QATAR UNIVERSITY

COLLEGE OF BUSINESS AND ECONOMICS

HARNESSING MACHINE LEARNING IN CLINICAL DECISION SUPPORT: THEORY AND PRACTICE

BY

AHMAD ABDEL LATIF ABUJABER

A Dissertation Submitted to

the College of Business and Economics

in Partial Fulfillment of the Requirements for the Degree of

Doctorate of Philosophy in Business Administration

June 2021

COMMITTEE PAGE

The members of the Committee approve the Dissertation of
Ahmad Abujaber defended on 12/04/2021.

Professor Adam M. Fadlalla
Dissertation Supervisor

Professor Belaid Aouni
Committee Member

Dr. Manar Alhassan
Committee Member

Professor Guodong (Gordon) Gao
Committee Member

Approved:

Dr. Adam Mohamed Fadlalla, Dean, College of Business and Economics

ABSTRACT

ABUJABER, AHMAD, A.M., Doctorate : June : [2021:],

Doctorate of Philosophy in Business Administration

Title: Harnessing Machine Learning in Clinical Decision Support: Theory and Practice

Supervisor of Dissertation: Adam M. Fadlalla.

Decision making is a central activity in all clinical professions. Clinical decisions bear wellbeing and economic risks and consequences for patients, families, employers, and national economies. Thus, clinicians should employ sound scientific knowledge to promote optimum decision outcomes. Evidence-based medicine organizes clinical decision-making activities with philosophical, ethical, and methodological foundations to ensure accessibility to the best scientific knowledge to inform clinical decision-making. But high-quality evidence could be lacking or methodically, ethically, or economically unfeasible, compelling clinicians to make risk-bearing decisions without an ideal evidence-base. In uncertain clinical conditions, decisions` outcomes are probabilistic and the clinicians'` knowledge about outcomes are limited which make the clinical decision in significant need for aid.

The advent of electronic health records and the data warehouse technology boosted healthcare industry capacity to generate and store vast amounts of diverse data types. Advanced analytical and computational techniques in artificial intelligence and medical informatics provide unprecedented opportunities to galvanize knowledge deployment in clinical decision, bridging gaps in clinical knowledge. The ability of machine learning approaches to handle the large-scale and diverse data addresses some evidence-based medicine challenges, providing real-time, cost-effective evidence.

Nevertheless, despite our beliefs that artificial intelligence and data science may

have the potentials to transform the clinical practice, the utilization of artificial intelligence in healthcare is still poor, and the full benefits are not reaped. The adoption of machine learning in healthcare faces several epistemological, methodological, and ethical challenges that make the integration between the machine learning and the evidence-based medicine a hard mission.

This dissertation adopted a literature-based reconceptualization to help comprehensively understand the two paradigms, to guide the paradigms reconciliation agenda and to determine the reasons behind the slow adoption of machine learning in healthcare industry. Secondly, we followed an interpretive research design in order to propose a roadmap that aims to enhance the adoption of machine learning in clinical decision support. To validate the theoretical work, we conducted five empirical studies in collaboration with trauma surgery section at Hamad Medical Corporation. In these studies, we developed several ML predictive algorithms to address real-life clinical issues that are faced by the trauma surgery clinicians and predict the prognosis of patients who suffered from Traumatic Brain Injury which include mortality, prolonged mechanical ventilation, ventilator associated pneumonia and prolonged in-hospital length of stay.

Subsequently, this dissertation determines how machine learning and evidence-based medicine can be reconciled through proposing a novel pragmatic reconciliation framework that guides the clinicians and the scholars on how to benefit from the synergistic effect of both paradigms. In addition, this dissertation determines the factors that negatively affect the adoption of machine learning in clinical decision support and proposes an original theoretical framework that draws a strategic road map towards the effective adoption of machine learning in healthcare. Furthermore, the dissertation

sheds the light on the future research directives that may enhance the compatibility between the data science and the evidence-based medicine paradigms in order to augment the clinicians' capacity to make high-quality informed decisions.

DEDICATION

Dedicated to my father who taught me that pursuing education is never too late. To my wife and children who sacrificed during the past 4 years and to Professor Adam Fadlalla who taught me that a PhD is more about uniqueness rather than completeness.

ACKNOWLEDGMENTS

I would like to acknowledge the medical research center at Hamad Medical Corporation for supporting this project. I would also like to acknowledge the Trauma Surgery Section at Hamad General Hospital for their support and faith in my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION AND PROBLEM STATEMENT

Data Science in Healthcare

The interest in data science approaches in the healthcare industry is growing significantly, and with the growing capabilities of handling large data sets, there are unprecedented research opportunities that may change the face of care delivery (1). The growing use of electronic health records has contributed to the generation of massive volumes of diverse types of data, but utilizing such data effectively remains problematic. For instance, improving the quality of care using electronic health records is still an under-researched area (2). Fortunately, researchers are increasingly assisted by the availability of increasingly effective analytical tools that enable the analysis of large-scale data sets, such as machine learning (ML) approaches (3). These help practitioners improve the quality of their decisions and as a result contribute to improving the treatment outcomes and their organizations' economic performance (4). The advancement in modern computational capacity, specifically ML, encouraged industrial organizations to make considerable investments in data science analytics in order to improve decision-making capabilities, and reduce risks and costs (3, 5).

However, despite the massive quantities of data accumulating from patient records (6), and the rise of the health information technology and health informatics (7), the potential benefits of the modern analytics in the healthcare industry have not been fully reaped, which results in both economic and wellbeing opportunity costs for healthcare systems and service users (6, 8, 9). The utilization of ML analytics in other business sectors helped organizations improve efficiency and profitability (10). Therefore, there is an economic cost

associated with not utilizing ML in managing healthcare organizations. One of the key areas where ML can benefit the healthcare industry is in clinical decision support. Clinical decisions are the fundamental activity of all clinical professions, whose outcomes influence patient health status, the quality of life (QoL) of patients and their families, and the reputation of healthcare institutions, as well as the cost of healthcare for patients and their families, hospitals, and the entire economy. Therefore, the great capacity of ML approaches represents an unprecedented opportunity to uncover potentially useful knowledge in clinical data that can influence clinical decision making and treatment outcomes.

Research Problem

The opportunity cost of not utilizing data science in healthcare is significant, and in the era of electronic health records, investment in data science is arguably an obligation. However, developing huge data warehouses without building the capacity to leverage such data is meaningless. Accordingly, the critical question of concern to this research is why the potential benefits of data science and ML analytics have not been realized in supporting clinical decision making.

Research Question and Objectives

The key research question for this thesis is "***why are the benefits of ML in supporting clinical decision making not fully realized?***"

To answer this question, we break it down into the following sub-questions:

(a) How can ML be integrated with EBM to support clinical decision making?

(b) Why is ML adoption in healthcare relatively slower than in other sectors?

(c) How ML adoption in clinical decision support can be enhanced?

Answering this question mandates a meticulous analysis of literature in order to explore the current state of research in this area, to identify the gaps and to determine the future research directions that help both practitioners and academics benefit from utilizing ML in the healthcare sector. Therefore, this study aims to:

- Understand the nature, strengths and challenges of clinical decision making in the era of evidence-based medicine (EBM).
- Review the recent ML literature to understand the current status of ML in supporting the clinical decision making, identifying related opportunities and challenges.
- Determine the gap in knowledge in order to address the reasons for the poor utilization of ML in biomedical sciences.
- Propose a theoretical framework to guide future efforts to reconcile ML and EBM in a way that addresses philosophical, methodical, and ethical differences.
- Propose a conceptual framework that provides a roadmap for policy makers to address the problem of slow ML adoption in clinical decision making.

CHAPTER 2: CLINICAL DECISION MAKING – THEORY AND PRACTICE

Within medicine, there are more than ten major disciplines consisting of various subspecialties, dealing with more than 10,000 specific illnesses. Before starting any treatment for all these diseases, a diagnosis must be established first (11). The wellbeing and economic consequences that are associated with clinical decisions influence numerous stakeholders in addition to patients, which is part of the reason substantial amounts of state subsidies are devoted to supporting the healthcare sector. Thus, providing support to clinical decision making is of utmost importance due to the inherent importance of healthcare services for patients and society in general.

Djulbegovic et al. (12, 13) discussed that although substantial expenditure and subsidies are devoted to the healthcare sector, health outcomes remain poor. They attributed this to suboptimal decision making, which is regarded as a leading cause of death, and which is responsible for more than 80% of healthcare expenditure. In 2011, the USA spent around 18% of GDP ($3.2 trillion) on healthcare, of which around 30% was considered inappropriate and therefore wasteful (14). Qatar spent only 2-3% of GDP on the healthcare sector in the past decade (15); while no comparative data is available about the suitability of expenditure on national healthcare, it can be assumed that a great deal of wasteful expenditure exists, which could be mitigated by enhancing the quality of decision making.

The objectives of this section are (a) to explore the theoretical and the philosophical foundations of the clinical decision making and the EBM paradigm, (b) to understand how the EBM defines the guiding principles for the clinicians' clinical decisions and scientific activities, and (c) to identify the

challenges that face the EBM and impede its further progress in the era of the 4th industrial revolution.

What Do the Terms “Decision” and “Clinical Decision Making” Mean?

This section explores the theoretical and philosophical foundations of clinical decision making, and how the EBM paradigm guides the clinicians’ clinical decisions and scientific activities. According to the Oxford Dictionary, the word “decision” refers to “the choice or the judgement that you make after thinking and talking about what is the best thing to do”. Dowie (16) argued that a decision is the “cognitive preliminary to the action” (p. 8), referring to the choice of what to do now or the choice to wait and see. A decision cannot be articulated without subjecting multiple options to judgement, which refers to the assessment of alternatives. In other words, the concept of a decision refers to assessing the several alternatives and then choosing to operate or not to operate on the basis of one of them. Regarding to particular concept of medical decision making, Karni (17) defined it as the “choice of an action following a diagnosis of a patient’s condition, the medical treatment itself; the facility in which it is to be administered; and, if perceived relevant, the individuals who administer it”. More broadly, Gladstone (18) stated that clinical decision making is used interchangeably with clinical judgement, referring to “a cognitive process concerned with problem recognition through the identification of cues and clinical features, data gathering, integration, analysis, evaluation and choice to produce an informed decision” (p. 66).

It is very important to understand that any clinical decision is associated with an element of assessment of the future (19). In the book *Analysing How We Reach Clinical Decisions*, Dowie (16) discussed that the sensibility of a

decision is determined, to a great extent, by the amount of knowledge the clinician has about how the future may look as a result of making the decision. Without this future element, the decision will be made without thought.

The future is an integral component in any clinical decision. The future that clinicians aim to influence by their decisions is related to patient outcomes and healthcare system interests. Since clinicians cannot know the future with certainty, nor whether their diagnosis is absolutely accurate, they cannot know whether their prescribed interventions and treatments will cure the patients (20). Furthermore, the probabilistic nature of the decision outcomes adds to the challenge that the clinicians face (21). The clinician is expected to make decisions that guarantee the best possible outcomes for patients, their families, and healthcare organizations. However, the inherent lack of knowledge of the future scenario resulting from the clinical decision places a great responsibility on the clinician, who must evaluate all conceivable risks. Clinical decisions may negatively affect the patient's health or QoL and may unnecessarily increase the healthcare bill and burden of families and the entire economy. Furthermore, the timeframe within which the clinical decisions are required are of significant importance (19, 20). Subsequently, clinicians are under pressure to make decisions that maximize gains for patients and healthcare institutions, while minimizing risks, all in the context of situations with a great degree of inherent uncertainty, represented in the limited amount of relevant knowledge about the future that may help evaluate various decision alternatives.

Several scholars asserted that the clinical decision making is a rational process that consists of observations, critical thinking, and clinical judgement and information processing, relying greatly on hypothetico-deductive reasoning

(22-24). When applied in clinical decision making under uncertainty, hypothetico-deductive reasoning is considered a method for choosing the best alternative (among available options) based on the rationality criterion (13, 24).

Rationality usually refers to analytical decision making to achieve desired goals. Rational decisions are evaluated based on their ability to optimally maximize the expected utility from available choices (13, 25). Nonetheless, rationality does not guarantee that a decision is error-free, rather it accounts for false negative or false positive types of errors (13). In other words, if rational reasoning is followed, a conclusion will be sought irrespective of the accuracy of the premises. If a premise is false, the conclusion, although the process is rationally valid, will be false.

Theories of rationality in clinical decision making can be broadly classified into three categories; descriptive, normative, and prescriptive theories (12, 13, 19, 26-28). Descriptive theories are concerned with ***how*** and ***why*** decisions are made (13, 19). Originating from psychology and behavioral sciences (29), descriptive theories that have direct relevance to medical decision making include (but are not limited to) bounded rationality, argumentative theory of reasoning, adaptive rationality, and pragmatic/ substantive rationality (12). Descriptive theories put great emphasis on the context, ecology, and environment where decisions are made, with no restrictions on whether people are rational or irrational (19). On the other hand, normative theories depict how ***rational*** people ***should*** make decisions in an optimal world (19). They stem from the disciplines of mathematics, statistics and economics (13, 27, 29). Normative theories that influence the clinical decision making include epistemic theories that direct the practice of EBM and

expected utility theory, which provides the basis for widely used clinical decision analyses (12, 13). The prescriptive theories basically deal with how people ***can*** think and decide (28, 30). They stem from the fields of operations research and management science (29). Prescriptive models are usually variants from both descriptive and normative theories. The prescriptive theories are set out to aid the decision making process in the form of policies, protocols and clinical guidelines (19). Table 1 summarizes some of the common theories of rationality that are related to the medical decision making.

Table 1. Major Rationality Theories Relevant to Clinical Decision Making

Descriptive theories	
Bounded rationality	Rational behavior relies on the satisficing process (finding a good enough solution) (12).
Argumentative theory of reasoning	People make decisions because they can find reasons to support them. These decisions are not necessarily the best or the most satisfying to rationality criteria; rather they favor socially acceptable attitudes, being easy to justify and less likely to be criticized (31).
Adaptive/ecological rationality	A variation from bounded rationality in which human decision making relies on the context and environmental cues. Rational decision making requires adaptation to environmental/ patient circumstances (20, 32).
pragmatic/substantive rationality	The decision depends on the content of the decision, not only on the process, and is assessed in light of the short- and long-term purpose (33).
Normative theories	
Epistemic theories	Rationality is based on acquisition of true/ fit-for-purpose knowledge. EBM is an example of a normative approach to medical decisions (12, 13, 34).

Normative Theories	
Expected utility theory (EUT)	Based on Bayesian probability calculus, this is the most dominant approach to rational decisions in medicine (13). The rational decision is associated with selection of the alternative with higher expected utility (expected utility is the average of all possible results weighted by their corresponding probabilities) (12, 13).
Prescriptive theories	
Bounded rationality	Variation from the descriptive approach providing mechanisms for improving human rationality given the human cognitive architecture (12, 13).
Dual processing theories of rational thought/ Meliorism	Human cognition architecture can be viewed as a function of two types of processes: (a) old mind – intuitive and emotional based; and (b) new mind – analytical, related to the future. Rationality should take into account both the intuitive and analytical processes. Prescriptive models are variations of descriptive and normative approaches, proposing that individuals can be taught to be rational.

Furthermore, we identified five principles that have been agreed upon across all the theories of rationality and decision making that apply to the clinical decision making.

- **Benefit-harm/ gain-loss interaction:** all the major rationality and decision-making theories aim to maximize gain and minimize loss. This applies to the basic principles of clinical decision making, where clinician's decisions aim to achieve the best outcomes for the patients and their families while being cost effective.
- **Uncertainty:** decisions occur under uncertain conditions and the future/decision outcome is probabilistic. Therefore, rational decisions require reliable evidence to face the uncertainties.
- **Cognitive architecture:** there is consensus among the three theoretical approaches that rational decisions have to be informed by the human

cognitive architecture, which is composed of two types of processes; (a) old mind, fast, intuitive, and emotional process; and (b) new mind, analytical, and effortful process.

- **Context:** rational decisions are contextual, depending on the context with respect to epistemological, environmental, and computational constraints of the human brain (13).
- **Ethics and morality:** rational decisions are expected to serve the best interests of the individual (as per the duty-bound obligations of clinical professionals), while putting high emphasis on the greatest benefit for the majority (utilitarian ethics) (35).

Consequently, clinical decision making can be defined as a cognitive and rational reasoning process that takes place under uncertain conditions, which relies greatly on the hypothetico-deductive reasoning, that defines the way a healthcare professional processes the patients information (36). This forward-reasoning approach is considered beneficial not only because it guides clinicians' analytical process and diagnoses, but also helps prescribe the decision (18, 37). This mode of reasoning and deciding follows four essential steps:

- **Acquisition of cues:** Primary data about the patient and other sensory stimuli which direct clinicians' thought processes and help generate hypotheses.
- **Hypotheses generation:** Based on the collected cues and patients' baseline data, clinicians set differential diagnoses.
- **Interpretation of cues:** Clinicians collect further data to re-explore and re-interpret cues and use them in evaluating hypotheses. This could

include diagnostics such as laboratory and imaging studies and/or consulting specific specializations.

- **Hypothesis evaluation:** The cues are then evaluated and applied to an overall hypothesis that directs the decision and the subsequent interventions.

In conclusion, clinical decision making follows a strict rational reasoning process that depends on true, reliable, and justifiable knowledge that results from thorough hypothetico-deductive methodologies, which gives confidence to clinicians to rely on when making decisions in uncertain conditions, when knowledge about the future is limited. These principles set the direction, values, and standards for the EBM paradigm.

Evidence-Based Medicine (EBM)

Conceptualization and Role

EBM is the fundamental basis of clinical practice in modern healthcare (38). As described by Djulbegovic et al. (12, 13), EBM, as a rational decision making paradigm, embraces the normative approach of rationality through determining how rational clinicians should make the decisions based on what they believe to be true (34, 38). The belief of truth is the function of the trustworthiness of evidence and the degree to which we believe that the process that led to the evidence is credible (12, 38). Emphasis on the reliability and the credibility of the process of reaching the evidence can be seen from the perspective of proximity to the truth. In other words, the reliable process takes us closer to the truth and therefore is capable of generating trustworthy evidence. Thus, the decision that is based on that evidence will better help us calibrate the gains and losses/ benefits and harms (13, 38). To enhance

clinicians' capacity to guide their decisions by high-quality evidence, EBM is developed to include the prescriptive aspects of rationality. One of the most important principles of the prescriptive rationality is that rationality can be taught or dictated (12). Therefore, it is very common nowadays to see the evidence-based clinical pathways and guidelines presented in decision tree algorithms (DTAs) that guide the clinician throughout the clinical reasoning and the decision process. Also, the modern healthcare system puts great emphasis on the continuing professional education that focuses on the process of generating, reviewing, evaluating and employing best available evidence in clinical practice. Healthcare organizations in today's modern healthcare system encourage activities such as journal clubs, multidisciplinary case reviews, and morbidity and mortality conferences, in order to review the degree to which the performance of the system and the individuals can be improved through evaluating the degree of conformance with EBM standards.

EBM is defined as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (39). The development of the evidence-based paradigm has been enabled by evidence from high quality randomized controlled trials (RCTs) and observational studies, communicated via further analysis in systematic reviews and meta analyses, providing clinical practitioners with distillations of extensive clinical research which they can interpret with their clinical expertise to tailor care to individual patient needs; this has revolutionized modern healthcare from the traditional paradigm of treatment traditions based on anecdotal evidence and theoretical reasoning (40).

The main concept behind the evidence-based practice is that the

activities of practitioners should be guided by the best empirical findings in their fields (41). According to Gambrill (42), evidence-based practice integrates the practice and research. This explains why healthcare providers are under pressure to employ research-based interventions in their practice (43). From another perspective, clinical decisions entail critical safety considerations and risks (44). This explains why clinical research puts significant emphasis on controlled experiments and the elimination of bias, with great reliance on randomized controlled clinical trials (1, 45). Indeed, the EBM paradigm is fundamentally concerned with criteria to evaluate the quality of evidence, which is organized in a hierarchical based on the research methods used in generating data (41, 46). The logic behind this pyramidal evidence model is that certain methods are more rigorous, and thus provide more confidence in the clinical decision making (47). Accordingly, methods that generate knowledge from data (e.g. ML) are in a critical position with regard to influencing clinical decisions. Figure 1 and Table 2 explain the hierarchy of evidence and grading of recommendations.

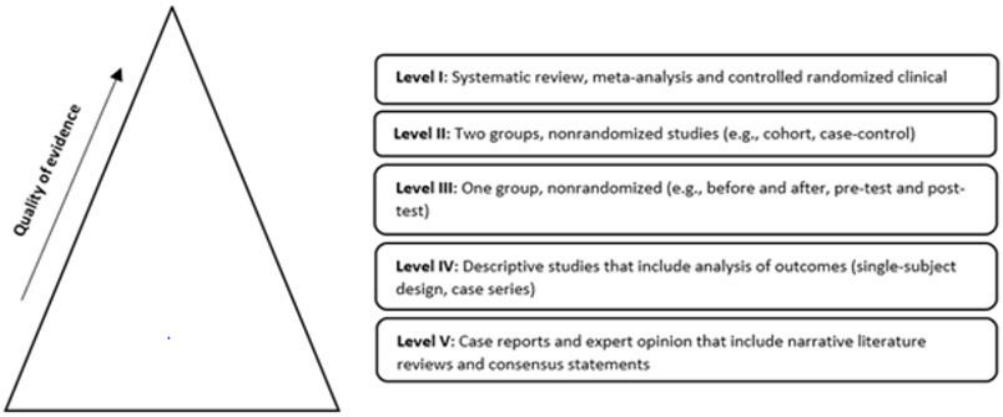

Figure 1. Traditional single-hierarchy evidence model; Tomlin and Borgetto (46).

Table 2. Levels of Evidence and Grade of Recommendations

Level of evidence	Grade of recommendation
Level I	Grade A
Level I	Results of single RCT with lower limit of confidence interval for the treatment effect exceeding the minimal clinically important benefit.
Level I+	Results come from a meta-analysis of RCTs in which the treatment effects from individual studies are consistent, and the lower limit of the confidence interval for the treatment effects exceeds the minimal clinically important benefit
Level I-	Results come from a meta-analysis for RCTs in which the treatment effects from individual studies are widely dispersed, but the lower limit of the confidence interval for the treatment effect still exceeds the minimal clinically important benefit.
Level II	Grade B
Level I	Results come from a single RCT in which the confidence interval for the treatment effect overlaps the minimal clinically important benefit.
Level I+	Results come from a meta-analysis of RCTs in which the treatment effects from individual studies are consistent and the confidence interval for the treatment effect overlaps the minimal clinically important benefit.
Level I-	Results come from a meta-analysis of RCTs in which the treatment effects from individual studies are widely dispersed, and the confidence interval for the treatment effect overlaps the minimal clinically important benefit.
Level III	Grade C
	Results come from non-randomized concurrent cohort studies.
Level IV	Grade C
	Results come from non-randomized historic cohort studies.
Level V	Grade C
	Results come from case series.

Adopted from Cook et al. (46).

Although EBM changed the face of medicine and contributed to

significant medical advancement, it faces significant challenges. First, the very nature of the clinical decision outcomes is probabilistic, due to the uncertainty and the bounded rationality of clinicians and clinical scenarios; put simply, there is always risk associated with clinical decisions (13, 44). This is a strong reason why clinical decision making is in a great need for aid by advanced decision support tools. From another angle, to achieve the goals of the EBM, practitioners need to build skills to critically evaluate the literature to find the proven effective interventions (best evidence) that are generated from meta-analysis of the RCTs (42, 43). This, according to Greenhalgh et al. (40), is one of the biggest challenges to the EBM. Clinicians' basic training does not concentrate on building academic skills to critically review and judge the quality of research or to abstract evidence. Therefore, proponents of EBM advise the institutionalization of the production of the clinical evidence and clinical guidelines rather than leaving them to the variable individual skills and abilities of practitioners (43).

Furthermore, to translate evidence into practice, the evidence is usually transformed into algorithmic top-down clinical guidelines that serve as a reference for practitioners and help standardize practice, to obtain optimal treatment outcomes. However, in many instances normative clinical evidence may be unavailable, impossible, unethical, impractical, or too expensive (48). The lack of high-quality evidence represents a very significant challenge to the EBM. In such scenarios, practitioners find themselves compelled to make risk-bearing clinical decisions in the absence of strong and reliable background knowledge. Therefore, they rely on lower-level evidences (i.e. level V - expert opinions) or on their experience, which may not necessarily lead to the best

possible outcomes. This risky trade-off led some scholars to take the initiative to propose solutions to help clinicians benefit from the scientific advancement of other disciplines, such as data science, management information systems, and health informatics.

Further to the aforementioned challenges, the process of evidence generation is very slow (1), as well as the adoption of evidence in clinical and managerial practice (49, 50). It has been estimated that the time lag between the generation of evidence until it is effectively deployed in clinical settings is approximately 17 years (51, 52). Westfall et al. (52) argued that only 14% of new discoveries will get access to day-to-day clinical activities. It has been reported that USA citizens receive 50% of the recommended preventive, acute, and long-term healthcare solutions theoretically available under EBM (53). Ebell et al. (54) aimed to determine the degree to which the primary care practices are informed by high quality evidence, and found that only 18% of primary care recommendations are based on high-quality and patient-oriented evidence, while around 50% are based on expert opinions (which in turn are based on the usual way of treating medical conditions). Bradly et al. (55) argued that although the evidence of the beneficial use of β-blockers in myocardial infarction has been in place for more than 20 years, this medication is still underused in clinical practice. Westfall et al. (52) ascribed this to several reasons: (a) the vast majority of patients are treated in healthcare facilities, while high quality research is conducted in the academic institutions; (b) the tight inclusion and exclusion criteria that are practiced in clinical research create an artificial sample of patients that do not represent the vast majority of patients; and (c) the majority of clinical guidelines are based on evidence that is based

on a very small number of patients who basically present to tertiary healthcare facilities. Consequently, the applicability of evidence in other settings such as primary healthcare facilities is questionable.

The above challenges mean, to a great extent, that in some scenarios the clinicians will find no high-quality evidence to support their clinical decisions, thereby posing significant risks to patients, families, and healthcare institutions, in addition to clinicians themselves. On the other hand, the healthcare industry is experiencing the revolutionary development of high-throughput data-generating technologies. This significant development in data collection and storage capacity requires building up capacity to enable the development of sophisticated algorithms that outperform the traditional computation and inferential techniques, enabling practitioners to obtain and analyze evidence much more quickly and efficiently (56).

Moreover, patients themselves pose challenges for EBM. The spread of social media and the use of smart devices have changed patients' access toward health information and options, and their attitudes towards managing their own health (57-59). Beckmann and Lew (60) maintained that we are living in the most profound periods of advancement in biology and medicine, leading to a medical revolution that will transform healthcare and contribute to precision medicine. Perhaps this is the reason that motivated several scholars to argue that the application of the big data analytics in the healthcare is inevitable and to suggest that the reconciliation between the EBM and the data science could be the remedy for the EBM challenges (1, 43, 61, 62). However, due to the lack of literature that tackles this subject, it is very difficult to abstract a serious literature-based framework that can guide this research agenda. In addition,

there is a serious lack of literature that informs the reconciliation between EBM and data science paradigms.

Challenges to EBM Implementation

In summary, despite the remarkable contributions of the EBM to the advancement of healthcare and enhancing clinical decision quality, through emphasizing on the importance of basing clinical decisions on reliable knowledge and high-quality evidence, there are several challenges to the achievement of the main purposes of EBM. These include:

1. Uncertainty, the bounded rationality of clinicians, and the probabilistic nature of the clinical decision outcomes.

These increase the margin of risk embedded in clinical decisions. Therefore, there is an increasing need for clinical decision support tools. Therefore, we argue that the modern analytics (i.e. ML) that can handle large volumes of data collected in electronic health records can provide a great support to the clinical decision making.

2. The availability of high-quality evidence is not guaranteed.

In several situations the normally expected evidence for EBM may be unavailable. Furthermore, the generation of evidence entails huge costs and ethical implications, requiring significant stretches of time. This is another reason why data science and artificial intelligence (AI) approaches that utilize the readily available data can be viewed as a remedy.

3. The generalizability of evidence is based on a relatively small number of patients.

Precision medicine relies to a great extent on large data volumes and ML approaches, which can provide an avenue to overcome the “one-size fits

all" challenge. It is worth noting that we do not propose data science or ML as a panacea for all EBM challenges. Rather we present the ML as a solution to address *some* of the challenges that EBM suffers from, knowing that ML and other AI modalities suffer their own challenges that can be partially addressed by EBM standards.

Data Science Applications in EBM

The potential applications of data science in the era of EBM can be categorized under four rationales.

First, clinical decisions must always be supported by strong arguments and thoroughly evaluated models. However, data science approaches are usually free from hypotheses or theory, which questions the positivist, empirical value of data-driven knowledge in EBM. Second, clinical experiments are costly and add significant discomfort for the participants, and tend to involve diseased people, therefore, data science approaches provide a relatively cheap, innovative, and a powerful alternative to the RCTs in specific circumstances and under specific conditions. Third, the data sets at hand may be irreproducible in controlled conditions. Therefore, controlled trials may not always be the appropriate method to generate evidence in certain conditions. Fourth, the process of evidence generation and adoption in EBM is very slow. Therefore, ML approaches may present an opportunity to enhance the timeliness of evidence availability and interestingly generate new hypotheses for future RCTs and experiments. This argument forms the base of the proposed pragmatic framework that is presented at the end of the next chapter, which discusses the applications of ML in supporting clinical decision making.

CHAPTER 3: UTILIZATION OF MACHINE LEARNING IN CLINICAL DECISION SUPPORT IN THE ERA OF EBM – A CRITICAL EVALUATION

Background

Motivated by facilitating fast and cost-effective decision processes, several business sectors have outsourced the process of decision making to AI and ML algorithms. It is evident in the literature that the utilization of ML in supporting decision making has resulted in significant benefits in several industrial and academic fields. Therefore, we believe that there is significant potential to enhance the quality of clinical decision making through the adoption of the ML in supporting clinical decisions and contributing to the process of generating reliable clinical evidence. The value of supporting the clinical decision making reflects on several aspects, including patient treatment outcomes, improved patient QoL, patient and family satisfaction, improved healthcare organization performance, and reduced healthcare costs for patients/families, healthcare systems, and national economies. Nevertheless, it has been reported in several academic and business reports that the benefits of the modern analytics in the healthcare industry are not fully reaped, and that the adoption of the AI and ML approaches in healthcare are particularly slow compared to other industrial sectors.

Therefore, it is crucial to determine the factors that impede realizing the maximum benefits of ML and which inhibit the speed of ML adoption in supporting clinical decisions. Hence, we decided to conduct a scoping review of literature that tackles the use of ML in clinical decision support in the era of EBM. The purpose of this review is to determine the current status of the literature to understand in depth the challenges that face ML in the healthcare

industry in the era of EBM, and which contribute to the ineffective adoption and prevent the realization of the full potential of ML in supporting clinical decisions. Furthermore, we aim to understand the avenues wherein the ML can address some of the key challenges that face the EBM. Thus, we can propose a literature-based pragmatic reconciliation framework that achieves the required synergy to address some of the challenges in both paradigms. Furthermore, we aim to provide some directives for future research to help the professionals and the scholars optimally utilize ML to enhance clinical decision making.

Methodology

Literature Search Strategy

PubMed database was searched to identify review articles that were published in the period from January 2016 to September 2020 discussing ML applications in various clinical fields in the era of EBM. The literature search was limited to the reviews that were published in the past five years because we believe that this field of research is rapidly evolving and changing. Therefore, older reviews may reflect obsolete issues that are addressed by newer research. Accordingly, the reviews that are published in the past five years provide the most recent challenges that face the ML in the clinical research and would help us identify the current position and the progress of the research agenda. We applied the following key terms to search in the titles and the abstracts (machine learning OR artificial intelligence OR big data) AND (evidence-based medicine) and the selection was limited to review articles. This approach was followed in order to: (a) optimize the literature search and selection procedures through limiting the key search terms to the main key words that relate directly to the core of the review; and (b) understand the

current position of this field of research by using the most recent reviews, enabling this study to address literature gaps and explore emerging research directions. This process resulted in 189 articles.

Article Selection

The initial step in articles selection was through scanning the abstracts to retain only studies discussing ML, AI or big data as key subjects, which resulted in 69 articles that were carried through to the next phase.

The articles' inclusion criteria were formulated to ensure that the selected articles were relevant to the study's objective. Articles that thoroughly discussed one or more of the following subjects were selected; (i) common ML methods in clinical research; (ii) ML potentials in supporting clinical decision making; (iii) challenges that face ML in clinical fields; (iv) epistemological aspects of ML in clinical research; and/or (v) recommendations to benefit from ML in supporting EBM. Articles that referred to ML techniques as a future opportunity for addressing some computational issues in clinical research without providing a thorough discussion were not selected. This step resulted in 52 articles.

Outlet Quality Measures

All the 52 articles that met the study inclusion criteria were again evaluated based on the citation metrics of the publishing journals. Impact factor and Source-Normalized Impact per Paper (SNIP) were considered for selecting high quality journals. The selected journals had to have an impact factor of 1 and SNIP of 1 at least to be considered. Despite the debate and the limitations, impact factor is still a reasonable measure for the medical journal's quality (63). Further, SNIP metric that is included in Elsevier's SCOPUS database was used

as it corrects for the differences in the impact factor across the scientific disciplines (64). This step resulted in 15 papers.

Results

The database search returned 189 unique review articles. After conducting the initial screening to the abstracts, 69 papers were selected for the next phase. The inclusion criteria were applied on the selected articles which resulted in 52 papers. of them, 37 papers were disqualified. Accordingly, 15 papers were ultimately included in the review. The 15 papers cover several medical disciplines as shown in Table 3.

Table 3. Clinical Disciplines Covered by Selected Reviews

Discipline(s)	Review article(s)
Cardiology	Shameer et al. (65)
Oncology, radiotherapy and cancer genomics	Marka et al. (66); Resteghini et al. (67); Vogelius et al. (68); Xu et al. (69)
Critical care medicine	Sanchez-Pinto et al. (70)
Immunology	Mersha et al. (71); Saglani and Custovic (72)
Neuroscience	Kim and Na (73)
Occupational health	Six Dijkstra (74)
Orthopedics	Cabitza et al. (75); Helm et al. (76)
Mental health	Torous et al. (77)
Pulmonology	Pepin et al. (78)
Rheumatology	Gossec et al. (79)

Discussion

The selected papers cover variety of subjects concerning ML in various medical disciplines. All the papers discuss the great computational capacity of ML approaches in supporting clinical decision making, and their potential in

boosting efforts towards precision or personalized medicine, and the future revolutionary impacts of ML that will change the way we practice medicine. Nevertheless, there is consensus amongst the scholars that there are serious fundamental limitations in AI in general and in the ML in particular that need to be addressed. These limitations make medical practitioners and scholars skeptical about the validity of the ML output and its ability to play a real role in shaping or influencing clinical decisions. This may explain why the EBM paradigm, in spite of its well documented challenges, will continue to have the upper hand in formulating the treatment strategies and influencing the clinical decision making, as opposed to purely data-driven approaches.

This review is structured as follows. The first of the following sections covers the historical background and a glossary of terms that help the reader who does not have a background about the ML obtain an overview of the subject. The second section discusses the ML conventional tasks and techniques predominantly used in the clinical research. The third section provides a summary of the main subjects that are covered by the reviewed articles which include the avenues of ML in medicine; the need for ML in supporting clinical decisions; and the philosophical, methodological, and the ethical challenges that face ML in clinical sciences. The fourth section discusses the proposed pragmatic reconciliatory framework that seeks to activate ML to exercise a true influence on clinical decision making and contributing to EBM. The fifth section presents a list of recommendations to address the ML limitations and challenges.

Historical Background and Glossary of Terms and Definitions

ML is an umbrella term for a set of algorithms that enable computers to

uncover patterns and make decisions from data (65). Viewed with a larger scope, ML is a subset of AI (76) that is a subset of the data science (80). ML term was first used in scientific publications in 1959 in an article published in the *IBM Journal of Research and Development*, which defined ML as the "field that gives the computer the ability to learn without being explicitly programmed" (81). However, years before that, Alan Turing designed the "Turing test" to answer the question "Can machine think?", which is considered to be the beginning of the AI (82). In 1957, Frank Rosenblatt was inspired by the human thought process to propose the first neural network for computers (83). The progress in the ML continued until 1990 when the work shifted from learning *from knowledge* to learning *from data* (in large volumes), which started to produce robust outcomes in various fields, including medicine (84). To better understand ML, it is crucial to define a set of the key terms that will be used throughout this review. Sanchez-Pinto et al. (70) provided a comprehensive glossary of definitions for the key terms used in data science fields (Table 4).

Table 4: Glossary of terms

Term	Definitions
Big data	"Digital data that are generated in high volume and high variety and that accumulate at high velocity, resulting in datasets too large for traditional data-processing systems" (70)
supervised learning	"Algorithms that are used to uncover the relationship between a set of features and one or more known outcomes" (70)
Unsupervised learning	"Algorithms that are used to uncover naturally occurring patterns or groupings in the data, without targeting a specific outcome" (70)
Semi-supervised learning	"Machine learning from input data, where only a subset of input data is paired with output data, that is, an approach that mixes supervised and unsupervised learning" (68)

Term	Definitions
Ensemble learning	“Is an umbrella term for methods that combine multiple inducers or base learner algorithms to make a decision. The objective is to minimize the error of a single algorithm. So, the overall prediction performance can be improved” (85)
High dimensionality dataset	“A general term used to describe datasets that contain large numbers of features per patient, including genomic data and image features” (68)
Feature	“A variable used in a machine learning algorithm or an aspect of a dataset that is of some relevance” (86)
Model training	“The process through which machine learning algorithms develop a model of the data by learning the relationships between features and, in supervised learning, between features and outcomes. This is also referred to as model derivation or data fitting” (70)
Model validation	“The process of measuring how well a model fits new, independent data. For example, evaluating the performance of a supervised model at predicting an outcome in new data. This approach is also referred to as model testing” (70)
Predictive model	“A model generally trained to predict the likelihood of a condition, event, or response” (70)
Prognostic model	“A model specifically trained to predict the likelihood of a condition-related endpoint or outcome such as mortality. In general, the goal is to estimate a prognosis given a set of baseline features, regardless of what ultimately leads to the outcome” (70)
Overfitting	“The phenomenon that occurs when an algorithm learns from idiosyncrasies in the training data, usually referred to as noise which leads to poor performance of the model” (70)
Structured data	“Data that are easy to search, summarize, sort, and quantify” (70)
Unstructured data	“Data that do not conform to a prespecified structure, such as a written narrative, images, video, or audio. Unstructured data are generally harder to search, sort, and quantify” (70)

Machine Learning (ML) Algorithms

The included reviews covered several ML algorithms that are commonly used in the clinical literature, categorized into three categories based on the intended goal or function and the characteristics of the outcomes: supervised learning, unsupervised learning, and reinforcement learning (65, 70, 73, 75).

Semi-supervised algorithms, a hybrid of supervised and unsupervised algorithms, are used less frequently in healthcare (70). Cabitza et al. (73) used the terms "*conventional ML*" to differentiate between supervised and unsupervised learning on one hand, and "deep learning" on the other (75). Similar to Shameer et al. (65), in this study, we categorize the ML algorithms into four classes: supervised, unsupervised, deep, and reinforcement learning.

Supervised ML Algorithms

As described by Deo (87), supervised learning starts with the goal of predicting known output or target. For example, if researchers need to know if a set of features can predict certain outcomes (e.g. mortality), they may apply a supervised ML algorithm where there is an outcome variable (e.g. survived or deceased). The most common supervised algorithms that are highlighted in the included articles are described below.

Regression-Based Algorithms

This class includes both the classic regression (i.e. linear and logistic regression models) and the regularized regression algorithms (i.e. Least Absolute Shrinkage and Selection Operator (Lasso), elastic net and ridge regression), which perform both the feature selection and regularization to enhance the accuracy of the prediction. This is achieved through imposing penalties on the fitted model to reduce its complexity and risk of overfitting (70).

Tree-Based Algorithms

This class includes variety of algorithms that serve classification prediction functions such as C.5 decision tree (C.5DT) and classification and regression algorithm (C&RT). Generally, DTs are sequence of "if-then-else" splits derived by iteratively separating the data into groups based on the

relationship of the features with the outcome. A DT is a “classification algorithm in which each non-leaf node indicates a test on an attribute of the input cases; each branch corresponds to an outcome of the test; and each leaf node indicates a class prediction” (88). Generally, they are powerful, logical and easy to interpret and to understand classification algorithms (89). This class also includes random forest (RF), an example of ensemble tree models, which combine the output of many trained models to estimate an outcome. For example, RF uses bootstrapping to grow a forest of uncorrelated trees with a high degree of randomness in feature selection, which contributes to significantly reducing errors (90).

