Assessment of Healthcare Claims Rejection Risk Using Machine Learning

Anundhara Chimmad
Anundhara.chimmad@edifecs.com

Prasad Saripalli
Prasad.saripalli@edifecs.com

Venu Tirumala
Venu.tirumala@edifecs.com

Edifecs Inc.
Bellevue, WA. USA

***Abstract* — Modern healthcare service records, called Claims, record the medical treatments by a Provider (Doctor/Clinic), medication advised etc., along with the charges, and payments to be made by the patient and the Payer (insurance provider). Denial and rejection of healthcare claims is a significant administrative burden and source of loss to various healthcare providers and payers as well. Automating the identification of Claims prone to denial by reason, source, cause and other deciding factors is critical to lowering this burden of rework. We present classification methods based on Machine Learning (ML) to fully automate identification of such claims prone to rejection or denial with high accuracy, investigate the reasons for claims denial and recommend methods to engineer features using Claim Adjustment Reason Codes (CARC) as features with high Information Gain. The ML engine reported is first of its kind in Claims risk identification and represents a novel, significant enhancement to the state of practice of using ML for automating and containing claims denial risks.**

Keywords-Healthcare, Claim, Rejection, Denial, Risk, Classification, Support Vector Machines

I. Introduction

Medical billing process is an essential interaction system between Healthcare Provider and the Insurance company (Payer), initiated by the Provider, registering patient details like name, age, gender, address, health insurance plan, reason for visit and a unique id such as the Social Security Number. During the treatment, type of diagnosis done, treatments and medication prescribed are all recorded to maintain a patient history. A trained coding specialist converts these into alphanumeric codes, using several standards accepted universally, like ICD codes to record diagnosis and CPT codes for treatments, and the resulting claims are sent to a third party adjudication system (clearinghouse or insurance company) for further processing.

Millions of claims across various service providers are generated and processed accurately. However, a significant fraction of claims is either incorrect or has some missing fields which cannot be processed as is and needs to be reprocessed. According to AMA, between 1.38 percent and 5.07 percent of claims are denied by insurers on the first submission. Even the best-performing medical practices experience a denial rate of 5%. Without an effective strategy for identifying and managing denial risk, providers are more apt to see denials unfavorably resolved or, eventually written off as bad debt. This extra administrative work to reprocess claims is called rework, which impacts the total revenue of the service provider. In an average 300-bed hospital, a 1% - 5% denial rate can mean $2 million to $10 million dollars a year - significant by any standards [1].

Analysis of Rejected and Denied Claims

Assessment of claim rejection risk is essential for increasing accuracy and reducing errors in processing the claims, which has a major impact on the revenue cycle. Reasons for the denial of a claim include incorrect patient identifier information, claim billed to wrong insurance company, lack of technical experience of medical biller in charge, filing a claim after the deadline, medical biller filling the claim with incorrect CPT and ICD codes [2] or leaving out codes altogether, patient not covered by insurance policy for the service, missing referrals from primary service provider for certain procedures etc. A claim denied by the Payer must be resubmitted by the Provider after necessary changes. Until recently, there were no satisfactory methods available to automatically identify claims according to the risk of denial.

Claims adjudication systems historically employed random manual audits and rule-based queries to identify the claims with potential risk of rejection, which require specialized domain experts to and are not very accurate. ML offers a promising, new approach to improve this. Kumar et al. [3] used Binary Classification for predicting claims that will need to be reworked, using claims data from a large US health insurer, and found an order of magnitude better precision (hit rate) over existing approaches. Ghani and Kumar [4] reported on an interactive prioritization component built on top of the batch classifier from [3] and showed significant reduction in claim audit results obtained from applying this system at two major US health insurance companies indicating. Wojtusiak *et al.* [5] reported on using attributional rules for predicting potential discrepancies in claims which were able to detect abnormal claims prior to their submission. Other works reported using ML methods such as rule-based learning and NLP to analyze claims focused on fraud detection, which is a different problem from claims rework assessment. [6,7,8,9]

