

Final Project of : Research Problem and Research Objective for

Developing a Predictive Model for Health Insurance Claim Acceptance and Denial in Medical Centers Using Cost-Sensitive XGBoost and Probability

Course Title: **Research Methodology in Computer Science**

Course Code: **AI800**

Weighting: **50%**

Lecturer: **Dr. Mohammed Anbar**

Hand-out Date: **15-Nov-2025**

Hand-in Date: **20-Dec-2025**

Hand-in Method: **Online Submission through E-Learning.**

Submitted by : Hussam Ahmed Almokayed

Student ID : 253001038

Table of Content

Contents

Course Title:	Research Methodology in Computer Science	1
Course Code:	AI800	1
Weighting:	50%	1
Lecturer:	Dr. Mohammed Anbar	1
Hand-out Date:	15-Nov-2025	1
Hand-in Date:	20-Dec-2025	1
Hand-in Method:	Online Submission through E-Learning.	1
Abstract		3
Introduction:		4
Background		5
Problem Statement		7
Research Objectives:		8
Research Scope		9
Literature review		10
Research Methodology		14
Ethical Considerations		16
Limitations		16
Expected Outcomes		17
Gantt Chart (Proposed Timeline)		17
Reference		18

Abstract

Health insurance claim adjudication remains a major administrative challenge within the Saudi healthcare system, particularly in private medical centers where manual procedures, operational inefficiencies, and exposure to fraud-related risks continue to affect both decision quality and processing speed (Ho et al., 2024). Although digital transformation initiatives have improved data availability and reporting consistency, existing claim adjudication systems still struggle to deliver accurate, reliable, and timely decisions (Al-Quayed et al., 2023).

This research aims to address these challenges by proposing a cost-sensitive predictive model based on XGBoost, enhanced with probability calibration techniques to improve the reliability of claim acceptance and denial decisions. The proposed approach targets key limitations observed in current machine learning applications, including inadequate probability calibration and limited consideration of the financial consequences of misclassification errors (Permai & Herdianto, 2023; Toffaha, 2024). By explicitly incorporating asymmetric misclassification costs, the model seeks to reduce operational losses and support more informed decision-making in health insurance claim processing.

The study builds upon existing literature and evaluates the proposed model using Saudi-specific health insurance claim data. The expected outcomes include improved fraud detection, faster claim processing, and increased trust in automated decision-support systems, supporting the broader digital transformation objectives of Saudi Vision 2030 (Somajohassula, 2025)

Introduction:

The health insurance sector plays a central role in supporting healthcare delivery in Saudi Arabia, particularly through mandatory insurance schemes designed to ensure broader access to medical services. As part of the country's ongoing healthcare modernization efforts, digital transformation has become a strategic priority aimed at improving efficiency, transparency, and service quality across the insurance ecosystem. Despite these initiatives, the adjudication of health insurance claims continues to represent a significant administrative challenge, especially within private medical centers (Ho et al., 2024).

Traditional claim adjudication practices remain heavily dependent on manual reviews and rigid rule-based systems. These approaches often lead to delayed decisions, inconsistent claim outcomes, and increased exposure to fraudulent or unjustified claims, placing substantial operational and financial pressure on both insurers and healthcare providers (Ho et al., 2024; Johnson et al., 2023). Such inefficiencies do not only result in financial losses but also affect patient access to timely care and strain relationships between medical centers and insurance companies.

Recent advances in machine learning have demonstrated strong potential for improving insurance analytics and supporting automated decision-making processes. Models such as gradient boosting techniques have shown superior predictive performance compared to traditional statistical methods (Permai & Herdianto, 2023). However, many existing machine learning applications in health insurance fail to adequately account for the financial impact of misclassification errors. In practice, denying a legitimate claim and approving a fraudulent one

carry fundamentally different consequences, highlighting the need for predictive models that explicitly incorporate cost considerations (Brati et al., 2025; Toffaha, 2024).

Addressing these limitations is essential for developing reliable and trustworthy decision-support systems that can enhance claim adjudication efficiency while aligning with Saudi Arabia's broader digital transformation objectives and healthcare strategy under Vision 2030 (Somajohassula, 2025).