Support Vector Machine (SVM)

SVM is a powerful classification algorithm that can be used for linear and non-linear data sets. SVM represents the data in a multidimensional space and then fits a hyperplane that optimally separates the data based on the intended outcome. To achieve this, SVM uses an optimal kernel function (e.g. linear, polynomial, or radial basis function) to map the input data into a higher dimensional feature space (70, 91, 92).

K-Nearest Neighbor (KNN)

KNN is a supervised learning algorithm that represents the data in a multidimensional feature space and uses the Euclidean distance to predict the class of the unknown example based on its closeness from training examples (70, 93).

Bayesian-Based Algorithms

A class of algorithms that uses Bayes theorem of conditional probability assumes the prediction of the occurrence of something given that something

else has already happen (posterior probability). This class includes naïve Bayes (NB) and Bayesian network.

Neural Networks

Neural networks are class of nonlinear algorithms which can be viewed as a set of connected input/ output units in which each connection has an associated weight. During the learning phase, the network learns by adjusting the weights to be able to predict the correct class label of the input items. The networks can be shallow, with two layers, or deep, with multiple layers, as in the field of deep learning (70, 93). This class of algorithms include the artificial neural network (ANN).

Unsupervised ML Algorithms

Unsupervised learning is a synonym for clustering (93). It is basically used to uncover naturally occurring patterns or groupings in the input data without pre-existing labels or particular outcome (68, 87). The most common examples of the unsupervised learning algorithms are the clustering algorithms (hierarchical clustering and k-means clustering) and principal component analysis (PCA) (65, 70).

Clustering Algorithms

Hierarchical clustering. This is a clustering algorithm that applies an iterative process of grouping similar observations in clusters based on similarity or the chosen distance function. Hierarchical clustering is classified into one of two types: (a) agglomerative approach (bottom-up), where the algorithm treats every observation as a cluster, and it keeps merging similar observations in an iterative process until only one cluster remains (80); and (b) divisive approach

(top-down), which starts with all objects as one cluster, which it then iteratively splits into smaller clusters, until eventually each object is in one cluster (93).

K-means clustering. Different from hierarchical clustering, which looks into similarities between instances and links them together, K-means algorithm focuses on the clusters themselves, through representing them by their centers or centroids (arithmetic mean) (80). The algorithm starts with the K initial cluster centers, formed based on the closeness of each point to the centroid. Next, the centroid of every cluster is recalculated by finding the new centroid after the changed cluster membership. The K-means procedure keeps iterating until there is no change in the clusters' membership (80, 93).

Dimension Reduction

Dimension reduction algorithms work on transforming a large data set into a smaller set that retains the important and pertinent information of the antecedent larger set (80). There are several dimension reduction methods, such as PCA, Wavelet transforms, and Linear Discriminant Analysis (70, 93). PCA is the most commonly used dimension reduction algorithm (73).

Principal Component Analysis (PCA). PCA's main function is to reduce a large set of dimensions into a smaller set of artificial dimensions, usually referred to as principal components (73). PCA combines dimensions that have relations which were not previously revealed resulting in dimensionality reduction (93). The principal components can then be used as input in further analysis, such as regression.

Deep Learning

Some scholars consider deep learning to be a subset of unsupervised ML (65), while others considers it to be a separate class that belongs to

unconventional ML algorithms (75). Deep learning is an emerging subset of ML that utilizes the multi-hidden layers neural networks to facilitate fast learning across a large number of samples (65, 94). Every layer in the neural network is used to encode input data into a something salient about the features that are contained within the data (86). Deep learning is basically suited for computer analysis, as it usually uses convolutional neural network-based representation to carryout image interpretation (95). This could be the reason why deep learning is popular in the fields of orthopedics and neuro-imaging (73, 75, 86).

Reinforcement Learning

Reinforcement learning is emerging subset of ML that is based on behavioral psychology (65). As described by Kim and Na (73), the key concept of the reinforcement learning is the interaction with the environment. Unlike conventional ML algorithms and deep learning, reinforcement learning is a dynamic process where the software agent aims to maximize the reward while interacting within a pre-specified environment. The agent learns the appropriate behavior by utilizing reward maximization criteria to handle the decision-making function (65). Reinforcement learning is used in several avenues such as image analytics, disease screening, and personalized prescription selection. Nevertheless, there are no promising outcomes yet for the application of the reinforcement learning, particularly in neuroimaging (65, 73).

ML Avenues and Opportunities Supporting Clinical Decisions

The main difference between the ML and the conventional statistical methods is that the latter helps us understand a relationship between small number of variables. However, the ML algorithms go way beyond that and help us handle large number of variables, handle diverse data forms, and engineer

features from data and perform predictions. With the advancement in data warehousing technologies, the advent of electronic health records, and the growing use of sensors and wearables in healthcare, the speed of data generation and the complexity of data have grown exponentially. Subsequently, the challenge has shifted from collecting data to obtaining potentially useful insight from the data (69). The big size and the complex diversity of the collected data (e.g. videos, waves, images, narratives, etc.) mandate the adoption of modern analytical methods that have the capacity to unveil the potentially useful information latent within the vast volume of available data. Therefore, the potentials of ML approaches in the medical field are undeniable. This review identified several avenues where ML algorithms prove useful in enhancing the clinical decision making.

Precision Medicine

Precision or personalized medicine is a field where the majority of scholars foresee that data science and AI can realize their maximum potential in healthcare (65, 69, 71, 78). AI is regarded as the main driver of the transformation of healthcare towards precision medicine due to its ability handle vast amounts of complex and diverse data, leveraging pattern recognition to enhance timelines and accuracy in clinical decision making (69). Precision, personalized, or individualized medicine are used interchangeably to describe "the treatments targeted to the needs of individual patients on the basis of genetic, biomarker, phenotypic, or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations" (96) (p. 2229). Precision medicine aims to tackle diseases through designing targeted treatments based on genomic, lifestyle, and environmental characteristics of

each patient.

Xu et al. (69) discussed that the advent of the AI and ML methods revolutionized oncology. The advancement in ML technology, especially dimension reduction algorithms, helped scientists analyze vast amounts of data with next-generation sequencing (NGS). Before NGS, Sanger sequencing technology was used to decipher the human genome, but required over a decade to deliver the final draft (97).

Shameer et al. (65) argued that the field of cardiovascular medicine produces huge amounts of data that are stored in different forms of repositories that are un-utilizable for research purposes. They argued that precision medicine has significant potential in the cardiovascular medicine. Precision medicine helps the physicians produce personalized guidelines that fit particular patients, thereby enabling individual-based decision making as part of tailored, patient-centered care. They suggested that there is a great opportunity for ML algorithms to be embedded in electronic health records to perform timely precision decisions and predictions.

Pepin et al. (78) studied the potentials of ML and big data analytics in diagnosing and designing personalized treatments for obstructive sleep apnea and other sleep disorders. They discussed that there are several avenues where ML can support precision medicine. Big data generated by telemonitoring help clinicians detect particular patients' phenotypes and understand the patient ecosystem, which supports personalized medicine efforts. Furthermore, the powerful clustering capacity of ML helps scholars determine homogenous patient groups, and generate hypotheses that can be utilized in targeted clinical trials to investigate responses to targeted therapies

directed to clusters that share similar molecular properties and pathobiological characteristics (70, 72, 78).

Torous et al. (77) discussed the potentials of ML methods in capturing and analyzing the huge amounts of data that are generated by social media in order to develop individualized suicidal risk profiles. They argued that ML clustering methods can use the social media data to cluster individuals into novel patient subgroups that may provide more accurate suicidal risk assessment at the individual patient level. For example, instead of using a few variables as currently utilized in clinical risk assessments (e.g. access to weapons and history of prior suicide attempts for suicide risk), researchers can now look at hundreds of variables to generate a more personalized risk profile.

Classification, Prediction and Prognostic Modeling

The growth in the use of the electronic health records, biomedical data bases, and data warehouse technologies and the advancement in the modern analytical approaches such as ML made disease or complication detection much easier than before (98). Moreover, the advent of the precision medicine that utilizes the AI techniques in genetics and the human genome made it possible to execute personalized predictions at the level of individual patients. One of the most important avenues where ML technologies may enhance clinical decision making is predictive and prognostic modeling (70).

Sanchez-Pinto et al. (70) stated that the purpose of predictive models is to identify patients with specific conditions or those who are likely to respond to specific treatments. They emphasized the supremacy of ML-based predictive power over conventional logistic regression-based predictive models. They reported that in several instances, ML-based models outperformed the mortality

prediction of one of the most common predictive tools in critical care settings, the Acute Physiology and Chronic Health Evaluation (APACHE) score, which developed in the 1980s.

In a similar vein, Shameer et al. (65) reported that ML can solve complex classification problems in cardiovascular medicine, especially in phenotypically difficult patients, such as differentiating between athletes' hearts and hypertrophic cardiomyopathy, or between constrictive pericarditis and restrictive cardiomyopathy. They argued that the use of ML in cardiovascular medicine enables risk stratification, which has significant implications for the quality of healthcare delivery and patient outcomes.

Torous et al. (77) highlighted on the great potential of ML for psychiatry. They discussed the application of ML and smartphone sensing technologies in the early detection and prevention of suicide. The authors reported that the majority of previous literature in this field focused on long-term suicidal risk due to the lack of real-time data and the computational capacity to effectively handle real-time data (where it exists). Therefore, they foresee that the utilization of ML in analyzing real-time data generated by smartphone and social media usage will enhance short-term suicidal risk detection (e.g. within 30 days).

In the field of orthopedic medicine, scholars discussed several avenues where ML predictive analytics can enhance the clinical decision making. ML has great potential to support orthopedic medicine through the field of image-based pathology detection and prediction (76). Helm et al. and Cabitza et al. (75, 76) reviewed several papers that tackle the applications of ML in several clinical and operational aspects of the orthopedic field, including cost estimation and reduction, length of stay, disease prognosis, and other outcomes.

Marka et al. (66) discussed the applications of ML classification potentials in the dermatology medicine. In their review, they compared between the ML and conventional analytics in handling the data complexity and dynamicity. They reported that the utilization of ML classification algorithms in digital images to classify non-melanoma skin cancer achieved results with higher accuracy (72%-100%) and discrimination power (81%-100%) that outperformed certified dermatologists.

Clustering, Phenotyping and Endotyping

Most well-known guidelines are implicitly based on the average patient, with limited tailoring toward individual patient needs (65). Thus Mersha et al. (71) and Saglani and Custovic (72) discussed the potentials of ML in supporting the field of immunology particularly through clustering and endotypes detection. Endotypes are defined as the pathophysiological mechanisms at the cellular and molecular levels that differentiate the response of individuals to certain disease process (72) . Similarly, Mersha et al. argued that it is not necessarily the case that individuals with similar clinical diagnoses will have similar disease etiologies, natural histories, or responses to treatment. Therefore, they argued that the advent of the high throughput molecular omics, immunophenotyping, and bioinformatics methods, including ML algorithms, will enhance the development of endotype-based diagnosis to help clinicians devise personalized treatment plans (71). Saglani and Custovic argued that big data assets provide the foundation for personalized asthma treatment. To achieve that goal, they emphasized the importance of multidisciplinary efforts whereby clinicians, epidemiologists, bio-informatics experts, and data scientists work together to obtain actionable insights from the data and reshape the clinical

evidence and make it more personalized to patients' specific characteristics (72). In the same field, Saglani and Custovic (72) discussed that ML has great potential to advance the endotype discovery through powerful clustering mechanisms that facilitate endotype-based diagnosis, enabling clinicians to devise treatment plans that take individualized characteristics and pathophysiological responses into account.

In the field of pulmonology and sleep disorders, Pepin et al. (78) discussed that the modern big data analytics, particularly clustering algorithms, would advance the field through phenotype identification. A phenotype is defined as "observable characteristics that result from a combination of hereditary and environmental influences" (99). It is noticeable that ML clustering is associated with different forms of advancements in medicine, such as next generation sequencing, phenotype and endotype discovery, and overall precision medicine, which transforms medicine towards devising personal care plans that are based on individuals' intrinsic and extrinsic factors to determine patients' risk of and response to certain diseases.

Real-World Evidence

EBM esteems RCTs and meta analyses because they satisfy the clinicians' quest to eliminate bias from their clinical decisions. Randomization and control are considered the golden standards of empirical research, to ensure that the observed variations between the study groups are caused by the intervention (100, 101). RCTs are "quantitative, comparative, controlled experiments in which a group of investigators study two or more interventions by administering them to groups of individuals who have been randomly assigned to receive each intervention" (102). Nevertheless, RCTs suffer from

several challenges and limitations, such as their significant cost, slowness, involvement of diseased individuals, and limited external validity (103, 104). These limitations have serious influences on the availability of evidence and negatively affect clinicians' ability to safely make necessary treatment decisions based on reliable evidence (48).

Due to the long time needed for RCTs to ultimately translate into EBM, many analysts have argued that modern ML approaches can be used in "real-world" applications to fast-track improvements in EBM. In an era where data and evidence about the real world are available in registries, databases, and longitudinal electronic health records, such scholars argue that studies which utilize ML and large volumes of real-world data can comprise "real-world studies" that circumvent the slow and tedious traditional processes of RCTs and EBM to offer immediate insights and guidance for clinical practice.

There are two perspectives about the relationship between the real-world studies and RCTs, the extremist and conservative views. The former view considers the real-world studies a substitute for RCTs, arguing that modern analytical techniques have the capability to achieve control and to analyze prospective data to mimic the RCT methodology, rendering evidence more quickly for use by clinicians at the point of care delivery (78, 104-106). However, this view is objected by several scholars who see that modern data science analytics are still inferior to RCTs because of their inability to control for confounders and to reveal causality. They argue that the majority of relationships revealed by data science methods are basically reflections of association rather than causation (68, 78).

Therefore, scholars such as Pepin et al. and Vogelius (68, 78) discussed

that the real-world studies cannot replace RCTs, and that data science outcomes should always be investigated by RCTs to ensure benefits. Nonetheless, such authors called for a complementary view that enables practitioners to yield the benefits of both approaches. The advancement in and the powerful computational capacity of ML methodologies can enable real-world studies to address the main two limitations in the mainstream RCTs: (a) the inability to gather a large sample in a reasonable timeframe, and (b) the inability to follow-up all the randomly assigned participants.

This coincides with our intuition that data science methods, through their pattern recognition ability, can provide hypotheses to be examined by RCTs. From another perspective, the great power of ML clustering represents a great potential to increase the success of RCTs. ML-based clustering capabilities reduce the heterogeneity of populations, enabling clinical researchers to select patients who are more suited to respond to the study's intervention.

Challenges to ML in the Era of EBM

It is undeniable that the modern computational methods will have great and perhaps revolutionary impacts on the ways medicine will be practiced in the coming years. Nevertheless, to reap the full potential of such technologies, we need to take a serious look at the challenges that face ML in the clinical fields, where clinical decision making remains dominated by the EBM paradigm for various legal, ethical, professional, and methodological reasons. Shedding light on these challenges is a crucial step to address them and to allow the data science paradigm to have a true influence on clinical decision making, to realize potential benefits for patient treatment outcomes and healthcare organizational performance in general.

The reviewed articles shed light on several challenges that can be classified as methodological and ethico-legal. However, there is a strong reason to believe that the origin of all the challenges is epistemological in nature. There are fundamental differences between the EBM and data science paradigms in the way in which they practice science and produce knowledge. These differences are reflected in the methodologies of the two paradigms and their ethical frameworks. Since EBM dominates the scene in clinical sciences, it poses numerous challenges to data science adoption.

Methodological Challenges

Control Over Confounders

The golden standard in EBM is control over confounders in order to eliminate bias and infer causal relationships. Several scholars argued that ML cannot control the confounding effect of other covariates, whereby it is difficult to infer a relevant causal relationship where bias is eliminated or controlled (65, 68). This is why, in the EBM paradigm, the most robust method is the RCT, while in ML randomization is not easily feasible (68). This is a serious limitation in ML; if this issue was addressed properly, ML could revolutionize clinical sciences to a greater extent.

Limited Generalizability

The ML algorithms train on a portion of a data set and then produce future predictions for the remaining portion of the dataset. This makes the model subject to learning from idiosyncrasies and noise in the data set that could lead to overfitting (73). Accordingly, there is no *a priori* reason to think that predictions made on the basis of the specific context wherein the model learned are applicable to other contexts (70). Model overfitting and lack of

control affect the generalizability of results and may lead to inappropriate conclusions (79).

Lack of Reproducibility

The combination between huge data sets and the complex computational and analysis processes (i.e. data cleaning, preparation, modelling, and validation) makes full reproducibility hard to achieve (107). Interestingly, in a paper published in *Nature*, a group of scholars and statisticians concluded that approximately 50% of published papers in the field of microarray gene expression that greatly utilized data mining are irreproducible or un-repeatable (108). Thus, several scholars advocate for the open availability of datasets, algorithms, study procedures, and even codes, to promote reproducibility (69, 70).

Complexity and Lack of Transparency

ML approaches are considered black-box analytical techniques for clinicians (69). This complicates interpretability and makes it difficult for clinicians to base their clinical decisions on ML based evidence (68, 78). Part of this problem is related to the awareness among the clinicians about the modern analytical methods (78). The other part is related to the very nature of ML projects, which lack transparency about the study procedure, making it difficult to compare the performance of the ML algorithm with the performance of human experts in order to demonstrate the value of the ML to the clinicians (69).

Lack of Standardization

Shameer et al. (65) noted that evidence in mainstream medicine is synthesized through summarizing multiple studies in standard procedures,

such as meta-analysis or systematic reviews. This process is essential in EBM to develop clinical guidelines that inform the clinical decisions. It is difficult to do the same in the ML field, because there is no policy to regulate the dissemination and the reporting of ML models in medicine. In many instances, researchers do not report data elements and model parameters, or declare the transformations that are carried out on the data. All these challenges complicate the synthesis of ML-based evidence and negatively affect the reliability of outcomes. This may partially explain why relatively few ML projects have succeeded (70).

Ethical and Legal Challenges

Six Dijkstra et al. (74) explored the ethical considerations and potential consequences of using ML-based decision support tools in the context of occupational health. They argued that the use of ML in occupational health will change the nature of interactions between clients and care providers, with unknown consequences. They discussed the ethical considerations in light of the four predominant ethical principles of medicine:

- ***Autonomy,*** which refers to respecting the individual's right of choice;
- ***Beneficence,*** which entails that the clinicians should act in the best interests of their patients;
- ***Non-maleficence,*** which means "do no harm" as per the Hippocratic Oath; and
- ***Justice,*** which refers to fairness and equality among patients.

In the context of clinical sciences, ML faces challenges in terms of respecting the autonomy principle. Several scholars claim that open consent does not specifically explain how data will be used. In addition, the lack of

transparency of predictive analytical techniques makes prediction unverifiable. Therefore, they concluded that patients feel pushed to choosing certain treatment options, which contradicts the principle of autonomy. Similarly, Torous et al. (77) studied the potentials of using the social media data to predict suicidal events and discussed the challenge of consent to use data generated by smart phones and circulated in social media. Also, Gossec et al. (79) raised the issue of patients' role in big data. They argued that active patient participation in data collection, interpretation, and decisions related to big data analytics is an ethical obligation. However, consent to use, analyze, and interpret the data is still challenging in the field of big data analytics.

In terms of beneficence, Six Dijkstra et al. (74) opined that the validity of ML algorithms' output is uncertain. This, to a great extent, is related to the lack of control and to the inability of ML to produce causal relationships. Hence, there is significant uncertainty around the consequences of basing clinical decisions on ML-based evidence, which raises issues concerning the principle of beneficence.

The issues related to privacy, confidentiality, and data ownership are tackled in the majority of data science literature. Six Dijkstra et al. (74) shed light on the ambiguity of the privacy and data ownership rights that are at risk as a result of the technological advancement. They mentioned the employees fear the powerful predictive functions of the ML that may reveal information to their employers about future conditions, and which may affect their employment status (74). In a similar vein, Torous et al. (77) argued that in the world of open data, ML algorithms may predict risks that may not exist, or carry out false negative predictions which may cause serious harm to individuals. These

issues pose serious challenges to the principle of non-maleficence.

With regard to the fourth principle (justice), Six Dijkstra et al. (74) discussed that patients' anonymity remains at risk even with data de-identification. They argued that patients are profiled according to similarities with others to know things they do not like others to know about. The risk increases when the data comes from open sources such as social media, where information about job, age, workplace, and years of experience could reveal individuals' identity. Besides the harm that may result from unveiling the identity, this challenge risks the principle of justice, as it may not control conflicts of interest (e.g. employers seeking information about employees that the latter do not desire their employers to know about).

Risking the very ethical principles is thought to negatively influence patients` and the clinicians` trust in ML. The power of ML to unveil the patients` identity, predict "possibly unwanted" future conditions, and the potential misuse of patients` sensitive information may lead to patients` mistrusting the ML. We believe the patients` mistrust in ML affects the clinicians` trust in the ML. In addition, the clinicians` concerns that artificial intelligence would replace the clinicians increases the mistrust feelings among the clinicians. We suggest that lack of trust in ML that results from the methodical and ethical challenges contributes to the slow adoption of ML in healthcare that hinders the healthcare industry's capacity to reap the full benefits of ML in supporting clinical decision making. Therefore, to address that, policy makers may need to address the methodical and ethical challenges that negatively influence the trust in ML in order to facilitate the adoption of ML in clinical decision support.

Epistemological Challenges

We believe that all the challenges that face the data science in the field of clinical sciences are essentially epistemological, as the epistemology of a certain paradigm defines all aspects related to knowledge production, thereby determining acceptable methods of data collection, processing, analysis, and interpretation, based on ontological and axiological assumptions. Nonetheless, despite their fundamental underlying importance, the epistemological challenges that face data-driven science in clinical sciences are not given adequate attention in related literature. The reviewed studies identified two key epistemological challenges that face data science and its subsets in clinical sciences dominated by the EBM paradigm: the challenge of inductive reasoning, and the challenge of passive observation.

Originating from the deductive post-positivism school of thought, medical knowledge is created through theory refutations and falsificationism (45). This means that the medical scholars have to guide their scientific inquiry by a hypothesis that is subject to refutation (109). The theoretical framework defines the process of data collection, processing, analysis, and interpretation (110). Accordingly, the quality of the evidence in clinical sciences is, to a great extent, determined by the soundness and the thoroughness of the research methodology. This explains why the EBM paradigm adopts the pyramidal representation of evidence quality, which explicitly means that certain methods are more rigorous than others, and therefore capable of providing more confidence in the clinical decision making (47, 111) (Figure 1).

On the other hand, the great reliance of data science paradigm on secondary data that are collected basically for reasons other than research not

only poses ethical and methodical problems, but also reflects a serious philosophical concern (43, 61). In the field of data science and its subsets, data collection is not always driven by a specific hypothesis to refute or a research question to answer (3, 44, 110, 112). ML algorithms learn from and identify patterns in past data that are collected without a guiding refutable conjecture, with an implicit assumption that future events can be anticipated by observing what happened in the past. This is a thorny epistemic assumption, especially when the related decisions are risk-bearing and regulated by strict philosophical foundations that form the EBM paradigm. There is a widespread belief that inductivism is a mistaken philosophy of science (110). Consequentially, data science is criticized for being non-scientific and not subject to refutation. Scholars argue that a theory or hypothesis is always required to guide scientific enquiry in EBM, while data-driven science leads to apophenia, due to assuming relations between unrelated things. Therefore, Fricke (110) concluded that fields that produce knowledge from data are chimerical. In the fields of clinical sciences, where decisions are critical in nature and are associated with high risks, an interpretable model grounded in explainable theories would stand a better chance to be accepted by clinical scientists and professionals than models that learn from spurious data patterns but which cannot lead to generalizable performance guidelines (113).

The second argument is the passive observation versus the active experimentation. Passive observations can be misleading and cannot affirm causal relationships, largely due to confounding effects. Fricke (110) defined confounders as “other conditions or variables which correlate either with the causes or with the effects to mask what is really happening at a causal level”.

Therefore, control in the EBM paradigm is the golden standard. The rigor of any research methodology in EBM is determined by its ability to control confounders, and therefore its ability to produce reliable knowledge based on which evidence can be abstracted. According to Gerstein et al. (105), RCT is the preferred methodology in clinical sciences because it achieves the three key criteria that ensure that the abstracted relationships are causal and generalizable in nature: (a) large sample size, to ensure that meaningful baseline imbalance does not occur by coincidence; (b) randomization, to ensure that the effect of the intended intervention on every participant is analyzed regardless of adherence to the intervention or treatment with additional therapies; and (c) near complete follow-up of all the participants until the end of the study.

Mainstream scholars' basic criticism of ML is based on their belief that the lack of theoretical-based protocolization in designing ML algorithms risks the loss of control over confounders (68, 70, 78), and that the data that are collected for ML algorithms are subject to incompleteness, errors, and systematic bias, which may thus lead to erroneous predictions (114). The methodological, ethical, and epistemological differences between the two paradigms lead to a phenomenon we name "*Normative non-Congruence*", which gives the perception to clinicians and researchers that the two paradigms are incompatible. Therefore, in order to facilitate the utilization of ML in clinical decision support we have to (a) propose recommendations that help address the methodical and ethical challenges, and (b) propose a pragmatic reconciliatory framework that guides clinicians and scholars on how to build on the synergy that results from capitalizing on the strengths of the two paradigms.

The next section presents our recommendations to help address the identified methodical and ethical challenges and presents our view on how to achieve the reconciliation between the two paradigms despite their fundamental epistemological differences. We adopted a ***pragmatic*** stance that aims to transcend the epistemological challenges and the normative incompatibilities between the two paradigms in order to enable the maximum realization of utility of each paradigm, and to achieve what we call "***perceived normative congruence***". This view provides a complementary prospect in which we believe that both paradigms have strengths and suffer weaknesses, and our goal is to capitalize on the strengths and mitigate the weaknesses where possible. The framework provides practical guidance to clinical practitioners and researchers on how and when to balance the emphasis on theory-driven or data-driven knowledge in order to obtain the ***best possible/ feasible evidence***.

We believe that although EBM has revolutionized the clinical sciences in the last three decades, it inherited the challenges from which the hypothetico-deductive paradigm suffers. Scholars ascertain that deductive conjectures lack the ability to predict the future with precision. Hence, inductive reasoning is needed to create causal inferences about the observed phenomena (115). ML approaches can address some of the key EBM challenges. Accordingly, if we adopt a synergistic stance, each paradigm can address some of the other paradigms' challenges to pragmatically provide robust evidence in a time-efficient way to support clinical decision making. This argument is the cornerstone of the proposed reconciliation agenda.

Data and Theory: A Pragmatic Paradigm Reconciliation Framework

It is obvious that both EBM and data science paradigms represent two extremes of knowledge discovery based on their sources of scientific information: theory-based knowledge vs. data-based knowledge. The EBM puts great emphasis on scientific knowledge that results from the accepted methodologies in the paradigm. Therefore, the clinical decision maker has to support clinical decisions with the ***best available evidence***. On the other hand, data science depends greatly on the knowledge contained in data that results from pattern recognition and other predictive or descriptive algorithms. Figure 2 illustrates the different emphasis on data vs. theory between the EBM and data driven science.

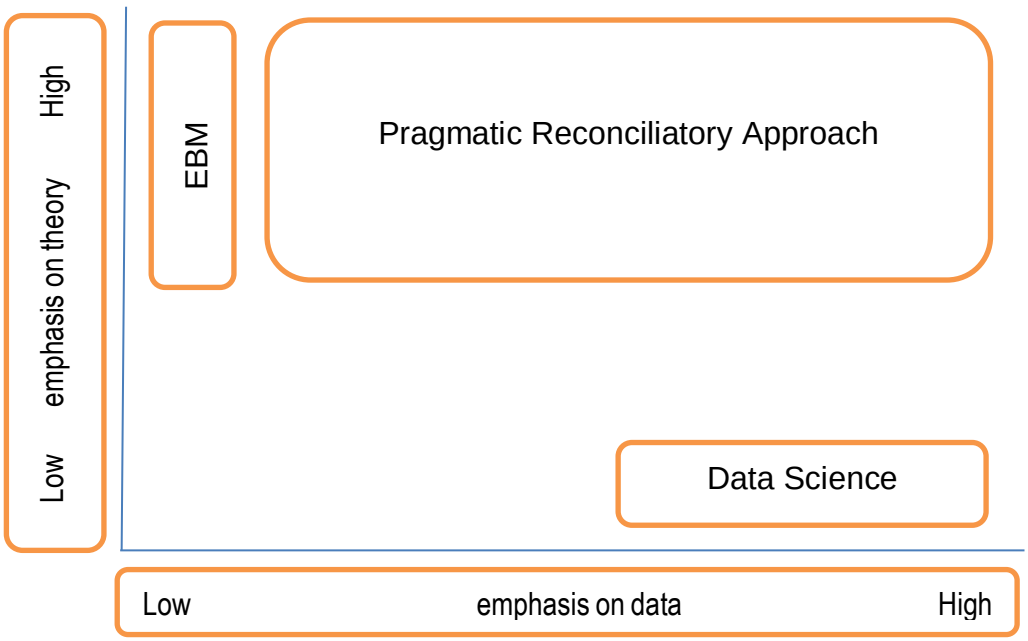

Figure 2. Emphasis on theory vs. data.

There is no doubt that the two paradigms' ultimate goal is to provide high-quality evidence to inform clinical decisions and ensure the best possible treatment outcomes. Nevertheless, the word "evidence" holds different meanings in every paradigm. Evidence in EBM refers to the reliable scientific knowledge that results from specific robust methods that adopt the hypothetico-deductive reasoning approach (116). In data science, evidence comes from knowledge hidden in data through the utilization of sophisticated algorithms which have the potential to detect patterns of associations and correlations; in this paradigm it is unnecessary that the scientific inquiry is guided by a background theory.

To better understand this argument, it is instructive to refer to Kuhn's theory of knowledge, which entails that the practices of scientists are strictly guided by core theoretical frameworks, guiding principles and techniques that structure a paradigm. In *The Structure of Scientific Revolutions*, Kuhn discussed the term "normal science" to describe scientists' actions as a process of solving a puzzle, aiming in the first place to prove and support the key theoretical framework of their scientific paradigm (117). The key practices and guiding principles in the EBM and the data science paradigms represent two extremes in terms of the emphasis on the theory-based knowledge vs. the data-based knowledge. Thus, the term "evidence" definitely means different things in each paradigm, and it is epistemologically impossible to achieve synthesis between paradigms, because one cannot operate on more than one paradigm simultaneously (118).

Therefore, our goal is not to synthesize a third paradigm as a result of merging two paradigms. Instead, the key principle of our view is that each

paradigm has strengths and challenges. Therefore, (i) neither paradigm is considered sufficient for knowledge discovery in complex scientific applications, and (ii) it is a mistaken assumption that data science will address all the EBM challenges, and vice versa. Consequently, a pragmatic, balanced view is crucial. The main goals of this pragmatic approach are (a) to transform the emphasis from the "***best available evidence***" to the "***best possible/feasible evidence***", and (b) to enhance the clinicians' "***perceived normative congruence***" that may boost willingness to accept the ML-based knowledge as a source of evidence to enhance the quality of clinical decisions.

Perceived normative congruence refers to the clinicians' perception that there is an acceptable degree of compatibility between the data-based and theory-based paradigm. We stress on the normative component because EBM sets strict rules on ***how*** clinical practitioners and researchers should think and decide. Therefore, it is crucial to approximate the way data-based approaches work in EBM. We strongly believe that this will contribute to boost clinicians' willingness to accept the data-based paradigm in their decision making process. We suggest that the congruence can be achieved through providing a framework that legitimizes the use of ML in scientific clinical knowledge production through addressing some of the key challenges from which ML approaches suffer. Therefore, our proposed framework and recommendations play an integral role in securing a foothold for ML approaches within the clinical sciences, which will contribute to enhancing congruence with the EBM paradigm.

On the other hand, the focus on the "***best possible/feasible evidence***" stems from the fact that there are several scenarios where the evidence does

not exist, or is impossible, unethical, time-consuming, or extremely costly (48). In such cases there is a need to support clinical decisions without compromising the quality of evidence. Thus we propose that when “high-quality evidence” is not possible through mainstream methodologies, or it requires long time to be formulated, the knowledge generated by applying ML approaches on a large amount of patient data will be utilized as an evidence base until higher quality evidence comes into existence. This is crucial to bridge the gap between clinical knowledge and clinical decisions in the EBM context. This also keeps a sense of tentativeness of evidence, which encourages clinicians and scholars to keep searching for better evidences to support clinicians make higher quality evidence-based decisions.

The pragmatic stance is summarized by viewing the two paradigms as a continuum, as shown in Figure 3, where clinicians and scholars can move right or left according to feasibility, time, and cost constraints. This does not in any way mean ignoring epistemological constraints. Knowledge produced in either direction must always be refereed by experts, considering the known body of knowledge, to ensure the reliability of knowledge and safety in supporting clinical decisions.

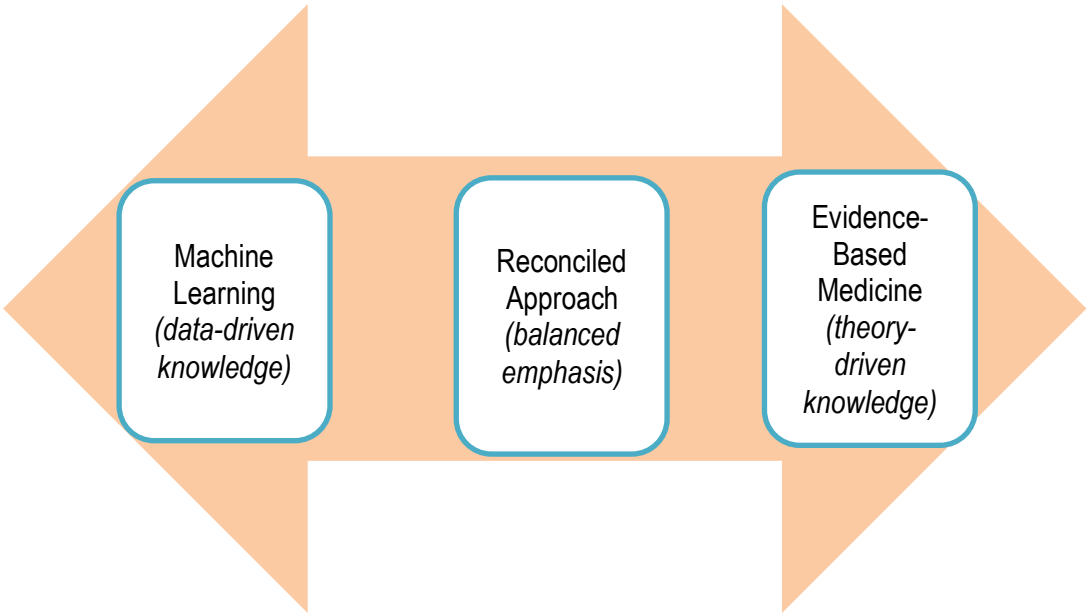

Figure 3. Data-driven, theory-driven knowledge continuum.

We believe that this framework provides practical guidance to the clinicians and scholars on how to balance between data- and theory-driven knowledge. We propose three conditions that may guide when and how clinicians and scholars can move along the continuum to yield the best possible/feasible evidence:

1. When evidence does not exist or is difficult to obtain due to ethical, legal, or cost constraints.

Utilization of ML algorithms to uncover potentially useful patterns in large volumes of data (e.g. from electronic health records) may be of a great value to bridge the gap in knowledge and to help clinicians make safe and reliable informed clinical decisions. For example, Toussi et al. (48) developed a C5.0 decision-tree learning algorithm to fill the evidence gap in identifying the best combination of the oral anti diabetic drugs for patients with type 2 diabetes mellitus. Also, Abujaber et al. (119) applied SVM algorithm on a dataset of

patients with moderate to severe traumatic brain injury (TBI) who received mechanical ventilation, and provided a new perspective to redefine the prolonged mechanical ventilation (> 10 days) in order to help critical care physicians timely consider the tracheostomy. These findings coincide with a recent Cochrane review which found that mechanically ventilated patients will benefit significantly if the tracheostomy is considered early (no later than 10 days from the start of mechanical ventilation) (120).

2. When mainstream clinical research methodologies are feasible, scholars can consider the utilization of the ML approaches as alternative analytical techniques.

In this way, the EBM best research practices such as randomization, control, research ethics, and the presence of a guiding background theory will be maintained, while the analytical power of ML techniques is properly utilized.

3. Electronic health records generate vast volume of diverse data that is thought to hide interesting patterns that are potentially insightful.

Furthermore, the nature of the electronic health records provides an outstanding opportunity for retrospectives longitudinal observation when the prospective longitudinal studies are not possible. The significant capacity of ML approaches to uncover interesting patterns in data can be of great value. These patterns can trigger questions about previously unknown associations between variables. These interesting patterns can easily serve as hypotheses for future RCT whenever possible.

Figure 4 shows the proposed reconciliation framework that guides the scholars how to benefit from ML in EBM. Importantly, to achieve the reconciliation purposes, the known body of knowledge should be respected.

This can be governed by three conditions.

First, institutional review boards (IRBs) should consider the provision of access to data scientists or ML experts who can collaborate with epidemiologists, biostatisticians and clinical scholars to ensure clinical researches that contain ML element respect the theoretical and the ethical frameworks.

Second, the output of ML algorithms should be interpreted in accordance with the known body of knowledge and the existing theoretical framework, to produce generalizable insights that influence clinical decision making.

Third, proponents of data science have to understand that the key word in data science is the "science", not the data. As discussed by Vogelius (68), clinical utility is not measured by discovering another prognostic tool or merely an interesting pattern in data or sophisticated mathematical technique, it is measured by the ability to support the clinical decision making or by the novel clinical insight.

Recommendations

The pragmatic reconciliation approach paves the way towards enabling the utilization of ML in supporting clinical decision and to ensure the timely availability of evidence to inform clinical and scientific activities. Besides that, the following recommendations aim to facilitate the adoption of the proposed framework and the achievement of the perceived normative congruence through addressing some of the main identified methodical and ethical challenges.

Data Management, Sharing, and Enhanced Generalizability

Healthcare policy makers have to consider the need for policies that: (a)

regulate and encourage the optimal utilization of electronic health records and data warehouses; (b) encourage open access to data provided the privacy and confidentiality are maintained; (c) encourage data sharing at the national and international levels, to facilitate multicenter studies which enhance the generalizability of research outcomes; and (d) encourage the sharing of data sets that are used in research studies to facilitate governance.

Analysis Procedure and Enhanced Interpretability and Reproducibility

It is imperative for the procedural governance, transparency, interpretability, and reproducibility that researchers should explicitly report study procedures, codes, missing data imputation, and algorithm tuning, etc. (121). In another aspect, there is a need to encourage a multidisciplinary approach in analyzing data in ML projects. This may help scholars utilize best epidemiological practices (e.g. randomization and controlling confounding variables). Intuitively, the availability of large volume of data may make the randomization and control in ML achievable tasks, given that the ethical, privacy, and confidentiality requirements are respected. It was suggested by May (106) that scholars can achieve a virtual control in ML studies, but no explanation was provided on how to achieve that. We argue that scholars can always create a randomly selected control group of patients to provide comparative perspectives between study and the control groups. Furthermore, there is a serious need to develop standards that ensure reproducibility, and cross-validation procedures to control the over/underfitting problems.

Research Ethics

ML study designs should respect privacy, anonymity, and confidentiality in all project phases, including data collection, processing, storage, analysis,

and interpretation. In the context of clinical research, ML-based clinical research should comply with the ethical standards of IRBs and health research ethics committees. Scholars must prove that their study protocols provide all possible means to secure patient data and maintain privacy. One of our key recommendations is that an ML expert or a data scientist should be a member in the IRB whenever ML project is presented. This is essential to inhibit any opportunity for potential profiling, function creep, or discrimination. This is also important for the maintenance of privacy and ownership of data.

Furthermore, there is a need to reconsider the process by which we obtain informed consent, especially as data science analytics holds risks of data misuse. We recommend that the IRB sets rules that force the scholars to restrict the use of data only to the purposes that are listed in the consent. Thus, whenever a need for further analysis arises that may risk the principles of patients or privacy, permission by the IRB should be granted.

Provision of Academic and Professional Training

Policy makers in healthcare institutions have to identify the opportunity to train healthcare professionals on modern data analytical approaches. The same applies for academic institutions that may consider the need to include data analytics and AI approaches in the curriculum of clinical sciences. This contributes to raising awareness among clinical students, which contributes to building a strong foundation that considers the ML as a reliable alternative to produce scientific knowledge. Furthermore, the complexity of ML methods may serve as a limiting factor. Therefore, academic and professional training may help clinicians master the use of the ML and realize their value in supporting clinical decision quality.