We report on a comprehensive classification engine using multiple ML algorithms (Binary Classification Trees, Neural Networks (NN), Support Vector machines (SVM) and Naive Bayes Classification, to predict whether a given claim is likely to involve rework (i.e., has a high risk of rejection or denial) via binary classification. Engineering a dataset of high information value, by selecting and/or deriving features based on CARC codes, which have a strong influence on the classification of claims is a key contribution of this work. There are no prior reports, which provide similarly comprehensive engineering of features based on Claim Adjustment Reason Codes (CARC) [10], for the first time. Further, synthetically engineering features is shown to add to improving the accuracy of predictions.

II. FEATURE ENGINEERING AND DATA

There are many reasons for claims denial, chief among them being the claimed service or procedure not being covered by the insurance provider for the particular patient's policy. For any authorized surgical procedures or use of expensive medical equipment, there is a timeframe within which payments must be made. However, if a hospital or physician misses the specified deadline, then payments will not be made further causing rejection of a claim. In this paper, we present methods of classifying the claims according to their likelihood of being accepted/rejected. These classifications are trained machine learning algorithms help us analyze denied claims and see if there are trends that cause denials. For example, claims with one particular insurer are rejected more often than others or a particular diagnosis code leads to more denials. A simple change in insurer requirements that goes unnoticed could possibly be an important reason for rejection. We predict claims that are more prone to denial and rectify before sending it off to associated parties for payments, to effectively reduce administrative costs, time and increase the performance of overall medical billing process. Objective of the ML methods presented is to classify claims before submission as highly likely to be rejected (*i.e.,* high risk of rework – labelled as 200) or unlikely to be rejected (normal claims – labelled as 100). Features contributing to this classification are derived from domain analysis with an expert, who prescribed the reasons for claims denial or rework (Table 1.)

Most common reasons for claims denial are: the procedure/revenue code is inconsistent with the patient's age; diagnosis is inconsistent with the procedure; authorization number is missing, invalid, or does not apply to the billed services or provider; claim lacks information needed for adjudication. Such reasons are converted into Claim Adjustment Reason Codes (CARC) by CMS (Table 1).

A. Dataset Preparation:

We built an ensemble of ML algorithms to classify claims against denial risk, using a dataset of 10000 accepted claims and 3000 denied claims from a prominent health insurance companies in the U. S. Days of service is the number of days elapsed starting from the date of service and payment date.

One of the authors, an expert medical claims analyst on our team inspected all of the claims manually, to identify and codify the specific reasons for denial of the claims using the CARC claim rejection codes. The analyst has prepared the training dataset by identifying thousands of medical claims, matching each Medical Claim document from a Provider to a Payer with the corresponding Remittance Advice document generated by the Payer to the Provider in response. This later document contains the Reason Code descriptions shown in Table 1. A total of 8607 such claims were included in the dataset. While this data regarding Payers and Providers is of exploratory interest, details cannot be provided to preserve anonymity. Such details are not critical for the risk calculations and prediction reported here.

The original dataset so constructed included the following 14 features: Patient Name, PatientCtrl #, Subscriber ID, ICN/DCN, date of Service (DoS), Provider Name, Payer Name, Batch ID, Payment Date, Upload Date, Payment #, Total Charge, Total Pay, Status and RSN1. Among these DoS represents Date of Service, RSN is the reason code for claim denial (per Table 1) and ICN/DCN is the unique Provider (Physician ID). Shown in Figure 1 is a Pie Chart showing the relative frequencies of each error reason code (CARC) in the total sample of 8607 claims.

