Detecting Healthcare Fraud using Machine Learning: Excluding Provider Labels for Improved Accuracy

Kondragunta Rama Krishnaiah, Professor, Department of Computer Science and Engineering, R K College of Engineering, Vijayawada - 521456, Andhra Pradesh, India, email: kondraguntark@gmail.com

Alahari Hanumant Prasad, Professor, Department of Computer Science and Engineering, R K College of Engineering, Vijayawada - 521456, Andhra Pradesh, India, email: hanuma.alahari@gmail.com

Abstract

As the elderly population continues to grow, it brings along with it a greater demand for medical services and increased healthcare expenses. To address these needs, the United States has a healthcare program called Medicare, designed to provide insurance primarily to individuals aged 65 and older, easing some of the financial burdens related to medical care. However, despite this initiative, healthcare costs remain high and continue to rise. One significant factor contributing to this problem is fraud. Our research paper tackles this issue head-on by employing machine learning techniques to identify fraudulent Medicare providers. We conducted a comprehensive study using publicly available Medicare data and fraud labels for provider exclusions. By building and evaluating three different learners, we sought to develop effective fraud detection methods. One challenge we faced was the class imbalance in the data, as there were very few actual fraud labels available. To address this, we adopted random under sampling, which allowed us to create four different class distributions. By doing so, we could mitigate the impact of the class imbalance and enhance the accuracy of our models. The results of our study were promising. Among the three learners we tested, the C4.5 decision tree and logistic regression stood out for their exceptional fraud detection performance. Particularly impressive was their performance with an 80:20 class distribution, where both models achieved average AUC scores of 0.883 and 0.882, respectively, along with low false negative rates. Our research demonstrates the effectiveness of using machine learning in combination with random under sampling to detect Medicare fraud. By identifying fraudulent providers more accurately, we can help reduce healthcare expenses and ensure that Medicare resources are utilized more efficiently, ultimately benefiting those in need of medical care.

Keywords: Healthcare, Fraud detection, Supervised methods, Unsupervised methods

1. Introduction

The viability of most healthcare systems revolves around competent and capable medical providers and a solid financial infrastructure. Both aspects can be irrevocably damaged by fraud, waste, and abuse. The financial backbone, in particular, is subject to fraudulent activities incurring potentially large losses. Healthcare programs, in the United States (U.S.), have experienced tremendous growth in patient populations and commensurate costs. The elderly community continues to grow with a 28% increase in 2014 versus a rate of just 6.5% for individuals under 65 years of age (U.S. Administration for Community Living 2015). Moreover, in 2015, spending on healthcare-related activities reached $3.2 trillion, which is about 17% of the total U.S. budget (Backman 2017). Medicare is one such U.S. healthcare program created to assist the elderly and other individuals with certain medical conditions (Medicare 2017). Medicare alone accounts for about 15% in spending (net of $588 billion), per year of the total healthcare budget and is expected to increase to 18% within the next decade (Backman 2017). Given the increase in the elderly population, with their need for increased healthcare and financial assistance, programs like Medicare are critical and, as such, must reduce program expenses and costs to allow for accessible healthcare. One way to accomplish this is to lessen the impact of fraud. The impact of healthcare fraud is estimated to be between 3% to 10% of the nation's total healthcare spending continuing to adversely impact the Medicare program and its beneficiaries (NHCAA 2017). There are programs, such as the Medicare Fraud Strike Force (OIG 2017), enacted to help combat fraud, but continued efforts are needed to better mitigate the effects of fraud.

