Time to Response Prediction for Following up on Account Receivables in Healthcare Revenue Cycle Management

Rupanjali Chaudhuri(✉), Sai Phani Krishna Parsa, Divya Nagpal, and Kalaivanan R

Cerner Corporation, Bangalore 560045, India
{rupanjali.chaudhuri,saiphanikrishna.parsa,divya.nagpal, kalaivanan.r}@cerner.com
https://www.cerner.com/

Abstract. Healthcare systems find it difficult to decide on how long they should wait between submitting insurance claims and following up on the Accounts Receivables (AR) for the submitted claims. The solution to this can be the study of payer-specific historical data to understand payment trends of the submitted claims. This data may include censored data points where the responses were not received from the payers and hence is appropriate to be analyzed with Survival Analysis. To aid the follow-up process on submitted claims, we developed and validated a Time to Response (TTR) survival models using the machine learning method Random Survival Forests (RSF) for Medicare, Medicaid, and Commercial payers. TTR models aim to streamline the insurance claim follow-up process for smoother functioning of healthcare systems. These models capture the previous response time patterns based on the covariates and predict the probability of getting a response for each claim filed, from the day of claim file submission. We studied the effect of demographic, geographic, date-time related field, diagnoses, procedural and clean claim covariates on TTR. The model performance was assessed in terms of Concordance Index (C-index) and Integrated Brier score (IBS). Train and test C-index of Medicare was 0.68 and 0.67, Medicaid 0.74 and 0.74, Commercial 0.76 and 0.75 respectively. Test IBS was reported as 0.02, 0.01 and 0.06 for Medicare, Medicaid and Commercial payers respectively.

Keywords: Survival analysis · Time to response · Medicare · Revenue cycle management · Healthcare

1 Introduction

Managing Accounts Receivables becomes challenging for healthcare systems due to day by day increasing number of healthcare services being offered to patients. The payment entry initiates with healthcare systems generating a bill for the medical care rendered to patients. In cases where the patient is covered with insurance, the generated invoices are sent out to insurance companies for payment. The outstanding bills form the healthcare system's Accounts Receivables (AR).

M. Singh et al. (Eds.): ICACDS 2022, CCIS 1614, pp. 124–137, 2022.
https://doi.org/10.1007/978-3-031-12641-3_11

Following up on the AR on a timely basis holds immense importance for healthcare systems to ensure uninterrupted cash flow. Timely payments will further help hospitals in planning the recovery of the patient's owed amount. Any delay in payments from the insurance payers will hamper the entire health care system's functioning as they have to bear operational and maintenance costs [1].

Staff members of healthcare systems should keep a tab on the AR and see if the payments reach on time, for eliminating aged AR. An efficient AR management team will keep track of all insurance claims that have been filed. Further, it will execute an action plan for claims whose payment is delayed beyond a set time-limit [2–4].

However, healthcare systems usually find it difficult to understand the timing of collecting AR thereby causing inefficient deployment of collection resources and difficulty estimating cash flow. Analyzing payer-specific (insurance companies) historical claims to understand the factors affecting the payment response behavior and developing payer-specific AR follow-up prediction models can help overcome this difficulty.

Studying the payer-specific historical data for identifying the factors affecting the corresponding payer's response behavior is challenging due to non-uniformity in response patterns. There may be claims which saw rejections by payers for which response was never received. Also, there may be claims for which response was not received during the time of the study. This leads to censored data in the dataset which can be appropriately modeled using Survival Analysis.

For efficient decision-making in AR follow-up of outstanding claims, we propose "Time to Response" models based on Survival Analysis. "Time to Response" (TTR) means the duration between the claim submission time until a response was heard from the insurance companies. We have developed the TTR models with the aim to recover maximum amounts of AR in as short a time as possible, thereby reducing AR days and maximizing the hospital revenues. These models capture the factors affecting the payer-specific response time patterns and predict the probability of getting a response for each claim filed, from the day of claim file submission. TTR model results along with the bill amount can aid prioritize claims to be followed up [3].

