Table 1: Tabulated Research Summary

Year; Article Authors	Objective	Methods / Algorithms used	Data Sources	Software used	Fraud labels assigned	Performance Metrics used	Conclusions
2006; Peng et al.[63]	To assess out of box clustering algorithms' results and performance using two different tools	K-means clustering	Health insurance claims data from one state in US	SAS Enterprise Miner and CLUTO (developed by Univ of Minnesota)	N/A	N/A	Based on external quality measures, the authors conclude that clustering algorithm in SAS Miner performs better and gives more meaningful clusters than CLUTO.
2012; Shin et al.[64]	To generate an index/rank for all providers based on their abusive behavior from 187 features	Correlation analysis Logistic Regression t-test Entropy reduction Discriminant analysis Chi-square test	2007 (third and fourth quarter) claims filed with South Korea's Health Insurance Review and Assessment Services (HIRA)	N/A	Intervention decisions made manually on the fourth-quarter data by HIRA	Confusion matrices	A simplified scoring mechanism is proposed by the authors based on different tests and algorithms. This helps quantify the degree of abusiveness performed by a provider. A single index that ties the degree of anomalies from different features into one score helps the investigation process. An added benefit is the application of a decision tree to help interpret the single index that is comprised from different anomaly measures.

2013; Ekin et al.[65]	To detect healthcare fraud between providers and beneficiaries using a Bayesian co-clustering framework	Markov chain Monte Carlo (MCMC)	Simulated data was used to test feasibility of the framework	N/A	N/A	N/A	A Bayesian framework of co-clustering using Gibbs sampler was used to assess posterior probability of a provider-beneficiary co-cluster. A conceptual model to identify fraudulent associations or links was proposed by the authors.
2013; Eldardiry et al.[66]	To identify suspicious pharmacies and provide leads to auditors and investigators	Term frequency-inverse document frequency	2012 government healthcare program pharmacy claims	N/A	N/A	N/A	A probabilistic outlier detection technique based on rule (developed by domain experts and data) violations were developed and was evaluated based on domain expert inputs on the labeled pharmacies.
2013; Thornton et al.[67]	To predict different types of healthcare fraud in Medicaid	N/A	N/A	N/A	N/A	N/A	Design of a multidimensional data models and analysis techniques to reveal conspirator and stand-alone fraud types are discussed by these authors.
2014; Bowblis et al.[68]	To test geographic reimbursement effect against skilled nursing facilitys' (SNFs) profit margin and to identify likelihood of upcoding in SNFs resource utilization groups (RUG) codes B)	Linear regression Mixed logit model Multivariate regression model					Up-coding likelihood increases in high cost of living areas due to the geographic payment differentials per RUG code.

2014; Joudaki et al.[69]	To provide a literature review on healthcare fraud and abuse data mining techniques	Range of classifier algorithms	N/A	N/A	N/A	N/A	21 different studies are briefly outlined based on their review goals. These studies span all methods applicable to different healthcare systems in different countries.
2016; Bauder et al.[70]	To detect anomalous billing behavior of physicians by categorizing physician's field of practice	Multinomial Naïve Bayes 5 fold cross validation	CMS Medicare Provider Utilization and Payment Data: Physician and Other Supplier 2013 calendar year - State of Florida	Weka	Physician specialty available in CMS input data	Precision, Recall, F-score	A multi-class naïve Bayes model was used to classify physicians based on their number of instances of billed procedure code. An inherent limitation to this study is not accounting for physician types who can perform the same set/group of procedure codes.
2016; Joudaki et al.[71]	To identify features in general physicians' drug prescription claims delineating between fraudulent and non-fraudulent providers	Hierarchical clustering, Linear Discriminant Analysis (LDA)	2011 Physician data and Drug Prescription data from Social Security Organization, Iran.	N/A	N/A	Accuracy	The clusters based on the indicators were used to evaluate performance of LDA classification of physicians.
2016; van Capelleveen et al.[72]	To detect outliers in the Medicaid dental domain	Descriptive analytics such as box-plots, peak-analysis. Regression and Clustering (K-means)	One state's dental Medicaid data from Medicaid Management Information System and federal data extract of the same state from Medicaid Statistical Information System	R	N/A	N/A	This study detects outliers using univariate and multivariate outlier procedures. The features in case of General Linear Models were based on expert input and derived from claim details with a set of inclusion and exclusion criteria for the provider claims.

