Available online at www.sciencedirect.com
ScienceDirect
The Journal of Finance and Data Science 8 (2022) 138–144
http://www.keaipublishing.com/en/journals/jfds/

Audit data analytics, machine learning, and full population testing

Feiqi Huang [a], Won Gyun No [b,*], Miklos A. Vasarhelyi [b], Zhaokai Yan [c]

[a] *Lubin School of Business, Pace University, One Pace Plaza, New York, NY 10038, USA*
[b] *Rutgers Business School, One Washington Park, Newark, NJ 07102, USA*
[c] *School of Management, Marist College, 3399 North Road, Poughkeepsie, NY 12601, USA*

Received 25 January 2022; revised 1 April 2022; accepted 8 May 2022
Available online 20 May 2022

Abstract

Emerging technologies like data analytics and machine learning are impacting the accounting profession. In particular, significant changes are anticipated in audit and assurance procedures because of those impacts. One such potential change is audit sampling. As audit sampling only provides a small snapshot of the entire population, it starts to lose some of its meaning in this big data era. One feasible solution is the usage of audit data analytics and machine learning to enable an analysis of the entire population rather than a sample of the transactions. This paper presents an approach for applying audit data analytics and machine learning to full population testing and discusses related challenges.

Keywords: Audit data analytics; Full population testing; Machine learning

1. Introduction

Technologies are evolving at an unprecedented pace and pose significant challenges and opportunities to companies and related parties, including the accounting profession. In today's business environment, it is inevitable for companies to react quickly to changing conditions and markets. Many companies are seeking better ways to utilize emerging technologies to transform how they conduct business. We live in an age of information explosion, with technologies capable of making revolutionary changes in various industries and reshaping business models. At present, many companies view data as one of their most valuable assets.[1] They amass an unprecedented amount of data from their daily business operation and strive to harness the power of data through analytics. Emerging technologies like robotic process automation, machine learning, and data analytics also impact the accounting profession. It is important for the profession to understand the impacts, opportunities, and challenges of these technologies.

Specifically, in audit and assurance areas, data analytics and machine learning will lead to many changes in the foreseeable future. Audit sampling is one such potential change. The use of sampling in audits has been criticized

* Corresponding author.
E-mail addresses: fhuang@pace.edu (F. Huang), wgno@business.rutgers.edu (W.G. No), miklosv@business.rutgers.edu (M.A. Vasarhelyi), zhaokai.yan@marist.edu (Z. Yan).
Peer review under responsibility of China Science Publishing & Media Ltd.

https://doi.org/10.1016/j.jfds.2022.05.002

since it only provides a small snapshot of the entire population. To address this major issue, this study introduces the idea of applying audit data analytics and machine learning for full population testing through the concept of “audit-by-exception” and “exceptional exceptions.” In this way, the emphasis of audit work shifts from “transaction examination” to “exception examination” and prioritizes the exceptions based on different criteria. Consequently, auditors can assess the associated risk based on the entire population of the transactions and thus enhance the effectiveness and efficiency of the audit process. Finally, this paper discusses potential challenges faced by full population testing.

2. Audit sampling issues

In 2020, Amazon delivered approximately 4.2 billion packages in the U.S., equivalent to more than 11 million packages per day.[d] These transactions comprise a significant portion of Amazon's revenue and serve as one of Amazon's biggest expenses, shipping costs. Now, imagine an audit team trying to perform a substantive test over the 4.2 billion packages delivered all over the U.S. Due to the large number of transactions, the most feasible way for performing the audit procedure would be to examine a sample of transactions (i.e., audit sampling). Auditing Standards 2315 defines that “[a]udit sampling is the application of an audit procedure to less than 100 percent of the items within an account balance or class of transactions for the purpose of evaluating some characteristic of the balance or class”.[2] Although audit sampling is well established in auditing practice and improves the efficiency of the audit, it has some limitations and thus starts to lose some of its meaning in this big data era. Sampling only provides a small snapshot of the entire population. It is hard to get a representative sample that accurately reflects the characteristics of the enormous population. Therefore, sampling risk arises from the possibility that auditors draw a different conclusion if the entire population is subject to the audit procedure rather than a sample.[2] In general, the smaller the sample size, the greater the sampling risk. The ratio between the sample size and the population size can be a proxy for sampling risk.

