfrac{\sum_{p \in \psi_q} WT_p}{P^{TOTAL}}$
13: Calculate doctor idle time per patient, $DIT = \dfrac{\max(0, k \times |\mathcal{S}| - \sum_{p \in \psi_q} \tau_p \times \theta_p)}{P^{TOTAL}}$
14: Calculate doctor overtime per patient, $DOT = \dfrac{\max(0, \sum_{p \in \psi_q} \tau_p \times \theta_p - k \times |\mathcal{S}|)}{P^{TOTAL}}$
15: Compute the total cost for schedule ψ_q, $TC(\psi_q) = C_p \times PWT + C_d([DIT + \pi \times DOT])$
16: **end for**
17: Compute average total cost across all $|\mathcal{Q}|$ schedules, $TC = \dfrac{\sum_{q \in \mathcal{Q}} TC(\psi_q)}{|\mathcal{Q}|}$
18: **return** (TC)

4.2 Results and Analysis

In this section, the proposed predict-then-schedule framework is evaluated using the EMR data obtained from the cardiology clinic at the University of Missouri Hospital. In addition, the impact of integrating ML predictions into the AS design is also assessed for several simulated clinical environments. All the models were developed and implemented on a computer configured with Intel Core i7 3.4GHz processor, macOS Sierra operating system, and 32 GB RAM.

4.2.1 Case Study to Evaluate the Predict-then-Schedule Framework

4.2.1.1 Case Study Background

The cardiology clinic under study operates for 8 hours a day and divides the clinic session into 5-minute slots. Patient calls arrive sequentially and average to 50 calls per day. The chi-square test (at $\alpha = 0.05$) and probability plots of the patient call volume data indicate Poisson distribution to be a good fit, as shown in Figure 2. The patient population is categorized into two classes – Class A (new) and Class B (return) patients. Among the total patient calls, 35% of patients are determined as Class A based on the analysis of historical data. With regard to the service time, the Anderson-Darling test (at $\alpha = 0.05$) suggests normal distribution with parameters $(30, 10^2)$ and $(15, 5^2)$ to be an appropriate fit for Class A and Class B patients, respectively. Figure 3 shows the probability plots for both the classes and the corresponding p-values. However, to avoid unrealistic service times (such as negative or very high positive values), the associated data for Class A and Class B patients were truncated on both sides with bounds (15, 75) and (5, 45), respectively. The

clinic sequences patients on an FCFA basis and the allocated appointment duration for each patient is equal to the average service time of that patient's class. Moreover, to maintain continuity of care, the clinic typically allows a provider to serve only his/her own patients. Therefore, similar to prior works, this clinic can be represented as a single-stage single-server system [18,153]. The clinic experiences an average no-show rate of 30% and double-books proportional to it to compensate for missed appointments. In summary, the AS design adopted by the clinic can be represented as FCFA-AD$^{\mu}$-NS$^{\mu}$-PD.

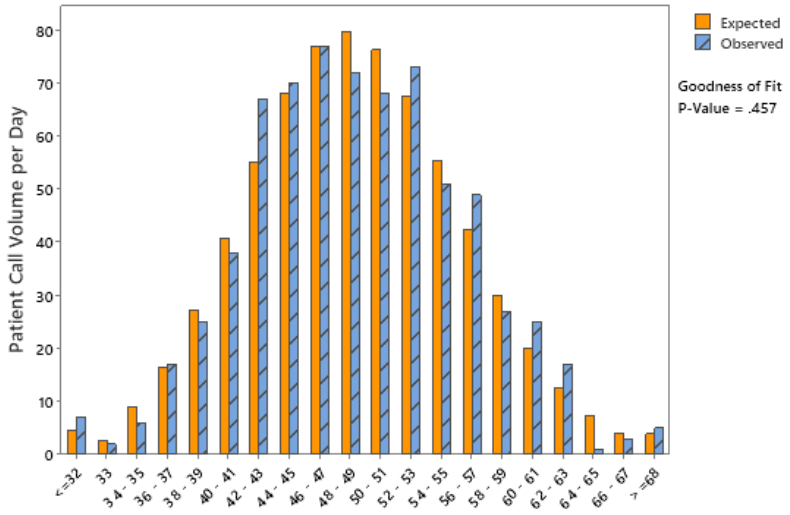

Figure 8: Probability plot for patient call volume per day

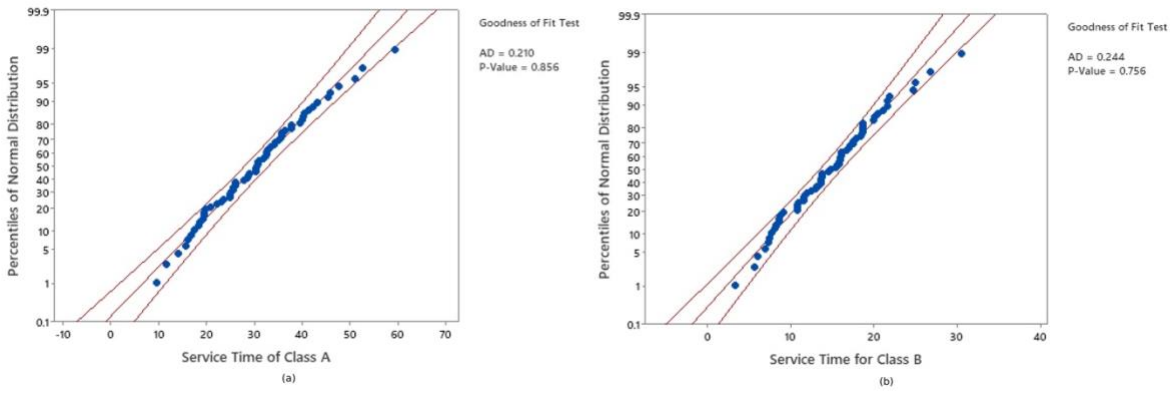

Figure 9: Probability plots for consultation duration of Class A and Class B patients

4.2.1.2 Performance of ML algorithms for no-show and appointment duration prediction

We used the consolidated EMR dataset to develop the ML models. Each visit included the 16 features (as shown in Table 1) and a corresponding outcome variable. To develop a generalized ML model, the dataset is randomly split into two parts – 70% for training and 30% for testing. Besides, a 10-fold cross-validation is applied to the training dataset to avoid overfitting. The hyperparameters of the three ML models under consideration (RF, XGB, DNN) are tuned using the grid search. Tables 6 and 7 show the performance of the three algorithms for no-show classification and appointment duration prediction, respectively. Specifically, the three algorithms achieve comparable performance for both no-show and service time prediction, with DNN achieving superior predictions than RF and XGB. Besides, the performance of ML algorithms is similar on both cross-validation and testing datasets, thereby suggesting the generalization ability of the ML models to predict new unseen instances with similar predictive performance.

Table 5: Performance of ML algorithms for no-show classification on cross-validation and testing dataset

ML Algorithm	Cross-Validation Dataset		Testing Dataset	
	AUC	**MCR**	**AUC**	**MCR**
RF	0.867	12.3	0.852	13.2
XGB	0.894	11.3	0.883	11.9
DNN	0.926	9.9	0.901	10.1

Table 6: Performance of ML algorithms for service time prediction on cross-validation and testing dataset

ML Algorithm	Cross-Validation Dataset		Testing Dataset	
	RMSE	**MAPE**	**RMSE**	**MAPE**
RF	7.86	10.89	8.36	11.14
XGB	7.00	10.33	7.58	10.72
DNN	6.65	9.75	7.09	10.20

4.2.1.3 Impact of integrating ML-based predictions on AS performance

The efficiency of different AS designs for the cardiology clinic under study is evaluated by simulating the arrival and service process, as discussed in Section 4.1.4. The simulation model is replicated for 100 clinical days by generating the same patient call-in sequence observed in the EMR. The best-performing ML algorithm (i.e., DNN) is used to obtain the no-show probability and service time duration for each call-in. To evaluate the schedule, the actual service time and no-show status for each patient call-in is retrieved from the EMR data. The thresholds β_1 and β_2 for PD were set as 0.8 and 0.5, respectively, based on empirical analysis. To estimate C_p and C_d costs, the relative importance between these costs C_d/C_p) is used and is set to 1. In addition, consistent with previous works, the doctor's overtime was considered as 50% higher than the idle time (i.e., π=1.5) to calculate the total cost [24,25].

Figure 4 compares the performance of the AS designs ranging from no ML-based integration (AD^{μ}-NS^{μ}) to integrating both no-show and appointment duration predictions (AD^{ML}-NS^{ML}). It can be observed that irrespective of the sequencing rules and overbooking policies, integrating ML-based prediction even for one type of clinical uncertainty (i.e., no-show or appointment duration) substantially reduces the average total cost associated with the AS design. Specifically, for the PD overbooking policy, the integration of only patient-specific appointment duration predictions (AD^{ML}-NS^{μ}) yielded a 30% average improvement in AS efficiency across all sequencing rules when compared to using class-specific average service time. On the other hand, leveraging only the ML-based no-show

predictions (AD^{μ}-NS^{ML}) in the AS design achieved about 51% reduction in total cost across all sequencing rules as opposed to the common approach of relying on the clinic's average no-show rate. When both no-show and appointment duration predictions were used to design the AS, the improvement attained was over 68%. In the case of the PS overbooking policy, the trends in the performance were similar to the PD policy, but the improvement gap was smaller. In particular, AD^{ML}-NS^{μ}, AD^{μ}-NS^{ML}, and AD^{ML}-NS^{ML} achieved 16%, 30%, and 48% improvement, respectively, over AS design with no ML-based integration. Thus, the proposed predict-then-schedule framework can significantly improve the performance achieved by the clinic's existing AS design (FCFA-AD^{μ}-NS^{μ}- PD) as well as other existing AS designs typically adopted in the literature. With respect to sequencing rules, FCFA resulted in the highest total cost, while B^{BND} contributed to lower costs in all AS designs, irrespective of the overbooking policies. Besides, double-booking (PD policy) showed better performance than shortening the appointment intervals (PS policy).

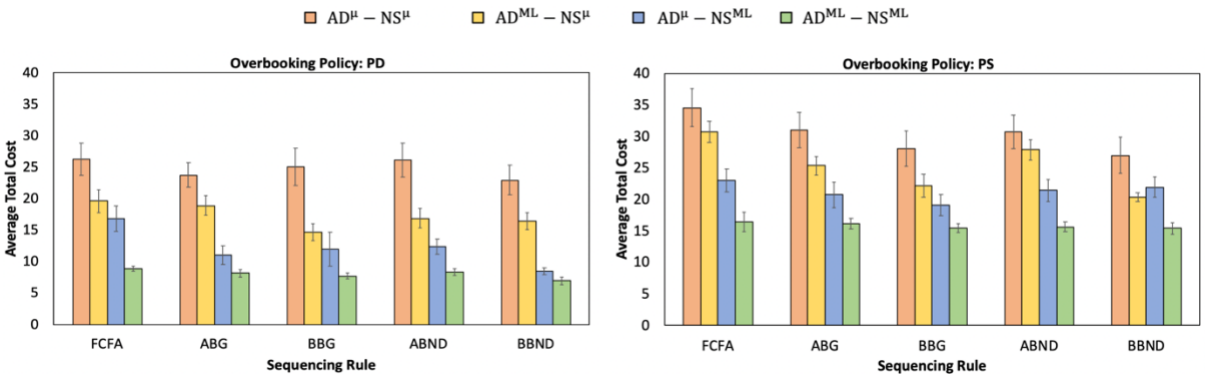

Figure 10: Impact of integrating ML-based prediction on total cost

4.2.1.4 Impact of new sequencing rules on AS performance

To assess the performance of the proposed ML-based sequencing rules, we benchmarked it to the existing sequencing rule that achieved the lowest cost (i.e., B^{BND}). Moreover, the appointment duration and no-show predictions from ML were used for both the proposed and B^{BND} sequencing rule since they yield the best performance. Figure 5 shows the reduction in total cost achieved by the proposed sequencing rules. It can be observed that all four proposed rules were able to improve the efficiency of AS design, irrespective of the overbooking policy. The S_3^{ML} rule resulted in the highest improvement of 9.5% and 7.5% for PD and PS policies, respectively. On the other hand, the benefit of adopting S_2^{ML} and S_4^{ML} were marginal but still would yield substantial cost savings when adopted over a longer time period.

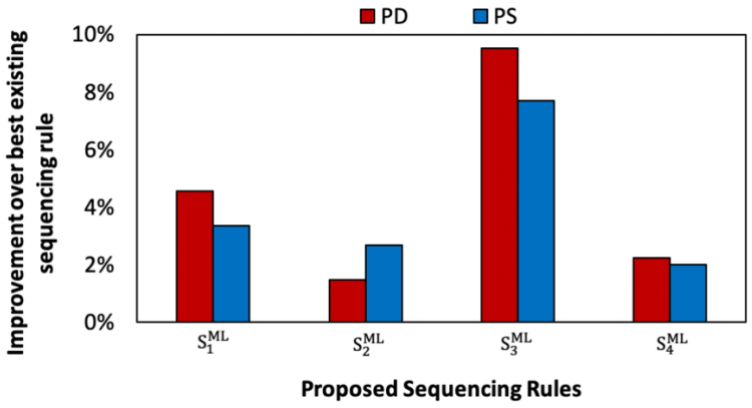

Figure 11: Improvement achieved by proposed ML-based sequencing rule

4.2.2 Impact of using patient-specific predictions for different clinical environments

In this section, the robustness of the proposed framework and sequencing rules is evaluated by considering different simulated clinical environments. To construct realistic clinical environments, we consider five factors - clinic's average no-show rate (*PN*), coefficient of service time variation for Class A patients (*CV-A*), coefficient of service time variation for Class B patients (*CV-B*), percent of Class A requests (λ^A), and the ratio of the mean consultation time of Class A patients to the mean consultation time of Class B patients (μ^A/μ^B). A clinic environment represents a combination of all aforementioned factors and is consistent with the factors considered in the literature [24,25]. We examine two levels (low and high) for each of the five environmental factors, as shown in Table 7. These levels are selected based on the values observed in the EMR data as well as prior literature [25]. Thus, a total of 32 (or 2^5) clinic environments are generated based on the two levels of the five environmental variables. Furthermore, to account for the influence of the relative importance of doctors' time to patients' time on AS efficiency, the following values of C_d/C_p are considered to determine the TC of each clinic environment - 1,2, 5. We used the EMR data as the baseline to establish the key parameters needed to simulate the different clinic environments. We leveraged the same mean and distribution to generate service time values for the different clinic environments but varied the standard deviation to achieve the desired CV value under consideration (i.e., 0.35 or 0.7). The no-show status for a patient is simulated as Bernoulli distribution, where the probability of that patient missing an appointment is set based on the clinic environment (either 0.3 or 0.5).

Table 7: Environmental factors and associated levels

Environmental Factors	Levels	Settings
PN	2	0.3, 0.5
CV- A	2	0.35, 0.7
CV- B	2	0.35, 0.7
$\boldsymbol{\mu_A/\mu_B}$	2	1.33, 2.00
$\boldsymbol{\lambda^A}$	2	35%, 50%

Table 8 presents the configuration of the best AS and the corresponding total cost for each clinic environment under consideration. It can be observed that all the AS designs that yield the lowest cost have employed ML-based predictions of both appointment duration and no-shows (AD^{ML}-NS^{ML}), thereby indicating the dominance of the proposed approach. For all the cost ratios under consideration, the B^{BND}-AL^{ML}-NS^{ML}-PD is most frequently ranked (over 30% of the environments considered) as the best AS configuration, followed by B^{BG}-AL^{ML}-NS^{ML}-PD configuration (over 20% of cases). Notably, the PD approach is observed to be the predominant choice of overbooking strategy in the best AS design, as PS policy results in the least cost only in three instances. Besides, the FCFA sequencing rule is never adopted in any of the best AS designs. To evaluate the impact of adopting the proposed sequencing rules (S_1^{ML} – S_5^{ML}) on different clinic environments, we benchmark its performance against the best AS in each clinic environment (i.e., Table 5). The improvement achieved by adopting the proposed sequencing rule ranged from 1% - 40.8% when compared to the best-performing AS that uses existing sequencing rules, as shown

in Table 10. The S_3^{ML}-AL$^{\mathrm{ML}}$-NS$^{\mathrm{ML}}$-PD configuration achieves the best performance in most cases. Besides, the AS based on the proposed sequencing rule yields the lowest cost in over 40% of the cases for C_d/C_p of 1. This value increases to 59% and 81% for C_d/C_p of 2 and 5, respectively. In other words, the number of environments dominated by the proposed rule increases as the cost associated with doctor's time is increased. This may indicate that these new rules are targeted more towards minimizing doctor's time as opposed to patient's waiting time. Thus, these findings demonstrate the significance of adopting ML predictions for sequencing patients.

