DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

MACHINE LEARNING MODELS FOR INPATIENT FALL PREDICTION: A RETROSPECTIVE STUDY USING ELECTRONIC HEALTH RECORD DATA IN TÜRKİYE

by

Veysel Karani BARIŞ

January, 2026

İZMİR

MACHINE LEARNING MODELS FOR INPATIENT FALL PREDICTION: A RETROSPECTIVE STUDY USING ELECTRONIC HEALTH RECORD DATA IN TÜRKİYE

A Thesis Submitted to the Graduate School of Natural and Applied Sciences of Dokuz Eylül University in Partial Fulfillment of the Requirements for Master of Science in Department of Statistics, Data Science Program

by

Veysel Karani BARIŞ

January, 2026

İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"MACHINE LEARNING MODELS FOR INPATIENT FALL PREDICTION: A RETROSPECTIVE STUDY USING ELECTRONIC HEALTH RECORD DATA IN TÜRKİYE"** completed by **VEYSEL KARANİ BARIŞ** under supervision of **PROF.DR. BURCU HÜDAVERDİ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of.Master of Science.

Prof.Dr. Burcu HÜDAVERDİ

Supervisor

Doç. Dr. Engin YILDIZTEPE

Jury Member

Doç. Dr. Aslı SUNER KARAKÜLAH

Jury Member

Prof. Dr. Abdullah SEÇGİN

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENT

I extend my deepest gratitude to my esteemed advisor, Prof. Dr. Burcu HÜDAVERDİ, whose invaluable knowledge and experience guided me throughout my master's education, who supported my professional growth, and who served as an exceptional role model with her discipline, communication skills, and leadership. I also thank all hospital staff who assisted in the collection of the research data.

Most importantly, I offer my heartfelt appreciation to my beloved spouse, my dearest family, and my cherished friends, whose unwavering love, encouragement, and unconditional support have uplifted me not only during this demanding academic journey but throughout every stage of my life.

Veysel Karani BARIŞ

MACHINE LEARNING MODELS FOR INPATIENT FALL PREDICTION: A RETROSPECTIVE STUDY USING ELECTRONIC HEALTH RECORD DATA IN TÜRKİYE

ABSTRACT

Patient falls are among the most common yet preventable hospital adverse events, leading to significant physical, psychological, and economic consequences. This study aimed to develop and validate machine learning models using electronic health record data to predict inpatient falls in a tertiary hospital in Türkiye. We conducted a retrospective case–control study (2017–2023) including 3,639 adult inpatients: 518 fallers and 3,121 matched non-fallers. Predictors were selected based on literature and clinical expertise, encompassing demographics, Charlson Comorbidity Index, fall-risk medications, and laboratory values. Class imbalance (fallers = 14.2%) was addressed using SMOTE-NC. Seven methods were tested: Support Vector Machine, Random Forest, XGBoost, Logistic Regression, Decision Tree, Artificial Neural Network, and Random Forest Quantile Classifier (RFQ). Among methods, RFQ achieved the best performance (AUC: 0.863; accuracy: 0.863; recall: 0.806; specificity: 0.805). Other methods showed lower discrimination (AUC: 0.543–0.784). Feature importance analysis identified Charlson-comorbidity-index, eosinophil count, age, and number of high-risk medications as leading predictors, with additional contributions from basophil count, sodium, blood-urea-nitrogen, and right-blood-cell count. This study demonstrates the first application of RFQ for inpatient fall prediction, offering a robust alternative to traditional resampling approaches. By leveraging routinely available electronic health record data, RFQ enables scalable, automated risk assessment without increasing nursing workload. Integrating such machine learning models into hospital systems could enhance early detection of high-risk patients and strengthen patient safety initiatives.

Key Words: Patient safety, patient falls, machine learning

YATAN HASTALARIN DÜŞME TAHMİNİ İÇİN MAKİNE ÖĞRENMESİ MODELLERİ: TÜRKİYE'DEKİ ELEKTRONİK SAĞLIK KAYIT VERİLERİNİ KULLANAN RETROSPEKTİF BİR ÇALIŞMA

ÖZ

Hastanelerde en sık görülen advers olaylardan birisi hasta düşmeleridir ve bu durum önemli fiziksel, psikolojik ve ekonomik sonuçlara yol açmaktadır. Bu çalışmanın amacı, Türkiye'deki bir üniversite hastanesindeki hasta düşmelerini öngörmek için elektronik sağlık kayıtları verilerine dayalı makine öğrenimi modelleri geliştirmek ve doğrulamaktır. Retrospektif olgu-kontrol araştırmasına 2017–2023 yılları arasında hastanede yatan 3.639 yetişkin hasta dahil edilmiştir (518 düşen hasta ve 3.121 düşmeyen hasta). Tahmin değişkenleri, literatür ve klinik uzmanlık ışığında seçilmiş olup demografik veriler, Charlson Komorbidite İndeksi, düşme riski yüksek ilaç kullanımı ve laboratuvar değerlerini kapsamaktadır. Sınıf dengesizliği (düşenler = %14,2) SMOTE-NC yöntemiyle düzeltilmiştir. Yedi yöntem test edilmiştir: Destek Vektör Makineleri, Rastgele Orman, XGBoost, Lojistik Regresyon, Karar Ağacı, Yapay Sinir Ağı ve Rastgele Orman Kantitatif Sınıflandırıcısı (RFQ). Modeller arasında RFQ en yüksek performansı göstermiştir (AUC: 0,863; doğruluk: 0,863; duyarlılık: 0,806; özgüllük: 0,805) ve hem SMOTE-NC tabanlı modellerden hem de geleneksel Rastgele Orman algoritmasından daha iyi sonuç vermiştir. Diğer yöntemler daha düşük ayırt edicilik göstermiştir (AUC: 0,543–0,784). Özellik önemlilik analizinde en önde gelen belirleyiciler, Charlson Komorbidite İndeksi, eozinofil sayısı, yaş ve yüksek riskli ilaç sayısı olarak tanımlanmış; bazofil sayısı, sodyum, kan-üre-azotu ve kırmızı-kan-hücresi sayısı da ek katkı sağladığı görülmüştür. Bu çalışma, RFQ'nun yatan hasta düşmelerinin öngörülmesinde ilk uygulamasını ortaya koyarak geleneksel yeniden örnekleme yaklaşımlarına güçlü bir alternatif sunmaktadır. Rutin olarak mevcut elektronik sağlık verilerinden yararlanılması sayesinde RFQ, hem ölçeklenebilir hem de hemşire iş yükünü artırmadan otomatik risk değerlendirmesi yapılmasına imkân tanımaktadır. Böyle ML modellerinin hastane sistemlerine entegrasyonu, yüksek riskli hastaların erken tanımlanmasını kolaylaştırabilir ve hasta güvenliği girişimlerini güçlendirebilir.

Anahtar Kelimeler: Hasta güvenliği, hasta düşmeleri, makine öğrenmesi

CONTENTS

Page

M.Sc THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGMENT iii
ABSTRACT iv
ÖZ v
CONTENTS vi
Page vi
LIST OF FIGURES viii
LIST OF TABLES ix
LIST OF ABBREVIATIONS x

CHAPTER ONE - INTRODUCTION 1

1.1 Medical Errors in Healthcare Services 1
1.2 Patient Falls 2
1.3 Risk Factors of Patient Falls 4
1.4 Consequences of Patient Falls 6
1.5 Prediction of Patient Falls 8

CHAPTER TWO - SUPERVISED MACHINE LEARNING 12

2.1 Supervised Machine Learning Methods 13
2.1.1 Support Vector Machine 13
2.1.2 Decision Trees 16
2.1.3 Random Forest 18
2.1.4 Random Forest Quantile Classifier 19
2.1.5 Extreme Gradient Boosting (XGBoost) 20
2.1.6 Logistic Regression 22
2.1.7 Artificial Neural Networks (ANNs) 25
2.2 Cross-Validation 28

2.3 Evaluation Metrics 29

CHAPTER THREE - APPLICATION 31

3.1 Aim 31
3.2 Study Population and Data 31
3.3 Features 32
3.4 Removing Correlated Variables 36
3.5 Missing Value Analysis 37
3.6 Data Scaling 40
3.7 Resampling 40
3.8 Model Training and Prediction 42
3.9 Data Analysis 43

CHAPTER FOUR - RESULTS 45

4.1 Descriptives of Study Participants 45
4.2 Predictive Performance of Models 48
4.3 Feature Importance 51

CHAPTER FIVE - CONCLUSION 53

REFERENCES 56

LIST OF FIGURES

Pages

Figure 2.1 Structure of decision tree 18
Figure 3.1 Corrplot of numeric variables 37
Figure 3.2 Study workflow 44
Figure 4.1 ROC curves and AUC values of models 51
Figure 4.2 Importance matrix of variables 52

LIST OF TABLES

Page

Table 3.1 Features and definitions 32
Table 3.2 Variables and number of missing values 38
Table 4.1 Comparisons of faller and non-faller patients 46
Table 4.2 Performance measures of models before SMOTE_NC 49
Table 4.3 Performance measures of models after SMOTE-NC 50

LIST OF ABBREVIATIONS

AHRQ	: Agency for Healthcare Research and Quality
ANN	: Artificial Neural Network
ATC	: Anatomical Therapeutic Chemical Classification System
AUC	: Area Under the Curve
BUN	: Blood Urea Nitrogen
CCI	: Charlson Comorbidity Index
HER	: Electronic Health Record
HCT	: Hematocrit
HGB	: Hemoglobin
ICD	: International Classification of Diseases
JCI	: Joint Commission International
ML	: Machine Learning
MLE	: Maximum Likelihood Estimation
OECD	: Organisation for Economic Co-operation and Development
RBC	: Red Blood Cell Count
RBF	: Radial Basis Function
RF	: Random Forest
RFQ	: Random Forest Quantile Classifier
ROC	: Receiver Operating Characteristic
SMOTE	: Synthetic Minority Over-sampling Technique
SMOTE-NC	: Synthetic Minority Over-sampling Technique – Nominal Continuous
WHO	: World Health Organization
XGBoost	: Extreme Gradient Boosting

CHAPTER ONE
INTRODUCTION

1.1 Medical Errors in Healthcare Services

The fundamental principle of health care is to avoid causing harm, yet the inherent complexity of modern health systems often creates conditions where errors can emerge. Multiple providers, fragmented communication, high patient acuity, and the increasing use of advanced technologies all contribute to environments where mistakes are more likely to occur. Medical errors, which are recognized as one of the most serious consequences of care delivery, are commonly defined as either the failure to carry out an intended action correctly or the execution of an inappropriate plan to achieve a clinical objective. Landmark evidence has demonstrated that tens of thousands of patients lose their lives each year as a direct result of such errors, positioning medical error as a major public health problem that extends beyond individual institutions or health systems (Kohn et al., 2000). Despite decades of concentrated attention to patient safety and the adoption of numerous strategies, such as standardized protocols, reporting systems, and safety culture initiatives, the overall burden of medical errors has not declined to a satisfactory degree. Current estimates indicate that approximately one in ten hospitalized patients continues to experience harm related to lapses in patient safety, underscoring the persistent and urgent nature of this challenge (Naseralallah, Stewart, Azfar, & Paudyal, 2022).

Medical errors, which can occur at nearly every stage of health care delivery, from diagnosis and treatment to medication administration and follow-up, remain a major threat to patient safety. These events arise from diverse causes, including communication breakdowns, system inefficiencies, human factors, and deficiencies in care coordination, and they collectively affect millions of people each year. The World Health Organization (WHO) estimates that in low- and middle-income countries alone, approximately 134 million adverse events occur annually, resulting in 2.6 million deaths, illustrating the immense global burden of unsafe care (World Health Organization [WHO], 2025). In high-income settings, the problem is similarly alarming. Evidence has shown that medical

errors account for more than 250,000 deaths annually, making them one of the leading causes of mortality in hospital settings (Makary & Daniel, 2016). Such findings underscore that patient harm due to medical errors is not only a challenge in resource-limited contexts but a pervasive and systemic issue across health systems worldwide.

1.2 Patient Falls

Patient falls are recognized as one of the most frequent and preventable medical errors within hospital settings (Morris et al., 2022). The World Health Organization defines a fall as the "inadvertent coming to rest on the ground, floor, or another lower level, excluding purposeful movements to rest against furniture, walls, or other objects" (WHO, 2025). Beyond their frequency, falls are considered one of the most serious adverse events in clinical care because of their potential to cause injury, extend hospital stays, increase costs, and worsen overall outcomes. According to the Joint Commission International (JCI), patient falls continue to represent the most commonly reported sentinel event, reflecting their significance as a persistent safety concern (The Joint Commission, 2025). Evidence indicates that falls account for a substantial proportion of hospital-based medical errors, with estimates ranging from 3.7% to 80% depending on the type of institution and reporting system used (Bernet, Everink, Hahn, Muller & Schols, 2024; Lee et al., 2021; García-Hedrera et al., 2021; Anderson et al., 2015). The magnitude of this issue is further highlighted by estimates suggesting that more than one million hospitalized patients in the United States experience at least one fall each year (Staggs, Mion & Shorr, 2014). Similarly, health services in other countries have reported a high burden, with hundreds of thousands of inpatient falls recorded annually (Royal College of Physicians [RCP], 2017). Together, these findings emphasize that patient falls remain a global and systemic challenge in health care safety.

Patient falls are among the most common and serious patient safety incidents, often classified as medical errors due to their preventable nature and potential to cause significant harm. Although reported incidence rates vary depending on patient population and care setting, studies have consistently demonstrated their high frequency. For

example, fall rates have been reported as 1.61 – 2.51 in recent study (García-Hedrera et al., 2021). Beyond their frequency, patient falls are also recognized for their severity. According to the JCI, which tracks adverse events leading to death, permanent disability, or serious injury requiring life-sustaining intervention, falls were the second most frequently reported medical error in 2024 (JCI, 2025). Broader analyses reinforce this concern: the Organization for Economic Co-operation and Development (OECD) identified falls, together with medication errors, pressure injuries, and misdiagnoses, as among the most frequent adverse events in hospitals (OECD, 2018). Similarly, data from the Agency for Healthcare Research and Quality (AHRQ) showed that in 2017, more than 2.5 million patient safety events were reported among hospitalized adults in the United States, with falls ranking as the third most common incident type (Agency for Healthcare Research and Quality [AHRQ], 2019). Taken together, these findings provide strong evidence that patient falls represent a widespread and persistent problem across health systems, underscoring their importance as a critical focus of patient safety initiatives at both national and international levels.