TABLE 1. Summary of rejection Codes with Reason

[CARC] Reason for claim rejection
[45] Charge exceeds fee schedule/maximum allowable or contracted/legislated fee arrangement.
[15] The authorization number is missing, invalid, or does not apply to the billed services or provider.
[22] This care may be covered by another payer per coordination of benefits.
[29] The time limit for filing the claim has expired.
[97] The benefit for this service is included in the payment/allowance for another service/procedure that has already been adjudicated.
[58] Treatment was deemed by the payer to have been rendered in an inappropriate or invalid place of service.
[16] Claim/service lacks information or has submission/billing error(s), which is needed for adjudication.
[18] Exact duplicate claim/service

Exploratory Data Analysis

Results from statistical analysis on the claims dataset containing a total number of 8,604 claims (rejected or faulty claims only) are summarized in Figure 1. Also summarized below Figure 1 are the numbers of key metrics related to the dataset. To preserve confidentiality, all references to Provider and Payer entities contained in the dataset are anonymized. Frequencies of error reason codes are shown in Figure 1, from 8604 claims rejected for various reasons.

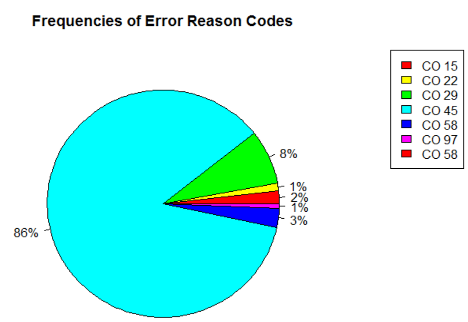

Figure 1. Frequencies of error reason codes

Number of Service Provider families (Groups) 2
Total number of service providers 15
Total number of Payers 27
Total number of Doctors (Unique ICN DCN): 10131
Total number of patients 13000

By mixing data tuples from denied claims and accepted (Normal) claims into a single dataset, a Training and Test datasets are prepared for the ML experiments which follow.
It can be seen that Distribution of frequencies of error reason codes (RSN) are as follows [Error code (%): 45(86%), 29(8%), 58(5%), 97(1%), 22(1%)]. Thus, Code 45 (Charge exceeds fee schedule/maximum allowable or contracted/legislated fee arrangement.) is a major reason for claims rework, followed by 29 (outdated claim). This is consistent with the Industry observation that outdated claims and excessive charges are key reasons for claims denial.

III. BINARY CLASSIFICATION APPROACH

ML experiments were conducted using three different approaches to feature engineering as follows: (1) CARC Codes (RSN1) included in the dataset as a single 'Status' feature without one-hot coding; (2) same as above but also converting the CARC codes as eight distinct features using one-hot (or Dummy) coding and (3) engineering 4 additional (synthetic) features added to the dataset. These are respectively referred to as Classification Run (CR) 1, 2 and 3. This detail is summarized in Table 2.

TABLE 2. Summary of ML Simulations

Classification Run #	Feature construction approach
CR1. CARC codes used as a single Status feature	Assumes that the individual differences in CARC codes do not influence classification
CR2 CR1 dataset plus 4 additional synthetic features:	Synthetic features (Non-covered charges, Coding errors, Duplicate claims and Overlapping claims) engineered to assess how they enhance classification accuracy.
CR3 Adding the CARC codes as distinct features using one-hot (or Dummy) coding	Assumes that the individual differences in CARC codes significantly influence classification

Three different Binary Classification algorithms (Classification Trees, SVM and Neural Networks) were used to classify the claims into High risk or Low risk for Denial. Training data sets were prepared by adding a status field of 100 (accepted) for all the claims with reason number (RSN) 100 and 200 (denied) for the remaining claims. This means the various error reason codes in Table 1 are al combined into either a 100 or 200 (accepted/denied) dichotomous variable. Mohit Kumar et al. [2] reported high hit rate using a similar approach. Results from the Classification Run 1 (CR 1) are presented in Table 2.