The next section is an attempt to synthesize a literature-based conceptual framework in order to answer the question of how clinicians' adoption of ML in clinical decision making can be enhanced. Several challenges have been highlighted in the literature that negatively influence the willingness of clinicians and clinical researchers to utilize ML in their scientific activities. Nonetheless, none of the past literature provided a comprehensive framework that presents a roadmap to guide policy makers to set strategic directions to enhance the utilization of ML in supporting clinical decisions.

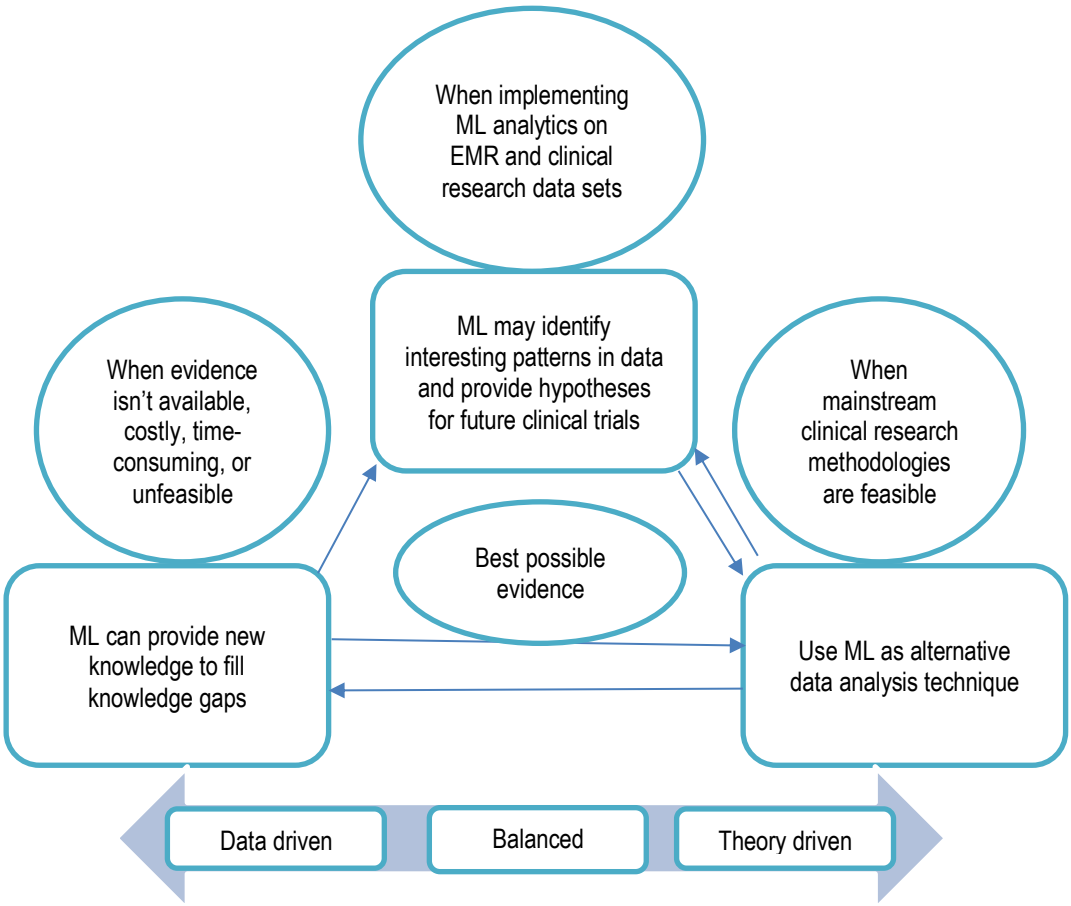

Figure 4. Pragmatic paradigm reconciliation framework

CHAPTER 4: USING TISM AND MICMAC ANALYSIS TO MODEL ENABLERS OF CLINICIANS INTENTIONS TO ADOPT ML IN CLINICAL DECISION MAKING – A CONCEPTUAL FRAMEWORK.

Background

It has been evidenced in the literature that the full potential benefits of modern analytics and data science approaches in healthcare are not fully realized (6, 8). We showed in the previous chapter that this is a multifactorial problem, highlighting the significant epistemological incommensurability that influences the normative congruence between the data science and the EBM paradigms. To address this gap, we proposed a pragmatic framework that provides practical guidance to clinical practitioners and researchers on how to balance their emphasis on data and on theory based on certain circumstances in order to provide the *best possible evidence*. In addition, we provided some recommendations that could address some of the main challenges faced by ML in EBM, which enhances the *perceived normative congruence* of the paradigms. However, one of the reasons that negatively affects the realization of the full benefits of ML in clinical decision support is the slow adoption of AI approaches, including ML, in the healthcare field.

In a market study conducted by the McKinsey Global Institute in 2019, the healthcare industry, relative to all other industries, was found to have the slowest rate of adoption of the five key AI technologies, including the ML (122). This poor adoption is a fundamental reason for not realizing the full benefits of ML in healthcare. The poor adoption of ML represents a lost opportunity to support clinical decision making, which is a key reason for poor healthcare outcomes and substantial wasteful expenditure in healthcare (13). Therefore, it

is crucial to determine and model the factors that lead to the slow adoption of ML technology in the healthcare context, particularly to support clinical decision making.

The purpose of this chapter is to propose a Total Interpretive Structural Modeling (TISM) conceptual framework that helps us: (a) model the factors that may enable the adoption of ML in the healthcare industry; (b) understand the nature of and the interrelationships between adoption enablers; and (c) develop a diagraph that orders the enablers in a hierarchical fashion, to present a roadmap that informs policy makers on how to enhance clinicians' intentions to adopt ML approaches in healthcare, particularly in supporting clinical decision making.

The scoping review identified several challenges that face ML in the field of clinical decision support. We believe that addressing these challenges will encourage clinical practitioners and researchers to adopt the ML in their clinical decisions and scientific activities. Furthermore, we consulted the technology acceptance model (TAM) to borrow the original constructs (perceived ease of use, perceived usefulness, and users' behavioral intentions to adopt). The reasons behind using TAM are that it addresses the potential users' intentions to use a technology, which we believe that it is a key for enhancing the adoption of the ML at the industry level; and the original constructs of TAM fit perfectly with this study context. The benefits and the complexity of ML were highlighted as challenges to adopting ML in several studies. Therefore, we consider that enhancing the perceived usefulness and the perceived ease of use will work as enablers to adopting ML in clinical decision support. On the other hand, although we do not treat the ML as a technology in the same way TAM treats

the technological innovations, we still see ML as an innovative technological approach to support clinical decision making, which offers great potential functionality in theory, but which stumbles at the intersection of actual effective use in clinical practice.

This work presents an original conceptual framework that provides strategic directives to the policy makers towards enhancing clinicians' intentions to adopt ML in their clinical decisions, thus reaping the maximum benefits through raising the quality of clinical decision making, while reducing healthcare expenditure. Although many identified variables were highlighted by previous literature, no previous studies presented a theoretical framework that explains the dynamicity amongst the enablers and the path towards achieving the ultimate outcome, which is enhancing the adoption of the ML in supporting the clinical decision making. Thus, we argue that this work is novel and adds value to the body of knowledge.

Why Total Interpretive Structural Modeling (TISM)?

According to Watson (123), structural models are necessary for dealing with the modeling process that helps scholars identify the elements of a model and analyze the interactions among the identified elements. Interpretive Structural Modeling (ISM) is defined as "a process that transforms unclear and poorly articulated mental models of systems into visible, well-defined models useful for many purposes" (124) (p. 87). The ISM helps develop a hierarchy of interacting variables to graphically represent transitive relationships between a system variables (125). Haleem et al. (126) argued that ISM is an interactive process that helps structure a set of interacting elements that are relevant to the problem at hand in a comprehensive systematic model. Also, ISM helps

scholars order and identify the direction of a complex system of relationships between elements of a system (126, 127). As described by Haleem et al. (126), ISM is interpretive in the sense that it helps scholars determine whether or not and how the variables are related. ISM is widely used in several industrial fields, including supply chain management (128, 129), airline performance (130), waste management (131), strategic decision making (132), and in the healthcare industry (133). Sushil (124) identified four key limitations of ISM:

- It can only be used by those who are trained to interpret the data.
- It relies greatly on computers, which makes it difficult if computers are not available.
- It provides partial interpretation of links between variables, which leads to multiple possible interpretations.
- It does not provide a causal explanation for the links between the variables, which limits theory-building potential.

Jena et al. (134) added a fifth limitation, which is that ISM does not consider the transitive linkage between the variables. Accordingly, Sushil presented TISM as an extended ISM approach to address the identified limitations of ISM (124, 135). According to Deshmukh and Mohan (135), the use of the TISM results in the a diagraph that represents the complex system among the variables of interest which was described as a novel qualitative multi-attribute decision modeling method (136). Jena et al. (134) summarized the salient characteristics of TISM as the following:

1. TISM is interpretive in nature, as it utilizes expert opinion to determine why linkages between connected variables exist in the way they do (i.e. inferring causality).

2. TISM is a modeling technique that depicts the contextual relationships, entire structure, and interpretation for direct and significant transitive links between variables by a diagraph.
3. TISM simplifies the portrayal of complicated systems.
4. TISM addresses ISM's poor interpretability of links through employing an interpretive matrix.
5. TISM facilitates theory-building through answering *what*, *why*, and *how* questions.
6. TISM interprets both the links and nodes in the structural model.

Methodology

The aim of this chapter is to develop a TISM-based hierarchical relationship among a set of literature-based identified enablers of ML adoption to support clinical decision making. TISM methodology is composed of nine steps, as described below. To achieve the objectives, a panel of 4 expert clinical scholars with experience and knowledge of the potential role of ML in supporting clinical decision making was formulated to conduct the TISM process. The panel consists of two experienced physicians, and two experienced nurses who are all well published researchers. It was agreed that the contextual pairwise relationships would be blinded and done by every panelist separately, then all pairwise relationships were discussed openly. The final decision about the pairwise relationships was reached by majority voting, if no consensus was achieved.

Step I: Identify and Define Enablers of ML Adoption in Clinical Decision Making in the EBM Era

This preliminary TISM step sources elements from previous literature

(129, 134), and brainstorming with a panel of experts (128). In this study, the adoption enablers were sought basically from the previous literature.

Perceived Normative Congruence

Based on the results of the scoping literature review, we identified several factors that pose challenges to the utilization of the ML in the clinical field. The epistemic incommensurability between the EBM paradigm, which emphasizes theory-based knowledge, and the data science paradigm, which puts great emphasis on data-driven knowledge is regarded a key challenge to the willingness of the clinicians to adopt ML. Within a scientific paradigm, scientists and practitioners are mandated to comply with the paradigm's norms. This understanding led us to propose the construct "*perceived normative congruence*" that reflects the pressure of complying to the paradigm's norms on the user's intentions to use an innovation. In this study's context, the perceived normative congruence refers to the degree to which the potential users (clinicians and clinical scholars) believe that the ML approaches are consistent with the values, standards, and scientific traditions of the dominant paradigm in the clinical sciences, EBM.

In other words, the perceived normative congruence reflects the degree to which clinicians believe that the ML approach satisfies the normative approach of decision making, which includes but is not limited to ensuring the best methodological standards, such as control and elimination of bias, respecting medical ethical principles, and respecting patient privacy and anonymity. We suggest that the perceived normative congruence contributes positively in enhancing the clinicians' intentions to adopt ML in their clinical decisions. This is consistent with the theory of planned behavior, which

indicates that a person's behavioral intention is determined by subjective norms, which refer to "the perceived social pressure to do or not to do" (137) (p. 188).

In a similar vein, Wang (138) studied the influence of the technological congruence on consumers' attitudes towards price change and technology adoption. The study found that the perceived technological congruence significantly influences consumers' adoption intentions. Wang defined the perceived technological congruence as "the degree to which a new mobile application is perceived by users as being consistent or compatible with their values and travel needs" (p. 23). Furthermore, in his book *Diffusion of Innovations* (139), Everett Rogers discussed the impact of perceived attributes of the innovation on the rate of adoption, including the attribute of compatibility. One way of evaluating compatibility was through the values and belief system. Rogers argued that incompatibility with the values and beliefs slows down the adoption rate. This is consistent with our proposal that the incompatibility that results from the incommensurability between the EBM and the data science paradigms in the way normal science is practiced plays a role in the slow adoption of AI (particularly ML approaches) in clinical decision making.

Clinicians' Awareness of ML Potential

Oh et al. (140), who investigated the awareness of Korean doctors of AI in medicine, found that less than 6% of the 669 surveyed doctors have a sort of familiarity of the AI in medicine. Kelly et al. (141) argued that raising the awareness amongst clinicians contributes to improving the adoption rate. The importance of raising the awareness of the users about the potentials of a new technology was also discussed by Kamal et al. (142). They differentiated

between the non-users due to resistance and due to lack of awareness. Therefore, they recommended raising awareness as a measure to overcome resistance and to improve adoption. They suggested awareness campaigns to help reduce resistance and to enhance knowledge about the potential benefits of adopting the technology, to contribute to improved adoption.

Similarly, it was reported that the awareness of the benefits of the technology was a limitation in studying the critical care nurses' intentions to use the eICU telemedicine technology (143). Interestingly, Oh et al. (140) came across a very important perspective which is the concern that AI will replace physicians' roles in the future, which could be a reason for resistance. In our opinion, this is a consequence of poor awareness. In this work's context, awareness is meant to target practicing clinicians who face patients and clinical researchers who are well-versed with EBM standards. There is a significant role to be played by government health ministries in encouraging the advent of AI in the healthcare industry, and in formulating strategies that ensure the readiness of the healthcare industry for the 4th industrial revolution.

Patients' Awareness of ML

Patients are the heart of the healthcare system. This is why there are numerous calls to build healthcare delivery systems that are patient-centered (144). As a result, patient engagement in developing care plans is becoming increasingly common in modern healthcare systems. In the context of the ML, there are several concerns that must be addressed to ensure patients' trust (145). Addressing these concerns requires raising patients' awareness of the potentials of ML to improve the quality of clinical decisions and therefore treatment outcomes. In a study conducted by Jaiswal et al. (146) to assess

patients' awareness of the role of the AI in dentistry, 77% of the surveyed patients reported that they do not know that AI can help in clinical decision making. In another study conducted by Jutzi et al. (145) to assess patients' perspectives about the role of AI in skin cancer, the majority of respondents were not amenable to the use of AI as a standalone system, and they preferred a joint system where AI supports doctors in clinical decisions related to skin cancer classification. Therefore, raising patient awareness is essential to eliminate misconceptions about the role that ML can play in improving healthcare delivery (146). Similar to raising the clinicians' awareness, this is a strategic role that health ministries can assume and encourage healthcare organizations to participate in.

Clinicians' Perceived Usefulness

Perceived usefulness is a concept of TAM referring to the degree to which users perceive or believe that the adopted technology will improve their performance (147). Davis attached the perception of usefulness to the motivation of employees to get promotions or pay increases. In our context, we consider perceived usefulness a function of the clinicians' perception that the utilization of the ML approaches will improve their capacity to make better clinical decisions, which will contribute to better treatment outcomes, improved service user satisfaction, and minimized healthcare bills for clients, healthcare systems, and national economies. It is evident in the literature that there is a significant positive relationship between users' perceived usefulness and their behavioral intentions to use a technology (142, 148, 149). We believe that the awareness of the ML potentials contributes to perceived usefulness.

Clinicians' Perceived Ease of Use

Davis (147) defined perceived ease of use as "the degree to which a person believes that using a particular system would be free of effort" (p. 319). He related it to the relative "ease" of the new technology when compared with the old way of doing things (in the absence of the innovation). In our context, we refer the perceived ease of use as clinicians' knowledge of how to use ML approaches, interpret results, and employ the obtained knowledge in their clinical decisions and scientific activities. ML techniques are usually referred to as a black-box, and there have been several calls to unlock them (150, 151). This contributes to the spread of the perception of the complexity of ML, which slows down adoption. The positive impact of the perceived ease of use on the users' intentions to use technology or innovation is well established in the literature (142, 148, 149).

Academic Foundation

Generally, clinicians are trained to conduct their clinical decisions and judgements according to the standards of EBM. However, there is no formal training in modern analytical techniques in clinical schools or healthcare organizations to orientate the future clinicians about the use of ML to enhance quality of clinical decisions. Lack of training was identified as a barrier to electronic health record adoption among physicians (152), and a barrier to adopting telemedicine technology among ICU nurses (143). Several recent studies stressed the importance of providing training to the clinicians and of introducing AI in the clinical sciences curricula to prepare the healthcare industry for the AI age (153, 154), moving on from the information age to the 4th industrial revolution (155). This step requires serious collaboration between the

healthcare, education, and government sectors to upgrade curricula with AI inclusion as a precursor to the effective use of ML in healthcare. The proposed reconciliatory framework forms the basis of an updated curriculum that legitimizes the utilization of the data-based paradigm alongside the EBM paradigm.

Multidisciplinary Collaborative Work Environment

This enabler is related to healthcare organizations who can enable ML adoption by fostering a climate conducive to collaboration among different disciplines (i.e. clinical researchers, epidemiologists, biostatisticians, data scientists, ML experts, etc.). By achieving this, healthcare organizations can facilitate clinicians' access to data science and ML experts, who can help clinicians and the clinical researchers unblock the ML black-box and help them better understand pertinent clinical evidence (72). Also, the collaboration between data scientists, epidemiologists, and clinical researchers can help the former develop deeper understanding of EBM best practices and standards.

Clinicians' Trust in ML

Although ML can help improve risk calculation in uncertain conditions, the use of ML raises several clinical, ethical, and legal concerns due to the lack of understanding how ML produces output (156). The complexity of ML approaches, lack of procedural transparency, and lack of interpretability increase the ambiguity around the ML, which undermine clinicians' trust in ML outputs (68, 78, 121, 156), and raise questions about accountability for wrong decisions (140).

According to Shahbaz et al. (149), trust is a state of mind in which a person has a confidence in the information provided by a system. In our context,

the trust refers to the perception of clinicians that ML is capable of providing valuable insights from the clinical data in a way that improves patient outcomes. Several studies prove that trust is a key reason for the success or the failure of information systems adoption (142, 149, 157, 158).

Patients' Trust in ML

Patients' perspectives have been thoroughly discussed in the context of ML in healthcare, identifying key concerns related to the huge potential of ML and the risks of breaching patients' privacy, confidentiality, and anonymity. Several scholars discussed that applications of ML may jeopardize the bioethical framework that stresses the four key ethical principles of autonomy, beneficence, non-maleficence, and justice. EBM particularly stresses autonomy and respecting patients' right of choice. Nevertheless, some scholars argued that the lack of transparency of the ML procedures leaves patients with no choices, and makes them feel pushed to choosing certain treatment options, which jeopardizes the very principle of autonomy (74). Furthermore, several scholars raised concerns about the ability of the ML approaches to uncover very sensitive information that patients may not like to be uncovered, which risks the patients confidentiality (79, 159).

According to He et al. (9), improving the interpretability of the ML output enhances patient trust in AI in general, and holds AI manufacturers more accountable. Therefore, there is great emphasis on the importance of improving the transparency of the patient consenting procedures and to design strict rules to save the patients' rights from function creep and data misuse, which will subsequently enhance patient trust in ML approaches (74, 159, 160).

Clinicians' Behavioral Intention to adopt ML

Behavioral intentions refer to someone's intention to achieve a specific behavior in the future (149). This TAM variable is largely used as a predictor for the actual acceptance of an innovation or technology (142, 161-163). This assumption is based on Ajzen's theory of planned behavior, which posits that behavioral intention can predict actual behavior (137). In other words, it is more likely that people will adopt a technology if they have the intention to use it (164). Consequently, we propose that clinicians' intentions to use ML in their practice will affect their actual adoption of ML.

Step II: Determine Contextual Relationships

This step is imperative for developing a structure of interacting elements, represented in the structural self-interaction matrix (SSIM) shown in Table 5, where expected contextual relationships between the enablers are presented in the form whereby enabler a influences or enhances enabler b, or enabler a helps achieve enabler b, and so forth (124, 128). The contextual relationships are identified from the literature and the expert panelists opinion. We use "V" if enabler a (in the column) enhances enabler b (in the raw); "A" if enabler b (in the raw) enhances enabler a (in the column); "X" if both variables enhance each other; and "O" if there is no relation.

Table 4. SSIM Matrix

	Enablers	10	9	8	7	6	5	4	3	2	1
1	Perceived normative congruence	V	O	V	V	A	O	V	O	A	X
2	Clinicians' awareness	V	V	V	X	A	V	V	O	X	
3	Patients' awareness	O	V	O	O	O	O	O	X		
4	Perceived usefulness	V	O	V	X	A	X	X			
5	Perceived ease of use	V	O	V	A	A	X				
6	Academic foundation	V	O	V	V	X					
7	Multidisciplinary collaborative work Environment	V	O	V	X						
8	Clinicians' trust	V	V	X							
9	Patients' trust	V	X								
10	Clinicians' intentions to adopt ML	X									

Step III: Interpret the Relationships

This step helps explain the way in which enabler a helps achieve or enhance enabler b, which helps scholars achieve in-depth knowledge about the relationships between the elements (124, 134).

Step IV: Interpretive Logic of Pair-Wise Comparison

This step identifies the possible directional links (a-b and/or b-a) to achieve an interpretive knowledge base. Making use of the concept of the interpretive matrix, the scholar in this step fully interprets each paired comparison in terms of how that directional relationship operates in the system under consideration by answering the interpretive query, as explained in step

III (124). The total number of pair-wise comparisons equals to n x (n-1) (128). As explained by Sushil (124), there are two possible directional links for every pair of elements (a-b or b-a). The entry could be yes (Y) or no (No) for every a-b link. If it is (Y), further interpretation is required.

Step V: Development of Reachability Matrix and Transitivity Test

As explained by Jena et al. (134), the reachability matrix is developed with the help of the interpretive knowledge base by transforming every Y to "1" and N to "0". Also, reachability matrix is checked for transitivity rule and updated until full transitivity is established. For every new transitive link, N is to be changed to Y. Table 6 shows the reachability matrix including the significant transitive relationships (1*).

Table 5. Reachability Matrix

	Enablers	1	2	3	4	5	6	7	8	9	10	Driving power
E1	Perceived normative congruence	1	0	0	1	1*	0	1	1	1*	1	8
E2	Clinicians' awareness	1	1	0	1	1	0	1	1	1	1	8
E3	Patients' awareness	0	0	1	0	0	0	0	0	1	0	2
E4	Perceived usefulness	0	0	0	1	1	0	1	1	1*	1	6
E5	Perceived ease of use	0	0	0	1	1	0	1*	1	1*	1	6
E6	Academic foundation	1	1	0	1	1	1	1	1	1*	1	9
E7	Multidisciplinary collaborative work Environment	0	1	0	1	1	0	1	1	1*	1	7
E8	Clinicians' trust	0	0	0	0	0	0	0	1	1	1	3
E9	Patients' trust	0	0	0	0	0	0	0	0	1	1	2

	Enablers	1	2	3	4	5	6	7	8	9	10	Driving power
E10	Clinicians' intentions to adopt ML	0	0	0	0	0	0	0	0	0	1	1
	Dependence power	3	3	1	6	6	1	6	6	9	9	

Step VI: Level Partition on Reachability Matrix

This step is important to place the enablers in the appropriate digraph level (134). In this step, we determine the intersection set, which is made of the common enablers between the antecedents set and the reachability set. The reachability set is composed of the enabler itself and the enablers that it enhances. The antecedents set is the enabler itself and the enablers that enhance it. Consequently, the elements where the intersection set equals the reachability set will sit in the top of the hierarchy. The same procedure is repeated until all enablers' levels are determined. Tables 7 to 12 show all the iterations that resulted in the enablers' level determination. The symbol (*) is assigned to the enabler(s) where the reachability and the intersection sets are equal.

Table 6. Partition Level (a)

Enabler	Reachability set	Antecedent set	Intersection set
E1	1,4,5,7,8,9,10	1,2,6	1
E2	1,2,3,4,5,7,8,9,10	2,6,7	2,7
E3	3,9	2,3	3
E4	4,5,7,8,9,10	1,2,4,5,6,7	4,5,7
E5	4,5,7,8,9,10	1,2,4,5,6,7	4,5,7

Enabler	Reachability set	Antecedent set	Intersection set
E6	1,2,4,5,6,7,8,9,10	6	6
E7	2,4,5,7,8,9,10	1,2,4,5,6,7	2,4,5,7
E8	8,9,10	1,2,4,5,6,7,8	8
E9	9,10	1,2,3,4,5,6,7,8,9	9
E10	10*	1,2,4,5,6,7,8,9,10	10*

(10*): clinicians' behavioral intentions to adopt ML

Table 7. Partition Level (b)

Enabler	Reachability set	Antecedent set	Intersection set
E1	1,4,5,7,8,9	1,2,6	1
E2	1,2,3,4,5,7,8,9	2,6,7	2,7
E3	3,9	2,3	3
E4	4,5,7,8,9	1,2,4,5,6,7	4,5,7
E5	4,5,7,8,9	1,2,4,5,6,7	4,5,7
E6	1,2,4,5,6,7,8,9	6	6
E7	2,4,5,7,8,9	1,2,4,5,6,7	2,4,5,7
E8	8,9	1,2,4,5,6,7,8	8
E9	9*	1,2,3,4,5,6,7,8,9	9*

(9*): patient trust

Table 8. Partition Level (c)

Enabler	Reachability set	Antecedent set	Intersection set
E1	1,4,5,7,8	1,2,6	1
E2	1,2,3,4,5,7,8	2,6,7	2,7
E3	3*	2,3	3*
E4	4,5,7,8	1,2,4,5,6,7	4,5,7
E5	4,5,7,8	1,2,4,5,6,7	4,5,7
E6	1,2,4,5,6,7,8	6	6
E7	2,4,5,7,8	1,2,4,5,6,7	2,4,5,7
E8	8*	1,2,4,5,6,7,8	8*

(3*& 8*): patient awareness and clinicians' trust

Table 9. Partition Level (d)

Enabler	Reachability set	Antecedent set	Intersection set
E1	1,4,5,7	1,2,6	1
E2	1,2,4,5,7	2,6,7	2,7
E4	4,5,7	1,2,4,5,6,7	4,5,7
E5	4,5,7	1,2,4,5,6,7	4,5,7
E6	1,2,4,5,6,7	6	6
E7	2*,4*,5*,7*	1,2,4,5,6,7	2*,4*,5*,7*

(2*, 4*,5* & 7*): clinicians' awareness, usefulness, ease of use and the collaborative multidisciplinary environment

Table 10. Partition Level (e)

Enabler	Reachability set	Antecedent set	Intersection set
E1	1*	1,6	1*
E6	1,6	6	6

(1*): perceived normative congruence

Table 11. Partition Level (f)

Enabler	Reachability set	Antecedent set	Intersection set
E6	6*	6	6*

(6*): building academic foundation

Step VII: Develop TISM Diagraph

This step aims to portray the elements in the form of a directed graph in which the enablers are ordered in levels based on the relationships that are shown in the reachability matrix (128, 134). Before developing the TISM diagraph, we conducted MICMAC analysis in order to determine the driving and the dependence power of the enablers (134). MICMAC analysis also helps us position the enablers in the right level within the TISM diagraph, which

contributes significantly to drawing the roadmap that guides the strategic directions to enhance the clinicians' intentions to adopt ML in their clinical decisions. MICMAC analysis clusters the enablers into four clusters based on their driving and dependence power as described below.

Cluster A: Autonomous Enablers

These enablers have weak driving and dependence power; compared to other enablers, they are detached from the system (128). Raising patient awareness is found to be an autonomous enabler with a driving power of 2 and dependence power of 1.

Cluster B: Dependent Enablers

These enablers have very weak driving power but strong dependence power. They usually sit atop the hierarchical diagraph as an outcome. We have three variables in this cluster; clinicians' trust, patients' trust, and clinicians' intentions to adopt ML. The most dependent enabler is the clinicians' intentions to adopt ML, which is basically the outcome of the TISM-based conceptual framework.

Cluster C: Linkage Enablers

These enablers have strong dependence and driving power. They are considered unstable enablers, because any action is done on these enablers will influence the whole system (134). They usually sit in the middle levels of the diagraph. We have three linkage enablers: perceived usefulness, perceived ease of use, and building a multidisciplinary collaborative environment.

Cluster D: Independent Enablers

These enablers have strong driving power and weak dependence power. Therefore, they are considered the key or the strategic enablers. They

usually sit at the bottom of the TISM diagraph. We have three dependent enablers: perceived normative congruence, clinicians' awareness, and building academic foundation, which have the highest driving power of 9 and the least dependence power of 1. Figure 5 shows the MICMAC analysis of the enablers.

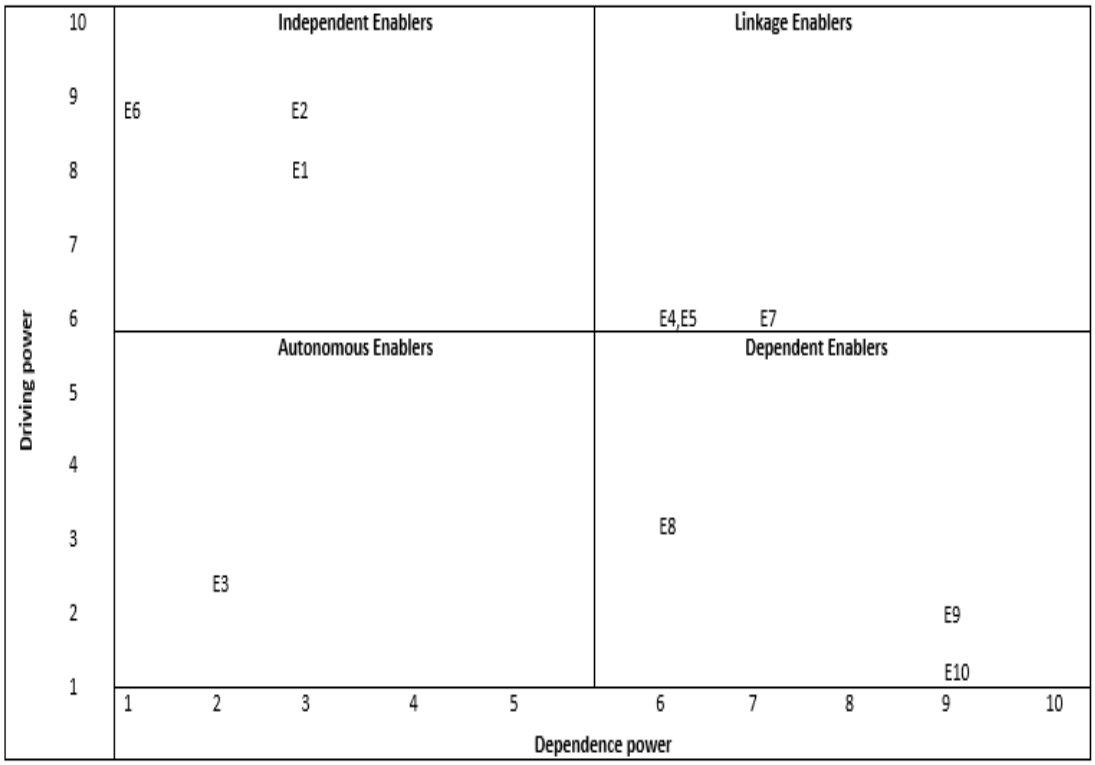

Figure 5. MICMAC analysis of enablers.

Step VIII: Develop the Interpretive Diagraph (Interaction Matrix)

In this step, the scholars convert the final diagraph into a binary interaction matrix, using "1" to indicate direct and significant transitive link (134). In addition, the scholars will develop an interpretive matrix through providing relevant interpretation, from the interpretive knowledge base, for every cell that

contains a “1” entry (124, 128).

Step IX: Total Interpretive Structural Model

With the help of the diagraph and the interpretive matrix, the TISM with the identified elements is developed. According to Sushil (124), the interpretation of the elements that are placed in the box replace the nodes in the diagraph. This leads to total interpretation of the structural model in terms of the nodes and the interpreting links (134). Figure 6 shows the final TISM diagraph.

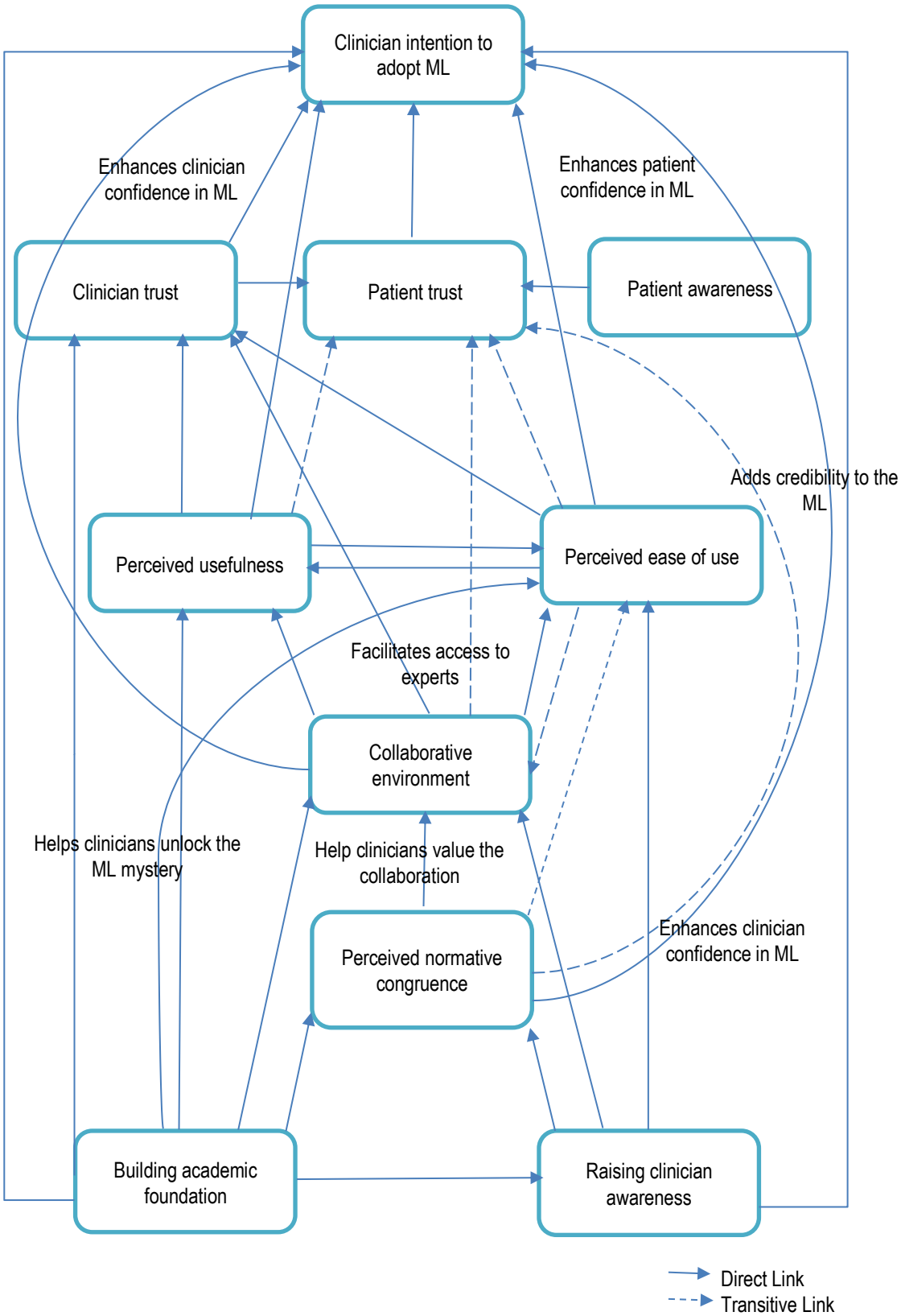

Figure 6. Final TISM diagraph.

Discussion

The purpose of this chapter is to propose a conceptual framework that (a) determines the interrelationship between the enablers for the adoption of the ML to support clinical decision making; and (b) depicts the enablers in a hierarchical diagraph to provide a roadmap that assists policy makers, clinical researchers, and clinicians to enhance ML adoption in supporting clinical decision making. The driving power and dependence power provide valuable insights for the policy makers to set strategies to enhance the willingness of clinical practitioners and researchers to consider ML approaches in evidence production and clinical decision support.

The enablers are ordered in the diagraph based on their driving and dependence power. The enablers that have the highest driving power with the lowest dependence power sit at the bottom, while the enablers that have the lowest driving power and the highest dependence power sit on the top (127, 128). Also, the position of the enabler in the hierarchy determines which among the enablers is considered strategic, and achieving these enablers may require interventions from ministries or academic institutions. Moving to higher positions indicates that the responsibility moves to the healthcare organizations or individual clinicians, until we reach to the outcome enablers that in turn leads to achieving the ultimate goal of enhancing ML adoption in healthcare, particularly in supporting clinical decision making.

The diagraph has six levels based on the driving and the dependence powers that were identified by the reachability matrix and MICMAC analysis. It is obvious that patient awareness is an autonomous enabler. This means that this enabler has weak dependence on the other enablers to achieve its impact

on the outcome. Accordingly, policy makers may need to think of strategies to raise patient awareness in line with building academic foundations and raising clinicians' awareness, which sit at the bottom of the diagraph.

Figure 6 shows that building an academic foundation and raising clinicians' awareness have the maximum driving power of 9 and the lowest dependence power (1 and 3, respectively). This means that these two enablers are the building blocks of any strategy aiming to enhance the adoption of the ML in healthcare. It is important to differentiate between these two enablers. Building an academic foundation starts with introducing the ML and AI to the clinical sciences curricula; subsequently, clinicians learn from an early stage that evidence can be obtained in several ways, including data-driven methods, wherever theory-driven methods are not feasible. On the other hand, awareness should target the clinical practitioners and researchers to demonstrate evidence that ML can improve their decisions, and improve patient and organizational outcomes. These two building blocks have a lot to do with the health ministries and the higher education ministries. Accordingly, more research is required to study the perspectives of the health and education ministries to embrace the AI and ML as means to improve clinical decisions and to effectively contribute to the overall economy, through effectively managing the substantial healthcare expenditure via improved clinical decision making.

The perceived normative congruence is another independent enabler. Nevertheless, its driving power is 8 and dependence power is 3, which puts it in the 5th level. The TISM model proposes that building the academic foundation and raising clinicians' awareness, along with adopting the pragmatic reconciliatory framework, will contribute to enhancing the perception of the

normative congruence, which will contribute to enhancing clinicians' willingness to use data-driven evidence in their clinical decision making. We conclude that the perceived normative congruence will be achieved through enhancing the clinicians' awareness and building the academic foundations, which simultaneously plays a significant driving role in influencing individual clinicians and organizational perspectives towards ML adoption.

At the operational level, healthcare organizations have to encourage collaboration between clinical researchers, ML experts, data scientists, and epidemiologists in order to: (a) exchange experiences; (b) facilitate mutual understanding of the potentials and the challenges of both paradigms; and (c) enhance clinicians' access to experts and improve the interpretability and the operability of clinical research findings. Accordingly, building a multidisciplinary collaborative research environment sits on the 4th level, with driving power of 7 and dependence power of 6. All the mentioned enablers facilitate clinicians' perceived ease of use and perceived usefulness, which are key linkage enablers for enhancing the behavioral intentions to adopt ML. Therefore, perceived usefulness and perceived ease of use sit in the 3rd level with driving power of 6 and dependence power of 6.

The MICMAC analysis (Figure 5) identified three dependent enablers (clinicians' trust, patients' trust, and clinicians' intentions to adopt ML). Both clinicians' trust and patients' trust sit on the second level, as they both drive clinicians' intentions to adopt ML, with driving power of 3 and 2 respectively, and dependence power of 6 and 9 respectively. Patient awareness was also placed in the second level, because it is related only to patient trust, with driving power of 2 and dependence power of 1. Finally, the clinicians' intention to adopt

ML serves as the outcome enabler that facilitates the actual adoption of the ML, with a driving power of 1 and dependence power of 9. The diagraph demonstrates a strong logical path towards the enablement of the ML adoption in supporting the clinical decision. This direction requires collaboration between ministries, healthcare, and academic institutions, as well as clinical practitioners and researchers.

