Archives of Business Research – Vol. 13, No. 06
Publication Date: June 25, 2025
DOI:10.14738/abr.1306.18994.
Rodriguez, A. E. & Ahmed, M. F. (2025). Identification of Misclassified Medicaid Audits. *Archives of Business Research, 13*(06). 149-159.

Identification of Misclassified Medicaid Audits

A. E. Rodriguez
ORCID: 0000-0003-2779-1300
Pompea College of Business
University of New Haven

M. F. Ahmed
Pompea College of Business
University of New Haven

ABSTRACT

Medicaid auditors typically use statistically valid samples from the total claims filed by service providers to detect overpayments or underpayments. These samples often contain many "zeros" (compliant claims) and a few "ones" (flagged issues), but errors in the auditing process can occur due to human mistakes, biases, or even AI and machine learning errors. Such mistakes may lead to false-positive results where a claim is wrongly flagged as an overpayment. Is it possible to identify and isolate false positives in defending impugned claims filed by Medicaid providers without the need for expensive forensic audits of the entire sample? To address this, we test the effectiveness of noise-filtering algorithms to isolate false-positive claims from legitimate ones. By artificially injecting noise (representing false positives) into synthetically generated data we create a realistic litigation environment resembling what happens during subpoenas or document requests. We then apply three noise-filtering algorithms and find that these filters reduce the audit data to a smaller, more focused sample, making it easier to identify false positives during any subsequent manual review. Our study does not propose a novel noise-filtering method; rather, we demonstrate how existing techniques can help forensic analysts concentrate false-positive claims in a reduced sample. While the specific focus is on Medicaid audits, the findings are applicable to any situation where audits are conducted on claims filed by service providers and paid by third parties.

Keywords: Imbalanced data, noise filters, anomaly detection, forensic analysis, label noise.

"Nothing can be known without there being an appropriate "instrument" in the makeup of the knower." E.F. Schumacher

INTRODUCTION

Medicaid is a federal public insurance program that is administered by the states. It pays for health care claims for services rendered by health care providers and health plans. Audits and enforcement of the contractual terms agreed upon between Medicaid and service providers are conducted by the Medicaid Fraud Control Units (MFCUs) typically housed within a State Attorney General's office (National Association of Attorneys General, 2025).

Performance audits of paid insurance claims often reveal instances where the claims are not in compliance. Put differently, audits may reveal that there are mismatches between the documented, paid amounts and the MFCU scrutinized amounts.

Audits of claims in Medicare, insurance or any other system that relies on a third-party payer can be mishandled reflecting deliberate equivocation, systemic error, arithmetic error, or judgmental bias (Brody, DeZoort, Gupta, & Hood, 2022; Harvin & Killey, 2021; Ioannidis, 2021; Rodriguez & Kucsma, Appraising Audit Error in Medicaid Audits, 2023). Many of these audits are contested, many litigated. For example, since 2007, the Department of Justice's Health Care Fraud Unit has charged over 5,400 defendants with fraudulently billing Medicare, Medicaid and private health insurers more than 27 billion dollars (Argentieri, 2024).

Importantly, and possible consequential development in Medicare auditing is the increased use of AI and machine learning to analyze large volumes of Medicare claims data to identify potential improper billing practices (Emanual, 2025) (Zimiles & Fontecilla, 2023). These AI-powered tools can flag unusual patterns or anomalies that might indicate fraud. But this shift towards AI-driven audits may be a harbinger of increased algorithmic errors or inconsistencies that could easily lead to increased levels of false positives.

Because the audited instances flagged as improper can constitute the foundation for legal action it is important that they be characterized properly. Distinguishing the legitimate, overpaid – albeit properly filed - claims from the misclassified ones is obviously critical for the correct appraisal of recoverable funds, overbilling estimates, or pecuniary damages. In fact, an unclean, uncorrected base series may result in improper compensation amounts in associated damages or monies recovered from defendants.

Although the scrutinized service provider, defendants in the legal proceedings, can manually re-examine any contested claims, this can be prohibitively costly. The research question for us is whether it is possible to identify and separate misclassified, false positive claims from legitimate ones, without resorting to costly forensic audits? In the alternative, is it possible to eliminate those claims that are not considered questionable so as to reduce the sample of claims? An affirmative answer – even to the latter question - goes a long way towards supporting an allegation of auditor error, a plausible defense.