The remainder of this paper is organized as follows. The related work is detailed in Section 2. Section 3 provides a data summary, Sects. 4 and 5 summarizes the features considered and the method followed for data preparation and model development. The results from the implementation and discussion are summarized in Section 6. Section 7 provides information on how the TTR model can be integrated within the AR workflow. Finally, concluding remarks and future scope are summarized in Section 8.

2 Related Work

2.1 Traditional Approach

This approach involves a fixed follow-up time ranging from 20–40 days after the initial claim submission based on the practice. This fixed follow-up time does not account for the fact that different payers have different time to response patterns [2].

2.2 Data-Driven Rule-Based Approach

This approach involves prioritizing claim for follow-up using descriptive data analysis on historical data set to understand the general time taken by different payers in combination with binning of claims based on charge amount. Payers can be grouped in different high-level bins based on historical TTR. However, this approach does not account for other covariates influencing TTR [2].

3 Data Summary

3.1 Key Terminologies

Institutional Claims are the claims, that are generated for services rendered by healthcare institutions such as hospitals, skilled nursing facilities, and any other outpatient and inpatient services. Institutional services include any use of equipment and supplies, laboratory and radiology services, etc., and others. The paper and electronic forms of the institutional claim are "UB-04" and "837I" respectively [5].

Professional claims are the claims, that are generated for services rendered by healthcare professionals and partners such as physicians, suppliers, and other non-institutional providers for both outpatient and inpatient services. The paper and electronic forms of professional claim are "CMS-1500" and "837P" respectively [6].

A Subscriber is a person, who signs for a healthcare insurance plan with insurance company. Subscriber should be uniquely identified to an information source by a Member Identification Number. Subscriber can also enroll his/her dependents in the insurance plan. Hence, it is not always necessary that the subscriber is the patient.

Payment floor is defined as the minimum amount of time a claim must be held by insurance companies before payment can be released to the healthcare institutions [7].

3.2 Data Source

TTR models were developed using data from "Electronic Data Interchange" (EDI) files named 837I for Institutional claims and 837P for Professional claims and the corresponding remits named 835 [8, 9]. The information available in the EDI file is summarized in Table 1. All the data files used for training our model are in x12 data (version 5010) format, which is the American National Standards Institute (ANSI) standard transaction set [10]. The use of this standard data format as an input source gives us the benefit of developing a vendor-agnostic model which can be deployed across different platforms.

3.3 Data Creation

The claims and remits data provided in x12 format was converted into XML format. Further, a sequence of data transformation steps were performed to convert the data in XML format to the pandas data frame. This pandas data frame was the input for the next steps towards model development.

The data is created in such a way that the service line items within a claim are in different rows and the claim-related information is repeated for all the line items. Line

Table 1. EDI file summary

Segment	Description
Transaction header	Contains information about the type of claim (whether it is a professional, institutional, or a dental claim), transaction set control numbers,etc
Billing provider detail	Contains information about provider's name, demography, identification, and currency details
Subscriber detail	Contains information about the subscriber's name, identification, gender, demography, etc
Patient detail	Contains information on the patient's name,gender, demography, and relationship with the subscriber
Claim information loop	Contains information about claim submitter's identifier, claim level charge amount, facility code, claim frequency code, date fields, etc
Service line loop	Contains line level information like line-item control number, Current Procedural Terminology (CPT) codes, date of service, service facility, etc

item, here refers to the entry made for each service provided to the patient insured under a claim. This gives us the flexibility to create claim-level aggregated features. Further, as our, goal is to predict TTR for a given claim, the line level information was converted to claim level information after creating the aggregate features based on the line items data.

3.4 Ground Truth Creation

For model development, the ground truth required is the "time to response" for a claim submitted. Below are the fields considered from claim-remit pair for ground truth creation:

1. BHT04 Data Element in both 837I and 837P files contain the Transaction Set Creation Date - "the date that the original submitter created the claim file from their business application system."
2. BPR16 Data Element in 835 file contain the Cheque Issue or EFT Effective Date (Remit File) - "the date the originating company intends for the transaction to be settled (i.e., Payment Effective Date)."
3. The difference in days between "BPR16" and "BHT04" was taken to derive the label of "time to response".