Background

Health Insurance Claim Adjudication in Saudi Arabia

The health insurance system in Saudi Arabia has undergone substantial reforms aimed at expanding coverage, improving service quality, and ensuring financial sustainability. Mandatory health insurance schemes, regulated by the Council of Cooperative Health Insurance (CCHI), have played a central role in organizing relationships between insurers, healthcare providers, and beneficiaries. As part of these reforms, digital transformation initiatives have been introduced to enhance data standardization, transparency, and operational efficiency. Despite these efforts, health insurance claim adjudication remains one of the most complex and resource-intensive administrative processes, particularly within private medical centers (Ho et al., 2024).

Operational Challenges and Data Standardization

Private medical centers continue to face operational challenges related to claim processing, including overbilling, duplicate submissions, prolonged investigation cycles, and delayed reimbursements. These issues contribute to increased administrative workloads and financial losses for both insurers and healthcare providers, ultimately affecting the sustainability of healthcare services. To address inconsistencies in clinical and administrative reporting, Saudi Arabia introduced the Minimum Data Set (MDS), which standardized essential claim-related information across healthcare institutions (Al-Hamad et al., 2025). While the MDS has improved data quality and interoperability, it functions primarily as a descriptive reporting framework and does not provide predictive or automated decision-support capabilities required for proactive claim management.

Machine Learning in Health Insurance Analytics

Globally, machine learning techniques have been increasingly adopted to enhance insurance analytics and overcome the limitations of traditional rule-based systems. Gradient boosting models, particularly XGBoost, have demonstrated superior performance in predicting claim denials, investigations, and payment delays compared to conventional logistic regression approaches (Permai & Herdianto, 2023; Brati et al., 2025). These models offer strong classification accuracy and the ability to handle complex, high-dimensional datasets commonly found in insurance environments.

However, despite their predictive strength, many machine learning models applied to health insurance claims generate poorly calibrated probability estimates. This limitation reduces their practical usefulness in real-world decision-making, where accurate and trustworthy risk estimation is critical. Poor probability calibration can lead to misguided financial decisions, especially in high-stakes environments such as health insurance, where the consequences of misclassification vary significantly (Toffaha, 2024).

Limitations of Existing Solutions and Research Gap

Within the Saudi context, emerging InsurTech initiatives and artificial intelligence–based architectures have been proposed to modernize claim adjudication systems. Although these approaches highlight the potential of automation and advanced analytics, many remain conceptual or lack sufficient adaptation to local regulatory requirements, data standards, and healthcare workflows (Al-Quayed et al., 2023; Somajohassula, 2025). As a result, there is a clear gap between global advancements in machine learning–based insurance analytics and their effective deployment within Saudi private medical centers.

This gap underscores the need for a Saudi-specific predictive framework that integrates cost-sensitive learning and reliable probability calibration. Such a framework is essential for improving the accuracy, efficiency, and trustworthiness of health insurance claim adjudication systems, while supporting the broader digital transformation objectives of Saudi Vision 2030.

Problem Statement

This research falls within the domain of health insurance claim adjudication and predictive analytics in Saudi private medical centers, where reliable, data-driven decision methods are increasingly needed (Al-Hamad et al., 2025).

In Saudi Arabia, processing health insurance claims remains one of the most challenging administrative tasks. Manual reviews and rigid rule-based systems lead to slow decisions, inconsistent approvals, heavier workloads, and a greater risk of undetected fraud (Ho et al., 2024). Although digital documentation has improved, claim processing has not kept pace; existing systems still lack modern machine-learning capabilities and fail to consider the financial impact of misclassification, such as approving fraudulent claims or rejecting legitimate ones (Toffaha, 2024).

Improving how claims are processed is essential because any delay or error directly affects patient access to care, increases operational pressures on healthcare providers, and undermines trust in the overall system. A smarter, more accurate decision-support model is therefore critical for maintaining healthcare quality and economic efficiency in Saudi Arabia.

Multiple attempts have been made to enhance claim processing. The XGBoost model and neural networks outperform traditional methods but still lack reliable probability outputs for real-world decision-making (Permai & Herdianto, 2023; Ho et al., 2024). Locally, the Minimum Data Set (MDS) improved data quality but remains a static framework without predictive or automated capabilities (Al-Hamad et al., 2025). InsurTech concepts involving AI and blockchain show promise but remain largely theoretical and not adapted to Saudi healthcare workflows (Somajohassula, 2025). As a result, current solutions still fall short of achieving accurate prediction, cost sensitivity, and operational practicality.

