How far is too far? Identifying suspicious travel patterns in healthcare claims using machine learning

Kai Wang
Deloitte & Touche LLP, United States
kawang@deloitte.com

Daniel Lasaga
Deloitte& Touche LLP, United States
dlasaga@deloitte.com

John Helms
Deloitte& Touche LLP, United States
johhelms@deloitte.com

Edward Bowen
Deloitte& Touche LLP, United States
edbowen@deloitte.com

Sanmitra Bhattacharya
Deloitte& Touche LLP, United States
sanmbhattacharya@deloitte.com

***Abstract*—Fraud in healthcare services and claims poses a significant threat to healthcare expenditure, accessibility to health services, and quality of care of members. One important type of member-provider collusion is where members travel unreasonable distances seeking healthcare services. Such activities could be indicators of "pill mills", doctor shopping, or referral kickback schemes. Previous research on the identification of suspicious travel distances have focused mostly on the billed amount and considered select diagnosis conditions and travel distances at zip code or county levels. Compared to these studies, our proposed framework focuses on claims across various diagnoses and takes into account population densities of members' zip codes, and provider densities for various specialties, among other features, which are critical to the prediction of travel distances. We experiment with two approaches – i) a regression model paired with a statistical anomalous distance detector, and ii) a neural network-based model paired with a likelihood estimator for anomalous distance detection. The evaluation of these models on a manually annotated dataset shows that the second approach outperforms the first one in identifying anomalous travel distances.**

***Index Terms*—Geographic Information Systems, GIS, Machine Learning, Neural Networks, Anomaly Detection**

I. Introduction

Medicare and Medicaid are two US Federal health insurance programs for those who are elderly or disabled and those who have very low income, respectively. According to the Centers for Medicare & Medicaid Services (CMS)[1], as of June 2021, the total number of enrollees in Medicare was over 63 million while for Medicaid it was over 75 million. The National Health Expenditure was over $4 trillion in 2020 which accounted for nearly 20% of the US Gross Domestic Product[2] and is estimated to reach over $6 trillion by 2028. The National Health Care Anti-Fraud Association estimates that 3% to 10% of total healthcare expenditures are lost to healthcare fraud[3] which conservatively amounts to over $100 billion. Healthcare fraud manifests in various forms [1]–[3], ranging from identity theft and using services under another member's name, to providers billing for unadministered services and pharmacists running "pill mills". Such fraud has negative consequences not just on healthcare expenditure but also on accessibility and reach of health services, and quality of care of members.

In this paper we focus on an important subset of member-provider collusions that result in members seeking healthcare providers at unreasonable distances from their homes. Examples of fraud schemes that tend to exhibit distance anomalies are "pill mills", doctor shopping, and referral kickback schemes for Durable Medical Equipments (DME). Referral kickback schemes are when doctors receive incentives such as money or gifts in exchange for remote patient referrals to receive DMEs (e.g., wheelchairs, prosthetic inserts) [4]. "Pill mills" are when healthcare providers and their staff are willing to accommodate or encourage members' drug use by prescribing the drugs (typically opioids) to help treat exaggerated or nonexistent symptoms [5]. Doctor shopping is when individuals travel from doctor to doctor often posing as out-of-town visitors until they find select drug prescriptions, duplicate prescriptions, or preference for treatments that are not within normal treatment guidelines [6]. "Pill mills" and doctor shopping contribute significantly to drug diversion, costing an estimated $72.5 billion annually[4].

All three of the above schemes tend to favor longer distances between members and providers and we'd expect them to show up as distance anomalies with members traveling exceedingly long distances to see providers. However, there are many legitimate reasons why members may need to travel long distances to see a medical provider. For example, members may need to see a specialist for procedures that are rarer and not commonly available, or members may live in rural or underserved areas and may need to travel longer distances even for more routine visits. A member may also need emergency medical care while far from home. Some members may assume higher quality of care from certain providers and hence travel further than usual. Many of these factors could indeed play a critical role in distances traveled by beneficiaries; but similar to [7], [8], we assume that these cases are either unobservable (e.g., member's assumption of quality of care)

[1]https://www.cms.gov/newsroom/news-alert/cms-releases-latest-enrollment-figures-medicare-medicaid-and-childrens-health-insurance-program-chip

[2]https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet

[3]https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud/

[4]https://insurancefraud.org/wp-content/uploads/drugDiversion.pdf

1946-0759/23/$31.00 ©2023 IEEE
DOI 10.1109/ICMLA58977.2023.00090

or extremely rare (e.g., non-emergency health services during travels). The methodologies we consider ideally condition what we expect to be an unreasonable distance based on the information from three primary sources – members, providers, and claims. Our overall objective is not to fully automate travel distance-based fraud identification but to flag high risk claims to fraud investigators for further review and determination of actual fraud.