Table 8: Best appointment system and corresponding total cost for different clinic environments

Clinic Environment	λ^A	$\frac{\mu_A}{\mu_B}$	PN	CV-A	CV-B	Best AS Design (TC)		
						$\frac{C_d}{C_p} = 1$	$\frac{C_d}{C_p} = 2$	$\frac{C_d}{C_p} = 5$
1	35	2	30	0.35	0.35	B^{BND}-AL^{ML}-NS^{ML}-PD (7.02)	B^{BND}-AL^{ML}-NS^{ML}-PD (12.55)	B^{BND}-AL^{ML}-NS^{ML}-PD (29.14)
2	35	2	30	0.35	0.70	B^{BG}-AL^{ML}-NS^{ML}-PD (7.71)	A^{BND}-AL^{ML}-NS^{ML}-PD (12.40)	A^{BND}-AL^{ML}-NS^{ML}-PD (24.38)
3	35	2	30	0.70	0.35	B^{BND}-AL^{ML}-NS^{ML}-PD (8.41)	A^{BND}-AL^{ML}-NS^{ML}-PD (16.03)	A^{BND}-AL^{ML}-NS^{ML}-PD (31.39)
4	35	2	30	0.70	0.70	B^{BND}-AL^{ML}-NS^{ML}-PD (8.12)	A^{BND}-AL^{ML}-NS^{ML}-PD (14.97)	B^{BG}-AL^{ML}-NS^{ML}-PD (34.17)
5	35	2	50	0.35	0.35	B^{BG}-AL^{ML}-NS^{ML}-PD (5.78)	B^{BG}-AL^{ML}-NS^{ML}-PD (10.59	B^{BG}-AL^{ML}-NS^{ML}-PD (25.00)
6	35	2	50	0.35	0.70	A^{BG}-AL^{ML}-NS^{ML}-PD (6.09)	A^{BG}-AL^{ML}-NS^{ML}-PD (6.09)	A^{BND}-AL^{ML}-NS^{ML}-PD (25.18)
7	35	2	50	0.70	0.35	B^{BND}-AL^{ML}-NS^{ML}-PD (8.45)	B^{BND}-AL^{ML}-NS^{ML}-PD (16.35)	B^{BND}-AL^{ML}-NS^{ML}-PD (34.65)
8	35	2	50	0.70	0.70	B^{BND}-AL^{ML}-NS^{ML}-PD (7.54)	B^{BND}-AL^{ML}-NS^{ML}-PD (14.17)	B^{BND}-AL^{ML}-NS^{ML}-PD (30.49)
9	50	2	30	0.35	0.35	B^{BG}-AL^{ML}-NS^{ML}-PD (6.30)	B^{BG}-AL^{ML}-NS^{ML}-PD (11.10)	B^{BG}-AL^{ML}-NS^{ML}-PD (25.52)
10	50	2	30	0.35	0.70	A^{BG}-AL^{ML}-NS^{ML}-PD (6.02)	A^{BG}-AL^{ML}-NS^{ML}-PD (11.29)	A^{BND}-AL^{ML}-NS^{ML}-PD (24.45)
11	50	2	30	0.70	0.35	B^{BND}-AL^{ML}-NS^{ML}-PD (9.14)	A^{BND}-AL^{ML}-NS^{ML}-PD (15.19)	A^{BND}-AL^{ML}-NS^{ML}-PD (30.55)
12	50	2	30	0.70	0.70	A^{BG}-AL^{ML}-NS^{ML}-PD (7.80)	B^{BND}-AL^{ML}-NS^{ML}-PD (13.16)	B^{BND}-AL^{ML}-NS^{ML}-PD (28.24)
13	50	2	50	0.35	0.35	B^{BG}-AL^{ML}-NS^{ML}-PD (6.30)	B^{BG}-AL^{ML}-NS^{ML}-PD (11.10)	B^{BG}-AL^{ML}-NS^{ML}-PD (25.52)
14	50	2	50	0.35	0.70	A^{BG}-AL^{ML}-NS^{ML}-PD (6.02)	A^{BG}-AL^{ML}-NS^{ML}-PD (11.29)	A^{BND}-AL^{ML}-NS^{ML}-PD (25.66)
15	50	2	50	0.70	0.35	B^{BND}-AL^{ML}-NS^{ML}-PD (9.14)	B^{BND}-AL^{ML}-NS^{ML}-PD (17.04)	A^{BG}-AL^{ML}-NS^{ML}-PD (38.75)
16	50	2	50	0.70	0.70	B^{BG}-AL^{ML}-NS^{ML}-PD (7.30)	B^{BG}-AL^{ML}-NS^{ML}-PD (13.83)	B^{BG}-AL^{ML}-NS^{ML}-PD (31.03)

Table 9 (continued): Best appointment system and corresponding total cost for different clinic environments

Clinic Environment	λ^A	$\frac{\mu_A}{\mu_B}$	PN	CV-A	CV-B	Best AS Design (TC)		
						$\frac{C_d}{C_p} = 1$	$\frac{C_d}{C_p} = 2$	$\frac{C_d}{C_p} = 5$
17	35	1.33	30	0.35	0.35	B^{BND}-AL^{ML}-NS^{ML}-PD (8.54)	B^{BND}-AL^{ML}-NS^{ML}-PD (14.28)	B^{BND}-AL^{ML}-NS^{ML}-PD (31.50)
18	35	1.33	30	0.35	0.70	B^{BND}-AL^{ML}-NS^{ML}-PD (7.33)	A^{BG}-AL^{ML}-NS^{ML}-PD (13.98)	A^{BG}-AL^{ML}-NS^{ML}-PD (33.08)
19	35	1.33	30	0.70	0.35	B^{BG}-AL^{ML}-NS^{ML}-PD (6.61)	B^{BG}-AL^{ML}-NS^{ML}-PD (12.79)	B^{BG}-AL^{ML}-NS^{ML}-PD (31.32)
20	35	1.33	30	0.70	0.70	B^{BND}-AL^{ML}-NS^{ML}-PD (11.32)	B^{BND}-AL^{ML}-NS^{ML}-PD (19.29)	**A^{BND}-AL^{ML}-NS^{ML}-PS (39.28)**
21	35	1.33	50	0.35	0.35	A^{BG}-AL^{ML}-NS^{ML}-PD (8.36)	A^{BG}-AL^{ML}-NS^{ML}-PD (15.40)	B^{BND}-AL^{ML}-NS^{ML}-PD (30.82)
22	35	1.33	50	0.35	0.70	A^{BG}-AL^{ML}-NS^{ML}-PD (9.41)	A^{BG}-AL^{ML}-NS^{ML}-PD (17.47)	A^{BG}-AL^{ML}-NS^{ML}-PD (41.64)
23	35	1.33	50	0.70	0.35	B^{BG}-AL^{ML}-NS^{ML}-PD (7.77)	B^{BG}-AL^{ML}-NS^{ML}-PD (14.83)	B^{BG}-AL^{ML}-NS^{ML}-PD (36.02)
24	35	1.33	50	0.70	0.70	B^{BND}-AL^{ML}-NS^{ML}-PD (10.50)	**A^{BND}-AL^{ML}-NS^{ML}-PS (16.02)**	**A^{BND}-AL^{ML}-NS^{ML}-PS (32.01)**
25	50	1.33	30	0.35	0.35	B^{BND}-AL^{ML}-NS^{ML}-PD (7.05)	B^{BND}-AL^{ML}-NS^{ML}-PD (12.76)	B^{BND}-AL^{ML}-NS^{ML}-PD (29.88)
26	50	1.33	30	0.35	0.70	B^{BG}-AL^{ML}-NS^{ML}-PD (6.13)	B^{BG}-AL^{ML}-NS^{ML}-PD (11.28)	B^{BG}-AL^{ML}-NS^{ML}-PD (26.08)
27	50	1.33	30	0.70	0.35	A^{BND}-AL^{ML}-NS^{ML}-PD (7.42)	A^{BND}-AL^{ML}-NS^{ML}-PD (13.27)	A^{BND}-AL^{ML}-NS^{ML}-PD (30.82)
28	50	1.33	30	0.70	0.70	A^{BND}-AL^{ML}-NS^{ML}-PD (7.16)	A^{BND}-AL^{ML}-NS^{ML}-PD (12.66)	A^{BND}-AL^{ML}-NS^{ML}-PD (29.15)
29	50	1.33	50	0.35	0.35	B^{BND}-AL^{ML}-NS^{ML}-PD (7.14)	B^{BND}-AL^{ML}-NS^{ML}-PD (11.40)	B^{BND}-AL^{ML}-NS^{ML}-PD (24.18)
30	50	1.33	50	0.35	0.70	A^{BND}-AL^{ML}-NS^{ML}-PD (6.94)	A^{BND}-AL^{ML}-NS^{ML}-PD (11.85)	A^{BND}-AL^{ML}-NS^{ML}-PD (26.56)
31	50	1.33	50	0.70	0.35	B^{BND}-AL^{ML}-NS^{ML}-PD (7.97)	B^{BND}-AL^{ML}-NS^{ML}-PD (13.69)	B^{BND}-AL^{ML}-NS^{ML}-PD (30.85)
32	50	1.33	50	0.70	0.70	B^{BND}-AL^{ML}-NS^{ML}-PD (6.94)	B^{BND}-AL^{ML}-NS^{ML}-PD (12.04)	B^{BND}-AL^{ML}-NS^{ML}-PD (27.33)

Table 9: Improvement achieved by adopting proposed scheduling rules

Clinic Environment	Best Sequencing Rule (Improvement Achieved)		
	$C_d/C_p = 1$	$C_d/C_p = 2$	$C_d/C_p = 5$
2	S_3^{ML} (8.95%)	—	—
3	S_3^{ML} (39.95%)	S_3^{ML} (40.80%)	S_3^{ML} (27.30%)
4	S_3^{ML} (23.40%)	S_3^{ML} (26.85%)	S_3^{ML} (26.37%)
7	S_3^{ML} (19.64%)	S_3^{ML} (20.43%)	S_3^{ML} (8.54%)
12	S_3^{ML} (2.30%)	—	—
17	S_3^{ML} (14.4%)	S_3^{ML} (14.30%)	S_3^{ML} (16.89%)
18	—	S_3^{ML} (6.44%)	S_3^{ML} (13.18%)
19	—	—	S_1^{ML} (8.33%)
20	S_4^{ML} (14.75%)	S_3^{ML} (10.89%)	S_3^{ML} (4.68%)
21	S_3^{ML} (11.24%)	S_3^{ML} (15.45%)	S_3^{ML} (11.32%)
22	—	S_3^{ML} (4.92%)	S_3^{ML} (9.10%)
23	S_1^{ML} (1.16%)	S_1^{ML} (14.23%)	S_1^{ML} (22.71%)
24	S_4^{ML} (37.71%)	S_4^{ML} (25.91%)	S_4^{ML} (12.99%)
25	S_4^{ML} (4.26%)	S_4^{ML} (12.23%)	S_4^{ML} (20.45%)
26	S_2^{ML} (1.00%)	S_2^{ML} (4.80%)	S_2^{ML} (5.33%)
27	S_2^{ML} (7.55%)	S_2^{ML} (16.28%)	S_2^{ML} (22.65%)
31	—	—	—
32	S_3^{ML} (1.15%)	—	—

Note: Proposed sequencing rule combined with **AL^{ML}-NS^{ML}-PD** yielded the best improvement in all cases.

4.3 Managerial implications and practical implementation

Even though the numerical experiments identify the best-performing rules for different clinic environments, redesigning an AS is a strategic decision for healthcare organizations. As demonstrated in Section 4, the design of an AS (allocation, sequencing, overbooking decisions) has a substantial impact on patient- and doctor-related performance measures, and the best approach could vary depending on the clinic environment. Besides, translating the proposed approaches into practice is a critical task and may be associated with some practical challenges. Therefore, in this section, we present the major implications of this study for healthcare administrators and provide some insights on the practical implementation.

The key takeaway is that clinics must transition to a data-driven patient-specific approach to estimate no-shows and service time since these uncertain parameters have a substantial influence on the AS performance. Our analysis shows that the benefit achieved by integrating ML-based predictions of clinical uncertainty into AS design can be significant as opposed to the common practice of using the statistical averages associated with each patient class. Another important finding with practical relevance is the insights on overbooking strategies for no-show adjustment. Our findings show that double-booking leads to higher AS efficiency as opposed to compressing the appointment duration proportional to the predicted no-show rate. Specifically, the proposed double-booking strategy, which achieves a trade-off between patient waiting time and doctor idle time, can enable the practice to efficiently manage no-shows. This strategy can be integrated into the

clinical decision support system (CDSS) to seamlessly automate the overbooking decision in real-time. Finally, we show that most clinics could benefit by classifying patients into distinct groups and leveraging them for sequencing appointment requests. Our results revealed two sequencing policies (B^{BG} and B^{BND}) to consistently perform well in most clinic environments. Furthermore, the proposed sequencing rules, which leverage the ML-based prediction of service time, can outperform traditional sequencing policies in many clinic settings, especially when the doctor's time is considered to be more important than the patient's time.

With regard to practical implementation, the proposed predict-then-schedule framework can be adopted and integrated with the existing health information system, as illustrated by the three layers in Figure 6. The system layer allows the extraction of relevant data needed to determine clinical uncertainties and appointment slots for any patient. The process layer will contain the proposed framework as two modules - predict and schedule. The "predict" module embeds the capability of ML algorithms into a CDSS, and facilitates real-time patient-specific no-show and service time predictions. In practice, clinics can use the procedure described in Section 4.1.1 to determine the best-performing ML algorithm for that clinic's patient population. The "schedule" module determines the appointment time and interval in real-time by integrating the AS design (allocation, sequencing, overbooking) into the CDSS. Upon determining the best AS design for the clinic (as described in Sections 4.1.2 - 4.1.4), the schedule construction algorithm (Algorithm 1) proposed in this research can be adapted for developing the "schedule" module. Note that the sequential scheduling approach described in Algorithm 1 is motivated by the

characteristics of the clinic under study (high demand, no-shows, service time variations, two patient classes, patient accepts available appointment), and may need to be tailored for specific clinic attributes. Finally, the experience layer acts as a front-end user interface and relays information from the CDSS.

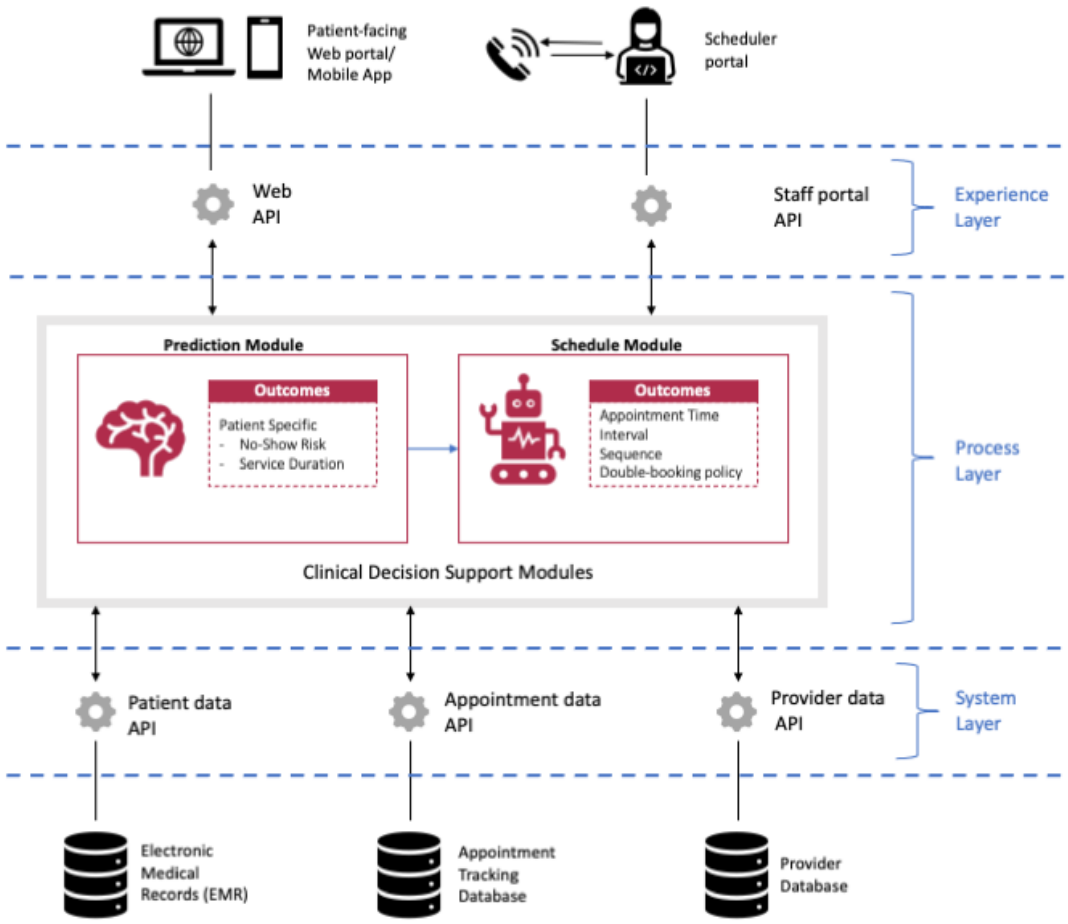

Figure 12: Illustration of health information system architecture for implementing the predict-then-schedule framework

Once the two modules are integrated into the CDSS, it can enable the scheduler to book the sequential appointment requests across the scheduling horizon. In particular, when a

patient calls for an appointment, the "prediction" module will estimate that patient's no-show probability and appointment duration, and feed it as an input to the "schedule" module. Subsequently, the "schedule" module will determine the appointment time and duration of this patient according to the AS design. Note that the clinic may schedule any slots within the scheduling horizon as long as it meets the AS design, and need not necessarily fully book a given day before accepting appointments for the next day. For instance, let us consider a situation where the scheduling horizon is 5 days and the first day currently has a 30-minute unscheduled slot. If the next patient requires a 45-minute appointment, then the proposed approach will assign that patient to a feasible slot in the upcoming day (while respecting the allocation, sequencing, and overbooking decisions pertaining to the AS design). The 30-minute unscheduled slot on the first day will be considered again for the next appointment request. Since the demand volume is high, it is likely that all slots will be booked before the day of the appointment.