More information on healthcare fraud, to include different types of fraud, can be found in (Joudaki et al. 2015; Bauder, Khoshgoftaar, and Seliya 2017). In this paper, we propose a machine learning approach for Medicare fraud detection using publicly available claims data and labels for known fraudulent medical providers, across all medical specialties or provider types (e.g. dermatology or cardiology). We do not build a distinct model per specialty, but rather one model to predict a fraudulent provider regardless of specialty. Specifically, we use the Medicare Provider Utilization and Payment Data: Physician and Other Supplier, available from the Centers for Medicaid and Medicare Services (CMS), which provides information, by physicians and other healthcare providers, on services and procedures provided to Medicare beneficiaries (CMS 2017). The Medicare data does not contain labels indicating fraudulent providers or procedures. In order to build models, or learners, to detect fraudulent providers, we use the information found in the List of Excluded Individuals and Entities (LEIE) database (LEIE 2017). This database contains a list of individuals and entities who are excluded from participating in federally funded healthcare programs due to fraud. We detail a process for merging the Medicare data and the LEIE labels that accounts for differing lengths of exclusions, matching providers by unique identification numbers. The final dataset has significantly more non-fraud versus fraud labels, thus is a considered highly imbalanced. In or der to mitigate the adverse effects of class imbalance on detecting fraud, we employ random undersampling (RUS) which retains all fraud labels while randomly reducing the number of non-fraud labels. Because the Medicare data is big data, with over 37 million instances, using oversampling methods would further increase the dataset size making many machine learning approaches impractical. We create and test four different class distributions, or ratios, to assess the best mixture of majority (non-fraud) and minority (fraud) class labels. For each distribution, we build and assess three different learners (C4.5 decision tree, logistic regression, and support vector machine) using 5-fold crossvalidation, repeated 10 times to reduce bias caused by bad draws during sampling. In order to fairly assess fraud detection performance, we use several measures which include the Area Under the ROC (Receiver Operating Characteristic) Curve (AUC), false positive rate (FPR), and false negative rate (FNR). Our results indicate that the C4.5 decision tree and logistic regression learners have the best overall AUC performance, particularly for the 80:20 and 75:25 (majority:minority) class distributions. To the best of our knowledge, no other work provides a study that directly incorporates the entire Medicare dataset plus LEIE exclusion labels to detect fraudulent providers for any specialty, using differing RUS class distributions on a diverse set of learners.

The remainder of the paper is organized as follows. The Related Works section discusses works related to the current research. In the Methodology section, we discuss our research methodology detailing the Medicare and LEIE data, learners, performance metrics, class imbalance, and experimental design. The results of our research are examined in the Results and Discussion section. Finally, the Conclusion section summarizes our conclusions and plans for future work.

2. Related Works

With the limited number of easily accessible, documented Medicare fraud cases and the relatively recent availability of data, a lot of the existing Medicare fraud detection research uses unsupervised machine learning via anomaly detection methods. A recent study by Sadiq et al. (Sadiq et al. 2017) employs the Patient Rule Induction Method (PRIM) based bump hunting (unsupervised) method to identify anomalies in the 2014 Florida Medicare data. Studies, such as those by our research group, employ unsupervised methods to detect anomalies in Medicare payments leveraging regression techniques and Bayesian modeling (Bauder and Khoshgoftaar 2017; 2016). In our work, we employ supervised Medicare detection methods using publicly available excluded, or fraudulent, provider information, which is the focus on the remainder of the related works.

In a preliminary study, Chandola et al. (Chandola, Sukumar, and Schryver 2013) use Medicare claims data and provider enrollment data from private sources to detect healthcare fraud. The authors employ several different techniques including social network analysis, text mining, and temporal analysis. Using features derived from the temporal analysis, the authors build a logistic regression model to detect known fraudulent cases using labeled data from the Texas Office of Inspector General's exclusion database only. Moreover, details are limited with regards to data processing and mapping fraud labels to the Medicare data. It is important to note that none of these studies deal with the problem of class imbalance. Our research group presents an exploratory study, using 2013 Florida Medicare data, that looks to predict fraudulent providers by using only the number of procedures performed via a Multinomial Naive Bayes model (Bauder et al. 2016).

If the predicted provider type does not match what is expected, then this provider is performing outside of normal practice patterns and should be investigated. There are only two related works found that address class imbalance in the detection of Medicare fraud, using the LEIE database. In a study by Herland et al. (Herland, Bauder, and Khoshgoftaar 2017), the authors validate and improve upon their previous model which detects possibly fraudulent behavior by predicting a provider's specialty based on the number of procedures performed. They use 2013 Medicare data (Florida only) and the LEIE database for fraud labels. The authors propose three strategies to improve their previous model that include the following: feature selection and sampling, removal of low scoring specialties, and grouping similar specialties. Class imbalance was mitigated using both random undersampling and Synthetic Minority Over-sampling Technique (SMOTE) for 82 specialties. Branting et al. (Branting et al. 2016) create a graph of providers, prescriptions, and procedures. The authors use two algorithms where one calculates the similarity to known fraud and non-fraud providers, and the other estimates fraud risk via shared practice locations.