In this paper we examine whether noise filtering methods are useful in identifying false positive instances of scrutinized audits in a manner simulating the litigation environment of a Medicaid audit. To our knowledge, there are no known audit data with identified false-positives publicly available for scrutiny. Part of this, of course, is due to the legal protections accorded sensitive health information (US Dept of Health & Human Services, 2025). Thus, we resort to examining prototypical simulations using synthetically generated data. Simulations facilitate controlling the type of noise injected into the data as well as its amount and characteristics and thus draw relevant conclusions based on their similarity to real-life conditions.

This paper reports the result of this inquiry. It is organized as follows. A succinct review of the various class misclassification identification approaches is discussed in the next section. To provide realistic context, we set forth in the third section an archetypical situation simulating a portfolio of claims resulting from the scrutiny of a hypothetical defendant, a Medicaid services provider. In the fourth section we describe how we assemble the synthetic data used to simulate the forensic environment. Section five provides results. The last section sets forth limitations of our work and discusses next steps.

The main contributions of our work are listed below:

- It furthers the study of the importance of noisy class labels in the field of machine learning-assisted fraud detection.
- sets forth an approach to inject seeming false positive errors simulating those found in real-world audits.
- enhances the understanding of new algorithms applied to misclassified forensic audit data.
- verifies how existing noise filtering methods perform when some adverse effect causes inaccuracies in the data.
- creates synthetic data with controlled errors to test the effectiveness of filtering methods for label noise treatment in small, imbalanced datasets.

We believe our work is generalizable to similar instances whereby an auditor reviews claims filed by a service provider, paid by a third party. However, its more immediate application is for economic, financial and accounting forensic experts for use in litigation. In addition, we believe this work illustrates the relevance and usefulness of noise filtering methods on imbalanced data sets in the forensic arena for machine learning researchers.

ANALYZING AUDITOR ERROR

Medicaid auditors rely on statistically valid samples drawn from the totality of claims filed by a provider to determine whether providers are appropriately billed for Medicaid services (Kvanli & Schauer, 2018).

"Sampling avoids the cost and practical challenge of examining a large number of claims" (Office of the Inspector General, Health and Human Services, 2018)

A Medicaid auditor reviewing an individual claim from a drawn sample cannot avoid the possibility of incurring one of two errors: the auditor can incorrectly flag a truly correctly filed claim as fraudulent or inappropriate; this type of misclassification is known as a false-positive. A second type of audit-error occurs when the auditor incorrectly fails to flag a truly fraudulent or inappropriate filed claim. This latter type of misclassification error is known as a false-negative (Rodriguez & Kucsma, Appraising Audit Error in Medicaid Audits, 2023)

False positives are inevitable side-effects of both inductive and deductive fraud detection protocols. Inductive fraud detection processes scrutinize extant patterns and anomalies observed in specific instances of data and infers those patterns onto subsequent data instances. Deductive fraud detection relies on spotting violations or departures from well-defined rules. Despite the seeming dichotomy most Medicaid investigations conflate both types of audits; both tracks are unable to avoid false positives. Generally, false positives can be time consuming, distracting, reputation killers amidst other sundry hardships for businesses, for auditors. In fact, their eradication or minimization has spawned a cottage industry within the broader field of fraud detection protocols.

Label noise, also known as class noise is a relatively common data artifact in applications of machine learning to fraud (Frenay & Verleysen, 2014; Villuendas-Rey, Tusell-Rey, & Camacho-Rey, 2024; Saez, Noise Models in Classification: Unified Nomenclature, Extended Taxonomy and Pragmatic Categorization, 2022; Walauskis & Khoshgoftaar, 2025). Unaddressed, false positives ultimately may lead to poor detection model performance.