Nevertheless, clinics may also face certain issues when adopting the proposed predict-then schedule approach. In particular, the schedule constructed may not utilize all available slots in some cases. Since the appointment duration varies for each patient, the clinic capacity is not filled at a constant rate. As a result, when the schedule is constructed sequentially, few slots (especially smaller durations) may be left vacant as the incoming requests may require a longer duration than the available capacity. The hospital administration may allow the empty slots in the schedule to compensate for any potential delay (e.g., late service completion, doctor interruptions) from accumulating throughout the schedule. Alternatively, an incoming request can be scheduled in the vacant slot to fully utilize the

capacity even if the patient is predicted to require a longer duration, thereby increasing access to care at the expense of a potentially higher patient waiting time and doctor overtime. Furthermore, a patient may insist on an appointment slot that violates the sequencing rule or overbooking policy associated with the AS design. In such situations, the scheduler could adopt one of the following options: (i) communicate the appointment time provided by the CDSS (in accordance with the AS design) and let the patient either accept/reject the appointment or (ii) oblige to the patient's request by not adhering to the CDSS recommendations. To establish the best decision for such situations, an extensive analysis to quantify the impact of both options on the AS efficiency must be performed.

4.4 Conclusions

In this chapter, we consider the problem of designing an AS for scheduling patients in the presence of sequential call-ins, no-shows, service time uncertainty, and distinct patient groups. Also, unlike most prior works, this study considers the two sources of clinical uncertainty, namely, no-shows and service time, to be heterogeneous and patient-specific. A predict-then-schedule framework is proposed, where ML algorithms are leveraged to design the AS. The first phase of the proposed approach is to predict the likelihood of a patient missing an appointment and his/her expected service time at the clinic using ML algorithms. A total of 16 predictors pertaining to patient demographics, appointment detail, medical history, and patient behavior are used as inputs to the ML algorithm. In the second phase, the predictions are leveraged to determine no-show adjustment and appointment duration during the real-time scheduling of patients. In addition, this research also

capitalizes on the ML-based prediction of service time to propose four new rules for sequencing patients. The proposed approach and sequencing rules are evaluated by considering a real-life case study of a cardiology clinic. Besides, we also simulated 32 different clinic environments to evaluate the impact of integrating ML-based no-show and service time prediction on AS efficiency. The results demonstrate the capability of ML algorithms to accurately predict patient-specific service time and no-show. Moreover, our comprehensive analysis showed that integrating the ML-based estimates into the AS design via the proposed predict-then-schedule framework achieves a substantial improvement in AS efficiency for all clinic environments tested when compared to the traditional approach of utilizing historical averages to estimate no-shows and service time.

For future work, we have identified the following research directions. First, the clinic environments considered in this study can be extended to include other characteristics such as multi-day scheduling horizon, walk-ins and late arrivals. Furthermore, the impact of accommodating patient preferences and defying sequencing decisions on AS efficiency can be investigated to obtain managerial insights for practical implementation. Second, the impact of integrating the ML-based predictions with existing scheduling approaches developed for other clinical environments, such as multi-server clinics [154,155] and multi-stage multi-server environments [59] can be considered. Third, the proposed approach can be extended to accommodate more than two patient classes. Finally, other approaches for uncertainty prediction (e.g., stacked generalization) and sequential scheduling (e.g., Markov decision process) can be considered within the scope of the predict-then-schedule framework.

Chapter 5. Early Detection of Cardiovascular Disease Risk Using Explainable Machine Learning Framework

As discussed in Chapter 2, existing association studies and prediction models have several limitations. First, most previous research focuses on the impact of a single risk factor (such as gender, total cholesterol, HDL cholesterol, systolic blood pressure, and smoking) on cardiovascular disease, thereby providing limited scope for risk assessment among adolescents. Second, studies that consider the impact of more than one risk factor (such as Framingham risk-score model and ASCVD Risk Estimator Plus) on CVD use different forms of regression or multivariate analysis and assume the risk factors are related to CVD in a linear pattern. As a result, the complex synergistic interaction of risk factors is not recognized. Third, almost all the existing prediction models use factors such as age, gender, race, cholesterol, blood pressure, smoking status, and presence of diabetes to estimate the 10-year or 30-year risk of heart disease or stroke and don't consider other behavioral and lifestyle's factors as inputs to their models. Most importantly, all existing risk prediction models are applicable only for adults above the age of 30 years and are not suitable for determining the long-term impact of unhealthy behavior in the earlier adolescent years. Finally, very little academic research is devoted to the development of a predictive model that is able to categorize adolescents as high or low risk of CVD in adulthood. This research aims to overcome the above-mentioned limitations in the literature. Specifically, the ML-based CVD risk predictor developed in this research applies to the adolescent population. Moreover, the proposed risk scoring method is unconstrained, assuming all possible forms of relatedness of risk factors and incidence of CVD risk. Herein, we propose to evaluate

whether the long-term CVD risk score can be predicted based on adolescent-related factors by using ML algorithms. Thus, this chapter aims to answer the following research questions:

i. Can adolescent risk factors (i.e., socioeconomic, demographic, lifestyle, stressful life event, positive mood, self-image, and depressive symptoms) predict adulthood CVD risk?
ii. Which adolescent risk factors are key predictors of adulthood CVD risk?
iii. Can the decision-making process of the ML-based long-term CVD risk prediction be explained or interpreted?

5.1 Research Methodology

An overview of the methodology is shown in Figure 13. We leveraged the Waves I and II Add Health data to identify potential adolescent risk factors (predictors) and Wave IV data to estimate the CVD risk (outcome). The data is first pre-processed and then partitioned into training and testing subsets. The training subset is employed to train the ML models, while the testing set is used to evaluate the trained algorithm. In addition, the results from ML models are explained using the Shapley Additive exPlanations (SHAP) method.

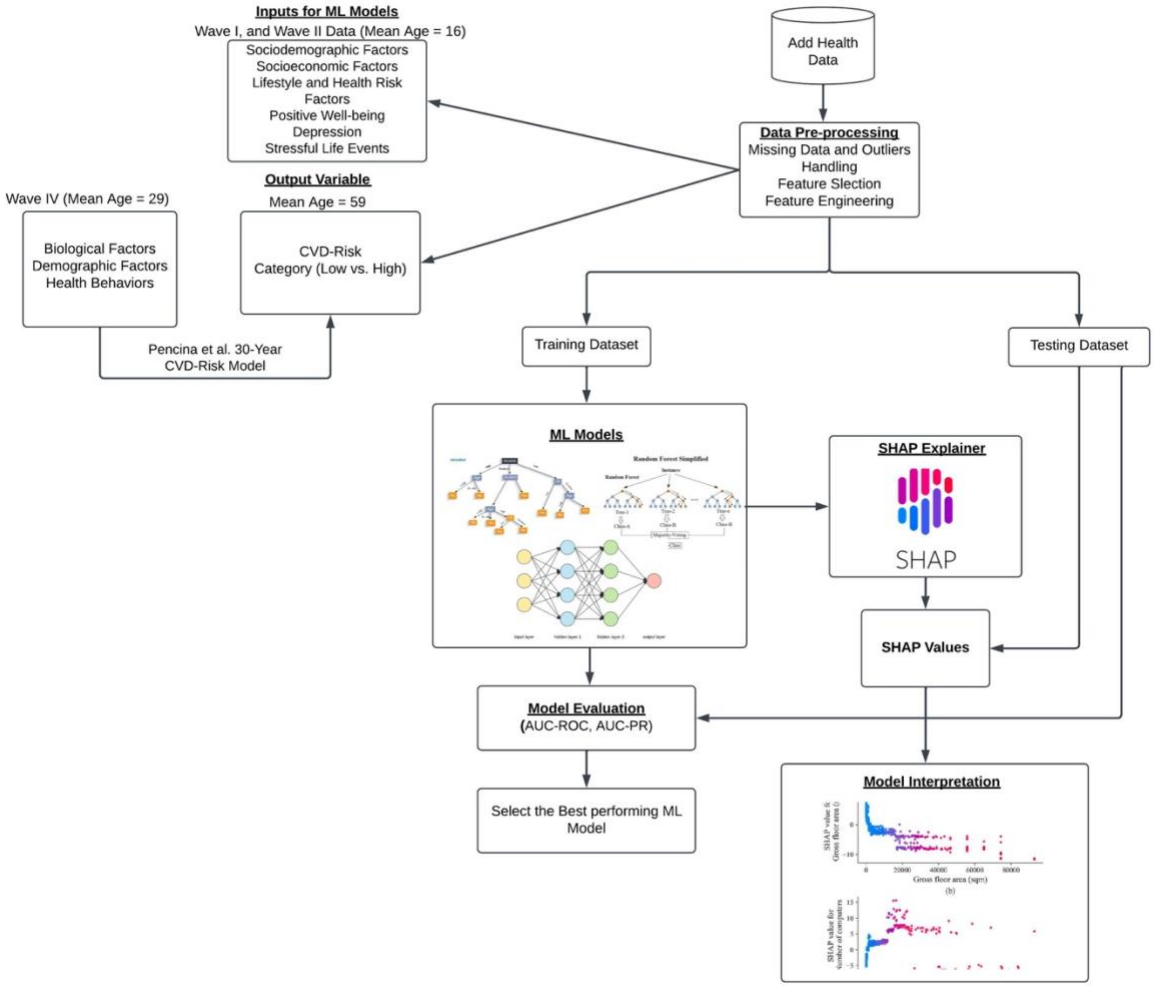

Figure 13: Methodology overview

5.1.1 Data Description

This study uses data from a nationally representative sample of adolescents taking part in the National Longitudinal Study of Adolescent to Adult Health (Add Health)[156]. The study followed over 20,000 individuals from adolescence to adulthood, starting with a school questionnaire and home interview for students in grades 7 through 12 in 1994 -1995 (Wave I). The Add Health cohort was followed into young adulthood with follow-up multi-

wave in-home interviews: Wave II (1996), Wave III (2001-2002), Wave IV (2007 -2008). For Wave I, the in-school questionnaire asked adolescents about their social demographics, their parents' education, occupations, self-esteem, health status, and risk behaviors. The in-home interviews included questions regarding nutrition, family composition and dynamics, substance use, and criminal activities. In addition, a parent, preferably the resident mother, was asked to complete an interviewer-assisted questionnaire on topics such as inheritable health conditions, relationships, education, employment, and income. From all participants in Wave I, around 15000 was followed up in Wave II. The data collected in this wave was similar to Wave I, but also included more detailed nutrition information. Moreover, the height and weight of each participant were recorded. In Wave IV, when the participants where adults, several health-related biomarkers such as height, weight, and waist circumference, cardiovascular measurements including systolic blood pressure, diastolic blood pressure, pulse, metabolic measures from lipids, glucose, and glycosylated hemoglobin (HbA1c), measures of inflammation and immune function were recorded. For more details about the Add Health study and design, readers can refer to [156]. In this study, we use data from Waves I and II to identify potential adolescent risk factors (predictors) and Wave IV data were used to estimate the adulthood CVD risk.

5.1.2 Data Preparation

The raw data is pre-processed and prepared for predictive modeling. The predictors included both continuous and categorical variables. However, certain variables contained missing values because the participant did not know the most appropriate option for that

item or refused to provide an answer. All missing values are imputed using chained equations [157], where the distribution of unobserved value is estimated based on the observed values. In particular, if there are *M* independent variables, then the variable (e.g., x_1) with missing values is regressed on the other independent variables, (x_2, x_3…, x_M), by considering only the observed values, and subsequently, the missing values in x_1 are obtained using the predictions from the fitted model. The procedure is repeated for each variable containing one or more missing values to obtain a complete dataset. Subsequently, each categorical variable is one-hot encoded and transformed into multiple numeric fields. The pre-processed dataset includes 14,083 complete records, which is then used for ML model development. The procedure adopted for preparing the predictors and outcome variable is described in the following subsections.

5.1.2.1 Input Variables and Feature Engineering

Since Add Health survey questions are not specifically targeted towards CVD risk factors, there are many survey items that are not pertinent to this research. Therefore, we selected relevant questions based on expert opinion (e.g., endocrinologist or cardiologist) and prior research findings [158,159,168–173,160–167]. Survey items from Waves I and II that reflected the following factors are selected as input variables for ML model development – sociodemographic, socioeconomic, lifestyle and health risk, stressful life events, positive well-being, and depression. While some independent variables (e.g., gender, age) can be directly obtained from Waves I and II survey questionnaires, some predictors must be inferred from one or more survey items. We rely on established survey items that are

validated and employed in prior literature to infer these predictors (see Table 10). For instance, adolescents' physical activity is obtained from the seven survey questions where individuals are asked to report how many times they are engaged in a specific activity in the last week. The responses from these questions are aggregated to create the adolescents' total physical activity variable [160]. For sedentary behavior, the total screen-time is captured based on the following questions – "How many hours a week do you watch television?", "How many hours a week do you watch videos?" and "How many hours a week do you play video or computer games?" Participants' responses are summed to obtain the total number of hours of screen-time per week [162].

Positive well-being factors such as positive mood and self-image are created from participants' responses to the 10 questions that came from the Center of Epidemiologic Studies Depression (CES-D) Scale [170]. Four of these items asked about the following feelings experienced in the last week – happiness, feeling as good as other people, enjoying life, and hopefulness. The answers for these four questions are summed to generate a single positive mood factor [165,166,174]. The other six questions asked participants whether they– have good qualities, have a lot to be proud of, like themselves, do things right, are socially accepted, feel loved and wanted. The answers to these questions are added to measure the self-image of each participant [167–169,174]. Adolescent depression is self-reported and measured based on 15 questions from the CES-D questionnaire. This includes whether individuals felt sad, couldn't shake off the blues, felt lonely, thought their life was a failure, bothered by things that usually don't bother them, felt disliked by other people, trouble of concentration, felt fearful, lost appetite, felt people unfriendly to them, felt exhausted, spoke less than usual, found it hard to get started doing things, and felt life

doesn't worth living. The responses to these 15 questions are summed to create a depressive score that ranged from 0 – 45, with 45 indicating higher depression [159,162,164,171,174].

Table 10: Input variables and their related topics based on literature

Adolescent Factors	Variables	Reference
Sociodemographic	• Gender • Age • Race	[158–162]
Socioeconomic	• Parental Education • Parental income • Family structure	[159–161]
Lifestyle and health risk	• Self-rated health • Physical activities • Sedentary behaviors • Fast-food consumption • Eating breakfast • Alcohol use • Marijuana use • Smoking status • Obesity • Sleep duration • Parental obesity • Parental diabetes	[159–164]
Positive well-being	• Positive mood • Self-image	[165–169,174]
Depression	• Depressive symptoms	[159,162,164,170,171,174]
Stressful life events	• Saw violence • Threatened by knife or a gun • Was stabbed • Was jumped • Skipped necessary medical care • Suffered a serious injury • Was raped • Friend attempted suicide • Family member attempted suicide • Was injured in a physical fight • Hurt someone in a physical fight • Romantic relationship ended • Contacted a STD • Run away from home • Suffered verbal abuse in romantic relationship	[159,164,172,173]

5.1.2.2 Output Variable

The 30-year adulthood CVD risk category (low or high risk) is estimated using Wave IV survey data collected 14 years after the initial interview (participants mean age of 29). In addition, the survey collected information related to participants' demographics, anthropometric measures, and health test results. We used the risk prediction function, a modified Cox model, derived by Pencina et al. [99] to compute the adulthood CVD risk over a 30-year time frame. The model leverages factors from Wave IV, including – age, gender, systolic blood pressure (SBP), BMI, smoking status, use of antihypertensive medications, and presence of diabetes to estimate the 30-year CVD risk score. Similar to previous research, an individual with a risk score over 20% is classified as high-risk and low-risk otherwise [175,176]. Thus, the outcome variable is the long-term CVD risk (low or high) of an adolescent.