Conclusion

The present work helps explain the *what*, *why*, and *how* questions that are essential for theory-building exercises. The TISM-based framework answers the “why” question through presenting the factors that impede the utilization of ML in clinical decision-making. Therefore, reversing these factors facilitates the enablement of the adoption of ML in supporting the clinical decision-making process. It also answers the “what” question through defining every enabler and explaining the mutual relationships between the enablers to comprehensively understand the dynamicity of the interactions. Therefore, understanding the type and the dynamicity of the relationships between the various enablers answers the “how” question through providing a systematic approach to guide the strategic efforts to enhance the adoption of ML in supporting clinical decision-making.

Through the MICMAC analysis, we were able to determine the level of every enabler in the TISM diagraph. This categorization of enablers is crucial for determining what enablers have to be addressed first. It is obvious that significant work is required from academic institutions to instill the seeds of data-driven approaches early during the undergraduate phase, to prepare a generation of clinicians ready to cope with the 4^{th} industrial revolution and

advance the healthcare industry. Then, step-by-step the ML-EBM reconciliation can be achieved and institutionalized to form the new way of producing clinical evidence that informs the clinical decision-making.

Limitations and Future Research Opportunities

This work provided a novel conceptual framework that guides the efforts to enhance the adoption of ML in healthcare. The significance of this work is that it answers the questions of why healthcare is the slowest industry to adopt ML approaches, and how ML adoption can be enhanced in the healthcare industry. The diagraph provides a specific path that guides policy makers in their journey to support clinical decision making, which has both wellbeing and economic benefits for patients, organizations, and the whole economy. Nevertheless, the interrelationships between the enablers and the resultant driving and dependence power were determined based on the expert panelists' opinions, which are subject to bias. Accordingly, there is a need for further work to validate the framework and ensure that this path will lead to improving the adoption of ML in the real business world, which is out of the scope of this work. Furthermore, the study addresses the enablers/ barriers from clinicians' perspectives. Therefore, to fully answer the question about the poor adoption in the healthcare industry, significant work is needed to address the matter from organizational perspectives, considering issues like infrastructure and managerial support, etc.

The next section presents the outcomes of collaborating with the Trauma Surgery Section in Hamad Medical Corporation (HMC) to develop ML prognostic and predictive models to help address specific clinical problems that concern the trauma surgery clinical and administration personnel. We adhered

to the reconciliatory framework that we proposed in the previous chapter to provide empirical evidence about the utility of the framework in supporting clinical decision.

CHAPTER 5: UTILIZING ML IN SOLVING REAL CLINICAL PROBLEMS – EMPIRICAL EVIDENCE

Background

Hamad Medical Corporation (HMC) started to adopt electronic health records during the past decade as part of Qatar's initiative to digitize the public healthcare system. The advent of electronic health records allows the collection of vast volumes of data in a data warehouse that brings opportunities to provide real-time evidence that could support frontline clinicians to make better clinical decisions. The Trauma Surgery Section in HMC that administers the level 1 trauma system established the National Trauma Registry (NTR) in 2007. This step contributed to a leap in HMC's capacity to benchmark with the most reputable trauma centers in the world. In addition, the Registry provided great opportunity to enhance the education and research activity for the trauma surgery and the corporation as whole.

The NTR is a member in the National Trauma Data Bank (NTDB), which is considered the largest aggregation of trauma registries in the world. NTDB provides a standard data dictionary with operational definitions for all collected data elements. Trauma surgery registration is an essential element in leading performance improvement efforts, making data available to the trauma surgery leadership to monitor provider and system performance and determine improvement needs.

The collaboration with Trauma Surgery Section aimed to address selected real clinical problems through the utilization of ML approaches. The objectives of this collaboration were to: (a) design a prognostic model that helps quickly predict mortality in patients with TBI in order to provide insight to

clinicians to consider early mortality preventive measures; (b) design an ML model to predict TBI patients at risk of requiring prolonged mechanical ventilation (PMV), in order to help redefine PMV and provide insights into the optimal time for tracheostomy; (c) design a prognostic model to predict the patients at risk for developing ventilator-associated pneumonia (VAP) following TBI, in order to provide an actionable insight that supports the preventive measures; and (d) design a model that predicts patients at risk for prolonged length of stay (PLOS) in the Trauma Surgery Section following TBI, which is important for enhancing the bed turnover rate, reducing unnecessary costs, and preventing medical complications that are associated with the PLOS.

Methodology

The project was approved by the Institutional Review Board (IRB MRC-01-19-106) of HMC. This project targeted all adult patients registered in the NTR who were admitted to the Level 1 Trauma Center at (L1TC) at Hamad General Hospital (HGH), a not-for-profit state healthcare service operated by HMC in Qatar, in the period from January 2014 to February 2019. All the studies in this project were conducted in accordance with the Cross-Industry Standard Process for Data Mining (CRISP-DM), which provides definitions of six typical phases of data mining projects: business and data understanding, data preparation, modeling, evaluation, and deployment (44). A total of 2318 patients with TBI were registered in the NTR for the given period. All patients' data were anonymized and saved only with the principal investigator. Figure 7 provides a summary of the methodology used in the whole project.

Business and Data Understanding

Not all the NTR data were usable in this project. Therefore, to better

understand and chose meaningful variables, the research team explored the definition of each variable in the NTR data dictionary, and reviewed the literature in order to determine which among the enormous number of recorded variables need to be considered predictors, and which among them to be imputed if in case they have missing values (88). Pediatric patients (< 14 years old) were excluded. This was important for understanding and interpreting the results, as some of the important parameters (i.e. vital signs) are different between the pediatric and adult patients.

Data Preparation

Only adult patients (≥ 14-years-old) who sustained TBI were included in the study. All variables that have no predictive power (e.g. health record number, date of admission, and date of disposition were excluded. Missing data may seriously affect predictive models' performance (165). Several approaches to handle data omissions have been used, such as the elimination of incomplete records (166) or imputing missing values, which is a widely used approach (165). Therefore, apart from the "time from injury until arrival to emergency" variable, which was considered crucial in predicting the VAP, we did not consider imputation in any of the included studies. Instead, the records with missing data were eliminated. Subsequently, 1620 eligible patients' records were included in the project. Figure 8 shows the records inclusion/ exclusion process.

The retrieved data included the following variables: age, gender, mechanism of injury, mode of arrival, alcohol blood level, blood pressure, heart rate, Glasgow Coma Score (GCS), CT findings, intubation status and location, date/time of injury, time of admission to the emergency department (ED),

injuries characteristics, injury severity score (ISS), abbreviated injury score (AIS), known comorbidities, performed procedures, blood transfusion, administration of venous thromboembolism (VTE) prophylaxis, in-hospital complications, outcome, and date of disposition.

Additional variables were secondarily generated from the retrieved variables: shift of admission (D: 7am to 6:59 pm and N: 7 pm to 6:59 am) (167, 168); and time from injury to emergency room, which was used in predicting VAP (169).

Outcome Measures

The Trauma Surgery Section identified four outcome measures: mortality, prolonged mechanical ventilation (PMV), ventilator associated pneumonia (VAP), and prolonged length of stay (PLOS). In addition, we conducted a subgroup analysis to predict the mortality in patients who sustained TBI and received mechanical ventilation (MV). All the outcome variables were binary, whereby the occurrence of the outcome was coded 1, and non-occurrence was coded 0.

Prediction Models

We utilized a variety of ML algorithms to allow us to compare their performance with each other and with past literature, in order to identify the optimal model for deployment. We adopted the standard practices to train the model and to prevent overfitting through data partitioning and optimizing the models' hyperparameters. We benchmarked our data partitioning procedure with the published literature. Commonly, 70-75% of the data is used for training the model and 25-30% is used for testing the model. The proportions are changeable according to the overfitting evaluation. We used SPSS modeler

18.1 to design the algorithms and conduct the analysis.

Models Evaluation Metrics

The literature presented various metrics to evaluate ML models' prediction performance. In this work, we used standard set of evaluation metrics to ensure that we evaluated predictive performance comprehensively. The evaluation process starts with developing the confusion matrix, which shows the difference between the actual and predicted classification (true positive/negative and false positive/negative). Table 13 shows the evaluation metrics and their formulae.

- True positive (TP): An outcome that the prediction model correctly labels positive.
- True negative (TN): An outcome that the prediction model correctly labels negative.
- False positive (FP): An outcome that the prediction model incorrectly labels positive.
- False negative (FN): An outcome that the prediction model incorrectly labels negative.

Table 12. Evaluation Metrics of Prediction Models

Metric	Formula
Accuracy	(TP + TN) / (TP+TN+FP+FN)
Sensitivity	TP / (TP+ FN)
Specificity	TN / (TN + FP)
Precision (positive predictive value)	TP / (TP + FP)
Negative predictive value (NPV)	TN / (TN + FN)
F-score	(2 * Precision*Sensitivity) / (Precision + Sensitivity)

In addition to the above metrics, we used the Area Under the Receiver Operating Characteristics Curve (AUC), which is another metric for evaluating models' accuracy and discrimination capacity (93). Models with 100% wrong predictions have an AUC of 0, while those with 100% accurate predictions have an AUC of 1; when the AUC is 0.5, the model has no separation or discrimination capacity.

Presentation of Results

Results were presented in several ways based on the study objectives and the chrematistics of the prediction model. The presentation of results is an essential step to effectively communicate potentially actionable insights that may support clinical decision making. For example, the results of the LR models were tabulated to show the predictors' significance and odds ratio. Furthermore, we illustrated the predictors' importance for all prediction models consistent with the common practice in ML studies. In ML, the contribution of every predictor to the overall model's capacity to produce accurate predictions is usually presented in the form of predictors' importance (90). The first predictor is

usually the most important predictor relative to the model's capacity, after which other predictors' importance values are ranked in relation to the first ranked predictor.

Collaborative Work Environment

We emphasized the involvement of HMC trauma surgery clinicians, the NTR administrator, and the Trauma Surgery Research Director in every step in this project, to satisfy the key principles for the reconciliation framework that we proposed in the previous chapter. All the studies were conducted based on their aim to obtain a new insight about real clinical problems they are facing in the Trauma Surgery Section. Mutual understanding was reached about every variable, and insights were exchanged throughout the project. Importantly, all the findings and reported outcomes were refereed by senior trauma surgery consultants to determine their consistency with the past literature and their potential to influence the clinical practices. As a result of this collaboration, there are serious discussions to embed selected ML algorithms in the NTR and electronic health records to provide real-time clinical decision support to the trauma clinicians in selected scenarios.

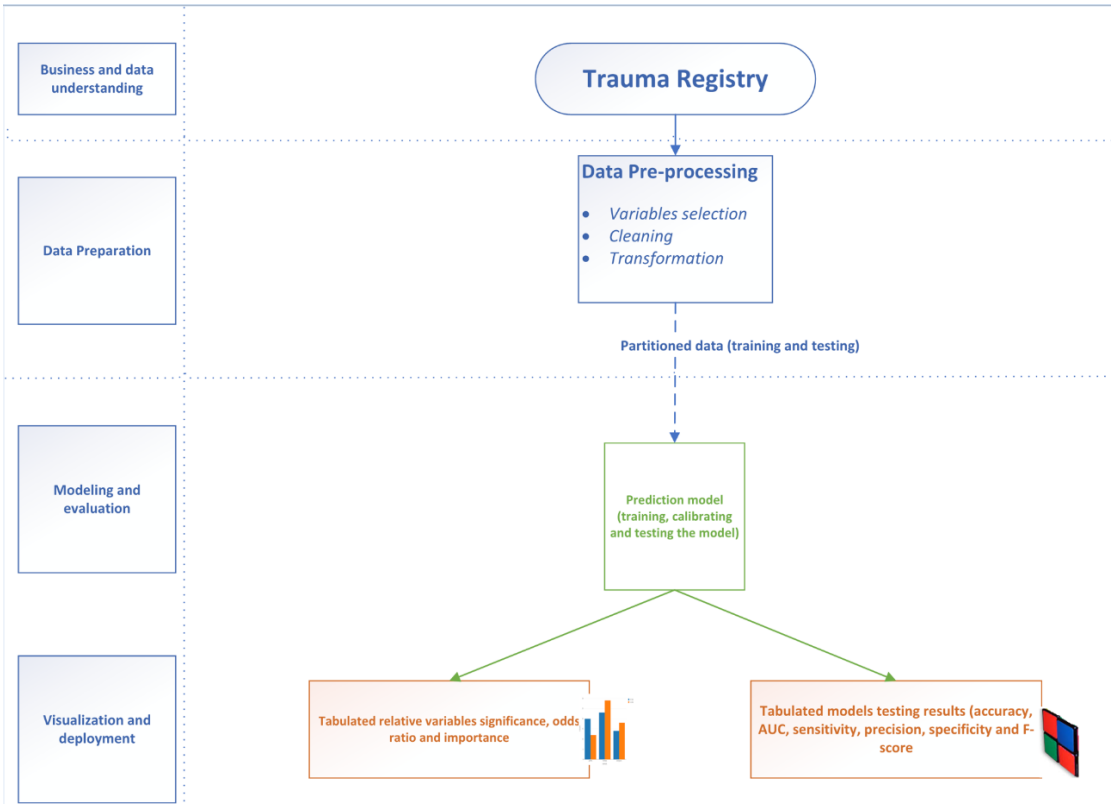

Figure 7. Summary of research methodology.

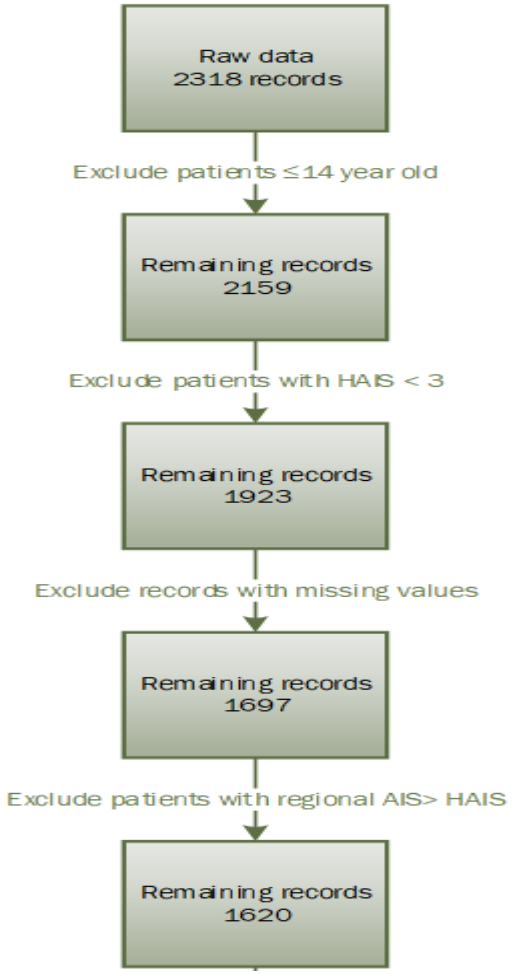

Figure 8. Records inclusion and exclusion process.

Limitations

One of the most important limitations in this project was faced during data processing and preparation. Although the data repository in the NTR is complaint with the standards of the NTDB and the Trauma Quality Improvement Program (TQIP) of the American College of Surgeons-Committee on Trauma (ACS-COT), which provide uniform data elements, several variables in the NTR are recorded as text-free and are prone to dictation errors, which complicates the data preparation process. Therefore, there were several unobtainable but potentially useful predictors, such as laboratory results, received medications,

surgical procedure and diagnostic study details.

Secondly, the empirical work was limited to the predictive modeling, because the agreement with the trauma surgery was to design predictive models to help address some real clinical problems. Therefore, the sample was limited to TBI patients, which left a relatively small sample. Hence, the potential of conducting reasonable quality clustering algorithms was minimal.

Furthermore, the deployment of the ML models to support clinical decision making was another significant challenge due to several reasons, such as the questionable reliability of the non-traditional predictive techniques, attributable to a certain extent to the lack of awareness among clinicians about AI potentials to support clinical decision-making processes. Very importantly, unlike the logistic regression for instance, the standardized coefficients and the odds ratios pertaining to each predictor in the SVM or in the ANN were not obtainable, which makes the interpretation of results more complex compared to using traditional computational techniques.

The sample size in some of our studies was small. For example, the sample in predicting mortality in mechanically ventilated patients was 785 records, this is considered small for studies that use ML techniques. The relatively small sample size complicated data processing and partitioning, and model training, validation, and testing. Nevertheless, the size of data set used and the variables included in the study are still comparable with the previous studies, supporting the validity of the results.

Furthermore, the lack of standard definitions was a challenge in some studies. For example, there is no agreed upon definition for prolonged mechanical ventilation. This complicates the classification mission. There are

no previous studies that define PMV to be 10 days. This cut-off was determined based on the result that tracheostomy conducted after 10 days of mechanical ventilation is considered late tracheostomy, and is associated with unfavorable outcomes. However, we utilized that definition in an attempt to showcase that ML is capable of providing novel insight that may influence the clinical practice, to reap the evident benefits of early tracheostomy and to overcome the well-known complications of prolonged mechanical ventilation.

CHAPTER 6: PREDICTION OF IN-HOSPITAL MORTALITY IN TBI PATIENTS USING NTR AND ML APPROACH

Background

TBI is defined as a brain injury caused by external trauma (170). TBI causes death and disabilities more than any other trauma (171), significantly reducing life expectancy and increasing the mortality rate by 30-70% (170, 172). TBI affects millions of people around the world yearly, causing a major global burden (172). Globally, 64-74 million individuals around the world are estimated to sustain TBI every year, with the greatest disease burden in the Southeast Asia and Western Pacific regions (171). Mortality is associated with the age and severity of TBI. It was found that the 14-day in-hospital mortality following severe TBI reaches up to 24.5% in adults between 16-65 years, and is greater than 40% in patients over 65 years old (173). There are several published and widely used prognostic/ outcome predictive models that demonstrate good predictive and discrimination power. Table 14 shows some of the widely used prognostic models and their performance as measured by AUC (5-15).

In addition, scholars designed several predictive models that aim to help the clinicians and the researchers predict TBI prognosis and outcomes. Jacobs et al. (174) designed a predictive model to predict the outcomes of moderate to severe TBI using demographic and clinical data (e.g. vital signs, pupil reaction, and GCS) and radiological parameters (brain CT scan findings). The study found that age, pupil responses, GCS score, the occurrence of a hypotensive episode post-injury, and having several CT scans are good predictors of TBI outcomes.

The use of ML techniques to predict disease outcomes has grown significantly in the last decade. Several studies prove that the ML predictive techniques outperform classical multivariate techniques (175, 176). In a systematic review of 30 studies that used ML techniques to predict several neurosurgical outcomes (including mortality) following TBI, ML techniques outperformed several well-known classical predictive tools and performed similar or better than field experts in some instances (177). Rau et al. (166) used ML techniques to predict the moderate to severe TBI mortality. The authors used age, sex, use of helmet, co-morbidities, GCS, and vital signs as predictors. They used LR, ANN, DT, SVM, and NB to classify patients based on the survival outcomes. They compared the performance of the models in terms of accuracy, sensitivity, specificity and AUC. ANN yielded the best performance amongst all tested classifiers, with 0.968 AUC, 92% accuracy, 84.4 sensitivity, and 92.8 specificity.

Hale et al. (178) used ANN ML technique to predict six-month favorable/ unfavorable outcomes (including mortality) among 565 pediatric patients who sustained TBI. They used GCS, pupil reactivity to light, blood glucose level, blood hemoglobin concentration, mass lesion, traumatic sub-arachnoid hemorrhage (SAH), cistern status, and midline shift to build the predictive model. They compared the performance of the ANN-based predictive models with three known classical predictive models, namely Helsinki, Rotterdam, and Marshall. The ML model not only achieved profound accuracy (> 94%), but also outperformed the three classical predictive tools. This finding supports Eftekhar et al. (175), who found that ANN significantly outperforms LR-based predictive models in predicting diseases outcomes, with AUC of 0.965 vs. 0.954

(respectively).

This study aims to design a supervised ML predictive model for early prediction of in-hospital mortality in adult patients who sustained TBI and who were admitted to L1TC.

Table 13. Examples of Popular TBI Prognostic Models

Objective(s)	Variables	Performance
Trauma Injury Severity Score (TRISS) – Trauma patients treated at hospitals with or without TBI (179)		
Calculates the probability of survival.	Age, revised trauma score (GCS, systolic blood pressure, respiratory rate), trauma type and injury severity score (ISS) (179).	Good discrimination power. Not specifically designed for TBI (166). Prone to poor performance in severe TBI (179). AUC in previous studies: 0.89 (179), 0.9 (180) and 0.92 (181).
International Mission for Prognosis and Analysis of Clinical Trials in TBI (IMPACT) – Adult patients (age ≥14 years) with TBI and GCS ≤ 12		
Predicts 6-month mortality and unfavorable outcomes (179).	Age, GCS motor scale, pupils reactivity, hypoxia, hypotension, CT results (epidural or subarachnoid hemorrhage), lab values (blood glucose level and hemoglobin concentration) (182).	Good discrimination power. Accurate outcome prediction when large sample size is utilized (179, 182). Poor precision at the individual patient level (183). AUC in previous studies: 0.8 (181), 0.83 (184), 0.85 (180) and 0.86 (182).

Objective(s)	Variables	Performance
Corticosteroid Randomization After Significant Head injury (CRASH) – Adult patients (age ≥ 16 years) with TBI and GCS ≤14 (182)		
Predicts probabilities of 14-day mortality and 6-month unfavorable outcome (182).	Age, GCS, pupil reactivity, major extracranial hemorrhage and CT findings (midline shift, obliteration of third ventricle, subarachnoid hemorrhage, petechial hemorrhage, and non-evacuated mass) (181).	Good discrimination power (182). Accurate outcome prediction when large sample size is utilized (181, 182). Poor precision at the individual patient level (183). AUC in previous studies: 0.86 (180), 0.87 (182) and 0.89 (181).
Marshall Scale – Patients who sustained TBI		
Grades the TBI and predicts the TBI outcomes on the basis of CT scan findings.	Presence of mass lesion, midline shift, and status of the peri-mesencephalic cisterns.	Simple to use. Reasonable discrimination power. Narrow scope (limited to 3 variables). Limited applicability to clinical practice (178). AUC in previous studies: 0.71 (185), 0.635 (186) and 0.78 (178).
Rotterdam CT scoring – Patients who sustained TBI		
Grades the TBI and predicts the TBI outcomes on the basis of CT scan findings.	Presence of mass lesion, midline shift, status of the peri mesencephalic cisterns and the presence of traumatic intra-ventricular or sub-arachnoid hemorrhage (SAH) (185).	Reasonable discrimination power. Does not differentiate between the type and size of the mass lesion (178). AUC in previous studies: 0.698 (186) 0.84 (178) and 0.85 (187).

Helsinki Computerized Tomography Score Chart – Patients who sustained TBI		
Grades the TBI and predicts the TBI outcomes on the basis of CT scan findings.	Mass lesion type and size, presence of intraventricular hemorrhage, suprasellar cistern.	Superior to Marshall and Rotterdam scales. Good accuracy and discrimination power. Lower performance when used alone as a predictive method (178, 186). Reported AUC: 0.717 (178) and 0.746 (186).

Methodology

This study aims to utilize NTR data to design an ML predictive model to predict in-hospital mortality in adult patients who sustained moderate to severe TBI and who were admitted to L1TC.

Only adult patients (aged ≥ 14-years-old) who sustained moderate to severe TBI (HAIS ≥ 3) were included in the study. Patients who sustained other systematic injuries with AIS > HAIS were excluded, to ensure that the primary injury was TBI. Records with missing data were eliminated. Subsequently, 1620 eligible patients were included in the study. The detailed information about the methodology is described in chapter 5.

Outcome Measure

The outcome measure is the in-hospital mortality during the initial hospitalization following moderate to severe TBI. It is a dichotomous variable (0 = alive and 1 = dead). Patients who were discharged from the Trauma Surgery Section or who were transferred to another hospital were considered alive.

Prediction Models

Two of the powerful supervised ML techniques were utilized to allow us

to compare their performance with each other and with previous studies in order to recommend the model that achieves the optimal performance and highest practicality in supporting clinical decisions. ANN and SVM are widely used in predicting in-hospital mortality, therefore they were selected to provide base line comparative performance. SPSS modeler 18.1 was used to conduct the analysis. To prevent overfitting and to validate the models' performance, we partitioned the data into training and testing sets, with the overfit prevention set at 30%. Table 15 explains the data partitions.

Table 14. Data Partitions

Set	Proportion	No. of cases	No. of alive patients	No. of dead patients
Training set	70%	1120	977	143
Testing set	30%	500	440	60
Total	100%	1620	1417	203

Artificial Neural Network (ANN)

ANN is a widely used ML technique that performs powerfully in classification and pattern identification (92). When used for classification, ANN is seen as a set of connected input/output units in which each connection has an associated weight representing the strength of the connection between the units (93). Although scholars consider ANN to be a black-box analytical model, it has great potential to support clinical practice through engagement with EBM (178). Usually, the performance of the neural network is optimized through partitioning data into training and testing data sets, which helps prevent

overfitting. The training continues until the error is not further reducible (188). Once trained, the ANN can be used for future cases where the outcome is unknown (189).

Support Vector Machine (SVM)

SVM is a powerful classification ML algorithm that can be used for linear and non-linear data sets (92). When using SVM for classification purposes, it is very important to decide which kernel function better achieves the optimal hyperplane that separates the classes (190). Linear kernel was used in this study, as it optimized the predictive performance in the preliminary assessment compared to other functions (i.e. polynomial, sigmoid or Radial Basis functions).

Results

Among the 1620 patients who were included in this study, 203 (12.5%) died in the hospital during their initial hospitalization. The mean age was 34.4 years, and the mean age at death was 37.2 years. The most common mechanism of injury was fall from height (34%) followed by motor vehicle crash (30%). The most common CT finding/mass lesion was subdural hemorrhage (28.1%), followed by extradural hemorrhage (22.9%) with 22% of the patients sustaining midline shift. Tables 16 and 17 show the sample characteristics and the descriptive statistics for the study sample.

Table 15. Sample Characteristics – Continuous Variables

Variable	N	Mean	SD	Mean at death
Age	1620	34.4	13.9	37.2
ED systolic blood pressure (SBP)	1620	127.66	22.6	118
ED heart rate (HR)	1620	93	22.9	108.5

Table 16: Sample Characteristics – Nominal and Ordinal Variables

Variable	Category	Count/%	With Outcome 0 (Alive)/%	With Outcome 1 (Dead)/%
Race	Asian	977/60.3	858/87.8	119/12.2
	Other	643/39.7	559/86.9	84/13.1
Mechanism of injury (MOI)	Motor vehicle crash	486/30	413/85	73/15
	Fall	551/34	495/89.8	56/10.2
	Pedestrian	268/16.5	216/80.6	52/19.4
	Other	315/19.4	293/93	22/7
Arrival mode	Ambulance	1350/83.3	1167/86.4	183/13.5
	Other	270/16.7	250/92.6	20/7.4
Midline shift	No	1260/77.8	1155/91.7	105/8.3
	Yes	360/22.2	262/72.8	98/27.2
CT findings/mass lesion	Subdural hemorrhage	455/28.1	380/83.5	75/16.5
	Epidural hemorrhage	371/22.9	352/94.9	19/5.1
	Subarachnoid hemorrhage	152/9.4	114/75	38/25
	Hemorrhagic contusions	321/19.8	303/94.4	18/5.6
	Diffuse axonal injury	120/7.4	99/82.5	21/17.5
	Other	201/12.4	169/84.1	32/15.9
Cerebral edema	No	1517/93.6	1370/90.3	147/9.7
	Yes	103/6.4	47/45.6	56/54.4
Facial bones fracture	No	981/60.6	857/87.4	124/12.6
	Yes	639/39.4	560/87.6	79/12.4
Lung contusion	No	1273/78.6	1152/90.5	121/9.5
	Yes	347/21.4	265/76.4	82/23.6
Hemothorax	No	1482/91.5	1319/89	163/11
	Yes	138/8.5	98/71	40/29
Pneumothorax	No	1387/85.6	1251/90.2	136/9.8
	Yes	233/14.4	166/71.2	67/28.8

Variable	Category	Count/%	With Outcome 0 (Alive)/%	With Outcome 1 (Dead)/%
Abdominal Organ injuries	No	1417/87.5	1278/90.2	139/9.8
	Yes	203/12.5	139/68.5	64/31.5
GCS category	13-15	893/55.1	875/98	18/2
	9-12	122/7.5	113/92.6	9/4.4
	≤ 8	605/37.3	429/70.9	176/29.1
Shift	7am-6:59pm	858/53	758/88.3	100/11.7
	7pm-6:59pm	762/47	659/86.5	103/13.5
Known comorbidities	No	1328/82	1167/87.9	161/12.1
	Yes	292/18	250/85.6	42/14.4
Intubation	No	848/52.3	847/99.9	1/0.1
	Yes	772/47.7	570/73.8	202/26.2
Venous thromboembolism prophylaxis	No	656/40.5	537/81.9	119/18.1
	Yes	964/59.5	880/91.3	84/8.7
Blood transfusion	No	1013/62.5	989/97.6	24/2.4
	Yes	607/37.5	428/70.5	179/29.5
	Total/%	1620/100	1417/87.5	203/12.5

Performance of ML models

To calculate the models' performance metrics, we first constructed the confusion matrix that displays the relationship between the actual observations and the predicted conditions.

Table 18 shows the performance evaluation metrics for the two ML techniques in the test data partition. Both models achieved accuracy greater than 91%. Nevertheless, since accuracy alone is an insufficient measure to evaluate model performance, AUC, precision, NPV, sensitivity, specificity, and F-score measures were taken into consideration. SVM achieved the best performance.

Table 17. Performance of the Classification Models

Model	No. of predictors	Accuracy %	AUC	PPV	NPV	Sensitivity	Specificity	F-Score
SVM	21	95.6	0.96	0.88	0.97	0.73	0.99	0.8
ANN	21	91.6	0.93	0.66	0.96	0.62	0.96	0.64

In-Hospital Mortality Risk Factors

SVM utilized all the 21 variables in predicting the in-hospital mortality. Figure 9 shows the predictors' importance. SVM revealed that receiving endotracheal intubation during resuscitation plays the most important role in predicting in-hospital mortality.

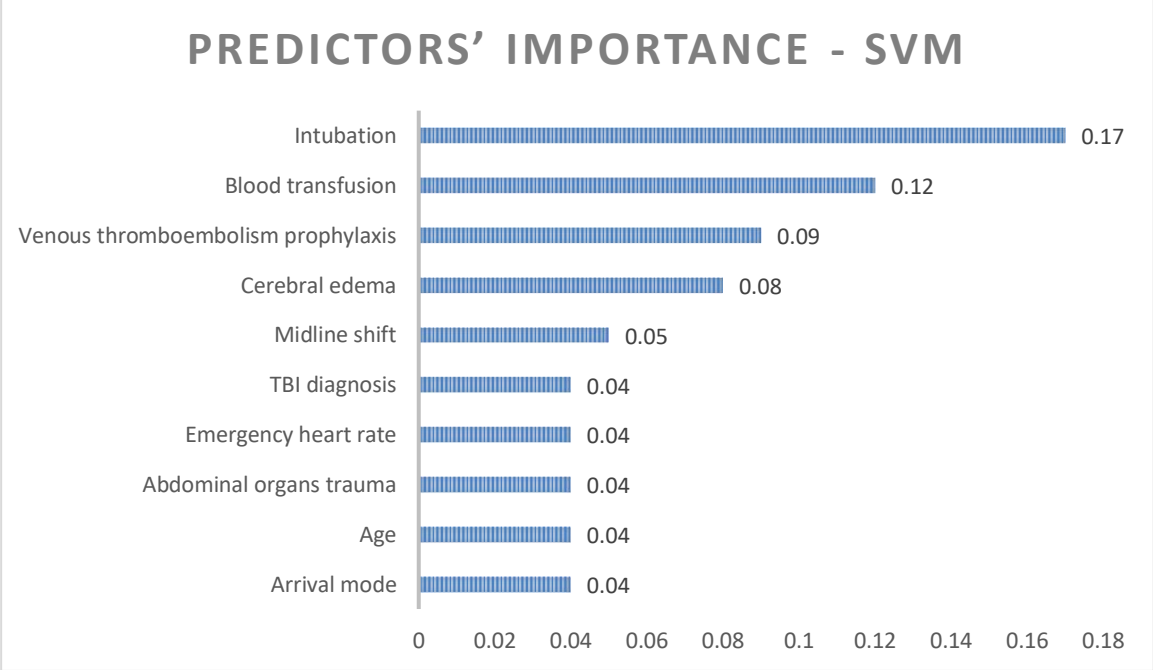

Figure 9. Predictors' importance in SVM.

Discussion

The early prediction of in-hospital mortality in TBI patients is of utmost importance. Early and powerful prediction of mortality helps clinicians and healthcare managers optimize the management of medical resources, initiate appropriate diagnostics and interventions in a timely manner, conduct comparative audits, and ensure that the patients' families and significant others receive appropriate guidance (166, 172). However, the prediction of disease prognosis and outcomes requires developing good prognostic models that include large samples and enjoy high external and internal validity in order to be generalizable beyond a specific research setting (191). Many prognostic models were published over the years, but few of them achieved sample validity requirements (192). Usually, clinicians use certain prognostic factors such as Glasgow Coma Scale (GCS) to guide their therapeutic decisions and to estimate prognostic outcomes (172, 192). Nonetheless, such predictors may be affected by several factors such as alcohol use, which negatively affects prediction success and the discriminatory power of the model (178, 184). Thus, for accurate outcome prediction, multiple risk factors (e.g. age, GCS, and others) need to be considered jointly in developing a prognostic model (192, 193).

In terms of models' performance, SVM outperformed the ANN in all the performance evaluation metrics (Table 18). Therefore, SVM is the chosen model for deployment.

On a wider scale, in this study, the SVM outperformed the conventional multivariate LR based models that utilize the conventional TBI prognostic models, as reported in Table 14. The highest reported AUC when using the

conventional prognostic models was 0.92 (174, 178, 181). Furthermore, when comparing this study's ML models' performance with the published literature on TBI, we found that the performance of the SVM model was higher or similar to the performance of the ML models in similar studies (166, 177). This comparison is crucial when considering the external validity of this study's findings.

This study ranked intubation as the most important predictor for post-TBI in-hospital mortality. Almost 26% of patients who were intubated in the first 24 hours post-injury died during their initial hospitalization, compared to 0.1% of those who were not intubated. This could be attributed to the severity of TBI, as the more severe the injury, the higher the likelihood of patient intubation. Moreover, intubation increases the length of stay in the hospital and increases the risks of in-hospital complications (e.g. VAP), which contribute significantly to increasing mortality (194). The need for blood transfusion during resuscitation has a significant relationship with in-hospital mortality. 29% of patients who received blood transfusion during resuscitation died compared to 2.4% mortality among those who did not receive blood. The need and the consequences of blood transfusion in TBI are still debatable. Several studies reported that blood transfusion in TBI is associated with unfavorable outcomes (195, 196). This could also be explained by the fact that patients who needed blood transfusion are those who had more severe injuries and had lost significant amounts of blood; such patients are inherently predisposed to poor TBI outcomes.

Consistent with the previous literature, this study found that patients who receive VTE prophylaxis have better survival rates compared to those who do

not (197). 18.1% of those who did not receive the VTE prophylaxis died during their initial hospitalization, compared to 8.7% of those who did received it. Further explanation about the relationship between the VTE and TBI outcomes is provided in the next chapter. Also, this study found that 54.4% of patients who developed CED following the primary TBI died in-hospital, compared to 9.7% of those who did not develop CED. This finding is consistent with Jha et al., who reported that cerebral edema is a leading cause of in-hospital mortality, as it occurs in more than 60% of patients with mass lesions, including post-TBI hemorrhage (198). Cerebral edema is a secondary complication of TBI in which brain tissue water increases following injury. This is why significant efforts in TBI management are devoted to the prevention of the secondary brain injury and maintaining adequate cerebral perfusion pressure (CPP) (199, 200).

Midline shift is a major post-traumatic complication that leads to serious unfavorable effects, including mortality (178, 182, 201). Around 27% of patients who had midline shift died compared to 8.3% of those who had no midline shift reported in their CT scan. TBI diagnosis as per brain CT scan result plays an integral role in predicting post-TBI in-hospital mortality. Interestingly, 25% of those who had SAH following the TBI died, compared to 17.5% and 16.5% for those with DAI and SDH respectively. It is documented in the literature that SAH has a significant effect on in-hospital mortality (177, 178, 182).

Presenting heart rate (HR) is an indicator of the patient's hemodynamic stability following any type of trauma, particularly TBI. High HR (> 100 bpm) (202), especially when associated with low SBP (< 90 mmHg) (200), may indicate hypovolemic shock state, which leads to poor CPP. This study found a positive relationship between the HR and in-hospital mortality. The HR in this

study was collected upon arrival to the ED following trauma. The mean HR upon admission was 93 bpm. The mean HR for those who survived was 90.8 while it was 108.5 bpm for those who later died in the hospital. Interestingly, the mortality rate increases significantly when patients with TBI have associated abdominal injuries (203). Mortality among those with associated abdominal injury is 31.5%, compared to 9.8% among those with no associated abdominal injury.

Finally, the tenth most-important variable was the arrival mode. Patients who arrived at the trauma center via ambulance had higher mortality compared to those who arrived to the trauma center via another mode (13.5% vs. 7.4%). This is consistent with previous literature, which found that mortality patterns are affected by the mode and the time of arrival to the emergency room following TBI or polytrauma (204, 205). This could be due to the assumption that the time between the injury and the arrival of the ambulance at the hospital is relatively longer than in cases of arrival via a private vehicle (205), or simply the assumption that the more severe the injury, the higher the likelihood that a patient gets transported to the hospital via ambulance.

Conclusion

This study demonstrated that the performance of the ML techniques is superior to the conventional multivariate models. Furthermore, the results were consistent with the known body of knowledge. Thus, with the availability of massive data sets in the electronic medical records and other structured registries, clinical evidence could be made available quickly and with less effort.

From another perspective, the results of this study may encourage decision makers in the trauma surgery to integrate the ML techniques with the

NTR and the electronic medical records. This may help clinicians plan their preventive efforts and mobilize the necessary resources in an early stage of patient treatment, which could improve the care outcomes.

CHAPTER 7: PREDICTION OF IN-HOSPITAL MORTALITY IN TBI PATIENTS ON MECHANICAL VENTILATION – ML APPROACH

Background

This study is a subgroup analysis which aims to design an ML algorithm to help predict in-hospital mortality in patients with TBI who received mechanical ventilation (MV). More than 70 million people worldwide sustain TBI every year (171). Compared to other injuries, TBI leads to the highest mortality and permanent disability rates (170, 172). Mortality in TBI is known to be highly associated with the severity of the TBI and patient age (173). Severe TBI is one of the common causes for the use of MV (206, 207). Although MV is a common intervention in intensive care units, and has saved countless lives since it was first used in 1950s (176), patients receiving MV are prone to several complications and have higher mortality rates compared to other patients (208). The early prediction of in-hospital mortality in patients with TBI is of utmost importance. Early and powerful prediction of mortality helps clinicians and healthcare managers optimize the management of medical resources, initiate appropriate diagnostics and interventions in a timely fashion, conduct comparative audits, and ensure that patients' families and significant others receive appropriate guidance (166, 172).

Many prognostic models were published over the years, but none of these models was designed exclusively to predict the mortality in TBI patients who receive MV. For example, Trauma Injury Severity Score (TRISS) aims to calculate the probability of survival and the outcomes in admitted trauma patients with or without TBI or MV (179). Similarly, the International Mission for Prognosis and Analysis of Clinical Trials in TBI (IMPACT), Marshal scale,

Helsinki CT score, Corticosteroid Randomization After Significant Head injury (CRASH), and Rotterdam CT score are all prognostic models that aim to predict mortality in adult patients with TBI, but which are not exclusive for patients with moderate to severe TBI who received MV (Table 14).

Methodology

This study aims to utilize the NTR data to design an ML predictive model to predict in-hospital mortality in adult patients who received MV following moderate to severe TBI who were admitted to L1TC.

Only adult patients (aged ≥ 14-years-old) who sustained moderate to severe TBI (HAIS ≥ 3) and who had received MV following TBI were included in the study. Patients who sustained other systematic injuries with AIS > HAIS were excluded, to ensure that the primary injury of participants was TBI. Records with missing data were eliminated. Subsequently, 785 eligible patients were included in the study. Detailed information about the methodology is explained available in chapter 5.

Outcome Measure

The outcome measure was in-hospital mortality during the initial hospitalization following moderate to severe TBI. It is a dichotomous variable (0 = alive and 1 = dead). Patients who were discharged from the Trauma Surgery Section or who were transferred to another hospital were considered alive.