Filtering protocols are ubiquitous today, involving the including, excluding, or moderating information according to individual choice or domain-specific rules or criteria (Diakopolous, 2016). Protocols emerge to arrest intrusive or offensive social media posts at the user-interface level. Examples include news-reading applications such as Google News, Microsoft Edge, platforms such as Facebook, X, Reddit, or any other similar apps. Moderation and filtering are crucial elements when publishing or processing human-in-the-loop contributions – like email, or social media commentary or online communities' contributions. Online comments and reviews are routinely filtered algorithmically to apprise their relevancy and propriety for public consumption (Akash, Amalan, Kumar, & Ram, 2021).

Within this broad umbrella, much research has been accorded to developing and operationalizing noise abatement, filtering or removal techniques (Blachnik, 2017; Villuendas-Rey, Tusell-Rey, & Camacho-Rey, 2024). A core element of this assembled repertoire for sanitizing noisy labels are the filtering algorithms based on k-nearest neighbor (kNN) predictors. kNN methods are preferred for their simplicity and intuitiveness (Torgo, 2011). At their core, kNN methods predict a case, or instance class based on its similarity to existing (training) cases – for which it earns them the appellate "lazy learner." Formally, this approach, known as 'instance-based learning" entails a characterization of "similarity" typically in the form of a distance metric (Aggarwal, 2014).

THE APPRAISAL OF MISCLASSIFICATION IN SMALL, IMBALANCED DATASETS

There is a vast and extensive research program examining the impact of label or class misclassification on classification accuracy (Schennach, 2016; Saez, Noise Models in Classification: Unified Nomenclature, Extended Taxonomy and Pragmatic Categorization, 2022; Frenay & Verleysen, 2014). We draw much from those efforts for our present analysis especially the literature on noise filtering of imbalanced data sets (Szeghalmy & Fazekas, 2024; Van Hulse, Khoshgoftaar, & Napolitano, 2011; Saez, Luengo, & Herrera, Predicting Noise Filtering Efficacy with Data Complexity Measures for Nearest Neigbor Classification, 2012). Our contribution tries to hone the takeaways and insights from the noise-filtering literature and apply them to identifying instances of false-positive, misclassified labels within small samples of imbalanced data. In classification analysis, noise can be found in both the attributes and labels. However, our focus is not on the impact of noisy attributes; rather, careful understanding of attribute noise on misclassified labels will remain the topic of later work.

In the specialized literature there exist two main approaches to deal with label noise (Frenay & Verleysen, 2014). Algorithm level approaches attempt to create robust classification algorithms that are little influenced by the presence of noise. This includes approaches where existing algorithms are modified to cope with label noise by either modeling it in the classifier construction (Northcutt, Athalye, & Mueller, 2019), by applying pruning strategies to avoid overfitting or by diminishing the importance of noisy instances with respect to clean ones (Northcutt, Athalye, & Mueller, 2019).

Alternatively, data level filters try to develop strategies to cleanse the dataset by iteratively filtering noisy instances, computing metrics on the data or even hybrid approaches that combine several of these strategies (Saez, Luengo, Stefanowski, & Herrera, 2014).

There exist recent proposals that combine these two approaches, which model the noise and give less relevance to potentially noisy instances in the classifier building process (Bouveyron & Girard, 2009).

Given that our interest is merely to identify a subset which contains most false positives, we have little use for the algorithm level approach. Instead, our focus is on noise removal and noise reparation strategies. The first option removes the noisy instances, whereas the second relabels these instances with the more likely label on the basis of the available information. The hybrid approaches carry out relabeling when they have enough confidence on the new label. Otherwise, they remove the noisy instance (Frenay & Verleysen, 2014).

Numerous possible filters, hybrid, similarity-based, and saturation ones, are available for this exercise; the NoiseFiltersR package alone lists 30 filters: 13 ensemble-based filters, 14 similarity-based and 3 based on data complexity measures (Morales, et al., 2017). Specifically, we use the following three methods: Condensed Nearest Neighbor (CNN), Edited Nearest Neighbor ENN), and Ensemble Filter (EF).

Condensed Nearest Neighbor (CNN) identifies a subset of the training data that can classify the original dataset using a one-nearest neighbor rule almost as accurately as the full dataset. Put differently, CNN removes majority class samples that are far from the decision boundary while retaining those that are close to the decision boundary (Morales, et al., 2017).