5.1.3 ML Algorithms and Evaluation

The problem of predicting the CVD risk (i.e., categorizing individuals as low and high risk of CVD) is modeled as a supervised classification problem. Recent research has demonstrated the capability of decision trees (DT), random forest (RF), extreme gradient boosting (XGBoost), and deep neural networks (DNN) to accurately predict binary variables in the healthcare domain, such as disease risk [177–179], heart disease [180–183], and mortality risk [184–186]. Therefore, we employed and evaluated these four ML algorithms for CVD risk classification. Stratified random sampling is performed to divide

the data into two parts – 75% is used for training the classification models, and the remaining 25% is held-out for evaluation. To eschew overfitting (learning noise) in the learning phase and optimize the parameters of ML models, a 10-fold cross-validation procedure is performed on the training data [112,113]. A brief description of the four ML techniques is provided in the following subsections.

5.1.3.1 Decision Tree (DT)

Decision Tree (DT) is a supervised ML technique used for both classification and regression problems. The algorithm uses a tree-like structure and includes root nodes representing the input features or the dataset, branches representing the decision rules, and the leaf/terminal node representing the outcome (or class) [187]. Decision trees are built using a recursive partitioning heuristic, where each sub-node following the root node is split into several nodes. The idea for DT is to divide the dataset into dense and sparse regions and continue splitting until the data is sufficiently homogenous. DT determine the attribute to split at each stage based on the information gain (IG) [188], which measures how well a specific feature divides the dataset according to its classification output [189]. IG measures the reduction in entropy (uncertainty in the data) after splitting the dataset based on a given attribute. Upon evaluating the IG for all potential attributes, the one with the highest IG is chosen for splitting. The formulas for entropy and IG are given in Equations (1) and (2), respectively, where *N* is a training dataset and p_i is the probability of randomly selecting $class_i$ in *S.* Also, *A* is attribute selected for splitting, *v* is a possible value of attribute *A* and N_v the subset of *S* for which attribute *A* has a value *v.*

$$E(N) = \sum_{i=1}^{c} - p_i \log_2 p_i \quad (1)$$

$$IG(N, A) = E(N) - \sum_{v \in A} \frac{\lceil N_v \rceil}{N} E(N_v) \quad (2)$$

The steps involved in training a DT are as follows:

i. It starts with the training set *N* as the root node.

ii. With each iteration, the algorithm calculates Entropy and IG for attribute A in N that has not been selected.

iii. Then, it selects attribute *A*, which has the largest IG.

iv. Using the selected attribute, the training set N is then divided into subsets.

v. The algorithm continues to run for each subset, considering only the attributes that have not been selected before.

5.1.3.2 Random Forest (RF)

Random forest is an ensemble technique that employs a multitude of uncorrelated decision trees [114]. Each tree is usually trained using the bagging method, where independent trees are constructed using bootstrap sampling [190]. Each tree gives a class prediction, and the class with the most votes will be selected as the final prediction. In standard trees, the best split is chosen from the set of all variables, whereas in RF, the node is split using the best among a subset of randomly selected variables. This strategy performs well compared to other classifiers such as support vector machines and discriminative analysis [114]. In addition, such an approach reduces the likelihood of overfitting [114]. The key steps

associated with training an RF algorithm for classification are as follows:

i. Randomly select *n* samples from the *N* training dataset.

ii. For each *n* sample, construct a classification tree. At each node, randomly sample *m* variables and select the best split among these *m* predictors.

iii. Classify new data by taking the majority vote (the most predicted class) of the constructed trees (*T*).

5.1.3.3 Extreme Gradient Boosting (XGBoost)

In gradient boosting machines, the learning process is based on the idea of consecutively fitting new base learners to provide a more accurate estimation of the output variable such that the gradient of the loss function of the whole ensemble (group of base learners) is reduced [191]. Extreme gradient boosting was developed by Chen and Guestrin [192], which is an optimized version of the gradient boosting machine. The main improvement in XGBoost was the normalization of the loss function, which reduces the modeling complexity and hence the likelihood of overfitting [193]. The algorithm for XGBoost is an iterative process that follows the following steps [117]:

- Initialize model with a constant value $f_0(x) = \bar{y}$
- For each boosted model *t*
    - Compute the residual of the previous model as $h_t(x) = y - f_{t-1}(x)$
    - Fit a tree ($g_t(x)$) on the residuals of the previous tree by using a subset of the training data drawn without replacement
    - Compute the gradient descent step size ($\rho_t(x)$)

- Combine the trained trees to obtain a boosted model, $f_t(x)$, which has lower loss function than $f_{t-1}(x)$. Also, to prevent overfitting a shrinkage parameter ($J \in (0,1]$) is used to control the learning rate. Therefore, the combined model is represented as $f_t(x) = f_{t-1}(x) + J \times \rho_t(x) \times g_t(x)$
- The procedure is repeated until the reduction in the loss function is insignificant.

5.1.3.4 Deep Neural Networks (DNN)

Deep learning models are based on the architecture of the neural networks, which adopts a hierarchal organization of interconnected neurons. These neurons pass a signal to other neurons based on the perceived input and form a complex network [194]. In this research, we adopt a deep feed-forward artificial neural network architecture containing three types of layers – input, hidden, and output. Each layer has a set of nodes, and each node has an activation function that calculates the output to the next node. The connection between two nodes of successive layers has an associate weight, which defines the strength of the connection between nodes. For instance, if the weight from node *i* has a larger value than the weight from node *j*, it means that node *i* has a greater influence over node *j*. The term "deep neural networks" evolves from many hidden layers that facilitate learning complex relationships between input and output data and provide better prediction [119].

In DNN, the learning process starts when input data is passed to the first hidden layer, which is then transformed using the activation function, then this information is forwarded

to the next layer. This process, known as forward-propagation, continues until the information from the last hidden layer is transformed and passed to the output layer to make a prediction. The training data is fed into the DNN in small batches, where a batch contains one or more training records for learning purposes. The connection weights between nodes *i* and j (w_{ij}) of DNN are initialized randomly and progressively altered (as shown in Equation 3) for every successive batch (b) based on the difference between the actual and predicted output (E) obtained from a previous batch (b – 1) and network's learning rate (η) - referred to as backward pass.

$$w_{ij}(b) = w_{ij}(b-1) + \eta \frac{\partial E}{\partial w_{ij}} \quad (3)$$

Moreover, DNN may require several epochs, number of passes through the entire training data, where each epoch has one or more batches before approximating the mapping function.

5.1.3.5 Hyperparameters Tuning

The hyperparameters of the four ML models are tuned using the grid search technique as it was well-suited for models with few parameters and easy-to-implement [120]. The hyperparameter space screened for each algorithm is based on the commonly used values in the literature [121–126]. The number of trees (T) considered for RF, DT, and XGBoost were (2^5, 2^6, 2^7, 2^8). Besides, the number of variables randomly sampled at each candidate split (m) in the RF algorithm are varied from 5 to 15. For XGBoost, the learning rate (J) explored was between 0.025 to 0.1, and the minimum number of observations in a trees'

terminal node (*C*) is varied from 2 to 8. For DNN, the number of hidden layers (*H*) examined were (1, 2, 3, 4, 5), while the number of nodes in each hidden layer (n_k) is varied from 1 to 16. Moreover, the values of learning rate (η), number of epochs (*P*) and batch size (*b*) searched were (0.01, 0.05, 0.1), (10, 20, 30, 40, 50), and (32, 64 ,..., 512), respectively.

5.1.3.6 Evaluation metrics

The classification models are compared based on three measures, namely, area under the receiver operating characteristic curve (AUC-ROC), area under the precision-recall curve (AUC-PR) and misclassification rate (MCR). The AUC-ROC is a single measure for evaluating the overall discriminative performance of the ML model [127]. Besides, AUC-ROC has been consistently used in prior research dealing with classification [59]. On the other hand, AUC-PR is considered to be a robust metric for evaluating models dealing with class imbalances [128]. The MCR is the ratio of total number of incorrect classifications to the total predictions. The value for these metrics ranges from 0 to 1, where a higher score is preferred for AUC-ROC and AUC-PR and a lower score is better for MCR.

5.1.4 ML interpretability and explainability

While RF, XGBOOST and DNN have demonstrated high-prediction accuracy than simpler models such as DT in prior studies, they are also regarded as 'black-box' models since it is difficult for humans to comprehend their behavior in predicting the outcome.

Interpretability of the ML model is the degree to which a model can explain its output based on a set of inputs [195]. The ability to interpret or explain the ML model predictions is crucial for data-driven decision-making, especially for adopting targeted interventions in healthcare. The scope of ML model interpretation can be global or local. Global interpretability identifies how the model makes its predictions based on a holistic view of its features, parameters, and structure. In other words, it explains the global output of the model on an abstract level [196]. On the other hand, local interpretability is achieved by designing more justified model architectures that provides explanation for a single prediction [196].

Existing methods for interpreting ML models can be categorized into two groups – intrinsically interpretable methods and model–agonistic methods [82]. Intrinsically interpretable methods are limited to self-explainable models, such as linear regression and DT models. These models are less complex and easy to explain since a mathematical rule can represent their internal structure. On the other hand, agonistic methods are not limited to specific ML models but can be applied to any ML model. In addition, agonistic methods work by analyzing the relationship between the inputs and output rather than analyzing the internal structure as in the intrinsically interpretable methods [82]. In this research, we use Shapley Additive Explanations (SHAP), a model-agonistic method that adopts the concepts from cooperative game theory [197], [198]. It calculates the contribution of each variable based on the Shapely value [199]. The contribution of feature *i* on the prediction using the Shapely value can be calculated as shown in Equation (4), where *F* is the entire

feature set, and S denotes a subset, $S \cup \{i\}$ is the union of subset S and feature i, $(v(S \cup \{i\}) - v(S))$ is the marginal contribution of feature i.

$$\varphi_i(v) = \sum_{S \subseteq F\{i\}} \frac{|S|!\,(F - |S| - 1)!}{F!}\,(v(S \cup \{i\}) - v(S)) \tag{4}$$

5.2 Results

The procedure for predicting CVD-risk category using ML algorithms is implemented in Python software on a computer configured with Intel Core i7 3.4 GHz processor, macOS Sierra operating system, and 32 GB RAM. The pre-processed dataset contained 14083 records, and 75% of it is used for training the ML algorithms while the remaining 25% is used for evaluation. The percentage of high CVD risk subjects in training and testing datasets are 17%, and 16%, respectively. The best set of hyperparameters for classification models are summarized in Table 11.

Table 11: Best hyperparameter values for classification models from grid-search

Classifier	Hyperparameter	Value
RF	*T*	32
	m	7
XGBoost	*T*	64
	C	6
	J	.05

DNN	H	3
	n_k	10
	P	50
	η	.0001
	b	8

5.2.1 Predictive Performance of Classification Models

The performance of classification models on 10-fold cross-validation and testing datasets is illustrated in Table 12. The evaluation metrics indicate a good discriminative capability of “Low” and “High” CVD-risk of all the classification models. In particular, XGBoost yielded the best average values for all the classification evaluation metrics under consideration. It achieved 4% better AUC-ROC, 3% higher AUC-PR, and 3% lower MCR than RF. Similarly, when compared to DNN, XGBoost showed about 6.0%, 2.0%, and 2.0% improvement in AUC-ROC, AUC-PR, and MCR values, respectively. The performance of ML models on the testing dataset is comparable to the cross-validation results, thereby suggesting the ML models' generalization capability. Besides, XGBoost and RF consistently outperformed the other two algorithms for CVD-risk classification, respectively.

Table 12: Performance of classification models on cross-validation and testing dataset

ML Model	Cross-validation Dataset			Testing Dataset		
	MCR	**AUC-ROC**	**AUC-PR**	**MCR**	**AUC-ROC**	**AUC-PR**
DT	11.3%	88.4%	97.7%	14.5%	78.9%	92.2%
RF	14.0%	90.1%	98.0%	15.3%	84.2%	96.7%
XGB	11.0%	90.5%	98.3%	12.6%	84.5%	96.9%
DNN	13.1%	82.8%	96.2%	13.5%	81.2%	95.6%

5.2.2 ML models interpretation

5.2.2.1 Variable Importance Plot — Global Interpretability

The predictions obtained by the ML models can be interpreted by SHAP in different ways. For instance, the feature importance plots, as shown in Figure 14**,** explain the general influence of each feature on the prediction [199]. It is calculated as the average of absolute shapely values across the entire dataset. It can be seen from the figure that gender, BMI, and age were in the top five most important features in all ML models except for DNNs where Age was less important than gender and BMI. For XGBoost, smoking and sedentary duration also influenced the CVD risk. In the case of RF, besides smoking, using marijuana was an important predictor of high CVD risk. In addition, parental obesity was one of the top five features for CVD-risk prediction. Other key factors that appeared to contribute to the prediction of CVD are self-image and eating habits in the case of DNN. Parental income is also considered one of the top five crucial CVD predictors for all ML algorithms except RF. In addition, depressive symptoms were among the top significant factors for CVD-risk prediction for DT.

As mentioned earlier, the importance plots only show the global influence of each feature on the prediction. However, they do not indicate how each predictor's contribution is either positively or negatively affecting the prediction. For that reason, summary plots, which provides a global macro-level explanation of how the input variables contribute to the prediction, is employed. Figure 15 presents the summary plot that demonstrates importance, impact, original value, and correlation of the adolescent factors to the high adulthood CVD risk. Note that the importance is demonstrated by the decreasing order of the variables. In particular, the impact (positive vs. negative) is shown in the x-axis. The color shows the value of a specific variable, in which red signifies high value and blue implies low value. In the case of categorical predictors, a red color indicates the presence of the factor (or true), while blue denotes the value of that variable to be false. For instance, "Gender_female" represents the encoded gender variable where red represents the female gender and blue indicates the male participants. Similarly, "Obese_parents" indicate parents who are obese if the value is "yes" and non-obese otherwise. The correlation of each variable with the target can be inferred when considering both the impact and the color of the observations for a specific variable [200]. The importance plots for all ML models show that males are more likely to be in the high CVD risk category as opposed to females. As expected, the likelihood of being categorized as high-risk increases with age, BMI, and cigarette smoking. On the other hand, the XGBoost (Figure 3a) shows that less sedentary durations (or being physically active) and higher parental income tend to have a lower influence on the prediction of high-risk category as opposed to the low-risk category. For RF, using marijuana and having an obese parent increase the probability of being classified as a high-risk. According to DNN plots, having low self-image and less

occurrence of eating breakfast increase the likelihood of being classified as a high-risk. The higher the depressive symptoms, the higher the chance of being classified as high-risk according to DT.