Despite the choice of sampling approaches, such as monetary unit sampling and judgmental sampling, the biggest challenge in this Amazon scenario is how to draw a sample that appropriately represents the characteristics of the entire 4.2 billion transactions. Auditors use sampling since examining the entire population would be too expensive and time-consuming. By nature, unusual or suspicious material transactions are rare. Therefore, the likelihood of identifying such transactions in a sample is low. Sampling is also ineffective in examining splitting large purchase orders across multiple purchase orders. Split purchase orders would only be observed if all purchase orders were randomly selected as part of the sample.

Consequently, there have been questions regarding the efficiency of audit sampling.[3,4] The fundamental question is whether a sample is a good representation of the entire population and sufficient enough to arrive at an adequate conclusion. In other words, is the audit evidence obtained from a trivial portion of the population reliable and sufficient? For instance, a sample of 500,000 transactions is only 0.012% of Amazon's entire population, even though it is a considerably large sample.

Due to the advance in information technologies that allows collecting, storing, and processing large volumes of data, transaction data in a client's system increases dramatically, making the size of samples that auditors examine in a substantive test trivial. Consequently, financial statement users expressed disappointment that the sample sizes were small.[5] Regulators often question the sample sizes used by auditors and identify deficiencies in projecting sample errors even though auditors generally use judgmental sampling or statistical sampling to assure an acceptably low sampling risk.[6] Moreover, according to a survey conducted by KPMG in 2017, 80% of the respondents believe that auditors should increase their sample sizes.[e] Nevertheless, increasing sample sizes means more transactions for auditors to examine and higher costs to the clients. Prior literature offers several suggestions for improving the effectiveness of audit sampling while not adding extra burdens to auditors and clients (e.g., Wurst et al[7]; Hoogduin et al[8]; Hoogduin et al[9]). One feasible solution is utilizing audit data analytics and machine learning to enable an analysis of the entire population rather than a sample of the transactions.[10]

[d] For more detail, visit http://news.pb.com/article_display.cfm?article_id=6005.

[e] For more detail, visit https://assets.kpmg/content/dam/kpmg/us/pdf/2017/03/us-audit-2025-final-report.pdf.

3. Audit data analytics and machine learning

Audit data analytics is defined as “the analysis of data underlying financial statements, together with related financial or non-financial information, for the purpose of identifying potential misstatements or risks of material misstatement”.[11] It encompasses traditional file interrogation, analytical procedures, and statistical-based analytics. However, auditors are often unfamiliar with many analytics methods.[12] A recent survey conducted by Deloitte also identifies that inadequate skills and knowledge are the key barriers to applying data analytics in auditing.[13] Although it requires additional knowledge and cost for initial implementation, audit data analytics can enhance audit efficiency and effectiveness. Cao et al[14] argue that audit data analytics can improve risk assessments about material misstatements, bankruptcy, and managerial fraud. Appelbaum et al[15] provide a summary of using data analytics in external audits and find that many analytic tools are widely applied to different phases of engagement, such as planning/risk assessment, substantive and compliance testing, and opinion formulation and reporting. For instance, auditors use regression techniques, descriptive statistics, and expert systems in the planning/risk assessment stage and ratio analysis and visualization in the opinion formulation and reporting stage. The application of audit data analytics also enhances the effectiveness of internal audits. The 2021 Deloitte survey identifies several key functions of data analytics used during an internal audit, such as audit planning, outlier identification, and audit reporting.[13]

As a subset of Artificial Intelligence (AI), machine learning is a computational program that automatically learns patterns and trends from historical data without being explicitly programmed by humans.[16,17] Unlike audit data analytics, which is primarily descriptive, machine learning involves more advanced predictive analytics.[18] There is a growing literature in accounting applying machine learning to learn data patterns and predict financial statement fraud and future earnings. For instance, Perols[19] and Bao et al[20] use multiple machine learning algorithms such as Support Vector Machines (SVMs), Artificial Neural networks (ANN), and decision trees to predict accounting frauds and irregularities. Hunt et al[21] find that random forest (a nonparametric machine learning technique) outperforms traditional regression in earnings prediction. In addition, machine learning techniques have been applied to audit procedures to support decision-making. For example, Halo, a machine learning platform developed by PwC, can read unstructured data in leases and contracts and identify potentially problematic journal entries with questionable keywords or from unauthorized sources.[22] Also, based on machine learning algorithms, Deloitte's Omnia DNAV establishes a new way of performing investment valuations and automating the audit of securities and investments.[23]