Edited Nearest Neighbor (ENN): This method reduces the size of the majority class (labeled as 0's) by undersampling. It carefully selects and deletes instances from the majority class if at least 2 of its 3 nearest neighbors belong to the minority class. It works by removing instances from the majority class (labeled as 0's) that are misclassified by their k-nearest neighbors (Morales, et al., 2017). This action effectively reduces the number of majority class instances. In doing so, it reduces the influence of the mislabeled majority, isolating the instances of misclassification among the instances flagged as fraud.

Ensemble Filter (EF): A label noise ensemble filter uses an ensemble of three different base classifiers (C4.5, 1-KNN, LDA) to identify and filter out instances with incorrect or noisy labels from a training or initial dataset. The algorithm leverages the collective wisdom of the three classifiers to flag instances where a high proportion of them indicate potential mislabeled data (Morales, et al., 2017).

URL: http://doi.org/10.14738/abr.1306.18994

We neither conduct nor appraise the usefulness or the capabilities of the other available noise filters. Our interest is largely on the applicability of the concept: introducing the possibility and demonstrating how to deploy noise-filtering algorithms to isolate any false positives in an audit data set.

EMPIRICAL METHODOLOGY

Medicare audits are designed to establish the total amount of ineligible claims. If at fault on a particular claim, the amount of overpayment (or underpayment) is documented. For the most part, compliance violations are a small proportion of the total.

We generate simulated multivariate data with five continuous predictors with variance equal to one and a co-variance equal to 0.65. Varying the feature covariance (between 0.25 and 0.75) proved to have negligible impact on the detection of false positives; we do not show those results here. We specify an equally-weighted logistic regression model to generate probabilities representing the audit outcome for each particular instance. We assume even (log) odds for each of the predictor coefficients. Finally, we use a binomial regression of size equal to one to generate an imbalanced Bernoulli series. The resulting representative tranche consists of a small sample consisting of an exceptionally significant percentage of zero values reflecting the fact that most audited claims are in compliance.

Table 1 displays a stylized representation of the ground-truth composition of audit results. The ellipsis is meant to represent an imbalanced dataset, a significant larger proportion of zeros. Zeros represents an audited claim not found to be problematic; symmetrically, ones represent flagged claims. Audit sample sizes ranging from 100 to 300 is not atypical. Accordingly, the tables represent a stylized depiction of the audit sample that in reality may range in size from 150 to 300.

Table 1: Latent (Unknown) Compliance Status (Ground Truth)

0	0	0	0	0	…	0	0	0	1	1	1

Upon conclusion of the auditing process, the assembled dataset is described in Table 2. Again, the ellipsis represents the imbalance. Table 2 is aligned with Table 1. The fact that a zero in the first cell on the left is a zero, reveals no mismatch between the provider billed amount and the amount paid by Medicaid for that particular instance. The zero matches the ground-truth value in Table 1. A one in the second cell identifies an audited claim that has been found wanting but is – in reality, unimpaired.

Table 2: Result of Random Audit

0	1	0	0	1	…	0	1	0	0	1	0

Note that by extrapolating the findings from a sample to the entire population of claims – Table 2 above represents a particular realization - allows auditors to determine the potential financial impact of errors or fraud without reviewing every single claim (Kvanli & Schauer, 2018; Office of the Inspector General, Health and Human Services, 2018).

Table 3 is aligned with Table 2. The fact that the results of the audit in Table 2 match the ground truth results in no misclassification flagged. On the other hand, the second column in Table 2, is coded as a one, indicating that the auditor considered that instance a problem. The one is inconsistent with the ground-truth set forth in Table 1 and is therefore considered a False Positive. Similarly, the mismatch between a ground truth of one (in Table 1) in Column 10, and an audit result of zero, as noted in Table 2 also in Column 10, indicated an audit mistake, in this instance a False Negative. FP and FN are thus identified in Table 3.

Table 3: Misclassified Instances (Unknown)

0	FP	0	0	FP	…	0	FP	0	FN	0	FN

This construction allows us to assemble a data set where we know, ahead of time, which outcomes are FPs. The analytical procedure entails determining and appraising the performance of the proposed algorithms in isolating the FPs. We analyze the performance of the different methods using predictor accuracy as our selection metric. Accuracy is the ratio of correct predictions to the total number of predictions (Torgo, 2011).