To this end, we discuss two approaches for the identification of suspicious travel distances in healthcare claims. In the first approach, we build regression models to predict travel distances, which are then used in a statistical model over residuals (i.e. difference between actual and predicted distances) to identify suspicious travel distances. In the second approach, we use Mixture Density Network (MDN) [9], a neural network-based approach, to output parameters of a probability density function (PDF) which approximates the underlying data distribution. The PDF is then used to compute the likelihood of a distance being an outlier. For both approaches we consider various claim-based features, such as diagnosis codes and provider specialty; census tract data, such as population density and average household income; and other data-driven derived features, such as density of providers with certain specialty in a certain geolocation. Using manually labeled data for model evaluation, we demonstrate the feasibility of our approaches in identifying suspicious travel distances.

We make several novel contributions in this paper. First, we experiment with a novel modeling approach which uses the MDN architecture with a Gamma distribution to approximate the underlying data distribution, which is in turn used to identify outlier travel distances. Second, while most comparable studies primarily focus on features such as payment amount and distance traveled, we employ a more extensive set of features that includes member, provider, claim, and census-level attributes. Lastly, in contrast to prior studies, our method is not limited to particular diagnostic conditions.

The rest of the paper is structured as follows: Section II covers related research, and Section III delves into the problem definition, data, and methods. Our findings are presented in Section IV, followed by discussions and conclusions in Section V.

II. Related Research

Musal [7] identified the expected quality of care from a provider, and the distance between a provider and beneficiary as two primary attributes a beneficiary considers in choosing of a provider. Since the expected quality of care from the perspective of a beneficiary cannot be observed, the paper proposed a method for identifying potentially fraudulent providers where members may be traveling impractical distances to receive healthcare services. The authors focused on beneficiaries residing in a Metropolitan Statistical Areas (MSA) and flagged providers when more than 10% of their beneficiaries traveled over a 99 percentile distance threshold, or their earnings are skewed toward beneficiaries who traveled over certain predefined distance brackets. In contrast, our approach is more general as it considers not only MSAs but also rural areas, and distances at a higher resolution between beneficiaries and providers compared to zip codes in [7].

Using CMS' anonymized claims data from 2010 Liu and Vasarhelyi [8] built a model to predict geo-location-based fraud. Since the exact address information for beneficiaries and providers were not available, latitudes and longitudes of county and state centroids were used to measure beneficiary-provider Euclidean distance. With a focus on select medical conditions (such as pneumonia, rehabilitation services, and septicemia) and payment amounts, clusters were created to identify claims with reasonable vs unreasonable travel distances and payment amounts. The general principle of defining outlier travel distances is similar to [7], and suffers from the same drawbacks stated above.

The Xerox Program Integrity Validator [10] featured various graph analysis techniques to identify fraud, waste and abuse (FWA) in real-world healthcare datasets. The authors leveraged an ego-net graph approach to identify anomalous geo-spatial relationships. Specifically, they computed the geographical distance between physician and pharmacy pairs, and derived an empirical cumulative distribution function (*cdf*). Then they applied a density-based clustering algorithm (DBSCAN) to the *cdf* to define a baseline. *Cdf*s that are similar to each other are clustered together while those that deviate significantly from the norm are flagged as anomalous. This approach of only considering the *cdf*s of the distances, and clustering based on *cdf*s assumes that deviations from the norm of provider proximities is the only factor that can be used to identify anomalies. This study did not consider factors such as population density and provider density of a particular specialty, which may play a critical role in distance traveled. Also, this study focused only on distances between physician-pharmacy pairs whereas we consider distances between beneficiary-provider pairs at a claim level.

Rosenblum *et al.* [11] examined commuting patterns of over twenty thousand methadone members enrolled in 84 opioid treatment programs. While the majority of members traveled <10 miles, 8% of members traveled across state borders. Using a multivariate model, the authors found that the factors most significantly associated with travel distances were whether they resided in the Southeast or Midwest, low urbanicity, certain demographic factors, prescription opioid abuse, and heroin usage patterns. There are various assumptions made in the modeling approach adopted in this study. The multivariate mixed-effects model assumes linearity between the features and response variable as well as considers various assumptions on the random effects which may not apply to a large heterogeneous real-world dataset like ours.

