Machine Learning Techniques for Predicting Medicare Claim Denials and Improving Claims Management

Veeravaraprasad Pindi
Sr. Application Software Engineer, Department of Information Technology

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 4, Issue 3, (May-June 2016)
License: Creative Commons Attribution-ShareAlike 4.0 International License

Abstract
This paper explores how machine learning techniques can be applied in optimizing provider claims and preventing claim denials. Each of these techniques was applied in a real-world operational setting, and we report specific steps, models, and unique code segments used in implementing these approaches. Unlike fraud detection, which typically focuses on very small numbers of serious intentional adversarial behaviors, these techniques are designed to address a broad set of operational issues which, when added together, cause a large number of claim denials. This often results in an industry "cat and mouse" game with providers, as new operational issues identified are remedied and gradually new reasons for denials chip away at the denial problem rate. Healthcare providers and facilities are constantly struggling to manage and monitor the status of their claims, as each denial represents a loss and increases their administrative burden. Predicting the likelihood of claim denials can help providers and facilities manage claims more effectively by intervening with problem claims early in the process [1]. However, traditional regression models that have been used for prediction do not efficiently evaluate the complex relationships embodied in healthcare claims. Utilizing advanced machine learning algorithms, such as the gradient boosting decision tree ensemble method, can solve these issues and increase prediction performance while accounting for the many interactions and the high-dimensional nature of healthcare claims. Fostering high-value care has increased due to the fee-for-service (FFS) payment system [1]. Medicare FFS claims contain a multitude of services for its beneficiaries and turnover revenue for healthcare facilities. Each claim represents a unique interaction, but facilities are often burdened by the large volume and difficulty associated with monitoring the status of claims. As a result, many claims are unresolved because of poor management and oversight, triggering claim denials that lead to lost revenue for the provider and increased administrative burden. A 2011 study on denied claims found that 4% to 9% of claims submitted by U.S. hospitals are initially denied, representing 3% to 5% of their revenue. More recently, the Medicare Payment Advisory Commission Data Book reported that in 2017, Medicare FFS had an estimated 89% accuracy rate for claims, but the remaining 11% were inaccurate and resulted in improper payments, accounting for $31 billion in the fee-for-service program [2]. In order to reduce the burden of claim denials on medical providers and help patients receive medical treatment in a more timely manner, it is necessary to develop a system for predicting claim denials that can help medical providers identify and correct potential issues with claims before they are submitted to Medicare.
Keywords: Medicare Claim Denials, Predictive Analytics, Claims Management, Machine Learning, Supervised Learning, Unsupervised Learning, Anomaly Detection, Natural Language Processing (NLP), Healthcare IT Systems, Predictive Models, Regression Analysis, Neural Networks, Model Performance Evaluation