RESULTS

Table 4 below provides the estimated accuracy of the various filtering algorithms for the randomly drawn data sets of varied sizes. The table sets forth the simulation parameters across instances of non-zero proportions and data size.

The proportion of false positives injected into the synthetic data is in the 1st column, labeled False Positives. This represents the proportion of claims that are likely to be erroneously classified as fraudulent. Sample sizes measure false positives, in Column 2.

The third column, "EF" is the resulting accuracy from identified the proportion of ones, False Positive identification. EF is ensemble of three different classifiers (C4.5, KNN, LDA) with consensus voting.

Table 4: Simulation Results

False Positives	Sample Size	EF	CNN	Class SF
0.03	150	0.987	0.941	0.987
0.05	150	0.980	0.933	0.980
0.10	150	0.947	0.907	0.867
0.03	225	0.991	0.958	0.973
0.05	225	0.982	0.933	0.964
0.10	225	0.951	0.896	0.889
0.03	300	0.993	0.957	0.993
0.05	300	0.987	0.935	0.980
0.10	300	0.967	0.915	0.943

k nearest neighbor-based filters constitute simple and effective algorithms, especially in problems such as a forensic audit process where interpretability and simplicity are indispensable. The logic of kNN is easy to explain to non-technical people, making it straightforward to communicate results to a trier-of-fact.

URL: http://doi.org/10.14738/abr.1306.18994

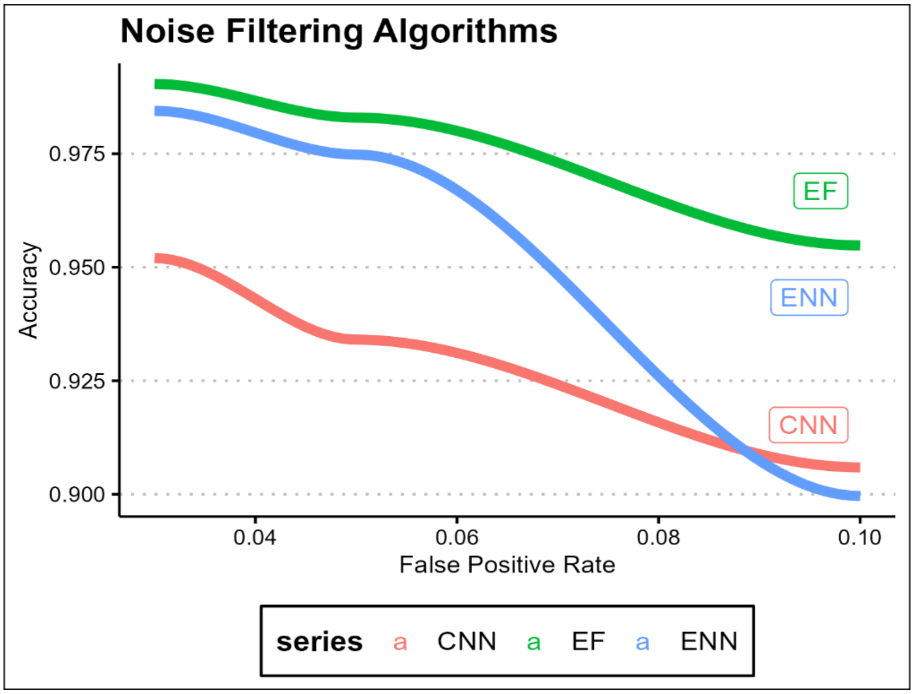

Figure 1: Filtering Accuracy by FP Rate

The results displayed in Figure 1 suggest notable levels of accuracy across the two parametrized dimensions for all three filter and for the Edited Nearest Neighbor. And amidst overall good performance the ENN method stands out. ENN maintain accuracy levels above 95 percent even with a 10 percent false positive rate. In this iteration the sample size remained constant at 300 claims audited.