CR #1A: After preparing the dataset of size (# 13000 (10000 accepted and 3000 denied claim instances) with 3 major attributes – days of service, total charge amount and total paid amount, we randomly sampled 75% of this dataset and trained this subset using a Classification Tree (CART) algorithm using the R ML package. This algorithm classifies the dataset of mixed mode variables by recursively partitioning the claims data space and fitting a simple prediction model within each partition, which are graphically represented as a decision tree [12]. Claims with days of service greater 140 were denied. Repeating the simulation CR #1A several times, we found that on an average, claims with days of service greater than 140 were denied and needed rework, which is consistent with Figure 1 and justified by the regression tree below. Further, it can be seen (Figure 2) that Total Charged Amount on the claim is the key deciding feature, followed by the Total Amount Paid and the Duration of service and claim submission cycle (Days). It stands to reason that claims involving excessive charge amounts or long delays in processing are flagged for rework.

CR #1B and #1C: On the same dataset, Neural Networks are used in CR #1B and Support Vector Machines was used to classify the dataset in CR #1C, using the libraries native in R. Results from these experiments indicate an Accuracy of 77% and a Precision of 77%, in contrast to a 70% Accuracy and an 81% Precision with the CART Decision Tree algorithm (CR #1A). In Figure 2, Classification of claims as potential Denial or Acceptance (100) is influenced mainly by the Total Charge Amount of the claim, followed by the Total paid Amount and the Duration of the process. Influence of CARC codes' as separate features is not seen in these results, confirming that better feature engineering by separating out their influence is necessary (seen from CR2 results in the following sections).

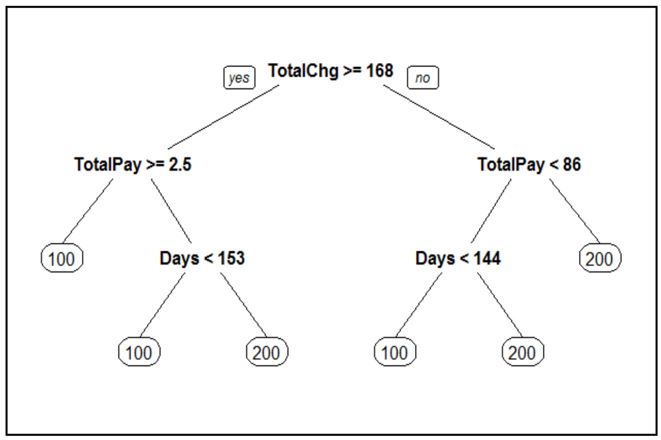

Figure 2. Binary Classification Tree (Run CR #1A)

Accuracy Evaluation

For all simulations in this study, standard metrics for Accuracy and Precision were used [11].

TABLE 2. Accuracy Metric for the Predictions: Run CR1

Algorithm		**Predicted: 100**	**Predicted: 200**
Classification Tree [0.70 /0.81]	**Actual: 100**	2474	569
	Actual: 200	569	192
Neural Network [0.77 / 0.77]	**Actual: 100**	2512	0
	Actual: 200	738	0
SVM [0.77 / 0.77]	**Actual: 100**	2511	0
	Actual: 200	739	20

Accuracy ((TP+TN)/Total) and Precision (TP/Total Predicted 200)) are shown in parenthesis as [A/P] in Table 2.

In Figure 3, results from the SVM simulation (CR 1B) are shown in a 2D space of the two key determining features – Total Chare Amount (TotalChrg) and the Duration (Days). It can be seen that the algorithm is able to partition and separate the rejection-prone (200) and normal (100) claims. Mark ‘x’ in figure 3 represents the support vectors and ‘o’ are the data points other than support vectors. We see two colors for support vectors as there are two values to the dependent variable (100 or 200); while majority of the claims are normal (Blue: 100), the Magenta (200) partition indicates high total charge amount as prone to rejection.