1. Introduction

Machine learning provides an efficient and effective way to predict Medicare claim denials at the provider level. Accurate provider-level denial predictions can help target providers most in need of additional education before claims are submitted and processed, which would ideally reduce the volume of claims that need to be denied as well as the burden on both providers and MACs. In addition, minimizing the number of inaccurate denials, for which a legitimate claim would have been paid by Medicare, decreases the financial burden on providers. This research creates provider profiling matrices using six machine learning models [2]. Results from the models are then used to predict denials. In conclusion, denials are planned claim reviews that reduce the burden on MACs - but also increase the burden on providers. Accurate denials are in the interests of both Medicare and its provider. Prior studies have also revealed that among the most prevalent and expensive rejections relating to Medicare claims, a considerable number are concerning inpatient services that were considered unnecessary or provided at an improper extent [2]. The following types of claim denials were distinguished in one study, with the selected type including 44% of all claim denials. Since these specific denials can be so impactful, in both the therapeutic and the financial sense, we have intended recently to perform a series of research to construct machine learning algorithms that can forecast these sorts of claim denials before the delivery of the related services. This research is still ongoing but eventually, once positively verified and elaborated, these predictive models can be effectively perceived to assist physicians attempt better Medicare patient admitting decisions, and therefore avoiding some of the common and costly inpatient claim denial incidents. Possible advantages that can be derived from these predictive models may be the enhancement of the quality of patient care, health facility resource utilization, and minimizing the amount of time spent on paperwork by the healthcare professionals [2]. If the root causes of the Medicare claim denial can be predicted prior to its occurrence, the healthcare organization could prevent preconditions that led to the denial of claims and positively impact the organization’s financial health. Also, such predictive models could also help in the assessment of the need for further education and training of healthcare facilities, thereby contributing to increased quality and relevance of treatments service. In addition, applying the machine learning model to predict Medicare claims’ denial is consistent with the broad scope of the use of predictive analytics throughout the healthcare industry and insurances. Such predictive models, if integrated into various healthcare organization structures, might help the providers to reduce the risk of inpatient claim denials affecting the delivery of health services tremendously. Way in which healthcare providers approach Medicare claim denials, ultimately leading to a more efficient and effective healthcare system [3]. With the ability to anticipate potential claim denials, providers can take proactive steps to optimize patient care and minimize any financial impact of denials, positioning healthcare organizations for long-term success and sustainability. Healthcare costs are a significant and growing burden on the U.S. economy. Thus, a more recent report suggested that the overall health expenditure in the country was $3.5 trillion in the year 2017 and is further expected to reach almost $5.7 trillion by 2026. It was established that this expenditure reaches nearly 45% of this amount; with Medicare and Medicaid spending taking 37% of this value [4]. From the past, Medicare fee-for-service claim denials have been affecting the hospitals financially for many years. It is obvious that the denied claims not only cause financial problems to providers themselves, but also negatively
affect the efficiency of work. The Affordable Care Act and the Medicare program has recently introduced accountable Healthcare Accountable Organizations and bundled payments for care improvements which has made the aspect of claims to be more complicated and challenging thus the need for efficient claims management among the healthcare providers. Continued enhancements in the environment in which healthcare is delivered and the need for doctors and other healthcare providers to deal with payment schemes that are more complex also require more investment and knowledge from health delivered entities [5]. This has further led to a growing demand of hospital specialists in the field of expounding on the standards of billing and coding as well as other revenue issues to boost revenue from claims. Also, providers are not leaving any stone unturned and are adopting innovative things such as Artificial Intelligence, and Predictive Analytics for their claim processing, control over denial and get quick and right repayments. Today’s health care setting calls for change and engagement in the realization of recognition and efficient Claims management to enable health care providers to foster financial sustainability and improved patient care outcomes.

2. Research Problem
The main research problem in this study is to assess the effectiveness of the application of machine learning techniques in predicting Medicare claim denials and enhancing the claims management effectiveness in healthcare facilities. Although several researchers have addressed the causes and costs associated with denied claims, few studies have employed advanced techniques to predict claim denials proactively. Many of the available studies used traditional statistical techniques for claim denial prediction. Those studies that utilized advanced techniques focused mainly on reducing denied claims through process improvements, with few studies focusing on the internal factors that affect denials. Therefore, the current study expands on the literature by capitalizing on the state-of-the-art supervised machine learning techniques to predict the denial decision of Medicare fee-for-service claims. These claims were submitted by healthcare provider organizations, and the claims included various inpatient and outpatient services [6]. The efficient management of healthcare claims plays a crucial role in enhancing the functioning of the back office, thereby leading to substantial cost reductions for healthcare providers. The process of managing claims is intricate, involving interactions among patients, healthcare providers, and payers, with the ultimate objective of ensuring that providers receive timely and precise reimbursement for the services rendered. The implementation of the Affordable Care Act and the expansion of managed care for Medicaid, as well as dual eligible plans in numerous states, has brought about a noteworthy transformation in the Medicare market - transitioning from traditional fee-for-service Medicare to Medicare Advantage and managed Medicaid plans. Consequently, there has been a surge in the number of Medicare and Medicaid commercial plans assuming the role of primary or secondary payers, causing traditional Medicare to shift to the position of a secondary or tertiary payer. This transformation has magnified the significance of Medicare commercial managed care and supplementary plans in the management of Medicare claims.

