Exploiting Machine Learning Bias: Predicting Medical Denials

Stephen Russell[1], Fabio Montes Suros[2], Ashwin Kumar[3]
Jackson Health System, 1500 N.W. 12th Avenue, Miami FL 33136
stephen.rusell@jhsmiami.org[1], fabio.montessuros@jhsmiami.org[2], ashwin.kumar@jhsmiami.org[3]

Abstract

For a large healthcare system, ignoring costs associated with managing the patient encounter denial process (staffing, contracts, etc.), total denial-related amounts can be more than $1B annually in gross charges. Being able to predict a denial before it occurs has the potential for tremendous savings. Using machine learning to predict denial has the potential to allow denial-preventing interventions. However, challenges of data imbalance make creating a single generalized model difficult. We employ two biased models in a hybrid voting scheme to achieve results that exceed the state-of-the art and allow for incremental predictions as the encounter progresses. The model had the added benefit of monitoring the human-driven denial process that affect the underlying distribution, on which the models' bias is based.

Keywords

machine learning, bias, classification, healthcare, data imbalance, human-machine teaming

Introduction

Denial management has become one of the most challenging activities in healthcare. Increasing numbers of healthcare providers are developing denial management processes and analytics to counter the effects of diminishing reimbursements. Multiple thousands of medical insurance claims are submitted by hospitals each day, and payers initially deny about 5-11% of hospital claims. For the average hospital in the US, this statistic means about $5 million in payments are at risk each year. Moreover, while 63% of denials can be recovered, it approaches $120 per claim in administrative costs to recoup the monies owed (Carroll 2020; Change Healthcare 2017). On average, hospitals are spending $8.6 billion nationally in denial-related administrative costs (Change Healthcare 2017). The process of managing claims in the face of potential and real denials represents a significant financial drain on providers' operations and resources, ultimately costing a typical small-medium sized healthcare system 3.3% of net patient revenue, or an average of $4.9 million per hospital and where the numbers are significantly larger for bigger and more complex providers.

The denial of a claim, hereafter referred to as a denial, can be defined as the act of refusing a request by an individual or a provider to pay for the services obtained from a professional. Claims are complex documents that integrate medical coding as relates to illnesses, diagnoses and related services provided to deliver medical care across the spectrum of injuries and/or sicknesses and human variability, and numerous paying entities. Such claims also pass through a complex submission and review process that often involves several third parties between the provider and payer, in addition to the provider's and payer's own processes. There is no single root cause for denials, nor is there any one single problem area. Rather, issues that lead to a denied claim occur throughout each medical encounter and revenue cycle process.

Strategically addressing denials with a data-analytic driven approach can improve the efficiency of the entire revenue cycle, in addition to providing continuous alignment of care delivery to positive outcomes. Considering the spectrum of denial reasons, it is apparent that this is a problem that has organization-wide impact. To this point, identifying and addressing the root causes of denials can have a larger financial benefit than appealing and overturning denials. Taking this notion further, predicting encounters that have a high probability of being denied before a patient is discharged allows procedural and augmenting interventions that can lower the likelihood of a denial. Moreover, an ability to accurately predict denial can be integrated into quality and compliance systems that provide key feedback and monitoring for hospital operational processes. Lastly, even patient experience can be affected not only because of the aforementioned reasons. The alignment of clinical outcomes, with operational outcomes, and ultimately financial outcomes, are the basis for an overall positive patient

experience. This would also suggest that the effects of denials can also be measured in patient experiences. Thus, preventing claim denials *before* claims are submitted to insurers improves profitability, accelerates the revenue cycle, *and* supports patients' well-being (M. Johnson, Albizri, and Harfouche 2021).

Managing denials should begin with using available data to analyze where errors and slowdowns occur, prioritizing those causes, and then addressing them. However, and not to minimize the importance of doing so, it is insufficient to simply *manage denials* because that is a post-hoc activity. Managing denials should be augmented by accurate predictions, so that interventions have the most time to be effective and processes can be measured as close to real time as possible.