Patient falls are commonly classified into three major categories, each reflecting distinct underlying causes and prevention opportunities (Rowe, 2013). **Accidental falls** occur unintentionally as a result of environmental hazards such as wet floors, cluttered walkways, poor lighting, or improperly placed equipment. Because these incidents are directly linked to the patient's surroundings, they are often preventable through systematic environmental safety checks and hazard reduction strategies. **Unanticipated physiologic falls** are triggered by sudden, unexpected medical events such as seizures, strokes, or syncopal episodes. These falls are considered "unanticipated" because no risk factors are typically evident during standard assessments, which makes both prediction and prevention particularly challenging. **Anticipated physiologic falls** arise from pre-existing conditions that are identifiable through clinical evaluation, including impaired mobility, muscle weakness, poor vision, balance difficulties, or urinary incontinence. Since these risks can be recognized in advance, anticipated falls are regarded as predictable and largely preventable through targeted interventions such as mobility support, vision aids,

or toileting programs. This classification framework provides a comprehensive approach for understanding nearly all types of inpatient falls and highlights the need for tailored prevention strategies that address both modifiable environmental risks and patient-specific clinical factors.

1.3 Risk Factors of Patient Falls

Extensive research has demonstrated that patient falls are multifactorial events shaped by a wide range of intrinsic and extrinsic factors. Individuals who present with one or more of these risk factors are considerably more likely to experience a fall than those without them, making systematic assessment a critical component of prevention. Effective fall prevention, therefore, requires a personalized approach in which each patient is carefully evaluated for their unique set of risks and matched with targeted interventions. In this context, risk factors serve a dual role: they contribute to the likelihood of falls occurring, but when identified accurately, they also provide valuable opportunities for prevention. Incorporating structured risk assessments into routine care enables clinicians to implement evidence-based strategies that not only reduce fall incidence but also minimize the severity of fall-related harm (Najafpour, Godarzi, Arab & Yaseri, 2019).

A substantial body of evidence has identified a wide range of risk factors that contribute to patient falls in hospital settings. Among the most consistently reported are a prior history of falls, which strongly predicts future incidents, and advanced age, which is often associated with frailty, reduced muscle strength, and slower reflexes. Cognitive impairment, including delirium and dementia, also increases vulnerability by affecting judgment, attention, and orientation. Physiological conditions such as balance disorders, dizziness, and gait abnormalities further heighten the likelihood of instability, while frequent toileting needs can increase exposure to unsafe mobility, particularly during the night. Sensory deficits, especially visual impairment, reduce patients' ability to detect hazards in their environment. Medication-related factors, including the use of antipsychotics and polypharmacy, are repeatedly highlighted as major contributors due to their potential to cause sedation, confusion, or orthostatic hypotension. In addition,

underlying health conditions such as depression and cancer have been linked to greater fall risk, reflecting the interplay between chronic disease, functional decline, and treatment effects (Shao et al., 2023; Li et al., 2023; Kakara, 2023).

Although numerous risk factors for patient falls have been identified, there is still no universal agreement on their exact type or number, and findings often vary depending on the population studied. Nevertheless, certain predictors have consistently emerged across the literature. Prior fall history, cognitive impairment, and the use of sedatives or antidepressants have been shown to significantly increase fall risk (Lee at al., 2021). Large-scale analyses of hospital risk assessments have further highlighted recent falls within the past six months, changes in consciousness, and toileting-related issues—such as urgency, frequency, or diuretic use—as particularly strong predictors (Moe, Brockopp, McCowan, Merritt, & Hall 2015). Additional evidence indicates that prolonged hospital stay, use of high-risk medications including chemotherapy agents, sedatives, anticonvulsants, benzodiazepines, and ACE inhibitors, as well as physical and functional conditions such as poor vision, balance impairment, reliance on assistive devices, urinary incontinence, and cancer diagnosis, substantially elevate fall likelihood (Najafpour et al., 2019). Other studies have also emphasized the contribution of equipment dependence, assisted mobilization, and frequent nighttime toileting to heightened fall risk (Ishikuro et al., 2017). Collectively, these findings underscore the multifactorial nature of falls, demonstrating that both patient-related and treatment-related factors interact to shape overall vulnerability.

Tailored interventions that specifically address identified risk factors have been shown to significantly reduce fall incidence and improve patient outcomes (Cameron et al., 2018; Dykes et al., 2017). Such interventions may include environmental modifications, medication reviews, mobility and balance training, vision correction, or toileting assistance, depending on the patient's unique profile of risks. As a result, systematic assessment of patients and accurate identification of those at heightened risk form the cornerstone of effective fall-prevention strategies. Incorporating individualized

risk evaluations into routine clinical practice not only allows for timely and targeted intervention but also enhances the overall efficiency of prevention programs, ultimately reducing fall-related harm and resource utilization (Montero-Odasso et al., 2022).

1.4 Consequences of Patient Falls

Although many inpatient falls do not result in harm, they can lead to a wide spectrum of injuries ranging from minor abrasions, cuts, and hematomas to severe outcomes such as fractures, dislocations, head or spinal injuries, and, in rare cases, death (Baris, Intepeler & Yeginboy, 2018). Evidence from large-scale incident analyses has shown that the majority of falls cause no injury, while a considerable proportion result in minor harm, and a smaller percentage lead to moderate or severe injuries or fatalities (Hignett, Sands & Griffiths, 2013). When falls do result in serious harm, fractures are among the most frequently reported consequences. Data indicate that hip and pelvic fractures, upper and lower extremity fractures, and vertebral injuries account for a substantial proportion of severe outcomes, often requiring surgical intervention or prolonged hospitalization. Other serious consequences include subdural hematomas, facial fractures, and combined injury patterns, which can further complicate recovery (Wong et al., 2011). Additional studies have also highlighted the prevalence of hip fractures, upper extremity injuries, and head or spinal trauma as common outcomes of fall-related harm, emphasizing the potential for long-term disability and increased care needs (Baris et al., 2018). Collectively, these findings demonstrate that while not all falls result in injury, those that do can have devastating physical, psychological, and economic consequences for patients, families, and healthcare systems.

In addition to physical harm, patient falls often result in substantial psychological consequences. Even in cases where no injury occurs, the experience of falling can undermine a patient's sense of security and lead to emotional distress for both patients and their families. Falls and fall-related disabilities have been strongly associated with depression, anxiety, and a persistent fear of falling. This fear frequently leads patients to restrict their activities as a protective measure, which can in turn cause deconditioning,

loss of confidence, and diminished social participation. Over time, these behavioral and emotional responses may contribute to social isolation, reduced quality of life, and an elevated risk of subsequent falls. In severe cases, the combined effects of physical decline and psychological burden can even contribute indirectly to increased mortality (Graham, 2012).

Among the psychological consequences of falls, fear of falling is the most prevalent and widely reported. It is commonly defined as a loss of self-confidence in one's ability to carry out daily activities without falling. This diminished confidence frequently leads to restriction of routine activities, even in the absence of physical limitations. Older adults are particularly vulnerable, often avoiding tasks they are otherwise capable of performing, which can result in confinement to the home or even prolonged periods of bed rest. Such behavioral avoidance contributes to reduced independence, social withdrawal, and emotional isolation (Graham, 2012). Evidence suggests that fear of falling has a substantial impact on daily functioning; for example, nearly half of patients receiving home care services have reported limiting their activities because of this concern. Over time, the resulting sedentary lifestyle promotes physical deconditioning, worsens social isolation, and intensifies the very fear it is meant to avoid, ultimately creating a self-perpetuating cycle that further increases vulnerability to future falls (MacKay, Ebert, Harbidge & Hogan, 2021).

Beyond their physical and psychological impact, patient falls also place a substantial economic burden on health systems. Fall-related injuries often require additional diagnostic tests, treatments, surgical interventions, and rehabilitation, which significantly increase resource utilization. These events also extend hospital stays, heighten demands on healthcare personnel, and generate costs associated with long-term care or readmissions. The financial impact is considerable: for example, it has been estimated that the annual cost of inpatient falls in England exceeds €630 million (RCP, 2017). Research further demonstrates that falls can prolong hospital stays by 7.5 to 34 days and increase hospitalization costs by $1,139 to $13,806 per patient, depending on

injury severity and care needs (Morello et al., 2015; Spetz, Brown & Aydin, 2015; Baris et al., 2018). Collectively, these findings highlight that the economic consequences of falls extend well beyond the individual patient, placing a sustained financial strain on hospitals and health systems, and underscoring the importance of prevention as both a clinical and economic priority.

Evidence indicates that patients who sustain injuries as a result of a fall remain hospitalized an average of four days longer and generate approximately $4,522 more in direct medical costs compared to those who fall without injury (Morello et al., 2015). However, the true economic burden of falls extends far beyond hospital bills. In addition to direct healthcare expenditures, falls often lead to substantial indirect costs, including legal expenses, compensation for pain and suffering, and the psychological toll experienced by patients and their families. Long-term disabilities arising from fall-related injuries can also contribute to productivity losses, especially when individuals are unable to return to work or require extended rehabilitation and caregiving support. Despite their significance, accurately quantifying these indirect costs remains challenging, and current evidence provides only limited estimates in this area (Baris et al., 2018). Collectively, these findings highlight that falls represent not only a major clinical and safety concern but also a complex economic issue with both visible and hidden costs for patients, families, and health systems.

1.5 Prediction of Patient Falls

According to evidence from a Cochrane review, systematic identification of patients' fall risk factors combined with the implementation of targeted preventive interventions can reduce fall rates by approximately 31% (Cameron et al., 2018). These findings underscore the importance of risk assessment as the foundation of fall-prevention strategies. However, the overall effectiveness of prevention efforts depends heavily on the quality and accuracy of the assessment tools employed. For a tool to be clinically useful, it must not only identify the presence of relevant risk factors but also reliably distinguish between patients at high and low risk of falling (Strini, Schiavolin & Prendin, 2021).

Inaccurate or poorly calibrated tools may lead to underestimation of risk, leaving vulnerable patients without adequate safeguards, or overestimation, which can result in unnecessary interventions, wasted resources, and staff burden. Therefore, rigorous validation and continual refinement of fall risk assessment tools remain essential to maximizing the effectiveness of fall-prevention programs.

Currently, fall risk is most often assessed using standardized instruments such as the Morse Fall Scale, the Johns Hopkins Fall Risk Assessment Tool (JHFRAT), the Hendrich II Fall Risk Model, and STRATIFY (Park, 2018). While these instruments are widely adopted in clinical practice, their predictive accuracy is often limited, which reduces their ability to reliably forecast inpatient falls (Strini et al., 2021). Furthermore, many of these tools require manual data collection and scoring, making them time-intensive and adding to the workload of already overburdened healthcare staff. Another important limitation is that they frequently fail to capture the full spectrum of intrinsic and extrinsic risk factors that contribute to falls, leaving gaps in patient safety assessments (Baris & Seren Intepeler, 2019). These shortcomings highlight the need for more precise, efficient, and comprehensive approaches to fall risk prediction that can integrate seamlessly into routine clinical workflows.

With advances in digital technology, electronic health records (EHRs) now capture an extensive range of clinical data, providing a valuable foundation for fall risk prediction. Many established risk factors—such as age, comorbidities, medication use, and mobility status—are routinely documented in EHRs, and additional variables available in these systems may further enhance predictive capability (Dormosh et al., 2023). However, the sheer volume, heterogeneity, and complexity of EHR data make it difficult to derive actionable insights using traditional analytic methods. In this context, machine learning (ML) has emerged as a powerful approach, offering greater predictive accuracy and the ability to uncover subtle, non-linear patterns that conventional models may overlook. When integrated into EHRs, ML-based algorithms can automate the analysis of both structured and unstructured data, thereby reducing the need for manual data entry and

minimizing staff workload. Importantly, these models have the potential to identify patients at high risk of falling in real time and to generate actionable insights that can be directly incorporated into clinical workflows (Ladios-Martin et al., 2022). A further advantage of ML approaches is their reliance on routinely collected and easily accessible clinical variables, which makes them well suited for integration into clinical decision support systems. Embedding such models into routine practice enables the delivery of timely, evidence-based recommendations, ultimately improving both the quality and efficiency of patient care.

Although numerous studies have applied ML techniques to predict fall risk, most have notable limitations in either population scope or predictor range For instance, Dormosh et al. (2023) and Chu et al. (2022) concentrated almost exclusively on older adults, with Chu et al. (2022) uniquely incorporating geriatric assessment measures such as cognition and functional status, while Dormosh et al. (2023) relied on a narrower set of clinical and demographic variables. Some recent efforts have sought to move beyond these constraints.–Shim et al. (2022) evaluated a broader inpatient population using diverse EHR-derived features, and Jahandideh et al. (2024) expanded the scope further by integrating organizational and staffing-related variables alongside patient characteristics. Other contributions have emphasized still different dimensions, such as sensor-based physiological measures (Bargiotas et al., 2023), trauma outcomes (Young et al., 2021), fall severity (Wang et al., 2019), inclusion of nursing prevention interventions (Ladios-Martin et al., 2022), and dynamic EHR-based risk assessment across hospitalization stages (Liu, Hu & Lin, 2021). Collectively, these approaches demonstrate innovation but remain constrained by focusing on particular subpopulations, predictor categories, or outcomes, rather than capturing the full complexity of inpatient fall risk.

While it is well established that older adults are at heightened risk of falls due to frailty, comorbidities, and polypharmacy, falls remain a critical safety concern across all age groups in hospital settings. Younger and middle-aged patients may also be vulnerable, particularly in the presence of acute illness, high-risk treatments, mobility restrictions, or

medication side effects. This underscores the need for more comprehensive and inclusive predictive models that integrate a wide range of patient, clinical, and organizational factors. By fully leveraging the analytical capabilities of ML, such models can incorporate diverse variables, capture complex interactions, and enhance predictive performance, thereby offering stronger support for fall-prevention initiatives across varied patient populations and clinical environments.

CHAPTER TWO
SUPERVISED MACHINE LEARNING

Supervised learning is about adjusting a model's parameters with the help of labeled data, so that the refined model can make accurate predictions when applied to new, unseen datasets. We can view this from another perspective. Suppose we have *n* candidate models: $y = f_{w1(x)}$, $y = f_{w2(x)}$, …, $y = f_{wn(x)}$ Using a labeled dataset for training and validation, we determine which model $y = f_{wi(x)}$ performs best. Once the optimal model is chosen, its effectiveness is assessed on a separate test dataset. From this standpoint, supervised learning can be broken down into four main goals: (1) tuning model parameters, (2) generating algorithms for tuning, (3) improving the models to work with unseen data, and (4) applying efficient quantitative and qualitative measures for tuning.

- **Training Algorithm:** A training algorithm serves as a structured method for producing model parameters and identifying the most effective ones through the use of labeled data, known as the training set. The goal is to minimize the gap between the model's predicted labels and the true labels in this dataset. Accomplishing this requires a reliable quantitative metric to evaluate performance and guide parameter optimization.

- **Testing Algorithm:** The testing algorithm offers a structured method to verify whether the model, once optimized with its best parameters, performs well on a separate labeled dataset known as the test set. This verification ensures that the tuned model maintains strong classification accuracy beyond the training phase.

- **Validation Algorithm:** A validation algorithm provides a structured way to expose models to different training and testing scenarios, with the goal of improving their ability to handle unseen data. This involves splitting the labeled dataset into multiple combinations of training and validation subsets, then applying the training and testing steps to each combination. In simple terms, validation can be

thought of as a “practice test” for the model before the final evaluation on the test set (Ayyadevara, 2018).