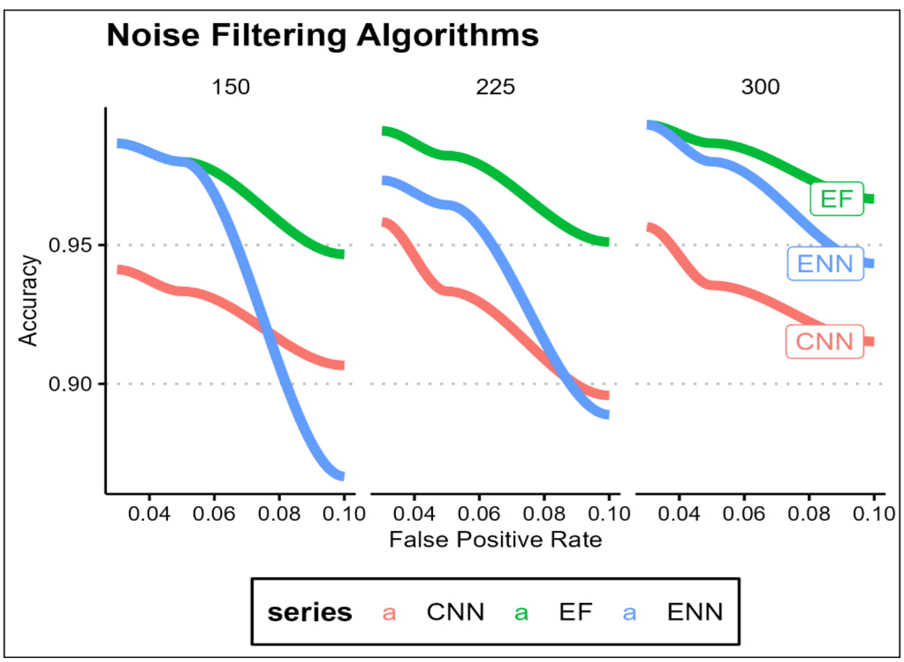

Figure 2: Filtering Accuracy by FP Rate and Sample Size

As one would surmise, an increase in the size of the audit sample results in improved filter performance. Again, the performance of the Edited Nearest Neighbor Stands when the sample size is varied.

CONCLUDING COMMENTS AND DISCUSSION

We study the robustness of performance of three noise filtering models across variations of two different parameters: the percentages of false positive rates and different sample sizes. And we did so by carefully mimicking conditions characterizing a Medicaid fraud audit. Specifically, the presence of small, imbalanced, synthetic datasets which contain a set of correlated predictors specifically designed to contribute to the class determination. The parameter variation setting is intended to simulate the possible litigation environments of a Medicaid fraud audit.

Naturally, one wonders whether fewer predictors would return acceptable accuracy. Moreover, the presence of attribute noise on resulting accuracies was left unexamined. Both of these issues, the relevance of the number of predictors and the impact of attribute noise, may be a topics for later work which may enhance the understanding of the proposed method (Pau, Perniciano, Pes, & Rubattu, 2023). The sensitivity of other parameters – such as the covariance between predictors or the level of imbalance on classification performance - may also be a fruitful, later inquiry.

Our results find that the three noise filters examined worked remarkably well in reducing the choice set of claims that can be manually re-examined by defendants. To then confirm the existence of false positives amidst an audit raises a rebuttable presumption that may enhance a defendant's legal position.

Most data used heretofore in litigation proceedings consisted solely of a Bernoulli series showing cleared and impugned audits and no predictors. Thus, defendants would have to extend the dataset to incorporate predictors for the methods proposed here to have any sense of working as proven. Left to itself this practice limits the usefulness of the filtering methods examined here. However, the advent of AI-enhanced audit procedures by Medicaid examiners may necessarily rely on multiple features in its algorithmic protocols. This extension conveys the necessary breadth required by the proposed noise filters.

In principle, any number of other anomaly detection procedures can be deployed given the availability of the particulars of any feature-enhanced AI and machine learning assisted audits disclosed in litigation. Methods that can be plausibly used to similarly reduce the number of claims in attempts to identify those instances most likely to be false positives (Torgo, 2011). For example, one can imagine productive uses for this task of, *inter alia*, semi-supervised methods, isolation forests, the more traditional filters such as naïve bayes and more uncommon ones such as PRIDIT and local outlier probabilities improve on the methods here when placed within a litigation context (Walauskis & Khoshgoftaar, 2025). We chose to highlight kNN-based methods not because they are optimal, but rather because they are simple, widely available, easy to implement, and relatively straightforward to interpret compared to some other machine learning approaches. Importantly, kNN methods are likely to overcome any apprehensions proceeding from the increasing reach of General Data Protection Regulation-type regulations. GDSR-like encroachments of algorithmic sophistication place a premium on

URL: http://doi.org/10.14738/abr.1306.18994

interpretable, parsimonious and easy to explain algorithms (Rodriguez, Duman, & Kuntze, Two General Data Protection Regulation (GDPR) Compliant Approaches to Scoring Firm Financial Frailty in Business Litigation, 2025).