Prediction Models

The objective of this study is to develop a supervised ML predictive model to predict the in-hospital mortality in patients who received MV following moderate to severe TBI. Two supervised ML techniques were utilized to

compare their performance with each other and with previous studies in order to recommend the model that achieves the optimal performance and highest practicality in supporting clinical decisions. LR and ANN are widely used in predicting in-hospital mortality. Therefore, they were selected to provide base line comparative performance. SPSS modeler 18.1 was used to conduct the analysis. Data were partitioned into three sets: the training set (60%), validation set (20%), and testing set (20%). Furthermore, the overfit prevention was set at 30%. Table 19 explains the data partitions.

Table 18. Data Partitions

Set	Proportion	No. of cases	No. of alive patients	No. of dead patients
Training set	60%	462	336 (72.7%)	126 (27.3%)
Validation set	20%	161	123 (76.4%)	38 (23.6%)
Testing set	20%	162	122 (75.3%)	40 (24.7%)
Total	100%	785	581 (74%)	204 (26%)

Logistic Regression (LR)

LR is a typical technique for predicting binary, binomial, or multinomial outcomes (88). It usually describes the relationship between a dichotomous dependent variable and a set of predictor variables that can be either numerical or categorical/ dummy variables. Typically, LR is used for the prediction of the probability of the occurrence of an event by fitting data to a sigmoidal (S-shape) logistic curve. Usually, LR uses a numerical cutoff value (0.5), whereby cases > 0.5 are classified as 1 = success, and the rest are categorized as 0 = failure (189). Thus, logistic regression is an appropriate procedure for predicting

mortality in TBI patients who received MV. Bi-directional step-wise LR was used in this study to control the effect of confounding variables and to measure the independent risk factors for post TBI mortality (166).

ANN

In this study, the ANN architecture was a standard feed-forward, back-propagation multi-layer perception (MLP) ANN. This MLP ANN consists of three layers; one input layer that had the study predictors, one hidden layer that consisted of six inaccessible neurons, and one output layer. We opted to design ANN using the MLP as it outperformed Radial Basis Function (RBF) during the initial assessment, with accuracy/AUC of 80.9%/0.875 vs. 77.9%/0.795.

One of the most important caveats in ANN is that it is prone to overfitting compared to LR, because the training makes the model perfectly fit the data set. Thus, with new data sets, prediction might be poor (189).

Ayer et al. compared between the two methods in several aspects and concluded that using LR requires more statistical knowledge than ANN, but ANN is more powerful in capturing complex relationships and determining interesting patterns in data. LR is easier to interpret and to identify important predictors compared to the black-box ANN (209). The discrimination power and the prediction performance for both methods are good in general, which makes it difficult to determine the superiority of one method over the other. Although the majority of studies that compared the performance of the two methods reported that one of them outperformed the other, the performance in general was similar (175, 189, 209, 210).

Results

This study of 785 participants with a mean age of 33 years included 204

patients (26%) who died in the hospital during their initial hospitalization., with a mean age at death of 36.9 years. The most common mechanism of injury was motor vehicle crash (37.5%) followed by fall from height (25.4%). The most common CT finding was subdural hemorrhage (29%), followed by extradural hemorrhage (21%), and 33.6% of patients sustained a midline shift. Tables 20 and 21 show the sample characteristics and the descriptive statistics for the study sample.

Table 19. Sample Characteristics – Continuous Variables

Variable	N	Mean	SD	Mean at death
Age	785	33	13.4	36.9
Injury severity score (ISS)	785	28.2	10.4	33.8
ED systolic blood pressure (SBP)	785	126.34	27.7	119
ED heart rate (HR)	785	102.8	25	107.7

Table 20. Sample Characteristics – Nominal and Ordinal Variables

Variable	Category	Count/%	With Outcome 0 (Alive)/%	With Outcome 1 (Dead)/%
Race	Asian	456/58.1	337/73.9	119/26.1
	Other	329/41.9	244/74.2	85/25.8
Mechanism of injury (MOI)	Motor vehicle crash	294/37.5	222/75.5	72/24.5
	Fall	199/25.4	142/71.4	57/28.6
	Pedestrian	162/20.5	110/67.9	52/32.1
	Other	130/16.6	107/82.3	23/17.7
Mode of arrival	Ambulance	639/81.4	455/71.2	184/28.8
	Other	146/18.6	126/86.3	20/13.7
Multiple rib fractures	No	600/76.4	454/75.7	146/24.3
	Yes	185/23.6	127/68.6	58/31.4
Lung contusion	No	509/64.8	387/76	122/24
	Yes	276/35.2	194/70.3	82/29.7

Variable	Category	Count/%	With Outcome 0 (Alive)/%	With Outcome 1 (Dead)/%
Hemothorax	No	678/86.4	514/75.8	164/24.2
	Yes	107/13.6	67/62.6	40/37.4
Pneumothorax	No	594/75.7	456/76.8	138/23.2
	Yes	191/24.3	125/65.4	66/34.6
Midline shift	No	521/66.4	416/79.8	105/20.2
	Yes	264/33.6	165/62.5	99/37.5
TBI diagnosis/ CT findings	Subdural hemorrhage	207/26.4	142/68.6	65/31.4
	Extradural hemorrhage	155/19.8	138/89	17/11
	Subarachnoid hemorrhage	67/8.5	46/68.7	21/31.3
	Hemorrhagic contusions	98/12.5	89/90.8	9/9.2
	Diffuse axonal injury	104/13.2	85/81.7	19/18.3
	Cerebral edema	84/10.7	29/24.5	55/65.5
	Other	70/8.9	52/74.3	18/25.7
Head AIS (HAIS)	3	241/30.7	218/90.5	23/9.5
	4	187/23.8	140/74.9	47/25.1
	5	357/45.5	223/62.5	134/37.5
Face AIS (FAIS)	0	399/50.8	276/69.2	123/30.8
	1	85/10.8	70/82.4	15/17.6
	2 (AIS 2-5)	301/38.3	235/78.1	66/21.9
Chest AIS (CAIS)	0	353/45	282/79.9	71/20.1
	1 (AIS 1-2)	120/15.3	82/68.3	38/31.7
	2 (AIS 3-5)	312/39.7	217/69.6	95/30.4
Abdomen AIS (AAIS)	0	610/77.7	473/77.5	137/22.5
	1 (AIS 1-2)	104/13.2	67/64.4	37/35.6
	2 (AIS 3-5)	71/9	41/57.7	30/42.3
Spine AIS (SAIS)	0	538/68.5	402/74.7	136/25.3
	1 (AIS 1-5)	247/31.5	179/72.5	68/27.5
Extremities AIS (EAIS)	0	416/53	316/76	100/24
	1 (AIS 1-2)	262/33.4	194/74	68/26
	2 (AIS 3-5)	107/13.6	71/66.4	36/33.6
Known comorbidities	No	659/83.9	496/75.3	163/24.7
	Yes	126/16.1	85/67.5	41/32.5
Intubation location	In-hospital	267/34	210/78.7	57/21.3
	Pre-hospital	518/66	371/71.6	147/28.4
Venous thromboembolism prophylaxis	No	180/22.9	60/33.3	120/66.7
	Yes	605/77.1	521/86.1	84/13.9
Blood transfusion	No	252/32.1	228/90.5	24/9.5
	Yes	533/67.9	353/66.2	180/33.8
Total/%		785/100	581/74.1	204/25.9

Performance of ML models

Table 22 shows the performance evaluation metrics for the two ML techniques in the test data partition. To obtain a comprehensive overview of the models' performance, the analysis considered their accuracy, AUC, precision, negative predictive value, sensitivity, specificity, and F-score measures. LR achieved better performance than ANN, with AUC of 0.905 and accuracy of 87%.

Table 21. Performance of the Classification Models

Model	No. of predictors	Accuracy	AUC	Precision	NPV	Sensitivity	Specificity	F-Score
LR	6	87%	0.90	0.81	0.89	0.63	0.95	0.7
ANN	24	80.9%	0.87	0.6	0.89	0.68	0.85	0.64

In-Hospital Mortality Risk Factors

LR identified six predictors (administration of VTE prophylaxis, HAIS, TBI diagnosis/CT finding, the need for blood transfusion during resuscitation, ED HR, and age) as independent risk factors for the in-hospital mortality of the intubated patients with moderate to severe TBI (Table 23). The administration of VTE prophylaxis was ranked first in the predictors' importance (0.37), followed by severity of head injury as measured by AIS (HAIS) (0.21). Figure

10 shows the predictors' importance ranking.

On the other hand, ANN used all the 24 predictors to predict the in-hospital mortality. ANN achieved 80.9% accuracy and 0.875 AUC, with ISS ranked first in the predictors' importance (0.12). Figure 11 ranks the top ten predictors based on their importance.

Table 22. Significant Predictors Estimation and Likelihood Ratio Assessment

Predictor	B coefficient	P value	EXP (B)	95% confidence interval
Venous thromboembolism (VTE) prophylaxis	3.435	< 0.05	31.035	14.198-67.837
No				
Yes: reference				
HAIS				
3	-2.153	< 0.05	0.116	0.045-0.299
4	-0.311	0.403		
5: reference				
TBI diagnosis/CT finding				
(EDH)	0.654	0.263	7.733	2.088-28.635
SDH	-1.22	0.066		
EDH	0.609	0.435		
SAH	-0.788	0.33		
Contusions	-0.313	0.644		
DAI	2.045	< 0.05		
Cerebral edema				
Others: reference				
Blood transfusion				
No	-1.655	< 0.05	0.191	0.086-0.422
Yes: reference				
ED HR	0.028	< 0.05	1.028	1.015-1.042
Age	0.035	< 0.05	1.036	1.013-1.059

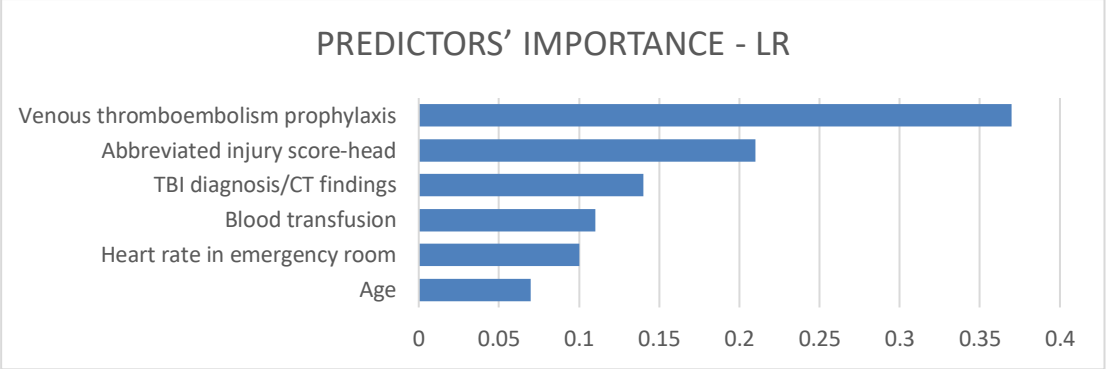

Figure 10. Predictors' importance in logistic regression.

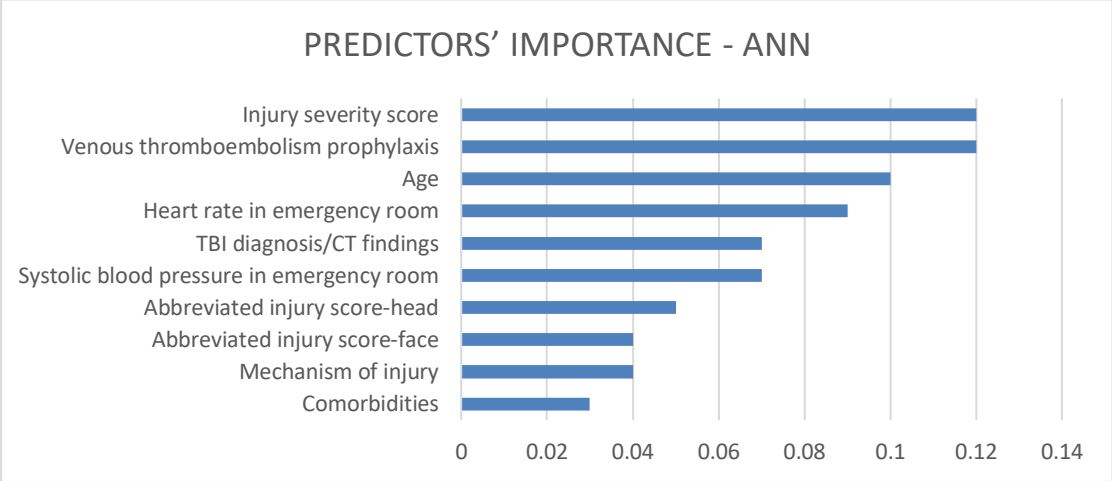

Figure 11. Predictors' importance in ANN.

Discussion

Consistent with the previous literature, this study found that patients who receive venous thromboembolism (VTE) prophylactic agents have better survival rates compared to those who do not (197). 66.7% of the patients who did not receive VTE prophylaxis died, compared to 13.9% of those who did receive it. The odds of mortality given that a patient does not receive VTE prophylaxis following moderate to severe TBI increases 31-fold compared to

those who receive it. The previous literature reported a significant association between TBI and VTE (e.g. deep vein thrombosis and pulmonary embolism) (211, 212). Although there is a debate concerning the potential to increase hemorrhagic risk among patients with TBI by administering VTE prophylaxis, the literature suggests that administering VTE prophylaxis 24 to 48 hours following TBI is safe, following a proper risk assessment (197, 213). Thus, we argue that the lack of or delayed administration of VTE prophylaxis may lead to VTE events that may contribute to increased mortality. Therefore, considering the early administration of VTE prophylaxis could improve TBI patients' prognosis and reduce the risk of mortality that is associated with coagulation related complications.

Additionally, we observed that patients who sustained more severe TBI (as measured by AIS) were more likely not to receive VTE prophylaxis. 17.4% of those with HAIS 3, 17.6% with HAIS 4, and 29.4% with HAIS 5 did not receive VTE prophylaxis. This is consistent with Nathens et al. (214) who concluded that doctors are conservative in administering VTE prophylaxis for patients with severe TBI. They found that patients with severe TBI, no injuries in their extremities, and those who received blood transfusions were more prone to delayed VTE prophylaxis administration. Accordingly, we conducted secondary analysis to test the association between the severity of TBI (HAIS), blood transfusion during resuscitation, and severity of extremities injury with the administration of VTE prophylaxis in order to understand the doctors' VTE prophylaxis prescription behavior. Chi square analysis revealed significant association between the HAIS and VTE ($Chi^2 = 15.57$ and P value < 0.05), which is consistent with past research (214). On the other hand, we found no

significant association between blood transfusion or extremities' AIS with VTE prophylaxis administration ($Chi^2 < 5$ and P value > 0.05). We further evaluated if there was any association between the TBI diagnosis and the administration of the VTE prophylaxis. We found a significant association ($Chi^2 = 13.498$ and P value < 0.05).

Patients with cerebral edema are more likely not to receive VTE prophylaxis (34.5%) compared to patients with subdural hemorrhage (24.6%), extradural hemorrhage (22.6%), subarachnoid hemorrhage (20.9%), brain contusions (21.4%), and diffuse axonal injury (20%). It is important to note that AIS is not necessarily available information for the treating doctor at the time of making the decision to administer VTE prophylaxis. Therefore, we argue that the perceived severity of TBI as per patient presentation and clinical examination influence physicians' VTE prescribing behavior. Importantly, this finding may support the argument that the relation between the VTE prophylaxis and the mortality does not reflect causality, but it is more associated with the severity of injury that influences the doctors' VTE prophylaxis prescription behaviors.

It is widely accepted that the more severe the head injury, the higher the probability of mortality and unfavorable outcomes (174, 179, 215). This study proved that patients with higher HAIS (AIS 5) have higher likelihood of mortality compared to those with lower HAIS (AIS 3). Only 9.5% of the patients who had HAIS 3 died, compared to 25.2% and 37.5% of patients who had HAIS 4 and 5 (respectively). The odds of mortality given that a patient have severe TBI (HAIS = 5) increase significantly by 88.4% compared to those with HAIS 3. We found no significant difference in the mortality between patients who had HAIS 4 and

HAIS 4.

The TBI diagnosis as identified by the initial brain CT scan plays a role in determining patients at risk of in-hospital mortality. Cerebral edema was found to play a significant role in predicting mortality in ventilated patients following TBI. More than 65% of the patients who had cerebral edema died compared to patients with subdural hemorrhage (31.4%), subarachnoid hemorrhage (31.3%), diffuse axonal injury (18.3%), extradural hemorrhage (11%), and brain contusions (9.2%). The odds of mortality given that a patient sustained cerebral edema are seven times higher than for patients with other TBI findings.

Interestingly, it is found that patients who underwent blood transfusion have higher odds of in-hospital mortality compared to those who have not received blood during resuscitation. Of those who received blood transfusion during resuscitation, 33.8% have died compared to 9.5% of those who did not need blood transfusion during resuscitation. The odds of mortality given that a patient receives blood during resuscitation increase significantly by 80.9% compared to those who do not require blood transfusion during resuscitation.

The need and the consequences of blood transfusion in TBI are debatable. Several studies have reported that blood transfusion in TBI is associated with unfavorable outcomes (195, 196). This could be explained by the fact the patients who needed blood transfusion are those who had more severe injuries and had lost significant amounts of blood which makes the patients prone to poor TBI outcomes. We argue that blood transfusion *per say* does not have a direct causal relationship with the mortality, but the reasons that indicate the need for blood transfusion during resuscitation (i.e. bleeding or

hypovolemia) should be considered predictors of mortality.

The patient heart rate upon arrival to the ED is an indicator of the organ perfusion adequacy. The mean heart rate upon admission to the emergency room following the TBI was 102.8 while the mean heart rate upon arrival for those who died was 107.7 beats per minute. An increase in HR by one unit may change the odds of mortality by approximately 3% (odds ratio = 1.028, $P< 0.05$).

Age is also found to play a significant role in predicting in-hospital mortality in patients with TBI who receive MV (166). An increase by one year of age increases the likelihood of mortality by more than 3.6%. (Odds ratio = 1.036, $P <0.05$). The patients' mean age in this study was 33 years. However, the mean age of those who died during their initial hospitalization was 36.9 years.

Limitations

785 patients in five years is considered a small sample in the field of ML. The size of this sample posed challenges in several respects, including class imbalance, management of missing data, and cross validation. Qatar has a relatively small population, therefore regional or international multicenter studies could help overcome this limitation. Furthermore, some of the potentially important predictors such as time to surgical procedures and other unfavorable outcomes were not captured in the data set. The availability of such variables may enhance the predictive performance and improve the clinical insight that can be obtained by the findings of the study.

Conclusion

Although plenty of literature focuses on predicting mortality in TBI patients, there is a dearth of research exploring the deployment of ML

techniques to predict in-hospital mortality in intubated patients with TBI. Accordingly, this study provides a valuable contribution to this pioneering field of research.

This study demonstrates that LR provides better performance than ANN in predicting the in-hospital mortality for patients who received MV following moderate to severe TBI.

The results are encouraging and provide an opportunity to integrate the ML techniques with the NTR and electronic medical records to provide instant clinical decision support to healthcare providers. In addition, with limited data size, ML algorithms demonstrate powerful predictive power, which opens the door for integrating the AI modalities with medical practice to enhance patients' treatment outcomes.

CHAPTER 8: USING NTR DATA TO PREDICT PROLONGED MECHANICAL VENTILATION IN PATIENTS WITH TBI – ML APPROACH

Background

Patients with severe TBI are prone to impaired arousal, which warrants protecting their airway by MV (216). Therefore, they are at higher risk of prolonged mechanical ventilation (PMV) than any other critical patients (207). In 2007, the *European Respiratory Journal* published guidelines on weaning from mechanical ventilation, to describe the entire process of liberating patients from the ventilator (217). Nonetheless, due to the lack of robust evidence in the literature, there were no clear recommendations about the weaning process in neurocritical care settings, and the decision to extubate patients remains a complex issue (207).

Although MV is a lifesaving intervention, it has several complications, such as ventilator-induced lung injury, VAP, prolonged hospitalization, and increased mortality (208, 218). These risks increase with PMV (218, 219). Approximately 30% of critically ill patients will require PMV at some stage (206, 218, 220), and more than 600,000 patients were expected to require PMV in 2020 (221). Several strategies, such as minimizing sedation and performing daily spontaneous breathing trials, have been adopted to mitigate the risks associated with the MV and to prevent PMV (222, 223).

Hence, predicting patients at risk for PMV is of utmost importance to help clinicians design individualized plans of care that mitigate the risk of PMV. This includes the decision of early use of tracheostomy, which has proven beneficial when MV is still required (220, 224-226). There are several studies that aimed to determine the significant predictors of PMV. However, it remains difficult to

determine a set of key predictors due to the differences in patients' clinical features and clinical settings. Furthermore, there is no consensus on the definition of PMV. The PMV period in the published literature ranges from 5 hours to 1 year, with > 21 days being the most common definition (227). Table 24 shows examples of the previously published literature in predicting PMV highlighting the patients' characteristics, PMV duration, predictors used, and predictive models' performance measures.

In a recent Cochrane systematic review, the early tracheostomy (<10 days from the start of MV) was found to be associated with significant improvement of patient treatment outcomes (120). This finding supports the previous randomized clinical trial by Young et al., which found that early tracheostomy replacement (< 10 days) is beneficial to patients and is associated with improved outcomes (228). Besides favorable clinical outcomes, early tracheostomy is associated with improved economic outcomes, such as reduced ICU costs (229) and hospital length of stay (120). Furthermore, early tracheostomy was found to significantly improve patient QoL compared to the endotracheal ventilation when prolonged ventilation is required (230). Therefore, defining the PMV to be longer than 10 days could be of a great value in term of early liberation from MV, early tracheostomy replacement, improving QoL, and cost-effectiveness.

Most previously published studies that aimed to predict PMV used conventional multivariate techniques, particularly LR, and yielded low to moderate accuracies (0.53- 0.75) and AUC (0.65-0.75) (220, 226). The implementation of the ML to predict the PMV has achieved a relatively higher performance than the conventional predictive models, with accuracy of 83.2%

and AUC of 0.82 (220). Accordingly, we decided in this study to evaluate the predictive performance of selected ML models when PMV is defined as > 10 days. At the same time, we will conduct another two sets of predictive models in which PMV is defined as > 7days and > 14 days in order to compare the predictive performance of the ML models in the three sets.

Methodology

This study utilized the NTR data to design supervised ML algorithms to predict the PMV (> 7 days, > 10 days, and > 14 days) for patients who received MV following moderate to severe TBI. We hypothesized that the ML algorithms would outperform the conventional multivariate predictive techniques in terms of accuracy, sensitivity, specificity, and precision, negative predictive value (NPV), F-score, and AUC. Also, consistent with the Cochrane's systematic review results, we hypothesized that defining PMV to be greater than 10 days would optimize the prediction performance of the ML models used in this study. We constructed three sets of predictive models based on the definition of the PMV. Set A defines PMV as > 7 days, set B defines PMV as > 10 days, and set C defines PMV as >14 days.

Data Preparation

To achieve the objectives of this particular study, the following inclusion and exclusion criteria were applied in the selection of patient records.

Patient Inclusion Criteria

- Adult patients older than 14 years with TBI.
- Patients whose Abbreviated Injury Score for head region (HAIS) ≥ 3

- Patients who underwent intubation following the injury either at the scene by the ambulance crew or in the hospital within the first 24 hours from the arrival to the hospital.

Patient Exclusion Criteria

- Patients with any regional injuries with AIS greater than the HAIS, to ensure that the TBI assumes the highest effects on the dependent variable.
- All patients who died or were discharged within 7 days for set A, within 10 days for set B, and within 14 days for set C.

Importantly, due to the criticality of the subject, we decided that all records with missing values would be deleted. Therefore, no imputation was needed. All variables with no predictive power (e.g. health record number, date of admission, and date of disposition) or those that were severely imbalanced (e.g. gender, where female patients were approximately 4% in the three data sets) were excluded. Subsequently, 674 records in set A, 643 records in set B, and 622 records in set C were eligible for the study. Figure 12 explains the records' inclusion and exclusion procedure. Further details about the methodology are provided in chapter 5.

Outcome Measure

The dichotomous outcome measure for this study is prolonged mechanical ventilation (PMV). PMV is defined as the stay on mechanical ventilation support for > 7 days in set A, > 10 days in set B, and > 14 days in set C, from the initial intubation that was performed within the first 24 hours from the injury. PMV0 means that the patient was extubated before the sets' period,

and PMV1 means that the patient stayed on MV longer than the sets' period.

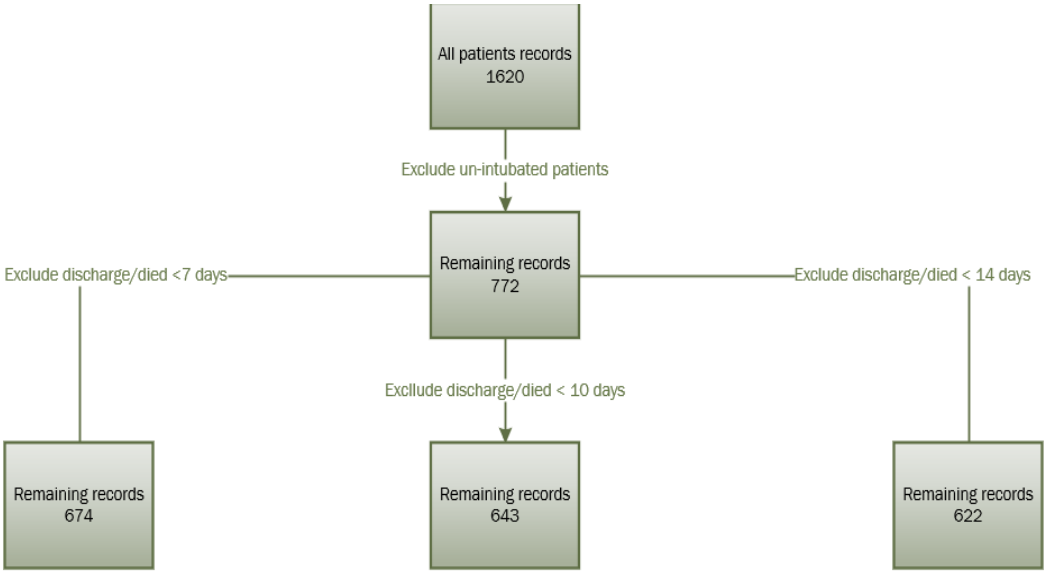

Figure 12. Records' inclusion and exclusion process.

Table 23. Examples of Previous PMV Prediction Studies

Patient group	PVM duration	Predictors	Predictive technique	Model's performance (AUC)
Parreco et al. (2018) (220); PMV > 7 days; predictive: Gradient-boosted DTA				
All ventilated level 3 ICU patients (2001-2012)		Oxford Acute Severity of Illness Score (OAISIS), Sequential Organ Failure Assessment, Simplified Acute Physiology Score (SAPS), Simplified Acute Physiology Score II (SAPS II), Acute Physiology Score III, Logistic Organ Dysfunction Score (LODS), Sepsis Related Organ Failure Assessment (SOFA)		Mean AUC 0.820 ± 0.016
Chang et al. (2018) (231); PMV > 21 days; predictive: LR				
ICU patients who survived sepsis/ septic shock and respiratory failure		Demographics Acute Physiology, Age, Chronic Health Evaluation (APACHE II) Comorbidities Lab findings (hematology, liver function, coagulation, urea electrolytes, arterial blood gases),Ventilator settings		AUC 0.725
Agle et al. (2006) (224); PMV > 14 days; predictive: LR				
Torso trauma patients who met specific criteria for shock resuscitation and required 48 hours of MV		Demographics, facial trauma, chest trauma severity (AIS), ventilator settings		AUC 0.79
Clark and Lettieri (2013) (218); PMV > 14 days; predictive: LR				
Adult patients requiring MV support in a medical ICU		Demographics, vital signs, laboratory values (hematology, renal and liver function tests, HCO3), APACHE 2		AUC 0.75
Dimopoulou et al. (2003) (225); PMV > 7 days; predictive: LR				
Adult patients with thoracic trauma requiring MV support in ICU		Demographics, injury characteristics, injury severity score, AIS of other associated injuries (head, neck, face, pelvis and extremities) and ventilator settings		Not declared
Figueroa-Casas et al. (2015) (226); PMV > 7 days; predictive: LR				
ICU patients receiving MV		Demographics, SOFA score on intubation, comorbidities, location before ICU admission, diagnosis category		AUC 0.65-0.70

ML Prediction Models

A group of supervised ML techniques were utilized to compare their performance with each other and with previous studies in order to recommend the model that achieves the optimal performance and highest practicality in supporting the clinical decision. LR, RF, ANN, C.5DT and SVM were selected to provide base line comparative performance.

To prevent overfitting and to validate the models' performance, we partitioned the data into training set (70%) and testing set (30%), and overfit prevention was set at 30%. The data partitioning was executed automatically by the analytical software based on the partition command we provided. Table 25 explains the data partitions.

Table 24. Data Partitions

Set A PMV >7 days	Set	Proportion	No. of cases	Ventilator days ≤ 7	Ventilator days > 7
	Training set	70%	472	183	289
	Testing set	30%	202	87	115
	Total	100%	674	270	404
Set B PMV >10 days	Set	Proportion	No. of cases	Ventilator days ≤ 10	Ventilator days > 10
	Training set	70%	446	239	207
	Testing set	30%	197	114	83
	Total	100%	643	353	290
Set C PMV >14 days	Set	Proportion	No. of cases	Ventilator days ≤ 14	Ventilator days > 14
	Training set	70%	432	312	120
	Testing set	30%	190	138	52
	Total	100%	622	450	172

Logistic Regression (LR)

Bidirectional step-wise LR was used in this study to control the effect of confounding variables and to measure the independent risk factors for post-TBI PMV (166).

Random Forest (RF)

RF is a powerful supervised ML technique that is used widely for classification problems (88, 232). RF is proven to have improved accuracy in comparison to other ML techniques. The reason is that RF uses bootstrapping to grow a forest of uncorrelated trees, with a high degree of randomness in feature selection, which contributes to reducing errors significantly (90)

SVM

We used Linear Kernel function in this study as it provided better predictive performance in the preliminary assessment.

ANN

We used MLP ANN architecture as it optimized the predictive performance compared to the other architectures.

C.5 Decision Tree (C.5DT)

C.5DT is the successor of the C.4.5 DT classification data mining algorithm. A DT is a “classification algorithm in which each non-leaf node indicates a test on an attribute of the input cases; each branch corresponds to an outcome of the test; and each leaf node indicates a class prediction” (88). Generally, DT classification algorithms are powerful, logical, and easy to interpret and understand (89).

Results

In set A (PMV> 7days), 674 eligible cases were included, of which 404 (59.9%) had more than seven ventilator days. The overall mean age was 32.3 years, and the mean age for patients with PMV was 33.6. Fifty-three percent of the patients sustained chest trauma, of whom 61.7%, 24%, and 46.5% sustained lung contusion, hemothorax, and pneumothorax, respectively. The most common TBI diagnosis was subdural hemorrhage (SDH) (25%), and more than 30% of them developed midline shift on the computed tomography head images.

In set B (PMV >10 days), 643 eligible cases were included, of which 290 (45%) had more than 10 ventilator days. The overall mean age was 32.1 years, and the mean age of patients with PMV was 33.6 years. Almost half of the patients sustained chest trauma, including 63%, 22%, and 43% with lung contusion, hemothorax, and pneumothorax, respectively. SDH was the most common TBI diagnosis (25%), and 29% of patients had midline shift. Tables 26 and 27 show the sample characteristics in set B.

In set C (PMV >14 days), 622 eligible cases were included, of which 172 (28.5%) had more than 14 ventilator days. The overall mean age was 32 years, and the mean age for patients with PMV was 33.9 years. There were 329 patients (52.9%) who sustained chest trauma, of whom 64% suffered lung contusion, 22.8% had hemothorax, and 42% had pneumothorax. One quarter of the patients sustained SDH, and 28.8% of them had midline shift.

Table 25. Sample Characteristics (set B) – Continuous Variables

Variable	N	Mean	SD	Mean when PM> 10
Age	643	32.1	12.9	33.6
ISS	643	26.8	9.5	29.5
ED SBP	643	128.2	26.1	126.6
ED HR	643	101.8	24.6	103.4

Table 26. Sample Characteristics (Set B) - Nominal and Ordinal Variables

Variable	Category	Count/%	PMV = 0	PMV = 1
Race	Asian	365/56.8	218/59.7	147/40.3
	Other	278/43.2	135/48.6	143/51.4
Mechanism of injury (MOI)	Motor vehicle crash	247/38.4	128/51.8	119/48.2
	Fall	159/24.7	100/62.9	59/37.1
	Pedestrian	125/19.4	59/47.2	66/52.8
	Other	112/17.4	66/58.9	46/41.1
Multiple rib fractures	No	497/77.3	294/59.2	203/40.8
	Yes	146/22.7	59/40.4	87/59.6
Lung contusion	No	427/66.4	246/57.6	181/42.4
	Yes	216/33.6	107/49.5	109/50.5
Hemothorax	No	567/88.2	325/57.3	242/42.7
	Yes	76/11.8	28/36.8	48/63.2
Pneumothorax	No	495/77	297/60	198/40
	Yes	148/23	56/37.8	92/62.2
Midline shift	No	456/70.9	264/57.9	192/42.1
	Yes	187/29.9	89/47.6	98/52.4
TBI diagnosis/ CT findings	Subdural hemorrhage	162/25.2	89/54.9	73/45.1
	Epidural hemorrhage	141/21.9	100/70.9	41/29.1
	Subarachnoid hemorrhage	53/8.2	24/45.3	29/54.7
	Hemorrhagic contusions	93/14.5	63/67.7	30/32.3
	Diffuse axonal injury	93/14.5	28/30.1	65/69.9
	Cerebral edema	45/7	15/33.3	30/66.7
	Other	56/8.7	34/60.7	22/39.3
Head AIS (HAIS)	3	225/35	153/68	72/32
	4	157/24.4	86/54.8	71/45.2
	5	281/40.6	114/34.2	147/65.8
Face AIS (FAIS)	0	313/48.7	169/54	144/46
	1	74/11.5	36/48.6	38/51.4
	2 (AIS 3-5)	256/39.8	148/57.8	108/42.2

Variable	Category	Count/%	PMV = 0	PMV = 1
Chest AIS (CAIS)	0	302/47	197/65.2	105/34.8
	1 (AIS 1-2)	100/15.5	42/42	58/58
	2 (AIS 3-5)	241/37.5	114/47.3	127/52.7
Abdomen AIS (AAIS)	0	515/80.1	301/58.4	214/41.6
	1 (AIS 1-5)	128/19.9	52/40.6	76/59.4
Spine AIS (SAIS)	0	443/68.9	262/59.1	181/40.9
	1 (AIS 1-5)	200/31.1	91/45.5	109/54.5
Glasgow Coma Score (GCS) category	13-15	85/13.2	56/65.9	29/34.1
	9-12	72/11.2	53/73.6	19/26.4
	≤ 8	486/75.6	244/50.2	242/49.8
Known comorbidities	No	537/83.5	311/57.9	226/42.1
	Yes	106/16.5	42/39.6	64/60.4
Intubation location	In-hospital	231/35.9	142/61.5	89/38.5
	Pre-hospital	412/64.1	211/51.2	201/48.8
Blood transfusion	No	235/36.6	189/80.4	46/19.6
	Yes	408/63.5	164/40.2	244/59.8
Ventilator associated pneumonia (VAP)	No	478/74.3	311/65	167/35
	Yes	165/25.7	42/25.5	123/74.5
Sepsis	No	583/90.7	343/58.8	240/41.2
	Yes	60/9.3	10/16.7	50/83.3
Total		643/100	353/54.9	290/45.1

Performance of the ML models

Table 28 shows the performance evaluation metrics for the five ML techniques in the test data partition. All models achieved moderate accuracy (66%-79%). Nevertheless, since accuracy alone is insufficient measure to evaluate the overall model's performance, AUC, precision, NPV, sensitivity, specificity, and F-score measures were taken into consideration. In set A, LR and SVM achieved relative similar performance in all performance metrics. Nonetheless, LR is the preferred model for deployment, as it demonstrated higher discrimination power, which is of great importance to the classification function (AUC 0.83 vs. 0.8), as well as its parsimony. LR achieves similar performance with fewer dimensions. In Set B, SVM achieved the highest

performance. In set C, ANN and SVM achieved similar performance, but ANN was the preferred model, as it gave higher accuracy, specificity, and positive predictive power (precision).

Comparing the discrimination power between the three sets, set B (which defines PMV to be greater than 10 days) performs better than Sets A and C, with AUC ranging from 0.77 to 0.84, while set A (PMV > 7days) AUC ranges from 0.65 to 0.83, and set C (PMV > 14 days) AUC ranges from 0.65 to 0.75. This implies that the discrimination power and the accuracy were more optimized when PMV was defined to be greater than 10 days.

Table 27. Performance of the Prediction Models

	No. of predictors	Accuracy%	Area under curve	Precision	NPV	Sensitivity	Specificity	F-score
Set A (PMV > 7days)								
LR	7	75	0.83	0.77	0.72	0.80	0.68	0.78
SVM	23	76	0.80	0.77	0.74	0.83	0.67	0.79
RF	23	73	0.77	0.77	0.69	0.76	0.70	0.76
ANN	23	69	0.78	0.72	0.66	0.77	0.60	0.74
C.5DT	19	66	0.65	0.70	0.60	0.70	0.61	0.70
Set B (PMV > 10 days)								
SVM	23	79	0.84	0.75	0.82	0.76	0.82	0.75
ANN	23	77	0.84	0.71	0.81	0.76	0.77	0.73
LR	7	75	0.82	0.70	0.78	0.69	0.79	0.70
RF	23	75	0.80	0.67	0.84	0.81	0.71	0.73
C.5DT	17	71	0.77	0.66	0.75	0.65	0.75	0.65
Set C (PMV > 14 days)								
ANN	23	76	0.72	0.64	0.77	0.27	0.94	0.38
SVM	23	74	0.74	0.54	0.77	0.29	0.91	0.38
LR	6	73	0.75	0.52	0.77	0.29	0.90	0.37
RF	23	71	0.73	0.47	0.80	0.46	0.80	0.47
C.5DT	10	71	0.65	0.43	0.76	0.25	0.88	0.32

Prolonged Mechanical Ventilation Predictors

In set A (PMV > 7 days), the LR model used seven predictors to classify the patients into two classes based on their mechanical ventilation dependency period. Figure 13 shows the predictors' importance in LR. Receiving blood during resuscitation scored the highest predictors' importance value (0.24).

In set B (PMV > 10 days), the SVM used all 27 predictors to classify the patients into the two outcome classes. Development of VAP scored the highest predictors' importance value of 0.16. Figure 14 ranks the top ten predictors based on the predictors' importance index.

In set C (PMV > 14 days), ANN used all the 27 predictors to classify the patients into the two classes. Injury Severity Score (ISS) scored the highest predictors' importance value (0.12). Figure 15 ranks the top ten predictors based on their importance.

Obviously, ML algorithms in set B demonstrated a higher discrimination power as presented by AUC. It is uncommon to define PMV to be greater than 10 days. This selection was related to the fact that the earlier liberation from MV is associated with improved patient outcomes. In addition, 10 days was identified to be the optimal period to perform a tracheostomy if ventilator support was still required (228). Therefore, defining PMV as >10 days seems to be more beneficial to devise an early individualized patient treatment plan.

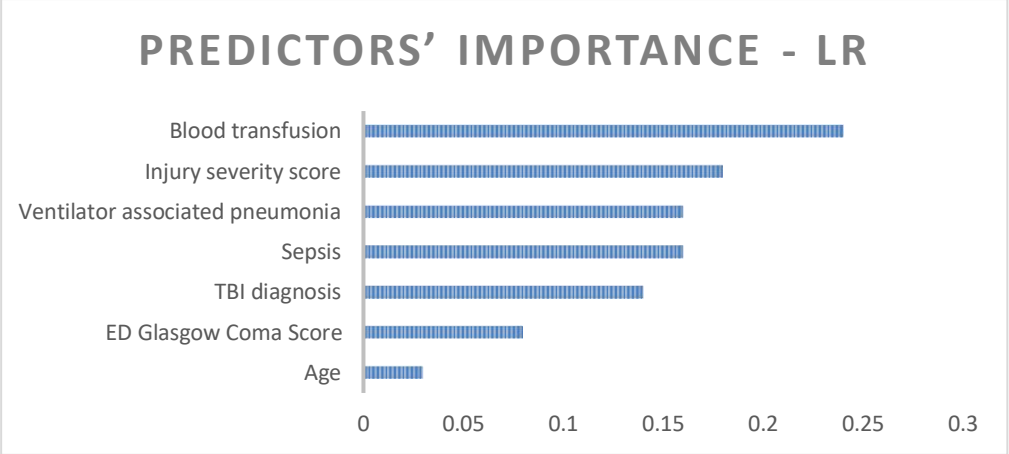

Figure 13. Predictors' importance chart – logistic regression (PMV> 7 days).