2.1 Supervised Machine Learning Methods

2.1.1 Support Vector Machine

Support Vector Machines (SVMs) are a robust group of supervised learning techniques mainly designed for classification tasks, though they can also be extended to regression problems. Their core principle is the use of hyperplanes to divide data points into distinct classes (Kotsiantis, Zaharakis & Pintelas, 2007).

Some definitions are given as:

Maximal margin classifier /Hyperplane

In the simplest definition, a hyperplane is a flat subspace of dimension (p−1) in a p-dimensional space. For example, if our space is two-dimensional, the hyperplane will be a one-dimensional straight line. In a three-dimensional space, the hyperplane will be two-dimensional.

The equation of the Maximal Margin Classifier (Hyperplane) is:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p = 0 \quad (2.1)$$

If an observation $X = (X_1, X_2, \ldots, X_p)$ is substituted into the above equation and the equality holds (the result is equal to zero), then the observation lies on the hyperplane. If the value of X makes the following expression greater than 0, then it lies on one side of the plane:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p > 0 \quad (2.2)$$

If the result is less than 0, then it lies on the other side of the plane:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p < 0 \quad (2.3)$$

The determination of the maximal margin classifier is an optimization problem:

$$y_i(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p) \geq M, for\ i = 1, \ldots, n \quad (2.4)$$

Here, M represents the margin. We aim to achieve the largest possible margin. Therefore, during the optimization process, the parameters $\beta_0, \beta_1, \ldots, \beta_p$ are chosen to maximize M (James et al., 2021).

Support vector classifier

In practice, perfect separation may not be possible. The Support Vector Classifier (soft margin) allows some misclassifications by introducing slack variables (ε_i) and a tuning parameter (C). This soft margin approach balances bias-variance tradeoff and improves generalization.

The Support Vector Classifier is formulated as an optimization problem that maximizes the margin (M) while allowing for some misclassifications through slack variables (ε_i). The problem can be expressed as follows:

Maximize: M

1. Normalization constraint:

$$\sum(\beta_j^2) = 1, for\ j = 1, \ldots, p \quad (2.5)$$

2. Classification constraint with slack variables:

$$y_i(\beta_0 + \beta_1 x_{\{i1\}} + \beta_2 x_{\{i2\}} + \ldots + \beta_p x_{\{ip\}}) \geq M(1 - \varepsilon_i), for\ i = 1, \ldots, n \quad (2.6)$$

3. Slack variable constraints:

$$\varepsilon_i \geq 0, \ \sum \varepsilon_i \leq C \quad (2.7)$$

Here, C is a tuning parameter that controls the tradeoff between maximizing the margin and allowing classification errors. The ε_i variables permit certain observations to violate the margin or even be misclassified (Cerulli, 2023).

Support Vector Machines (Nonlinear Boundaries)

SVM extends the Support Vector Classifier to handle nonlinear decision boundaries using the kernel trick. Kernels implicitly map data into higher-dimensional spaces without explicit computation, enabling flexible decision boundaries. The main kernels are:

Linear Kernel

$$K(X_i, X_i') = \sum \left(X_{\{ij\}} * X'_{\{ij\}}\right), for\ j = 1, \dots, p \tag{2.8}$$

The linear kernel corresponds to the standard support vector classifier and is suitable when the classes are linearly separable.

Polynomial Kernel

$$K(X_i, X_i') = \left(1 + \sum \left(X_{\{ij\}} * X'_{\{ij\}}\right)\right)^d \tag{2.9}$$

Here, d is a positive integer that determines the flexibility of the decision boundary. Higher values of d allow for more complex boundaries.

Radial Basis Function (RBF) Kernel

$$K(X_i, X_i') = \exp\left(-\gamma \sum \left(X_{\{ij\}} - X'_{\{ij\}}\right)^2\right) \tag{2.10}$$

The RBF kernel is widely used because it captures local patterns effectively. The parameter γ controls the influence of each training example, with higher values leading to tighter decision boundaries (Kotsiantis et al., 2007).

2.1.2 Decision Trees

A decision tree is a supervised learning model designed to predict qualitative (categorical) outcomes. It works by hierarchically partitioning the feature space into subdomains, where each terminal node corresponds to a predicted class label. Each observation is assigned to the most frequently occurring class within its region. Decision trees are widely used due to their interpretability and ability to capture non-linear relationships (An et al., 2023).

Splitting Criterion and Information Gain

Decision trees are grown through recursive partitioning of the feature space. At each node, the algorithm selects a predictor and an appropriate split (binary or multiway, depending on the specific algorithm) to maximize information gain and improve class separation in the resulting subdomains. The choice of split is guided by quantitative measures of impurity, with the goal of creating purer nodes.

Classification Error Rate

$$E = 1 - \max_k (\hat{p}_{mk}) \tag{2.11}$$

where $\hat{p}_{mk}$ is the proportion of training observations in node m that belong to class k.

- Reflects the proportion of misclassified samples in a node.
- Simple but less sensitive for tree growth.
- Often used for evaluating pruned trees rather than growing them.

However, it turns out that classification error is not sufficiently sensitive for tree-growing, and in practice some other measures are preferable.

Gini Index

The Gini index measures the degree of impurity or diversity in a node. It is defined as:

$$G = \sum_{k} \hat{p}_{mk}(1 - \hat{p}_{mk}) \tag{2.12}$$

where $\hat{p}_{mk}$ is the proportion of samples of class k in node m. A Gini index of 0 indicates perfect purity, meaning all samples belong to one class. Commonly used in practice for growing trees (Charbuty & Abdulazeez, 2021).

Entropy

Entropy quantifies the amount of uncertainty or randomness in a node. It is given by:

$$D = -\sum_{k} \hat{p}_{mk} \log(\hat{p}_{mk}) \tag{2.13}$$

where $\hat{p}_{mk}$ represents the proportion of class k in node m. An entropy value of 0 indicates a pure node, while higher values indicate greater class mixing. Closely related to information gain, where a split is chosen to reduce entropy.

Parametrization

The construction of a classification tree requires parametrization of:

- Feature selection at each node.
- Threshold selection for splitting.
- Maximum depth or stopping criterion of the tree.

These parameters define the model's complexity and influence its generalization ability.

Optimization

Optimization refers to the search for the best feature and split that maximize information gain or minimize impurity. This iterative process continues until no significant improvement can be achieved, or until stopping criteria are met. Pruning is later applied to reduce overfitting and improve predictive performance (Ayyadevara, 2018).

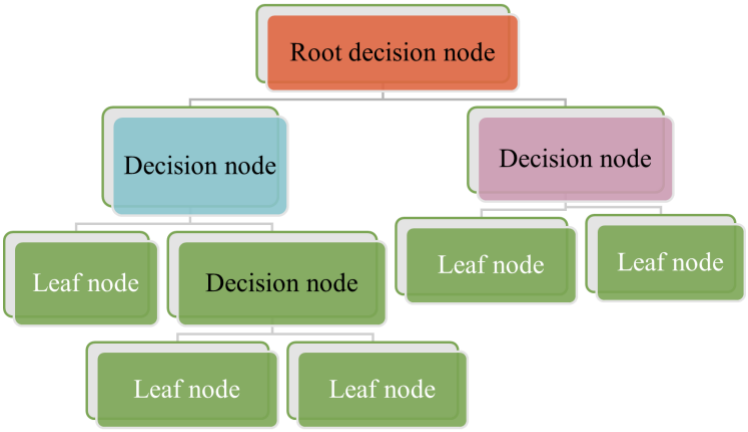


Figure 2.1 Structure of decision tree

2.1.3 Random Forest

Random Forest is an ensemble learning method that extends the decision tree algorithm by mitigating its major drawback—overfitting. While a single decision tree tends to adapt too closely to the training data, resulting in poor generalization, Random Forest improves predictive performance by combining the outputs of multiple decision trees.

Each tree in the forest is constructed on a bootstrap sample of the original dataset, meaning the training data for every tree is drawn with replacement. This ensures that each tree is trained on a slightly different version of the dataset, introducing variability across the ensemble. The principle of bootstrap aggregation (bagging) reduces variance, making the combined model more stable and less prone to overfitting (Shalev-Shwartz, & Ben-David, 2014).

Beyond bootstrapping, Random Forest introduces an additional layer of randomness through feature randomization**.** At each split in the tree-building process, instead of evaluating all available predictors, the algorithm selects a random subset of features. The splitting decision is then based only on this subset. This strategy prevents trees from repeatedly selecting the same “strong” predictors, thereby decorrelating the

trees and encouraging diversity in the ensemble. The aggregation of these diverse trees typically produces a more accurate and robust predictor than any individual tree.

A central tuning parameter in Random Forest is m**,** which represents the number of features considered at each split. The choice of m has direct implications for the bias–variance trade-off:

- When m is small**,** the trees are more decorrelated, since each tree is likely to be constructed on different sets of predictors. This generally reduces variance but may increase bias because individual trees become weaker and less precise.

- When m is large (closer to the total number of predictors p), trees resemble those grown using the full feature set, resulting in lower bias but higher variance.

Thus, m plays a role analogous to a regularization parameter and must be carefully tuned to optimize performance. A common heuristic is to set $m \approx \sqrt{p}$ for classification tasks, though cross-validation provides a more principled approach to selecting this parameter.

In practice, Random Forest predictions are obtained by aggregating the outputs of the ensemble. For classification tasks**,** the final prediction is determined by majority voting across trees. For regression tasks**,** predictions are averaged across all trees in the forest (Zollanvari, 2023).

2.1.4 Random Forest Quantile Classifier

Random Forests are widely used because of their ability to handle high-dimensional data, capture nonlinear interactions, and remain robust to noise. However, in imbalanced classification problems traditional Random Forests often become biased toward the majority class. To address this limitation, O'Brien and Ishwaran (2019) proposed the Random Forest Quantile Classifier (RFQ), a modification of Random Forests that applies quantile-based decision thresholds, improving performance in imbalanced settings.

Quantile Classifier Framework

a) *Bayes Classifier:* The Bayes classifier minimizes misclassification error by assigning an observation to class 1 if:

$$p(x) = P(Y = 1 \mid X = x) \geq 0.5 \tag{2.14}$$

This is optimal under balanced priors, but in imbalanced data, the minority class prior probability is typically much smaller than 0.5, making this threshold inappropriate.

b) *q-Classifier:* The quantile classifier generalizes the Bayes rule by introducing a threshold q:

$$\delta_{q(x)} = 1\{ p(x) \geq q \}, \ 0 < q < 1 \tag{2.15}$$

If q = 0.5, this reduces to the Bayes rule. If $q = \pi = P(Y = 1)$ the classifier achieves TNR (true negative rate) + TPR (true positive rate) optimality, maximizing the sum of sensitivity and specificity. This formulation is called the q-classifier.

The RFQ adapts the quantile classifier to Random Forests. The procedure is:

1. Train a Random Forest with bootstrap samples and random feature selection.
2. For each tree, estimate p(x) using class proportions in terminal nodes.
3. Aggregate probabilities across trees to obtain p_RF(x).
4. Apply the quantile rule: $\delta_{RFQ(x)} = 1\{ p_{RF(x)} \geq \pi \}$ where π is the minority class prior.

This approach avoids oversampling or under sampling and leverages all available data (Ishwaran, Robert, Lu, & Kogalur, 2023).

2.1.5 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a powerful and scalable machine learning algorithm that belongs to the family of gradient boosting methods. It was introduced by Tianqi Chen and Carlos Guestrin in 2016 as an optimized gradient tree boosting system (Chen & Guestrin, 2016). XGBoost has gained significant popularity due to its exceptional performance in

terms of speed, accuracy, and generalization across a wide range of machine learning problems, particularly those involving structured or tabular data.

XGBoost builds on the concept of boosting, where multiple weak learners (usually decision trees) are combined sequentially to create a strong model. Each subsequent tree is trained to minimize the residual errors of the previous trees, thereby improving the model iteratively. Unlike bagging, which trains trees independently, boosting creates a dependency among trees, resulting in cumulative error correction (Anand & Selvaganesan, 2025).

The objective function of XGBoost can be expressed as:

$$L(\Phi) = \sum l(\hat{y}_i, y_i) + \sum \Omega(f_k) \tag{2.16}$$

where:

- $l(\hat{y}_i, y_i)$ is the loss function (e.g., squared error for regression, log loss for classification)
- $\Omega(f_k)$ is the regularization term to control model complexity
- $\hat{y}_i$ is the predicted value, y_i is the true label
- f_k represents the k-th tree

The XGBoost engine begins by defining an initial prediction value, typically the mean of the target values in regression tasks or the log odds for classification problems. This provides a baseline estimate for the model. Next, it calculates the initial residuals**,** which represent the difference between the observed target values and the initial prediction. These residuals quantify the error that the model must correct. Once residuals are established, the boosting iterations commence. In each iteration (also known as a boosting round), a new decision tree is trained on the residuals of the previous prediction. The new tree attempts to capture the patterns that explain the remaining errors. Each subsequent tree thus incrementally improves the model's predictive ability by focusing on the mistakes of its predecessors. This iterative improvement is guided by the gradient of the loss function**,** which provides the direction and magnitude of adjustment needed. Each tree contributes to the final model proportionally to a learning rate (shrinkage parameter)**.**

A smaller learning rate slows down learning but improves generalization, while a larger rate accelerates training but may risk overfitting (Bonaccorso, 2018).

To further enhance generalization, regularization techniques such as L1 (Lasso) and L2 (Ridge) penalties are applied. These penalties discourage overly complex trees, ensuring that the model balances bias and variance effectively. XGBoost also supports mechanisms like column subsampling, row subsampling, and maximum depth control, which collectively limit model complexity.

Finally, optimization methods are employed to minimize the objective function**,** which combines the training loss (measuring how well predictions match actual outcomes) with the regularization term (controlling complexity). As emphasized by Chen & Guestrin (2016), this combination ensures that the model not only fits the training data but also generalizes well to unseen data. Through this iterative refinement process**,** XGBoost continually updates predictions, reduces residual errors, and converges toward a highly accurate model. The synergy of gradient-based optimization, shrinkage, and regularization makes XGBoost exceptionally robust across a wide range of predictive modeling tasks (Ayyadevara, 2018).

2.1.6 Logistic Regression

Logistic regression is one of the most widely used statistical models for classification problems. Despite the name 'regression,' logistic regression is a classification method that predicts the probability that a given input belongs to a particular class. It is especially effective for binary classification, where the dependent variable takes only two possible outcomes (e.g., 0 or 1, success or failure, churn or no churn). The central idea is to model the log-odds of the probability as a linear function of the independent variables (Ray, 2019).

Linear regression assumes a continuous outcome and can predict values outside the [0,1] range, which makes it unsuitable for modeling probabilities. For example, in a binary classification problem such as predicting whether a customer churns or not, linear regression may predict probabilities greater than 1 or less than 0, which are not meaningful. Moreover, linear regression does not model the S-shaped relationship often observed in classification tasks. To overcome the limitations of linear regression, logistic regression uses the sigmoid (logistic) function. The sigmoid maps any real-valued number into the range (0,1), making it suitable for probability estimation. The sigmoid function is defined as:

$$\sigma(t) = \frac{1}{(1 + e^{-t})} \tag{2.17}$$

This function produces an S-shaped curve, where values close to $-\infty$ approach 0, and values close to $+\infty$ approach 1.