In a recent paper [12] Chen *et al.* focused on the complex interactivity of various members, provider, and incident features that goes into prediction of travel distances of members to access healthcare. In contrast to conventional statistical and econometrics approaches that may not consider confounding variables and only apply to balanced label sampling approaches, the authors proposed a convolutional neural network-based approach to predict travel distances. Specifically, they focused on members with respiratory infections, flu, or needing emergency services, and predicted whether they will travel for <5 km, 5-10 km, 10-15 km or >15 km. Compared to [12], our approach has a few advantages. First, this study categorizes distances into four levels, suggesting a classification approach for predicting travel distances. Although this might be suitable for addressing healthcare access, such a broad categorization could result imprecise predictions in a healthcare FWA context. Therefore, we concentrate on more granular modeling approaches, such as regression models. Second, we consider the full sample population of a client state Medicare/Medicaid members instead of focusing on particular conditions. Third, while the authors propose that bucketing the distance traveled leads to a more generalized solution, some obvious drawbacks are sizes of each bucket, number of buckets, etc. And finally, the provider-member distances are only approximations since they were calculated based on centroids of districts of providers and members, and suffers from similar drawbacks as the aforementioned studies.

III. Problem Definition, Data and Methods

A. Problem Definition

We propose two approaches to identify anomalous travel distances. For the first approach, we build a regression model to predict travel distances given claims data, census data, and data-derived features, such as provider density. The difference between actual and predicted values are used to compute residuals to generate a truncated normal *cdf*. The *cdf* is then used to calculate whether certain distances are normal or anomalous. For the second approach, we use a MDN to approximate the underlying data distribution and compute parameters of a PDF, which is then used to identify anomalous travel distances.

While unsupervised machine learning methods such as clustering or autoencoders are often applied for anomaly detection [13], [14], there are various challenges in applying them to very high dimensional data like ours. Clustering in high dimensional data often leads to objects being nearly equidistant from each other, completely masking individual clusters [15]. While there are alternatives such as feature selection, dimensionality reduction and subspace clustering [16], they often involve either manual parameter selections that are challenging, or force sub-optimal model-driven parameter selection [17]. Other methods such as autoencoders, which learn some low-dimensional representation space from which input data can be well reconstructed, are biased by presence of outliers and infrequent regularities in the training data [14]. In contrast, our proposed probabilistic methods offer several advantages, including the ability to quantify uncertainly in predictions, robustness to outliers, and the capability for end-to-end training.

B. Dataset for Model Training

1) Claims: In the state Medicare/Medicaid dataset we use, each claim typically contains information about members (member id, address, demographics, etc.); providers (National Provider Identifier (NPI) id[5], practice address, specialty, etc.); details of diagnosis (in the form of International Classification of Diseases (ICD-10-CM) codes[6]); procedures (in the form of Current Procedural Terminology (CPT) codes[7]); qualifiers/modifiers; billed amount; etc. Due to sparsity of provider practice addresses in our dataset we use the NPI registry to identify missing addresses. Since the NPI registry is frequently updated and provider practice addresses may change over time, to ensure data quality we limit our analysis to claims since January, 2021. We also filter out claims that are labeled as void or invalid, has $0 billed, provider is *locum tenens*[8], or has invalid addresses of members or providers (e.g., zip codes: 00000/9999, PO Box addresses).

Besides these claim-level features, we also derive an additional feature to account for members' accessibility of providers of various specialties. Since we have access to the full state Medicare/Medicaid data, we can estimate provider density within N square miles of a member. As we discuss later, the median travel distance of members is around 30 miles so we calculate the provider density of specialties within 30 square miles of a member's address.

The complete dataset comprises of over 40 million claims with 48,570 distinct providers and 1.7 million distinct members. Applying the filters described above resulted in a dataset of 379,009 claims with 29,790 providers and 268,980 members. The top 5 most frequent ICD-10-CM categories in our dataset are: R69: Illness, unspecified, F41: Other anxiety disorders, F11: Opioid related disorders, F33: Major depressive disorder, recurrent, and F43: Reaction to severe stress, and adjustment disorders. In this dataset, provider specialty is coded using a 10-character code that designates type and classification. The top 5 most frequent physician taxonomy codes are: 282N00000X: Hospitals/ General Acute Care Hospital, 2084P0800X: Allopathic & Osteopathic Physicians/ Psychiatry, 363LF0000X: Physician Assistants & Advanced Practice Nursing Providers/ Nurse Practitioner, Family, 207Q00000X: Allopathic & Osteopathic Physicians/ Family Medicine, and 2085R0202X: Allopathic & Osteopathic Physicians/ Radiology, Diagnostic Radiology.