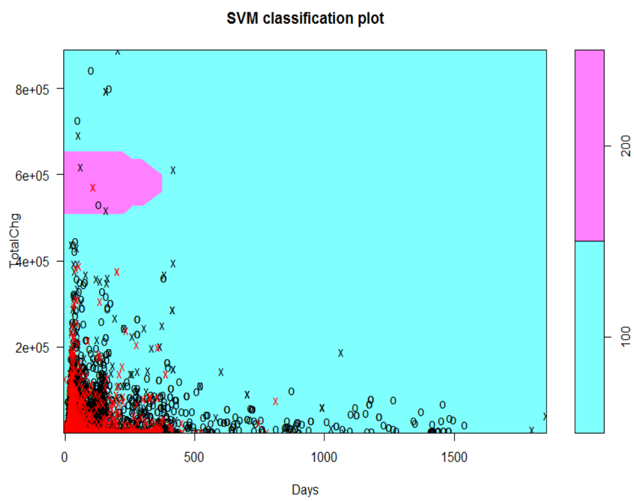

Figure 3. SVM Classification (Run CR #1B)

Classification tree algorithm outperforms other algorithms in predicting the total number of accepted claims (RSN: 100) and rejected claims (RSN: 200). Findings from CR #1A indicate a key rule for a claim being highly probable is when the total Charge amount is > 168K AND Total Payment amount is > 86K OR the number of days > 144 days (140 on an average). Overall, results indicate that binary classification approaches help automate claim denial identification, but the Accuracy needs improvement from the current 70 to 77% result. To achieve this, we have designed additional features including the CARC Codes (Table 1).

IV. Enhancements via Additional Attributes

Enginering Additional (Synthetic) Features

When an expert analyst is available to identify and codify the specific reasons for denial of the claims using the CARC claim rejection codes, it is feasible to identify and categorize the claims as potential denial risks with a reason code. However, such transcription is very cumbersome, costly and requires expensive and seldom available expert scrutiny. Most of the providers such as Clinics and Small Practices do not have access to such expertise. Goal of the ML engines and apps such as described in this paper as well as is to automate this process with as much less expert intervention as possible. To this end, it would be helpful to further automate the feature generation process explained in the earlier sections.

Inspection of the different reasons in practice why claims are denied or sent for rework would be helpful in this process. From the – CARC Codes available [8], we have assigned a probability rank (H/M/L) to each feature for its likely influence on claim denial. Our goal is then to engineer at least one feature (attribute) which can be derived from the claims dataset. These features are also described below, against each reason for denial.

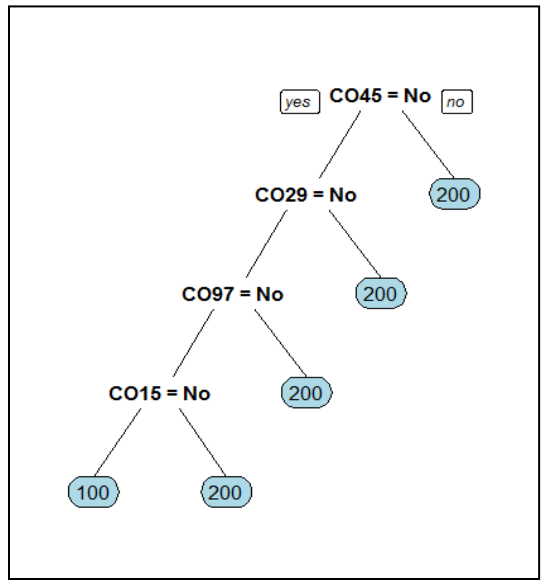

Figure 4. Classification Tree with Synthetic Features

1. Non-covered Charges: CMS rules clearly specify which services are covered by a certain plan and which are not. Typically, ICD-9 or ICD-10 codes are used to specify these eligible charges. An Attribute can be derived from 837 claims by combining the ICD (or DRG) code, the patient's Plan ID checked against a payer's database of eligible services, to return a Covered/Not Covered (C/NC) decision.