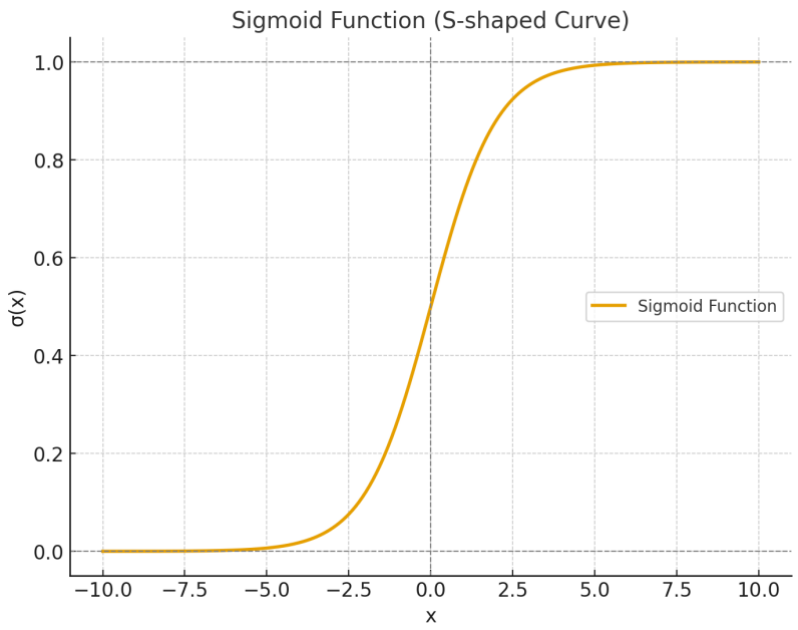


Figure 2.2 Sigmoid function

In logistic regression, the probability that an observation belongs to class 1 is given by:

$$P(Y = 1 \mid X) = 1 / (1 + \exp(-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n))) \tag{2.18}$$

The model estimates the log-odds (logit), which is the natural logarithm of the odds ratio:

$$\log\left(\frac{p}{(1-p)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \quad (2.19)$$

here p is the probability of Y=1. The coefficients β are estimated using Maximum Likelihood Estimation (MLE).

The parameters in logistic regression are estimated using MLE, which finds the set of parameters that maximizes the probability of observing the given data. This is achieved by maximizing the log-likelihood function or equivalently minimizing the cross-entropy loss. Optimization is often performed using iterative algorithms such as Newton-Raphson or stochastic gradient descent (James, Witten, Hastie & Tibshirani, 2021).

Binary Logistic Regression

This is the simplest form, used when the dependent variable has two categories (e.g., diseased vs. healthy, churn vs. no churn). The model estimates the probability of one class (often coded as 1) versus the other class (coded as 0). It is widely used because many real-world problems involve binary decisions.

Multinomial Logistic Regression

When the dependent variable has more than two unordered categories (e.g., predicting which brand a consumer will choose among three options), binary logistic regression is insufficient. Multinomial logistic regression generalizes the binary case by modeling the log-odds of each category relative to a baseline category. Instead of the sigmoid function, it often uses the softmax function to ensure that the predicted probabilities across all categories sum to 1.

Ordinal Logistic Regression

This extension applies when the dependent variable is categorical but the categories have a natural order (e.g., rating levels: poor, fair, good, excellent). While

multinomial logistic regression treats categories as unordered, ordinal logistic regression takes the order into account. The proportional odds model is a commonly used framework here, which assumes that the relationship between each pair of outcome categories is the same. This makes ordinal logistic regression especially useful in survey analysis, Likert-scale data, and clinical severity ratings (Suthaharan, 2016).

2.1.7 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are a class of machine learning models inspired by the structure and function of the human brain. They consist of interconnected layers of artificial neurons that process information in a parallel and distributed manner. ANNs have become fundamental tools in modern artificial intelligence, especially in deep learning, where multi-layered networks achieve state-of-the-art performance in tasks such as image recognition, natural language processing, and predictive analytics.

Artificial Neuron (Perceptron)

The simplest form of an ANN is the perceptron, introduced by Rosenblatt (1958). A perceptron is a mathematical model of a biological neuron that performs binary classification. It takes multiple input signals, multiplies each with an associated weight, sums them, and applies an activation function to produce an output.

Mathematically, the perceptron can be expressed as (Cerulli, 2023):

$$y = 1\ if \sum w_j x_j + b > \theta;\ else\ 0 \tag{2.20}$$

where:

- x_j is the Input features
- w_j is the weights
- b is the bias

Structure of an Artificial Neural Network

An ANN is composed of layers of interconnected artificial neurons. The main components include:

Input Layer: This is the entry point of the data into the network. Each neuron in this layer represents one feature of the dataset. It does not perform any computation but simply forwards the input values to the next layer.

Hidden Layers: These layers contain neurons that apply weights, biases, and activation functions to the inputs. They are responsible for detecting and learning patterns in the data. Multiple hidden layers allow the network to capture complex relationships.

Output Layer: This layer produces the final prediction. The number of neurons depends on the nature of the task (e.g., one neuron for binary classification, multiple neurons with softmax activation for multi-class problems).

Weights: Numerical parameters that define the strength of connections between neurons. They are updated during training to minimize prediction errors.

Biases: Additional trainable parameters added to the weighted sum to allow more flexibility in learning complex patterns.

Activation Functions: Nonlinear transformations applied to the weighted input, enabling the network to learn complex, nonlinear decision boundaries (Zollanvari, 2023).

Activation Functions

Activation functions introduce non-linearity into the network, making it capable of learning complex patterns rather than simple linear relationships. They determine whether a neuron should be activated based on the input signal.

Commonly used activation functions include:

• *Sigmoid:* Maps input values into the range (0,1). Useful for probabilities but prone to vanishing gradients in deep networks.

$$F(x) = \frac{1}{(1 + e^{-x})} \tag{2.21}$$

• *Hyperbolic Tangent (tanh):* Outputs values in the range (-1,1). It is zero-centered, making optimization easier compared to sigmoid.

$$F(x) = \frac{(e^{x} + e^{-x})}{(e^{x} + e^{-x})} \tag{2.22}$$

• *Rectified Linear Unit (ReLU):* Outputs zero for negative inputs and the raw input for positive values. It is computationally efficient and reduces vanishing gradient issues.

$$F(x) = \max(0, x) \tag{2.23}$$

•*Linear:* Often used in regression problems. Produces continuous values without nonlinearity, making it suitable for outputs that are not categorical (Shalev-Shwartz, & Ben-David, 2014).

$$F(x) = x \tag{2.24}$$

Architectures of Artificial Neural Networks

ANN architectures define how neurons are connected. Common types include:

Single-Layer Feedforward Networks: The simplest form of ANN with one input and one output layer. Suitable for linear problems but limited in handling complex data.

Multi-Layer Perceptrons (MLPs): Contain one or more hidden layers between input and output. They are the foundation of deep learning and can approximate complex functions.

Recurrent Neural Networks (RNNs): Designed to process sequential data. Neurons have feedback loops that allow them to retain information from previous inputs, making them effective for tasks such as time series prediction and natural language processing.

Convolutional Neural Networks (CNNs): Primarily used for image and spatial data. They use convolutional layers to automatically extract local features such as edges, textures, and shapes, leading to high performance in image recognition.

Learning in ANNs

The training of an ANN involves adjusting weights and biases to minimize the error between predicted and actual outputs.

Forward Propagation: Forward propagation is the process of passing input data through the layers of a neural network to generate predictions. Each input is multiplied by weights, added to a bias, and passed through an activation function. The outputs are then propagated layer by layer until the final prediction is produced.

Backward Propagation (Backpropagation): Backward propagation is the process of updating weights and biases by propagating the error backwards from the output layer to earlier layers. It computes gradients using the chain rule and optimizes parameters using gradient descent. Purpose: To adjust weights and biases so that predictions get closer to the true values (James et al., 2021).

2.2 Cross-Validation

a) N-fold Cross-Validation

In n-fold cross-validation, the dataset is partitioned into n equally sized subsets, known as folds. The model is then trained and tested n times, with each fold serving once as the validation set while the remaining folds are used for training. This ensures that every observation is used for both training and validation exactly once. The results from all iterations yield a set of performance measures, which are averaged to provide a reliable estimate of the model's overall error (Bonaccorso, 2018).

b) Leave-One-Out

In Leave one out, the validation process is repeated as many times as there are observations in the dataset. At each iteration, a single observation is set aside as the validation set, while all remaining observations are used for training. The model is trained and tested repeatedly in this manner until every observation has served once as the validation instance. This procedure produces a set of error and accuracy values, which are

then averaged to provide an overall estimate of model performance. These aggregated results can be used to compare different models or determine whether additional retraining is required (Bonaccorso, 2018).

c) Leave-p-Out

In leave p out approach, the dataset is repeatedly divided into two parts: one portion is used for training and the other for validation. A fixed number of observations are randomly assigned to the validation set, while the remainder form the training set. This partitioning is carried out multiple times, creating several distinct training–validation pairs. Each model is trained and evaluated across all these pairs, producing a set of error and accuracy values. By averaging the results, an overall performance estimate is obtained, which can then be used to identify the better-performing model or to decide whether further retraining is necessary (Bonaccorso, 2018).

d) Random Subsampling

In the random subsampling method, the dataset is repeatedly divided into training and validation subsets through random selection. For each iteration, a portion of the data is randomly assigned to the training set, while the remainder forms the validation set. This random partitioning is carried out multiple times, generating several distinct training–validation pairs. The model is then trained and evaluated across all these pairs, and the results are aggregated to estimate overall performance (Bonaccorso, 2018).

2.3 Evaluation Metrics

To rigorously assess the classification models, several standard evaluation measures were computed: Accuracy, Precision, Recall, F1 Score, and the Area Under the ROC Curve (AUC). These metrics are based on the confusion matrix, which summarizes the predicted and actual class outcomes (Naidu, Zuva & Sibanda, 2023).

- **TP:** True Positives (correctly predicted high turnover intention)
- **FP:** False Positives (incorrectly predicted high turnover intention)
- **TN:** True Negatives (correctly predicted low turnover intention)

- **FN:** False Negatives (incorrectly predicted low turnover intention)

The metrics are defined as follows;

a) Accuracy

Represents the overall proportion of correct predictions:

$$Accuracy = (\text{TP} + \text{TN}) / (TP + TN + FP + FN) \quad (2.25)$$

b) Precision

Indicates the reliability of positive predictions:

$$Precision = TP/(TP + FP) \quad (2.26)$$

c) Recall (Sensitivity)

Measures the ability of the model to correctly identify positive cases:

$$Recall = TP/(TP + FN) \quad (2.27)$$

d) F1 Score

The harmonic mean of Precision and Recall, balancing both metrics:

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (2.28)$$

e) Area Under the ROC Curve (AUC)

A threshold-independent metric reflecting the model's discriminative ability. It is calculated as the area under the ROC curve, bounded by:

$$0.5 \leq \text{AUC} \leq 1 \quad (2.29)$$

CHAPTER THREE
APPLICATION

3.1 Aim

The aim of this study is to develop a model that can predict patient falls using various machine learning methods. With this model, potential patient falls that may occur in hospitals will be predicted with high accuracy. In this way, patient falls and the physical, psychological, and economic problems they cause can be prevented.

3.2 Study Population and Data

This retrospective study was conducted all adult in patient units in a University Hospital in Türkiye. The hospital has 1100 bed capacity. All adult inpatient units were included in the study. The study population consisted of adult patients (aged 18 years and older) who experienced at least one fall during their hospital stay, as well as adult patients who were hospitalized in the same unit and within a similar time frame but did not experience a fall. Ethics approval was obtained from the University Noninterventional Ethic Board, Türkiye (Reference 2023/15-05). Institutional approval was obtained from where the study was conducted.

In this study, a patient fall was defined as “an unplanned descent to the floor (or an extension of the floor, such as a trash can or other equipment), with or without injury to the patient” (NQF, 2011). Patients who experienced falls were identified retrospectively by reviewing incident reports archived by the hospital’s quality management department. Reports between 2017 and 2023 were reviewed to identify faller patients. If a patient experienced multiple falls during a single hospitalization, only the first event was included in the analysis. To construct and validate the machine learning models, approximately five control patients were randomly selected for each faller case. These control patients were hospitalized in the same clinical unit and during the same general time period (+/- 6 months), but did not experience any falls. This matching strategy was used to account for potential variations in infrastructure or fall-prevention protocols across different hospital units, which may influence the likelihood of patient falls.

3.3 Features

The variables included in this study were selected based on a review of the relevant literature and the research team's clinical expertise (Seppala et al., 2021; Clercq et al., 2021; Xu et al., 2022). The primary objective was to develop an automated fall risk assessment model using machine learning algorithms applied to EHR data. Consequently, the availability of structured data within the hospital information system was a key criterion for variable selection. A summary of all selected variables, their characteristics, and the workflows used for extracting these variables from the EHR is provided in Table 3.1.

Table 0.1 Features and definitions

Variable	**Definition**
Age	Indicates the patient's age at the time of hospitalization. It is a continuous variable.
Gender	Indicates the patient's gender. It is a categorical variable (Female/Male).
Length of Hospital Stay (Days)	Shows the total duration of the patient's hospital stay. It is a continuous variable, calculated using admission and discharge dates.
Unit	Indicates the hospital unit where the patient received inpatient care. Units are grouped as: medical, surgical, intensive care unit and emergency.
Visual Impairment	Indicates whether the patient has a visual impairment. (Categorical: 0 = No, 1 = Yes)
Balance Disorder	Indicates whether the patient has a balance disorder. (Categorical: 0 = No, 1 = Yes)
Dizziness	Indicates whether the patient experiences dizziness. (Categorical: 0 = No, 1 = Yes)
Incontinence	Indicates whether the patient has any incontinence problems. (Categorical: 0 = No, 1 = Yes)
Muscle Weakness	Indicates whether the patient has muscle weakness. (Categorical: 0 = No, 1 = Yes)

Table 3.1 Features and definitions - *continue*

Hypotension	Indicates whether the patient has hypotension. (Categorical: 0 = No, 1 = Yes)
Mental Disorder	Indicates whether the patient has a mental disorder. (Categorical: 0 = No, 1 = Yes)
Depression	Indicates whether the patient has depression. (Categorical: 0 = No, 1 = Yes)
Alzheimer's Disease	Indicates whether the patient has Alzheimer's disease. (Categorical: 0 = No, 1 = Yes)
Anemia	Indicates whether the patient has anemia. (Categorical: 0 = No, 1 = Yes)
Fluid-Electrolyte Imbalance	Indicates whether the patient has a fluid-electrolyte imbalance. (Categorical: 0 = No, 1 = Yes)
Osteoporosis	Indicates whether the patient has osteoporosis. (Categorical: 0 = No, 1 = Yes)
Movement Disorders	Indicates whether the patient has a movement disorder. (Categorical: 0 = No, 1 = Yes)
Parkinson's Disease	Indicates whether the patient has Parkinson's disease. (Categorical: 0 = No, 1 = Yes)
Total CCI	The total Charlson Comorbidity Index score calculated based on the patient's comorbidities. It is a continuous variable.
Fall	Indicates whether the patient experienced a fall during hospitalization. (Categorical: 0 = No, 1 = Yes; Target variable)
Use of High Fall Risk Medications	Indicates whether the patient used medications associated with a high fall risk (Categorical: 0 = No, 1 = Yes). For patients who fell, only the pre-fall period was examined; for non-fallers, the entire hospital stay was analyzed.
Number of High Fall Risk Medications Used	Indicates the total number of high fall risk medications used by the patient. It is a continuous variable. For patients who fell, only the pre-fall period was analyzed; for non-fallers, the entire hospital stay was examined.
Sodium	Serum sodium level. Continuous variable.
Creatinine	Serum creatinine level. Continuous variable.
Estimated glomerular filtration rate. (e-GFR)	Estimated glomerular filtration rate. Continuous variable.
Albumin	Serum albumin level. Continuous variable.
Neutrophils (NEU)%	Serum neutrophil percentage. Continuous variable.