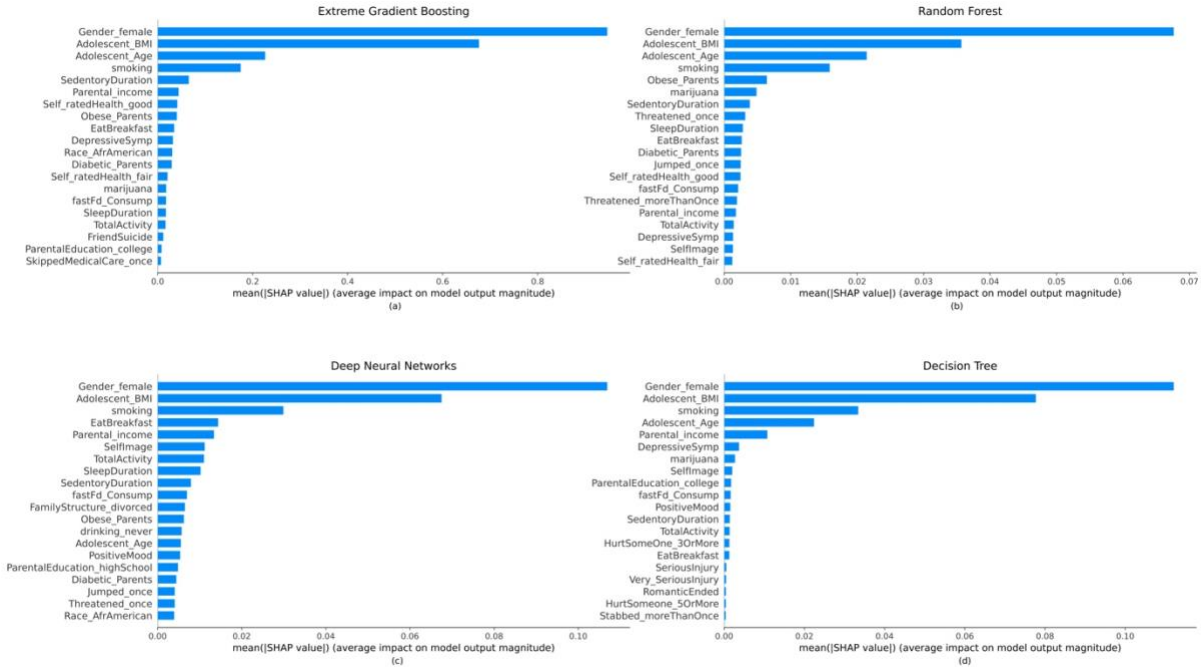

Figure 14: Global interpretation of ML models – SHAP variable importance plots: (a) XGBoost, (b) RF, (c) DNN, (d) DT

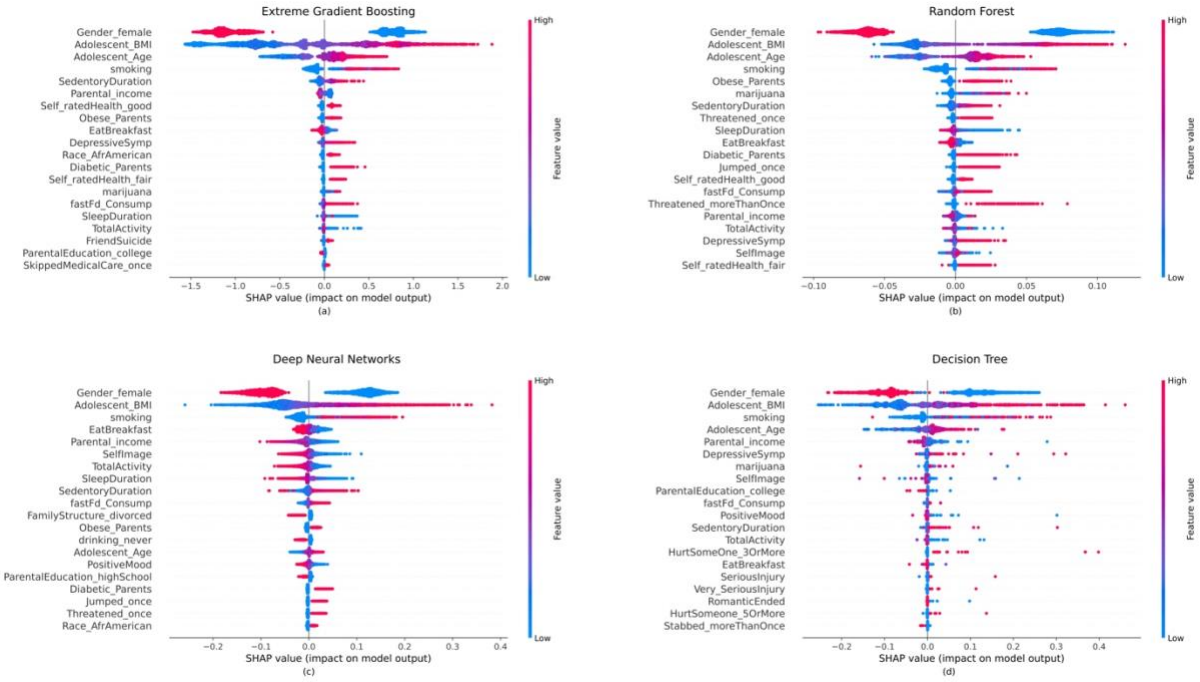

Figure 15: Global interpretation of ML models – SHAP summary plots: (a) XGBoost, (b) RF, (c) DNN, (d) DT

5.2.2.2 SHAP Dependence Plot — Global Interpretability

Other than the demographic variables (i.e., age and gender), some lifestyle and health risk variables that appear to affect the risk of CVD are – BMI, smoking, sedentary duration, and weekly breakfast frequency. To show the marginal effect of these features on the outcome of the ML models, dependence plots are used [201]. These plots view the relationship between the feature and the feature's impact on the model. In addition, they include another variable for coloring (red or blue) to highlight possible interactions. They plot the feature value versus the SHAP values of that feature across many samples. As the SHAP values increase with the increasing values of the feature, that would indicate a

positive correlation between the feature and the predicted outcome; otherwise, it would signify a negative relationship. For instance, Figure 16 (a) shows an approximately positive and linear relationship between BMI and the high CVD risk, and that BMI interacts mainly with gender. Similarly, as the instance of smoking increases, this also increases CVD risk, as shown in Figure 16 (b). The figure also shows that males (in blue) who smoke more than two cigarettes a month have a higher risk of CVD than females (in red) who also smoke the same number of cigarettes. Sedentary durations seem to interact mostly with age, as shown in Figure 16 (c). It also can be seen that individuals who are 15 years and older and have more than 24 hours of inactive durations a week have a positive and approximately linear relationship with higher CVD risk. Figure 16 (d) illustrates that eating breakfast more often decreases CVD risk. It also shows that individuals who eat breakfast more than once a week have a lower BMI than those who have breakfast once a week or do not have it.

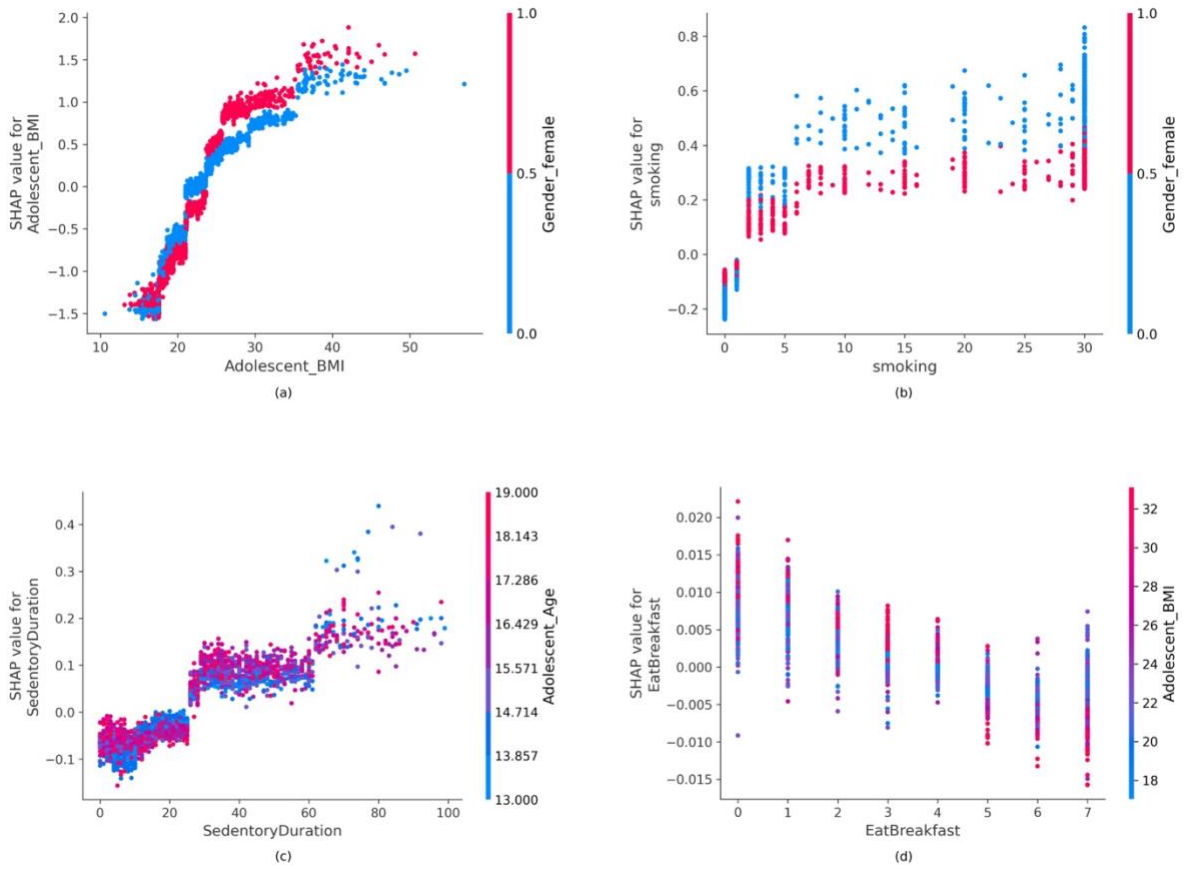

Figure 16: Partial dependance plots: (a) adolescent BMI, (b) smoking, (c) sedentary duration, (d) eat breakfast

5.2.2.3 Individual SHAP Value Plot — Local Interpretability

In addition to the global interpretation of the entire dataset, SHAP also provides local interpretation for each sample. The individual plot, as shown in Figures 17 (a) and (b) illustrate the classification of two samples as high risk and low risk, respectively. The individual/force plot shows how each feature influences the classification of each observation as high or low risk, as well as the direction and magnitude of the influence. In the context of classification, the red color represents features that drive the

classification to be in the high-risk category, while the blue color shows those nudging the prediction to be in the low-risk category. The length of the bar denotes the magnitude of influence for the corresponding feature. For instance, features with a more extended bar indicate having more influence on the output [200]. The bold value is the probability of the output being predicted as a particular risk category. Higher probability than the cutoff (in our case, the cut-off is kept at the default setting of 0.5) leads the model to classify it as a high risk, and a lower probability than the threshold leads to a classification of low risk [202]. For instance, the selected individual in Figure 17 (a) is a male who is 18 years old and smoked every day in the past 30 days. The individual plot explains how this individual is perceived by the model. It can be seen from the figure that age, smoking, and gender, in this case, are pushing the model to classify it as a high risk. It can also be seen that gender has a larger influence, followed by smoking and age for this individual. On the other hand, Figure 17 (b) shows a male individual who is 14 years old, has a normal BMI and does not smoke. In this case, the BMI, Age, and smoking status push the model to classify it as a low risk, whereas the male gender pushes it to be classified as a high risk. The combined influence from the normal BMI, younger age, and non-smoking leads to a prediction of 0.10, which is the probability of being classified as high-risk. Since this is less than the threshold 0.5, the predicted risk category is low-risk.

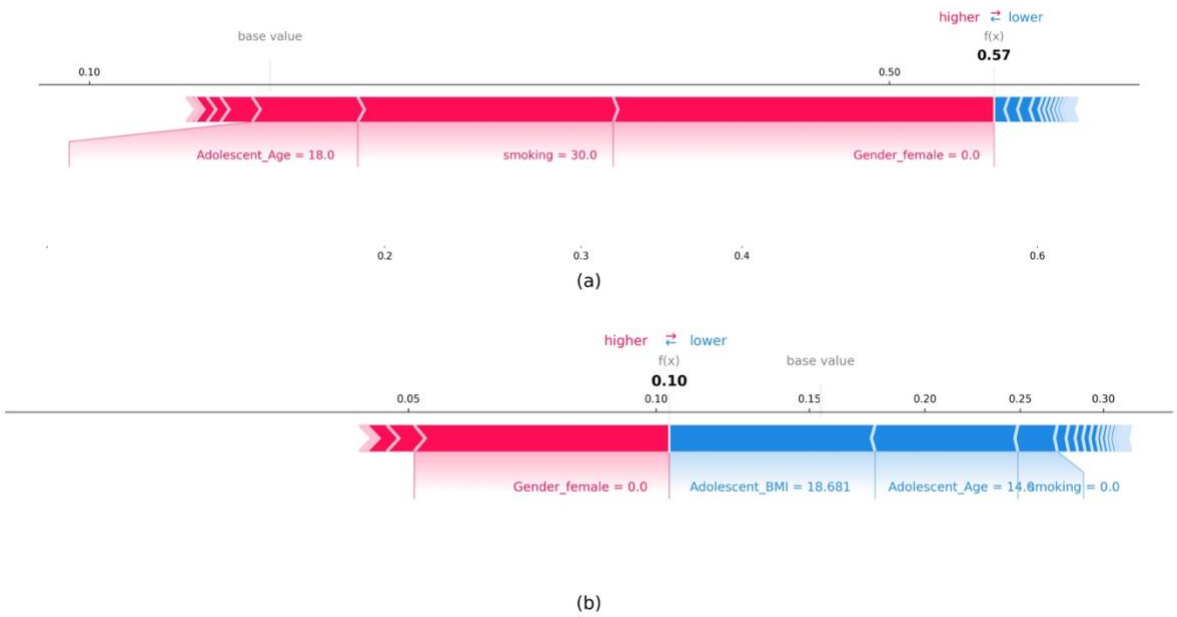

Figure 17: Local interpretation – individual plots: (a) high risk, (b) low risk

5.3 Discussion and Conclusion

In this chapter, an ML-based approach is implemented to determine if adolescent risk factors can accurately predict CVD risk in adulthood. Currently, there is no tool available to predict CVD risk among young adults. Instead, almost all of the available tools are only applicable to adults and use cross-sectional data. This study provides the first long-term ML-based CVD risk prediction model among adolescents based on a longitudinal dataset. We trained four ML models, namely XGBoost, RF, DDNs, and DT, to predict CVD risk, using 37 adolescent risk factors as predictors, and compared the model performances measured by MCR, AUC-ROC, and AUC-PR. The results of the prediction models indicated that adolescent risk factors were able to predict the CVD risk with high accuracy. This finding supports the prior research highlighting the importance of adolescent risk

factors in developing CVD events later in life [32,84,85]. Moreover, our results suggest ML models to be capable of accurately predicting long-term risk of CVD among adolescents. This finding complements other research works that have employed ML to predict the long-term risk of other diseases such as type 2 diabetes [203], kidney [204], cancer [205].

This study also identified adolescents' risk factors that were significant for predicting long-term CVD risk. Consistent with previous research, our findings reveal that gender, age, BMI, and smoking were important for predicting CVD risk [30,35,36]. Earlier studies have established an association between parental income [206], sedentary duration [207], skipping breakfast [208], self-image [209] , and depressive symptoms [210] and a higher likelihood of CVD risk, and our study results further substantiated this as these factors are found to be critical predictors of CVD risk. In addition, some predictors emerged as important for specific algorithms but not for others. This could be due to the learning pattern of the ML algorithm and the way they select and rank features as important [211]. For instance, "FamilyStructure_divorced" was ranked as one of the key predictors for CVD by DNN as opposed to the other ML models. Also, stressful life events appeared to have little to no influence on the CVD risk as they were not identified as important and low ranked by all ML models.

Most prior works that focus on interpretability/association use algorithms such as LR, DT, whereas studies focusing on achieving higher prediction accuracy use black-box ML models and compromise interpretability. This research is among the first to provide the

local and global interpretation uncovered by the black-box ML model for predicting adulthood CVD risk using adolescent risk factors. For instance, results from dependance plots revealed how risk factors such as, weekly breakfast frequency and BMI interact with each other. Our results indicate that adolescents who eat breakfast more than once a week have a lower BMI compared to those who have it once a week or skipped it. This finding support prior work highlighting the association between high BMI and skipping breakfast for adolescents [212]. Contrary to the previous work about smoking and its effect on developing CVD among adults, our results indicate that female adolescents who smoke same number of cigarettes as males have less chance of developing CVD risk [213].

The findings of this study have several implications. Once the proposed tool is validated on new data sources, it can be used to develop primordial prevention plans that promote youth health and enable individuals to seek care at an early stage. Developing such plans could improve the quality of life, avoid psychological stress, functional impairment, medication-related side effects, and premature death [214]. Moreover, early intervention could reduce healthcare costs by up to 70% [215]. Therefore, standardizing the proposed approach for other diseases and adapting it to intervene at an earlier stage could achieve substantial cost savings. In addition, the proposed approach can be scaled up to prevent and manage other diseases such as type 2 diabetes, obesity, and arthritis, thereby improving the overall population health.

Although this study has many merits, it also has a few limitations that can guide future research. First, the ML models in this study uses the data collected as part of the Add Health

study. While the study uses a representative sample of adolescents in the US, the generalization of the proposed ML models for other adolescent cohorts is not evaluated and could be considered as a future research direction. In addition, the capability of our ML models to predict CVD risk for individuals who may fall outside the age ranges considered in this study is not known. Second, the impact of certain predictors such as adolescent waist circumference, heart rate and family history of CVD were not considered as it was not collected as part of the Add Health study. To this end, future studies should validate these findings in other large-scale and longitudinal cohorts.

Chapter 6. Conclusions and Future Research Directions

This dissertation addresses the problem of improving cardiac care delivery from two perspectives. First, the problem of managing high demand for cardiac care and limited available capacity, which is currently experienced by most cardiology clinics, is tackled through the integration of ML into AS design. Second, a decision support tool is developed to facilitate early detection of CVD cases, which has the potential to handle the problem of rising CVD cases through early targeted intervention.