4. Full population testing

To eliminate sampling risk, the intuitive idea is to analyze the entire population rather than a sample of the transactions. In practice, however, it is likely quite challenging to audit an entire population due to monetary or time constraints. Nevertheless, the idea of full population testing has become feasible in audit practice because of exponentially faster computers, powerful audit data analytics, and machine learning applications. The concept of applying audit data analytics and machine learning for full population testing is called “audit-by-exception.” This concept is first developed by Vasarhelyi and Halper,[24] which discusses the design of an automatic and real-time audit system. They argue that instead of examining transactions at the end of the fiscal year, auditors should define a set of rules and use a system to monitor the violations of rules as each transaction occurs. Those rules are primarily designed based on internal controls and, more likely, knowledge engineered by experienced auditors. When a transaction violates one or more rules, the system treats it as an exception, generates an alert, and notifies auditors for further examination. Hence, the emphasis of audit work shifts from “transaction examination” to “exception examination.”

While performing “audit-by-exception” increases the effectiveness of the audit, it often generates a large number of exceptions. Issa[25] proposes an approach called “exceptional exceptions.” The essential point of performing full population testing through “audit-by-exception” is to recognize unusual or suspicious transactions by analyzing all transaction records based on criteria so that auditors can further investigate the identified exceptions. The criteria can be identified based on transaction types, business rules, and auditors' prior experiences and should tie in with the audit objectives that are guided by risk assessment and materiality.

Audit data analytics and machine learning could play a critical role in identifying abnormal transactions and their deviation level. The gist of “exceptional exceptions” is to prioritize the exceptions based on criteria such as the severity of the violations and the dollar amounts involved. By performing audit procedures on the prioritized exceptions, auditors can focus on unusual or suspicious transactions that are most likely to go wrong and cause the most significant impact on the client's financial statements or denote critical weaknesses in controls. The rest of the transactions, which are not selected for further investigation, are theoretically assumed to be less risky. In this way, auditors can assess the associated risk based on the entire population of the transactions and thus obtain more substantial audit evidence than traditional audit sampling. Compared with traditional audit sampling, the "exceptional exceptions" approach provides a higher level of assurance while auditors are performing audit procedures on the same number of transactions.[f] Fig. 1 provides an overall comparison between traditional sampling and full population testing.

The left-hand side of Fig. 1 shows the traditional sampling process. Auditors select a representative sample from the entire population based on different sampling methods (e.g., attribute or monetary sampling), appropriate sample size, and several other characteristics. Then, audit procedures are applied to the audit sample and help auditors conclude on the entire population. The right-hand side of Fig. 1 displays the full population testing approach, which includes an audit data selection and outlier prioritization process. Steps 1 and 2 show multistep outliers detections based on audit data analytics and machine learning techniques. In step 3, the identified violations will be ranked according to their risk levels.

Specifically, in step 1, there are various ways to identify exceptions (i.e., unusual or suspicious transactions). One approach is to use a set of filters. Auditors design multiple filters based on the client's business processes and internal controls and apply them to the entire transaction set. For instance, in an expenditure cycle (i.e., procure-to-pay process), if the client has an internal control requiring that the management must authorize any purchase exceeding $200,000, a filter can be created to identify purchases exceeding $200,000 without appropriate authorization. Similarly, additional