Table 3.1 Features and definitions - *continue*

Hematocrit (HCT)	Proportion of red blood cells in blood volume. Continuous variable.
Mean Corpuscular Volume (MCV)	Mean corpuscular volume. Continuous variable.
Basophils (BASO)%	Serum basophil percentage. Continuous variable.
NEU#	Serum absolute neutrophil count. Continuous variable.
Lymphocytes (LYM)#	Serum lymphocyte count. Continuous variable.
Red Cell Distribution Width (RDW)	It shows the variation in the size of red blood cells. Continuous variable.
Eosinophils (EOS)#	It shows the number of eosinophils in the blood. Continuous variable.
Mean Platelet Volume (MPV)	It shows the average size of the platelets. Continuous variable.
Hemoglobin (HGB)	It shows the amount of hemoglobin in the blood. Continuous variable.
Plateletcrit (PCT)	It shows the proportion of platelets in the total blood volume. Continuous variable.
Red Blood Cell (RBC)	It shows the amount of red blood cells in the blood. Continuous variable.
Mean Corpuscular Hemoglobin (MCH)	It shows the average amount of hemoglobin in red blood cells. Continuous variable.
Monocytes (MONO)#	It shows the number of monocytes in the blood. Continuous variable.
EOS%	It shows the percentage of eosinophils in the blood. Continuous variable.
Mean Corpuscular Hemoglobin Concentration (MCHC)	It shows the hemoglobin concentration of red blood cells. Continuous variable.
MONO%	It shows the percentage of monocytes in the blood. Continuous variable.
BASO#	It shows the number of basophils in the blood. Continuous variable.
Blood Urea Nitrogen (BUN)	It shows the amount of blood urea nitrogen (BUN) in the blood. Continuous variable.
Potassium	It shows the amount of potassium in the blood. Continuous variable.

Table 3.1 Features and definitions - *continue*

Squamous epithelial cells	It shows the amount of squamous epithelial cells in the urine. Continuous variable.
Erythrose (Urine)	It shows the amount of red blood cells in the urine. Continuous variable.
pH (Urine)	It indicates the acidity of the urine. Continuous variable.
B12 Vitamin	It shows the level of vitamin B12 in the blood. Continuous variable.
Thyroid-Stimulating Hormone (TSH)	It shows the level of the TSH hormone. Continuous variable.

In the study, ICD codes were used to identify patients diagnosed with mental disorders, osteoporosis, Alzheimer's disease, and similar conditions. Laboratory results can provide important information for predicting falls. Accordingly, all laboratory results of the patients included in the study were incorporated into the dataset. For patients who experienced a fall, the laboratory and medication data closest to the fall date—among the values recorded before the fall—were used to establish a temporal relationship. If there were no available data before the fall, the closest value after the fall date was selected. For patients who did not experience a fall, if there were multiple values for the same laboratory parameter, the average of those values was used for each parameter.

It has previously been established that some medications can increase the risk of falls even when used alone (Seppälä et al., 2021). In line with this, information on the medications used by patients during their hospital stay was obtained from the hospital information system. The medications used by patients included in the study were grouped using the Anatomical Therapeutic Chemical (ATC) classification system developed by the WHO, which categorizes drugs based on their anatomical, therapeutic, and chemical properties. This approach allowed the identification of patients who used drug groups known in the literature to increase the risk of falling. In addition to identifying whether a patient used fall-risk-increasing drugs, the number of such medications was also calculated and included in the dataset as a separate variable.

In addition to the primary reason for hospitalization, patients' coexisting chronic conditions can also increase the risk of falling. The most commonly used approach to account for the burden of these comorbidities is the Charlson Comorbidity Index (CCI). The CCI includes 17 categories of chronic diseases, each weighted according to its severity level (Wang et al., 2023). Departments were categorized into 8 categories (Internal Medicine, Geriatric, Adult ICU, Surgical, Day treatment unit, Physical Therapy and Rehabilitation, Psychiatry, Oncology, Emergency).

3.4 Removing Correlated Variables

Within the scope of the study, some laboratory variables were found to be highly correlated with each other or represented different expressions (e.g., square, percentage) of the same biomarker. To prevent multicollinearity in the analyses, variables that either reflected the same value or showed strong correlations with one another were excluded from the dataset (Figure 3). Accordingly, the following variables were removed: BASO_Square, MONO_percentage, MCHC, PCT, EOS_percentage, HCT, LYM_percentage, NEU, NEU2, and MCH. In addition, the variables Alzheimer, Muscle Weakness, and Incontinence were excluded due to their very low frequency and limited potential to contribute meaningfully to model learning. Finally, the variable Depression was removed because the more comprehensive variable Mental Disorder was already included in the model (Table 4.1).

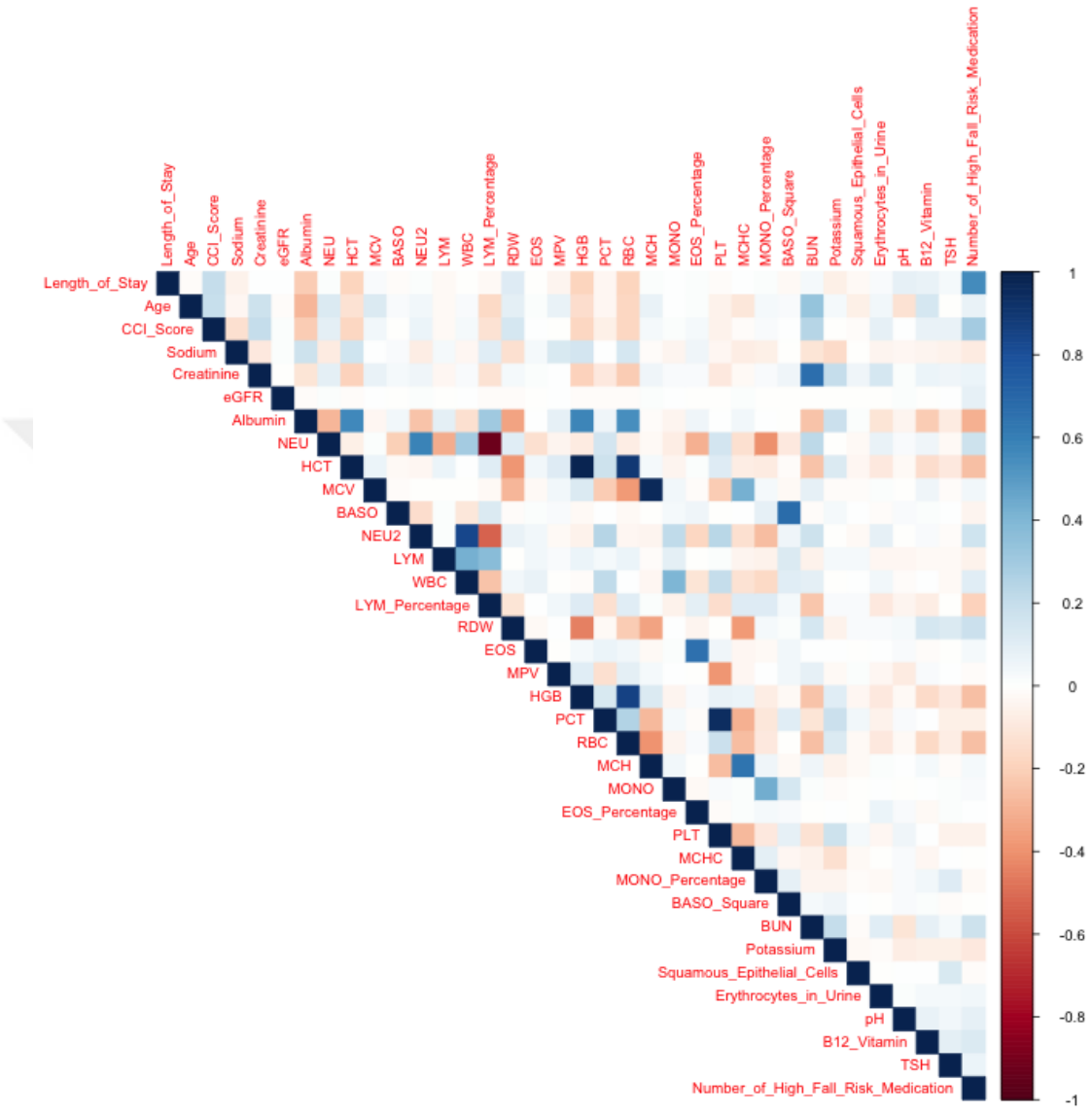


Figure 3.1 Corrplot of numeric variables

3.5 Missing Value Analysis

Descriptive analysis revealed that laboratory variables, in particular, had high rates of missing data. The percentage of missing values changed between 3.05 to 70.84 %. The number and percentage of missing values for each variable are presented in the Table 3.2. Seven variables with more than 15% missing data were excluded from the analysis (WBC, PLT, Squamous Epithelial Cells, RBC, pH, B12, TSH).

For the remaining 31 variables with lower rates of missingness, missing values were imputed using the Multiple Imputation method via Fully Conditional Specification. This approach is capable of modeling both continuous and categorical variables, capturing complex interrelationships among variables, and iteratively refining estimates for more consistent imputation. Moreover, by generating multiple completed datasets and pooling their results, this method enhances prediction accuracy and reduces bias (Lee & Carlin, 2010). The mice package was used to apply Multiple Imputation method via Fully Conditional Specification to handle missing values in the dataset (Lee Van Buuren & Groothuis-Oudshoorn, 2011).

Table 0.2 Variables and number of missing values

Variable	Number of Missing Values	Percentage of Missing Values
Length of Stay	0	0
Eye Disorder	0	0
Balance Disorder	0	0
Dizziness	0	0
Incontinence	0	0
Hypotension	0	0
Mental Disorder	0	0
Gender	0	0
Age	0	0
Anemia	0	0
Fluid and Electrolyte Imbalance	0	0
Osteoporosis	0	0
Movement Disorder	0	0
Parkinson	0	0
Unit	0	0
Total CCI	0	0

Table 3.2 Variables and missing values - *continue*

Sodium	168	4.61665293
Creatinine	163	4.47925254
e-GFR (CKD-EPI)	255	7.00741962
Albumin	429	11.788953
NEU%	132	3.62737016
MCV	111	3.05028854
BASO%	132	3.62737016
LYM#	132	3.62737016
WBC	1137	31.2448475
RDW	112	3.07776862
EOS#	131	3.59989008
MPV	111	3.05028854
HGB	111	3.05028854
RBC	111	3.05028854
MONO#	131	3.59989008
PLT	1137	31.2448475
BASO#	131	3.59989008
BUN	177	4.86397362
Potasium	182	5.001374
Squamous Epithelial Cells	1473	40.4781533
RBC	1451	39.8735916
pH	1442	39.626271
B12 Vitamin	2578	70.8436384
TSH	2294	63.0392965
Fall Status	0	0
High Fall Risk Medicine Usage	46	1.26408354
Number of High Fall Risk Medicine	46	1.26408354

3.6 Data Scaling

Variables such as age, length of hospital stay, total CCI score, and laboratory parameters in the dataset exhibited wide variation in value ranges. When variables differ substantially in scale, distance-based machine learning algorithms may give disproportionate weight to variables with larger numeric ranges, overshadowing those with smaller values (Ahsan et al., 2021).

Data scaling ensures that all variables contribute equally to the classification process. In this study, numerical variables were scaled using Z-score standardization. This method involves subtracting the mean of each variable from its value and dividing by the standard deviation. As a result, all variables are transformed to have a mean of 0 and a standard deviation of 1, while preserving their original distribution (de Amorim, Cavalcanti & Cruz, 2023).

3.7 Resampling

Statistical models often exhibit a tendency to predict the majority class, especially when the class distribution is imbalanced. As a result, the model's performance in predicting the minority class tends to be suboptimal (He and Garcia, 2009).

In the study dataset, the number of patients without falls was 6.02 times higher than the number of patients with falls. Therefore, the target group to be predicted represents the minority class. Since the dependent variable in the study was the minority class, random oversampling methods were employed to create a balanced dataset.

The primary goal of random oversampling is to increase the number of samples in the minority class. This is achieved by randomly selecting instances from the minority class and duplicating them within the training dataset. However, this approach may lead to overfitting, particularly in high-dimensional or mixed-type datasets. To mitigate this risk and effectively handle datasets that include both continuous and categorical variables,

the Synthetic Minority Over-sampling TEchnique–Nominal Continuous (SMOTE-NC) method was used to balance the training data.

Synthetic Minority Over-sampling Technique (SMOTE)

Developed by Chawla et al. in 2002, SMOTE is an oversampling method. In its simplest definition, SMOTE attempts to prevent overfitting by generating synthetic examples of the minority class in the dataset. SMOTE creates artificial minority samples along the line segments connecting the minority class instances and their 'k' nearest minority neighbors. It uses linear interpolation to generate synthetic training examples for the minority class. The steps of the method are as follows (Dey & Pratap 2023):

1. A random subset of minority class examples is selected.
2. For each example in the selected subset, its k-nearest neighbors are identified.
3. The distance between each example and its selected neighbor is calculated.
4. One of the k-nearest neighbors is chosen. The distance is computed as: $\|X - X_k\|$, where X is the instance and X_k is the chosen neighbor.
5. The product of the distance and a randomly chosen number between 0 and 1 is calculated.
6. This product is added to the existing example to generate a new synthetic sample X_new as follows: $X_{new} = X + rand(0,1) \times \|X - X_k\|$

The main purpose of the SMOTE method is to increase the number of instances in the minority class. In doing so, it ensures that the generated synthetic data are not identical to the existing examples while not being too different from them either. While the SMOTE algorithm performs successfully on datasets consisting of numerical variables, it does not achieve the same success on datasets containing both numeric and categorical variables. Therefore, the SMOTE-NC method was developed for datasets containing both numerical and categorical variables

Synthetic Minority Over-sampling Technique–Nominal Continuous (SMOTE-NC)

Since the SMOTE method cannot be applied to datasets containing categorical variables, Chawla et al. (2002) developed the Synthetic Minority Over-sampling

Technique–Nominal Continuous (SMOTE-NC) as an alternative. SMOTE-NC is an extension of the previously developed SMOTE method. The steps of the method are as follows:

1. The median of the standard deviations of the continuous features belonging to the minority class is calculated. If there is a difference in nominal (categorical) features between a sample and its potential nearest neighbors, this difference is included in the Euclidean distance and the calculated median is used to penalize this difference.