Medicare data from 2012 to 2014 was used with 12,153 excluded providers from the LEIE database. To address class imbalance, the authors only used a 50:50 class distribution. A J48 decision tree was built using 11 graph-based features and 10-fold cross-validation but no repeats. In relation to the last two very preliminary studies, which also use Medicare data with LEIE fraud labels, our research is more comprehensive in the breadth and depth of experimentation and results. We provide a comprehensive discussion of the data and the mapping of the fraud labels. We employ three different learners on four different class distributions to assess the effects of class imbalance. Moreover, our experimental design is robust using 5-fold cross-validation with 10 repeats per learner and class distribution combinations. Finally, we present results using several different metrics and discuss statistical significance of the results.

3. Methodology

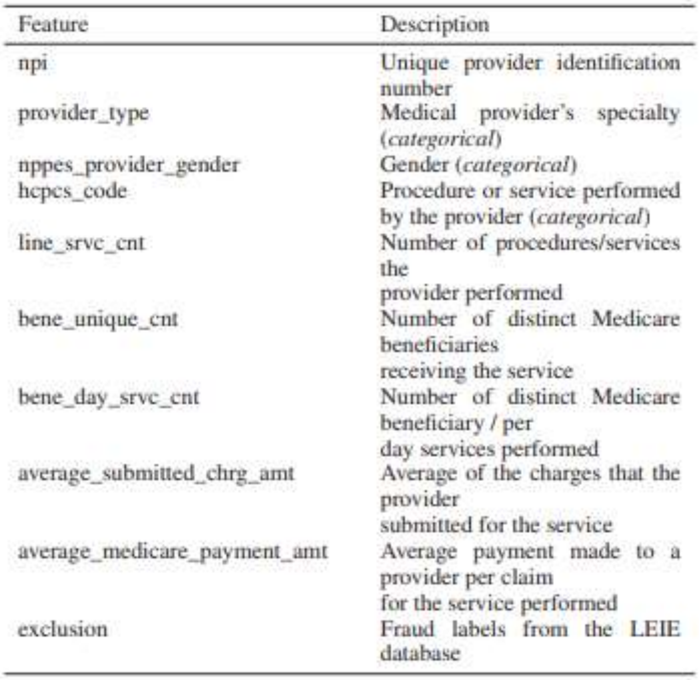
In this section, we detail the Medicare data, LEIE database, and the mapping of fraud labels. Additionally, we discuss the three learners, performance metrics, and class imbalance. Finally, we briefly outline our experimental design.

Data: The data in our experiment is from the Centers for Medicare and Medicaid Services (CMS) which encompass the 2012 to 2015 calendar years. The Medicare Provider Utilization and Payment

Data: Physician and Other Supplier describes payment and utilization claims data, with information on services and procedures provided to Medicare beneficiaries. The data was compiled and aggregated by CMS, grouping claims information by unique National Provider Identification (NPI) numbers, Healthcare Common Procedure Coding System (HCPCS) code, and place of service (e.g. office or hospital). The Medicare dataset contains values that are recorded after claims payments were made and with that, we assume that the Medicare dataset was appropriately recorded and cleansed by CMS (CMS Office of Enterprise Data and Analytics 2017). The combined Medicare dataset has 37,147,213 instances and 30 features, covering 89 specialties, and 1,080,115 distinct providers. We focus on detecting fraud using the features in Table 1.

Note that three features are categorical, with the remainder being numerical. The feature exclusion is the class variable that contains the fraud or non-fraud labels. NPI is not used in the model but retained for identification purposes. It is important to point out that because we merged all four years of Medicare data, the standardized payment variables are not included since these only appear in the 2014 and 2015 Medicare years. Similarly, the standard deviation variables were also excluded, because they pertain to 2012 and 2013 only. The possible use of the remaining variables, applying additional feature engineering, is left as future work.

