

Assignment 2: 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: **30%**

Lecturer: **Dr. Mohammed Anbar**

Hand-out Date: **01-Nov-2025**

Hand-in Date: **22-Nov-2025**

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

Submitted by : Hussam Ahmed Almokayed

Student ID : 253001038

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 focuses on testing the model on real or simulated Saudi claim data to confirm its reliability. also will 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)

Reference:

Ho, C.-L., Lai, G., Liao, C.-H., & Lonare, G. (2024). *Unveiling the black box: Determinants of claims investigation, delay payment, and claim denial in private health insurance*. SSRN.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5153822

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/profile/Noor-Aldossary/publication/389403735>

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>

Somajohassula, D. K. (2025). *Transforming healthcare claims adjudication through InsurTech innovations*. *Journal of Computer Science and Technology Studies*, 7, 737–745. <https://al-kindipublishers.org/index.php/jcsts/article/view/9871>

Al-Quayed, F., Humayun, M., & Tahir, S. (2023). *A secure technology-based healthcare architecture for smart health insurance systems*. *Healthcare*, 11(16), 2257.

<https://www.mdpi.com/2227-9032/11/16/2257>

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/S1877050923017775>