2) Census: While claims data provides us with details related to members' diagnoses and treatments, their addresses and properties related to their addresses may play an important role in their travel distances. As discussed before, members in rural areas may tend to travel longer for the same diagnosis compared to someone in an urban area. Members from wealthier neighborhoods may travel for longer distances to seek better quality of care compared to those from poorer

[5]https://npiregistry.cms.hhs.gov/

[6]https://www.cdc.gov/nchs/icd/icd-10-cm.htm

[7]https://www.ama-assn.org/amaone/cpt-current-procedural-terminology

[8]A provider working temporarily from another practice which could be in a different city or state from where they primarily work.

neighborhoods [18]. We use the 2020 US Census data[9] to get zip-code level information on overall population, population density, land area, number of occupied housing units, average home value, and average household income.

In our analysis of the census data, we find that the total number of unique zip-codes in our dataset is 1,450. The minimum population in a zip-code is 64 while the maximum is 105,549. The median average household income is $47,080, while the minimum is $2,499 and maximum is $160,631.

As model inputs, all numerical features (e.g., provider density, population density, etc.) are represented as they are, except for long-tailed distributions (e.g., average household income), which are log-transformed. Categorical features, such as diagnoses codes and provider specialty, are one-hot or multi-hot encoded.

3) Distance Traveled: Given the sensitive nature of member addresses, similar to [19], we derive the distance traveled between members and providers in multiple steps using offline maps data. First, we convert the addresses of members and providers to approximate latitudes and longitudes using Open Street Map (OSM) [20]. We then use the A* graph traversal algorithm [21] on OSM data to identify the optimal route – a set of nodes with corresponding latitude and longitude in the OSM graph. For each pair of OSM nodes, we calculate the distance using the Haversine formula [22] which are summed up to get the total distance between two addresses. The Haversine formula is used to calculate the distance between two points on the Earth's surface given their latitude and longitude.

Analysis of the travel distances between members and providers show that the minimum distance is 0 miles, while the median travel distance is around 30 miles, and the maximum distance is 7436 miles. Around 15% of claims have travel distances exceeding 100 miles.

C. Modeling Approaches

Since the dataset was not labeled with anomalous travel distances, we experimented with 2 semi-supervised approaches.

1) Approach 1: Travel Distance Prediction and Statistical Modeling for Anomalous Distance Detection: In the first approach, using features from claims data – which includes information about members, providers, and individual claims, and census data – which includes information on various geo-spatial and socio-economic features, we build regression models to predict the log transformed value of distance traveled. We then use statistical models to identify anomalous travel distances. An overview of our approach is shown in Fig 1 with details in the following sections.

a) Travel Distance Prediction: We experimented with various tree- and neural network-based regression models to predict travel distances. Below we present the hyper-parameters used in these models which were tuned using random search [23] of the hyper-parameter space.

Random Forest A Random Forest regression model uses an ensemble of decision trees to predict travel distances. We use

[9]https://data.census.gov/

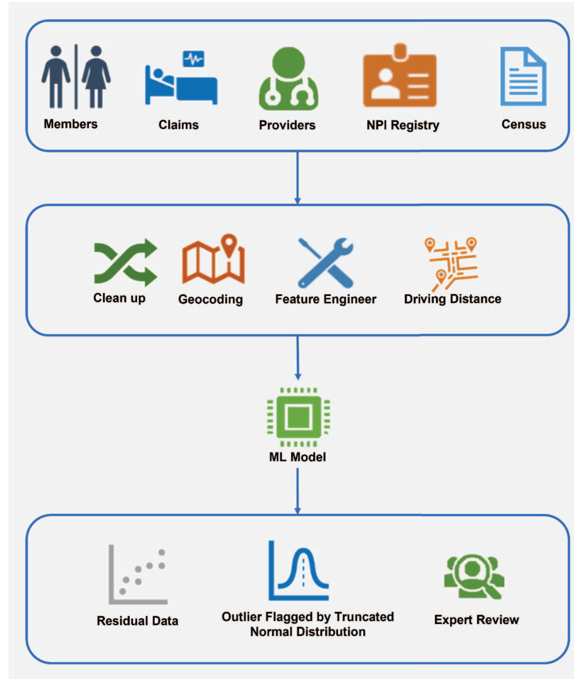

Fig. 1. Overview of Anomalous Travel Distance Identification Framework

this as a baseline with the number of trees = 100, maximum depth of trees = 8, number of features to consider for the best split = 0.1, with other parameters being default as defined in [24].

Extreme Gradient Boosting (XGBoost) We use a gradient boosting regression model, specifically XGBoost, which is an ensemble learning technique that combines the predictions of several weak learners like decision trees. It has been shown to perform well on data similar to ours. Here the optimal hyper-parameters are: number of gradient boosted trees/number of boosting rounds = 565, L1 regularization (alpha) = 0.396, L2 regularization (lambda) = 0.592, learning rate = 0.05, the number of nodes in the tree = 9, subsample ratio of columns when constructing each tree = 0.65, minimum loss reduction for tree partitioning = 3.14, minimum sum if instance weight needed in a child = 7, subsample ratio of training instances to prevent overfitting = 0.5, tree construction method = exact greedy algorithm, and other parameters are default as defined in [25].