3.5 Dataset Summary

The dataset prepared for model development after mapping claims to the corresponding remits primarily constituted the claims submitted to Medicare, Medicaid, and Commercial payers. Table 2 below lists down the dataset summary for each of the Health Care Plan.

Table 2. Description of dataset statistics.

Health care plan	Samples count	Event ratio	Time to response		
			Minimum	*Median*	*Maximum*
Medicare	24037	98%	1	14	26
Medicaid	14239	99%	3	5	56
Commercial	2781	78%	1	9	77

4 Features Considered

A wide range of features were considered for inclusion in this model. Broadly speaking, they fit into the categories as mentioned in Figure 1.

1. Patient particular details such as gender, age, weight, pregnancy indicator, death date, length of stay in hospital etc., are considered.
2. Geographic features such as patient's city and state, insurance company's city & state, billing provider's city & state are considered.
3. Diagnoses and Procedural features capture the Diagnoses and Procedural information related to the patient. The diagnoses and procedural codes were converted into their respective hierarchical groups using Clinical Classifications Software Refined (CCSR) to reduce cardinality of these features.
4. Date Time Related Features such as hour of the day, weekday, day of the month etc., are derived from the claim creation date and claim creation time data on the claim files. The underlying idea beneath creating these features is to capture any seasonal patterns in the way claim files are adjudicated by the insurance companies. Further, "count of holidays" feature was created specifically for Medicare payers as they follow a 13-day payment floor. However, in the event of a holiday on 14^{th} day, the ideal response date gets shifted based on the number of holidays. Hence, this feature was created to account for this pattern.
5. Code Value Features - There are a rich variety of code value segments such as Diagnosis Related Group (DRG), Condition Information Codes and etc. available in the claim files. All these code values were included into the features set.
6. Line Item Features are created by aggregating the information across all the line items for each claim file. Features such as count of line items per claim and sum, mean, and median values of line item quantities per each claim were created. All

the code values data such as Revenue codes, HCPCS codes and NDC codes were translated as Boolean features at a claim level, to capture whether a line item with a particular code value is available for a given claim or not.

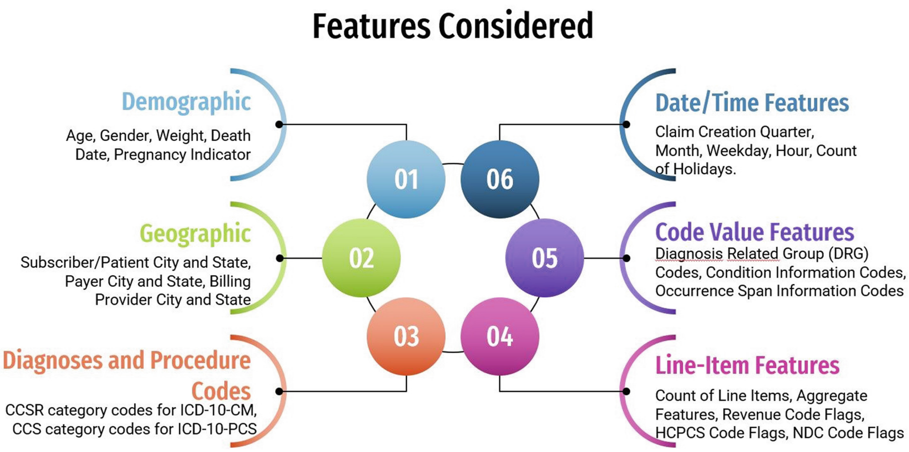

Fig. 1. Features considered

5 Method

5.1 Survival Analysis

Survival analysis is a branch of statistics, that is focused primarily on the time until the occurrence of an event. Survival analysis techniques have the capability to handle Censoring (censored data). Censoring is a form of missing data problem in which some of subjects in experiment do not have the time to event value [11, 12]. Censoring comes in different variants, while the most common being "Right Censoring". Right Censoring is said to occur when either the subjects choose not to continue with the study due to several reasons such as dropout, lost to follow-up, unavailable information, or the study ends before the event of interest occurs before the trial end date [13, 14].