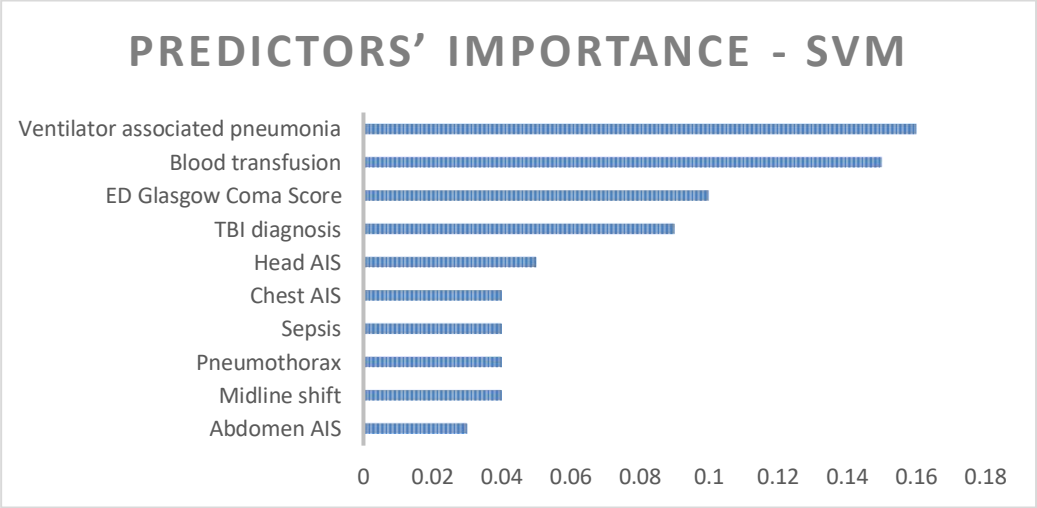

Figure 14. Predictors' importance chart – SVM (PMV> 10 days).

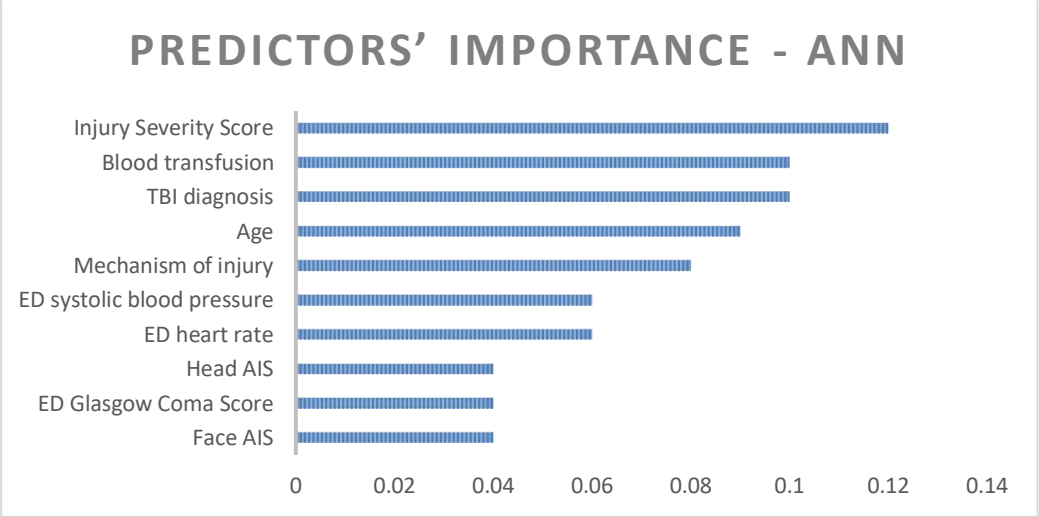

Figure 15. Predictors' importance chart – ANN (PMV> 14 days).

Discussion

Predicting PMV in patients with TBI is of utmost importance as early liberation from mechanical ventilation yields improved outcomes (120, 220). Nevertheless, predicting PMV is proven to be a difficult mission due to several factors such as lack of consensus on the PMV definition and the poor outcomes of the conventional analytical techniques used to make predictions. Traditional statistical techniques help clinicians predict PMV with only 59% accuracy (220, 226).

Hence, it is very important to help clinicians early identify patients at risk for PMV in order to design an individualized care plan and to decide on early tracheostomy, to help patients achieve better outcomes if MV is still required. The lack of a consensual definition of PMV makes the determination of the optimal MV duration and the tracheostomy decision very difficult. We opted for 10 days to be the cutoff point to differentiate between the PMV and the non-PMV, based on a prior Cochrane review which found that 10 days is the optimal

duration when moving patients to tracheostomy, which results in better outcomes (120).

This study demonstrated that the application of supervised ML techniques yields moderate overall performance for all the ML models in the three sets. Nevertheless, set B achieved more stable performance and higher discrimination power than sets A and C, with average AUC 0.813 for the five prediction models, compared to the average AUC values of 0.76 for set A and 0.72 for set C. SVM was the chosen prediction algorithm for set B, with accuracy of 79% and AUC of 0.84. The importance of this finding is that the optimal prediction performance is achieved when PMV is defined as >10 days, which is the optimal period for early tracheostomy (120, 228).

The ML models in the three sets achieved better performance than the traditional predictive techniques that recorded mild accuracies ranging between 0.60-0.69 and mild AUCs ranging between 0.52 and 0.67. This proves that the ML techniques outperform the conventional analytical techniques and can provide more support to the clinicians to make higher quality decisions that improve patient treatment outcomes.

In set B, SVM achieved the optimal performance with accuracy of 79% and AUC of 0.84. In addition, it performed moderately in the other measures (precision = 0.75, NPV = 0.82, sensitivity = 0.76, specificity = 0.82 and F-score = 0.75). The development of VAP ranked first in predictors' importance (0.16). Three-quarters of patients who developed VAP ended up with PMV> 10 days, compared to 35% of patients who did not develop VAP. VAP is known to be associated with poor outcomes, including PMV and mortality (194). This finding may not necessarily help early prediction of PMV, as VAP is not a condition that

a patient presents with when sustaining TBI. However, knowing that VAP contributes significantly to the PMV warrants the early implementation of the preventive measures, such as VAP bundle of care (233), which is becoming the standard of quality care in critical care medicine for ventilated patients.

The need for blood transfusion during resuscitation ranked second in predictors' importance, with a score of 0.15. 59.8% of patients who received blood for resuscitation needed PMV compared to 19.6% of the patients who did not require blood for resuscitation. The need to administer blood transfusion to resuscitate patients who sustain severe trauma could indicate the severity of injury, and perhaps a hypovolemic shock that contributes to poor patient outcomes. This affirms the findings of Ghiani et al. (234), who reported that blood transfusion is independently correlated with worse outcomes. Nevertheless, they concluded that blood transfusion is an indicator for disease severity rather than directly impacting the prognosis. Lai et al. found that the low hemoglobin level is associated with difficult weaning from MV and may lead to PMV (235). Also, Zubrow et al. (2018) found that the transfusion of Red Blood Cells (RBCs) in pediatrics with acute respiratory distress syndrome is associated with PMV (236).

The GCS of patients presenting to the ED is a significant predictor, with a predictors' importance score of 0.1. In this study, 49.8% of patients who presented to the ED with GCS ≤ 8 stayed on MV longer than 10 days, compared to 34.1% of patients who presented to the ED with GCS 13-15, and 26.4% of those who had GCS of 9-12. This finding is consistent with the previous literature which proves that patients with lower GCS are at higher risk of post-TBI complications, including death (225, 237).

TBI diagnosis (CT scan finding) scored predictors' importance of 0.09. 69% of patients who had diffuse axonal injury (DAI) and 66.7% of those who had cerebral edema stayed on a mechanical ventilator for longer than 10 days. Both cerebral edema and DAI are associated with significant mortality and morbidity (198, 238).

Furthermore, it is found that the greater the HAIS, the greater the risk of PMV. Almost 66% of patients who had HAIS = 5 ended with PMV, compared to 32% and 45.2% for HAIS 3 and 4 respectively. It is well documented that the more severe the TBI, the higher the risk of comorbidities and mortality (166). The same applies for predicting PMV (225).

Also, chest AIS scored 0.04 in predictors' importance. More than 50% of the patients who sustained chest injury with AIS between 1 and 5 had PMV compared to about 34% of the patients who didn't sustain chest trauma (chest AIS= 0). Previous literature found that chest AIS helps predict the PMV (224). Okabe found that the severity of blunt chest trauma is significantly associated with the risk of PMV (239).

Sepsis was also found to be among the top ten important predictors for the PMV (predictors' importance = 0.04). 83.3% of patients who suffered sepsis following TBI had PMV > 10 days. Like VAP, patients may not present to the ED with sepsis right after the TBI. However, knowing that sepsis contributes significantly to the PMV warrants the early implementation of the preventive measures i.e. six-hour sepsis bundle (240), which is a standard of critical care medicine.

Furthermore, 62.2% of patients who sustained pneumothorax stayed on MV longer than 10 days. Pneumothorax scored 0.04 in predictors' importance.

It is evident in the literature that pneumothorax and prolonged chest tube duration are associated with poor outcomes, such as PMV, increased ICU length of stay, and mortality (241).

In this study, midline shift is ranked number nine, with predictors' importance of 0.04. Midline shift is defined as the "displacement of septum pellucidum in relation to the midline in millimeters" (242). 52.4% of patients who sustained midline shift had PMV compared to 42.1% of those who did not have midline shift. Midline shift is a commonly used variable in predicting post-TBI unfavorable outcomes (e.g. CRASH and IMPACT tools) (181, 182, 201).

The tenth ranked predictor was abdominal AIS (predictors' importance 0.03). 59.4% of patients who had abdominal trauma with AIS (1-5) had PMV, compared to 41.6% of those who did not sustain abdominal injury. Although Blaser et al. found that there is a strong correlation between intra-abdominal hemorrhage and the ICU length of stay and the PMV (243), the severity of the abdominal injury as measured by the AIS was found to be an insignificant predictor of PMV in other studies (224, 225). This could be attributed to different data processing and inclusion criteria that were followed in every study.

Wellbeing and Economic Values

Predicting PMV is proven beneficial in several aspects. Besides the proven clinical value added, predicting PMV supports the decision of early tracheostomy, which is economically beneficial. PMV is associated with several complications that contribute to increasing ICU and hospital length of stay, and significantly escalated costs of care. Early tracheostomy is associated with reduced ICU and hospital length of stay (120, 244). It is estimated that the average daily cost of ICU in the USA ranges between $1,300 and $9,400,

depending on the specialization and complexity of patient disease (245). The daily cost in the ICU increases when mechanical ventilation is required (246). Therefore, it was found that the early tracheostomy contributes to a significant reduction in the ICU daily cost compared to delayed tracheostomy (229).

Usually, patients' treatment plans do not concentrate merely on addressing the acute and the chronic healthcare problems, but focus strongly on enhancing patient QoL. Patients who require mechanical ventilation suffer severe deterioration in QoL (247). It was found that early liberation from ventilator or early tracheostomy are associated with enhanced QoL (230). Therefore, this study adds value in several aspects that include the clinical, wellbeing, and economic aspects.

Conclusion

The importance of mechanical ventilation in critical care settings is increasing due to the increasing demand on the critical care intervention worldwide. Nonetheless, dependence on mechanical ventilation is associated with several serious outcomes. Therefore, the early liberation from mechanical ventilator is of utmost importance. Predicting patients at risk of PMV helps clinicians devise personalized care plans in order to mitigate the risk of PMV and to decide on tracheostomy in a timely fashion in case ventilator support is still required. Predicting patients at risk of PMV will not only help improve patients' clinical outcomes, but also help reduce critical care costs and enhance patients' QoL, while enabling the optimum deployment of healthcare services to meet user needs (e.g. freeing up ICU beds for other patients due to reduced hospitalization).

The study showed that it is possible to improve predictive power when

using ML approach, but the most important requirement is to enhance the quality of data in the NTR or electronic medical records to help improve the quality of predictions. Moreover, deploying such models into clinical practice and making them available in a user-friendly way to the clinicians to support their decision-making will be of great value.

CHAPTER 9: USING NTR DATA TO PREDICT VENTILATOR-ASSOCIATED PNEUMONIA IN PATIENTS WITH MODERATE TO SEVERE TBI: C.5 DECISION TREE APPROACH

Background

MV is one of the most common interventions in the critical care settings (176). Although it saves lives, MV is associated with several complications, including increased mortality (208). One of the most frequent complications that are associated with MV is VAP (248). VAP is “the pneumonia that occurs 48–72 hours or thereafter following endotracheal intubation, characterized by the presence of a new or progressive infiltrate, signs of systemic infection (fever, altered white blood cell count), changes in sputum characteristics, and detection of a causative agent” (249, 250). Approximately a fourth of all mechanically ventilated patients are at risk of VAP, which makes it one of the most common nosocomial infections in ICUs (251). More than 10% of deaths among mechanically ventilated patients are related to VAP (252, 253). Interestingly, the rate of VAP in trauma patients is four times higher than among non-trauma patients (254). This could be due to the fact that patients with trauma have more risk factors for developing VAP than non-trauma patients, such as pre-hospital or emergency intubation and polytrauma, including chest injuries (255). Several studies found that severe TBI is associated with increased risk of nosocomial pneumonia, which may lead to increased risk of mortality (256, 257). It was found that the risk for early onset VAP is frequent after severe TBI, with an overall rate of 61% (258).

VAP is associated with significant morbidity and mortality, and also with increased healthcare costs (254). The cost burden of treating VAP in the USA

was estimated in several studies to be between $10,000 to $40,000 per patient (259-261). In a study conducted in a cardiac intensive care unit in the UK, the cost of post-operative recovery after cardiac surgery for patients with VAP was at least two times greater than the cost of recovery for patients with no VAP (261).

The prediction of VAP in patients with TBI is of utmost importance. Early and powerful prediction of VAP helps clinicians and healthcare managers devise personalized intervention plans that help prevent the occurrence of VAP, thereby improving treatment outcomes and saving significant costs for patients and the healthcare system. The limited research predicting VAP in TBI mainly utilized conventional multivariate LR, with no reported area under the curve, accuracy, or positive predictive value. Aside from pioneering the ML predictive approach, this study is amongst the first studies to utilize NTR data to predict VAP in patients with TBI. This study designs a supervised ML model to predict the occurrence of VAP in adult patients who sustained moderate to severe TBI and who were admitted to L1TC.

Methodology

Data Preparation

Only adult patients (aged ≥ 14-years-old) who sustained moderate to severe TBI (HAIS ≥ 3) and who had received MV following the TBI were included in the study. Patients who sustained other systematic injuries with AIS > HAIS were excluded, to ensure that the primary injury was TBI. The retrieved data included the following variables: age, gender, mechanism of injury, mode of arrival, alcohol blood level, blood pressure, heart rate, GCS, CT findings, ISS, AIS per body region, intubation status and location, date/time of injury,

time of admission to ED, known comorbidities, performed procedures, blood transfusion, administration of VTE prophylaxis, in-hospital complications, outcome and date of disposition. Some other important variables such as time from injury until arrival to the ED were calculated secondarily. We hypothesized that the longer the time from injury until a patient reaches to the ED, the higher the likelihood of developing VAP, which could be due to the reduced level of consciousness and the increased risk of aspiration (255).

In this study, time from injury to ED is the only variable that was imputed. Patients who arrived via ambulance or helipad have their time to ED captured in the NTR. On the other hand, those who arrived via private vehicle (11%) did not have their time to ED captured in the NTR. Accordingly, the mean time to ED (118.7 minutes) was used for imputation when patients had VAP. Similarly, the mean time to ED (110.3 minutes) was used for imputation when patients had no VAP. On the other hand, records with missing values other than time to ED were eliminated from the study. Therefore, the final number of eligible patients was 772 patients. The detailed methodology is described in chapter 5.

Outcome Measure

The dichotomous outcome measure for this study is VAP. VAP0 indicates that the patient had no VAP reported during hospitalization, and VAP1 means that the patient had VAP during the hospitalization period.

Prediction Model

The objective of this study is to develop a supervised ML model that uses C.5DT to predict the occurrence of VAP in patients who received MV following moderate to severe TBI.

C.5 Decision Tree (C.5DT)

The DT is “a flowchart-like tree structure, where each internal node (non-leaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node”. DT helps extract IF-THEN rules which are considered easy to understand and interpret (93).

C.5DT is a widely used ML technique in the medical field (262). C.5DT is the successor of the C. 4.5 DT classification data mining algorithm. It works on the principle of information gain that reflects how much more information a predictor gives or contributes to reducing the entropy (impurity). The predictor that contributes more to reducing entropy will be the first split, and then the splitting continues until no more predictors significantly reduce entropy (263).

To prevent overfitting and to validate the models’ performance, we partitioned the data into training and testing sets (70% vs. 30% respectively). Data are portioned randomly by the analytical software based on the partition command. Furthermore, the overfit prevention was set at 30%. Table 29 explains the data partitions.

Table 28. Data Partitions

Set	Proportion	No. of cases	No. of no-VAP patients	No. of VAP patients
Training set	70%	541	423 (78.2%)	118 (21.8%)
Testing set	30%	231	180 (77.9%)	51 (22.1%)
Total	100%	772	603 (78.1%)	169 (21.9%)

Results

Of the 772 eligible patients (with a mean age of 33.1 years), 169 (21.9%) developed VAP during their hospitalization. The most common mechanism of injury was motor vehicle crash (37.8%), followed by fall from height (25.1%). The most common CT finding/ mass lesion was subdural hemorrhage (26.6%), followed by extradural hemorrhage (19.7%), with 34% of the patients sustaining midline shift. Tables 30 and 31 show the sample characteristics and the descriptive statistics for the study sample.

Table 29. Sample Characteristics – Continuous Variables

Variable	N	Mean	Mean with VAP
Age	722	33.1	33.7
Injury severity score (ISS)	722	28.2	28.9
ED systolic blood pressure (SBP)	722	126.3	126.3
ED heart rate (HR)	722	103	102.8
Time from injury to ED	722	112.1	118.6

Table 30. Sample Characteristics – Nominal and Ordinal Variables

Variable	Category	Count/%	With Outcome 0 (Alive)/%	With Outcome 1 (Dead)/%
Race	Asian	451/58.4	360/79.8	91/20.2
	Other	321/41.6	243/75.7	78/24.3
Mechanism of injury (MOI)	Motor vehicle crash	292/37.8	222/76	70/24
	Fall	194/25.1	157/80.9	37/19.1
	Pedestrian	161/20.9	124/77	37/23
	Other	125/16.2	100/80	25/20
Arrival mode	Ambulance	629/81.5	492/78.2	137/21.8
	Other	143/18.5	111/77.6	32/22.4
Multiple ribs fracture	No	589/76.3	463/78.6	126/21.4
	Yes	183/23.7	140/76.5	43/23.5
Lung contusion	No	500/64.8	384/76.8	116/23.2
	Yes	272/35.2	219/80.5	53/19.5
Hemothorax	No	666/86.3	519/77.9	147/22.1
	Yes	106/13.7	84/79.2	22/20.8
Pneumothorax	No	582/75.4	458/78.7	124/21.3
	Yes	190/24.6	145/76.3	45/23.7
Midline shift	No	509/65.9	404/79.4	105/20.6
	Yes	263/34.1	199/75.7	64/24.3
TBI diagnosis	Subdural hemorrhage	205/26.6	166/81	39/19
	Epidural hemorrhage	152/19.7	119/78.3	33/21.7
	Subarachnoid hemorrhage	67/8.7	51/76.1	16/22.1
	Hemorrhagic contusions	91/11.8	89/79.5	23/23.9
	Cerebral edema	84/11.9	66/78.6	18/21.4
	Diffuse axonal injury	104/13.5	70/67.3	35/32.7
	Other	69/8.9	55/79.7	14/20.3
Head AIS (HAIS)	3	231/29.9	186/80.5	45/19.5
	4	185/24	150/81.1	35/18.9
	5	356/46.1	267/75	89/25
Chest AIS (CAIS)	0	344/44.6	268/77.9	76/22.1
	1 (AIS 1-2)	119/15.4	94/79	25/21
	2 (AIS 3-5)	309/40	241/78	68/22
Abdomen AIS (AAIS)	0	601/77.8	463/77	138/23
	1 (AIS 1-5)	171/22.2	140/81.9	31/18.1
Spine AIS (SAIS)	0	531/68.8	420/79.1	111/20.9
	1 (AIS 1-5)	241/31.2	183/75.9	58/21.4
Extremities AIS (EAIS)	0	408/52.8	331/81.1	77/18.9
	1 (AIS 1-2)	258/33.4	191/74	67/26
	2 (AIS 3-5)	106/13.7	81/76.4	25/23.6

Variable	Category	Count/%	With Outcome 0 (Alive)/%	With Outcome 1 (Dead)/%
Glasgow Coma Score (GCS) category	13-15	93/12	69/74.2	24/25.8
	9-12	78/10.1	71/91	7/9
	≤ 8	601/77.8	463/77	138/23
Known comorbidities	No	646/83.7	520/80.5	126/19.5
	Yes	126/16.3	83/65.9	43/34.1
Venous thromboembolism prophylaxis	No	176/22.8	165/93.8	11/6.3
	Yes	596/77.2	438/73.5	158/26.5
Blood transfusion	No	249/32.3	225/90.4	24/9.6
	Yes	523/67.7	378/72.3	145/27.7
Total		772/100	603/78.1	169/21.9

Performance of the ML models

Table 32 shows the performance evaluation metrics for the C.5DT in the test data partition. C.5DT scored low sensitivity (0.43). This could be attributed to the quality of the data, particularly data imbalance. Two methods were considered to improve sensitivity. Both random over-sampling and random under-sampling were used to improve data quality and to improve sensitivity. Both techniques scored lower sensitivity (0.18 and 0.29, respectively), and more importantly significantly lowered the precision, F-score, and specificity. Therefore, we determined that the current form is the optimal form of data to achieve the best possible performance measures.

Table 31. Performance of the Classification Models

Model	No. of predictors	Accuracy%	AUC	Precision	NPV	Sensitivity	Specificity	F-Score
C.5DT	5	83.5	0.8	0.71	0.86	0.43	0.95	0.54

VAP Risk Factors

Out of 24 predictors, C.5DT identified five predictors as significant for predicting VAP in patients with moderate to severe TBI: time from injury to ED, blood transfusion during resuscitation, known comorbidities, ISS, and pneumothorax, in descending order of importance (Figure 16).

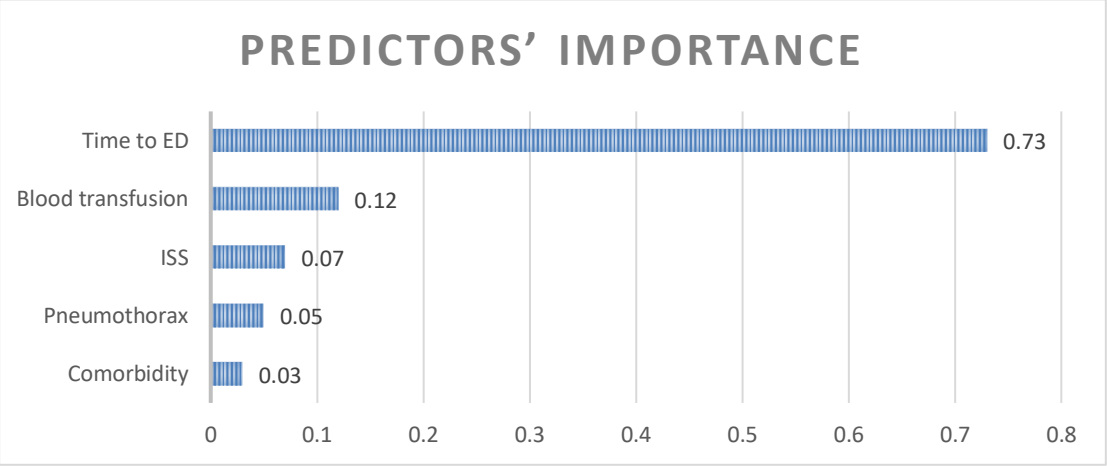

Figure 16. Predictors' importance in C.5DT.

Discussion

VAP is one of the most frequent complications in intubated patients that increase hospital length of stay and may lead to mortality. Therefore, early

prediction of VAP may guide early personalized preventive measures. Understanding VAP risk factors facilitates early preventive efforts, reducing the risk of VAP and its consequences and helping reduce the costs of healthcare while improving patient QoL and clinical outcomes. The prediction of disease prognosis and outcomes requires developing good prognostic models that include large samples and enjoy high external and internal validity in order to be generalizable beyond a specific research setting (191).

Previous literature identified several risk factors for developing VAP among trauma patients. The presence of polytrauma (256), the intubation location (prehospital vs. trauma resuscitation room) (255), failed prehospital intubation (264), the presence of chest injury (such as rib fractures and pulmonary contusion) (194, 255, 264, 265), severity of head and neck injury (266), severity of chest injury as measured by the Abbreviated Injury Scale Score (chest AIS) (258), injury severity score (ISS), coma upon admission, age, and GCS (194) are all significant predictors of VAP.

The C.5DT model yielded moderate predictive performance, with 83.5% accuracy, 0.805 AUC, 0.71 precision (PPV), 0.86 NPV, 0.43 sensitivity, and 0.95 specificity. In the training C.5DT model, the depth was 5, the number of nodes was 16, and the number of terminal nodes was 9. We identified five variables that significantly predict VAP among patients with moderate to severe TBI.

Similar to previous literature (169), time from injury until arrival to the ED was found to be a significant predictor for the development of VAP in patients with TBI. C.5DT identified time from injury to ED to be the most significant predictor in predicting VAP in patients with TBI. The average injury to ED time

was 112.1 minutes, while the time from injury to ED in patients who developed VAP was 118.6 minutes. The time from injury to ED was first split at 113.5 minutes. 47.6% (49 out of 103 patients) of those who arrived at the ED after 113.5 minutes following TBI developed VAP, compared to 15.8% (69 out of 438 patients) who arrived at the ED ≤ 113.5 minutes following the TBI. The second split was for those who arrived at ED in shorter or equal to 113.5 minutes at the level of 110.15 minutes, where all 69 patients (24.1%) who arrived at the ED in a shorter time than 110.15 minutes developed VAP; while none of those who arrived between 110.15 minutes and 118.8 minutes developed VAP. For the group who arrived at the ED after 113.5 minutes (103 patients), of the 32 who arrived in a time shorter than 118.8 minutes, 31 of them (96.9%) developed VAP, while of the remaining 71 patients who arrived at the ED after 118.8 minutes, only 25% (18 patients) developed VAP.

The second most important variable was blood transfusion during resuscitation. This finding is consistent with previous literature, which found a significant association between blood transfusion and risk of developing VAP (267-269). 217 patients arrived at the ED within 110.15 minutes, of whom 203 received blood during resuscitation, compared to 83 who did not receive blood. 62 patients among those who received blood (30.5%) developed VAP, while only 7 patients (8.4%) of those who did not receive blood developed VAP.

Pneumothorax was found to be a significant predictor for VAP. There were 71 patients who arrived at the ED in a longer time than 118.8 minutes following trauma. Of them, 18 patients (25.4%) developed VAP. Among the 71 patients, 11 sustained pneumothoraxes. 7 out of the 11 patients (63.6%) developed VAP compared to 11 out of the 60 patients who did not sustain

pneumothorax (18.3%) who developed VAP. This result is consistent with previous literature which found that pulmonary parenchymal injury caused by pneumothorax after chest tube insertion could increase the risk of developing VAP (270, 271). Interestingly, 100% of the 7 patients who sustained pneumothorax and took longer than 135.5 minutes to arrive to the ED following TBI developed VAP.

Patients with chronic morbid conditions were found to have higher risk of developing VAP (270, 272). Amongst patients who received blood during resuscitation, there were 34 with known comorbid conditions such as diabetes, hypertension, and coronary artery disease etc., of whom 18 patients (53%) developed VAP, compared to 44 out of 169 patients (26%) who developed VAP after receiving blood who were free of comorbidities. Finally, ISS was found to play a significant role in predicting VAP. This result is consistent with the previous research (194, 255, 266, 273). 18 out of 25 patients (72%) who had ISS of 34 or less and who were known to have a chronic morbid condition developed VAP, while none of those who were known to have comorbid conditions with an ISS greater than 34 developed VAP.

Finally, to summarize the identified potentially interesting patterns, VAP was developed by 72% of the patients who: had ISS ≤ 34; were known to have one or more comorbidities; received blood during resuscitation; and arrived at the ED within 110.15 minutes after the trauma developed. The second pattern was that 100% of those who arrived the ED after more than 135.5 minutes and who were diagnosed with traumatic pneumothorax developed VAP (Figure 17).

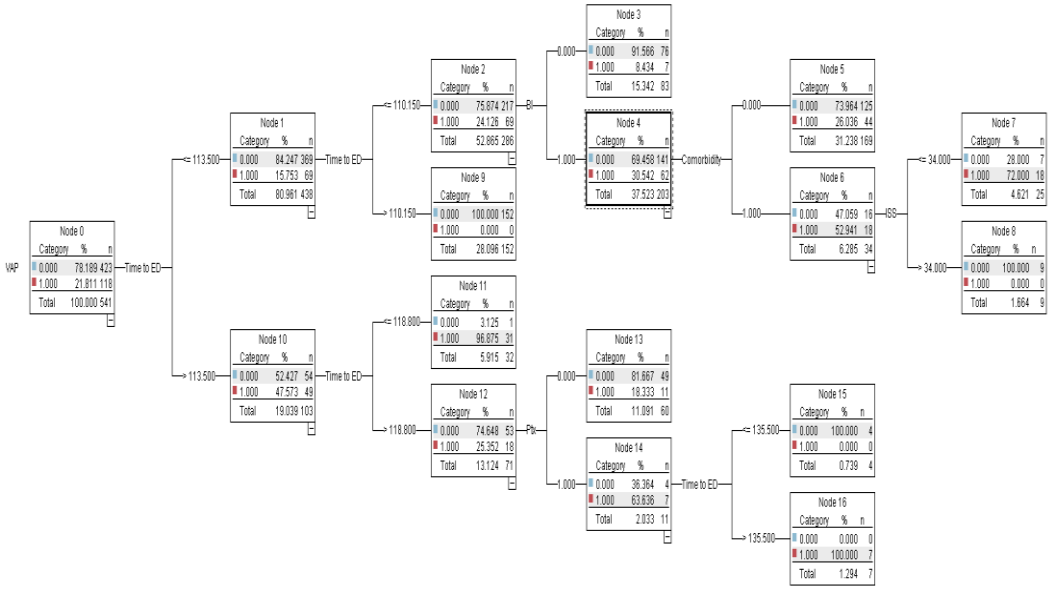

Figure 17. VAP predictors by C.5DT.

Conclusion

Implementing ML methods in predicting VAP in patients with moderate to severe brain injury is beneficial and may enhance the preventive efforts of the healthcare providers. Deploying C.5DT may require collaboration between clinicians and data scientists in order to integrate the ML predictive models with electronic medical records. This integration helps provide timely decision support to the care givers to improve patients' treatment outcomes. This integration also facilitates timely discharge from the hospital, and reduces healthcare costs for patients and their families, and healthcare systems.

CHAPTER 10: USING ARTIFICIAL NEURAL NETWORK TO PREDICT PROLONGED HOSPITAL LENGTH OF STAY IN PATIENTS WITH TBI

Background

Hospitals are entrusted to provide timely care to the patients while maintaining efficient resources utilization (274). One of the most essential resources in hospitals is beds, and to provide timely patient care, hospitals are always under pressure to keep beds available (275). PLOS and Bed Turnover Rate (BTR) are among the indicators that reflect hospitals' performance (276). BTR is an indicator of productivity. It is measured by dividing the total number of discharges (including deaths) over the total number of available beds. Accordingly, longer hospital LOS negatively affects BTR, and makes the timeliness of care delivery a challenging task (276, 277). Besides the efficiency and the timeliness of care delivery, reducing LOS helps reduce costs for patients and healthcare systems (278).

Several determinants of hospital PLOS in trauma have been identified in prior research. Moore et al. identified age, ISS, known comorbid conditions, and discharge destination as significant determinants of LOS following trauma (279). Chen et al. (280) found that the extent to which care providers interact with each other in the care of the trauma patients significantly affects LOS. Others found that in-hospital complications such as renal failure, respiratory failure, sepsis, ISS, blood transfusion, invasive ventilation, and initial GSC score (≤8) have significant impacts on ICU LOS for trauma patients (281, 282). The mechanism of injury also has a significant association with the extended length of stay for pediatric patients with trauma (283), as well as receiving blood during resuscitation and surgical procedures among adult trauma patients

(284). Also, age, ISS, pre-existing morbid conditions such as renal disorders and some infectious complications were found to be significant predictors for LOS greater than 30 days in the trauma ICU (285). On the other hand, Hwabejire et al. found that the PLOS is not necessarily associated with injury severity. They found that system and operational related issues are the leading cause of the PLOS (286).

An important aspect of PLOS is its cost considerations. In the USA, the annual cost of trauma care is greater than $37 billion (287). In Canada, more than 200,000 admissions following trauma were calculated to have cost more than $11 billion (279). The estimated daily cost in the ICU varies based on the hospital day and the required interventions, such as mechanical ventilation (246). For example, the average cost of day 1 in the ICU (without mechanical ventilation) is approximately $6,600 and with mechanical ventilation it is > $10,000. The cost reduces on day 2 to $3,500 and $4,700 for non-intubated and intubated patients, respectively. In addition, the cost changes significantly between medical and surgical inpatient unit admissions. Li et al. (288) found that the cost per day in the acute surgical inpatient units is significantly higher than in medical units, with $18,000 and $6,000, respectively. Furthermore, they found that the cost per additional day increases by $112 in surgical units compared to $79 in medical units. The cost varies among countries based on several factors, such as cost of living, average wage, healthcare subsidies, and healthcare professional to service user ratios, etc.

Very importantly, there is significant association between the length of stay and several unfavorable outcomes such as in-hospital mortality and morbidity (289, 290). Mathew et al. (291) found that every extra day of stay in

hospital after completing medical treatment increases the odds of developing complications by 5%. From another perspective, in-hospital LOS predicts patient and family satisfaction with trauma center services, which is another important hospital performance measure (292). Therefore, predicting PLOS early may help the hospital administration and clinicians mobilize the required resources to facilitate early discharge, which contributes to improved patient treatment outcomes, cost reduction, and improved patient and family satisfaction.

There is no single agreed upon definition for PLOS in the literature. Some scholars defined PLOS based on previous literature as ≥21 days (283, 293), while others defined it as ≥ two standard deviations (SDs) above the mean LOS of the diagnosis-related group (286). For example, the mean LOS for the forearm fracture without complications or comorbidities for patients older than 17 years is 3 days. So, 2 SDs above the mean was calculated to be 9 days, greater than the upper quartile of all the cases or greater than the 95th percentile of all the cases (278, 290, 294). Additionally, some defined PLOS to be greater than 30 days from the time of admission (285, 291). This study, similar to Krell et al. (295), defines PLOS to be greater than the 75th percentile of the entire cohort, which is 23 days.

Methodology

This study aims to design a supervised ANN ML model to predict in-hospital PLOS for adult patients who sustained moderate to severe TBI and admitted to L1TC.

Data Preparation

Only adult patients (≥ 14-years-old) who sustained moderate TBI (HAIS

= 3 or HAIS = 4) to severe TBI (HAIS = 5) were included in the study. Records with missing data were eliminated. Furthermore, all patients who died during their initial hospitalization were excluded. Subsequently, 1417 eligible patients were included in the study.

Outcome Measure

The outcome measure is the in-hospital PLOS, which is staying in the hospital more than 23 days from the day of initial admission following the trauma. 23 days is the 75^{th} percentile of all the cases. Staying more than 23 days was coded 1, while less than or equal to 23 days was coded 0. Transfer to another hospital for rehabilitation was considered as discharge.

ML Prediction Model

This study uses ANN for the early prediction of PLOS for patients with moderate to severe TBI. In this study, we used the MLP architecture, as it outperformed RBF architecture during the initial assessment (with accuracy of 85.6% vs. 81.9%, AUC of 0.91 vs. 0.862, and precision of 0.71 vs. 0.60, respectively).

Typically, MLP network consists of at least three layers of neurons: one input layer that represents the predictors, one or more hidden layer(s) representing computational neurons, and an output layer representing the outcome variable(s) (296). MLP usually uses sigmoidal activation function, and the training is conducted using backpropagation.

To prevent overfitting and to validate the models' performance, we partitioned the data into training set (75%) and testing set (25%), with the overfit prevention set at 30%. The partitioning was done randomly by the analytical software, based on the partitioning node specification. Table 33 explains the

data partitions.

Table 32. Data Partitions

Set	Proportion	No. of cases	No. of patients without PLOS	No. of patients with PLOS
Training set	75%	1055	790	265
Testing set	25%	362	277	85
Total	100%	1417	1067	350

Results

We included 1417 patients in the study, of whom 350 (24.7%) stayed in the hospital longer that 23 days (PLOS) from the day of admission. The mean age of all patients was 34 years. The most common mechanism of injury was fall from height (34.9%), followed by motor vehicle crash (29.1%). The most common CT finding/ mass lesion was subdural hemorrhage (26.1%), followed by extradural hemorrhage (24.7%) with 18.5% of the patients sustaining midline shift. Tables 34 and 35 show the sample characteristics and the descriptive statistics for the study sample.

Table 33. Sample Characteristics – Continuous Variables

Variable	N	Mean	SD	Mean at PLOS
Age	1417	34	13.5	34.6
Injury severity score (ISS)	1417	20.7	9.3	27.7
ED systolic blood pressure (SBP)	1417	129	20.3	128.5
ED heart rate (HR)	1417	90.8	21.6	99.8

Table 34. Sample Characteristics – Nominal and Ordinal Variables

Variable	Category	Count/%	With Outcome 0 (no PLOS)/%	With Outcome 1 (PLOS)/%
Gender	Male	1339/94.5	1011/75.5	328/24.5
	Female	78/5.5	56/71.8	22/28.2
Race	Asian	858/60.6	665/77.5	193/22.5
	Other	559/39.4	402/71.9	157/28.1
Mechanism of injury (MOI)	Motor vehicle crash	413/29.1	274/66.3	139/33.6
	Fall	495/34.9	414/83.6	81/16.4
	Pedestrian	216/15.2	141/65.3	75/34.7
	Other	293/20.7	238/81.2	55/18.8
Midline shift	No	1155/81.5	914/79.1	241/20.9
	Yes	262/18.5	153/58.4	109/41.6
CT findings/mass lesion	Subdural hemorrhage	370/26.1	277/74.9	93/25.1
	Epidural hemorrhage	350/24.7	291/83.1	59/16.9
	Subarachnoid hemorrhage	110/7.8	79/71.8	31/28.2
	Hemorrhagic contusions	285/20.1	241/84.6	44/15.4
	Cerebral edema	47/3.3	25/53.2	22/46.8
	Diffuse axonal injury	99/7	30/30.3	69/69.7
	Other	156/11	124/79.5	32/20.5
Head AIS (HAIS)	3	794/56	683/86	111/14
	4	302/21.3	208/68.9	94/31.1
	5	321/22.7	176/54.8	145/45.2
Face AIS (FAIS)	0	842/59.4	664/78.9	178/21.1
	1	151/10.7	108/71.5	43/21.5
	2 (AIS 3-5)	424/29.9	295/69.6	129/30.4
Chest AIS (CAIS)	0	960/67.8	808/84.2	152/15.8
	1 (AIS 1-2)	145/10.2	96/66.2	49/33.8
	2 (AIS 3-5)	312/22	163/52.2	149/47.8
Abdomen AIS (AAIS)	0	1271/89.7	1006/79.2	265/20.8
	1 (AIS 1-5)	146/10.3	61/41.8	85/58.2
Spine AIS (SAIS)	0	1091/77	863/79.1	228/20.9
	1 (AIS 1-5)	326/23	204/62.6	122/37.4
Extremities AIS (EAIS)	0	917/64.7	761/82.9	156/17.1
	1 (AIS 1-2)	380/26.8	244/64.2	136/35.8
	2 (AIS 3-5)	120/8.5	62/51.7	58/48.3
Glasgow Coma Score (GCS) category	13-15	875/61.8	799/91.3	76/8.7
	9-12	113/8	86/76.1	27/23.9
	≤ 8	429/30.3	182/42.4	247/57.6

Variable	Category	Count/%	With Outcome 0 (no PLOS)/%	With Outcome 1 (PLOS)/%
Known comorbidities	No	1167/82.4	896/76.8	271/23.2
	Yes	250/17.6	171/68.4	79/31.6
Intubation	No	847/59.8	803/94.8	44/5.2
	Yes	570/40.2	264/46.3	306/53.7
Venous thromboembolism prophylaxis	No	537/37.9	520/96.8	17/3.2
	Yes	880/62.1	547/62.2	333/37.8
Blood transfusion	No	989/69.8	878/88.8	111/11.2
	Yes	428/30.2	189/44.2	239/55.8
Total/%		1417/100	1067/75.3	350/24.7

Performance of the ANN Prediction Model

To calculate the model's performance metrics, we first constructed the confusion matrix that displays the relationship between the actual observations and the predicted conditions. The model's performance is evaluated in terms of several metrics: accuracy, Area Under the Curve (AUC), precision (positive predictive value), negative predictive value, sensitivity, specificity, and F-score.