2. Coding Errors: Applications focusing on claims adjudication are able to inspect whether or not the claim was properly coded by the Provider to a Payer are correct, before submission to a payer, by verifying attributes such as format, allowable values, required presence and data integrity. This processes determines whether or not the claim was properly coded by the provider. An Attribute can be derived by historic observations of claims submitted by a Provider or Agency for how error-prone the process is, on a 1-10 Likert scale. Past errors usually indicate future errors in a process.

3. ***Duplicate Claims:*** Duplicate claims are multiple claims pertaining to a single encounter with the same values for the date, provider, beneficiary, and service or item fields, which will be denied with error code CO18. Duplicates are one of the largest reasons for Medicare Part B claim denials, by as much as 32%. In some cases, the provider may be able to justify them valid claims for payment, by applying the

correct condition codes or modifiers in rework. An attribute can be derived by analyzing the dataset for duplicates.

4. ***Overlapping Claims:*** According to CMS, claims overlap when "... the date of service or billing period of one claim seems to conflict with the date on another claim, indicating that one of the claims may be incorrect." For example, an N347 code indicates that "... claim for a referred or purchased service cannot be paid because payment has already been made for this same service to another provider by a payment contractor representing the payer". Similarly, an M86 code indicates that "... payment already made for same/similar procedure ...." An attribute can be derived by analyzing claims using a Rule to identify such conditions.

5. ***Expired Time Limit:*** Time limits are set by all insurers on claims submissions. Medicare requires all claims be filed within 12 months following the date of service. By calculating the difference in time stamps (dates) between the service date and submission date of a claim, this could be used as an attribute tag to identify claims in denial risk. A derived attribute can be engineered as the time difference between the last prescription date and the service date, to determine if it exceeds the typical prescription currency period of days and flagged as a contributor to denial risk.

Following this approach, we created a synthetic dataset to include 4 additional attributes. Coding errors typically are addressed by Claim Scrubber software, intended for an intricate inspection and editing of a claim prior to submission. We assume that a good scrubbing step is applied before claims are submitted and hence we only add only four of the above five derived attributes (Table 3). After including these synthetic attributes, a random sample of 3000 claims from the denied dataset and 10000 claims from approved dataset forms a 13000 claims dataset and again the following algorithms are implemented. Shown in Table 4 are the results from (CR #2) using the synthetic data.

TABLE 3. Derived attribues added to the dataset in CR2

Duplicate Claims: 3% of claims duplicated from 8543 denied claims Total number of accepted claims - 36853
Expired Time Limit: 'ExpireTimeLimit' is added which is a 'Yes' if the number of days (delta for date of service and payment date) is greater than 45.
Likely29: RSN code 29 corresponds to claim expiration time limit. Setting 30 days as this limit, if the number of days is greater than 30, this value will be a 'Yes', else 'No'
Likely45: Flag as a 'Yes' if charges exceed fee schedule / max allowable amount ($45000), else flag as 'No'.

Results indicate improved Classification (Figure 4) compared to CR 1. Accuracy of all 3 algorithms improved but Precision deteriorated in case of SVM and Neural Networks. Both Amount Paid and Amount Charged have a marked influence on the classification. However, accuracy of prediction (Table 4) did not show significant improvements, indicating that the original feature set is adequate for the classification. SVM algorithms now correctly predicts a few more denied claims as denied. While this experiment confirmed the usefulness of enhancing the dataset with additional predictors, the particular synthetic data (4 features added) did not lead to higher hit rates. Adopting a different approach to feature engineering using CARC codes led to significantly better results, reported in the following section.