Despite digital advances, key gaps persist:

1. **Poor probability calibration** in current ML models (Toffaha, 2024).
2. **Lack of cost-sensitive decision-making**, even though misclassification carries major financial consequences (Somajohassula, 2025).
3. **No Saudi-specific predictive model** tailored to local claim patterns and regulations.
4. **Underutilization of the MDS**, which lacks predictive analytics (Al-Hamad et al., 2025).

These gaps justify the development of a cost-sensitive XGBoost model with calibrated probability outputs—one capable of delivering faster, fairer, and more reliable claim adjudication aligned with the digital transformation goals of Saudi Vision 2030.

Research Objectives:

This Research aims to develop a cost-sensitive predictive model using XGBoost, supporting by probability calibration methods, to enhance the accuracy and efficiency of health insurance claim acceptance and denial specially in private medical centers in Saudi Arabia.

- 1- To analyze the existing challenges in the current claim adjudication process.
 - To identify the operational, administrative, and fraud-related problems affecting claim decisions in Saudi healthcare settings, based on findings that highlight delays, high costs, and inconsistent processing (Al-Hamad et al., 2025).
- 2- To evaluate current machine learning techniques to use claim prediction
 - To examine the performance of models such as logistic regression, XGBoost, and neural networks in predicting denial, investigation, and delay outcomes (Permai & Herdianto, 2023).
 - To solve the gaps between poor probability calibration and the lack of cost-sensitive modeling (Toffaha, 2024)
- 3- To develop a cost-sensitive XgBoost model adapted to Saudi health insurance data
 - To build the model using related medical, financial, and demographic features consistent with Saudi claim structures (Al-Hamad et al., 2025)
- 4- To apply probability calibration techniques to improve model reliability.
 - To test calibration methods such as Platt scaling and isotonic regression and evaluate how they enhance the trustworthiness of predicted probabilities (Permai & Herdianto, 2023)
- 5- **To validate the proposed model using Saudi claim datasets.**

- To focus on testing the model on real or simulated Saudi claim data to confirm its reliability. also validation will use metrics such as precision, recall, F1-score, AUROC, and calibration curves to assess the model's accuracy and practical suitability. (Toffaha, 2024)

Research Scope

This research focuses on health insurance claim adjudication in private medical centers within Saudi Arabia. The study examines claim acceptance and denial decisions within the regulatory and operational framework governing private healthcare insurance in the Kingdom.

The research utilizes anonymized or simulated health insurance claim data that comply with data governance requirements and the regulations of the Council of Cooperative Health Insurance (CCHI). The analysis is limited to binary claim outcomes and relevant medical, financial, and demographic attributes commonly observed in Saudi insurance datasets.

To ensure a focused scope, claims from the public healthcare sector and real-time system deployment are excluded. The study is restricted to machine learning-based predictive modeling and probability calibration techniques. Model evaluation is conducted using standard performance metrics to assess predictive accuracy and the reliability of probability estimates.

Literature review

Ho et al. (2024): Determinants of Claims Investigation, Delay Payment, and Claim Denial in Private Health Insurance

Ho et al. (2024) conducted a large-scale empirical study examining the factors that influence claim investigation, payment delays, and claim denials within private health insurance systems. Using a dataset comprising more than half a million insurance claims, the authors applied multiple machine learning models, including logistic regression and neural networks, to identify key predictors of claim outcomes. Their findings demonstrated that advanced machine learning models significantly outperform traditional statistical approaches in predicting claim denials and investigations. The study highlights the operational value of predictive analytics in improving claim processing efficiency and fraud detection. However, the research focuses primarily on predictive accuracy and does not incorporate cost-sensitive learning, despite the significant financial implications associated with misclassification errors. This limitation restricts the applicability of the model in real-world insurance environments where different errors carry unequal consequences.