The model achieved good discrimination power (AUC = 0.911), positive predictive value (0.73), negative predictive value (0.88), and accuracy (0.854). Table 36 shows the performance metrics of the prediction model.

Table 35. ANN Performance Metrics

	No. of predictors	Accuracy %	AUC	Precision	NPV	Sensitivity	Specificity	F-Score
ANN	20	85.4	0.91	0.73	0.8	0.6	0.93	0.66

In-Hospital PLOS Risk Factors

ANN utilized all the 20 variables in predicting the in-hospital PLOS. Figure 18 ranks the predictors based on their PI. The most important predictor is the patient's age (0.13), followed by the receiving intubation (0.12).

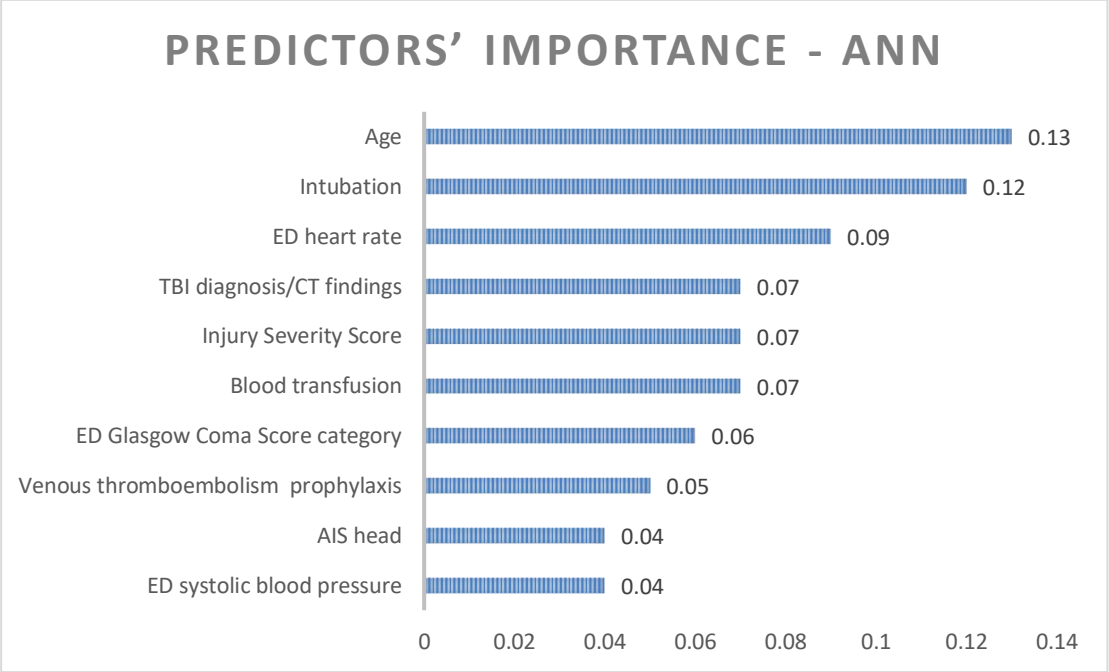

Figure 18. Predictors' importance

Discussion

Predicting PLOS is of utmost importance for the timeliness of care giving and for the healthcare economics. Researchers have identified several risk factors such as age, ISS, mechanism of injury, presenting vital signs, severity of TBI, and severity of associated injuries for PLOS. Moreover, scholars attempted to adjust several well-known mortality prediction models to predict LOS. Woods et al. evaluated the capacity of Acute Physiology and Chronic

Health Evaluation (APACHE) III system to predict ICU patients' LOS, but the overall model performance was poor (297). In a similar vein, Chattopadhyay and Chatterjee (298) evaluated the performance of APACHE IV in predicting the LOS for ICU patients with severe sepsis. They found that APACHE IV poorly predicts ICU patients LOS. Vasilevskis et al. (299) compared the performance of a recalibrated acute physiology and chronic health evaluation (APACHE) IV-LOS model with the mortality probability model III at zero hours (MPM (0)) and the simplified acute physiology score (SAPS) II mortality prediction model to predict LOS. Their study found that APACHE IV performs better than MPM0 and SAPS II. By examining SAPSII in predicting LOS for ICU LOS, they found that the model's performance is poor and tends to estimate patients' LOS, with low discrimination power (AUC = 0.62). However, the updated SAPSIII performed better than SAPSII in predicting LOS of ICU patients, with an AUC of 0.85 (300). MPM0 performance was also evaluated in predicting ICU LOS and performed poorly (299, 300).

This study identified several predictors for PLOS in patients with moderate to severe TBI. Patient's age ranks first in predictor's importance (PI = 0.13). This study found that the average age for patients who were discharged before 23 days was less than the average age for those who stayed longer than 23 days (34.6 years vs. 33.8 years). This finding is consistent with previous literature, which found age to be a significant determinant of LOS following trauma (279, 285). Intubation in the first 24 hours following the TBI ranks second in the predictors' importance (PI = 0.12). 54% of the patients who were intubated stayed in the hospital longer than 23 days compared to 5% only of those who were not intubated. Although intubation and mechanical ventilation

save lives, they are associated with increased length of stay and several serious complications, including VAP and mortality (208, 301). The third most important predictor was the initial heart rate recorded at the ED after the trauma (ED HR) (PI = 0.09). The average ED HR for patients who stayed in the hospital longer than 23 days was 99.8, compared to average ED HR of 87.9 for those who stayed shorter than 23 days. This result is consistent with Paterson et al. (302), who studied the impact of a standardized early warning score on predicting the in-hospital mortality and the length of stay. They found that the initial heart rate that is recorded at the point of entry to care plays a significant role in predicting the LOS.

The TBI diagnosis as per the CT scan finding was the fourth most important predictor (PI = 0.07). Around 70% of patients who sustained diffuse axonal injury (DAI) had PLOS, compared to 28% and 25% for those who sustained subarachnoid hemorrhage and subdural hemorrhage respectively. This is consistent with the previous literature which found that TBI diagnosis, particularly DAI, is associated with several outcomes, including extended ICU/hospital LOS (215, 238).

The ISS was the fifth most important predictor (PI = 0.07). The average ISS for patients with PLOS was approximately 28, compared to ISS of 18 for those with no PLOS. This is consistent with the previous literature that found ISS to be independent predictor for LOS (281, 285).

Blood transfusion is also found to play a significant role in predicting the PLOS (PI = 0.07). 56% of patients who received blood transfusion during resuscitation had PLOS, compared to 11% only of those who did not receive blood transfusion during resuscitation. Malone et al. (303) found that blood

transfusion in the first 24 hours following the trauma serves as an independent predictor of the ICU LOS. This finding perhaps reflects a correlation rather than a causation relationship. Usually, patients who need resuscitative blood transfusion are those who have severe injuries and lost significant amounts of blood, which is associated with serious unfavorable outcomes. This may help explain why patients who receive resuscitative blood transfusion are more prone to PLOS than those who do not require resuscitative blood transfusion.

The seventh important predictor is the initial Glasgow Coma Score category that is recorded in the emergency following trauma (ED GCS). Consistent with the previous literature (281, 282), this study found that around 58% of those who presented to the ED with GCS ≤ 8 had PLOS compared to 24% of those who presented with GCS between 9 and 12, and 9% of those who presented with GCS 13-15.

The administration of VTE prophylaxis was found to be a significant predictor for the PLOS. The study found that those who receive VTE prophylaxis within the first 48 hours from admission are more prone to PLOS compared to those who have not received VTE prophylaxis (37.8% vs. 3.2% respectively). Interestingly, although the administration of VTE prophylaxis is associated with a higher survival rate in patients with TBI (197, 301), there is no evidence in the literature that the administration of VTE prophylaxis reduces the overall hospital LOS. Accordingly, this relationship, similar to blood transfusion, could reflect a correlation rather than a causation relationship as explained in chapter 7.

The ninth ranked predictor (PI = 0.04) was HAIS. Patients with higher HAIS are at higher risk for PLOS. 45.2% of patients who had HAIS 5 had PLOS

compared to 14% and 31.1% for patients with HAIS 3 and 4 respectively. It is scientifically known that the more severe the TBI, the higher the risk of unfavorable outcomes (166), which is usually correlated with increased hospital LOS.

The initial systolic blood pressure recorded in the ED following trauma (ED SBP) was found to be a significant predictor of LOS (PI = 0.04). This is consistent with Paterson et al. (302), who found that the initial SBP recorded at the point of entry to care after trauma plays a significant role in predicting hospital LOS.

Healthcare Economics Implications

Early prediction of patients prone to PLOS following TBI helps healthcare teams devise personalized plans of care that address risk factors. Besides reducing LOS, this helps reduce the healthcare bill and boost patient and family satisfaction and QoL. One of the motives of this work is that the costs of healthcare in Qatar are the highest in the region. According to the World Health Organization Choosing Interventions that are Cost Effective project (WHO-CHOICE), the cost per bed day in the tertiary hospitals in Qatar is the highest in the Gulf Cooperation Council (GCC) countries. The cost per bed day in Qatar is around $270, compared to $224 in Bahrain, $223 in the United Arab Emirates, $209 in Kuwait, $170 in Saudi Arabia, and $158 in Oman. In 2013, Tuma et al. (304) estimated the cost of care per patient per ICU day in Qatar to be about $1,500. Furthermore, in a report published in the Qatari *Peninsula Newspaper* (305), the cost per night in the trauma surgery inpatient unit without the cost of surgery and medication ranges between $1,300-$1,900 and the cost per night in the trauma ICU is more than $2,700. In this study, the average LOS

for patients who stayed in the hospital for a period greater than 23 days was 47 days. The average cost of the excess 24 days for the 350 patients (≈$1,900/ day) is estimated to be greater than $15 million during the study period. Therefore, reducing one day only for the 350 patients with PLOS would have saved more than $8000 daily.

Beside the direct monetary cost per bed day, healthcare executives need to account for the opportunity cost of blocked beds, which delay specialized treatment for subsequent patients. This delay in care delivery is a key contributing factor to poorer disease prognosis, increased risk of comorbidity and mortality, reduced patient and family satisfaction, and very importantly a factor for increased PLOS. Sandmann et al. (306), calculated the opportunity cost of bed days in terms of health forgone for the second patient. The study found that the opportunity cost of the bed days consumed by the first patient is around £14,000 in terms of the net benefits forgone for the second patient. Unavailability of beds in a timely manner can also lead to higher healthcare costs to the individual or the country, as it can lead to seeking faster alternatives in other countries with major additional travel, healthcare, and social costs.

This economic perspective could be a good avenue for future research that utilizes ML modeling techniques to estimate the cost burden of trauma patients, and to estimate the opportunity cost of reduced hospital beds turnover rate.

Conclusion

The study proves that the application of ML to predict PLOS may bring wellbeing and economic benefits to the patients and the healthcare system. ANN achieved good predictive performance. Therefore, utilizing ML techniques

presents an opportunity to enhance hospitals' capacity to improve their bed turnover rate, timeliness of care delivery, and patient and family satisfaction, while reducing healthcare costs for service users, healthcare systems, and national economies. The results of this study may encourage decision makers in trauma surgery units to integrate ML techniques with the NTR and the electronic medical records. This may help clinicians plan their preventive efforts and mobilize necessary resources in an earlier stage of patient treatment, which could improve care outcomes and enable more effective and efficient deployment of healthcare system resources. Furthermore, the results were consistent with the known body of knowledge. Thus, with the availability of massive data sets in the electronic medical records and other structured registries, clinical evidence could be made available quickly and with less effort.

CHAPTER 11: PRACTICAL AND THEORETICAL SIGNIFICANCE AND IMPLICATIONS

The motive of this doctoral dissertation was to answer the compelling question of why the benefits of ML in supporting clinical decisions are not fully realized. There is overwhelming evidence that other industrial sectors benefited significantly from the powerful computational capacity of AI modalities, and specifically ML. Following an extensive literature review, we realized that in order to answer the question, we needed to answer two important sub-questions that reflect significant gaps in current knowledge:

(a) How can ML be integrated with EBM to support clinical decision making?

(b) Why is ML adoption in healthcare relatively slower than in other sectors?

(c) How can ML adoption support clinical decision-making?

Through reviewing the literature, we found that there is consensus that the potentials of ML in supporting clinical decision-making are undeniable. There are various avenues where ML can help advance clinical practice through engagement with EBM to provide real-time clinical evidence to better inform clinical decisions. Nevertheless, it was obvious that there are significant challenges that face the methods that abstract knowledge from data in clinical fields dominated by the EBM paradigm.

We discussed that on top of these challenges there are epistemological barriers. EBM determines how the clinicians base their decisions on sound evidence that guarantees, to a great extent, the best possible treatment outcomes. Therefore, we discussed that the epistemological differences between the EBM and the data science paradigms are largely irreconcilable. Thus, it is epistemologically impossible to synthesize a third paradigm that

integrates both EBM and data science.

Subsequently, we argued that the only way to benefit from the power of ML in supporting the clinical decision is through proposing a pragmatic and practical reconciliatory framework that transcends the epistemological differences of the two paradigms to capitalize on their synergistic effect, whereby one paradigm can address some of the challenges of the other.

The utility of the framework lies in its ability to provide conditional guidance to clinical practitioners and researchers on when and how to benefit from both paradigms to produce sound evidence to support their clinical decisions. This framework can overcome one of the key challenges that EBM suffers from, which is the lack of sound and high-quality evidence to inform clinical decision making in some scenarios. With the abundance of clinical data that originates from the electronic health records, sensors, and medical devices, etc. we became able to provide the real-time information that can guide clinicians' decisions with less efforts, time, and costs, using modern analytical approaches.

Nonetheless, we took into consideration that there must be guiding principles that ensure the judicious utilization of data-driven knowledge in clinical decision-making. On top of the guiding principles is the need to referee the knowledge by the subject matter experts considering the known body of knowledge. This is crucial to ensure the safety of the patients. We implemented this in our empirical work, whereby all the predictive models' outputs were evaluated by senior trauma surgery consultants and clinical researchers.

On the other hand, it was necessary to emphasize the pragmatic aspects of the framework, through admitting that this framework will not address all the

challenges from which EBM and ML suffer. Instead, it transparently acknowledges that EBM will continue to have the upper hand in shaping the clinical decision-making process, and that the ML can effectively fill the gap in knowledge within a certain pre-determined context. This explains why we put significant emphasis on transforming the focus from the best available evidence to the best possible or feasible evidence. The term “best possible/feasible evidence” respects the principle of scientific knowledge tentativeness. In conditions where high-quality evidence is not obtainable using mainstream methods, ML can help provide the knowledge which can support the clinical decision until better quality evidence can be reached. Accordingly, the practical pragmatic framework answers the first sub-question about how to integrate the ML with the EBM.

However, the practical framework will work only when the utility of the ML is realized by clinicians. It was found in several studies that clinicians’ awareness of the potentials of AI in general and ML in particular is poor. This was demonstrated in a recent market study conducted by McKinsey Global Institute in 2019, which reported that the healthcare industry is the slowest industry in adopting the AI (122). Thus, it was necessary to understand the underlying factors that lead to the slow adoption in order to theorize a framework that can guide the efforts to enhance the effective adoption of ML in clinical decisions support. Accordingly, we identified nine enablers that may lead to the desired outcome of enhancing clinicians’ intentions to adopt ML in their clinical practice.

The framework that we presented provides a genuine logical roadmap that guides the efforts towards enhancing clinicians’ adoption of ML through

interrelated steps that are ordered based on their driving and dependence power. Despite our belief that the framework provides a theoretical foundation for the enhancement of the adoption of ML in the healthcare industry, we believe that significant work is still required to validate the framework and to ensure its comprehensiveness and value. This would be an area for future research that may pave the way for the healthcare industry to realize the benefits of AI and to be prepared for facing the challenges (and utilizing the opportunities) of the fourth industrial revolution.

Consequentially, the answer to the second and third research questions were answered through providing the framework that identifies and defines the enablers of the way towards achieving the effective adoption of ML in supporting clinical decision making, which is an integral step towards reaping the full benefits of the ML in healthcare industry.

CHAPTER 12: MANAGERIAL IMPLICATIONS

This doctoral dissertation provides theoretical and practical guidance to the clinicians, clinical researchers and the policy makers on how to benefit from the potentials of the ML in supporting clinical decisions. Very importantly, the scope of the work is to enhance the clinical decision making. It is undeniable that decision making, decision effectiveness and decision support are among the hot research avenues in several business fields and particularly in the management field.

Accordingly, this dissertation is thought to contribute to the decision effectiveness in the healthcare industry which is one of the key sectors that influences the national economy performance. Clinical decisions are associated with resources allocation, financial costs, indirect and opportunity costs, and very importantly, wellbeing costs. Therefore, the opportunities that ML could provide to enhance the clinical decision effectiveness and success contribute not only to cost optimization but also in enhancing the overall national wellbeing which result in healthier communities and reduced expenditure on healthcare.

REFERENCES

1.Bellazzi R, Ferrazzi F, Sacchi L. Predictive Data Mining in Clinical Medicine: A Focus on Selected Methods and Applications. WIREs Data Mining and Knowledge Discovery. 2011;1:416-30.

2.Dixon-Woods M, Redwood S, Leslie M, Minion J, Martin G, Coleman J. Improving Quality and Safety of Care Using “Technovigilance”: An Ethnographic Case Study of Secondary Use of Data from an Electronic Prescribing and Decision Support System. Milbank Quarterly. 2013;91(3):424–54.

3.Bates D, Saria S, Ohno-Machado, L., Shah A, Escobar G. Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients. *Health Affairs*. 2014;7:1123–31.

4.Agarwal R, Gao G, DesRoches C, Jha A. The Digital Transformation of Healthcare: Current Status and the Road Ahead. Information Systems Research. 2010;21(4):796-809.

5.Chen H, Chiang R, Storey V. Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly. 2012;36(4):1165–88.

6.Krumholz H. Big Data And New Knowledge In Medicine: The Thinking , Training , And Tools Needed For A Learning Health System. Health Affairs. 2014;33(7):1163–70.

7.Wang Y, Hajli N. Exploring the Path to Big Data Analytics Success in Healthcare. Journal of Business Research. 2017;70:287–99.

8.Wang Y, Kung L, Byrd T. Big Data Analytics: Understanding its Capabilities and Potential Benefits for Healthcare Organizations. Technological Forecasting & Social Change. 2018;126:3–13.

9.He J, Baxter S, Xu J, Zhou X, Zhang K. The Practical Implementation of Artificial Intelligence Technologies in Medicine. Nature Medicine. 2019;25(1):30-6.

10.Olson D, Delen D. Advanced Data Mining Techniques. Berlin: Springer; 2008.

11.Croskerry P, Norman G. Overconfidence in Clinical Decision Making. The American Journal of Medicine. 2008;121(5):S24-S9.

12.Djulbegovic B, Elqayam S. Many Faces of Rationality: Implications of the Great Rationality Debate for Clinical Decision Making. Journal of Evaluation in Clinical Practice. 2017;23(5):915-22.

13.Djulbegovic B, Elqayam S, Dale W. Rational Decision Making in Medicine: Implications for Overuse and Underuse. Journal of Evaluation in Clinical Practice. 2018;24(3):655-65.

14.Berwick D, Hackbarth A. Eliminating Waste in US Health Care. The Journal of the American Medical Association. 2012;307(14):1513-6.

15.Macrotrends[Internet]. Qatar Healthcare Spending 2000-2020. 2020 [Available from: https://www.macrotrends.net/countries/QAT/qatar/healthcare-spending.

16.Dowie J. Clinical Decision Analysis: Background and Introduction. In: Llewelyn H, Hopkins A, editors. Analysing How We Reach Clinical Decisions. London: Royal College of Physicians 1993.

17.Karni E. A Theory of Medical Decision Making Under Uncertainty. Journal of Risk and Uncertainty. 2009;39(1):1-16.

18.Gladstone N. Comparative Theories in Clinical Decision Making and their Application to Practice: a Reflective Case Study. British Journal of

Anaesthetic and Recovery Nursing. 2012;13(3-4):65-71.

19.Shaban R. Theories of Clinical Judgment and Decision-Making: A Review of the Theoretical Literature. Australasian Journal of Paramedicine. 2015;3(1).

20.Marewski J, Gigerenzer G. Heuristic Decision Making in Medicine. Dialogues in Clinical Neuroscience. 2012;14(1):77-89.

21.Kaplan R, Frosch D. Decision Making in Medicine and Health Care. Annual Review of Clinical Psychology. 2005;1:525-56.

22.Standing M. Clinical Judgement and Decision-Making in Nursing and Interprofessional Healthcare Maidenhead- England: Open University Press-McGraw-Hill; 2010.

23.Elstein AS, Schwartz A. Clinical Problem Solving and Diagnostic Decision Making: Selective Review of the Cognitive Literature. British Medical Journal. 2002;324(7339):729-32.

24.Jefford E, Fahy K, Sundin D. Decision-Making Theories and Their Usefulness to the Midwifery Profession Both in Terms of Midwifery Practice and the Education of Midwives. International Journal of Nursing Practice. 2011;17(3):246-53.

25.Bekker H. Using Decision Making Theory to Inform Clinical Practice. In: Edwards A, Elwyn G, editors. Shared Decision Making in Healthcare: Achieving Evidence-based Patient Choice. 2nd ed: Oxford University Press; 2009. p. 45-52.

26.Muir N. Clinical Decision-Making: Theory and Practice. Nursing standard. 2004;18:47-52; quiz 4.

27.Hansson S. Decision Theory: A Brief Introduction. 2005. p. 1.

28.Baron J. Thinking and Deciding. 4th ed. New York-USA: Cambridge

University Press; 2000.

29.Bell D, Raiffa H, Tversky A. Descriptive, Normative, and Prescriptive Interactions in Decision Making. In: Tversky A, Bell DE, Raiffa H, editors. Decision Making: Descriptive, Normative, and Prescriptive Interactions. Cambridge: Cambridge University Press; 1988. p. 9-30.

30.McFall J. Rational, Normative, Descriptive, Prescriptive, or Choice Behavior? The Search for Integrative Metatheory of Decision Making. Behavioral Development Bulletin. 2015;20(1):45-59.

31.Mercier H, Sperber D. Why Do Humans Reason? Arguments for an Argumentative Theory. The Behavioral and brain sciences. 2011;34:57-74; discussion

32.Gigerenzer G, Brighton H. Homo Heuristicus: Why Biased Minds Make Better Inferences. Topics in Cognitive Science. 2009;1(1):107-43.

33.Schwartz B. What Does it Mean to be a Rational Decision Maker? Journal of Marketing Behavior. 2015;1(2):113-45.

34.Djulbegovic B, Guyatt G. Progress in Evidence-Based Medicine: A Quarter Century On. The Lancet. 2017;390(10092):415-23.

35.Bowen S. A Practical Model for Ethical Decision Making in Issues Management and Public Relations. Journal of Public Relations Research. 2005;17(3):191-216.

36.Norman G. Research in Clinical Reasoning: Past History and Current Trends. Medical Education. 2005;39(4):418-27.

37.Botti M, Reeve R. Role of Knowledge and Ability in Student Nurses' Clinical Decision-Making. Nursing & Health Sciences. 2003;5(1):39-49.

38.Djulbegovic B, Guyatt G, Ashcroft R. Epistemologic Inquiries in Evidence-

Based Medicine. Cancer Control. 2009;16(2):158-68.

39.Sackett D, Rosenberg W, Gray J, Haynes R, Richardson W. Evidence-Based Medicine: What It Is and What It Isn't. British Medical Journal. 1996;312(7023):71-2.

40.Greenhalgh T, Howick J, Maskrey N. Evidence Based Medicine: A Movement in Crisis? British Medical Journal. 2014;348:1-7.

41.Nevo I, Slonim-Nevo V. The Myth of Evidence-Based Practice:Towards Evidence-Informed Practice. British Journal of Social Work. 2011;41:1176–97.

42.Gambrill E. Evidence-Based Practice and Policy: Choices Ahead. Research on Social Work Practice. 2006;16(3):338–57.

43.Epstein I, Joubert L. Clinical Data Mining in the Age of Evidence-Based Practice: Recent Exemplars and Future Challenges. In: Syvajarvi, A., Stenvall J, editors. Data Mining in Public and Private Sectors: Organizational and Government Applications: IGI Global; 2010. p. 316 - 36.

44.Bellazzi R, Zupan B. Predictive Data Mining in Clinical Medicine: Current Issues and Guidelines. International Journal of Medical Informatics. 2008;77:81–97.

45.Goldenberg M. On Evidence and Evidence-Based Medicine: Lessons from the Philosophy of Science. Social Science & Medicine. 2006;62:2621–32.

46.Cook D, Guyatt G, Laupacis A, Sackett D, Goldberg R. Clinical Recommendations Using Levels of Evidence for Antithrombotic Agents. Chest. 1995;108(4):227-30.

47.Tomlin G, Borgetto B. Research Pyramid: A New Evidence-Based Practice Model for Occupational Therapy. American Journal of Occupational Therapy. 2011;65(2):189–96.

48.Toussi M, Lamy J, Le Toumelin P, Venot A. Using Data Mining Techniques to Explore Physicians' Therapeutic Decisions when Clinical Guidelines Do Not Provide Recommendations: Methods and Example for Type 2 Diabetes. BMC Medical Informatics and Decision Making. 2009;9(28):1-12.

49.Pagoto S, Spring B, Coups E, Mulvaney S, Coutu M, Ozakinci G. Barriers and Facilitators of Evidence-Based Practice Perceived by Behavioral Science Health Professionals. Journal of Clinical Psychology. 2007;63(7):695-705.

50.Walshe K, Rundall T. Evidence-Based Management: From Theory to Practice in Healthcare. The Milbank Quarterly. 2001;79(3):429-57.

51.Morris Z, Wooding S, Grant J. The Answer is 17 Years, What is the Question: Understanding Time Lags in Translational Research. Journal of the Royal Society of Medicine. 2011;104(12):510-20.

52.Westfall J, Mold J, Fagnan L. Practice-Based Research—“Blue Highways” on the NIH Roadmap. The Journal of the American Medical Association (JAMA). 2007;297(4):403-6.

53.McGlynn E, Asch S, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The Quality of Health Care Delivered to Adults in the United States. New England Journal of Medicine. 2003;348(26):2635-45.

54.Ebell M, Sokol R, Lee A, Simons C, Early J. How Good is the Evidence to Support Primary Care Practice? Evidence Based Medicine. 2017;22(3):88.

55.Bradley E, Herrin J, Mattera J, Holmboe E, Wang Y, Frederick P, et al. Quality Improvement Efforts and Hospital Performance: Rates of Beta-Blocker Prescription After Acute Myocardial Infarction. Medical Care. 2005;43(3):282-92.

56.Rath A, Olry A, Dhombres F, Brandt M, Urbero B, Ayme S. Representation

of Rare Diseases in Health Information Systems: The Orphanet Approach to Serve a Wide Range of End Users. Human Mutation. 2012;33:803–8.

57.Bourzac K. Participation: Power to the Patients. Nature. 2016;537:S66–S8.

58.Brownstein C, Brownstein J, Williams D, Wicks P, Heywood J. The Power of Social Networking in Medicine. *Nature Biotechnology*. 2009;*27*(10):888-90.

59.Steinhubl S, Muse E, Topol E. The Emerging Field of Mobile Health. Science Translational Medicine. 2015;7(283).

60.Beckmann J, Lew D. Reconciling Evidence-Based Medicine and Precision Medicine in the Era of Big Data: Challenges and Opportunities. Genome Medicine. 2016;8(134):1-11.

61.Epstein I. Reconciling Evidence-Based Practice, Evidence-Informed Practice, and Practice-Based Research: The Role of Clinical Data-Mining. Social Work. 2011;56(3):284-8.

62.Murdoch T, Detsky A. The Inevitable Application of Big Data to Health Care. Journal of American Medical Association. 2013;309(13):1351-2.

63.Saha S, Saint S, Christakis D. Impact factor: A Valid Measure of Journal Quality? Journal of the Medical Library Association. 2003;91(1):42-6.

64.Moed H. Measuring contextual citation impact of scientific journals. Journal of Informetrics. 2010;4(3):265–77.

65.Shameer K, Johnson K, Glicksberg B, Dudley J, Sengupta P. Machine Learning in Cardiovascular Medicine: Are We There Yet? Heart. 2018;104(14):1156.

66.Marka A, Carter J, Toto E, Hassanpour S. Automated Detection of Nonmelanoma Skin Cancer Using Digital Images: A Systematic review. BMC Medical Imaging. 2019;19(1):21.

67.Resteghini C, Trama A, Borgonovi E, Hosni H, Corrao G, Orlandi E, et al. Big Data in Head and Neck Cancer. Curr Treat Options Oncol. 2018;19(12):62.

68.Vogelius I, Petersen J, Bentzen S. Harnessing Data Science to Advance Radiation Oncology. Molecular Oncology. 2020.

69.Xu J, Yang P, Xue S, Sharma B, Sanchez-Martin M, Wang F, et al. Translating Cancer Genomics into Precision Medicine with Artificial Intelligence: Applications, Challenges and Future Perspectives. Human Genetics. 2019;138(2):109-24.

70.Sanchez-Pinto L, Luo Y, Churpek M. Big Data and Data Science in Critical Care. Chest. 2018;154(5):1239-48.

71.Mersha T, Afanador Y, Johansson E, Proper S, Bernstein J, Rothenberg M, et al. Resolving Clinical Phenotypes into Endotypes in Allergy: Molecular and Omics Approaches. Clinical Reviews in Allergy & Immunology. 2020.

72.Saglani S, Custovic A. Childhood Asthma: Advances Using Machine Learning and Mechanistic Studies. American Journal of Respiratory and Critical Care Medicine. 2018;199(4):414-22.

73.Kim Y, Na K. Application of Machine Learning Classification for Structural Brain MRI in Mood Disorders: Critical Review from a Clinical Perspective. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2018;80:71-80.

74.Six Dijkstra M, Siebrand E, Dorrestijn S, Salomons E, Reneman M, Oosterveld F, et al. Ethical Considerations of Using Machine Learning for Decision Support in Occupational Health: An Example Involving Periodic Workers' Health Assessments. Journal of Occupational Rehabilitation. 2020.

75.Cabitza F, Locoro A, Banfi G. Machine Learning in Orthopedics: A Literature Review. Frontiers in Bioengineering and Biotechnology. 2018;6:75.
76.Helm J, Swiergosz A, Haeberle H, Karnuta J, Schaffer J, Krebs V, et al. Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. Current reviews in musculoskeletal medicine. 2020;13(1):69-76.
77.Torous J, Larsen M, Depp C, Cosco T, Barnett I, Nock M, et al. Smartphones, Sensors, and Machine Learning to Advance Real-Time Prediction and Interventions for Suicide Prevention: a Review of Current Progress and Next Steps. Current Psychiatry Reports. 2018;20(7):51.
78.Pepin J, Bailly S, Tamisier R. Big Data in Sleep Apnoea: Opportunities and Challenges. Respirology. 2020;25(5):486-94.
79.Gossec L, Kedra J, Servy H, Pandit A, Stones S, Berenbaum F, et al. EULAR Points to Consider for the Use of Big Data in Rheumatic and Musculoskeletal Diseases. Annals of the Rheumatic Diseases. 2020;79(1):69-76.
80.Provost F, Fawcett T. Data Science for Business. 1st ed. Sebastopol, CA: O'Reilly Media, Inc.; 2013.
81.Samuel A. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development. 1959;3(3):210-29.
82.Saygin A, Cicekli I, Akman V. Turing Test: 50 Years Later. Minds and Machines. 2000;10(4):463-518.
83.Choi J, Shin K, Jung J, Bae H, Kim D, Byeon J, et al. Convolutional Neural Network Technology in Endoscopic Imaging: Artificial Intelligence for Endoscopy. Clinical endoscopy. 2020;53(2):117-26.

84.Marr B. A Short History of Machine Learning—Every Manager Should Read. Forbes. 2016.

85.Sagi O, Rokach L. Ensemble Learning: A Survey. WIREs Data Mining and Knowledge Discovery. 2018;8(4):e1249.

86.Cole J, Franke K. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. Trends in Neurosciences. 2017;40(12):681-90.

87.Deo R. Machine Learning in Medicine. Circulation. 2015;132(20):1920-30.

88.Zolbanin H, Delen D, Hassan Z. Predicting Overall Survivability in Comorbidity of Cancers: A Data Mining Approach. Decision Support Systems. 2015;74:150-61.

89.Lakshmi B, Indumathi T, Ravi N. A Study on C.5 Decision Tree Classification Algorithm for Risk Predictions During Pregnancy. Procedia Technology. 2016;24:1542-9.

90.Archer K, Kimes R. Empirical Characterization of Random Forest Variable Importance Measures. Computational Statistics & Data Analysis. 2008;52(4):2249-60.

91.Ahmed R, Temko A, Marnane W, Lightbody G, Boylan G. Grading Hypoxic–Ischemic Encephalopathy Severity in Neonatal EEG Using GMM Supervectors and the Support Vector Machine. Clinical Neurophysiology. 2016;127(1):297-309.

92.Dag A, Oztekin A, Yucel A, Bulur S, Megahed F. Predicting Heart Transplantation Outcomes through Data Analytics. Decision Support Systems. 2017;94:42–52.

93.Han J, Kamber M, Pie J. Data Mining: Concepts and Techniques. 3rd ed. San Francisco, CA , USA: Morgan Kaufmann Publishers Inc.; 2012.

94.Narula S, Khader S, Omar A, Sengupta P. Deep Learning With Unsupervised Feature in Echocardiographic Imaging. Journal of the American College of Cardiology. 2017;69:2101.

95.Shen D, Wu G, Suk H-I. Deep Learning in Medical Image Analysis. Annual review of biomedical engineering. 2017;19:221-48.

96.Jameson J, Longo D. Precision Medicine — Personalized, Problematic,and Promising. The New England Journal of Medicine. 2015;372(23):2229-34.

97.Behjati S, Tarpey P. What Is Next Generation Sequencing? Archives of Disease in Childhood Education and Practice Edition. 2013;98(6):236-8.

98.Chen M, Hao Y, Hwang K, Wang L. Disease Prediction by Machine Learning Over Big Data From Healthcare Communities. IEEE Access. 2017;5:8869-79.

99.Kuruvilla M, Lee F, Lee G. Understanding Asthma Phenotypes, Endotypes, and Mechanisms of Disease. Clinical reviews in allergy & immunology. 2019;56(2):219-33.

100.Kennedy H. The Importance of Randomized Clinical Trials and Evidence-Based Medicine: A Clinician's Perspective. Clinical cardiology. 1999;22(1):6-12.

101.Sullivan G. Getting Off the "Gold Standard": Randomized Controlled Trials and Education Research. Journal of Graduate Medical Education. 2011;3(3):285-9.

102.Stolberg H, Norman G, Trop I. Randomized Controlled Trials. American Journal of Roentgenology. 2004;183(6):1539-44.

103.Khozin S, Blumenthal G, Pazdur R. Real-world Data for Clinical Evidence

Generation in Oncology. JNCI: Journal of the National Cancer Institute. 2017;109(11).

104.Franklin J, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? Clinical Pharmacology & Therapeutics. 2017;102(6):924-33.

105.Gerstein H, McMurray J, Holman R. Real-world Studies No Substitute for RCTs in Establishing Efficacy. Lancet. 2019;393(10168):210-1.

106.May M. Twenty-Five Ways Clinical Trials Have Changed in the Last 25 Years. Nature Medicine. 25. United States2019. p. 2-5.

107.Baggerly K, Coombes K. Deriving Chemosensitivity from Cell Lines: Forensic Bioinformatics and Reproducible Research in High- hroughput Biology. Annals of Applied Statistics 2009;3(4):1309-34.

108.Ioannidis J, Allison D, Ball C, Coulibaly I, Xiangqin C, Culhane A, et al. Repeatability of Published Microarray Gene Expression Analyses. Nature Genetics. 2009;41(2):149-55.

109.Popper K. Conjectures and Refutations: The Growth of Scientific Knowledge. London: Routledge and Kegan Paul; 1963.

110.Fricke M. Big Data and Its Epistemology. Journal of the Association for Information Science and Technology. 2015;66(4):651–61.

111.Parkhurst J, Abeysinghe S. What Constitutes “Good” Evidence for Public Health and Social Policy-making? From Hierarchies to Appropriateness. Social Epistemology. 2016;30(5-6):665-79.

112.Roski J, Bo-Linn G, Andrews TA. Creating Value in Health Care Through Big Data: Opportunities and Policy Implications. *Health Affairs*. 2014;33(7):1115-22.

113.Karpatne A, Atluri G, Faghmous J, Steinbach M, Banerjee A, Ganguly A, et al. Theory-Guided Data Science: A New Paradigm for Scientific Discovery From Data. IEEE Transactions on Knowledge and Data Engineering. 2017;29(10):2318-31.

114.Scott I. Machine Learning and Evidence-Based Medicine. Ann Intern Med. 2018;169(1):44-6.

115.Kyriacou D. Evidence-Based Medical Decision Making: Deductive versus Inductive Logical Thinking. Academic Emergency Medicine. 2004;11(6):670-1.

116.Martini C. What “Evidence” in Evidence-Based Medicine? Topoi. 2020.

117.Kuhn T. The Structure of Scientific Revolutions. 2nd ed. USA: The University of Chicago Press; 1970. 210 p.

118.Burrel G, Morgan G. Sociological Paradigms
and Organisational
Analysis: Elements of the Sociology of Corporate Life. USA: Ashgate; 1979.

119.Abujaber A, Fadlalla A, Gammoh D, Abdelrahman H, Mollazehi M, El-Menyar A. Using Trauma Registry Data to Predict Prolonged Mechanical Ventilation in Patients with Traumatic Brain Injury: Machine Learning Approach. PLOS ONE. 2020;15(7):e0235231.

120.Andriolo B, Andriolo R, Saconato H, Atallah A, Valente O. Early Versus Late Tracheostomy for Critically Ill Patients. Cochrane Database for Systematic Reviews. 2015;1:Cd007271.

121.Piano S. Ethical Principles in Machine Learning and Artificial Intelligence: Cases from the Field and Possible Ways Forward. Humanities and Social Sciences Communications. 2020;7(1):9.

122.Institute TMG. Global AI Survey: AI Proves its Worth, But Few Scale

Impact. 2019.

123.Watson R. Interpretive Structural Modeling—A Useful Tool for Technology Assessment? Technological Forecasting and Social Change. 1978;11(2):165-85.

124.Sushil. Interpreting the Interpretive Structural Model. Global Journal of Flexible Systems Management. 2012;13(2):87-106.

125.Ramesh A, Banwet D, Shankar R. Modeling the Barriers of Supply Chain Collaboration. Journal of Modelling in Management. 2010.

126.Haleem A, Sushil, Qadri M, Kumar S. Analysis of Critical Success Factors of World-Class Manufacturing Practices: An Application of Interpretative Structural Modelling and Interpretative Ranking Process. Production Planning & Control. 2012;23(10-11):722-34.

127.Chaple A, Narkhede B, Akarte M, Raut R. Modeling the Lean Barriers for Successful Lean Implementation: TISM Approach. International Journal of Lean Six Sigma. 2018.

128.Biswas P. Modeling Reconfigurability in Supply Chains Using Total Interpretive Structural Modeling. Journal of Advances in Management Research. 2017.

129.Mohanty M. Assessing Sustainable Supply Chain Enablers Using Total Interpretive Structural Modeling Approach and Fuzzy-MICMAC Analysis. Management of Environmental Quality: An International Journal. 2018.

130.Singh A. Modeling Enablers of TQM to Improve Airline Performance. International Journal of Productivity and Performance Management. 2013.

131.Singh A. Developing a Conceptual Framework of Waste Management in the Organizational Context. Management of Environmental Quality: An

International Journal. 2017.

132.Bolanos R, Fontela E, Nenclares A, Pastor P. Using Interpretive Structural Modelling in Strategic Decision-Making Groups. Management Decision. 2005.