5.2 Survival Analysis Key Terminologies

Survival Function: Let T - a possibly infinite, but always non-negative random lifetime taken from the population under study. The survival function $S(t)$ defines the probability of surviving past time t, as represented in Eq. (1).

$$S(t) = Pr(T > t) \tag{1}$$

1. $0 \leq S(t) \leq 1$

2. $S(t)$ is a non-increasing function of t.

Cumulative Density Function: Cumulative Density Function (CDF) captures the probability of an event happening before time t. Given the survival probabilities $S(t)$, CDF can obtained as per the Eq. (2).

$$CDF = 1 - S(t) \tag{2}$$

Hazard Function: Hazard Function ($\lambda(t)$) captures the probability of a death event occurring occurring at time t, given that the death event has not occurred yet. Hazard function is computed as seen in the Eq. (3).

$$\lambda(t) = Pr(T = t | T \geq t) \tag{3}$$

Cumulative Hazard Function Cumulative Hazard Function (CHF) denoted by $\Lambda(t)$ captures the accumulated hazard for an event up to time t. CHF can be computed as per the Eq. (4).

$$\Lambda(t) = \int_0^t \lambda(\mathrm{i})\, di \tag{4}$$

Further information on survival terminologies and how each of the functions are inter-related can be found at Lifelines python package documentation [15].

5.3 Random Survival Forest

Random Survival Forest (RSF) algorithm is a survival tree based ensemble method for analyzing right-censored survival data [16]. RSF can handle both categorical and numerical features for detecting feature interaction. RSF also supports a large number of predictors and works with small sample size. Also, as demonstrated by Breiman et al. [17] forests based algorithms are robust to noise variables.

Below are the steps involved in the RSF algorithm [16, 18]:

1. Randomly select B bootstrap samples from the original dataset. On an average 37% of the data from the original dataset is excluded in each bootstrap sample, which is called out-of-bag data (OOB data).
2. Grow B survival trees, one from each bootstrap sample. While growing the tree, at each node, p candidate variables (covariates) are selected at random to identify the best candidate variable that maximizes the survival difference between daughter nodes.
3. Grow each tree to full size under the constraint that a terminal node should have no less than a specified number of unique response time values.
4. Compute an ensemble cumulative hazard by taking the average of cumulative hazard information from the B trees.
5. Using OOB data, calculate prediction error for the ensemble cumulative hazard function (CHF).

5.4 Feature Selection

When dealing with a huge set of variables, it is easy to pick up noisy variables that do not generalize well. Our goal is to extract a parsimonious set of features which captures the most meaningful information about "time to response". Redundant, rare, and irrelevant features were dropped by using missingness, variance and correlation analysis conditions. Prior to running RSF feature selection, stratified 10-fold cross validation was used to split the data into train and test sets. For each fold in the 10-fold groups, a "noise" feature was added from Poisson distribution and Gaussian distribution. Later feature importance scores were computed using permutation-combination approach [16, 19, 20]. Using RSF model, train and test set Concordance Index (C-Index) was further computed. Across 10-folds, aggregated feature importance, the number of times a feature was ranked above the "noise" and the number of times a feature was given a positive score were tracked to subset the important features. Further, 10-fold grid search was performed to identify the best RSF parameters using the finalized features. The final set of features and the best parameters selected for RSF model training is summarized in Table 3.

Table 3. RSF model parameters

Models	No. of features	No. of trees	Max depth of tree	Min samples at internal nodes	Min samples at leaf nodes
Medicare	6	200	10	20	2
Medicaid	4	200	5	20	2
Commercial	11	200	15	40	10

5.5 TTR Predictions

The central elements of the RSF algorithm are growing a survival tree and constructing the ensemble CHF [16].