TabNet TabNet [26] deviates from the above tree-based methods and uses sparse attention in a series of steps to model the output. The TabNet encoder comprises of a feature transformer, an attentive transformer and feature masking, while the decoder is composed of a feature transformer block in each step. Compared to other deep learning methods for instance-wise feature selection methods, TabNet uses a single deep learning architecture for feature selection and reasoning. TabNet is shown to outperform or on par with other tabular learning models on various benchmark datasets [26]. We use the default parameters values recommended in the paper for a regression task.

Neural Oblivious Decision Ensembles (NODE) NODE [27] uses neural equivalent of oblivious trees (like CatBoost trees [28]) as basis of architecture but benefits from end-to-end gradient optimization and multi-layer hierarchical representation learning. Across a large number of benchmark datasets NODE is shown to outperform other gradient boosted tree based methods [27]. In our model we set the number of oblivious decision tree layers to 8, number of trees to 128, embedding dropout of 0, with other parameters set to the recommended values from the paper.

Each of these models are trained on the dataset described in Section III(B). 80% of the data is used for training and the remaining 20% holdout data is used for testing.

b) Statistical Modeling for Anomalous Distance Detection: To identify and prioritize suspicious claims for further investigation, we quantify the risk of each claim using a statistical model on top of residuals from the regression models. For a given data point, the steps involved in this process are:

1) Calculate residuals (difference between actual and predicted distance) of the regression model.
2) Create log-transformed distribution of the residuals to reduce the skewness of the data.
3) Calculate the z-score from the log transformed residuals as below:
   $z\text{-}score = (residual - \mu)/\sigma$
   where μ and σ are the mean and standard deviation of the log-transformed residual distribution.
4) Calculate probability (p) of each claim being an outlier using a truncated normal distribution. The truncated normal distribution is derived from a normally distributed random variable by imposing upper or lower bounds (or both) on the range of the random variable [29]. In our case, we only consider one tail of the distribution. We apply an identify function which assigns 0 to all negative residual cases. The formula is:
   $p = \frac{\varphi(\mu,\delta;z_{score})-\varphi(\mu,\delta;z_{cutoff})}{\varphi(\mu,\delta;\infty)-\ \varphi(\mu,\delta;z_{cutoff})} * I_{residual}$
   where $\varphi(\mu, \delta; x)$ is the *cdf* function of the normal distribution; and

$$I_{residual} = \begin{cases} 0 & residual < 0 \\ 1 & residual \geq 0 \end{cases}$$

2) Approach 2: Anomalous Travel Distance Prediction Using MDN and PDF: In the second approach, instead of predicting the log transformed value of distance traveled, we use MDN [9] to predict the key parameters (such as mixing coefficient, μ and θ) that approximate the underlying distance data distribution. A MDN comprises of a neural network and a mixture model, where the neural network is an encoder to learn feature representations of the input, while the mixture model is a probability distribution that combines the weighted sum of multiple simpler distributions. The network is trained end-to-end while minimizing the negative log-likelihood. The output parameters of MDN define component PDFs in a mixture. Given a new data point, a trained MDN can be used to compute the likelihood of observing it under the learned mixture model and a p-value is computed. The significance level denoted by the alpha (α) is set at 0.05 and p-values of data points below this threshold are considered outliers.

TABLE I. Performance Metrics of Travel Distance Prediction Models

Models	MAE	RMSE	MAPE	R^2
Random Forest	1.04	1.47	2.68	0.33
XGBoost	**0.89**	**1.13**	**1.67**	**0.60**
TabNet	1.09	1.59	2.94	0.21
NODE	1.01	1.42	2.29	0.37

Our MDN architecture comprises of 2 layers with 256 nodes in each layer, a dropout of 0.2, kaiming initialization, hyperbolic tangent activation function, and with batch normalization. Typically, MDNs consider a mixture model that comprises of multiple Gaussian distributions. However, in our case, the underlying distribution of the training data fits a Gamma distribution i.e. the data is positively skewed where there is a long tail of high values. Therefore, we train the MDN on both Gaussian (MDN-Gaussian) and Gamma (MDN-Gamma) mixture models. Based on hyper-parameter tuning we find optimal performance when the number of components in the mixture is 1.