2017; Bauder et al.[73]	To survey on the state of healthcare up-coding fraud analysis and detection.	Review of methods from literature	N/A	N/A	N/A	N/A	A comprehensive literature review specifically focusing on up-coding scheme is discussed by these authors. They cover review on different domains such as nursing facility RUGs up-coding, E/M up-coding in emergency department, DRG up-coding, etc.
2017; Ekin et al.[74]	To detect anomalies in prescribed services in comparison to peers using concentration function.	Lorenz curve	CMS Medicare Provider Utilization and Payment Data: Physician and Other Supplier - subsetted		N/A	Likelihood ratio-based thresholds, Gini area of concentration and Pietra's index	A simple and elegant pre-screening tool to detect billing anomalies (number of services and percentage of services in comparison to peer providers) using graphical methods is proposed and demonstrated on a limited data from the source.
2017; Sadiq et al.[75]	To provide anomaly detection using PRIM (an unsupervised learning method) and compare model results with known classifiers	Patient Rule Induction Method (PRIM) based bump hunting	CMS Medicare Provider Utilization and Payment Data: Physician and Other Supplier, Prescriber and DMEPOS data 2012 – 2014 calendar year	R PRIMsrc package	N/A	F-Score and Accuracy	PRIM model achieves better F-score and accuracy results in comparison to Support Vector Machine, Naïve Bayes, Random Forest, Discriminant analysis and Logistic Regression classifiers
2018 May; Bauder et al.[76]	To evaluate the effect of class imbalance in fraud detection	Random Forest 5-fold cross validation with 10 time repetition	CMS Medicare Provider Utilization and Payment Data: Physician and Other Supplier 2012 – 2015 calendar year	Weka	LEIE Database (January 2018)	AUC	Training dataset with a 90 (non-fraud) - 10 (fraud) class distribution ratio is considered best for a good classifier to delineate between the two classes instead of the traditional 50

							(non-fraud) -50 (fraud) class distribution ratio
2018; Bauder et al.[77]	To evaluate the effect of class imbalance in fraud detection	C4.5 Logistic Regression Support Vector Machines 5 fold cross validation with 10-time repetition	CMS Medicare Provider Utilization and Payment Data: Physician and Other Supplier 2012 – 2015 calendar year	Weka	LEIE Database (2017)	AUC, FPR, FNR, Tukeys HSD, ANOVA	No single learner was considered best, however authors conclude that LOGISTIC REGRESSION and C4.5 had performed better with an 80 (non-fraud) - 20 (fraud) class distribution ratio
2018; Herland et al.[78]	To detect Medicare fraud using publicly available CMS data	Logistic regression, Gradient Boosted Trees and Random Forest	CMS Medicare Provider Utilization and Payment Data: Physician and Other Supplier (2012 - 2015), Prescriber (2013 - 2015)and DMEPOS data 2013 – 2015 calendar year	Apache Spark 2.3.0 Machine Learning Library	LEIE Database (2016)	Area under curve	The LEIE database values for providers were used as ground truth and logistic regression results had the best performance amongst the different algorithms tested for this data.
2019; Fan et al.[79]	To detect physician fraud using open data	Logistic regression, Naïve Bayes, SVM and decision tree	CMS Part B, Part D and DMEPOS, Social media data (Healthgrades.com)	N/A	LEIE Database (2015-16) and Board actions	F1 score	Decision tree with features based on social media rating had the best performance rate.

2019; Herland et al.[80]	To address effect of class imbalance in supervised learning algorithms using CMS data	Logistic regression, Gradient Boosted Trees and Random Forest	CMS Medicare Provider Utilization and Payment Data: Physician and Other Supplier (2012 - 2016), Prescriber (2013 - 2016) and DMEPOS data 2013 – 2016 calendar year	Apache Spark 2.3.0 Machine Learning Library	LEIE Database (2016)	Area under curve	They employed different random undersampling techniques to understand the effect of class imbalance in supervised learning algorithms. They concluded a 10:90 ratio of class is needed for a learner to discriminate well and that a 50:50 class ratio might not help with learning the discriminative patterns.
2019; Sadiq et al.[81]	To see if a deliberate fraudulent action causes a perturbation in the observational data, accounting for any co-variates or confounding factors that could lead to that fraudulent outcome	Bump hunting, Propensity matching, inertia based clustering, mixed variable cosine distance	CMS Medicare (2012 - 2015)	N/A	LEIE Database (2017)	Accuracy, F1-score, Precision and recall	The authors devised a framework using unsupervised statistical methods and compared their framework performance to that of other methods such as k-means clustering. They conclude that this framework outperforms other existent models in literature.
2019; Zafari and Ekin[82]	To detect prescription fraud using topic modelling	Structural topic modelling and concentration functions	CMS Medicare Prescriber Part D (2015)	R (STM package)	N/A	N/A	A novel unsupervised prescription fraud framework consisting of two main steps to identify fraudulent prescriptions; step 1 involves identifying associations between prescribers and their drug billing utilization (using STM) followed by step 2 that uses these groupings to detect outliers within a specialty.

2021; Ekin et al.[83]	To review the performance of health care fraud classifiers in practice	Random forest, Naïve Bayes, Support vector machines, k-nearest neighbor, Neural nets, Linear discriminant analysis, Quadratic discriminant analysis, Random walk oversampling, random under sampling, SMOTE, Majority weighted Minority oversampling, PCA, Remotion	CMS Medicare Provider Utilization and Payment Data: Physician and Other Supplier, CMS Medicare Geographic Variation Public Use File, CMS Physician Fee schedules	R (Mice, imbalance, smotefamily packages)	Synthetic and LEIE database	Area under curve	The authors experimented feature engineering, class-imbalance, computing time and several algorithm techniques in combination with the above. They conclude that any classifier with a combination of good sampling and collinearity approach will perform best with AUC upto 0.9994
2021; Ai et al.[84]	Systematic review of healthcare fraud	Range of classifier algorithms	N/A	N/A	N/A	N/A	27 different studies are briefly outlined based on their review goals such as assessing data mining methods, performance metrics used in evaluation, publication bias. These studies span all methods applicable to different healthcare systems in different countries.