An ideal first step towards achieving this capability is machine learning (ML) based system that enables healthcare providers to reliably predict which encounters are likely to be denied and ideally to forecast a payer's response to a subsequent claim, before the patient has been discharged and well-before a claim gets submitted. Predictions from such a system could benefit decision makers by guiding revenue cycle systems and staff; focusing attention on at-denial-risk encounters, by highlighting high-value denials; and monitoring/measuring denial management processes' efficiency and efficacy.

Predicting denials can be formulated as a traditional supervised binary classification problem. However, there are significant challenges in creating a single general model. Healthcare data tends to be highly dimensional and noisy. Further, it is often highly imbalanced, meaning few positive target examples compared to many negative target examples. This introduces problems in achieving a high degree of both precision and recall, while addressing the issues of training data imbalance, overfitting, and bias.

This paper proposes a method for denial prediction that solves issues of training data imbalance by exploiting model bias. The next section presents previous work in denial prediction and some background on algorithmic bias, overfitting and data imbalance. This is followed by a section that illustrates the technical approach, followed by results, implications, and conclusion.

Review of Prior Literature

Machine Learning (ML) and resulting Artificial Intelligence (AI) are igniting and fueling research into the early detection of disease, boosting diagnostic capabilities to enable improved treatment and care outcomes. Similarly, the use of AI to advance healthcare business and revenue cycle functions demonstrates the same promise. Of course, AI is still just a tool. However, if the adoption of this tool is beginning to demonstrate dramatic increases in predictive-accuracy and insight generation over previously available methods, it is a valuable tool; one that warrants attention. Many healthcare providers have already begun experimenting with weaving AI into their critical workflows. Common examples include using AI to predict length of stay, bed utilization, and determining the probability of readmission. Most providers find that it is possible to develop AI models with reasonable accuracy, but there is a nontrivial need to improve the way data is collected and processed across the organization in order to deliver models that are sustainable and durable in operational implementations (Sethi et al. 2021). As a result, AI solutions are still relatively nascent in healthcare provider operations by comparison to big tech organizations that are at the forefront of AI research such as Facebook, Amazon, and Google.

Coupled with electronic medical records, the administrative application of AI technologies is most relevant to process automation. Particularly in bill processing, clinical documentation, revenue cycle and medical records management, several types of AI are already used by payers and care providers (Kaavya 2021). These areas generate critical source data for claims and any subsequent denial processing. As such, these functional areas and related process form the basis for predicting the probability of denials. Table 1 summarizes the current state-of-the-art (SOTA) in denial prediction research.

Other research efforts target improving denials management through predictive analysis, such as error correction (Kumar, Ghani, and Mei 2010), multivariate regression analysis (Matson et al. 2020), and sparsity handling (Zhong et al. 2019; Bai et al. 2019). Johnson and Nagarur (M. E. Johnson and Nagarur 2016) document a framework for detecting provider fraud, which is noted here because it takes a different orientation on denials prediction from the payer perspective. Though only summarized here, it is evident there is ample work on the problem of predicting denials. However, at the time of this writing, the literature is surprisingly sparse and heavily skewed toward claims analysis, vice incremental pre-billing predictions.

Machine Learning, Bias, and Data Imbalance

Bias in the training of ML models is a well-investigated and active area of research (Yang et al. 2023). Machine learning bias, or algorithm bias is when an algorithm produces results that favor or disfavor algorithmic outcomes due to erroneous assumptions in the machine learning process. In this sense, bias is often thought of as fairness, a term which has human, social, and systemic connotations. However, the use of the term bias in this research is of the statistical and computational type. Statistical and computational bias results in effects such as amplification, selection, and sampling. In AI systems, these biases are present in the datasets and algorithmic processes used in the development of AI