D. Manually Annotated Dataset for Model Evaluation

Since labels for anomalous distances were unavailable, we create a manually annotated dataset for evaluation of models outlined in Approaches 1 and 2. We use stratified sampling on model results from both of the aforementioned approaches to get 438 records. To ensure that the sampling is not biased towards a specific model's output, we conduct three rounds of independent sampling. In each round, we do stratified sampling from a different model's output. A Program Integrity[10] subject matter specialist (SMS) performed a blinded reviewed of each record to classify the travel distances as normal or anomalous based on claim, census and data-derived features (as described in Section III(B)). This manually annotated dataset contains 22.8% anomalies and 77.2% normal data, which align with our assumptions that in a real-world scenario anomalies are a minority.

IV. Experimental Results

A. Results from Approach 1: Travel Distance Prediction Model and Statistical Modeling for Anomalous Distance Detection

We present results of the regression model for travel distance prediction on the holdout test dataset. For the statistical anomalous distance detection model we perform evaluation on the manually annotated dataset. Performance metrics on the holdout test dataset for the 4 models (namely Random Forest, XGBoost, TabNet, and NODE) used for prediction of travel

[10]https://www.medicaid.gov/medicaid/program-integrity/index.html

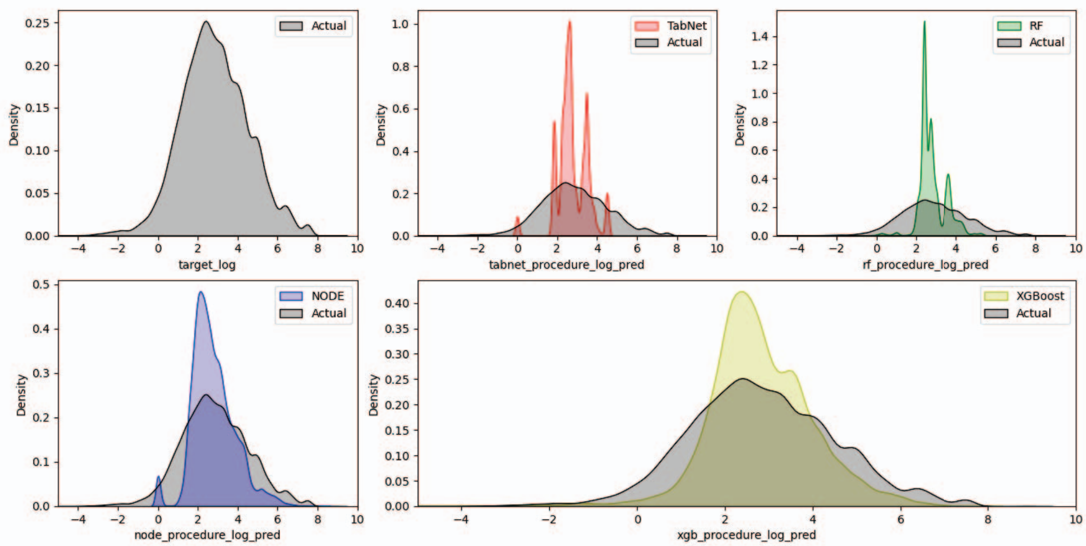

Fig. 2. Density Plots of Travel Distance Prediction Models

distances are shown in Table I. We use Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R^2) as performance metrics. For MAE, RMSE, and MAPE, lower values are better as they indicate smaller errors and better predictive performance, while for (R^2) higher values are better. The results indicate that the XGBoost model outperforms all other models in all metrics with MAE of 0.89, RMSE of 1.13, MAPE of 1.67, and R^2 of 0.6. The Random Forest regressor significantly lags behind the XGBoost regressor in performance. We speculate that that the high dimensionality of our feature space and the intricate non-linear relationships in the data are effectively captured by XGBoost, but not by the Random Forest model. Results for NODE are trailing XGBoost but better than other models across all metrics.

We also plot the density function for each model's results against the actual distance traveled (Fig 2). Similar to our observations from Table I, we find that the XGBoost model overlap the most with the original distribution compared to other models. We also notice that XGBoost predictions are more concentrated around the median compared to the actual distance distribution. This aligns with our target to identify suspicious records and lower false positive rates. Other models such as TabNet and Random Forest tend to underestimate the travel distances, which may lead to a high false positive rate.

Analysis of the feature importance of the XGBoost model shows that specialty of providers and diagnosis information in claims are key features. A majority of prominent features are associated with laboratories, medical suppliers, and nursing facilities. This is aligned with our expectations as well as previous studies on factors influencing travel distances [11]. For example, patients generally prefer going to nursing facilities that are closer to their residence. However, for certain specialty lab tests, members may have to travel for above average travel distances.