TABLE 4. Accuracy Metric for the Predictions: Run CR2

Algorithm		**Predicted: 100**	**Predicted: 200**
Classification Tree [0.82/0.25]	**Actual: 100**	2474	569
	Actual: 200	15	192
Neural Network [0.77/0]	**Actual: 100**	2512	738
	Actual: 200	0	0
Support Vector Machine [0.77/1]	**Actual: 100**	2511	739
	Actual: 200	0	0

V. ONE-HOT FEATURE CLASSIFICATION APPROACH

In this approach, we retain the separate Reason Codes (CARC) as individual attributes (CO 15, 22, 29, 45, 97, 58; OA 16, 18, 45; PI 22, 29 and PR 19, 119) thus adding 11 13 additional attributes to the dataset. We then repeat the equivalents of simulation runs CR 1 to 3, using Classification Trees, Neural Networks and SVM. This approach led to significant gains in both Accuracy and Precision of classification, especially with Decision Trees and SVM. It can be seen that classification tree identified claims based on CO45 and CO29 (Figure 5) as the key deciding attributes.

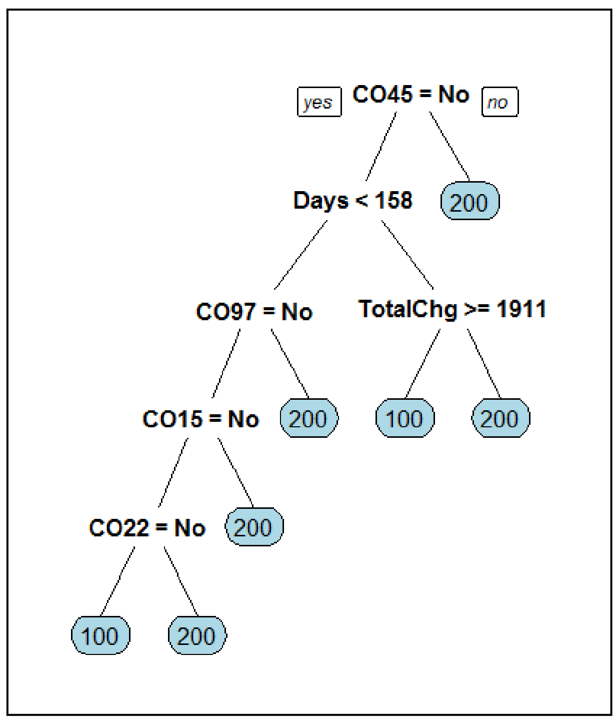

Figure 5. Multi- Classification Tree (Run CR2A)

Shown in Figure 3 and Table 4 are the results from simulation run set CR #2. For CR2 runs, adding the individual influence of CARC codes by separating them as one-hot coded attributes clearly improved the performance of the classification engines, in case of GLM and SVM models. Overall, this approach showed significant improvements in the accuracy of prediction.

TABLE 5. Accuracy Metric for the Predictions: Run CR3

Algorithm		Predicted: 100	Predicted: 200
Classification Tree [0.98/0.97]	Actual: 100	1668	70
	Actual: 200	13	1992
Neural Network [0.45/0]	Actual: 100	1681	2069
	Actual: 200	0	1
Support Vector Machine [1/1]	Actual: 100	1681	0
	Actual: 200	0	2069

Binary Classification Tree algorithm yielded an Accuracy of 0.98 and Precision of 0.97 (Table 5), whereas the SVM algorithm yielded 100% Accuracy and Precision. When the data patterns include multiple complex influences, SVM is known to perform better than decision trees.

Neural Networks did not yield good performance, as can be seen by the poor Accuracy and Precision metrics (Table 3). Reasons for this could be overfitting nature of Neural Networks in the absence of innate complexity of patterns in the dataset. In contrast, CART and SVM algorithms were able to identify the few key deciding features for classification as can be seen from Figure 3. SVM training always finds a global minimum, whereas the performance of neural networks may sometimes suffer from the existence of multiple local minima solutions [13], and SVM perform better comared to Decision Trees in earlier studies [14, 15]. Results from CR3 runs, especially the GLM and SVM algorithms, clearly indicate that one-hot coding the CARC Code as separate features significantly improves the classification performance. This is an encouraging, novel finding which is likely to be of significant practical use to the claims adjudication process automation using ML, which itself is a relatively new practice. There are only three other significant works reported using ML for this purpose [3-5, 16] which did not report using CARC codes as key features.