Authors	Methods	Primary Contribution & Results
Hoseini (2020)	Interpretable, White-box, Black-box Models (LR, Random Forest, Artificial Neural Network)	Propose an AI-based solution to identify quality issues on Medicaid claim forms that may result in claim denials, waste, abuse, or fraud. The method can detect quality issues with ~80% precision.
Kim et al. (2020)	Blackbox AI (Deep Learning)	Create a system that can represent effectively learned complicated dependencies in claims data to determine the insurers response (denied vs. accepted). The method can detect denied claims with 95% precision.
Kovach and Borikar (2018)	Interpretable AI (Logistic Regression, Statistical Analysis, and Lean Six Sigma)	Develop an improved emergency center registration system to handle missing and inaccurate information on claim data. Denial rates were reduced by 67%.
Khurjekar (2017)	Whitebox & Blackbox AI - (General Regression Neural Networks and RF)	Develop a prediction engine to predict denied claims and aid patient accounting team to manage them. The method can detect denied claims with ~80% accuracy.
Saripalli et al. (2017)	Whitebox & Blackbox AI - (Random Forest, Artificial Neural Network)	Propose a system to fully automate identification of the claims prone to rejection or denial. The method can detect denied claims with ~70% accuracy.
Johnson and Nagarur (2016)	Blackbox AI (Artificial Neural Network and Self Organizing Maps)	Cluster physicians based on their services, diagnoses, and charges to identify the characteristics of physicians with high denial rates. The method can detect denied claims with ~80% accuracy.
Wojtusiak et al. (2011)	White-box AI (Rule-based Methods)	Craft a method for deriving attributional rules that can be used to support the preparation and screening of claims prior to their submission to payers to reduce denial rates.

Table 1. Summary of research studies aiming to predict claim denials using AI. Adapted from M. Johnson, et al. (2021).

applications, and often arise when algorithms are trained on one type of data and cannot extrapolate beyond those data (Schwartz et al. 2022). Issues of statistical and computational bias are often associated with model overfitting, particularly in tabular data-based classification problems.

A ML model that achieves high accuracy on a training dataset can fare worse on unseen data. In this case, the model has "overfit" the training data, reflecting not only true underlying relationships, but also patterns arising purely by chance. Since overfitting in this sense decreases predictive accuracy, a great deal of effort has been expended in developing overfitting avoidance methods (Jabbar and Khan 2015). Overfitting avoidance methods include early stopping, penalty, and omission regimes. However, if overfitting avoidance methods improve the predictive accuracy of a model, they must do so by inducing amplification, selection, or sampling effects. Therefore, any overfitting avoidance strategy amounts to a kind of bias, and biases are only as helpful as they are appropriate to a domain of application (Schaffer 1993).

It may be argued that, in practice, bias is intrinsic in the training machine. Ground truth datasets, necessary to train models, are commonly cleansed or otherwise manipulated to obtain the best model results and generalizability. Moreover, training datasets are typically collected so that samples are maximized under time, budget and accessibility constraints. As such, the performance of ML classifiers is, among other factors, sensitive to the class proportions of the training dataset.

While the previous discussion could be generalized to supervised ML problems, we focus on classifiers, vice regression to give focus to the denial prediction context. The literature has shown that there is a clear relationship among the bias, overfitting, and training-data balance. The question arises if bias can be exploited to address data imbalance issues without problems of overfitting. In the next section we outline our approach that employs two biased models to address the impediment of the denial data imbalances.

Technical Approach and Results

As discussed above, denial transactions represent a significant amount of revenue, complicated by diverse patient characteristics and a competitive relationship with numerous payers. Specifically, this effort sought to create a model that can provide a real-time prediction of denial-risk as each patient encounter progresses, based on the known demographic characteristics and clinical events of that encounter.

Unlike much of the prior work on denials, our approach does not focus on the encounter claim; rather we adopt use of the encounter characteristics and clinical transactions, e.g., labs ordered, diagnoses, patient demographics, and derived ratios. We engineered 18 features; Table 2 shows the

list of features. Using these features, several machine learning methods were implemented to create models using an 80/20 train/test-holdout split, with cross validation. The whole dataset consisted of 22.5 encounters, of which 5,925 encounters were set aside for test -- 4532 accepted; 1393 denied.