TABLE II. Performance Metrics of Approach 1: Statistical Modeling for Anomalous Distance Detection

Models	Precision	Recall	F1- score
Random Forest	0.69	0.77	0.73
XGBoost	**0.84**	0.74	0.79
TabNet	0.70	0.78	0.74
NODE	0.80	**0.81**	**0.80**

The outputs from the regression models are further analyzed to identify anomalous travel distances. While the XGBoost regression model fit the data quite well, it may also have higher risk of over-fitting on outlier distances. Therefore, instead of using thresholds on the model predictions, we use the residuals from each model described earlier, and identify anomalous travel distances, as described in Section III(C)(2). Table II shows the performance metrics of the models on the manually annotated evaluation dataset (Section III(D)) using classification metrics – precision, recall, and F1-score. We find that NODE outperforms other models in recall (0.81) and F1-score (0.80), and closely behind XGBoost in Precision.

B. Results from Approach 2: Anomalous Travel Distance Prediction Using MDN and PDF

For both MDN-Gamma and MDN-Gaussian models, we calculate p-value of each record in the manually annotated evaluation dataset using the predicted PDFs.

TABLE III. Performance Metrics of Approach 2: Anomalous Travel Distance Prediction Using MDN and PDF

Models	Precision	Recall	F1- score
MDN-Gamma	**0.80**	**0.82**	**0.81**
MDN-Gaussian	0.78	0.53	0.63

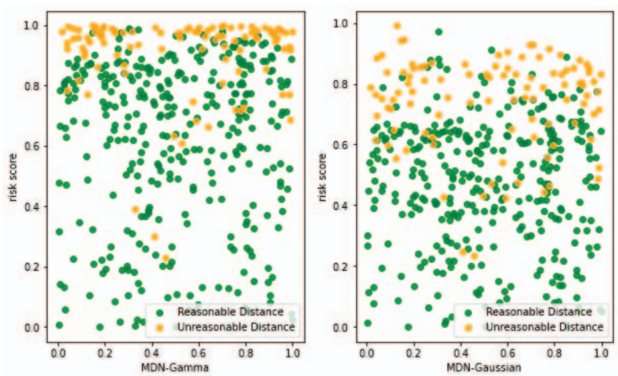

Fig. 3. Distribution of Risk Scores from the MDN Models

Table III shows the performance metrics of these models using precision, recall, and F1-score. We find that the MDN-Gamma model outperforms the MDN-Gaussian model with precision of 0.80, recall of 0.82, and F1-score of 0.81.

Across both approaches (Tables II and III), we find that MDN-Gamma outperforms other methods with F1-score of 0.81, closely followed by NODE with F1-score of 0.80. Although XGBoost has a higher precision compared to these models, it produces more false negatives and hence a lower recall. When comparing MDN-Gamma with MDN-Gaussian (Fig 3), we observe that MDN-Gamma more effectively distinguishes anomalous records from normal ones. Over 90% outliers are assigned a risk score ($\frac{1}{1+p\text{-value}}$ such that smaller p-values are closer to 1) higher than 0.6 by MDN-Gamma. While these results are on a relatively small manually labeled dataset, they are promising.

V. Conclusions

In this paper we propose two approaches for identification of high risk claims based on travel distances between members and providers, taking into account various aspects of claims, members, providers and census data, as well as other data-driven features. A neural network-based approach paired with a probabilistic outlier detection technique is able to predict anomalous travel distances with good results compared to other semi-supervised approaches. An example of such high risk claim in our data is when a member requiring interventional pain management travels for almost 700 miles to seek services from an out-of-state provider. Another such example is when a member requiring mental health services due to opioids travels for 950 miles to see a physician assistance specializing in Gastroenterology. Models for detecting healthcare FWA, like the one presented in this paper, are often employed in post-payment scenarios to identify high-risk claims that require additional scrutiny by Program Integrity SMSs and investigators. The primary goal is not just to detect FWA but to deter it, creating a healthcare system that's more efficient, trustworthy, and cost-effective.

In future, we would like to explore the temporal nature of claims that are at high risk for travel distances. We would also like to explore through graph analytics-based methods whether there are clusters of providers who share similar risk profiles for these high risk claims.

VI. Acknowledgments

We would like to thank Gautam Goel for his assistance with data profiling. We would also like to thank Deloitte's Program Integrity SMSs Dan Olson, Dave Dutton and Patricia Hansrote for their guidance and help with labeling of the data for the evaluation of the models that predict anomalous travel distances.

References

[1] B. Jones and A. Jing, "Prevention not cure in tackling health-care fraud," *World Health Organization. Bulletin of the World Health Organization*, vol. 89, no. 12, p. 858, 2011.