For making TTR predictions, each claim file's features are dropped down each tree in the forest until it reaches a terminal node. Terminal nodes are the most extreme nodes in a saturated tree [16]. Let τ denote all the terminal nodes in a tree. Let $(T_{1,h}, \delta_{1,h}), \ldots,$ $(T_{n(h),\,h}, \delta_{n(h),h})$ be pairs of the survival times (response times) and censoring information (was there a response for a claim or not) for individuals (cases) in a terminal node h δ. A claim file is said to be right-censored at time $T_{i,h}$ if $\delta_{i,h} = 0$, meaning that there is no response for the claim file by the end of experiment time; otherwise, if $\delta_{i,h} = 1$, meaning that a response was heard for the claim file at $T_{i,h}$. Let $t_{1,h} < t_{2,h} < < t_{N(h),h}$ be the $N(h)$ distinct response times. Define $d_{l,h}$ and $Y_{l,h}$ to be the number of deaths and individuals at risk at time $t_{l,h}$.

$$\hat{H}_h(t) = \sum_{t_{l,h} \leq t} \frac{d_{l,h}}{Y_{l,h}} \tag{5}$$

The CHF estimate of h denoted as $\hat{H}_h(t)$ is obtained using the the Nelson-Aalen [21] estimator as given in equation (5). All claim files within h have the same CHF. Each claim file i has a d -dimensional covariate x_i. Let $H(t|x_i)$ be the CHF for i. To determine this value, drop x_i down the tree. Because of the binary nature of a survival tree, x_i will fall into a unique terminal node $h \in \tau$.

$$H(t|x_i) = \hat{H}_h(t), \ \ if\ x_i \in h \tag{6}$$

Equation (6) defines the CHF for all cases and defines the CHF for the tree and is denoted as $H(t|x_i)$. This CHF is derived from a single tree and in order to compute an ensemble CHF, all the CHF values across B survival trees are averaged.

Since the problem statement discussed in this paper is aimed at interpreting the time to death as time until the occurrence of an event, for purpose of convenient interpretation the survival probabilities obtained for each claim file are translated into cumulative distribution function.

$$F(t) = P(T \leq t) = 1 - S(t) \tag{7}$$

Cumulative distribution values (denoted as $F\ (t)$) are obtained from survival values using equation (7), where S(t) is the survival function. The reason for CDF implementation instead of survival function is primarily because this problem statement demands probability of TTR event before time t for AR follow-up. The provider would be interested in following up with the payer if they are yet to receive response to the submitted claim. This time t of interest would be the time where predicted probability of response is high by the survival model indicating the provider should have ideally got a response for a claim submitted.

6 Results and Discussion

Payer-wise three RSF models were trained for Medicare, Medicaid, and Commercial payers with the final set of features and the best parameters using scikit-survival package [22]. The performance of the developed models was assessed using two measures- the concordance index (C-index) and the integrated brier score.

The C-index represents the global assessment of the survival model's discrimination power. In this use case C-index indicates model's ability to correctly provide a reliable ranking of the survival times based on the time to responses of claims. In general, when the C-index is close to 1, the model has an almost perfect discriminatory power; but if it is close to 0.5, it has no ability to discriminate between low and high time to response claims.

Brier score is a measure of the model's accuracy and evaluates the accuracy of a predicted survival function at a given time t. Integrated Brier Score (IBS) is an overall measure for the prediction performance of the model at all times. The lower the Brier score is the better is a model's performance.

6.1 Performance Evaluation

Stratified grouped 10-fold cross validation grouped by a patient was used to assess the model performance. This approach helps get a fair assessment of model performance ensuring minimal longitudinal correlation between train and test splits. The performance of the models is summarized in Table 4.

Table 4. Model performance

Model	Train C-index	Test C-index	OOB C-index	Train IBS	Test IBS
Medicare	0.68	0.67	0.67	0.05	0.02
Medicaid	0.74	0.74	0.74	0.01	0.01
Commercial	0.76	0.75	0.75	0.06	0.06

We further studied the discriminating capability of the selected features to model the time to response for the submitted claims from each of the payers. The feature importance of selected features is depicted in Fig. 2.