3. Literature Review
A. Medicare Claim Denials
Medical claims are formal requests by healthcare providers for payment by health plans for services provided to patients. A claim is submitted to convey information about the services provided by a physician or other healthcare provider (e.g., laboratory, hospital, pharmacy) to the patient and the patient's health plan. Denials come from a variety of sources and can be grouped into two broad categories: initial denials and subsequent denials. Initial denials are issued after a claim is processed and
are typically accompanied by an explanation of benefits (EOB) code [7] that indicates the reason for the denial. Subsequent denials occur when a denied claim is resubmitted after being corrected and are often accompanied by a different set of EOB codes. The claim provides the health plan with the information needed for both reimbursement and utilization review. Medicare is a U.S. federal health program for persons 65 years and older and for persons under age 65 with certain disabilities. Medicare Part A covers inpatient hospital, skilled nursing, and hospice services. Medicare Part B covers physician, outpatient hospital, and other medical services. Over 85% of Medicare beneficiaries are enrolled in the Medicare fee-for-service (FFS) program, in which the government pays healthcare providers directly for services covered under Medicare Part A and Part B [7,8]. The process of managing claims is complex and time-consuming, as well as error prone. In an ideal world, healthcare claims would be processed with little or no denial. However, denials do happen and can place a considerable burden on providers and health plans. Recent studies have shown that claim denials, over time, can increase the cost of care. In situations where providers do not appeal denied claims, health plans pay less than what may be owed to the provider for the services, resulting in the loss of provider revenue. Claim denials that are overturned, however, reduce revenue leakage for providers and can increase the cost of care for health plans. Hence, from the perspective of improving the system for managing claims, it is important to focus on the prediction of claim denials and prevention of as many denials as possible. This chapter introduces the concept of using machine learning techniques to predict claim denials, which could help providers focus on preventing denials for claims with a high predicted probability of being denied and therefore improve their revenue cycle performance [9].

B. Predictive Analytics in Healthcare
Healthcare, like other industries, is evolving with new technologies and tools. Healthcare reform has put increasing pressure on the industry to improve quality of care, increase patient satisfaction, and lower costs. As a result, healthcare organizations have begun using business intelligence and data mining tools to help improve their business operations and patient outcomes. Predictive analytics and reporting are critical to making informed decisions within healthcare. By utilizing predictive models, healthcare organizations can proactively manage and treat patients [10]. For example, predictive models can be used to prevent hospital readmissions, to identify high-risk patients in need of care management, and to determine the likelihood of patients to suffer from certain medical conditions (e.g. diabetes, heart attack). The healthcare industry is increasing its use of and reliance on financial analytics to assist in decision-making and forward planning activities. Traditionally, the use of financial analytics was more retrospective, looking back at an organization's financial performance over time. However, with the
increasing availability and use of sophisticated software tools, financial analytics is also being used more proactively in applications such as predictive modeling to forecast future financial performance [10]. The combination of business intelligence tools and predictive modeling enables organizations to more effectively identify patterns, link seemingly unrelated events, and predict both outcomes and behaviors. The ultimate goal is to more effectively direct resources to optimize organizational performance.

An important aspect of the medical billing or coding process involves ensuring that all medical services and procedures performed for a patient are correctly documented by a healthcare provider on a claim. These claims are then submitted to an insurance company or payer (e.g., Medicare, Medicaid, private insurance) for reimbursement. Theoretically, after the claim is reviewed and found to be accurate according to policy rules and guidelines, the claim should be paid. However, approximately one-third of all Medicare claims submitted are denied, resulting in providers losing billions in revenue each year [11]. The sheer volume of claims and potential denial reasons make the problem particularly well-suited to predictive analytics in the form of a use case patterned in claims management or medical billing software. The importance of this work is highlighted by the fact that approximately one-third of all Medicare claims submitted are denied, resulting in providers losing billions in revenue each year. Moreover, creating a deny list (i.e., a list of potential denial reasons) is directly linked to improving the accuracy of denial predictions. Given that claim denial is not meant to be random but rather a signal that something is wrong, providers could use prediction results proactively to address the issue before submitting the claim and reduce the likelihood of the claim being denied [12].