For developing an efficient AS, clinical uncertainties, such as consultation length and no-shows, should be correctly estimated. Thus, in chapter 3, we proposed an ML-based approach for predicting the consultation length and no-shows. Unlike the ML-based models for predicting consultation length in literature, this work presents an ML-based two-part model for predicting the consultation length while considering the possibility of no-shows. The proposed model includes – a classification model to predict no-shows (i.e., categorize the consultation length as zero or non-zero) and an ML regression algorithm to estimate consultation duration for instances classified as non-zero. Besides, a comprehensive list of features pertaining to patients, appointments, and doctors' information is considered. Furthermore, the proposed approach is validated using data obtained from the EMR system of a cardiology clinic, thereby demonstrating its feasibility. In addition, we assessed the impact of integrating consultation length and no-show predictions on the appointment system design, where two simulation models are developed to compare the clinic's existing approach with the proposed approach for estimating

patients' consultation length and compensating for no-shows. Specifically, the procedure to estimate the consultation length in the current approach depends on the patient status (new vs. return) and the reason for the visit. Whereas to compensate for no-shows, the clinic randomly overbooks based on the clinic's average no-show rate. The performance of the two simulation models is compared based on critical schedule outcome measures, namely, patient waiting time and doctor idle time. The results indicated that the proposed approach is significantly better than the current practice with respect to patient waiting time and physician idle time.

Considering these results, in chapter 4, we designed an AS for scheduling patients in the presence of sequential call-ins, no-shows, service time uncertainty, and distinct patient groups. Unlike most prior work, this study considers patients' service times and no-shows to be heterogeneous and patient-specific. A predict-then schedule framework is proposed, where predictions from ML algorithms are leveraged to determine no-show adjustment and appointment duration during the real-time scheduling of patients. Additionally, this work proposes four new rules for sequencing patients based on the ML-based prediction of service time. The proposed framework and the new sequencing rules are evaluated based on a case study of a cardiology clinic. Besides, we performed a simulation of 32 different clinic environments to evaluate the impact of integrating ML-based no-show and service time prediction on the efficiency of AS. Consistent with chapter 3 results, our comprehensive analysis showed that integrating the ML-based estimates into the AS design via the proposed predict-then-schedule framework substantially improves AS efficiency

for all clinic environments tested compared to the traditional approach of utilizing historical averages to estimate no-shows and service time.

In chapter 5, we developed a long-term ML-based model to predict CVD risk among adolescents based on a longitudinal dataset that facilitates the early detection of CVD. Almost all the existing CVD prediction models use traditional risk factors such as age, gender, race, blood pressure, and smoking status to estimate the 10-year or 30-year risk of heart disease and do not consider other behavioral and lifestyle factors as inputs to their models. Most importantly, all existing risk prediction models are applicable only for adults above the age of 30 years and are not suitable for determining the long-term impact of unhealthy behavior in the earlier adolescent years. Furthermore, very little academic research is devoted to developing a predictive model that can categorize adolescents as high or low risk of CVD in adulthood. This research overcame the above-mentioned limitations and developed an ML-based CVD risk predictor that considers a comprehensive list of risk factors, including sociodemographic, socioeconomic, lifestyle, positive well-being, depression, and stressful life events. Besides, the developed model in this research applies to the adolescent population. This study also identified adolescents' risk factors that were significant for predicting long-term CVD risk. Moreover, this research is among the first to provide the local and global interpretation uncovered by the black-box ML model for predicting adulthood CVD risk using adolescent risk factors. Finally, several implications are proposed based on our findings. For instance, the proposed model can be used to plan for early interventions for individuals who are at high risk of developing CVD,

which improves the quality of life, lowers healthcare costs, and reduces the demand for cardiac care services later.

This dissertation also identified directions for future work relating to consultation length prediction, AS design, and long-term CVD prediction. For consultation length study (chapter 3), future research should repeat the study procedure across multiple clinics, specialties, and locations to support the findings of this study. Second, the fields available with the clinic under consideration were used to build the prediction model. Therefore, it is plausible to have missed specific relevant predictors, such as medical reports and prior test results. The creation of extended databases and standardization of the fields collected across clinics would further enhance the capability of these ML algorithms. For the AS design (chapter 4), we have identified the following research directions. First, the clinic environments considered in this study can be extended to include other characteristics such as multi-day scheduling horizon, walk-ins, and late arrivals. Furthermore, the impact of accommodating patient preferences and defying sequencing decisions on AS efficiency can be investigated to obtain managerial insights for practical implementation. Second, the impact of integrating the ML-based predictions with existing scheduling approaches developed for other clinical environments, such as multi-server clinics and multi-stage multi-server environments, can be considered. Third, the proposed approach can be extended to accommodate more than two patient classes. Finally, other approaches for uncertainty prediction (e.g., stacked generalization) and sequential scheduling (e.g., Markov decision process) can be considered within the scope of the predict-then-schedule framework. In terms of long-term CVD prediction study (chapter 5), future studies should

validate our findings in other large-scale and longitudinal cohorts. In addition, the impact of other risk factors, such as waist circumference, heart rate, and family history of CVD, could be considered.

Bibliography

[1] Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, et al. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. vol. 139. 2019. https://doi.org/10.1161/CIR.0000000000000659.

[2] Center for Health Statistics N. National Ambulatory Medical Care Survey: 2014 State and National Summary Tables. Centers for Disease Control and Prevention 2014.

[3] National Center for Health Statistics. National Ambulatory Medical Care Survey: 2015 State and National Summary Tables. Centers for Disease Control and Prevention 2015:29–31. https://doi.org/10.1080/0308107031000075690.

[4] Rui P, Okeyode T. National ambulatory medical care survey: 2016 National Summary Tabbles. Centers for Disease Control and Prevention 2016:1–11.

[5] Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, et al. Forecasting the future of cardiovascular disease in the United States: A policy statement from the American Heart Association. Circulation 2011;123:933–44. https://doi.org/10.1161/CIR.0b013e31820a55f5.

[6] Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, et al. Explaining the Decrease in U.S. Deaths from Coronary Disease, 1980–2000. New England Journal of Medicine 2007;356:2388–98. https://doi.org/10.1056/NEJMsa053935.

[7] (EHN) TEHN. Early detection of cardiovascular disease – an update from the European Heart Network – 2020 2020:1–21.

[8] Lo A, Ryder K, Shorr RI. Relationship between patient age and duration of physician visit in ambulatory setting: Does one size fit all? Journal of the American Geriatrics Society 2005;53:1162–7. https://doi.org/10.1111/j.1532-5415.2005.53367.x.

[9] Srinivas S, Ravi Ravindran A. Systematic review of opportunities to improve Outpatient appointment systems. 67th Annual Conference and Expo of the Institute of Industrial Engineers 2017 2017:1697–702.

[10] McClellan M, Brown N, Califf RM, Warner JJ. Call to Action: Urgent Challenges in Cardiovascular Disease: A Presidential Advisory from the American Heart Association. Circulation 2019;139:E44–54. https://doi.org/10.1161/CIR.0000000000000652.

[11] Bailey NTJ. A Study of Queues and Appointment Systems in Hospital Out-Patient Departments, with Special Reference to Waiting-Times. Journal of the Royal Statistical Society: Series B (Methodological) 1952;14:185–99. https://doi.org/10.1111/j.2517-6161.1952.tb00112.x.

[12] Cayirli T, Yang KK. A Universal Appointment Rule with Patient Classification for Service Times, No-Shows, and Walk-Ins. Service Science 2014;6:274–95. https://doi.org/10.1287/serv.2014.0087.

[13] Klassen KJ, Rohleder TR. Scheduling outpatient appointments in a dynamic environment. Journal of Operations Management 1996;14:83–101. https://doi.org/10.1016/0272-6963(95)00044-5.

[14] Rohleder TR, Klassen KJ. Using client-variance information to improve dynamic appointment scheduling performance. Omega 2000;28:293–302. https://doi.org/10.1016/S0305-0483(99)00040-7.

[15] Oppenheim GL, Bergman JJ, English EC. Failed appointments: a review. Journal of Family Practice 1979;8:789–96.

[16] White MJB, Pike MC. Appointment systems in out-patients' clinics and the effect of patients' unpunctuality. Medical Care 1964;2:133–45. https://doi.org/10.1097/00005650-196407000-00002.

[17] Fetter RB, Thompson JD. Patients' waiting time and doctors' idle time in the outpatient setting. Health Services Research 1966;1:66–90.

[18] Cox TF, Birchall JP, Wong H. Optimising the queuing system for an ear, nose and throat outpatient clinic. Journal of Applied Statistics 1985;12:113–26. https://doi.org/10.1080/02664768500000017.

[19] Toland B. No-shows cost health care system billions 2013.

[20] Gupta D, Denton B. Appointment scheduling in health care: Challenges and opportunities. IIE Transactions (Institute of Industrial Engineers) 2008;40:800–19. https://doi.org/10.1080/07408170802165880.

[21] Orton PK, Pereira Gray D. Factors influencing consultation length in general/family practice. Family Practice 2016;33:529–34. https://doi.org/10.1093/fampra/cmw056.

[22] Strahl JP. Patient appointment scheduling system: with supervised learning prediction. Master's Thesis, Aalto University 2015.

[23] Åman A. Predicting consultation durations in a digital primary care setting Predicting consultation durations in a digital primary care setting. 2018.

[24] Cayirli T, Veral E, Rosen H. Designing appointment scheduling systems for ambulatory care services. Health Care Management Science 2006;9:47–58. https://doi.org/10.1007/s10729-006-6279-5.

[25] Cayirli T, Veral E, Rosen H. Assessment of patient classification in appointment system design. Production and Operations Management 2008;17:338–53. https://doi.org/10.3401/poms.1080.0031.

[26] Berenson GS, Wattigney WA, Tracy RE, Newman WP, Srinivasan SR, Webber LS, et al. Atherosclerosis of the aorta and coronary arteries and cardiovascular risk factors in persons aged 6 to 30 years and studied at necropsy (the Bogalusa Heart Study). The American Journal of Cardiology 1992;70:851–8. https://doi.org/10.1016/0002-9149(92)90726-F.

[27] Berenson GS, Srinivasan SR, Bao W, Newman WP, Tracy RE, Wattigney WA. Association between Multiple Cardiovascular Risk Factors and Atherosclerosis in Children and Young Adults. New England Journal of Medicine 1998;338:1650–6. https://doi.org/10.1056/nejm199806043382302.

[28] Shrestha R, Vascular MC-C medicine reviews in, 2015 U. Long-Term Effects of Childhood Risk Factors on Cardiovascular Health During Adulthood. Clinical Medicine Reviews in Vascular Health 2015;7:1–5. https://doi.org/10.4137/cmrvh.s29964.

[29] Magnussen CG, Smith KJ, Juonala M. What the Long Term Cohort Studies that Began in Childhood Have Taught Us about the Origins of Coronary Heart Disease. Current Cardiovascular Risk Reports 2014;8:1–10. https://doi.org/10.1007/s12170-014-0373-x.

[30] Mikkilä V, Räsänen L, Laaksonen MML, Juonala M, Viikari J, Pietinen P, et al. Long-term dietary patterns and carotid artery intima media thickness: The Cardiovascular Risk in Young Finns Study. British Journal of Nutrition 2009;102:1507–12. https://doi.org/10.1017/S000711450999064X.

[31] Juonala M, Viikari JSA, Kähönen M, Taittonen L, Laitinen T, Hutri-Kähönen N, et al. Life-time risk factors and progression of carotid atherosclerosis in young adults: The Cardiovascular Risk in Young Finns study. European Heart Journal 2010;31:1745–51. https://doi.org/10.1093/eurheartj/ehq141.

[32] Tirosh A, Shai I, Afek A, Dubnov-Raz G, Ayalon N, Gordon B, et al. Adolescent BMI Trajectory and Risk of Diabetes versus Coronary Disease. New England Journal of Medicine 2011;364:1315–25. https://doi.org/10.1056/nejmoa1006992.

[33] Pulkki-Råback L, Puttonen S, Elovainio M, Raitakari OT, Juonala M, Keltikangas-Järvinen L. Adulthood EAS-temperament and carotid artery intima-media thickness: The Cardiovascular Risk in Young Finns Study. Psychology and Health

2011;26:61–75. https://doi.org/10.1080/08870440903270690.

[34] Juhola J, Magnussen CG, Berenson GS, Venn A, Burns TL, Sabin MA, et al. Combined effects of child and adult elevated blood pressure on subclinical atherosclerosis: The international childhood cardiovascular cohort consortium. Circulation 2013;128:217–24. https://doi.org/10.1161/CIRCULATIONAHA.113.001614.

[35] Raitakari OT, Juonala M, Kähönen M, Taittonen L, Laitinen T, Mäki-Torkko N, et al. Cardiovascular Risk Factors in Childhood and Carotid Artery Intima-Media Thickness in Adulthood: The Cardiovascular Risk in Young Finns Study. Journal of the American Medical Association 2003;290:2277–83. https://doi.org/10.1001/jama.290.17.2277.

[36] Cheng HM, Ye ZX, Charng MJ. Association of pathobiologic determinants of atherosclerosis in youth risk score and carotid artery intima-media thickness in asymptomatic young heterozygous familial hypercholesterolemia patients. Acta Cardiologica Sinica 2011;27:152–7.

[37] Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. New England Journal of Medicine 2016;375:1216–9. https://doi.org/10.1056/nejmp1606181.

[38] Berglund E, Lytsy P, Westerling R. Adherence to and beliefs in lipid-lowering medical treatments: A structural equation modeling approach including the necessity-concern framework. Patient Education and Counseling 2013;91:105–12. https://doi.org/10.1016/j.pec.2012.11.001.

[39] Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. BMC Medical Informatics and Decision Making 2011;11. https://doi.org/10.1186/1472-6947-11-51.

[40] Ayer T, Chhatwal J, Alagoz O, Kahn CE, Woods RW, Burnside ES. Informatics in radiology: Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. Radiographics 2010;30:13–22. https://doi.org/10.1148/rg.301095057.

[41] Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: The Reynolds Risk Score. Journal of the American Medical Association 2007;297:611–9. https://doi.org/10.1001/jama.297.6.611.

[42] Cayirli T, Veral E. Outpatient scheduling in health care: A review of literature. Production and Operations Management 2003;12:519–49. https://doi.org/10.1111/j.1937-5956.2003.tb00218.x.

[43] Welcht JD. Appointments Systems in Hospitals and General Practice: Appointment Systems in Hospital Outpatient Departments. Journal of the Operational Research Society 1964;15:224–32. https://doi.org/10.1057/jors.1964.43.

[44] Rockart JF, Hofmann PB. Physician and patient behavior under different scheduling systems in a hospital outpatient department. Medical Care 1969;7:463–70. https://doi.org/10.1097/00005650-196911000-00005.

[45] Ho C-J, Lau H-S. Minimizing Total Cost in Scheduling Outpatient Appointments. Management Science 1992;38:1750–64. https://doi.org/10.1287/mnsc.38.12.1750.

[46] Fries BE, Marathe VP. Determination of optimal variable-sized multiple-block appointment systems. Operations Research 1981;29:324–45. https://doi.org/10.1287/opre.29.2.324.

[47] Choi S (Sam), Banerjee A (Andy). Comparison of a branch-and-bound heuristic, a newsvendor-based heuristic and periodic Bailey rules for outpatients appointment scheduling systems. Journal of the Operational Research Society 2016;67:576–92. https://doi.org/10.1057/jors.2015.79.

[48] Ho CJ, Lau HS, Li J. Introducing variable-interval appointment scheduling rules in service systems. International Journal of Operations and Production Management 1995;15:59–68. https://doi.org/10.1108/01443579510090345.

[49] Srinivas S, Ravindran AR. Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework. Expert Systems with Applications 2018;102:245–61. https://doi.org/10.1016/j.eswa.2018.02.022.

[50] Chen Y, Kuo YH, Fan P, Balasubramanian H. Appointment overbooking with different time slot structures. Computers and Industrial Engineering 2018;124:237–48. https://doi.org/10.1016/j.cie.2018.07.021.

[51] Robinson LW, Chen RR. Scheduling doctors' appointments: Optimal and empirically-based heuristic policies. IIE Transactions (Institute of Industrial Engineers) 2003;35:295–307. https://doi.org/10.1080/07408170304367.

[52] Yan C, Tang J, Jiang B, Fung RYK. Sequential appointment scheduling considering patient choice and service fairness. International Journal of Production Research 2015;53:7376–95. https://doi.org/10.1080/00207543.2015.1081426.

[53] Lehaney B, Clarke SA, Paul RJ. A case of an intervention in an outpatients department. Journal of the Operational Research Society 1999;50:877–91. https://doi.org/10.1057/palgrave.jors.2600796.

[54] Lau HS, Lau AHL. A fast procedure for computing the total system cost of an

appointment schedule for medical and kindred facilities. IIE Transactions (Institute of Industrial Engineers) 2000;32:833–9. https://doi.org/10.1080/07408170008967442.

[55] Erdogan SA, Gose A, Denton BT. Online appointment sequencing and scheduling. IIE Transactions (Institute of Industrial Engineers) 2015;47:1267–86. https://doi.org/10.1080/0740817X.2015.1011355.

[56] Laganga LR, Lawrence SR. Clinic overbooking to improve patient access and increase provider productivity. Decision Sciences 2007;38:251–76. https://doi.org/10.1111/j.1540-5915.2007.00158.x.