Feature Name	
Length of Stay	Medical Service
Division	Attending Physician
Admit Point	Lab Ratio
Race	ICD Ratio
Sex	ICU Flag
Patient Age	Surgery Flag
Diagnostic Related Group	Expired
Current FC	Denied Flag
Primary Insurance Plan	Trauma Flag

Table 2. Feature list

Historically, the healthcare provider who supplied the data typically experienced a ratio of 80% of their inpatient encounters accepted and 20% denied by payers. Knowledge of this 80% - 20% ratio was exploited as *bias* relative to the training data proportions. One might think of this bias as the proportional mixture of accepted and denied observations in the training set. Several models were trained, adjusting the accepted/denied proportions using random sampling to create the desired mix. Figure *2* shows the results of the three top performing ML methods: K-Nearest Neighbor (KNN), Random Forest (RF), and Extreme Gradient Boost (XGBoost) and their respective tuning approach: grid search (GS) and forest pruning (FP) for the accepted encounter results. Figure *1* shows the same for denied encounter results. In both cases, the proportions of accepted and denied encounters in the training data is shown along the x-axis. While shown in separate graphs for clarity, it is apparent that the training proportions impacted the quantity of type-2 errors from the models in the two target contexts.

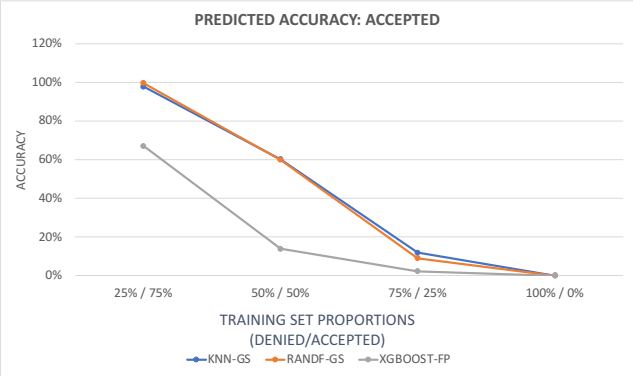

Figure 2. Proportional training model results for payer-accepted encounters

The models precision, recall, and F1 scores reflected this dynamic as shown in the two graphs.

Utilizing the historical knowledge-bias of operational denial proportions and the experimental results of the proportional training experiment, a "biased expert" model was created for each class -- expert denial and expert acceptance. Each of these employed the best-performing ML method, XGBoost and Random Forest respectively. Hereafter the XGBoost expert-denial model is referred to as the *deny-model*; and the Random Forest expert-acceptance model *accept-model*. The two models were employed an ensemble that passes all inputs to both models. Given the output of both models, the decision logic for a final prediction follows a truth table, shown in Table 3, that corresponds to the two models' biased expertise. Since the accept-model is an expert at predicting accepted encounters and the denial-model is an expert at predicting denied encounters when both models agree the expert is selected. This leaves two cases, where the experts disagree, one where they follow their expertise and another where they both predict their weakness. In this instance, the final prediction is a random selection, weighted towards the operational bias.

Deny-Model	Accept-Model	Final Prediction
A	A	A
A	D	Randomly 75% A 25% D
D	D	D
D	A	Randomly 75% A 25% D

Table 3. Ensemble final-prediction truth table.

The ensemble model was evaluated with the holdout dataset consisting of 5,925 encounters. Neither model had been exposed these encounters. The accept-model correctly predicted 4505/4532 accepted (99.404% accuracy) and correctly predicted 43/1393 denied. The deny-model correctly predicted 1388/1393 denied (99.641%) and correctly

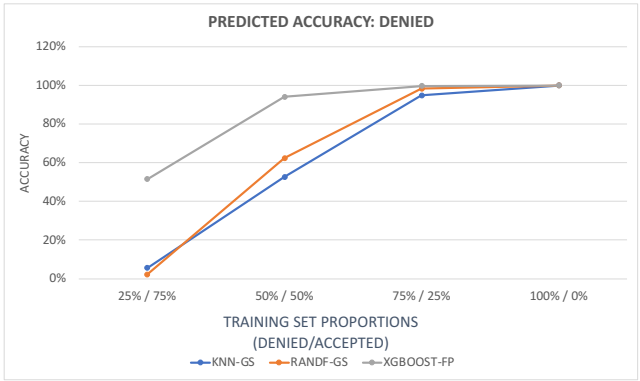

Figure 1. Proportional training model results for payer-denied encounters

predicted 101/4532 accepted. The overall ensemble accuracy was 97.570%.