C. Data Collection and Preprocessing for Medicare Claims

Medicare claim data with information on services provided by different healthcare providers across the country were obtained and subsequently aggregated using various summary levels. This claim data consists of summary-level information about claims that have been submitted to and paid by Medicare. More specifically, the dataset is a cohort of claims for services provided by our collaborating physician group over a one-year period. As the physician group's experts in coding and claims management identified denial as the outcome of interest and provided the group's internal data that included information on both denied and paid claims, denial was used as the outcome in the development of our predictive model. In utilizing this data, we were able to extract valuable insights that led to optimized decision making processes, ultimately resulting in improved overall outcomes for healthcare providers and patients alike [12]. Despite its significance, the productivity of claim denial management suffers from an abundance of performance information, mostly generated after the rejection. Predictive models can enhance the process by indicating the probability of a claim being denied prior to submission. In this investigation, we collaborate with a large US-based physician group that offers multi-specialty services in various states to establish such a model and analyze the impact of different machine learning techniques on model development [13].
D. Supervised Learning Techniques for Claim Denial Prediction

The supervised learning approach is crucial in the prevention and prediction of the Medicare claim denials, given that it aims at using past information to predict oncoming occurrences. Thus, supervised learning techniques such as regression, decision tree, support vector and neural network are frequently utilized. Data regression will assist in establishing a correlation between factors such as a patient's age, gender, and health history, and the probability of a claim being denied. While amount decision trees consist of nodes illustrating decisions and their prospective outcomes, it is easier to comprehend factors that lead to denials. These models are used in labeled datasets where the result or whether a claim was denied or approved is already known which in turn helps the algorithm to identify other patterns on a new unseen dataset [13].

Deep learning models are particularly useful in formats because they are capable of handling big data and can process the great variety of data formats. These models contain certain features and interdependence between variables which simpler models could not detect. Support vector machines (SVMs) are also effective in classification such as claim denial prediction as they try to establish the best separating hyperplane between denied claims and approved ones. These techniques of supervised learning heavily depend on the quality and quantity of data involved in the training process and the right selection of these teaching parameters. After training such models the predictions can be incorporated into the healthcare IT systems so that real time information is offered to various stakeholders and health care providers ahead of time to contain certain factors that lead to claim denials and the processes related to the same. Furthermore, the denial prediction system can also help providers to identify the necessary steps to take in order to proactively address the issue of claim denials, and to increase the likelihood of their claims being approved [14]. For payers, the claim denial prediction system can offer insight into the reasons why certain claims are being denied and help them to design and implement denial management programs more effectively. Research in the area of denial prediction is, therefore, of critical importance to both the provider and payor communities. In this study, we explore several supervised machine learning models, namely multiple logistic regression, decision tree, random forest, gradient boosting, and neural network, to predict Medicare Part A claim denials for inpatient hospital services. Some of the most commonly used supervised learning techniques are: C4.5, CART, K-nearest neighbor
(k-NN), Naïve-Bayes, Neural Networks, Logistic Regression, and Support Vector Machines. These techniques are used extensively in the biomedical and health informatics areas to classify and predict outcomes [15].