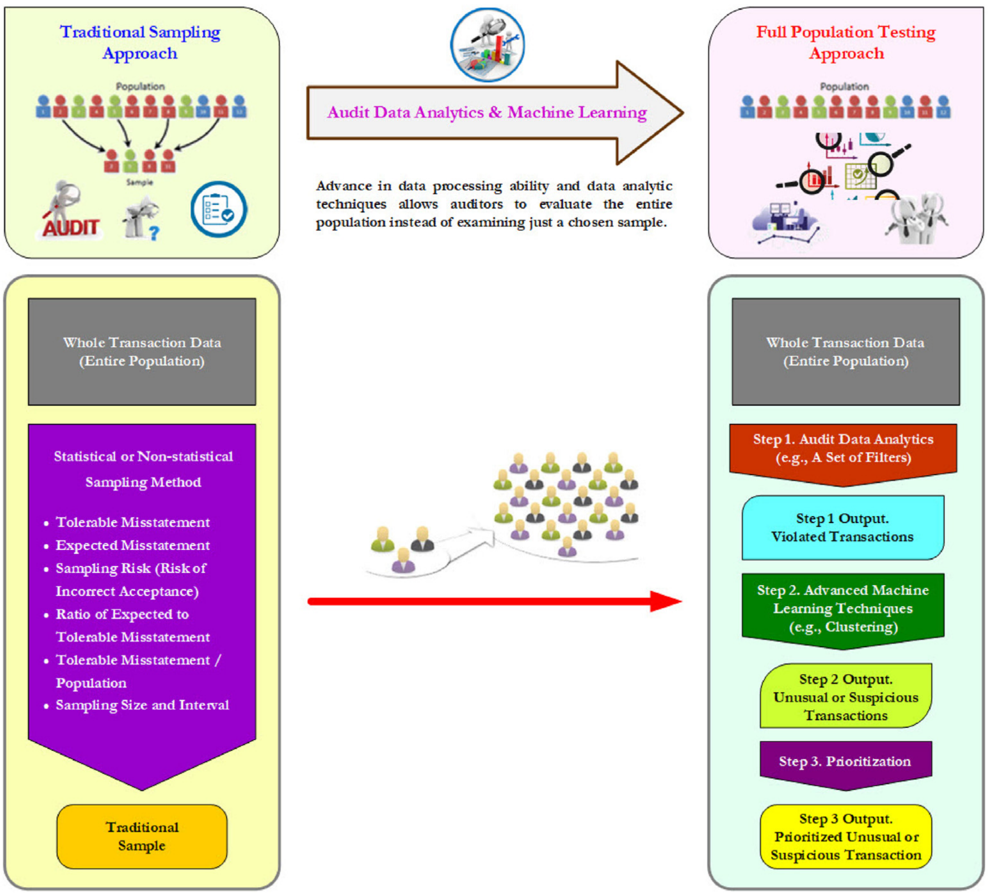

Fig. 1. Traditional audit sampling versus full population testing.

[f] The approach could be used throughout the audit processes such as substantive tests of details, tests of controls and internal audits.

filters can be created to identify purchases under the approval cut-off amount, purchases made with unusual order amount and quantity, and duplicate purchases. By applying these filters to the entire population of transactions, auditors identify a list of transactions that violate one or more filters (Step 1 Output in Fig. 1).

Based on the output of step 1, advanced machine learning algorithms can also be applied to identify unusual or suspicious transactions in step 2.[10] For example, clustering is an unsupervised machine learning approach[g] aiming to group transactions into a few clusters so that transactions within a group are similar while data across groups are different. As clustering is based on the similarity of data characteristics, a group with many transactions is considered normal, while a cluster with a relatively small number of transactions is viewed as a potential candidate group for outliers. Also, observations that are far away from the cluster centers or cannot be clustered into any groups may need further investigation, as they are significantly different from other transactions. By applying such advanced machine learning techniques to the step 1 output (i.e., violated transaction), auditors obtain a list of unusual or suspicious transactions (Step 2 Output in Fig. 1).

In the final step (i.e., prioritization), auditors then rank the identified transactions by utilizing prioritization methods such as calculating a suspicion score (i.e., a score representing the risky level) for each transaction. The suspicion score calculation is based on the auditor's professional judgment and could consider the number of violations triggered by the transaction and the significance of each violation. This prioritization step generates a list of prioritized unusual or suspicious transactions (Step 3 Output in Fig. 1). Lastly, auditors determine transactions for performing follow-up tests based on the list of prioritized unusual or suspicious transactions.

Although performing full population testing provides auditors with a higher level of assurance by allowing them to obtain a much-increased level of audit evidence, it is noteworthy that the approaches discussed in this paper are limited to data-oriented tests. Other tests that may involve more manual work, such as inquiry, confirmation, or inspection, may not be applicable for a full population test even though some tests can be performed using emerging technologies such as Robotic Process Automation for confirmation and Unmanned Aircraft Systems (i.e., drones) for inspection and observation.[26] Furthermore, there are several challenges to be addressed before the promised potential of full population testing can be realized.