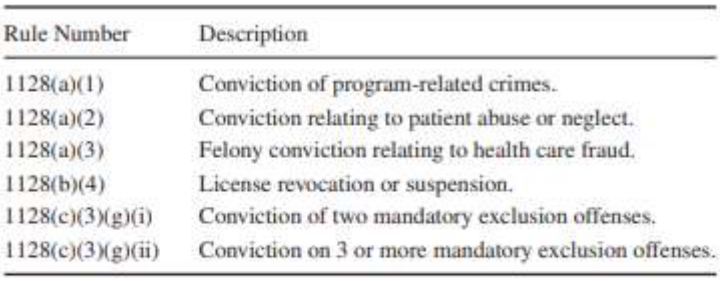
Table 1: Description of Medicare features



Feature	Description
npi	Unique provider identification number
provider_type	Medical provider's specialty (*categorical*)
nppes_provider_gender	Gender (*categorical*)
hcpcs_code	Procedure or service performed by the provider (*categorical*)
line_srvc_cnt	Number of procedures/services the provider performed
bene_unique_cnt	Number of distinct Medicare beneficiaries receiving the service
bene_day_srvc_cnt	Number of distinct Medicare beneficiary / per day services performed
average_submitted_chrg_amt	Average of the charges that the provider submitted for the service
average_medicare_payment_amt	Average payment made to a provider per claim for the service performed
exclusion	Fraud labels from the LEIE database

In order to obtain labels indicating fraudulent providers, we incorporate excluded providers from the List of Excluded Individuals/Entities (LEIE) database (LEIE 2017). The LEIE only includes NPI-level, or provider-level, exclusions, with no details on procedures (HCPCS codes) that contribute to the fraud. The exclusions are categorized by various rule numbers, which indicate severity as well as the length of time of each exclusion. We selected the providers excluded for more severe reasons, that are classified as mandatory exclusions by the Office of Inspector General (LEIE 2017), as seen in Table 2. The 1128(a) rules have five-year minimum periods, whereas rule 1128(c)(3)(g)(i) has a 10 year minimum period, and rule 1128(c)(3)(g)(ii) is permanent exclusion. More specifically, we label providers as excluded during the exclusion period only for the currently available Medicare years. These activities during the exclusion period can indicate a submission of claims for services by an excluded provider which are considered fraud under the federal False Claims Act (United States Code 2006). Even though the LEIE is limited in nature and does not contain National Provider Identification (NPI) number for most of the providers (Pande and Maas 2013), we decided to match on NPI only to accurately capture the known fraudulent exclusions. Moreover, due to the lack of detail in the LEIE database, we assume the excluded providers (NPI) include all of the corresponding procedures (HCPCS) performed for the exclusion period. Based on this assumption, all procedures performed by an excluded provider are considered fraudulent. Presently, there is no known publicly available dataset which includes fraud labels by provider and by each procedure performed, but future research will look at ways to mitigate this lack of data through majority voting or methods of NPI-level data aggregation.

Table 2: LEIE exclusion rules



Rule Number	Description
1128(a)(1)	Conviction of program-related crimes.
1128(a)(2)	Conviction relating to patient abuse or neglect.
1128(a)(3)	Felony conviction relating to health care fraud.
1128(b)(4)	License revocation or suspension.
1128(c)(3)(g)(i)	Conviction of two mandatory exclusion offenses.
1128(c)(3)(g)(ii)	Conviction on 3 or more mandatory exclusion offenses.

In combining the 2012 to 2015 Medicare datasets, we matched features and excluded those that did not match in all four years. For instance, in 2012 the standard deviations for charges and payments are available but discontinued for the later years. To provide fraud labels for the combined Medicare dataset, we cross-referenced NPI numbers in the Medicare data and LEIE database, to match any providers with past or current exclusions. In the LEIE database used for our study, only the 1128(a) rules were used which indicate a 5-year exclusion period. Note that only the year is available in the Medicare data not day or month, so we assumed that if a provider was excluded anywhere in a given year, all of those instances would get fraud labels. In order to map the LEIE fraud labels to the Medicare data, we first exclude providers who have been reinstated or have received waivers. Then, both start and end dates need to be set based on the maximum period of exclusion. In our case, five years was the maximum period, so we start five years prior to the first year of the Medicare dataset. This indicates that a provider could have been put on the exclusion list in 2008 and still be on the list in 2012 (which is the first year of the Medicare data), thus be labeled as fraud ulent for 2012. Similarly, we do the reverse process from the last year of the Medicare dataset and label providers accordingly. We take the disjunction of these start and end labels to get the list of excluded instances to be labeled as fraud. For examples, if a provider is placed on the exclusion list in 2009, then their claims are marked as fraudulent for 2012 and 2013, but not 2014 and 2015. Finally, we match this with the Medicare NPI numbers to generate the mapped fraud and non-fraud labels. These steps to map fraud labels help to mitigate over counting fraudulent providers due to overlapping or expired exclusion periods, thus we can be reasonably confident, with the stated assumptions, that we capture a fair number of fraud labels for the corresponding excluded providers. The final Medicare dataset, used in our experiments, has 3,331 instances labeled as fraudulent due to flagged providers with the remaining 37,143,882 instances being labeled as not fraudulent.