133.Ajmera P, Jain V. Modelling the Barriers of Health 4.0–the Fourth Healthcare Industrial Revolution in India by TISM. Operations Management Research. 2019;12(3):129-45.

134.Jena J, Sidharth S, Thakur L, Pathak D, Pandey V. Total Interpretive Structural Modeling (TISM): Approach and Application. Journal of Advances in Management Research. 2017.

135.Deshmukh A, Mohan A. Analysis of Indian Retail Demand Chain Using Total Interpretive Modeling. Journal of Modelling in Management. 2017.

136.Nasim S. Total Interpretive Structural Modeling of Continuity and Change Forces in e-Government. Journal of Enterprise Transformation. 2011;1(2):147-68.

137.Ajzen I. The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes. 1991;50(2):179-211.

138.Wang W. The Influence of Perceived Technological Congruence of Smartphone Application and Air Travel Experience on Consumers' Attitudes Toward Price Change and Adoption. Journal of Hospitality and Tourism Technology. 2019;10(2):122-35.

139.Rogers E. Diffusion of Innovations. 3rd ed. New York - USA: Macmillan Publishing Co., Inc; 1983.

140.Oh S, Kim J, Choi S, Lee H, Hong J, Kwon S. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. Journal of Medical Internet

Research. 2019;21(3):e12422.

141.Kelly C, Karthikesalingam A, Suleyman M, Corrado G, King D. Key Challenges for Delivering Clinical Impact with Artificial Intelligence. BMC Medicine. 2019;17(1):195.

142.Kamal S, Shafiq M, Kakria P. Investigating Acceptance of Telemedicine Services Through an Extended Technology Acceptance Model (TAM). Technology in Society. 2020;60:101212.

143.Kowitlawakul Y. The Technology Acceptance Model: Predicting Nurses' Intention to Use Telemedicine Technology (eICU). Computers, Informatics, Nursing. 2011;29(7).

144.Gusmano M, Maschke K, Solomon M. Patient-Centered Care, Yes; Patients As Consumers, No. Health Affairs. 2019;38(3):368-73.

145.Jutzi T, Krieghoff-Henning E, Holland-Letz T, Utikal J, Hauschild A, Schadendorf D, et al. Artificial Intelligence in Skin Cancer Diagnostics: The Patients' Perspective. Frontiers in Medicine. 2020;7:233.

146.Jaiswal M, Gupta N, Singh A. Study on Patient's Awareness Towards Role of Artificial Intelligence in Dentistry. 2019;9:35-9.

147.Davis F. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly. 1989;13(3):319-40.

148.Alhashmi S, Salloum S, Mhamdi C. Implementing Artificial Intelligence in the United Arab Emirates Healthcare Sector: An Extended Technology Acceptance Model. International Journal of Information Technology and Language Studies. 2019;3(3):27-42.

149.Shahbaz M, Gao C, Zhai L, Shahzad F, Hu Y. Investigating the Adoption of Big Data Analytics in Healthcare: The Moderating Role of Resistance to

Change. Journal of Big Data. 2019;6(1):6.

150.Zihni E, Madai V, Livne M, Galinovic I, Khalil A, Fiebach J, et al. Opening the Black Box of Artificial Intelligence for Clinical Decision Support: A Study Predicting Stroke Outcome. Plos one. 2020;15(4):e0231166.

151.Baselli G, Codari M, Sardanelli F. Opening the Black Box of Machine Learning in Radiology: Can the Proximity of Annotated Cases be a Way? European Radiology Experimental. 2020;4(1):30.

152.Ajami S, Bagheri-Tadi T. Barriers for Adopting Electronic Health Records (EHRs) by Physicians. Acta Inform Med. 2013;21(2):129-34.

153.Sapci A, Sapci H. Artificial Intelligence Education and Tools for Medical and Health Informatics Students: Systematic Review. JMIR Medical Education. 2020;6(1):e19285.

154.Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing Artificial Intelligence Training in Medical Education. JMIR Med Educ. 2019;5(2):e16048.

155.Wartman S, Combs C. Reimagining Medical Education in the Age of AI. AMA journal of ethics. 2019;21(2):146-52.

156.Diprose W, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician Understanding, Explainability, and Trust in a Hypothetical Machine Learning Risk Calculator. Journal of the American Medical Informatics Association. 2020;27(4):592-600.

157.Fang Y, Qureshi I, Sun H, McCole P, Ramsey E, Lim K. Trust, Satisfaction, and Online Repurchase Intention. Mis Quarterly. 2014;38(2):407-A9.

158.Kim K, Prabhakar B, Park S. Trust, Perceived Risk, and trusting Behavior

in Internet Banking. Asia Pacific Journal of Information Systems. 2009;19(3):1-23.

159.Cohen I, Amarasingham R, Shah A, Xie B, Lo B. Big Data And New Knowledge In Medicine: The Thinking , Training , And Tools Needed For A Learning Health System. Health Affairs. 2014;33(7):1139-47.

160.Cohen I. Informed Consent and Medical Artificial Intelligence: What to Tell the Patient? The Georgetown Law Journal. 2019;108:1425.

161.Holden R, Ben-Tzion K. The Technology Acceptance Model: Its Past and its Future in Health Care. Journal of Biomedical Informatics. 2010;43(1):159-72.

162.Gucin N, Berk O. Technology Acceptance in Health Care: An Integrative Review of Predictive Factors and Intervention Programs. Procedia - Social and Behavioral Sciences. 2015;195:1698-704.

163.Hsiao J, Chen R. Critical Factors Influencing Physicians' Intention to Use Computerized Clinical Practice Guidelines: An Integrative Model of Activity Theory and the Technology Acceptance Model. BMC Medical Informatics and Decision Making. 2016;16(1):3.

164.Pavlou P, Fygenson M. Understanding and Prediction Electronic Commerce Adoption: An Extension of the Theory of Planned Behavior. MIS Quarterly. 2006;30(1):115-43.

165.Garcia-Laencina P, Abreu P, Abreu M, Afonoso N. Missing Data Imputation on the 5-Year Survival Prediction of Breast Cancer Patients with Unknown Discrete Values. Computers in Biology and Medicine. 2015;59:125–33.

166.Rau C, Kuo P, Chien P, Huang C, Hsieh H, Hsieh C. Mortality Prediction

in Patients with Isolated Moderate and Severe Traumatic Brain Injury Using Machine Learning Models. PLOS ONE. 2018;13(11):e0207192.

167.Walker A, Mason A, Quan T, Fawcett N, Watkinson P, Llewelyn M, et al. Mortality Risks Associated with Emergency Admissions During Weekends and Public Holidays: An Analysis of Electronic Health Records. Lancet. 2017;390(10089):62-72.

168.Han L, Sutton M, Clough S, Warner R, Doran T. Impact of Out-of-Hours Admission on Patient Mortality: Longitudinal Analysis in a Tertiary Acute Hospital. BMJ Quality & Safety. 2018;27(6):445.

169.Roh Y, Kim H, Cha Y, Cha K, Kim H, Lee K, et al. Mortality Reduction in Major Trauma Patients after Establishment of a Level I Trauma Center in Korea:A Single-Center Experience. Journal of Trauma and Injury. 2017;30(4):131-9.

170.Andelic N, Anke A, Skandsen T, Sigurdardottir S, Sandhaug M, Ader T, et al. Incidence of Hospital-Admitted Severe Traumatic Brain Injury and In-Hospital Fatality in Norway: A National Cohort Study. Neuroepidemiology. 2012;38:259–67.

171.Dewan M, Rattani A, Gupta S, Baticulon R, Hung YC, Punchak M, et al. Estimating the Global Incidence of Traumatic Brain Injury. Journal of Neurosurgery. 2018;130(4):1-18.

172.Rached M, Gaudet J, Delhumeau C, Walder B. Comparison of Two Simple Models for Prediction of Short Term Mortality in Patients after Severe Traumatic Brain Injury. Injury. 2019;50:65–72.

173.Walder B, Haller G, Rebetez M, Delhumeau C, Bottequin E, Schoettker P, et al. Severe Traumatic Brain Injury in a High-Income Country: An

Epidemiological Study. *Journal of Neurotrauma*. 2013;30(23):1934–42.

174.Jacobs B, Beems T, van der Vliet T, van Vugt A, Hoedemaekers C, Horn J, et al. Outcome Prediction in Moderate and Severe Traumatic Brain Injury: A Focus on Computed Tomography Variables. Neurocritical Care. 2013;19(1):79-89.

175.Eftekhar B, Mohammad K, Ardebili H, Ghodsi M, Ketabchi E. Comparison of Artificial Neural Network and Logistic Regression Models for Prediction of Mortality in Head Trauma Based on Initial Clinical Data. BMC Medical Informatics and Decision Making. 2005;5:3.

176.Chen C, Shi H, Lee K, Huang T. In-Hospital Mortality Prediction in Patients Receiving Mechanical Ventilation in Taiwan. American Journal of Critical Care. 2013;22(6):506-13.

177.Senders J, Staples P, Karhade A, Zaki M, Gormley W, Broekman M, et al. Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review. World Neurosurgery. 2018;109:476-86.e1.

178.Hale A, Stonko D, Brown A, Lim J, Voce D, Gannon S, et al. Machine-Learning Analysis Outperforms Conventional Statistical Models and CT Classification Systems in Predicting 6-month Outcomes in Pediatric Patients Sustaining Traumatic Brain Injury. Neurosurgical Focus. 2018;45(5):E2.

179.Domingues C, Coimbra R, Poggetti R, Nogueira L, de Sousa R. New Trauma and Injury Severity Score (TRISS) Adjustments for Survival Prediction. World Journal of Emergency Surgery. 2018;13:12-.

180.Maeda Y, Ichikawa R, Misawa J, Shibuya A, Hishiki T, Maeda T, et al. External Validation of the TRISS, CRASH, and IMPACT Prognostic Models in Severe Traumatic Brain Injury in Japan. PloS one. 2019;14(8):e0221791-e.

181.Wong G, Teoh J, Yeung J, Chan E, Siu E, Woo P, et al. Outcomes of Traumatic Brain Injury in Hong Kong: Validation with the TRISS, CRASH, and IMPACT Models. Journal of Clinical Neuroscience. 2013;20(12):1693-6.

182.Han J, King N, Neilson S, Gandhi M, Ng I. External Validation of the CRASH and IMPACT Prognostic Models in Severe Traumatic Brain Injury. Journal of Neurotrauma. 2014;31(13):1146–52.

183.Carter E, Hutchinson P, Kolias A, Menon D. Predicting the Outcome for Individual Patients with Traumatic Brain Injury: A Case-Based Review. British Journal of Neurosurgery. 2016;30(2):227-32.

184.Reith F, Van den Brande R, Synnot A, Gruen R, Maas A. The Reliability of the Glasgow Coma Scale: A Systematic Review. Intensive Care Medicine. 2016;42(1):3-15.

185.Maas A, Hukkelhoven C, Marshall L, Steyerberg E. Prediction of Outcome in Traumatic Brain Injury with Computed Tomographic Characteristics: A Comparison between the Computed Tomographic Classification and Combinations of Computed Tomographic Predictors. Neurosurgery. 2005;57(6):1173-82.

186.Raj R, Siironen J, Skrifvars M, Hernesniemi J, Kivisaari R. Predicting Outcome in Traumatic Brain Injury: Development of a Novel Computerized Tomography Classification System (Helsinki Computerized Tomography Score). Neurosurgery. 2014;75(6):632–47.

187.Liesemer K, Riva-Cambrin J, Bennett KS, Bratton S, Tran H, Metzger R, et al. Use of Rotterdam CT Scores for Mortality Risk Stratification in Children with Traumatic Brain Injury. Pediatric Critical Care Medicine. 2014;15(6):554-62.

188.Das A, Ben-Menachem T, Cooper G, Chak A, Sivak M, Gonet J, et al. Prediction of Outcome in Acute Lower-Gastrointestinal Haemorrhage Based on an Artificial Neural Network: Internal and External Validation of a Predictive Model. Lancet. 2003;362(9392):1261-6.

189.Schumacher P, Olinsky A, Quinn J, Smith R. A Comparison of Logistic Regression, Neural Networks, and Classification Trees Predicting Success of Actuarial Students. Journal of Education for Business. 2010;85(5):258-63.

190.Cui S, Wang D, Wang Y, Yu P, Jin Y. An Improved Support Vector Machine-Base Diabetic Readmission Prediction. Computer Methods and Programs in Biomedicine. 2018;166:123–35.

191.Roozenbeek B, Chiu Y, Lingsma H, Gerber L, Steyerberg E, Ghajar J, et al. Journal of Neurotrauma. 2012;29:1306–12.

192.Mushkudiani N, Hukkelhoven C, Hernández A, Murray G, Choi S, Maas A, et al. A Systematic Review Finds Methodological Improvements Necessary for Prognostic Models in Determining Traumatic Brain Injury Outcomes. Journal of Clinical Epidemiology. 2008;61(4):331-43.

193.Teasdale G, Jennett B. Assessment of coma and impaired consciousness: A practical scale. Lancet. 1974;304(7872):81-4.

194.Jovanovic B, Milan Z, Markovic-Denic L, Djuric O, Radinovic K, Doklestic K, et al. Risk Factors for Ventilator-Associated Pneumonia in Patients with Severe Traumatic Brain Injury in a Serbian Trauma Centre. International Journal of Infectious Diseases. 2015;38:46-51.

195.Boutin A, Moore L, Lauzier F, Chassé M, English S, Zarychanski R, et al. Transfusion of Red Blood Cells in Patients with Traumatic Brain Injuries Admitted to Canadian Trauma Health Centres: A Multicentre Cohort Study.

BMJ Open. 2017;7(3):e014472.

196.East JM, Viau-Lapointe J, McCredie VA. Transfusion practices in traumatic brain injury. Current opinion in anaesthesiology. 2018;31(2):219-26.

197.Mohseni S, Talving P, Lam L, Chan L, Ives C, Demetriades D. Venous Thromboembolic Events in Isolated Severe Traumatic Brain Injury. Journal of Emergencies, Trauma, and Shock. 2012;5(1):11-5.

198.Jha R, Kochanek P, Simard J. Pathophysiology and Treatment of Cerebral Edema in Traumatic Brain Injury. Neuropharmacology. 2019;145:230-46.

199.Unterberg A, Stover J, Kress B, Kiening K. Edema and Brain Trauma. Neuroscience. 2004;129(4):1021-9.

200.Bratton S, Chestnut R, Ghajar J, McConnell Hammond F, Harris O, Hartl R, et al. Guidelines for the Management of Severe Traumatic Brain Injury. I. Blood Pressure and Oxygenation. Journal of Neurotrauma. 2007;24 Suppl 1:7-13.

201.Bartels R, Meijer F, van der Hoeven H, Edwards M, Prokop M. Midline Shift in Relation to Thickness of Traumatic Acute Subdural Hematoma Predicts Mortality. BMC Neurology. 2015;15(1):220.

202.Inoue A, Hifumi T, Kuroda Y, Nishimoto N, Kawakita K, Yamashita S, et al. Mild Decrease in Heart Rate During Early Phase of Targeted Temperature Management Following Tachycardia on Admission is Associated with Unfavorable Neurological Outcomes After Severe Traumatic Brain Injury: A Post Hoc Analysis of a Multicenter Randomized Controlled Trial. Critical Care. 2018;22(1):352.

203.Ramesh V, Thirumaran K, Raja M. A New Scale for Prognostication in

Head Injury. Journal of Clinical Neuroscience. 2008;15(10):1110-3.

204.Anselmi L, Meacock R, Kristensen S, Doran T, Sutton M. Arrival by Ambulance Explains Variation in Mortality by Time of Admission: Retrospective Study of Admissions to Hospital Following Emergency Department Attendance in England. BMJ Quality and Safety. 2017;26(8):613.

205.Aljerian N, Alhaidar S, Alothman A, AlJohi W, Albaqami F, Alghnam S. Association Between the Mode of Transport and In-Hospital Medical Complications in Trauma Patients: Findings from a Level-I Trauma Center in Saudi Arabia. Annals of Saudi medicine. 2018;38(1):8-14.

206.Esteban A, Anzueto A, Frutos F, Alía I, Brochard L, Stewart T, et al. Characteristics and Outcomes in Adult Patients Receiving Mechanical VentilationA 28-Day International Study. The Journal of the American Medical Association. 2002;287(3):345-55.

207.Asehnoune K, Roquilly A, Cinotti R. Respiratory Management in Patients with Severe Brain Injury. Critical care. 2018;22(1):76-.

208.Wang C, Lin H, Chang Y, Maa S, Wang J, Tang W, et al. Predictive Factors of In-Hospital Mortality in Ventilated Intensive Care Unit. Medicine. 2017;96(51):e9165-e.

209.Ayer T, Chhatwal J, Alagoz O, Kahn C, Woods R, Burnside E. Comparison of Logistic Regression and Artificial Neural Network Models in Breast Cancer Risk Estimation. RadioGraphics. 2010;30(1):13-22.

210.Dreiseitl S, Ohno-Machado L. Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. Journal of Biomedical Informatics. 2002;35(5):352-9.

211.Knudson M, Ikossi D, Khaw L, Morabito D, Speetzen L.

Thromboembolism After Trauma: An Analysis of 1602 Episodes from the American College of Surgeons National Trauma Data Bank. Annals of Surgery. 2004;240(3):490-6; discussion 6-8.

212.Geerts W, Code K, Jay R, Chen E, Szalai J. A Prospective Study of Venous Thromboembolism After Major Trauma. New England Journal of Medicine. 1994;331(24):1601-6.

213.Margolick J, Dandurand C, Duncan K, Chen W, Evans D, Sekhon M, et al. A Systematic Review of the Risks and Benefits of Venous Thromboembolism Prophylaxis in Traumatic Brain Injury. Canadian Journal of Neurological Sciences. 2018;45(4):432-44.

214.Nathens A, McMurray M, Cuschieri J, Durr E, Moore E, Bankey P, et al. The Practice of Venous Thromboembolism Prophylaxis in the Major Trauma Patient. Journal of Trauma and Acute Care Surgery. 2007;62(3):557-62; discussion 62-3.

215.Humble S, Wilson L, Wang L, Long D, Smith M, Siktberg J, et al. Prognosis of Diffuse Axonal Injury with Traumatic Brain Injury. The Journal of Trauma and Acute Care Surgery. 2018;85(1):155-9.

216.Cinotti R, Bouras M, Roquilly A, Asehnoune K. Management and Weaning from Mechanical Ventilation in Neurologic Patients. Annals of Translational Medicine. 2018;6(19):1-5.

217.Boles J, Bion J, Connors A, Herridge M, Marsh B, Melot C, et al. Weaning from Mechanical Ventilation. European Respiratory Journal. 2007;29(5):1033.

218.Clark P, Lettieri C. Clinical Model for Predicting Prolonged Mechanical Ventilation. Journal of Critical Care. 2013;28:880.e1–.e7.

219.Esteban A, Frutos-Vivar F, Muriel A, Ferguson N, Penuelas O, Abraira V,

et al. Evolution of Mortality Over Time in Patients Receiving Mechanical Ventilation. American Journal of Respiratory and Critical Care Medicine. 2013;188(2):220-30.

220.Parreco J, Hidalgo A, Parks J, Kozol R, Rattan R. Using Artificial Intelligence to Predict Prolonged Mechanical Ventilation and Tracheostomy Placement. Journal of Surgical Research. 2018;228:179-87.

221.Zilberberg M, de Wit M, Shorr A. Accuracy of Previous Estimates for Adult Prolonged Acute Mechanical Ventilation Volume in 2020: Update Using 2000-2008 Data. Critical Care Medicine. 2012;40(1):18-20.

222.Girard T, Kress J, Fuchs B, Thomason J, Schweickert W, Pun B, et al. Efficacy and Safety of a Paired Sedation and Ventilator Weaning Protocol for Mechanically Ventilated Patients in Intensive Care (Awakening and Breathing Controlled trial): A Randomised Controlled Trial. Lancet. 2008;371(9607):126-34.

223.Luetz A, Goldmann A, Weber-Carstens S, Spies C. Weaning from Mechanical Ventilation and Sedation. Current Opinion in Anaesthesiology. 2012;25(2):164-9.

224.Agle S, Kao L, Moore F, Gonzalez E, Vercruysse G, Todd R. Early Predictors of Prolonged Mechanical Ventilation in Major Torso Trauma Patients who Require Resuscitation. The American Journal of Surgery. 2006;192:822–7.

225.Dimopoulou I, Anthi A, Lignos M, Boukouvalas E, Evangelou E, Routsi C, et al. Prediction of Prolonged Ventilatory Support in Blunt Thoracic Trauma Patients. Intensive Care Medicine. 2003;29:1101–5.

226.Figueroa-Casas J, Dwivedi A, Connery S, Quansah R, Ellerbrook L,

Galvis J. Predictive Models of Prolonged Mechanical Ventilation Yield Moderate Accuracy. Journal of Critical Care. 2015;30:502-5.

227.Rose L, McGinlay M, Amin R, Burns K, Connolly B, Hart N, et al. Variation in Definition of Prolonged Mechanical Ventilation. Respiratory Care. 2017;62(10):1324.

228.Young D, Harrison D, Cuthbertson B, Rowan K. Effect of Early vs Late Tracheostomy Placement on Survival in Patients Receiving Mechanical Ventilation: The TracMan Randomized Trial. The Journal of the American Medical Association. 2013;309(20):2121-9.

229.Herritt B, Chaudhuri D, Thavorn K, Kubelik D, Kyeremanteng K. Early vs. late tracheostomy in intensive care settings: Impact on ICU and hospital costs. J Crit Care. 2018;44:285-8.

230.Pandian V, Bose S, Miller C, Schiavi A, Feller-Kopman D, Bhatti N, et al. Exploring Quality of Life in Critically Ill Tracheostomy Patients: A Pilot Study. ORL Head & Neck Nursing. 2014;32(1):6-8, 10-3.

231.Chang Y, Huang K, Chen Y, Wang C, Wang Y, Tseng C, et al. Ventilator Dependence Risk Score for the Prediction of Prolonged Mechanical Ventilation in Patients Who Survive Sepsis/Septic Shock with Respiratory Failure. Scientific reports. 2018;8(1):5650-.

232.Shaikhina T, Lowe D, Daga S, Briggs D, Higgins R, Khovanova N. Decision Tree and Random Forest Models for Outcome Prediction in Antibody Incompatible Kidney Transplantation. Biomedical Signal Processing and Control. 2019;52:456-62.

233.Hellyer T, Ewan V, Wilson P, Simpson A. The Intensive Care Society Recommended Bundle of Interventions for the Prevention of Ventilator-

Associated Pneumonia. Journal of the Intensive Care Society. 2016;17(3):238-43.

234.Ghiani A, Sainis A, Sainis G, Neurohr C. Anemia and Red Blood Cell Transfusion Practice in Prolonged Mechanically Ventilated Patients Admitted to a Specialized Weaning Center: An Observational Study. BMC Pulmonary Medicine. 2019;19(1):250.

235.Lai Y, Ruan S, Huang C, Kuo P, Yu C. Hemoglobin Levels and Weaning Outcome of Mechanical Ventilation in Difficult-to-Wean Patients: A Retrospective Cohort Study. PLoS One. 2013;8(8):e73743.

236.Zubrow M, Thomas N, Friedman D, Yehya N. RBC Transfusions Are Associated With Prolonged Mechanical Ventilation in Pediatric Acute Respiratory Distress Syndrome. Pediatric Critical Care Medicine. 2018;19(2):e88-e96.

237.Mena J, Sanchez A, Rubiano A, Peitzman A, Sperry J, Gutierrez M, et al. Effect of the Modified Glasgow Coma Scale Score Criteria for Mild Traumatic Brain Injury on Mortality Prediction: Comparing Classic and Modified Glasgow Coma Scale Score Model Scores of 13. The Journal of Trauma. 2011;71(5):1185-93.

238.Vieira R, Paiva W, de Oliveira D, Teixeira M, de Andrade A, de Sousa R. Diffuse Axonal Injury: Epidemiology, Outcome and Associated Risk Factors. Frontiers in Neurology. 2016;7:178.

239.Okabe Y. Risk Factors for Prolonged Mechanical Ventilation in Patients with Severe Multiple Injuries and Blunt Chest Trauma: A Single Center Retrospective Case-Control Study. Acute Medicine and Surgery. 2018;5(2):166-72.

240.Kim H, Park S. Sepsis: Early Recognition and Optimized Treatment. Tuberculosis and Respiratory Diseases. 2019;82(1):6-14.
241.Kao J, Kao H, Chen Y, Yu W, Pan S, Wang J, et al. Impact and Predictors of Prolonged Chest Tube Duration in Mechanically Ventilated Patients with Acquired Pneumothorax. Respiratory Care. 2013;58(12):2093.
242.Jacobs B, Beems T, van der Vliet T, Diaz-Arrastia R, Borm G, Vos P. Computed Tomography and Outcome in Moderate and Severe Traumatic Brain Injury: Hematoma Volume and Midline Shift Revisited. Journal of Neurotrauma. 2011;28(2):203-15.
243.Blaser A, Regli A, De Keulenaer B, Kimball E, Starkopf L, Davis W, et al. Incidence, Risk Factors, and Outcomes of Intra-Abdominal Hypertension in Critically Ill Patients-A Prospective Multicenter Study (IROI Study). Critical care medicine. 2019;47(4):535-42.
244.Trouillet J, Luyt C, Guiguet M, Ouattara A, Vaissier E, Makri R, et al. Early Percutaneous Tracheotomy versus Prolonged Intubation of Mechanically Ventilated Patients After Cardiac Surgery: A Randomized Trial. Annals of Internal Medicine. 2011;154(6):373-83.
245.Gershengorn H, Garland A, Gong M. Patterns of Daily Costs Differ for Medical and Surgical Intensive Care Unit Patients. Annals of the American Thoracic Society. 2015;12(12):1831-6.
246.Dasta J, McLaughlin T, Mody S, Piech C. Daily Cost of an Intensive Care Unit Day: The Contribution of Mechanical Ventilation. Critical Care Medicine. 2005;33(6):1266-71.
247.Huttmann S, Magnet F, Karagiannidis C, Storre J, Windisch W. Quality of Life and Life Satisfaction are Severely Impaired in Patients with Long-Term

Invasive Ventilation Following ICU Treatment and Unsuccessful Weaning. Annals of Intensive Care. 2018;8(1):38.

248.Kourbeti I, Vakis A, Papadakis J, Karabetsos D, Bertsias G, Filippou M, et al. Infections in Traumatic Brain Injury Patients. Clinical Microbiology and Infection. 2012;18:359–64.

249.Kalanuria A, Zai W, Mirski M. Ventilator-Associated Pneumonia in the ICU. Critical Care. 2014;18(2):208.

250.Guidelines for the Management of Adults with Hospital-Acquired, Ventilator-Associated, and Healthcare-Associated Pneumonia. American Journal of Respiratory and Critical Care Medicine. 2005;171(4):388-416.

251.Chastre J, Fagon J. Ventilator-Associated Pneumonia. Am J Respir Crit Care Med. 2002;165(7):867-903.

252.Melsen W, Rovers M, Groenwold R, Bergmans D, Camus C, Bauer T, et al. Attributable Mortality of Ventilator-Associated Pneumonia: A Meta-Analysis of Individual Patient Data from Randomised Prevention Studies. Lancet Infectious Diseases. 2013;13(8):665-71.

253.Larsson J, Itenov T, Bestle M. Risk Prediction Models for Mortality in Patients with Ventilator-Associated Pneumonia: A Systematic Review and Meta-Analysis. Journal of Critical Care. 2017;37:112-8.

254.Dudeck MA, Weiner LM, Allen-Bridson K, Malpiedi PJ, Peterson KD, Pollock DA, et al. National Healthcare Safety Network (NHSN) report, data summary for 2012, Device-associated module. American Journal of Infection Control. 2013;41(12):1148-66.

255.Arumugam S, Mudali I, Strandvik G, El-Menyar A, Al-Hassani A, Al-Thani H. Risk Factors for Ventilator-Associated Pneumonia in Trauma Patients: A

Descriptive Analysis. World Journal Emergency Medicine. 2018;9(3):203-10.
256.Zygun D, Zuege D, Boiteau P, Laupland K, Henderson E, Kortbeek J, et al. Ventilator-Associated Pneumonia in Severe Traumatic Brain Injury. Neurocritical Care. 2006;5:108–14.
257.Bronchard R, Albaladejo P, Brezac G, Geffroy A, Seince P, Morris W, et al. Early Onset Pneumonia: Risk Factors and Consequences in Head Trauma Patients. Anesthesiology: The Journal of the American Society of Anesthesiologists. 2004;100(2):234-9.
258.Esnault P, Nguyen C, Bordes J, D'Aranda E, Montcriol A, Contargyris C, et al. Early-Onset Ventilator-Associated Pneumonia in Patients with Severe Traumatic Brain Injury: Incidence, Risk Factors, and Consequences in Cerebral Oxygenation and Outcome. Neurocrit Care. 2017;27(2):187-98.
259.Zimlichman E, Henderson D, Tamir O, Franz C, Song P, Yamin C, et al. Health Care-Associated Infections: A Meta-Analysis of Costs and Financial Impact on the US Health Care System. JAMA Internal Medicine. 2013;173(22):2039-46.
260.Kollef M, Hamilton C, Ernst F. Economic Impact of Ventilator-Associated Pneumonia in a Large Matched Cohort. Infection Control & Hospital Epidemiology. 2012;33(3):250-6.
261.Luckraz H, Manga N, Senanayake E, Abdelaziz M, Gopal S, Charman S, et al. Cost of Treating Ventilator-Associated Pneumonia Post Cardiac Surgery in the National Health Service: Results from a Propensity-Matched Cohort Study. Journal of the Intensive Care Society. 2018;19(2):94-100.
262.Ahmadi E, Weckman G, Masel D. Decision Making Model to Predict Presence of Coronary Artery Disease Using Neural Network and C5.0

Decision Tree. Journal of Ambient Intelligence and Humanized Computing. 2018;9(4):999-1011.

263.Wiharto W, Kusnanto H, Herianto H. Interpretation of Clinical Data Based on C4.5 Algorithm for the Diagnosis of Coronary Heart Disease. Healthcare informatics research. 2016;22(3):186-95.

264.Mangram A, Sohn J, Zhou N, Hollingworth A, Ali-Osman F, Sucher J, et al. Trauma-Associated Pneumonia: Time to Redefine Ventilator-Associated Pneumonia in Trauma Patients. The American Journal of Surgery. 2015;210(6):1056-61; discussion 61-2.

265.Park H, Kang D, Moon S, Yang J, Kim S, Byun J. Risk Factors for Pneumonia in Ventilated Trauma Patients with Multiple Rib Fractures. The Korean journal of thoracic and cardiovascular surgery. 2017;50(5):346-54.

266.Cavalcanti M, Ferrer M, Ferrer R, Morforte R, Garnacho A, Torres A. Risk and Prognostic Factors of Ventilator-Associated Pneumonia in Trauma Patients. Critical Care Medicine. 2006;34(4).

267.Bochicchio G, Napolitano L, Joshi M, Bochicchio K, Shih D, Meyer W, et al. Blood Product Transfusion and Ventilator-Associated Pneumonia in Trauma Patients. Surgical Infections. 2008;9(4):415-22.

268.Shorr A, Duh M, Kelly K, Kollef M. Red Blood Cell Transfusion and Ventilator-Associated Pneumonia: A Potential Link? Critical Care Medicine. 2004;32(3):666-74.

269.Younan D, Delozier S, Adamski J, Loudon A, Violette A, Ustin J, et al. Factors Predictive of Ventilator-associated Pneumonia in Critically Ill Trauma Patients. World Journal of Surgery. 2019.

270.Wu D, Wu C, Zhang S, Zhong Y. Risk Factors of Ventilator-Associated

Pneumonia in Critically III Patients. Frontiers in pharmacology. 2019;10:482-.
271.Apostolopoulou E, Bakakos P, Katostaras T, Gregorakos L. Incidence and Risk Factors for Ventilator-Associated Pneumonia in 4 Multidisciplinary Intensive Care Units in Athens, Greece. Respiratory Care. 2003;48(7):681-8.
272.Chang L, Dong Y, Zhou P. Investigation on Risk Factors of Ventilator-Associated Pneumonia in Acute Cerebral Hemorrhage Patients in Intensive Care Unit. Canadian Respiratory Journal. 2017;2017:7272080.
273.Evans H, Zonies D, Warner K, Bulger E, Sharar S, Maier R, et al. Timing of Intubation and Ventilator-Associated Pneumonia Following Injury. Archives of Surgery. 2010;145(11):1041-6.
274.Barnes S, Hamrock E, Toerper M, Siddiqui S, Levin S. Real-Time Prediction of Inpatient Length of Stay for Discharge Prioritization. Journal of the American Medical Informatics Association. 2015;23(e1):e2-e10.
275.Turgeman L, May J, Sciulli R. Insights from a Machine Learning Model for Predicting the Hospital Length of Stay (LOS) at the Time of Admission. Expert Systems with Applications. 2017;78:376-85.
276.Nwagbara V, Rasiah R, Aslam M. An Approach Toward Public Hospital Performance Assessment. Medicine. 2016;95(36):e4688-e.
277.Cunningham J, Kernohan W, Rush T. Bed Occupancy, Turnover Intervals and MRSA Rates in English Hospitals. British Journal of Nursing. 2006;15(12):656-60.
278.Kelly M, Sharp L, Dwane F, Kelleher T, Comber H. Factors Predicting Hospital Length-of-Stay and Readmission After Colorectal Resection: A Population-Based Study of Elective and Emergency Admissions. BMC Health Services Research. 2012;12(1):77.

279.Moore L, Stelfox H, Turgeon A, Nathens A, Bourgeois G, Lapointe J, et al. Hospital Length of Stay after Admission for Traumatic Injury in Canada: A Multicenter Cohort Study. Annals of Surgery. 2014;260(1):179-87.

280.Chen Y, Patel M, McNaughton C, Malin B. Interaction Patterns of Trauma Providers are Associated with Length of Stay. Journal of the American Medical Informatics Association. 2018;25(7):790-9.

281.Bohmer A, Just K, Lefering R, Paffrath T, Bouillon B, Joppich R, et al. Factors Influencing Lengths of Stay in the Intensive Care Unit for Surviving Trauma Patients: A Retrospective Analysis of 30,157 Cases. Critical Care. 2014;18(4):R143.

282.Ingraham A, Xiong W, Hemmila M, Shafi S, Goble S, Neal M, et al. The Attributable Mortality and Length of Stay of Trauma-Related Complications: A Matched Cohort Study. Annals of Surgery. 2010;252(2).

283.Alghnam S, Towhari J, Al Babtain I, Al Nahdi M, Aldebasi M, Alyami M, et al. The Associations Between Injury Mechanism and Extended Hospital Stay Among Pediatric Patients: Findings from a Trauma Center in Saudi Arabia. BMC Pediatrics. 2019;19(1):177.

284.Staudt A, Gurney J, Valdez-Delgado K, Suresh M, Trevino J, Le T, et al. Factors Associated with Trauma Patients' Length of Stay at Role 2 Facilities in Afghanistan, October 2009 to September 2014. Journal of Trauma and Acute Care Surgery. 2018;85(1S).

285.Ong A, Omert L, Vido D, Goodman B, Protetch J, Rodriguez A, et al. Characteristics and Outcomes of Trauma Patients with ICU Lengths of Stay 30 Days and Greater: A Seven-Year Retrospective Study. Critical Care. 2009;13(5):R154.

286.Hwabejire J, Kaafarani H, Imam A, Solis C, Verge J, Sullivan N, et al. Excessively Long Hospital Stays After Trauma Are Not Related to the Severity of Illness: Let's Aim to the Right Target! JAMA Surgery. 2013;148(10):956-61.
287.Velopulos C, Enwerem N, Obirieze A, Hui X, Hashmi Z, Scott V, et al. National Cost of Trauma Care by Payer Status. The Journal of surgical research. 2013;184(1):444-9.
288.Li Y, Wong E, Sales A, Sharp N, Needleman J, Maciejewski M, et al. Nurse Staffing and Patient Care Costs in Acute Inpatient Nursing Units. Medical Care. 2011;49(8):708-15.
289.Zhang Z, Bokhari F, Guo Y, Goyal H. Prolonged Length of Stay in the Emergency Department and Increased Risk of Hospital Mortality in Patients with Sepsis Requiring ICU Admission. Emergency Medicine Journal. 2019;36(2):82-7.
290.Marfil-Garza B, Belaunzarán-Zamudio P, Gulias-Herrero A, Zuñiga A, Caro-Vega Y, Kershenobich-Stalnikowitz D, et al. Risk Factors Associated with Prolonged Hospital Length-of-Stay: 18-Year Retrospective Study of Hospitalizations in a Tertiary Healthcare Center in Mexico. PLOS ONE. 2018;13(11):e0207203.
291.Mathew P, Jehan F, Kulvatunyou N, Khan M, O'Keeffe T, Tang A, et al. The Burden of Excess Length of Stay in Trauma Patients. The American Journal of Surgery. 2018;216(5):881-5.
292.Wigder H, Johnson C, Shah M, Fantus R, Brasic J, Tanouye K, et al. Length of Stay Predicts Patient and Family Satisfaction with Trauma Center Services. The American Journal of Emergency Medicine. 2003;21(7):606-7.
293.Doctoroff L, Hsu D, Mukamal K. Trends in Prolonged Hospitalizations in

the United States from 2001 to 2012: A Longitudinal Cohort Study. The American Journal of Medicine. 2017;130(4):483.e1-.e7.

294.Passias P, Jalai C, Worley N, Vira S, Hasan S, Horn S, et al. Predictors of Hospital Length of Stay and 30-Day Readmission in Cervical Spondylotic Myelopathy Patients: An Analysis of 3057 Patients Using the ACS-NSQIP Database. World Neurosurgery. 2018;110:e450-e8.

295.Krell R, Girotti M, Dimick J. Extended Length of Stay After Surgery: Complications, Inefficient Practice, or Sick Patients? JAMA Surgery. 2014;149(8):815-20.

296.Shi H, Hwang S, Lee K, Lin C. In-Hospital Mortality After Traumatic Brain Injury Surgery: A Nationwide Population-Based Comparison of Mortality Predictors Used in Artificial Neural Network and Logistic Regression Models. Journal of Neurosurgery. 2013;118(4):746-52.

297.Woods A, MacKirdy F, Livingston B, Norrie J, Howie J. Evaluation of Predicted and Actual Length of Stay in 22 Scottish Intensive Care Units Using the APACHE III System. Anaesthesia. 2000;55(11):1058-65.

298.Chattopadhyay A, Chatterjee S. Predicting ICU Length of Stay Using APACHE-IV in Persons with Severe Sepsis – A Pilot Study. Journal of Epidemiological Research. 2016;2(1):1-8.

299.Vasilevskis E, Kuzniewicz M, Cason B, Lane R, Dean M, Clay T, et al. Mortality Probability Model III and Simplified Acute Physiology Score II: Assessing Their Value in Predicting Length of Stay and Comparison to APACHE IV. Chest. 2009;136(1):89-101.

300.Moralez G, Rabello L, Lisboa T, Lima M, Hatum R, De Marco F, et al. External Validation of SAPS 3 and MPM(0)-III Scores in 48,816 Patients from

72 Brazilian ICUs. Annals of intensive care. 2017;7(1):53-.

301.Abujaber A, Fadlalla A, Gammoh D, Abdelrahman H, Mollazehi M, El-Menyar A. Prediction of In-Hospital Mortality in Patients with Post Traumatic Brain Injury Using National Trauma Registry and Machine Learning Approach. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine. 2020;28(1):44.

302.Paterson R, MacLeod D, Thetford D, Beattie A, Graham C, Lam S, et al. Prediction of In-Hospital Mortality and Length of Stay Using an Early Warning Scoring System: Clinical Audit. Clinical Medicine (London, England). 2006;6(3):281-4.

303.Malone D, Dunne J, Tracy J, Putnam A, Scalea T, Napolitano L. Blood Transfusion, Independent of Shock Severity, Is Associated with Worse Outcome in Trauma. Journal of Trauma and Acute Care Surgery. 2003;54(5).

304.Tuma M, Acerra J, El-Menyar A, Al-Thani H, Al-Hassani A, Recicar J, et al. Epidemiology of Workplace-Related Fall from Height and Cost of Trauma Care in Qatar. International journal of critical illness and injury science. 2013;3(1):3-7.

305.Sidi M. Road Accidents A Huge Cost Burden to HMC; Most Victims are Below 30 Years. The Peninsula. 2017.

306.Sandmann F, Robotham J, Deeny S, Edmunds W, Jit M. Estimating the Opportunity Costs of Bed-Days. Health Economics. 2018;27(3):592-605.