Al-Hamad et al. (2025): Minimum Data Set (MDS) in Saudi Arabia's Medical Claims Insurance

Al-Hamad et al. (2025) explored the implementation of the Minimum Data Set (MDS) within Saudi Arabia's medical claims insurance framework. The study emphasizes the role of standardized clinical and administrative data in improving claim accuracy, interoperability, and fraud reduction. By enforcing structured reporting across healthcare providers, the MDS contributed to enhanced transparency and consistency in claim submissions. Despite these improvements, the authors acknowledge that the MDS functions as a static data framework and does not incorporate predictive analytics or automated decision-making mechanisms. Consequently, while the study provides a strong foundation for data quality in Saudi health insurance systems, it highlights a gap in leveraging this standardized data for machine learning-based claim adjudication, which the current research seeks to address.

Toffaha (2024): Using Machine Learning Approaches for Risk Prediction and Stratification in Healthcare

Toffaha (2024) investigated the application of machine learning models, including XGBoost, for risk prediction and stratification in healthcare settings. The study demonstrated that ensemble

learning models achieve high predictive performance when applied to complex healthcare datasets. A key contribution of this work is its emphasis on probability calibration, highlighting that raw probability outputs from machine learning models are often unreliable for real-world decision-making. Although the study focuses on clinical risk prediction rather than insurance claims, its findings regarding calibration techniques are highly relevant to claim adjudication systems. The research underscores the importance of calibrated probabilities for trustworthy decision support, a principle directly adopted in the proposed model of this study.

Somajohassula (2025): Transforming Healthcare Claims Adjudication through InsurTech Innovations

Somajohassula (2025) examined the role of InsurTech innovations, including artificial intelligence and blockchain technologies, in modernizing healthcare claim adjudication. The study argues that automation and decentralized architectures can significantly reduce administrative costs and improve transparency in insurance workflows. While the paper presents a compelling vision for the future of claim adjudication, its primary limitation lies in the lack of empirical validation. The proposed solutions remain largely conceptual and are not evaluated using real-world datasets, particularly within region-specific regulatory contexts. This highlights the need for applied, data-driven research that translates InsurTech concepts into practical and validated predictive models, especially in the Saudi healthcare environment.

Al-Quayed et al. (2023): A Secure Technology-Based Healthcare Architecture for Smart Health Insurance Systems

Al-Quayed et al. (2023) proposed a secure, technology-driven architecture for smart health insurance systems, integrating machine learning, cloud computing, and blockchain technologies to enhance fraud detection and system security. The study emphasizes the importance of secure data sharing and system interoperability in insurance ecosystems. While the proposed architecture improves transparency and security, it does not explicitly address cost-sensitive prediction or probability calibration. As a result, the model's ability to support nuanced claim acceptance and denial decisions remains limited. This research complements the current study by highlighting architectural considerations while reinforcing the need for cost-aware predictive decision models.

Permai and Herdianto (2023): Prediction of Health Insurance Claims Using Logistic Regression and XGBoost Methods

Permai and Herdianto (2023) compared logistic regression and XGBoost models for health insurance claim prediction, demonstrating the superior predictive performance of XGBoost in fraud detection tasks. Their findings confirm the effectiveness of ensemble learning methods in handling complex insurance datasets. However, the study primarily evaluates model performance based on classification accuracy and does not consider probability calibration or the financial cost of misclassification. This limitation restricts the practical usability of the model in operational insurance settings, where decision-makers require reliable probability estimates and cost-aware predictions.

Brati et al. (2025): Machine Learning Applications for Predicting High-Cost Claims Using Insurance Data

Brati et al. (2025) investigated the use of machine learning algorithms to predict high-cost insurance claims. The study demonstrated that ensemble models such as Random Forest and XGBoost outperform traditional approaches in identifying high-risk claims. The authors emphasized the importance of feature selection and risk assessment in insurance analytics. Although the study advances the understanding of cost-related claim prediction, it does not explicitly integrate cost-sensitive learning into the modeling process, nor does it address probability calibration. These gaps highlight opportunities for enhancing model reliability and economic relevance.

Owolabi (2025): Transforming Appeal Decisions through Machine Learning Triage

Owolabi (2025) focused on the application of machine learning models to support hospital admission appeal decisions. The study demonstrated that predictive models can improve decision consistency and reduce administrative burden in utilization management. A notable contribution of this research is its emphasis on fairness and bias reduction in automated decision-making. However, the study relies on traditional regularized models and does not explore ensemble techniques or calibrated probability outputs. The findings reinforce the need for trustworthy predictive models that balance accuracy, fairness, and operational practicality.