Learners: For our experiments, we built and test three different learners to classify fraudulent Medicare provider claims: C4.5 decision tree (C4.5), Support Vector Machine (SVM), and Logistic

Regression (LR). We chose these learners due to their popularity and relatively good performance in different classification-related domains. Each of these learners was built and tested using the Weka machine learning software (Witten et al. 2016). The default parameters are used and changes were made to these configurations when experimentation indicated increased performance based on preliminary analysis. The decision tree, C4.5, was trained using the J48 algorithm in Weka and configured with Laplace smoothing and no pruning as these have been shown to improve performance (Weiss and Provost 2003). Logistic Regression (LR) is a classification algorithm similar to linear regression except a different hypothesis class is used to predict the probability of class membership (Le Cessie and Van Houwelingen 1992). SVM in Weka incorporates sequential minimal optimization (SMO) for training the SVM models. We set the complexity parameter 'c' to 5.0 and the 'buildLogisticModels' parameter to true.

Performance Metrics: The classification models are evaluated using the AUC performance metric (Bekkar, Djemaa, and Alitouche 2013). AUC is a popular measure of model performance, providing a general idea of predictive potential of a binary classifier, and was chosen as the performance measure for our experiment because of the severe class imbalance of our testing data (Jeni, Cohn, and De La Torre 2013). The ROC curve is used to characterize the trade-off between true positive rate and false positive rate and depicts a learner's performance across all decision thresholds, i.e. a value between 0 and 1 that theoretically separate the classes. AUC is a single value that ranges from 0 to 1, where a perfect classifier provides an AUC value of 1. In order to gather more detail on learner performance, we also examine false positive rate (FPR) and false negative rate (FNR), with the instances labeled as fraud being the positive class. A classification threshold of 0.5 was used to assess these metrics for each learner. For the detection of Medicare claims fraud, a low FNR is most important since this indicates a higher detection rate for capturing actual fraudulent claims. Given the current manually intensive process in detecting fraud, we can generally accept a slightly higher FPR (i.e. claims predicted as fraud that are not actual fraud) as long as we obtain the lowest possible FNR. In practice, missing a substantial number of fraudulent events will render any fraud detection system ineffective, but, conversely, having too many false positives will make the system unusable. For our research, a learner with a low false negative rate and a reasonably low false positive rate is desired.