2. The Euclidean distance between a minority class sample and other samples is calculated using continuous features. For each differing nominal feature, the previously calculated median is added to the Euclidean distance.

3. To create a new synthetic sample, continuous features are determined using linear interpolation between two examples. For nominal features, the most frequently occurring value among the k-nearest neighbors is selected and assigned to the synthetic sample.

SMOTE-NC treats nominal variables differently from continuous variables and preserves the properties of categorical variables in the resampled data. However, it does not achieve successful results in cases with highly multi-class nominal variables (Mukherjee & Khushi, 2021).

3.8 Model Training and Prediction

For model development, the dataset was randomly split into a training set (75%) and an independent test set (25%). The training set was used to fit the models, while the test set was reserved exclusively for evaluating predictive performance. Hyperparameter tuning was conducted within the training set using a rigorous 10-fold cross-validation procedure to optimize model performance and reduce overfitting.

Study workflow is summarized in Figure 3.2.

3.9 Data Analysis

Continuous socio-demographic and occupational variables were summarized using means and standard deviations, while categorical variables were presented as frequencies and percentages. The distributional properties of continuous variables were evaluated using the Anderson–Darling normality test. Since all continuous variables deviated significantly from normality, comparisons between faller and non-faller groups for continuous variables were conducted using the Mann–Whitney U test. Categorical variables were compared using Pearson's chi-square test when expected cell counts were sufficient; otherwise, Fisher's exact test was applied. All statistical tests were two-sided, and a p-value < 0.05 was considered statistically significant.

All statistical analyses and supervised machine learning procedures were performed using R software (version 4.4.0). Data preprocessing and manipulation were conducted using the "dplyr" package. Model training and hyperparameter tuning were implemented within a cross-validation framework using the caret package. To address class imbalance in the outcome variable, oversampling (SMOTE-NC) was performed using the "themis" package. Supervised machine learning models were implemented using the following packages: "kernlab" for linear and nonlinear support vector machine kernels, "randomForest" for random forest models, "xgboost" for extreme gradient boosting, tree for decision tree models, h2o for artificial neural networks, and "randomForestSRC" for the random forest quantile classifier. Logistic regression analyses were performed using the "stats" package. Data visualization was carried out using the "ggplot2" package.

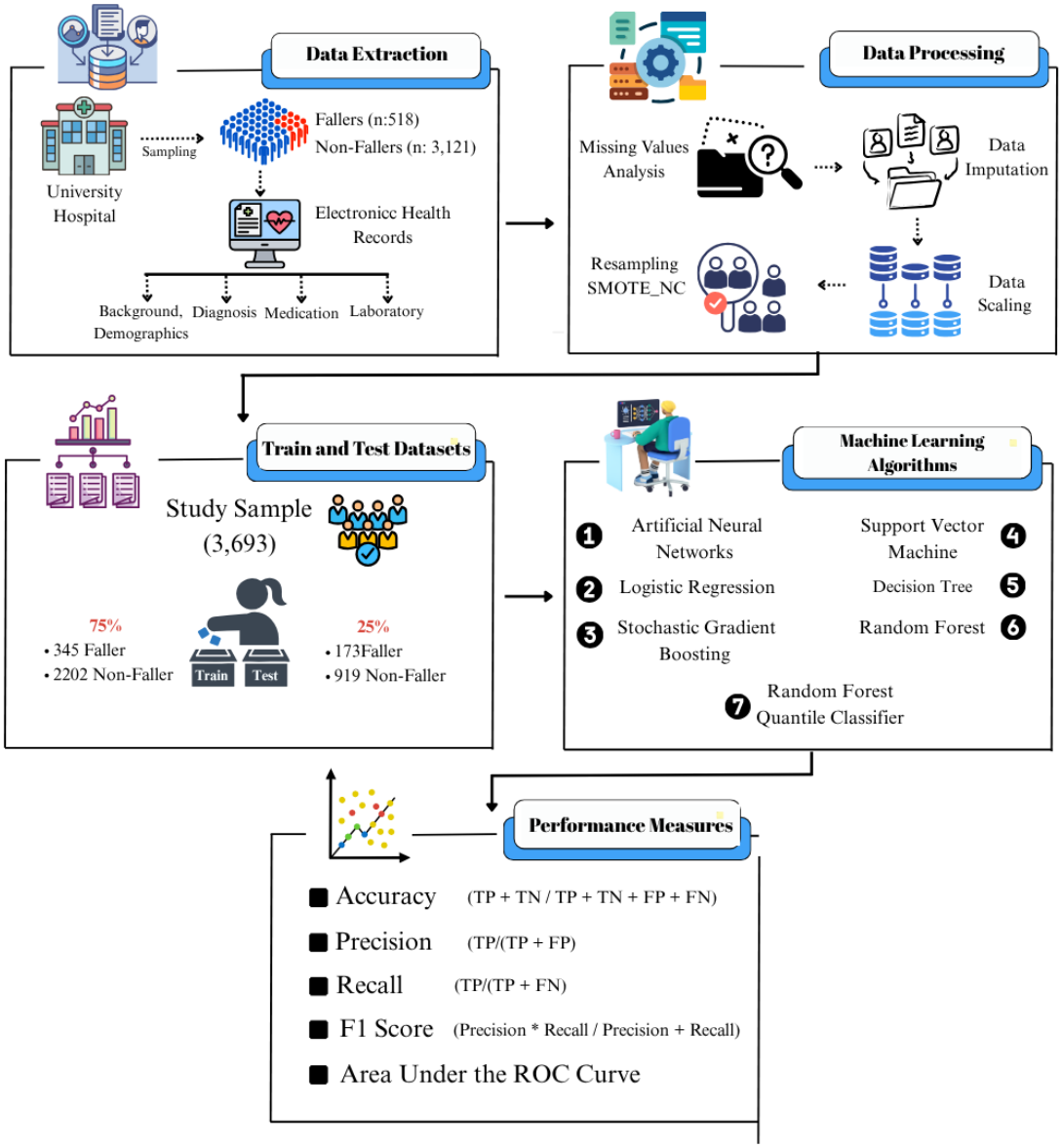


Figure 3.2 Study workflow

CHAPTER FOUR
RESULTS

4.1 Descriptives of Study Participants

The study cohort comprised 3,639 adult inpatients, of whom 518 (14.2%) experienced at least one fall, while 3,121 (85.8%) did not, as identified through incident-reporting and electronic health record systems. Compared with non-fallers, patients who experienced a fall were significantly older (median age 68.5 [IQR 18] vs. 64 [IQR 24], $p < 0.001$) and had longer hospital stays (median 16 [IQR 21] vs. 12 [IQR 16] days, $p < 0.001$). Although the median CCI was 1 in both groups, fallers demonstrated a higher overall comorbidity burden ($p < 0.001$). Fallers were prescribed a greater number of high fall-risk medications (median 3 [IQR 5] vs. 2 [IQR 3], $p < 0.001$), and a higher proportion of fallers received at least one high fall-risk medication compared with non-fallers (86% vs. 76%, $p < 0.001$). Male sex was more common among fallers than non-fallers (57.3% vs. 50.5%, $p = 0.004$). Regarding clinical characteristics, balance disorders (18% vs. 12%, $p < 0.001$), movement disorders (3.2% vs. 1.6%, $p = 0.023$), fluid and electrolyte imbalances (7.7% vs. 5.0%, $p = 0.017$), and mental disorders (26.3% vs. 17.2%, $p < 0.001$) were significantly more prevalent among fallers. Other conditions, including anemia, osteoporosis, Parkinson's disease, eye disorders, Alzheimer's disease, dizziness, depression, hypotension, muscle weakness, and incontinence, did not differ significantly between groups.

Several laboratory parameters showed statistically significant but clinically modest differences. Fallers had lower median sodium levels (137 [IQR 6] vs. 138 [IQR 4.1], $p = 0.001$) and lower estimated glomerular filtration rates (eGFR) (83.4 [IQR 51.7] vs. 88.3 [IQR 44.7], $p = 0.001$), alongside higher creatinine (0.84 [IQR 0.66] vs. 0.79 [IQR 0.49], $p = 0.014$) and blood urea nitrogen (BUN) levels (21 [IQR 18.7] vs. 18.8 [IQR 14.3], $p < 0.001$). Hematological differences included lower hemoglobin (10.7 [IQR 2.9] vs. 10.9 [IQR 3.19], $p = 0.035$), hematocrit (32.2 [IQR 8.6] vs. 33.1 [IQR 9.32], $p = 0.012$), and red blood cell counts (3.74 [IQR 1.05] vs. 3.9 [IQR 1.10], $p = 0.001$) among fallers.

Fallers also demonstrated lower lymphocyte counts and percentages ($p < 0.001$ for both), higher neutrophil percentages ($p = 0.012$), and small but significant differences in basophil and eosinophil parameters. Mean corpuscular hemoglobin (MCH) and mean corpuscular hemoglobin concentration (MCHC) were slightly higher among fallers, while mean corpuscular volume (MCV) did not reach statistical significance ($p = 0.052$). Platelet count and most platelet indices did not differ significantly between groups.

Other laboratory measures, including albumin, white blood cell count, potassium, vitamin B12, thyroid-stimulating hormone, and urine pH, showed either no clinically meaningful differences or small statistically significant variations.

Table 0.1 Comparisons of faller and non-faller patients

Variable	**Faller (n=518)**	**Non-Faller (n=3121)**	**p-value**
Age, median, (IQR)	68.5 (18)	64 (24)	<0.001
Length of Stay, median, (IQR)	16 (21)	12 (16)	<0.001
Total CCI, median, (IQR)	1 (2)	1 (1)	<0.001
Number of High Fall Risk Medicine, median, (IQR)	3 (5)	2 (3)	<0.001
High Fall Risk Medicine Usage, n (%)	445 (%86)	2335 (%76)	<0.001
Gender (Male), n (%)	1577 (%50.5)	297 (%57.3)	0.004
Anemia, n (%)	18 (%3.4)	104 (%3.3)	0.971
Osteoporosis, n (%)	6 (%1.1)	64 (%2)	0.231
Eye Disorder, n (%)	18 (%3.4)	87 (%2.8)	0.469
Parkinson, n (%)	10 (%2)	41 (%1)	0.365
Movement Disorder, n (%)	17 (%3.2)	53 (%1.6)	0.023
Balance Disorder, n (%)	93 (%18)	377 (%12)	<0.001
Fluid and Electrolyte Imbalance, n (%)	40 (%7.72)	158 (%5)	0.017

Table 4.1 Comparisons of faller and non-faller patients - *continue*

Alzheimer, n (%)*	1 (%0.1)	2 (%0.06)	0.369
Dizzinies, n (%)	7 (%0.7)	23 (%1.3)	0.242
Depression, n (%)*	40(%7.7)	214 (%6.8)	0.533
Mental Disorder, n (%)	136 (%26.3)	536 (%17.2)	<0.001
Hipotension , n (%)	2 (%0.3)	48 (%1.5)	0.059
Muscle Weakness, n (%)*	1 (%0.1)	1 (%0.03)	0.264
Inkontinence , n (%)*	5 (%0.9)	11 (%0.3)	0.111
Sodyum, median, (IQR)	137 (6)	138 (4.10)	0.001<
Kreatinin, median, (IQR)	0.84 (0.66)	0.793 (0.490)	0.014
estimated Glomerular Filtration Rate (eGFR), median, (IQR)	83.4 (51.7)	88.3 (44.7)	0.001
Unit			
• Medical Unit	351 (14.04)	2149 (85.96)	0.001
• Surgical Unit	127 (12.89)	858 (87.11)	
• Adult ICU	34 (28.33)	86 (71.67)	
• Emergency	6 (17.65)	28 (82.35)	
Albumin, median, (IQR)	3.24 (0.90)	3.29 (0.95)	0.151
NEU#, median, (IQR)*	71.3 (20.8)	70.1 (16.8)	0.012
HCT, median, (IQR)*	32.2 (8.6)	33.1 (9.32)	0.012
MCV, median, (IQR)	87.2 (7.40)	86.8 (7.80)	0.052
BASO#, median, (IQR)	0.4 (0.4)	0.412 (0.309)	0.001<
NEU^2, median, (IQR)*	5.6 (4.6)	5.8 (3.88)	0.314
LYM, median, (IQR)	1.2 (1.2)	1.43 (1.05)	0.001<
White Blood Cell (WBC), median, (IQR)*	8.4 (5.4)	8.29 (4.18)	0.789
LYM%, median, (IQR)*	16.7 (17.4)	18.9 (14.6)	0.001<
RDW, median, (IQR)	15.7 (3.4)	15.3 (3.4)	0.063
EOS, median, (IQR)	0.1 (0.2)	0.1 (0.16)	0.328

Table 4.1 Comparisons of faller and non-faller patients - *continue*

MPV, median, (IQR)	8.6 (1.6)	8.5 (1.38)	0.701
HGB, median, (IQR)	10.7 (2.9)	10.9 (3.19)	0.035
PCT, median, (IQR)*	0.20 (0.10)	0.20 (0.09)	0.348
RBC, median, (IQR)*	3.74 (1.05)	3.9 (1.10)	0.001
MCH, median, (IQR)*	29 (2.83)	28.9 (2.93)	0.021
MONO#, median, (IQR)	0.6 (0.5)	0.6 (0.35)	0.969
EOS%, median, (IQR)*	1 (2.3)	1.45 (2)	0.001<
PLT, median, (IQR)*	232 (148)	240 (128)	0.328
MCHC, median, (IQR)*	33.2 (1.30)	33.2 (1.09)	0.036
MONO%, median, (IQR)*	7.6 (4.4)	7.60 (3.42)	0.735
$BASO^2$, median, (IQR)*	0 (0)	0.009 (0.04)	0.001<
BUN, median, (IQR)	21 (18.7)	18.8 (4.3)	0.001
Potasyum, median, (IQR)	4.13 (0.8)	4.15 (0.54)	0.205
Squamous Epithelial Cells, median, (IQR)*	0 (1)	0 (1.18)	0.029
pH, median, (IQR)*	5.5 (1)	5.5 (1.33)	0.004
B12 Vitamin, median, (IQR)*	306 (325)	303 (283)	0.930
TSH, median, (IQR)*	1.32 (1.86)	1.4 (1.61)	0.670
Eritrosit RBC, median, (IQR)	1 (5)	2 (10.5)	0.019

* Removed

4.2 Predictive Performance of Models

Table 4.2 presents the discrimination and classification performance of the candidate algorithms prior to the application of SMOTE-NC. Overall, the results indicate substantial variability across models, with several algorithms exhibiting imbalanced performance between sensitivity and specificity.