We found conclusive evidence of seasonal pattern driven time to response for Medicare and Medicaid payers. Medicare has a payment floor of 13 days, wherein most claims get cleared on the 14^{th} day from the day it was filed [7]. However, if there are any public holidays in between the processing gets delayed by the number of holidays. We observed count of holidays as the feature with highest importance. Fig. 3 (a) captures the variance in predictions depending on the value of "count of holidays" feature. Further, Medicare claims TTR is also affected by other date-time related features and claim charge amount.

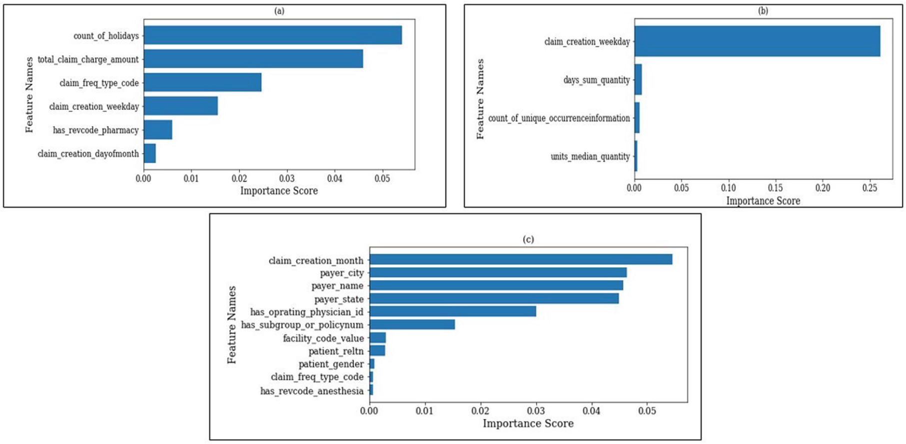

Fig. 2. Final model features. (a): Medicare feature importance; (b): Medicaid feature importance; (c) Commercial feature importance