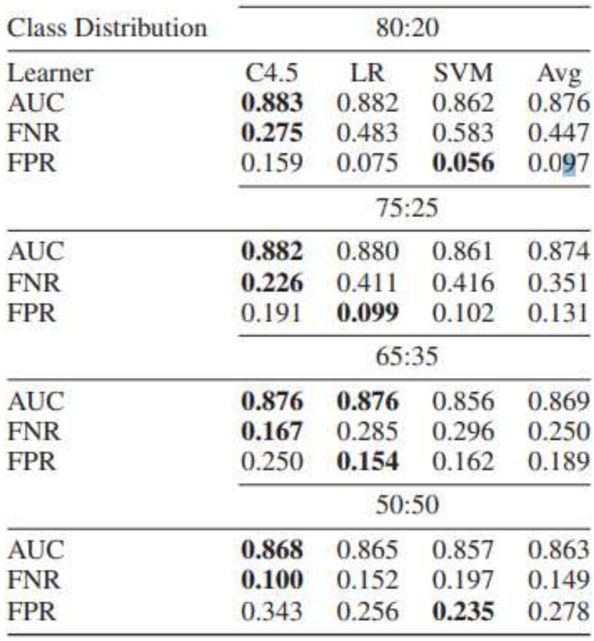
Class Imbalance: The Medicare claims data, with fraud labels, is a challenging dataset due to the skewed nature of the provider exclusions. With such class imbalance (Haixiang et al. 2017), the learner will tend to focus on the majority class (i.e. the class with the majority of instances), which is usually not the class of interest. In our case, the non-fraud labels are the majority class. An effective way to compensate for some of the detrimental effects of severe class imbalance is by changing the class distribution in the training data, to increase the representation of the minority class to help improve model performance. The sampling of data changes the class distribution of the training instances to minimize the effects of these rare events. Van Hulse et al. (Van Hulse, Khoshgoftaar, and Napolitano 2007) provide a comprehensive survey on data sampling techniques and their impact on various classification algorithms. There are two basic sampling methods: oversampling and undersampling. Oversampling is a method for balancing classes by adding instances to the minority class, whereas undersampling removes samples from the majority class. Oversampling can increase processing time by increasing the overall size. More critically, oversampling can overfit the data by making identical copies of the minority class. On the contrary, with undersampling, we retain all of the original fraud-labeled instances and randomly sample without replacement from the remaining majority class instances. In our study, we use random undersampling (RUS) with the following class distributions (majority:minority): 50:50, 65:35, 75:25, and 80:20. The selected class ratios retain a reasonable amount of the majority class and reduce loss of information relative to the minority (fraud labeled) class. In our experiment, we repeat the RUS process 10 times for each of the class distributions.

Experimental Design: We employ stratified 5-fold cross-validation to assess the performance of each of the learners (Witten et al. 2016). The reason we use 5-fold cross-validation is because of the extremely low percentage of fraud labels throughout the entire Medicare dataset. This reduces the likelihood that a fold has too few positive class instances and retains more equitable labeled data for fair evaluation. Moreover, to further reduce bias due to bad random draws and to better represent the claims data, we repeat the 5-fold cross-validation process 10 times and average the scores to get the final performance results.

4.Results and Discussion

In general, the results of our study do not necessarily point to one specific learner as the best overall performer across class distributions and performance metrics. Even so, C4.5 and LR both perform well, based on average AUC, across all class distributions, with C4.5 having the highest absolute AUC score. Table 3 details the performance results for all class distributions and learners, across all the performance metrics. From this, we can see that C4.5 and LR are indeed the best performing learners, with the general trend indicating worse performance as the minority class percentage increases. SVM has a deviation from this general trend with the lowest AUC at the 65:35 class distribution. At this point, based on AUC only, the best learner is C4.5 with an 80:20 class distribution.

Table 3: Performance results by class distribution



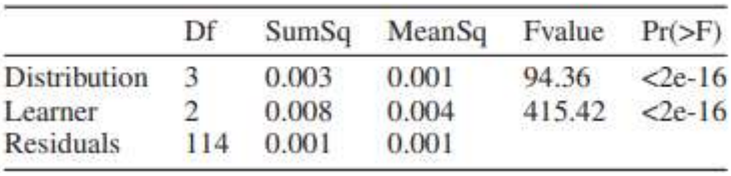
Class Distribution	80:20			
Learner	C4.5	LR	SVM	Avg
AUC	**0.883**	0.882	0.862	0.876
FNR	**0.275**	0.483	0.583	0.447
FPR	0.159	0.075	**0.056**	0.097
	75:25			
AUC	**0.882**	0.880	0.861	0.874
FNR	**0.226**	0.411	0.416	0.351
FPR	0.191	**0.099**	0.102	0.131
	65:35			
AUC	**0.876**	**0.876**	0.856	0.869
FNR	**0.167**	0.285	0.296	0.250
FPR	0.250	**0.154**	0.162	0.189
	50:50			
AUC	**0.868**	0.865	0.857	0.863
FNR	**0.100**	0.152	0.197	0.149
FPR	0.343	0.256	**0.235**	0.278