The SVM achieved the highest AUC (0.720) among the evaluated models, with high precision (0.867) and perfect recall (1.000), resulting in an F1-score of 0.929.

However, this performance was accompanied by a specificity of 0.000, indicating that the model classified all observations as positive and failed to correctly identify any negative cases. The Decision Tree classifier also demonstrated strong sensitivity, achieving a recall of 0.993, precision of 0.857, and an F1-score of 0.920, with an overall accuracy of 0.853. Nevertheless, its specificity was relatively low at 0.250, suggesting limited discrimination of negative cases despite high positive-class detection. Logistic Regression produced a more balanced performance, with an AUC of 0.604, precision of 0.903, recall of 0.563, and F1-score of 0.693. Its specificity (0.601) and accuracy (0.568) indicate moderate classification capability across both classes. In contrast, the Random Forest algorithm yielded a lower AUC of 0.635, with modest recall (0.276) and F1-score (0.418), but comparatively higher specificity (0.713), reflecting a tendency toward negative-class prediction. XGBoost recorded an AUC of 0.573, with high precision (0.913) but low recall (0.177), resulting in a low F1-score of 0.297. Despite a high specificity of 0.888, its overall accuracy remained limited at 0.270. The ANN showed the weakest discriminative performance, with an AUC of 0.507, precision of 0.315, recall of 0.468, F1-score of 0.377, specificity of 0.781, and accuracy of 0.740.

Table 0.2 Performance measures of models before SMOTE_NC

Algorithms	**AUC**	**Precision**	**Recall**	**F1 Score**	**Specificity**	**Accuracy**
SVM*	0.720	0.867	1.00	0.929	0.00	0.867
Random Forrest	0.635	0.864	0.276	0.418	0.713	0.333
XGBoost	0.573	0.913	0.177	0.297	0.888	0.270
Logistic Regression	0.604	0.903	0.563	0.693	0.601	0.568
Decision Tree	0.687	0.857	0.993	0.920	0.250	0.853
ANN	0.507	0.315	0.468	0.377	0.781	0.740

* With linear kernel

Table 4.3 summarizes the discrimination and classification metrics for the seven candidate algorithms. The Random Forest Quantile Classifier demonstrated the strongest overall performance, with an AUC of 0.863, precision of 0.446, recall of 0.806, F1-score of 0.540, specificity of 0.805, and accuracy of 0.863. The conventional Random Forest algorithm achieved an AUC of 0.784, precision of 0.859, recall of 0.991, F1-score of 0.920, and accuracy of 0.856; however, its specificity was notably low at 0.138. Support Vector Machine yielded an AUC of 0.660, precision of 0.914, recall of 0.574, F1-score of 0.705, specificity of 0.677, and accuracy of 0.588. Logistic Regression produced an AUC of 0.656, precision of 0.203, recall of 0.670, F1-score of 0.311, specificity of 0.564, and accuracy of 0.579. Artificial Neural Network reported an AUC of 0.665, precision of 0.239, recall of 0.503, F1-score of 0.365, specificity of 0.735, and accuracy of 0.703. XGBoost yielded an AUC of 0.543, precision of 0.160, recall of 0.980, F1-score of 0.274, specificity of 0.040, and accuracy of 0.177. The Decision Tree classifier recorded an AUC of 0.674, recall of 0.000, specificity of 1.000, and accuracy of 0.141; as no positive cases were predicted, precision and F1-score were not defined. The ROC curves and AUC values for each algorithm are presented in Figure 4.1.

Table 0.3 Performance measures of models after SMOTE-NC

Algorithms	**AUC**	**Precision**	**Recall**	**F1 Score**	**Specificity**	**Accuracy**
SVM*	0.660	0.914	0.574	0.705	0.677	0.588
Random Forrest	0.784	0.859	0.991	0.920	0.138	0.856
XGBoost	0.543	0.160	0.980	0.274	0.04	0.177
Logistic Regression	0.656	0.203	0.670	0.311	0.564	0.579
Decision Tree	0.674	NA	0.00	NA	1.000	0.141
ANN	0.665	0.239	0.503	0.365	0.735	0.703
RFQ	**0.863**	**0.446**	**0.806**	**0.540**	**0.805**	**0.863**

* With linear kernel

Collectively, the results in Table 4.2 highlight the challenges posed by class imbalance prior to SMOTE-NC, with several models exhibiting either excessively high sensitivity at the expense of specificity or vice versa. These findings provide a baseline for comparison and underscore the motivation for applying SMOTE-NC to improve balanced classification performance, as reflected in the results presented in Table 4.3.

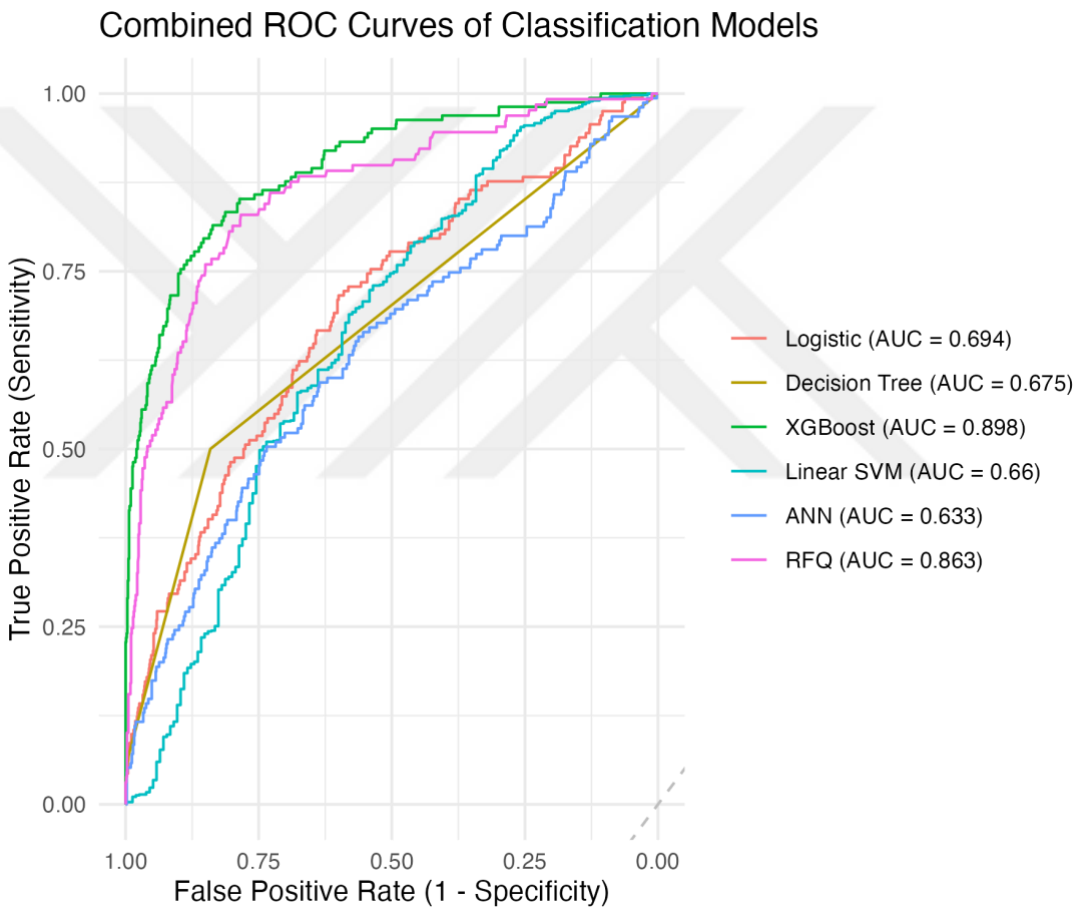


Figure 4.1 ROC curves and AUC values of models

4.3 Feature Importance

Feature importance was calculated using the Random Forest Quantile Classifier algorithm to identify the most influential variables contributing to model predictions. The top-ranked variable was the total Charlson Comorbidity Index (CCI) score, followed by

eosinophil count (EOS), age, and the number of high fall-risk medications used. Other high-ranking features included the number of high fall-risk medications, basophil count (BASO), sodium, blood urea nitrogen (BUN), and red blood cell count (RBC).

Several additional laboratory parameters, including hemoglobin (HGB), mean platelet volume (MPV), potassium, and albumin, also demonstrated moderate importance in the model. Non-laboratory features such as length of stay, medical unit type, and surgical unit type were included among the top 20 but had comparatively lower importance scores. The top 20 important variables and their relative importance values are provided in Figure 4.2.

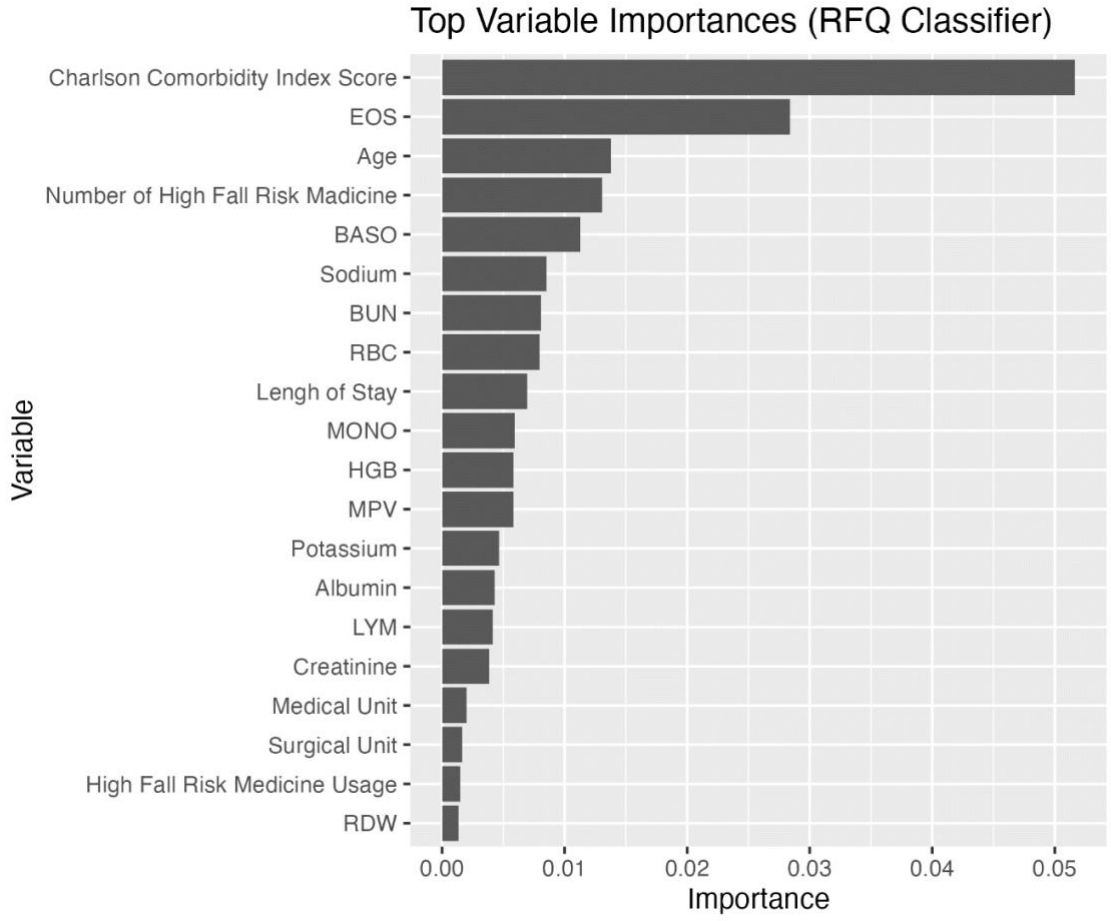


Figure 4.2 Importance matrix of variables

CHAPTER FIVE
CONCLUSION

This research designed and validated a machine learning framework to anticipate inpatient falls using structured data from electronic health records in a tertiary hospital in Türkiye. Fall incidents are relatively uncommon in hospital settings, which creates a pronounced imbalance between outcome classes and complicates the use of machine learning techniques. To mitigate this, prior studies frequently employed oversampling or undersampling strategies such as SMOTE (Ladios-Martin et al., 2022). Following similar practices, we applied SMOTE-NC, a variant designed to handle mixed data types containing both categorical and continuous features (Chawla et al., 2002). In addition, we tested an alternative approach using the RFQ, a method not previously reported in the context of fall risk prediction.

We constructed seven different machine learning models to estimate the likelihood of inpatient falls. Among these, the Random Forest Quantile Classifier (RFQ) demonstrated the most favorable overall performance (AUC: 0.863; accuracy: 0.863; recall: 0.806; specificity: 0.805), achieving a well-balanced trade-off between sensitivity and specificity. This balance is particularly critical in the context of fall prevention, where both false-negative and false-positive predictions carry important clinical consequences. False-negative classifications may result in missed opportunities for preventive interventions, potentially leading to avoidable patient harm, whereas excessive false-positive classifications may label a large proportion of patients as high risk, leading to inefficient use of preventive resources, alarm fatigue, and diminished trust among healthcare providers in predictive decision-support tools.

In comparison with models trained using SMOTE-NC, RFQ consistently demonstrated superior discriminatory performance without relying on synthetic data generation. While oversampling techniques may improve sensitivity, they can also inflate false-positive rates and increase the risk of overfitting by introducing artificial

observations. In contrast, RFQ addresses class imbalance by adjusting classification thresholds through a quantile-based decision rule that explicitly accounts for the true prevalence of rare events. This approach allows the full dataset to be utilized without synthetic augmentation and enables more clinically meaningful risk stratification. The ability of RFQ to maintain high sensitivity while preserving specificity underscores its suitability for rare-event prediction tasks such as inpatient falls, where sustainable clinical adoption depends on both predictive accuracy and operational feasibility.

When compared with established fall risk assessment instruments, the RFQ achieved substantially higher predictive power, suggesting its capacity to outperform traditional screening tools for identifying high-risk patients (Strini et al., 2021). Furthermore, because it relies solely on routinely available EHR data, it may alleviate nursing workload by eliminating additional manual assessments (Dormosh, Schut, Heymans, van der Velde & Abu-Hanna, 2023).

To uncover the most influential predictors of inpatient falls, we examined feature importance within the RFQ model. The Charlson Comorbidity Index emerged as the strongest predictor, followed by eosinophil count (EOS), age, and the number of high fall-risk medications. These findings emphasize both well-recognized and underappreciated determinants of fall risk. The prominence of CCI is consistent with prior evidence linking multimorbidity to increased fall likelihood via polypharmacy, physical decline, and reduced physiological resilience (Wang et al., 2023). Age, another established risk factor, also ranked among the top contributors, reinforcing its key role in fall susceptibility (Jehu et al., 2021).