In sum and to be sure, there are number of outstanding queries that should be addressed to ensure the vitality and robustness of what we propose here. There is much to be gained, however, by considering the benefits of this initial, critical survey of the readiness of noise filters for audit litigation.

References

Aggarwal, C. C. (2014). Data Classification. In C. C. Aggarwal, *Instance-based Learning: A Survey, Data Classification: Algorithms and Applications.* Chapman and Hall/CRC Press.

Akash, S., Amalan, R., Kumar, C. B., & Ram, O. B. (2021, April). Detecting and Characterizing Reviewer Groups in Online Product Reviews. *International Research Journal of Engineering and Technology, 8*(4), 835-839.

Argentieri, N. M. (2024, July Monday). Combating Health Care Fraud: 2024 National Enforcement Action. *Blog Post*. U.S. Department of Justice. From https://www.justice.gov/archives/opa/blog/combating-health-care-fraud-2024-national-enforcement-action

Bekkum, M. v. (2025, February 16). Using sensitive data to de-bias AI systems: Article 10(5) of the EU AI act. *Computer Law & Security Review, 56*. doi:https://doi.org/10.1016/j.clsr.2025.106115

Blachnik, M. (2017). Instance Selection for Classifier Performance Estimation in Meta Learning. *Entropy, 19*. doi:doi:10.3390/e19110583

Bouveyron, C., & Girard, S. (2009, February 12). Robust supervised classification with mixture models: Learning from data with uncertain labels. *Pattern Recognition*, 2649-2658. From https://hal.science/hal-00325263/file/RR_RMDA.pdf

Breunig, M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying Density-Based Local Outliers. *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 93-104. doi:https://doi.org/10.1145/342009.335388

Brody, R. G., DeZoort, F. T., Gupta, G., & Hood, M. B. (2022). The Effects of Cognitive Bias on Fraud Examiner Judgments and Decisions. *Journal of Forensic Accounting Research, 7*(1), 50-63. doi:https://doi.org/10.2308/JFAR-2020-030

Diakopolous, N. (2016, February). Accountability in Algorithmic Decision Making. *Communications of the ACM, 59*(2). doi:http://dx.doi.org/10.1145/2844110

Emanual, K. C. (2025, February 5). *AI Poised to Take the Reins for Medicare Audits*. From RACmonitor: https://racmonitor.medlearn.com/ai-poised-to-take-the-reins-for-medicare-audits/

Frenay, B., & Verleysen, M. (2014, May). Classification in the Presence of Label Noise: A Survey. *IEEE Transactions on Neural Networks and Learning Systems, 25*(5), 845-869. doi:doi.org/10.1109/TNNLS.2013.2292894

Harvin, O., & Killey, M. (2021). Do "Superstar" CEOs Impair Auditors' Judgement and Reduce Fraud Detection Opportunities? *Journal of Forensic and Investigative Accounting, 13*(3). From http://s3.amazonaws.com/web.nacva.com/JFIA/Issues/JFIA-2021-No3-7.pdf

Ioannidis, J. P. (2021). Over- and under-estimation of COVID019 deaths. *European Journal of Epidemiology, 36*(6), 581-588. doi:https://doi.org/10.1007%2Fs10654-021-00787-9

Kvanli, A., & Schauer, R. (2018). *Handbook for HealthCare Auditors.* Alan Kvanli.

Morales, P., Luengo, J., Garcia , L. P., Lorena, A. C., de Carvalho, A. C., & Herrera, F. (2017, June). The NoiseFiltersR Package: Label Noise Preproscessing in R. *The R Journal, 9*(1), 219-228.