As discussed, additional metrics, including FPR and FNR, are used to further assess learner performance across class distributions. It is important to use other measures of learner performance to help gauge actual detection capabilities, particularly when the correct detection of real fraud cases is more important than detecting non-fraud ones. From Table 3, we again note that C4.5 has the highest AUC for each class distribution, with LR being very close to C4.5 in average AUC. Because we wish to catch as many actual fraudulent providers as possible, we require a learner with a low FNR to correctly identify positive class instances. However, there is a tradeoff between the number of actual fraud instances detected and false positives. As stated, the detection of actual fraudulent providers is the primary purpose of any fraud detection approach, thus using the learner with a low FNR is critical, even at the cost of injecting additional false positives. The C4.5 decision tree learner has the lowest FNR for every class distribution, but also the highest FPR. The lowest FPR scores alternate between LR and SVM, depending on the class distribution. Even though LR and C4.5 have similar AUC scores, LR has higher false negative rates. Given our need for the accurate detection of actual fraud, the C4.5 learner is the best choice with the highest AUC and the lowest rate of false negatives. In order to provide additional rigor around our results and recommendations, we evaluated the statistical significance of our AUC results with a two-factor ANalysis Of VAriance (ANOVA) and Tukey's Honest Significant Difference (HSD) tests, at a 95% confidence level (Sargin and others 2009).

Table 4 shows that both the class distribution and learner factors are significant. To further elucidate the specifics per factor, we performed a Tukey's HSD test outlined in Figure 5.

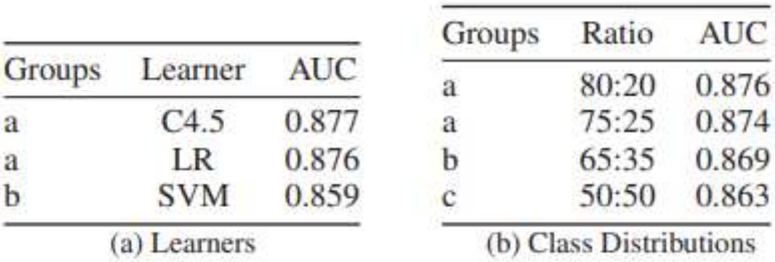
Table 5a, in the Tukey's HSD results table, confirms that C4.5 and LR are significantly better than SVM. Table 5b shows that the only significant difference is seen in the 65:35 and 50:50 class distributions. These results, as noted, are for AUC and do not directly reflect the FNR or FPR of each learner but do highlight the need to understand the domain and investigate other metrics to assess overall model performance and fraud detection capabilities.

Table 4: ANOVA results



	Df	SumSq	MeanSq	Fvalue	Pr(>F)
Distribution	3	0.003	0.001	94.36	<2e-16
Learner	2	0.008	0.004	415.42	<2e-16
Residuals	114	0.001	0.001		

Table 5: Tukey's HSD results



Groups	Learner	AUC
a	C4.5	0.877
a	LR	0.876
b	SVM	0.859

(a) Learners

Groups	Ratio	AUC
a	80:20	0.876
a	75:25	0.874
b	65:35	0.869
c	50:50	0.863

(b) Class Distributions

5. Conclusion

Medicare fraud is a major contributor to high overall healthcare expenses and costs, particularly for the growing elderly population. The reduction of fraud and the recovery of costs is of utmost importance to maintain proper health and well-being. In our study, we present an effective approach to detect Medicare fraud leveraging the LEIE database for provider fraud labels. Additionally, the merging of the Medicare data with LEIE fraud labels is outlined which reduces the potential for over representation of fraud labels. The use of random undersampling is highlighted in exhibiting good fraud detection capabilities with different learners. Our research demonstrates the efficacy of using known fraud labels coupled with RUS to detect fraudulent Medicare providers. Since our focus is on detecting actual fraudulent providers, we require a model with a high AUC and low false negative rate. We demonstrate that C4.5 is the best overall learner with the 80:20 class distribution, with an AUC of 0.883, and the lowest false negative rates. In our study, we show that using RUS with big data can successfully detect fraudulent Medicare providers. Continued research includes acquiring additional LEIE fraud labels using other methods, such as fuzzy string matching, and other data sources. Additionally, performing experiments by specialty to simulate real-word fraud detection performance will be pursued. Finally, we intend to account for NPI-level LEIE exclusions, rather than assumed NPI and procedure-level exclusions, in assessing fraud using the Medicare data

References

[1]. Backman, M. 2017. 10 jaw-dropping stats about medicare.