Alibrahim et al. (2025): Patterns of Claims and Determinants of Claim Rejections in Kuwait's National Health Insurance

Alibrahim et al. (2025) analyzed millions of insurance claims within Kuwait’s national health insurance system to identify patterns and determinants of claim rejection. The study revealed that demographic factors, service types, and administrative procedures significantly influence claim outcomes. Although conducted outside Saudi Arabia, the findings provide valuable regional insights applicable to Gulf healthcare systems. The authors emphasize the importance of standardized processes and data-driven decision-making, supporting the relevance of developing region-specific predictive models tailored to local regulations and claim patterns.

Johnson et al. (2023): Responsible Artificial Intelligence in Healthcare Insurance

Johnson et al. (2023) proposed a responsible artificial intelligence framework for predicting and preventing insurance claim denials. The study integrates ethical considerations with machine learning models to improve both economic and social outcomes. While the research advances the discussion on responsible AI, it does not explicitly address probability calibration or cost-sensitive learning in depth. Nevertheless, the study reinforces the importance of trust, transparency, and fairness in automated insurance decision-support systems.

Summary and Research Gap

Collectively, the reviewed studies demonstrate the effectiveness of machine learning models—particularly XGBoost—in improving health insurance claim prediction. However, several limitations persist across the literature. Most studies prioritize predictive accuracy while overlooking cost-sensitive learning and probability calibration. In addition, Saudi-specific implementations that integrate standardized data frameworks such as the MDS with advanced predictive analytics remain limited. These gaps justify the development of a cost-sensitive XGBoost model with calibrated probability outputs tailored to health insurance claim adjudication in Saudi private medical centers.

Research Methodology

This research adopts a quantitative, model-driven methodology to address the challenges of health insurance claim adjudication in Saudi private medical centers. The methodology is designed to achieve the stated research objectives by systematically analyzing existing literature, preparing and modeling insurance claim data, and validating the proposed predictive framework. The research process is structured into four main phases, as illustrated below.

Phase 1: Literature Review and Problem Analysis

The first phase focuses on conducting a comprehensive review of existing literature related to health insurance claim adjudication, machine learning–based claim prediction, cost-sensitive learning, and probability calibration techniques. Academic journals, conference proceedings, and recent studies published between 2019 and 2025 are reviewed to identify existing methodologies, limitations, and research gaps.

This phase aims to:

- Understand current practices in claim adjudication and predictive analytics.
- Identify limitations related to cost insensitivity, probability miscalibration, and lack of Saudi-specific adaptations.
- Establish a theoretical foundation that supports the development of the proposed model.

The findings from this phase directly inform the design of the research methodology and the selection of modeling techniques.

Phase 2: Data Collection and Preprocessing

In the second phase, anonymized or simulated health insurance claim datasets relevant to the Saudi private healthcare sector are utilized. The datasets are structured in accordance with local data governance requirements and the Council of Cooperative Health Insurance (CCHI) regulations.

This phase includes:

- Selection of relevant medical, financial, and demographic features.
- Data cleaning and preprocessing to handle missing values and inconsistencies.

- Addressing data imbalance between accepted and denied claims through appropriate preprocessing strategies.

This phase ensures that the data is suitable for predictive modeling and reflects real-world claim patterns observed in Saudi medical centers.

Phase 3: Model Development and Calibration

The third phase involves developing the predictive model using a cost-sensitive XGBoost algorithm. Cost-sensitive learning is applied to explicitly account for the asymmetric financial impact of misclassification errors, such as approving fraudulent claims or denying legitimate ones.

In addition, probability calibration techniques, including Platt Scaling and Isotonic Regression, are applied to improve the reliability of predicted probabilities. This phase aims to ensure that the model outputs are not only accurate in classification but also trustworthy for decision-support purposes.

Phase 4: Model Evaluation and Validation

The final phase focuses on evaluating and validating the proposed model using Saudi health insurance claim data. Model performance is assessed using standard evaluation metrics, including Precision, Recall, F1-score, AUROC, and calibration curves.

This phase aims to:

- Compare the calibrated cost-sensitive model with baseline machine learning models.
- Assess the effectiveness of probability calibration in improving decision reliability.
- Validate the practical applicability of the proposed framework within private medical centers.