Implications and Discussion

The overall ensemble accuracy of 97.6% exceed the SOTA from the reviewed literature. From a purely computational sense, the approach appears to have merit. However, it relies on operational bias – 25% denial rate, which like most bias has a probability of changing. In fact, should the ensemble model be operationally effective the bias is sure to change. While this may seem to be a significant weakness, there is little different from model drift that would be experienced by any model, where the solution is to simply retrain the model. The same would apply here. The ensemble has been running for several months and has experienced some degradation due to changes in the underlying features such as new insurance plans, diagnostics related group codes, and ratios that the model has never seen. Again, this is to be expected. Anecdotally, the ensemble was able to detect a change in human policy that affected the bias, which manifested as a sharp change in the predictive accuracy. There are several contributions of this research:

- Exploitation of computational bias and an ensemble approach to address training set data imbalances for machine learning
- The use of inter-encounter features that allow predictions prior to an encounter claim being produced, extending the timeline for human interventions
- Creation of a process sensitive ensemble model that can identify changes in human processes; implying utility as a process monitoring tool

There a several pragmatic limitations, besides those mentioned already, associated with re-training / model reliability. There are engineering considerations that should be explored to provide solutions to implementation challenges such as quantifiable uncertainty and interoperability. It is critical that denial predictions be accurate and reliably bounded because, in a real-time implementation, they will support decision-making that allocates or re-allocates resources, potentially leading to unanticipated negative effects.

Conclusion

Healthcare providers are chiefly concerned with ensuring patients receive the highest quality of care. However, to do so a provider must also be able to get compensated for delivering that care. Despite advances in medical technologies and a declining number of uninsured patients, healthcare providers still have challenges getting paid fully and in a timely manner. The capital involved in claim denials and related processes can exceed hundreds of millions of dollars, fundamentally putting care delivery at risk. There is ample justification, illustrated by a conservative 4-8X return-on-investment, to justify further investigation and development of a denial prediction solution.

The model described herein advances the SOTA with demonstrated improvement in predictive accuracy that exploits computational bias and allows the temporal scaling of prediction, such that interventions are possible. This differentiation from previous work may also yield further benefits to surrounding revenue cycle processes. Moreover, the approach sets the foundation for greater fidelity in the provided predictions. Critical to executing the development of a new, more robust, denial prediction model is addressing generalization and implementation concerns to ensure the reliability and robustness of the underlying model(s) and machine learning in a sustainable way.

References

Bai, Tian, Brian L. Egleston, Richard Bleicher, and Slobodan Vucetic. 2019. Medical Concept Representation Learning from Multi-Source Data. In *IJCAI: Proceedings of the Conference*, 2019:4897. NIH Public Access.

Carroll, Linda. 2020. More than a Third of U.S. Healthcare Costs Go to Bureaucracy. *Reuters*, January 6, 2020, sec. Healthcare & Pharma. https://www.reuters.com/article/us-health-costs-administration-idUSKBN1Z5261.

Change Healthcare. 2017. Front-End Revenue Cycle Processes Leading Cause of Denials. https://revenuecycleadvisor.com/news-analysis/report-front-end-revenue-cycle-processes-leading-cause-denials.

Hoseini, Cyrus. 2020. Leveraging Machine Learning to Identify Quality Issues in the Medicaid Claim Adjudication Process. PhD Thesis, Indiana State University.

Jabbar, H., and Rafiqul Zaman Khan. 2015. Methods to Avoid Over-Fitting and under-Fitting in Supervised Machine Learning (Comparative Study). *Computer Science, Communication and Instrumentation Devices* 70 (10.3850): 978–81.