[2]. Bauder, R. A., and Khoshgoftaar, T. M. 2016. A probabilistic programming approach for outlier detection in healthcare claims. In 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 347–354.

[3]. Bauder, R. A., and Khoshgoftaar, T. M. 2017. Multivariate outlier detection in medicare claims payments applying probabilistic programming methods. Health Services and Outcomes Research Methodology 1–34.

[4]. Bauder, R. A.; Khoshgoftaar, T. M.; Richter, A.; and Herland, M. 2016. Predicting medical provider specialties to detect anomalous insurance claims. In Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on, 784–790. IEEE. Bauder,

[5]. R. A.; Khoshgoftaar, T. M.; and Seliya, N. 2017. A survey on the state of healthcare upcoding fraud analysis and detection. Health Services and Outcomes Research Methodology 17(1):31–55.

[6]. Bekkar, M.; Djemaa, H. K.; and Alitouche, T. A. 2013. Evaluation measures for models assessment over imbalanced data sets. Iournal Of Information Engineering and Applications 3(10).

[7]. Branting, L. K.; Reeder, F.; Gold, J.; and Champney, T. 2016. Graph analytics for healthcare fraud risk estimation. In Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on, 845–851. IEEE.

[8]. Chandola, V.; Sukumar, S. R.; and Schryver, J. C. 2013. Knowledge discovery from massive healthcare claims data. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 1312– 1320. ACM. CMS Office of Enterprise Data and Analytics. 2017. Medicare Fee-For-Service Provider Utilization & Payment Data Physician and Other Supplier. CMS. 2017. Centers for Medicare and Medicaid Services: Research, Statistics, Data, and Systems.

[9]. Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; and Bing, G. 2017. Learning from class-imbalanced data: review of methods and applications. Expert Systems with Applications 73:220–239.

[10]. Herland, M.; Bauder, R. A.; and Khoshgoftaar, T. M. 2017. Medical provider specialty predictions for the detection of anomalous medicare insurance claims. In Information Reuse and Integration (IRI), 2017 IEEE 18th International Conference, 579–588. IEEE.

[11]. Jeni, L. A.; Cohn, J. F.; and De La Torre, F. 2013. Facing imbalanced data–recommendations for the use of performance metrics. In Affective Computing and Intelligent Interaction (ACII), 2013

[12]. Humaine Association Conference on, 245–251. IEEE. Joudaki, H.; Rashidian, A.; Minaei-Bidgoli, B.; Mahmoodi, M.; Geraili, B.; Nasiri, M.; and Arab, M. 2015. Using data mining to detect health care fraud and abuse: a review of literature. Global journal of health science 7(1):194.

[13]. Le Cessie, S., and Van Houwelingen, J. C. 1992. Ridge estimators in logistic regression. Applied statistics 191–201. LEIE. 2017. Office of inspector general leie downloadable databases. Medicare. 2017. US Medicare Program. NHCAA. 2017. The National Heath Care Anti-Fraud Association. OIG. 2017. Medicare Fraud Strike Force.

[14]. Pande, V., and Maas, W. 2013. Physician medicare fraud: characteristics and consequences. International Journal of Pharmaceutical and Healthcare Marketing 7(1):8–33.

[15]. Sadiq, S.; Tao, Y.; Yan, Y.; and Shyu, M.-L. 2017. Mining anomalies in medicare big data using patient rule induction method. In Multimedia Big Data (BigMM), 2017 IEEE Third International Conference on, 185–192. IEEE.

[16]. Sargin, A., et al. 2009. Statistics and data with r: An applied approach through examples. Journal of Statistical Software 30(b06). United States Code. 2006. Supplement 5, title 31, sec. 3729 - false claims. U.S. Administration for Community Living. 2015. Profile of older Americans: 2015.

[17]. Van Hulse, J.; Khoshgoftaar, T. M.; and Napolitano, A. 2007. Experimental perspectives on learning from imbalanced data. In Proceedings of the 24th international conference on Machine learning, 935–942. ACM.

[18]. Weiss, G. M., and Provost, F. 2003. Learning when training data are costly: The effect of class distribution on tree induction. Journal of Artificial Intelligence Research 19:315–354.

[19]. Witten, I. H.; Frank, E.; Hall, M. A.; and Pal, C. J. 2016. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.