5. Challenges

The initial implementation cost for full population testing can be high. Several advanced audit data analytics and machine learning algorithms involved in full population testing rely on advanced mathematical and statistical tools that require a high-level knowledge of the subject matter with difficult results to interpret. For instance, some powerful machine learning algorithms like Artificial Neural Network (ANN) and Vectors Support Machines (SVMs) are sometimes defined as “Black Box” since it is hard, if not impossible, to gain a good understanding of their inner working mechanisms. Therefore, the learning curve for auditors to master such methods would be steep. Potentially, “smart apps”[27] could be developed to supplement auditors' statistical and analytical competencies.

Documentation is another issue. When performing full population testing, auditors choose analytic techniques that will be applied to the entire population of transactions. Although various tools provide user-friendly interfaces to help auditors better implement and document audit data analytics, the choice of analytic techniques is often subjective. For instance, filters created to perform full population testing, for now, are often purely based on auditor judgment. Therefore, it is difficult for auditors to justify the use of analytic techniques and document the use in their working papers.

Furthermore, full population testing may impose an excessive liability burden on auditors. Currently, auditing standards state that auditors provide “reasonable assurance” that the financial statements are “fairly stated, in all material respects.” Potential errors, restricted by both occurrence probability and material level, are allowed to exist, and auditors are not held liable if they comply with auditing standards. However, performing full population testing would reduce the applicability of this “safe harbor” because the extent of audit procedures is expanded to the entire population of transactions to generate significantly strong audit evidence. Nevertheless, potential anomalies may still exist and be undiscovered even after applying audit data analytics since it is impractical for auditors to investigate every identified exception involved in unusual or suspicious transactions.

[g] Generally speaking, unsupervised machine learning refers to understanding what normally happens without pre-classified data.

Therefore, auditing standards are needed to clarify how audit data analytics fit into the current audit framework (i.e., risk assessment, tests of controls, substantive analytical procedures, and tests of details), to what extent each exception should be investigated, and auditors' responsibility if an entire population has been tested. Moreover, full population testing may cause a fundamental structural change in the current audit framework.

From a client's perspective, full population testing requires auditors to be provided with sufficient data, leading to several issues. Extracting a large amount of data from the client's system and making the data available for auditors will incur additional costs, especially for companies with large volumes of transactions. However, although well-implemented Enterprise Resource Planning (ERP) systems can significantly reduce the cost of data extraction, performing full population testing on extracted data certainly leads to lower costs than investigating and correcting data directly through the client's ERP systems. In addition, if the client uses multiple systems, data is often unstandardized and inconsistent. Preparing data in a format suitable for applying audit data analytics would require a significant amount of work from both auditors and the client.

Clients would initially resist, but once the routines are designed, the incremental cost of extraction is trivial. Imagine an extreme example such as Amazon. Obtaining and preparing 4.2 billion records ready for analysis would entail significant challenges to both auditors and database professionals. However, it is likely that some of these data already exist for operational purposes. Resistance may arise if the client does not see the benefits of full population testing due to the time and effort required for the initial retrieval and preparation of the data. In addition, the quality of the data stored in the client's database (i.e., data integrity issue) can also affect the result of full population testing. Furthermore, providing auditors access to the client's systems may lead to additional data security concerns. If IT security regulations are not properly followed, the risk of potential unauthorized access or breach of sensitive information increases.

There are also several challenges for standard setters. The American Institute of Certified Public Accountants (AICPA) acknowledges that applying data analytics and machine learning provides greater insights to auditors and improves audit efficiencies and quality. However, it also expresses challenges faced by applying AI or machine learning in audit practice.[28] For example, a machine cannot see the big picture because it is restricted only to the correct or incorrect data it has access to. Also, data analytic models or AI cannot consider moral or ethical concerns as auditors do. Therefore, audit standards usually lag in adopting emerging technologies due to those concerns or limitations. While the current auditing standards do not prohibit auditors from using audit data analytics and machine learning techniques, the standard setters do not provide specific standards or guidance on applying them for full population testing. Consequently, it is challenging for auditors to perform full population testing. Auditors are less likely to perform full population testing without appropriate auditing standards and guidance in their audits.[h] Furthermore, regulators need to weigh the balance between small and large audit firms since an expensive initial cost is unavoidable. While large audit firms and their clients can absorb the initial cost of full population testing and enjoy the benefits, small audit firms may find it too costly to invest in audit data analytics.