The outcomes of this phase provide empirical evidence of the model's effectiveness and its potential contribution to improving claim adjudication efficiency.

Ethical Considerations

Ethical considerations are a fundamental component of this research, particularly due to the sensitive nature of health insurance and medical claim data. To ensure ethical compliance, all data used in this study are anonymized or simulated, with no personally identifiable information related to patients, healthcare providers, or insurance beneficiaries being collected or processed. This approach aligns with local data governance policies and the regulations issued by the Council of Cooperative Health Insurance (CCHI).

The research is conducted strictly for academic and analytical purposes, with no intention of influencing real-time claim decisions or replacing human judgment in operational settings. The proposed predictive model is designed as a decision-support tool rather than an automated decision-maker, ensuring that final claim decisions remain under human oversight. Furthermore, care is taken to avoid algorithmic bias by adopting balanced evaluation strategies and emphasizing transparency and reliability in probability outputs. These measures aim to promote responsible use of machine learning techniques and maintain trust among stakeholders involved in the health insurance ecosystem.

Limitations

Despite the structured methodology adopted in this research, several limitations are acknowledged. First, the study relies on anonymized or simulated datasets that reflect Saudi private medical center claim patterns. While this approach ensures ethical compliance and data accessibility, it may not fully capture all complexities present in real-world operational environments.

Second, the scope of the research is limited to binary claim outcomes—acceptance and denial—excluding intermediate outcomes such as claim investigation or delay. Additionally, the study focuses exclusively on private medical centers, and therefore the findings may not be directly generalizable to public healthcare institutions or other insurance systems outside Saudi Arabia.

Finally, the research emphasizes machine learning–based predictive modeling and probability calibration techniques. Other advanced approaches, such as real-time system deployment, deep learning architectures, or integration with live insurance platforms, are beyond the scope of this study and represent opportunities for future research.