Several laboratory variables, including EOS, basophil count, BUN, sodium, hemoglobin, and RBC count also appeared among the top-ranked predictors. Although absent from most conventional fall risk tools, these measures may reflect underlying conditions such as inflammation, anemia, or renal dysfunction, which can impair balance, mobility, or cognition. For instance, hyponatremia is known to cause gait disturbances,

while anemia can lead to orthostatic hypotension and fatigue. Consistent with our findings, incorporating laboratory biomarkers has been shown to improve model performance in other ML studies (Kim et al., 2025; Dormosh et al., 2023). Interestingly, some biomarkers did not differ significantly between fallers and non-fallers in univariate tests, yet were highly ranked in the RFQ's feature importance. This discrepancy illustrates the contrast between traditional statistical approaches, which evaluate predictors in isolation, and machine learning models, which account for multivariate and nonlinear interactions. A feature may have limited individual predictive power but gain substantial importance when combined with others. Unlike p-values, RFQ's importance scores reflect a variable's contribution to overall classification performance, enabling the detection of subtle patterns that conventional analyses often miss.

Taken together, these results reinforce that fall risk is a multifaceted phenomenon influenced by clinical, demographic, and biochemical factors. Machine learning models, particularly RFQ, are well-suited to capture these complex interdependencies and provide advantages over conventional screening approaches.

REFERENCES

Agency for Healthcare Research and Quality (AHRQ). National scorecard on hospital-acquired conditions: content last reviewed November. Rockville MD, Agency for Healthcare Research and Quality, 2019; 1-25.

Ahsan, M. M., Mahmud, M. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, 9(3), 52.

An, Q., Rahman, S., Zhou, J., & Kang, J. J. (2023). A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. *Sensors*, *23*(9), 4178.

Anand, V., & Selvaganesan, C. (2025). Enhancing Ransomware Detection-A Comparative Review of XGBoost, Random Forest, and Neural Network Approaches. In *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC)* (pp. 710-715). IEEE.

Ayyadevara, V. K. (2018). *Pro machine learning algorithms*. Apress: Berkeley, CA, USA, 283-297.

Bargiotas, I., Wang, D., Mantilla, J., Quijoux, F., Moreau, A., Vidal, C., ... & Vidal, P. P. (2023). Preventing falls: the use of machine learning for the prediction of future falls in individuals without history of fall. *Journal of Neurology*, *270*(2), 618-631.

Baris, V.K., Seren Intepeler S., & Yeginboy EY. (2018). The cost of serious patient fall-related injuries at hospitals in Turkey: a matched case-control study, *Clinical Nursing Research, 27*(2):162-179.

Baris, V.K., & Seren Intepeler S. (2019). Views of key stakeholders on the causes of patient falls and prevention interventions: A qualitative study using the international classification of functioning, disability and health. *Journal of Clinical Nursing, 28*(3-4):615-628.

Bernet, N. S., Everink, I. H., Hahn, S., Müller, M., & Schols, J. M. (2024). Risk-adjusted trend in national inpatient fall rates observed from 2011 to 2019 in acute care hospitals in Switzerland: a repeated multicentre cross-sectional study. *BMJ open*, *14*(5), e082417.

Bonaccorso, G. (2018). *Machine Learning Algorithms: Popular algorithms for data science and machine learning*. Packt Publishing Ltd.

Cameron, I. D., Dyer, S. M., Panagoda, C. E., Murray, G. R., Hill, K. D., Cumming, R. G., & Kerse, N. (2018). Interventions for preventing falls in older people in care facilities and hospitals. Cochrane database of systematic reviews, (9).

Cerulli, G. (2023). *Fundamentals of supervised machine learning: With applications in Python, R, and Stata* (1st ed.). Springer.

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20-28.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Chu, W. M., Kristiani, E., Wang, Y. C., Lin, Y. R., Lin, S. Y., Chan, W. C., ... & Tsan, Y. T. (2022). A model for predicting fall risks of hospitalized elderly in Taiwan-A machine learning approach based on both electronic health records and comprehensive geriatric assessment. *Frontiers in Medicine*, *9*, 937216.

De Amorim, L. B., Cavalcanti, G. D., & Cruz, R. M. (2023). The choice of scaling technique matters for classification performance. *Applied Soft Computing*, *133*, 109924.

De Clercq, H., Naudé, A., & Bornman, J. (2021). Factors included in adult fall risk assessment tools (FRATs): A systematic review. Ageing & Society, 41(11), 2558-2582.

Dey, I., & Pratap, V. (2023, March). A comparative study of SMOTE, borderline-SMOTE, and ADASYN oversampling techniques using different classifiers. In 2023 3rd international conference on smart data intelligence (ICSMDI) (pp. 294-302). IEEE.

Dormosh, N., Schut, M. C., Heymans, M. W., van der Velde, N., & Abu-Hanna, A. (2022). Development and internal validation of a risk prediction model for falls among older people using primary care electronic health records. *The Journals of Gerontology: Series A*, *77*(7), 1438-1445.

Dykes, P. C., Duckworth, M., Cunningham, S., Dubois, S., Driscoll, M., Feliciano, Z., ... & Scanlan, M. (2017). Pilot testing fall TIPS (tailoring interventions for patient safety): a patient-centered fall prevention toolkit. *The Joint Commission Journal on Quality and Patient Safety, 43*(8), 403-413.

García-Hedrera, F. J., Noguera-Quijada, C., Sanz-Márquez, S., Pérez-Fernández, E., Acevedo-García, M., Domínguez-Rincón, R., ... & Camacho-Pastor, J. L. (2021). Incidence and characteristics of falls in hospitalized patients: A cohort study. *Enfermería Clínica (English Edition)*, *31*(6), 381-389.

Graham, B.C., (2012). Examining evidence-based interventions to prevent inpatient falls, *Medical Surgical Nursing,21*(5):267-270.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.

Hignett, S., Sands, G., & Griffiths, P. (2013). In-patient falls: what can we learn from incident reports?, *Age and Aging, 42*:527-531.

Ishikuro, M., Ubeda, S. R. G., Obara, T., Saga, T., Tanaka, N., Oikawa, C., & Fujimori, K. (2017). Exploring risk factors of patient falls: a retrospective hospital record study in Japan. *The Tohoku Journal Of Experimental Medicine*, *243*(3), 195-203.

The Joint Commission International-JCI. (2025, December). Sentinel event data 2024 annual review.

https://digitalassets.jointcommission.org/api/public/content/eac7511986c0442a9c1ae04b1aa02cc0?v=ad34daa0

Jahandideh, S., Hutchinson, A. F., Bucknall, T. K., Considine, J., Driscoll, A., Manias, E., ... & Hutchinson, A. M. (2024). Using machine learning models to predict falls in hospitalised adults. *International Journal of Medical Informatics*, *187*, 105436.

Ishwaran, H., Robert, O. B., Lu, M., & Kogalur, U. B. (2023, November). *randomForestSRC: Random Forests Quantile Classifier (RFQ) Vignette*.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer.

Kakara, R. (2023). Nonfatal and fatal falls among adults aged≥ 65 years—United States, 2020–2021. *MMWR. Morbidity and Mortality Weekly Report*, *72*.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). Errors in health care: a leading cause of death and injury. In To err is human: Building a safer health system. National Academies Press (US).

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*(1), 3-24.

Ladios-Martin, M., Cabañero-Martínez, M. J., Fernández-de-Maya, J., Ballesta-López, F. J., Belso-Garzas, A., Zamora-Aznar, F. M., & Cabrero-Garcia, J. (2022). Development of a predictive inpatient falls risk model using machine learning. *Journal of Nursing Management*, *30*(8), 3777-3786.

Lee, K. J., & Carlin, J. B. (2010). Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *American Journal of Epidemiology*, *171*(5), 624-632.

Lee, K. B., Lee, J. S., Jeon, I. P., Choo, D. Y., Baik, M. J., Kim, E. H., ... & Kim, J. S. (2021). An analysis of fall incidence rate and risk factors in an inpatient rehabilitation unit: a retrospective study. *Topics in Stroke Rehabilitation*, *28*(2), 81-87.

Li, Y., Hou, L., Zhao, H., Xie, R., Yi, Y., & Ding, X. (2023). Risk factors for falls among community-dwelling older adults: a systematic review and meta-analysis. *Frontiers in Medicine*, *9*, 1019094.

Liu, C. H., Hu, Y. H., & Lin, Y. H. (2021). A machine learning–based fall risk assessment model for inpatients. CIN: Computers, Informatics, Nursing, 39(8), 450-459.

MacKay, S., Ebert, P., Harbidge, C., & Hogan, D. B. (2021). Fear of falling in older adults: a scoping review of recent literature. *Canadian Geriatrics Journal*, *24*(4), 379.

Makary, M.A. & Daniel, M. (2016). Medical error-the third leading cause of death in the US. *BMJ, 353*:i2139.

Moe, K., Brockopp, D., McCowan, D., Merritt, S., & Hall, B. (2015). Major predictors of inpatient falls: a multisite study. *JONA: The Journal of Nursing Administration*, *45*(10), 498-502.

Morello, R. T., Barker, A. L., Watts, J. J., Haines, T., Zavarsek, S. S., Hill, K. D., ... & Stoelwinder, J. U. (2015). The extra resource burden of in-hospital falls: a cost of falls study. *Medical journal of Australia*, *203*(9), 367-367.

Montero-Odasso, M., Van Der Velde, N., Martin, F. C., Petrovic, M., Tan, M. P., Ryg, J., ... & Masud, T. (2022). World guidelines for falls prevention and management for older adults: a global initiative. *Age and ageing, 51*(9), afac205.

Morris, M. E., Webster, K., Jones, C., Hill, A. M., Haines, T., McPhail, S., ... & Cameron, I. (2022). Interventions to reduce falls in hospitals: a systematic review and meta-analysis. *Age and Ageing*, *51*(5), afac077.

Mukherjee, M., & Khushi, M. (2021). SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features. *Applied System Innovation*, 4(1), 18.

Naidu, G., Zuva, T., & Sibanda, E. M. (2023, April). A review of evaluation metrics in machine learning algorithms. In R. Silhavy (Ed.), *Computer Science On-line Conference* (pp. 15–25). Springer International Publishing.

Najafpour, Z., Godarzi, Z., Arab, M., & Yaseri, M. (2019). Risk factors for falls in hospital in-patients: a prospective nested case control study. *International Journal of Health Policy and Management*, *8*(5), 300.

Naseralallah, L., Stewart, D., Azfar Ali, R., & Paudyal, V. (2022). An umbrella review of systematic reviews on contributory factors to medication errors in health-care settings. *Expert Opinion on Drug Safety*, *21*(11), 1379-1399.

National Quality Forum (NQF), Serious Reportable Events In Healthcare—2011 Update: A Consensus Report, Washington, DC: NQF; 2011.

O'Brien, R., & Ishwaran, H. (2019). A random forests quantile classifier for class imbalanced data. *Pattern Recognition*, 90, 232-249.

Organisation for Economic Co-operation and Development (OECD). *Measuring patient safety: opening the black box*. Organisation for Economic Co-operation and Development, 2018;4-14.

Park, S. H. (2018). Tools for assessing fall risk in the elderly: a systematic review and meta-analysis, *Aging Clinical and Experimental Research, 30*(1), 1-16.

Ray, S. (2019, February). A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*(pp. 35-39). IEEE.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.

Royal College of Physicians (RCP). *National Audit of Inpatient Falls: audit report 2017*. London: Royal College of Physicians, 2017;13-22.

Rowe, R.J., (2013). Preventing patient falls: what are the factors in hospital settings that help reduce and prevent inpatient falls?, *Home Health Care Management & Practice, 25*(3), 98–103.

Seppala, L. J., Petrovic, M., Ryg, J., Bahat, G., Topinkova, E., Szczerbińska, K., ... & Van der Velde, N. (2021). STOPPFall (screening tool of older persons prescriptions in older adults with high fall risk): a Delphi study by the EuGMS task and finish group on fall-risk-increasing drugs. *Age and Ageing*, 50(4), 1189-1199.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.

Shao, L., Shi, Y., Xie, X. Y., Wang, Z., Wang, Z. A., & Zhang, J. E. (2023). Incidence and risk factors of falls among older people in nursing homes: systematic review and meta-analysis. *Journal of the American Medical Directors Association*, *24*(11), 1708-1717.

Shim, S., Yu, J. Y., Jekal, S., Song, Y. J., Moon, K. T., Lee, J. H., ... & Hong, J. H. (2022). Development and validation of interpretable machine learning models for inpatient fall events and electronic medical record integration. *Clinical and Experimental Emergency Medicine*, *9*(4), 345.

Spetz, J., Brown, D. S., & Aydin, C. (2015). The economics of preventing hospital falls: demonstrating ROI through a simple model. *JONA: The Journal of Nursing Administration*, *45*(1), 50-57.

Staggs, V. S., Mion, L. C., & Shorr, R. I. (2014). Assisted and unassisted falls: different events, different outcomes, different implications for quality of hospital care. *The Joint Commission Journal on Quality and Patient Safety*, *40*(8), 358-364.

Strini, V., Schiavolin, R., & Prendin, A. (2021). Fall risk assessment scales: A systematic literature review. *Nursing Reports*, *11*(2), 430-443.

Suthaharan, S. (2016) Logistic regression. In: Machine learning models and algorithms for big data classification, vol 36. Springer, pp 207–235

Xu, Q., Ou, X., & Li, J. (2022). The risk of falls among the aging population: a systematic review and meta-analysis. *Frontiers in Public Health*, 10, 902599.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, *45*, 1-67.

Wang, C., Zhang, X., Li, B., & Mu, D. (2023). A study of factors impacting disease based on the Charlson Comorbidity Index in UK Biobank. *Frontiers in Public Health*, 10, 1050129.

Wong, C. A., Recktenwald, A. J., Jones, M. L., Waterman, B. M., Bollini, M. L., & Dunagan, W. C. (2011). The cost of serious fall-related injuries at three Midwestern hospitals. *The Joint Commission Journal on Quality and Patient Safety*, *37*(2), 81-87.

World Health Organization (2025, December). WHO Global Report on Falls Prevention in Older Age, https://www.who.int/publications/i/item/9789241563536

World Health Organization. (2025, November). Patient safety [Fact sheet]. https://www.who.int/news-room/fact-sheets/detail/patient-safety

Zollanvari, A. (2023). *Machine learning with Python: Theory and implementation*. Springer Nature.

Wang, L., Xue, Z., Ezeana, C. F., Puppala, M., Chen, S., Danforth, R. L., ... & Wong, S. T. (2019). Preventing inpatient falls with injuries using integrative machine learning prediction: a cohort study. *NPJ digital medicine, 2*(1), 127.

Young, A. J., Hare, A., Subramanian, M., Weaver, J. L., Kaufman, E., & Sims, C. (2021). Using machine learning to make predictions in patients who fall. *Journal of Surgical Research, 257*, 118-127.