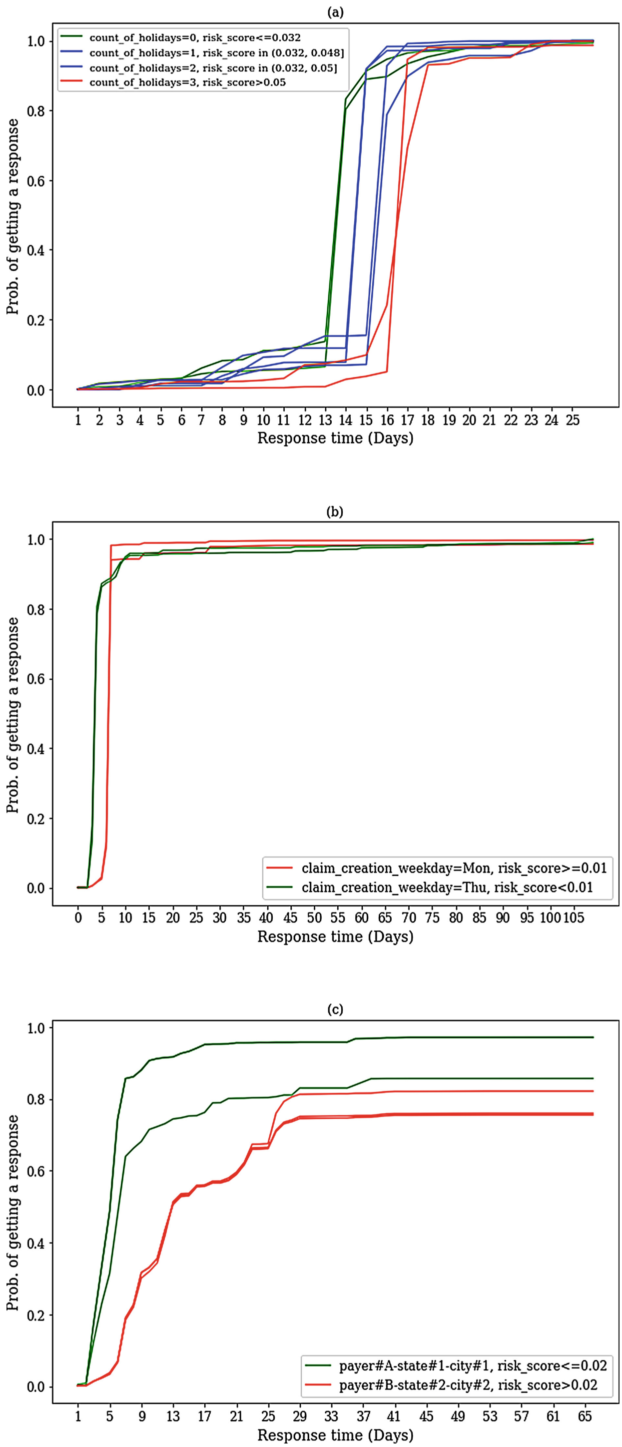

Fig. 3. Model predictions. (a): Medicare model predictions; (b): Medicaid model predictions; (c) Commercial model predictions

Medicaid claims showed a weekly seasonality pattern with claims having bulk clearance on Mondays. This seasonality may differ in different healthcare system's data based on location. This is primarily because Medicaid being a state- based government payer, the time to response and claim clearance patterns may be driven by state policies. Fig. 3 (b) captures the Monday clearance pattern, where in we can see that if the claim files were created on weekday's they have a quicker response time. However, Saturday's and Sunday's are exception as they will be cleared on the second Monday from claim creation date instead of the first Monday(as observed for other claim creation weekdays).

For Commercial payers geographic and demographic features were important as commercial model has several sub-payers showing mixed payer time to response behaviour. Fig. 3 (c) captures the variance in prediction plots given different values for payer name, payer state and payer city.

Please note that for the purpose of interpretation, inverse of actual risk score values are projected as risk scores in Fig. 3. As per survival analysis risk score will be high for subjects with a shorter time-to-event value. To ensure that a claim with higher response time value is appropriately aligned instead with high risk score values aiding to highlight claims at risk to follow up, inverse of risk score values are projected as risk scores.

7 TTR Model Integration in AR Workflow

Our proposed AR follow-up process includes payer specific TTR models with their corresponding features and is depicted in Fig. 4. Here is an outline of AR follow-up process with TTR models:

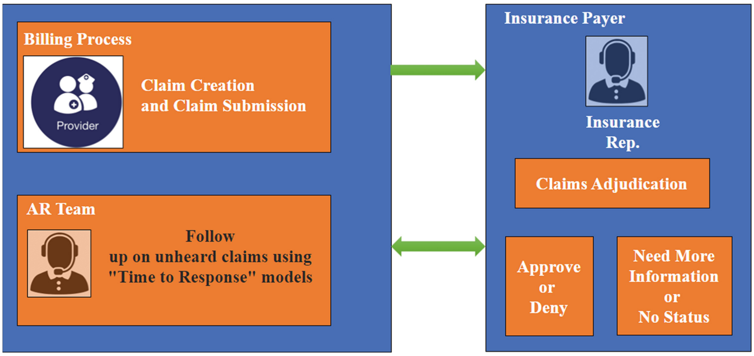

Fig. 4. Proposed RCM workflow

1. Billing Process: Healthcare providers or providers will create the claim file and submit it to the insurance companies for clearance.
2. Insurance Payer: Insurance companies can either choose to settle the claim provided the claim is clean and provides all the required information for settlement. Claims can also be denied, or insurance companies reach out to the healthcare providers in case they need more information.

3. AR Team: AR teams will leverage the payer specific TTR models directly or TTR models output in combination with the "total claim charge amount" to prioritize follow-up on outstanding claims for timely receiving the outstanding amounts.

8 Conclusion and Future Scope

We have developed and validated TTR models using the machine learning method RSF for Medicare, Medicaid, and Commercial payers based on data collected from a Healthcare system. RSF algorithm being simple and easy to make predictions with, is well suited for integration to production environment. However, in future, we would like to perform comparison study between RSF and other algorithms to conclude on research on better prediction performance for healthcare insurance claims.