[2] N. Q. Tran, G. Goel, N. Pudota, M. Suesserman, J. Helms, D. Lasaga, D. Olson, E. Bowen, and S. Bhattacharya, "Leveraging deep survival models to predict quality of care risk in diverse hospital readmissions," vol. 13, no. 1, p. 10479. [Online]. Available: https://doi.org/10.1038/s41598-023-37477-3

[3] S. Aguinaga, D. Lasaga, S. Danthuri, J. Helms, E. Bowen, and S. Bhattacharya, "Identification of providers with similar risk profiles in healthcare claims using graphs," in *2023 11th International Symposium on Digital Forensics and Security (ISDFS)*, pp. 1–6.

[4] C. Policastro and B. K. Payne, "An examination of deviance and deviants in the durable medical equipment (dme) field: Characteristics, consequences, and responses to fraud," *Deviant Behavior*, vol. 34, no. 3, pp. 191–207, 2013.

[5] W. Moreto, J. Gau, and E. Brooke, "Pill mills, occupational offending, and situational crime prevention: a framework for analyzing offender behavior and adaptation," *Security Journal*, vol. 33, 06 2020.

[6] R. Simeone, "Doctor shopping behavior and the diversion of prescription opioids," *Substance abuse: research and treatment*, vol. 11, p. 1178221817696077, 2017.

[7] R. M. Musal, "Two models to investigate medicare fraud within unsupervised databases," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8628–8633, 2010.

[8] Q. Liu and M. Vasarhelyi, "Healthcare fraud detection: A survey and a clustering model incorporating geo-location information," in *29th world continuous auditing and reporting symposium (29WCARS), Brisbane, Australia*, 2013.

[9] C. M. Bishop, "Mixture density networks," 1994.

[10] J. Liu, E. Bier, A. Wilson, J. A. Guerra-Gomez, T. Honda, K. Sricharan, L. Gilpin, and D. Davies, "Graph analysis for detecting fraud, waste, and abuse in healthcare data," *AI Magazine*, vol. 37, no. 2, pp. 33–46, 2016.

[11] A. Rosenblum, C. M. Cleland, C. Fong, D. J. Kayman, B. Tempalski, and M. Parrino, "Distance traveled and cross-state commuting to opioid treatment programs in the united states," *Journal of environmental and public health*, vol. 2011, 2011.

[12] L.-C. Chen, J.-T. Sheu, Y.-J. Chuang, and Y. Tsao, "Predicting the travel distance of patients to access healthcare using deep neural networks," *arXiv preprint arXiv:2112.03541*, 2021.

[13] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: a systematic review," *IEEE Access*, 2021.

[14] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–38, 2021.

[15] M. Steinbach, L. Ertöz, and V. Kumar, "The challenges of clustering high dimensional data," in *New directions in statistical physics*. Springer, 2004, pp. 273–309.

[16] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data: a review," *Acm sigkdd explorations newsletter*, vol. 6, no. 1, pp. 90–105, 2004.

[17] I. Assent, "Clustering high dimensional data," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 4, pp. 340–350, 2012.

[18] D. H. Peters, A. Garg, G. Bloom, D. G. Walker, W. R. Brieger, and M. Hafizur Rahman, "Poverty and access to health care in developing countries," *Annals of the new York Academy of Sciences*, vol. 1136, no. 1, pp. 161–171, 2008.

[19] N. R. Chopde and M. Nichat, "Landmark based shortest path detection by using a* and haversine formula," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 1, no. 2, pp. 298–302, 2013.

[20] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive computing*, vol. 7, no. 4, pp. 12–18, 2008.

[21] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, 1968.

[22] R. C. Das and T. Alam, "Location based emergency medical assistance system using openstreetmap," in *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*. IEEE, 2014, pp. 1–5.

[23] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization." *Journal of machine learning research*, vol. 13, no. 2, 2012.

[24] Scikit-learn, "Randomforestregressor," 2022, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html.

[25] XGBoost, "Xgboost," 2022, https://xgboost.readthedocs.io/en/stable/parameter.html.

[26] S. O. Arık and T. Pfister, "Tabnet: Attentive interpretable tabular learning," in *AAAI*, vol. 35, 2021, pp. 6679–6687.

[27] S. Popov, S. Morozov, and A. Babenko, "Neural oblivious decision ensembles for deep learning on tabular data," *arXiv preprint arXiv:1909.06312*, 2019.

[28] A. V. Dorogush, V. Ershov, and A. Gulin, "Catboost: gradient boosting with categorical features support," *arXiv preprint arXiv:1810.11363*, 2018.

[29] J. Burkardt, "The truncated normal distribution," *Department of Scientific Computing Website, Florida State University*, vol. 1, p. 35, 2014.