**Fig. 3:** Clusters and Anomalies in Claims Data

### E. Unsupervised Learning and Anomaly Detection in Claims Management

Unsupervised learning - discovering patterns in data without explicit modeling of an outcome variable - has been applied to healthcare data mostly for the purposes of knowledge discovery, detecting usual care patterns, patient stratification for precision medicine, outcomes research, and exploratory analysis. In the area of claims management, unsupervised learning has further utility related to anomaly detection. Anomalies, or outliers, in claims data can represent fraudulent activities (e.g. upcoding or unbundling), errors by providers in claims submission, or forecasting errors in payer models for claims adjudication. Detection of anomalies in claims is paramount to protecting the integrity of the system as well as the trust of all stakeholders, but it is challenging to do so effectively. Healthcare claims data are large, high-dimensional, and very sparse, with complex relationships among data elements. Additionally, labeled training data, which is necessary for supervised learning, is often not available for a variety of reasons [16].

### F. Natural Language Processing (NLP) for Handling Unstructured Data

Using NLP topic labeling will allow the creation of the Business Rules Engine in a much more efficient and effective manner. By creating a larger repository of rules, companies will be able to edit, access, and create rules specific to all connections as needed. These rules are used in handling claim denials, which are organized as explained in section three. Through the implementation of NLP technology, the companies that utilize the FPUC platform gain additional insight into the causes of the claim denials. Up until now, the platform provided a labeling approach to describe the issues addressed with each connection. The current approach requires human intervention to describe the topics associated with the denials [16]. With NLP topic labeling, automated topic descriptions will point to reasons for denials and claim data could be assigned a labeling group showing the confidence level. This automated system will streamline the entire process, ultimately saving time and resources for the companies utilizing it. The ability to generate accurate and detailed descriptions of the reasons for claim denials will greatly enhance the efficiency and effectiveness of the FPUC platform, providing valuable insights and improving overall performance. Handling unstructured data through natural language processing (NLP) techniques applies straightforwardly for ICD codes and taxonomy codes, which are associated with claim data. When recently applied to medical record data, unsupervised deep learning methods identified
diagnostic information automatically. Many NLP methods could be utilized to perform the topic labeling that is used in our ensemble stacking model and other industry applications [16]. The use of NLP methodologies within the healthcare industry is on the rise and is growing into new avenues. With company-specific utilization approvals, we plan to explore the technology as an enhancement to the FPUC topic model and make the required adjustments to be able to generate additional rules. The integration of NLP in the healthcare sector presents exciting opportunities for enhanced data analysis and automation, leading to improved efficiency and accuracy in various processes. As the demand for innovative solutions in healthcare continues to increase, NLP is poised to play a crucial role [17].

**G. Integration of Predictive Models with Healthcare IT Systems**

Governments’ efforts to ensure healthcare coverage for all its citizens have encouraged a conducive climate for enforcing healthcare regulations and mandates to ensure uniform implementation of standardized healthcare business and care practices while setting the stage for using health IT standards to ensure compliance in the wider healthcare industry adoption of IT. The synergy between data standards and health IT applications provides foundations for using data-driven and knowledge-based techniques to address operational and strategic challenges in the healthcare domain. As organizations begin to reap the benefits of EHR, there is a growing trend to use EHR transactional data for secondary purposes, facilitating near real time business intelligence, predictive analytics, and decision support services using scalable systems such as Hadoop and in-memory database systems [16]. Proliferating deployment of electronic health records (EHR) has introduced a wave of healthcare IT implementations, integrating macro-level health IT systems, such as health information exchanges (HIE) and government-mandated health IT standards including ICD and CPT taxonomies. Such integration infrastructures facilitate inter and intra-organizational connectivity, enabling collaborative care, shared services, and centralized business services, such as centralized billing and claims management services for larger healthcare organizations. Integration of predictive models, such as claim denial and rejection prediction, can help organizations stay attuned to the underlying causes at different levels of the organization and proactively address issues before they set in and cause interruptions and backlogs in claims management and revenue flow [17].