[57] Laganga LR, Lawrence SR. Appointment Scheduling With Overbooking To Mitigate Productivity Loss From No-Shows. Decision Sciences 2007;38:1–29.

[58] Muthuraman K, Lawley M. A stochastic overbooking model for outpatient clinical scheduling with no-shows. IIE Transactions 2008;40:820–37. https://doi.org/10.1080/07408170802165823.

[59] Srinivas S. A machine learning-based approach for predicting patient punctuality in ambulatory care centers. International Journal of Environmental Research and Public Health 2020;17:3703. https://doi.org/10.3390/ijerph17103703.

[60] Daggy J, Lawley M, Willis D, Thayer D, Suelzer C, Delaurentis PC, et al. Using no-show modeling to improve clinic performance. Health Informatics Journal 2010;16:246–59. https://doi.org/10.1177/1460458210380521.

[61] Zacharias C, Pinedo M. Appointment scheduling with no-shows and overbooking. Production and Operations Management 2014;23:788–801. https://doi.org/10.1111/poms.12065.

[62] Stephan P. Kudyba. Healthcare Informatics: Improving Efficiency and Productivity - Stephan P. Kudyba - Google Books 2010:215.

[63] Feldman B, Martin EM, Skotnes T. Big Data in Healthcare Hype and Hope. 2012.

[64] Fernandes LM, O'Connor M, Weaver V. Big Data, Bigger Outcomes. Journal of AHIMA 2012;83:38–43.

[65] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology 2017;2:230–43. https://doi.org/10.1136/svn-2017-000101.

[66] Abdel-Basset M, Manogaran G, Gamal A, Smarandache F. A Group Decision Making Framework Based on Neutrosophic TOPSIS Approach for Smart Medical Device Selection. Journal of Medical Systems 2019;43.

https://doi.org/10.1007/s10916-019-1156-1.

[67] Huntley DA, Cho DW, Christman J, Csernansky JG. Predicting length of stay in an acute psychiatric hospital. Psychiatric Services 1998;49:1049–53. https://doi.org/10.1176/ps.49.8.1049.

[68] Pradhan N, Rani G, Dhaka VS, Poonia RC. Diabetes prediction using artificial neural network. Deep Learning Techniques for Biomedical and Health Informatics, vol. 121, 2020, p. 327–39. https://doi.org/10.1016/B978-0-12-819061-6.00014-8.

[69] Waghulde NP, Patil NP. Genetic Neural Approach for Heart Disease Prediction. International Journal of Advanced Computer Research 2014;4:778–84.

[70] Viju Raghupathi WR. An Overview of Health Analytics. Journal of Health & Medical Informatics 2013;04. https://doi.org/10.4172/2157-7420.1000132.

[71] Bentayeb D, Lahrichi N, Rousseau LM. Patient scheduling based on a service-time prediction model: a data-driven study for a radiotherapy center. Health Care Management Science 2019;22:768–82. https://doi.org/10.1007/s10729-018-9459-1.

[72] David Blumenthal, M., Culpepper, L., Stafford, R., & Starfield B. Duration Of Ambulatry Visits. The Journal of Family Practice 1999;48.

[73] Hajebrahimi S, Janati A, Arab-Zozani M, Sokhanvar M, Haghgoshayie E, Siraneh Y, et al. Medical visit time and predictors in health facilities: a mega systematic review and meta-analysis. International Journal of Human Rights in Healthcare 2019;12:373–402. https://doi.org/10.1108/IJHRH-05-2019-0036.

[74] Pankevich V. Patient demographics as a predictive tool of consultation duration. London Journal of Primary Care 2014;6:79–83. https://doi.org/10.1080/17571472.2014.11493421.

[75] Stevens S, Bankhead C, Mukhtar T, Perera-Salazar R, Holt TA, Salisbury C, et al. Patient-level and practice-level factors associated with consultation duration: A cross-sectional analysis of over one million consultations in English primary care. BMJ Open 2017;7:1–7. https://doi.org/10.1136/bmjopen-2017-018261.

[76] Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: A systematic review of 67 countries. BMJ Open 2017;7. https://doi.org/10.1136/bmjopen-2017-017902.

[77] Kabeya Y, Uchida J, Toyoda M, Katsuki T, Oikawa Y, Kato K, et al. Factors affecting consultation length in a Japanese diabetes practice. Diabetes Research and Clinical Practice 2017;126:54–9. https://doi.org/10.1016/j.diabres.2016.12.020.

[78] Åman A. Predicting consultation durations in a digital primary care setting Predicting consultation durations in a digital primary care setting 2018.

[79] Bentayeb D, Lahrichi N, Rousseau LM. Patient scheduling based on a service-time prediction model: a data-driven study for a radiotherapy center. Health Care Management Science 2019;22:768–82. https://doi.org/10.1007/s10729-018-9459-1.

[80] Strahl JP. Patient appointment scheduling system: with supervised learning prediction. 2015.

[81] Belotti F, Deb P, Manning WG, Norton EC. twopm: Two-part models. Stata Journal 2015;15:3–20. https://doi.org/10.1177/1536867x1501500102.

[82] Sheikhpour R, Sarram MA, Gharaghani S, Chahooki MAZ. A Survey on semi-supervised feature selection methods. Pattern Recognition 2017;64:141–58. https://doi.org/10.1016/j.patcog.2016.11.003.

[83] Burnap A, Pan Y, Liu Y, Ren Y, Lee H, Gonzalez R, et al. Improving Design Preference Prediction Accuracy Using Feature Learning. Journal of Mechanical Design, Transactions of the ASME 2016;138. https://doi.org/10.1115/1.4033427.

[84] Ferreira I, Van De Laar RJ, Prins MH, Twisk JW, Stehouwer CD. Carotid stiffness in young adults: A life-course analysis of its early determinants: The Amsterdam growth and health longitudinal study. Hypertension 2012;59:54–61. https://doi.org/10.1161/HYPERTENSIONAHA.110.156109.

[85] Ferreira I, Twisk JWR, Van Mechelen W, Kemper HCG, Seidell JC, Stehouwer CDA. Current and adolescent body fatness and fat distribution: Relationships with carotid intima-media thickness and large artery stiffness at the age of 36 years. Journal of Hypertension 2004;22:145–55. https://doi.org/10.1097/00004872-200401000-00024.

[86] Van De Laar RJJ, Stehouwer CDA, Boreham CA, Murray LM, Schalkwijk CG, Prins MH, et al. Continuing smoking between adolescence and young adulthood is associated with higher arterial stiffness in young adults: The Northern Ireland Young Hearts Project. Journal of Hypertension 2011;29:2201–9. https://doi.org/10.1097/HJH.0b013e32834b0ecf.

[87] Ried-Larsen M, Grontved A, Kristensen PL, Froberg K, Andersen LB. Moderate-and-vigorous physical activity from adolescence to adulthood and subclinical atherosclerosis in adulthood: Prospective observations from the European Youth Heart Study. British Journal of Sports Medicine 2015;49:107–12. https://doi.org/10.1136/bjsports-2013-092409.

[88] Connelly CD, Hazen AL, Baker-Ericzén MJ, Landsverk J, Horwitz SMC. Is screening for depression in the perinatal period enough? The co-occurrence of depression, substance abuse, and intimate partner violence in culturally diverse pregnant women. Journal of Women's Health 2013;22:844–52. https://doi.org/10.1089/jwh.2012.4121.

[89] Devries KM, Mak JY, Bacchus LJ, Child JC, Falder G, Petzold M, et al. Intimate Partner Violence and Incident Depressive Symptoms and Suicide Attempts: A Systematic Review of Longitudinal Studies. PLoS Medicine 2013;10. https://doi.org/10.1371/journal.pmed.1001439.

[90] Chuang CH, Cattoi AL, McCall-Hosenfeld JS, Camacho F, Dyer AM, Weisman CS. Longitudinal association of intimate partner violence and depressive symptoms. Mental Health in Family Medicine 2012;9:107–14.

[91] Schultz WM, Kelli HM, Lisko JC, Varghese T, Shen J, Sandesara P, et al. Socioeconomic status and cardiovascular outcomes: Challenges and interventions. Circulation 2018;137:2166–78. https://doi.org/10.1161/CIRCULATIONAHA.117.029652.

[92] Mosquera PA, San Sebastian M, Waenerlund AK, Ivarsson A, Weinehall L, Gustafsson PE. Income-related inequalities in cardiovascular disease from mid-life to old age in a Northern Swedish cohort: A decomposition analysis. Social Science and Medicine 2016;149:135–44. https://doi.org/10.1016/j.socscimed.2015.12.017.

[93] Kucharska-Newton AM, Harald K, Rosamond WD, Rose KM, Rea TD, Salomaa V. Socioeconomic Indicators and the Risk of Acute Coronary Heart Disease Events: Comparison of Population-Based Data from the United States and Finland. Annals of Epidemiology 2011;21:572–9. https://doi.org/10.1016/j.annepidem.2011.04.006.

[94] Meneton P, Kesse-Guyot E, Méjean C, Fezeu L, Galan P, Hercberg S, et al. Unemployment is associated with high cardiovascular event rate and increased all-cause mortality in middle-aged socially privileged individuals. International Archives of Occupational and Environmental Health 2015;88:707–16. https://doi.org/10.1007/s00420-014-0997-7.

[95] Roux AVD, Merkin SS, Arnett D, Chambless L, Massing M, Nieto FJ, et al. Neighborhood of Residence and Incidence of Coronary Heart Disease. New England Journal of Medicine 2001;345:99–106. https://doi.org/10.1056/nejm200107123450205.

[96] Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the Prospective Cardiovascular Münster (PROCAM) study. Circulation 2002;105:310–5. https://doi.org/10.1161/hc0302.102575.

[97] Ridker PM, Paynter NP, Rifai N, Gaziano JM, Cook NR. C-reactive protein and parental history improve global cardiovascular risk prediction: The Reynolds risk score for men. Circulation 2008;118:2243–51. https://doi.org/10.1161/CIRCULATIONAHA.108.814251.

[98] Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. European Heart Journal 2003;24:987–1003. https://doi.org/10.1016/S0195-668X(03)00114-3.

[99] Pencina MJ, D'Agostino RB, Larson MG, Massaro JM, Vasan RS. Predicting the 30-year risk of cardiovascular disease: The framingham heart study. Circulation 2009;119:3078–84. https://doi.org/10.1161/CIRCULATIONAHA.108.816694.

[100] Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, Naghavi M. Machine learning outperforms ACC/AHA CVD risk calculator in MESA. Journal of the American Heart Association 2018;7. https://doi.org/10.1161/JAHA.118.009476.

[101] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can Machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE 2017;12. https://doi.org/10.1371/journal.pone.0174944.

[102] Kim JO, Jeong YS, Kim JH, Lee JW, Park D, Kim HS. Machine learning-based cardiovascular disease prediction model: A cohort study on the korean national health insurance service health screening database. Diagnostics 2021;11. https://doi.org/10.3390/diagnostics11060943.

[103] Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: A methodology review. Journal of Biomedical Informatics 2002;35:352–9. https://doi.org/10.1016/S1532-0464(03)00034-0.

[104] Sastry SH, Babu PMSP. Implementation of CRISP Methodology for ERP Systems 2013;2:203–17.

[105] Wirth R. CRISP-DM : Towards a Standard Process Model for Data Mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining 2000:29–39. https://doi.org/10.1.1.198.5133.

[106] Buuren AS Van. The mice Package 2007.

[107] U.S. Census Bureau. Explore Census Data. CensusGov 2019. https://data.census.gov/cedsci/ (accessed July 19, 2020).

[108] Bonito AJ, Bann C, Eicheldinger MS CL. Creation of New Race-Ethnicity Codes and Socioeconomic Status (SES) Indicators for Medicare Beneficiaries. Final

Report. Rockville (MD): Agency for Healthcare Research and Quality. 2008.

[109] Becker D. Using Categorical Data with One Hot Encoding 2018.

[110] Povak NA, Hessburg PF, Reynolds KM, Sullivan TJ, McDonnell TC, Salter RB. Machine learning and hurdle models for improving regional predictions of stream water acid neutralizing capacity. Water Resources Research 2013;49:3531–46. https://doi.org/10.1002/wrcr.20308.

[111] Kuhn M. Variable importance using the caret package. Journal of Statistical Software 2007.

[112] Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognition 2015;48:2839–46. https://doi.org/10.1016/j.patcog.2015.03.009.

[113] Ghojogh B, Crowley M. The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial 2019:1–23.

[114] Breiman L. Random forests. Machine Learning 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.

[115] Fawagreh K, Gaber MM, Elyan E. Random forests: From early developments to recent advancements. Systems Science and Control Engineering 2014;2:602–9. https://doi.org/10.1080/21642583.2014.956265.

[116] Wiener AL and M. Classification and Regression by randomForest. R News 2. ResearchgateNet 2003;3:18–22.

[117] Natekin A, Knoll A. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics 2013;7. https://doi.org/10.3389/fnbot.2013.00021.

[118] Svozil D, Kvasnieka V, Pospichal J. Chemometrics and intelligent laboratory systems Introduction to multi-layer feed-forward neural networks. vol. 39. 1997.

[119] Schmidhuber J. Deep Learning in neural networks: An overview. Neural Networks 2015;61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

[120] Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011 2011:1–9.

[121] Oshiro TM, Perez PS. How Many Trees in a Random Forest ? How Many Trees in a Random Forest ? 2016:154–68. https://doi.org/10.1007/978-3-642-31537-4.

[122] Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for

random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2019;9. https://doi.org/10.1002/widm.1301.

[123] Friedman JH. Stochastic gradient boosting. Computational Statistics and Data Analysis 2002;38:367–78. https://doi.org/10.1016/S0167-9473(01)00065-2.

[124] Karsoliya S. Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. International Journal of Engineering Trends and Technology 2012;3:714–7.

[125] Bengio Y. Practical recommendations for gradient-based training of deep architectures. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2012;7700 LECTU:437–78. https://doi.org/10.1007/978-3-642-35289-8-26.

[126] Keskar NS, Nocedal J, Tang PTP, Mudigere D, Smelyanskiy M. On large-batch training for deep learning: Generalization gap and sharp minima. 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings 2019:1–16.

[127] Narkhede S. Understanding AUC-ROC Curve. Towards Data Science 2018.

[128] Davis J, Goadrich M. The relationship between precision-recall and ROC curves. ACM International Conference Proceeding Series, vol. 148, 2006, p. 233–40. https://doi.org/10.1145/1143844.1143874.

[129] Botchkarev A. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology 2018.

[130] Srinivas S. Evaluating the impact of nature of patient flow and patient availability on the performance of appointment scheduling rules in outpatient clinics. International Journal of Operations and Quantitative Management 2016;22:93–118.

[131] Srinivas S, Ravindran AR. Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework. Expert Systems with Applications 2018;102:245–61. https://doi.org/10.1016/j.eswa.2018.02.022.

[132] Dantas LF, Fleck JL, Cyrino Oliveira FL, Hamacher S. No-shows in appointment scheduling – a systematic literature review. Health Policy 2018;122:412–21. https://doi.org/10.1016/j.healthpol.2018.02.002.

[133] Srinivas S. A Machine Learning-Based Approach for Predicting Patient Punctuality in Ambulatory Care Centers. International Journal of Environmental Research and Public Health 2020;17. https://doi.org/10.3390/ijerph17103703.

[134] Kempny A, Diller GP, Dimopoulos K, Alonso-Gonzalez R, Uebing A, Li W, et al.

Determinants of outpatient clinic attendance amongst adults with congenital heart disease and outcome. International Journal of Cardiology 2016;203:245–50. https://doi.org/10.1016/j.ijcard.2015.10.081.

[135] Michel J, Sangha D, III JE. Burnout among cardiologists 2017.

[136] Smith DF. Negative emotions and coronary heart disease: Causally related or merely coexistent? A review. Scandinavian Journal of Psychology 2001;42:57–69. https://doi.org/10.1111/1467-9450.00214.

[137] Steinwachs DM, Collins-Nakai RL, Cohn LH, Garson A, Wolk MJ. The Future of Cardiology: Utilization and Costs of Care. 2000. https://doi.org/10.1016/S0735-1097(00)00559-3.

[138] Duggal R, Brindle I, Bagenal J. Digital healthcare: Regulating the revolution. BMJ (Online) 2018;360:1–2. https://doi.org/10.1136/bmj.k6.

[139] Khan MFF, Sakamura K. Fine-grained access control to medical records in digital healthcare enterprises. 2015 International Symposium on Networks, Computers and Communications, ISNCC 2015 2015:1–6. https://doi.org/10.1109/ISNCC.2015.7238590.

[140] Holtzclaw J, Clear R, Dittmar H, Goldstein D, Haas P. Location efficiency: Neighborhood and socio-economic characteristics determine auto ownership and use - Studies in Chicago, Los Angeles and San Francisco. Transportation Planning and Technology 2002;25:1–27. https://doi.org/10.1080/03081060290032033.