Johnson, Marina, Abdullah Albizri, and Antoine Harfouche. 2021. Responsible Artificial Intelligence in Healthcare: Predicting and Preventing Insurance Claim Denials for Economic and Social Wellbeing. *Information Systems Frontiers*, April. https://doi.org/10.1007/s10796-021-10137-5.

Johnson, Marina Evrim, and Nagen Nagarur. 2016. Multi-Stage Methodology to Detect Health Insurance Claim Fraud. *Health Care Management Science* 19 (3): 249–60.

Kaavya, Saravanakumar. 2021. User Friendly AI In Technology In Hospitals. *Bodhi International Journal of Research in Humanities, Arts and Science* 5 (2).

Khurjekar, Neel Mandar. 2017. An Integrated Three Stage Predictive Framework for Health Insurance Claim Denials. PhD Thesis, State University of New York at Binghamton.

Kim, Byung-Hak, Seshadri Sridharan, Andy Atwal, and Varun Ganapathi. 2020. Deep Claim: Payer Response

Prediction from Claims Data with Deep Learning. *arXiv Preprint arXiv:2007.06229*.

Kovach, Jamison V., and Shrutika Borikar. 2018. Enhancing Financial Performance: An Application of Lean Six Sigma to Reduce Insurance Claim Denials. *Quality Management in Healthcare* 27 (3): 165–71.

Kumar, Mohit, Rayid Ghani, and Zhu-Song Mei. 2010. Data Mining to Predict and Prevent Errors in Health Insurance Claims Processing. In , 65–74. https://doi.org/10.1145/1835804.1835816.

Matson, Andrew P., Brandon E. Earp, Kyra A. Benavent, Katarina M. Geresy, Jamie E. Collins, and Philip E. Blazar. 2020. Predictors of Insurance Claim Rejection in Hand and Upper Extremity Surgery. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* 28 (15): e662–69.

Saripalli, Prasad, Venu Tirumala, and Anundhara Chimmad. 2017. Assessment of Healthcare Claims Rejection Risk Using Machine Learning. In *2017 IEEE 19th International Conference on E-Health Networking, Applications and Services (Healthcom)*, 1–6. IEEE.

Schaffer, Cullen. 1993. Overfitting Avoidance as Bias. *Machine Learning* 10 (2): 153–78. https://doi.org/10.1007/BF00993504.

Schwartz, Reva, Apostol Vassilev, Kristen Greene, Lori Perine, Andrew Burt, and Patrick Hall. 2022. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. *NIST Special Publication* 1270 (10.6028). https://view.ckcest.cn/AllFiles/ZKBG/Pages/264/c914336ac0e68a6e3e34187adf9dd83bb3b7c09f.pdf.

Sethi, Ganesh Kumar, Nazir Ahmad, Mohammed Burhanur Rehman, Hatim M. Elhassan Ibrahim Dafallaa, and Mamoon Rashid. 2021. Use of Artificial Intelligence in Healthcare Systems: State-of-the-Art Survey. In *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, 243–48. IEEE.

Wojtusiak, Janusz, Che Ngufor, John Shiver, and Ronald Ewald. 2011. Rule-Based Prediction of Medical Claims' Payments: A Method and Initial Application to Medicaid Data. In *2011 10th International Conference on Machine Learning and Applications and Workshops*, 2:162–67. IEEE.

Yang, Jenny, Andrew AS Soltan, David W. Eyre, Yang Yang, and David A. Clifton. 2023. An Adversarial Training Framework for Mitigating Algorithmic Biases in Clinical Machine Learning. *NPJ Digital Medicine* 6 (1): 55.

Zhong, Qiu-Yue, Andrew H. Fairless, Jasmine M. McCammon, and Farbod Rahmanian. 2019. Medical Concept Representation Learning from Claims Data and Application to Health Plan Payment Risk Adjustment. *arXiv Preprint arXiv:1907.06600*.