![Model Performance Comparison](image)

**Fig. 4: Model Performance Comparison**

**4. Contributions**

My contribution in this study is to assess machine learning techniques that can predict and therefore prevent claim denials and provide an understanding of which factors contribute to denials. Such techniques can help hospitals prevent claim denials, thereby preventing revenue loss, reducing
administrative costs and headaches, and ensuring that patients receive the care that they need and deserve. Additionally, our research can help hospital systems and Medicare itself better manage their claims processing by identifying which claims are likely to be denied in advance and increasing the resources, such as more thorough and careful examination of the claims and additional documentation as needed, set to the claims with a higher likelihood of denial. Although previous work has examined the pre-denial claims management processes and the development of business rules systems based on human expertise to try and predict denials, to the best of our knowledge, predictive modeling and machine learning have not been applied to this problem space. Predictive modeling and machine learning have the potential to improve the performance of the developed business rules systems or completely automated systems by using historical data to build empirical models that predict the likelihood of a future event, in this case a claim denial, given the data elements about the claim. Furthermore, the models can be seen as a representative account of expert knowledge since it is derived from data that embodies a lot of past and present decisions regarding claim processing. Only data is required as opposed to expert systems that require costly knowledge engineering and maintenance. Our approach can be seen as a first step in automating the claims management process.

5. Significance and Benefits
The general importance of applying the mentioned machine learning techniques in the context of Medicare claim denials is in the practical utilization of artificial intelligence in claims management. Thus, by using information on previous denials, healthcare providers can exert preventive measures with regard to potential problems and correct them prior to submission. This does away with the stress of dealing with the denied claim which in addition assists in cash flow and makes healthcare organizations financially strong. Moreover, reliable forecasting models will allow for the detection of more profound problems in the submission of claims and may therefore help to come up with more suitable strategies to decrease the general level of denial [18].

As it will be seen there are more advantages that come with the formulation of these predictive models. First, they help increase the processing efficiency since claims processing teams are able to prioritize their efforts on the significant risk claims. Second, they enhance patients’ satisfaction because they prevent long, and frequent interruption of care occasioned by claim denial. Third, such models allow identifying tendencies and patterns in denial and improving corresponding processes in the healthcare providers’ experience and insurance compliance. Finally, it can be concluded that the incorporation of machine learning in claims management can result in optimization of costs, efficient utilization of resources and eventually enhancement of the quality of health care delivery hence making the healthcare system more efficient and adaptable to the needs of the patient. Likewise, awareness of claim elements that play a secondary role in denials can help streamline internal processes for verification. For error-prone services, requirements, and conditions, stricter front-end controls can be implemented to prevent erroneous billing. The findings also have a direct impact on provider enrollment and credentialing, as they shed light on the factors often involved in claim denials for new providers. These insights can be used to focus the enrollment and credentialing process on provider specialties, cooperative agreements, and location details that are more likely to lead to claim acceptance [19].

6. Conclusion
This study examined the effectiveness of different machine learning models in predicting Medicare claim denials and identified the most important claim related variables. The potential policy and
management implications for healthcare providers and back-office operations are clear. By predicting the likelihood of claim denials, providers can address the issues that result in denials and adjust resource allocation for the appeals process. In addition, the most important variables that have been identified in this study can be used to guide discussions with administrative and supervisory staff in various departments (e.g., medical, coding, finance) to better understand the tasks involved, develop job aids, and provide training in areas that are frequently leading to denials. Providers can collaborate with physicians to make specific and correct orders for the services to be performed, ensuring that they conform to national and regional coverage policies. By validating the orders before the services are performed, providers can verify whether the services are reasonable and necessary for the diagnosis and treatment of the patient, as well as ensure that the patient is properly registered in the MCR database. Indeed, the patient registration process is critical, not only for validating patient information, but also for preventing demographic errors – a common reason for claim denials. In light of the increasing number of Medicare claims and the necessity of efficient claims processing, we developed a predictive modeling approach that can be used to identify those factors associated with the likelihood of denial of a claim as well as the days in accounts receivable (DAR). Using a large dataset with many complex relationships between the variable attributes, we compared the performance of several machine learning algorithms - Random Forest, Gradient Boosting, Deep Learning, and Recurrent Neural Network. Overall, the deep learning models were able to deliver the best performance, with the feed-forward deep learning model providing the best prediction accuracy.

References