CONCLUSIONS

We presented a novel classification framework for the identification of medical claims likely to be rejected during the submission process. Adoption of CARC codes as key features predicting such denial and rework in the classification datasets and its automation via one-hot coding, yielded significant improvements in prediction accuracy and precision. This represents a new advancement in the very recent efforts to automate the claims adjudication processing for rework identification using ML. Our attempt to include 4 additional synthetic attributes led to moderate additional improvement from classification using the base datasets. However, the one-hot coding approach using CARC codes as additional features led to very significant improvements in both accuracy and precision of classification (>99% Accuracy and Precision) Synthetic addition of features is now being investigated on weaker datasets, where the influence of CARC codes cannot be captured by detailed manual analyst inputs a priori. Overall, the reported work represents a novel, significant enhancement to the state of practice of using Machine Learning for automating and enhancing claims denial risks.

ACKNOWLEDGMENT

Authors thankfully acknowledge the support from Edifecs Inc., which made the collection of rare de-identified datasets and implementation of the reported work possible.

REFERENCES

[1] Lachney, K. (2016) Medical Billing Denials Are Avoidable: How to Help Prevent the Top 5 (2016) Mckesson Technologies Blog online.

[2] Guide to Common Claim Rejections, Careo, Inc. (2015).

[3] Kumar, M., Ghani, J. and Zhu-Song, M. (2010) Data Mining to Predict and Prevent Errors in Health Insurance Claims Processing Proc. 16th ACM SIGKDD.

[4] Ghani, J. and Kumar, M. (2011) Interactive Learning for Efficiently Detecting Errors in Insurance Claims, KDD'11, August 21–24, 2011, San Diego, California, USA.

[5] Wojtusiak, J., Ngufor, C., Shiver, J., Ewald, R., "Rule-based Prediction of Medical Claims' Payments: A Method and Initial Application to Medicaid Data" Proc. Int. Conf. on Machine Learning and Apps, Honolulu, HI, December 2011.

[6] A. Anand and D. Khots (2008). A data mining framework for identifying claim overpayments for the health insurance industry. INFORMS Workshop on Data Mining and Health Informatics.

[7] WS Yang, SY Hwang 92006) A process-mining framework for the detection of healthcare fraud and abuse, Expert Systems with Applications, 2006 – Elsevier.

[8] C.S. Hilas, P.A. Mastorocostas, "An application of supervised and unsupervised learning approaches to telecommunications fraud detection, Knowledge-Based Systems," 21, 7, 2008, 721-726.

[9] E.W.T. Ngai, Yong Hu, Y.H. Wong, Yijun Chen, and Xin Sun, "Th e application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature" Decision Support Systems: On quantitative methods for detection of financial fraud, vol. 50, iss. 3, February 2011, pp. 559-569.

[10] CMS Claim Adjustment Reason Codes (2017) Washington Publishing Co., Washington D. C.

[11] E. Frank, I H. Witten and Hall, M. A. (2016) Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann.

[12] Loh (2011), Classification and regression trees, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.

[13] Suykens, JAK et al. (2002) Least Squares Support Vector Machines, World Scientific Publishing Company, Singapore.

[14] Y. Shao, R. S. Lunetta, Comparison of SVM, neural network, and CART algorithms for land-cover classification using limited training data points, ISPRS J. Photo. Remote Sensing, 70(6):2012.2012.

[15] H.Z.M. Shafri and F.S.H. Ramle, 2009. A Comparison of Support Vector Machine and Decision Tree Classifications Using Satellite Data of Langkawi Island. Information Technology Journal, 8: 64-70.

[16] Abdullah, U. (2009) Comparative Study of Medical Claim Scrubber and A rule Based System, Proc. IEEE 2009 Int. Conf. Inf. Engg. Comp. Sci., Wuhan, China.