[141] Cole R, Leslie E, Bauman A, Donald M, Owen N. Socio-demographic variations in walking for transport and for recreation or exercise among adult Australians. Journal of Physical Activity and Health 2006;3:164–78. https://doi.org/10.1123/jpah.3.2.164.

[142] CAYIRLI T, VERAL E. OUTPATIENT SCHEDULING IN HEALTH CARE: A REVIEW OF LITERATURE. Production and Operations Management 2009;12:519–49. https://doi.org/10.1111/j.1937-5956.2003.tb00218.x.

[143] Han J. Data Mining: Concepts and Techniques-Chapter 6. 2018.

[144] Learned-miller EG. Introduction to Supervised Learning. I: Department of Computer Science, University of Massachusetts 2014.

[145] Srinivas S, Salah H. Consultation length and no-show prediction for improving appointment scheduling efficiency at a cardiology clinic: A data analytics approach. International Journal of Medical Informatics 2021;145:104290. https://doi.org/10.1016/j.ijmedinf.2020.104290.

[146] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 13-17- Augu, Association for Computing Machinery; 2016, p. 785–94. https://doi.org/10.1145/2939672.2939785.

[147] Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognition 2015;48:2839–46. https://doi.org/10.1016/j.patcog.2015.03.009.

[148] Ghojogh B, Crowley M. The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial 2019:1–23.

[149] Botchkarev A. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology 2018.

[150] Barrera Ferro D, Brailsford S, Bravo C, Smith H. Improving healthcare access management by predicting patient no-show behaviour. Decision Support Systems 2020;138. https://doi.org/10.1016/j.dss.2020.113398.

[151] Partridge JW. Consultation time, workload, and problems for audit in outpatients clinics. Archives of Disease in Childhood 1992;67:206–10. https://doi.org/10.1136/adc.67.2.206.

[152] Laganga LR, Lawrence SR. Appointment overbooking in health care clinics to improve patient service and clinic performance. Production and Operations Management 2012;21:874–88. https://doi.org/10.1111/j.1937-5956.2011.01308.x.

[153] Rising EJ, Baron R, Averill B. SYSTEMS ANALYSIS OF A UNIVERSITY-HEALTH-SERVICE OUTPATIENT CLINIC. Operations Research 1973;21:1030–47. https://doi.org/10.1287/opre.21.5.1030.

[154] Soltani M, Samorani M, Kolfal B. Appointment scheduling with multiple providers and stochastic service times. European Journal of Operational Research 2019;277:667–83. https://doi.org/10.1016/j.ejor.2019.02.051.

[155] Zacharias C, Pinedo M. Managing customer arrivals in service systems with multiple identical servers. Manufacturing and Service Operations Management 2017;19:639–56. https://doi.org/10.1287/msom.2017.0629.

[156] Harris KM, Richard J. Udry. National Longitudinal Study of Adolescent to Adult Health (Add Health) Wave I – Wave V, 1994-2018 2019.

[157] S VB, K G-O. mice : Multivariate imputation by chained equations in R. Journal of Statistical Software 2011;45.

[158] Kim J, Kim R, Oh H, Lippert AM, Subramanian S V. Estimating the influence of adolescent delinquent behavior on adult health using sibling fixed effects. Social

Science and Medicine 2020;265:113397. https://doi.org/10.1016/j.socscimed.2020.113397.

[159] Lee TK, Wickrama KAS, O'Neal CW. How Early Stressful Life Experiences Combine With Adolescents' Conjoint Health Risk Trajectories to Influence Cardiometabolic Disease Risk in Young Adulthood. Journal of Youth and Adolescence 2021;50:1234–53. https://doi.org/10.1007/s10964-021-01440-0.

[160] Noppert GA, Gaydosh L, Harris KM, Goodwin A, Hummer RA. Is educational attainment associated with young adult cardiometabolic health? SSM - Population Health 2021;13:100752. https://doi.org/10.1016/j.ssmph.2021.100752.

[161] Stewart SD, Menning CL. Family Structure, Nonresident Father Involvement, and Adolescent Eating Patterns. Journal of Adolescent Health 2009;45:193–201. https://doi.org/10.1016/j.jadohealth.2009.01.005.

[162] Brunet J, Sabiston CM, O'Loughlin E, Chaiton M, Low NCP, O'Loughlin JL. Symptoms of depression are longitudinally associated with sedentary behaviors among young men but not among young women. Preventive Medicine 2014;60:16–20. https://doi.org/10.1016/j.ypmed.2013.12.003.

[163] Hoyt LT, Chase-Lansdale PL, McDade TW, Adam EK. Positive youth, healthy adults: Does positive well-being in adolescence predict better perceived health and fewer risky health behaviors in young adulthood? Journal of Adolescent Health 2012;50:66–73. https://doi.org/10.1016/j.jadohealth.2011.05.002.

[164] Yıldız M. Stressful life events and adolescent suicidality: An investigation of the mediating mechanisms. Journal of Adolescence 2020;82:32–40. https://doi.org/10.1016/j.adolescence.2020.05.006.

[165] Pressman SD, Cohen S. Does positive affect influence health? Psychological Bulletin 2005;131:925–71. https://doi.org/10.1037/0033-2909.131.6.925.

[166] Sheehan TJ, Fifield J, Reisine S, Tennen H. The Measurement Structure of the Center for Epidemiologic Studies Depression Scale. Journal of Personality Assessment 1995;64:507–21. https://doi.org/10.1207/s15327752jpa6403_9.

[167] Rosenberg M. Society and the adolescent self-image. 2015. https://doi.org/10.2307/2575639.

[168] Resnick MD, Bearman PS, Blum RW, Bauman KE, Harris KM, Jones J, et al. Protecting adolescent's from harm: Findings from the national longitudinal study on adolescent health. Journal of the American Medical Association 1997;278:823–32. https://doi.org/10.1001/jama.278.10.823.

[169] Sandler AD. A prospective study of the role of depression in the development and persistence of adolescent obesity. Journal of Developmental and Behavioral

Pediatrics 2003;24:81. https://doi.org/10.1097/00004703-200302000-00026.

[170] Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement 1977;1:385–401. https://doi.org/10.1177/014662167700100306.

[171] Noppert Lauren;Harris, Kathleen Mullan;Goodwin, Andrea;Hummer, Robert A. GA ;Gaydos. Is educational attainment associated with young adult cardiometabolic health? 2021;13. https://doi.org/10.1016/j.ssmph.2021.100752.

[172] Hatzenbuehler ML, Slopen N, McLaughlin KA. Stressful life events, sexual orientation, and cardiometabolic risk among young adults in the United States. Health Psychology 2014;33:1185–94. https://doi.org/10.1037/hea0000126.

[173] Adkins DE, Wang V, Dupre ME, Van Den Oord EJCG, Elder GH. Structure and stress: Trajectories of depressive symptoms across adolescence and young adulthood. Social Forces 2009;88:31–60. https://doi.org/10.1353/sof.0.0238.

[174] Hoyt LT, Chase-Lansdale PL, McDade TW, Adam EK. Positive youth, healthy adults: Does positive well-being in adolescence predict better perceived health and fewer risky health behaviors in young adulthood? Journal of Adolescent Health 2012;50:66–73. https://doi.org/10.1016/j.jadohealth.2011.05.002.

[175] Clark CJ, Alonso A, Spencer RA, Pencina M, Williams K, Everson-Rose SA. Predicted long-term cardiovascular risk among young adults in the national longitudinal study of adolescent health. American Journal of Public Health 2014;104:e108–15. https://doi.org/10.2105/AJPH.2014.302148.

[176] Jama EEP on D-, 2001 U. Report of the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Archives of Internal Medicine 1988;148:36–69. https://doi.org/10.1001/archinte.1988.00380010040006.

[177] Ogunleye A, Wang QG. XGBoost Model for Chronic Kidney Disease Diagnosis. IEEE/ACM Transactions on Computational Biology and Bioinformatics 2020;17:2131–40. https://doi.org/10.1109/TCBB.2019.2911071.

[178] Rallapalli S, Suryakanthi T. Predicting the risk of diabetes in big data electronic health Records by using scalable random forest classification algorithm. Proceedings - 2016 3rd International Conference on Advances in Computing, Communication and Engineering, ICACCE 2016, 2017, p. 281–4. https://doi.org/10.1109/ICACCE.2016.8073762.

[179] Scoralick JP, Iwashima GC, Colugnati FAB, Goliatt L, Capriles PVSZ. A Extreme Gradient Boosting Classifier for Predicting Chronic Kidney Disease Stages.

Springer, 2021, p. 901–10. https://doi.org/10.1007/978-3-030-71187-0_83.

[180] Jabbar MA, Deekshatulu BL, Chndra P. Alternating decision trees for early diagnosis of heart disease. Proceedings of International Conference on Circuits, Communication, Control and Computing, I4C 2014, 2014, p. 322–8. https://doi.org/10.1109/CIMCA.2014.7057816.

[181] Jabbar MA, Deekshatulu BL, Chandra P. Prediction of heart disease using random forest and feature subset selection. Advances in Intelligent Systems and Computing, vol. 424, Springer Verlag; 2016, p. 187–96. https://doi.org/10.1007/978-3-319-28031-8_16.

[182] Rath A, Mishra D, Panda G, Satapathy SC. Heart disease detection using deep learning methods from imbalanced ECG samples. Biomedical Signal Processing and Control 2021;68. https://doi.org/10.1016/j.bspc.2021.102820.

[183] Budholiya K, Shrivastava SK, Sharma V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. Journal of King Saud University - Computer and Information Sciences 2020. https://doi.org/10.1016/j.jksuci.2020.10.013.

[184] Huang YC, Li SJ, Chen M, Lee TS, Chien YN. Machine-learning techniques for feature selection and prediction of mortality in elderly CABG patients. Healthcare (Switzerland) 2021;9. https://doi.org/10.3390/healthcare9050547.

[185] Stylianou N, Akbarov A, Kontopantelis E, Buchan I, Dunn KW. Mortality risk prediction in burn injury: Comparison of logistic regression with machine learning approaches. Burns 2015;41:925–34. https://doi.org/10.1016/j.burns.2015.03.016.

[186] Karthikeyan A, Garg A, Vinod PK, Priyakumar UD. Machine Learning Based Clinical Decision Support System for Early COVID-19 Mortality Prediction. Frontiers in Public Health 2021;9. https://doi.org/10.3389/fpubh.2021.626697.

[187] Kingsford C, Salzberg SL. What are decision trees? Nature Biotechnology 2008;26:1011–2. https://doi.org/10.1038/nbt0908-1011.

[188] Rokach L, Maimon O. Decision Trees. Data Mining and Knowledge Discovery Handbook, Springer-Verlag; 2006, p. 165–92. https://doi.org/10.1007/0-387-25465-x_9.

[189] Azhagusundari B, Thanamani AS. Feature Selection based on Information Gain. International Journal of Innovative Technology and Exploring Engineering (IJITEE) 2013;2:18–21.

[190] Liaw A, Wiener M. Classification and Regression by randomForest. R News 2002;2:18–22.

[191] Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of Statistics 2001;29:1189–232. https://doi.org/10.2307/2699986.

[192] Johnson-Carlson P, Costanzo C, Kopetsky DBT-NE. Predictive staffing simulation model methodology 2017;35:161+.

[193] Chang Y, Chang K, Computing GW-AS, 2018 undefined. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. Elsevier n.d.

[194] Moolayil J. An Introduction to Deep Learning and Keras. Learn Keras for Deep Neural Networks, Apress; 2019, p. 1–16. https://doi.org/10.1007/978-1-4842-4240-7_1.

[195] Miller T. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence 2019;267:1–38. https://doi.org/10.1016/j.artint.2018.07.007.

[196] Du M, Liu N, Hu X. Techniques for interpretable machine learning. Communications of the ACM 2020;63:68–77. https://doi.org/10.1145/3359786.

[197] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, vol. 2017- Decem, 2017, p. 4766–75.

[198] Shapley LS. A value for n-person games. Contributions to TheTheory of Games 1953;2:07–317. https://doi.org/10.2307/j.ctv173f1fh.12.

[199] Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems 2014;41:647–65. https://doi.org/10.1007/s10115-013-0679-x.

[200] Lubo-Robles D, Devegowda D, Jayaram V, Bedle H, Marfurt KJ, Pranter MJ. Machine learning model interpretability using SHAP values: Application to a seismic facies classification task. SEG Technical Program Expanded Abstracts, vol. 2020- Octob, 2020, p. 1460–4. https://doi.org/10.1190/segam2020-3428275.1.

[201] Dhamdhere K, Agarwal A, Sundararajan M. The shapley taylor interaction index. 37th International Conference on Machine Learning, ICML 2020, vol. PartF16814, International Machine Learning Society (IMLS); 2020, p. 9196–209.

[202] Steel M. SHAP Force Plots for Classification. MLearningAi 2021. https://medium.com/mlearning-ai/shap-force-plots-for-classification-d30be430e195 (accessed March 22, 2022).

[203] Fazakis N, Kocsis O, Dritsas E, Alexiou S, Fakotakis N, Moustakas K. Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction. IEEE Access 2021;9:103737–57. https://doi.org/10.1109/ACCESS.2021.3098691.

[204] Sekercioglu N, Fu R, Kim SJ, Mitsakakis N. Machine learning for predicting long-term kidney allograft survival: a scoping review. Irish Journal of Medical Science 2021;190:807–17. https://doi.org/10.1007/s11845-020-02332-1.

[205] Razavi AC, Wong N, Budoff M, Bazzano LA, Kelly TN, He J, et al. Predicting Long-Term Absence of Coronary Artery Calcium in Metabolic Syndrome and Diabetes: The MESA Study. JACC: Cardiovascular Imaging 2021;14:219–29. https://doi.org/10.1016/j.jcmg.2020.06.047.

[206] Wang SY, Tan ASL, Claggett B, Chandra A, Khatana SAM, Lutsey PL, et al. Longitudinal Associations between Income Changes and Incident Cardiovascular Disease: The Atherosclerosis Risk in Communities Study. JAMA Cardiology 2019;4:1203–12. https://doi.org/10.1001/jamacardio.2019.3788.

[207] Same R V., Feldman DI, Shah N, Martin SS, Al Rifai M, Blaha MJ, et al. Relationship Between Sedentary Behavior and Cardiovascular Risk. Current Cardiology Reports 2016;18:1–7. https://doi.org/10.1007/s11886-015-0678-5.

[208] Sakata K, Matumura Y, Yoshimura N, Tamaki J, Hashimoto T, Oguri S, et al. Relationship between skipping breakfast and cardiovascular disease risk factors in the national nutrition survey data. [Nippon Kōshū Eisei Zasshi] Japanese Journal of Public Health 2001;48:837–41.

[209] Keppel CC, Crowe SF. Changes to body image and self-esteem following stroke in young adults. Neuropsychological Rehabilitation 2000;10:15–31. https://doi.org/10.1080/096020100389273.

[210] Srinivas S, Anand K, Chockalingam A. Longitudinal association between adolescent negative emotions and adulthood cardiovascular disease risk: an opportunity for healthcare quality improvement. Benchmarking 2020;27:2323–39. https://doi.org/10.1108/BIJ-01-2020-0028.

[211] Sun X, Ram N, McHale SM. Adolescent Family Experiences Predict Young Adult Educational Attainment: A Data-Based Cross-Study Synthesis With Machine Learning. Journal of Child and Family Studies 2020;29:2770–85. https://doi.org/10.1007/s10826-020-01775-5.

[212] Keski-Rahkonen A, Kaprio J, Rissanen A, Virkkunen M, Rose RJ. Breakfast skipping and health-compromising behaviors in adolescents and adults. European Journal of Clinical Nutrition 2003;57:842–53. https://doi.org/10.1038/sj.ejcn.1601618.

[213] O'Neil A, Scovelle AJ, Milner AJ, Kavanagh A. Gender/sex as a social determinant of cardiovascular risk. Circulation 2018;137:854–64. https://doi.org/10.1161/CIRCULATIONAHA.117.028595.

[214] Schwappach DLB, Boluarte TA, Suhrcke M. The economics of primary prevention of cardiovascular disease - A systematic review of economic evaluations. Cost Effectiveness and Resource Allocation 2007;5. https://doi.org/10.1186/1478-7547-5-5.

[215] Miller S. Screenings and Early Intervention Can Reduce Medical Costs. Society for Human Resource Society 2012.

Vita

Haya Salah was born and raised in Tulkarm, Palestine. Before immigrating to the United States, she attended An-Najah National University, Nablus, and received a Bachelor of Science in Electrical Engineering in 2013. From 2015 to 2017, She attended Binghamton University, where she earned a Master of Science in Industrial Engineering and graduated with honors. While at the University of Missouri, Haya was nominated to be part of the Industrial Engineering Honor Society (Alpha Pi Mu) and offered graduate research assistant positions and teaching assistant for several courses, including Engineering and Predictive Modeling, Introduction to Industrial Engineering, and Engineering Quality Control. Currently, Haya is an Associate Data Scientist at Edward-Elmhurst Health in Naperville, Illinois.