National Association of Attorneys General. (2025, May 5). *About the Medicaid Fraud Control Units.* From National Association of Attorneys General: https://www.naag.org/about-naag/namfcu/about-the-medicaid-fraud-control-units/

Nigrini, M. J. (2020). *Forensic Analytics.* New York: John Wiley & Sons Inc.

Northcutt, C. G., Athalye, A., & Mueller, J. (2019, November 7). Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks. *arXiv: 2103.14749*. From https://arxiv.org/pdf/2103.14749

Office of the Inspector General, Health and Human Services. (2018). *Statistical Sampling: A Toolkit for MFCUs.* From https://oig.hhs.gov/fraud/medicaid-fraud-control-units-mfcu/files/MFCU-Sampling-Guidance.pdf

Pau, S., Perniciano, A., Pes, B., & Rubattu, D. (2023). An Evaluation of Feature Selection Robustness on Class Noisy Data. *Information, 14*. doi:doi.org/10.3390/info14080438

Rodriguez, A. E., & Kucsma, K. (2023). Appraising Audit Error in Medicaid Audits. *International Journal of Accounting and Financial Reporting, 13*(3), 2162-3082. doi:https://doi.org/10.5296/ijafr.v13i3

Rodriguez, A. E., Duman, G. M., & Kuntze, R. (2025, May). Two General Data Protection Regulation (GDPR) Compliant Approaches to Scoring Firm Financial Frailty in Business Litigation. *American Business Review, 28*(1), 1-14. doi:10.37625/abr.27.2.372-400

Saez, J. A. (2022, 10). Noise Models in Classification: Unified Nomenclature, Extended Taxonomy and Pragmatic Categorization. *Mathematics, 10*. doi:https://doi.org/10.3390/math10203736

Saez, J. A., Luengo, J., & Herrera, F. (2012, July 23). Predicting Noise Filtering Efficacy with Data Complexity Measures for Nearest Neigbor Classification. *Pattern Recognition*. doi:http://dx.doi.org/10.1016/j.patcog.2012.07.009

Saez, J. A., Luengo, J., Stefanowski, J., & Herrera, F. (2014). SMOTE–IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering. *Information Sciences*, 184-203. doi:http://dx.doi.org/10.1016/j.ins.2014.08.051

Schennach, S. M. (2016). Recent Advances in the Measurement Error Litarature. *Annual Review of Economics, 8*, 341-377. doi:10.1146/annurev-economics-080315-015058

Szeghalmy, S., & Fazekas, A. (2024, October). A Comparative Study on Noise Filtering of Imbalanced Data Sets. *Knowledged-Based Systems*, 1-17. doi:doi.org/10.1016/j.knosys.2024.112236

Torgo, L. (2011). *Data Mining with R.* Boca Raton, FL: Chapman & Hall/CRC.

US Dept of Health & Human Services. (2025, May 5). *Summary of the HIPAA Privacy Rule.* From US Dept. of Health and Human Services: https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html#:~:text=The%20Health%20Insurance%20Portability%20and%20Accountability%20Act%20of%201996%20(HIPAA,was%20published%20December%2028%2C%202000.

Van Hulse, J., Khoshgoftaar, T. M., & Napolitano, A. (2011). An Exploration of Learning When Data is Noisy. *Intelligent Data Analysis*, 215-236. doi:doi.org/10.3233/IDA-2010-0464

Villuendas-Rey, Y., Tusell-Rey, C. C., & Camacho-Rey, O. (2024, September 19). Simultaneous Instance and Attribute Selection for Noise Filtering. *Applied Sciences*. doi:https://doi.org/10.3390/app14188459

Walauskis, M. A., & Khoshgoftaar, T. M. (2025). Unsupervised Label Generation for Severely Imbalanced Fraud Data. *Journal of Big Data, 12*(63), 1-23.

Zimiles, E., & Fontecilla, R. (2023, June 9). *AI and machine learning – an intelligent approach to healthcare fraud prevention.* Retrieved May 5, 2025 from healthcare financial management association: https://www.hfma.org/cost-effectiveness-of-health/ai-and-machine-learning-an-intelligent-approach-to-healthcare-fraud-prevention/