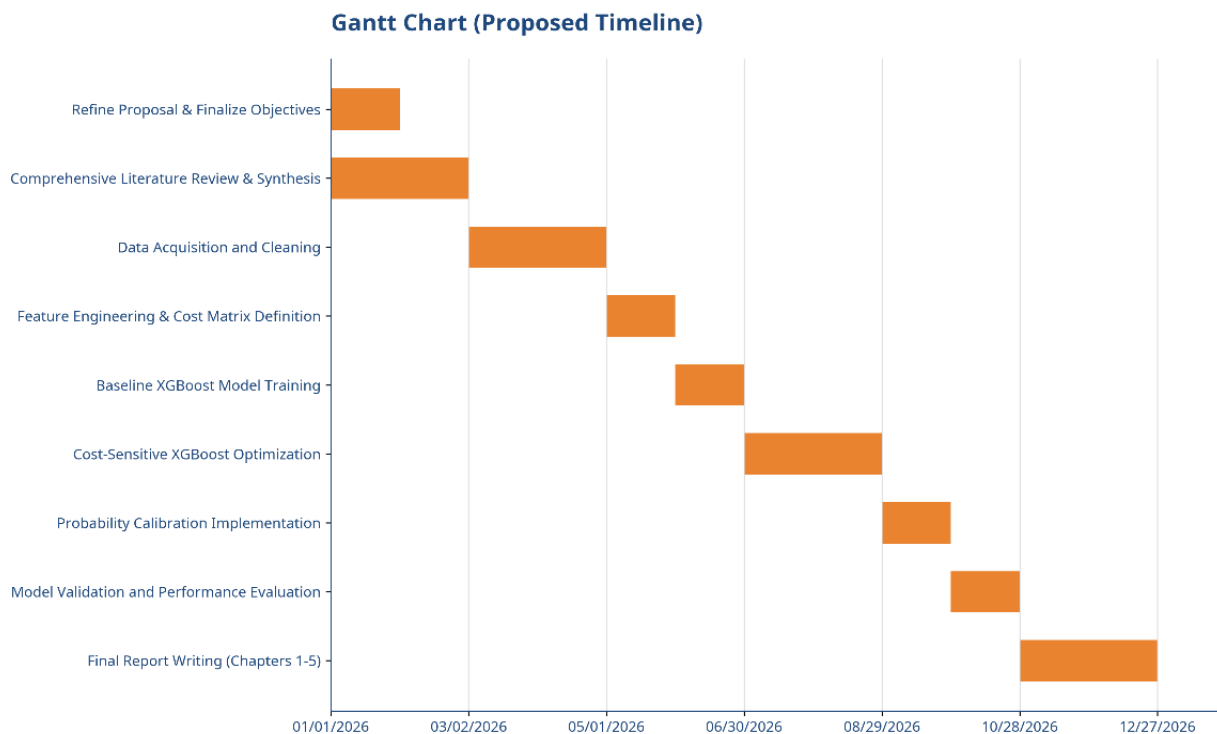
Expected Outcomes

This research is expected to contribute both academically and practically to the field of health insurance claim adjudication. One anticipated outcome is the development of a cost-sensitive predictive model capable of improving the accuracy and reliability of health insurance claim acceptance and denial decisions within Saudi private medical centers.

The study is also expected to demonstrate the importance of probability calibration in enhancing trust and interpretability of machine learning models used in high-stakes financial and healthcare environments. By addressing asymmetric misclassification costs, the proposed model is anticipated to support more informed decision-making, reduce operational losses, and improve efficiency in claim processing workflows.

From an academic perspective, the research is expected to fill existing gaps in the literature by integrating cost-sensitive learning and probability calibration within a Saudi-specific context. Practically, the findings may provide a foundation for future implementation of intelligent decision-support systems aligned with Saudi Arabia's digital transformation objectives under Vision 2030

Gantt Chart (Proposed Timeline)



Reference

- Al-Hamad, B., Al-Dossary, N., Al-Muhaisen, H., Gollapalli, M., Rahman, A., & Ahmed, M. S. (2025). Minimum Data Set (MDS) in Saudi Arabia's medical claims insurance. *International Journal of Computer Science and Network Security*, 25(1), 50–60.
https://www.researchgate.net/publication/389403735_Minimum_Data_Set_MDS_in_Saudi_Arabia%27s_Medical_Claims_Insurance
- Alibrahim, A. I., Kelendar, H., & Alhenaidi, A. (2025). Patterns of claims and determinants of claim rejections in Kuwait's national health insurance for retirees (AFYA): A comprehensive analysis. *Frontiers in Public Health*, 13, Article 1606980.
<https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2025.1606980/full>
- Al-Quayed, F., Humayun, M., & Tahir, S. (2023). A secure technology-based healthcare architecture for smart health insurance systems. *Healthcare*, 11(16), Article 2257.
<https://www.mdpi.com/2227-9032/11/16/2257>
- Brati, E., Braimllari, A., & Gjeçi, A. (2025). Machine learning applications for predicting high-cost claims using insurance data. *Data*, 10(6), Article 90. <https://www.mdpi.com/2306-5729/10/6/90>
- Ho, C.-L., Lai, G., Liao, C.-H., & Loncare, G. (2024). *Unveiling the black box: Determinants of claims investigation, delay payment, and claim denial in private health insurance*. SSRN Electronic Journal. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5153822
- Johnson, M., Albizri, A., & Harfouche, A. (2023). Responsible artificial intelligence in healthcare: Predicting and preventing insurance claim denials for economic and social wellbeing. *Information Systems Frontiers*, 25(6), 2179–2195.
<https://link.springer.com/article/10.1007/s10796-021-10137-5>
- Owolabi, T. (2025). Transforming appeal decisions: Machine learning triage for hospital admission denials. *JAMIA Open*, 8(1), Article ooaf016.
<https://academic.oup.com/jamiaopen/article/8/1/ooaf016/8042205>
- Permai, S. D., & Herdianto, K. (2023). Prediction of health insurance claims using logistic regression and XGBoost methods. *Procedia Computer Science*, 227, 1012–1019.
<https://www.sciencedirect.com/science/article/pii/S1877050923017842?via%3Dihub>
- Somajohassula, D. K. (2025). Transforming healthcare claims adjudication through InsurTech innovations. *Journal of Computer Science and Technology Studies*, 7, 737–745. <https://al-kindipublisher.com/index.php/jcsts/article/view/9871>

Toffaha, K. (2024). *Using machine learning approaches for risk prediction and stratification in healthcare* (Doctoral dissertation, Khalifa University).

<https://khazna.ku.ac.ae/ws/portalfiles/portal/27091975/file>