References

1. Kamenetzky, S.A.: Cash flow management. Arch. Ophthalmol. **111**(6), 757–60 https://doi.org/10.1001/archopht.1993.01090060045021, PMID:8512475
2. Manley, R., Satiani, B.: Revenue cycle management. J. Vasc. Surg. **50**(5), 1232 (2009). ISSN:0741–5214 PMID: 19782521 https://doi.org/10.1016/j.jvs.2009.07.065
3. McDaniel, J.W., Baum, N.: Putting the receive in accounts receivable. J. Med. Pract. Manag. **22**(1), 31–33 (2006). PMID: 16986638
4. Beaulieu-volk, D.: Accounts receivable. Strategies for better management. From eligibility verification to patient engagement, managing accounts receivable is a vital process to maintaining financial health. Med. Econ. **92**(11), 38 42–4 (2015). PMID:26298962
5. Center for Medicare and Medicaid Services(CMS) Medicare Billing Form CMS- 1450 and the 837 Institutional Medicare Learning Network Booklet. https://www.cms.gov/Outreach-andEducation/Medicare-Learning-Network-MLN/MLNProducts/Downloads/837I-FormCMS-1450-ICN006926.pdf
6. Center for Medicare and Medicaid Services (CMS) Medicare Billing Form CMS- and the 837 Professional Medicare Learning Network Booklet: https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/837P-CMS-1500.pdf
7. Center for Medicare and Medicaid Services(CMS) CMS Manual System Medicare Claims Processing. https://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/Downloads/R114CP.pdf
8. Romo, E.: Adapting to the ANSI 835 transaction set. Healthc. Financ. Manage **47**(1), 54–56 (1993). PMID:10145738
9. Moynihan, J.J.: Improving the claims process with EDI. Healthc. Financ. Manage **47**(1), 48–49 (1993). PMID: 10145737
10. Clement, J.M., Marc, O.J., Paul, D., Blaine, T., Jeffrey, G.S.: What is done, what is needed and what is realistic to expect from medical informatics standards. Int. J. Med. Inform. 48(1–3), 5-12. ISSN 1386-5056 (1998). https://doi.org/10.1016/S1386-5056(97)00102-0
11. Prinja, S., Gupta, N., Verma, R.: Censoring in clinical trials: review of survival analysis techniques. Indian J. Comm. Med. Official Publ. Indian Assoc. Prev. Soc. Med. **35**(2), 217–221 (2010). https://doi.org/10.4103/0970-0218.66859
12. Wikipedia Survival Analysis. https://en.wikipedia.org/wiki/Survival_analysis
13. Rich, J.T., et al.: A practical guide to understanding Kaplan-Meier curves. Otolaryngol Head Neck Surg. **143**(3), 331–336 (2010). https://doi.org/10.1016/j.otohns.2010.05.007

14. Enrique, B., Laura, B.: Effect of right censoring bias on survival analysis. arXiv2012.08649 (2020)
15. Lifelines Documentation. https://lifelines.readthedocs.io/en/latest/
16. Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S.: Random survival forests. Ann. Appl. Stat. **2**, 841–860 (2008). https://doi.org/10.1214/08-AOAS169
17. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
18. Janika, S.: Modelling Late Invoice Payment Times Using Survival Analysis and Random Forests Techniques. Financial and Actuarial Mathematics Curriculum Master's thesis, Institute of Mathematics and Statistics, University of Tartu
19. Eli5 Package Document. https://eli5.readthedocs.io/en/latest/blackbox/ permutation_importance.html
20. Lars, B., Gilles, L., et al.: API design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning, pp. 108–122 (2013)
21. Ørnulf, B.: Nelson-aalen estimator. Encycl. Biostatis. (2005). https://doi.org/10.1002/0470011815.b2a11054
22. Pölsterl, S.: scikit-survival: a library for time-to-event analysis built on top of scikit-learn. J. Mach. Learn. Res. **21**(212), 1–6 (2020)