6. Conclusions

The evolution in information technologies and business models and the dramatic expansion of population sizes have diminished the quality of audit evidence obtained through audit sampling. With the development in audit data analytics and machine learning, full population testing is now feasible and can serve as the potential solution for addressing issues regarding audit sampling. Full population testing is performed through the concept of “audit-by-exception” and “exceptional exceptions,” by which auditors use audit data analytics to identify and prioritize unusual or suspicious transactions. Using the prioritized unusual or suspicious transactions, auditors can enhance the effectiveness and efficiency of the audit process.

Several challenges need to be addressed to realize the promised potential of full population testing. First, from the perspective of cost, the initial implementation cost for full population testing can be high, and the learning curve for auditors to master such methods would be steep. Second, from the client's perspective, full population testing requires auditors to be provided with sufficient data, leading to an increased burden from the client for data cleaning and preparation. Finally, from the regulation perspective, auditing standards are needed to clarify how audit data analytics

[h] The Rutgers AICPA Data Analytics Research (RADAR) initiative proposes a Multidimensional Audit Data Selection (MADS) framework to guide auditors when performing a full population analysis. For more details, visit http://raw.rutgers.edu/radar.html.

and machine learning fit into the current audit framework, to what extent each exception should be investigated, and the auditors' responsibility if a full population test is applied.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Parkins D. The world's most valuable resource is no longer oil, but data. *Economist*; 2017. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.
2. Public Company Accounting Oversight Board (PCAOB). *Auditing Standards 2315: Audit Sampling*. 2014.
3. Kachelmeier SJ, Messier Jr WF. An investigation of the influence of a nonstatistical decision aid on auditor sample size decisions. *Account Rev*. 1990:209–226.
4. Power MK. Auditing and the production of legitimacy. *Account Org Soc*. 2003;28(4):379–394.
5. Gray GL, Turner JL, Coram PJ, Mock TJ. Perceptions and misperceptions regarding the unqualified auditor's report by financial statement preparers, users, and auditors. *Account Horiz*. 2011;25(4):659–684.
6. Public Company Accounting Oversight Board (PCAOB). *Report on the PCAOB's 2004, 2005, 2006, and 2007 Inspections of Domestic Annually Inspected Firms*. 2008.
7. Wurst J, Neter J, Godfrey J. Effectiveness of rectification in audit sampling. *Account Rev*. 1991:333–346.
8. Hoogduin LA, Hall TW, Tsay JJ. Modified sieve sampling: a method for single-and multi-stage probability-proportional-to-size sampling. *Auditing-J Pract Th*. 2010;29(1):125–148.
9. Hoogduin LA, Hall TW, Tsay JJ, Pierce BJ. Does systematic selection lead to unreliable risk assessments in monetary-unit sampling applications? *Auditing-J Pract Th*. 2015;34(4):85–107.
10. No WG, Lee K, Huang F, Li Q. Multidimensional audit data selection (MADS): a framework for using data analytics in the audit data selection process. *Account Horiz*. 2019;33(3):127–140.
11. Stewart T. *Data analytics for financial-statement Audits*. In: *Chapter 5 in AICPA, Audit Analytics and Continuous Audit: Looking toward the Future*. American Institute of Certified Public Accountants; 2015.
12. Appelbaum D. *Public Auditing, Analytics, and Big Data in the Modern Economy*. Rutgers University-Graduate School-Newark; 2017.
13. Deloitte. *Internal Audit Digital and Analytics Survey 2021*. 2021.
14. Cao M, Chychyla R, Stewart T. Big data analytics in financial statement audits. *Account Horiz*. 2015;29(2):423–429.
15. Appelbaum DA, Kogan A, Vasarhelyi MA. Analytical procedures in external auditing: a comprehensive literature survey and framework for external audit analytics. *J Account Lit*. 2018;40:83–101.
16. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev*. 1959;3(3):210–229.
17. Mitchell TM. Does machine learning really work? *AI Mag*. 1997;18(3):11.
18. Reavie V. *Do You Know the Difference between Data Analytics and AI Machine Learning?*. 2018.
19. Perols J. Financial statement fraud detection: an analysis of statistical and machine learning algorithms. *Auditing-J Pract Th*. 2011;30(2):19–50.
20. Bao Y, Ke B, Li B, Yu YJ, Zhang J. Detecting accounting fraud in publicly traded US firms using a machine learning approach. *J Account Res*. 2020;58(1):199–235.
21. Hunt JO, Myers JN, Myers LA. Improving earnings predictions and abnormal returns with machine learning. *Account Horiz*. 2022;36(1):131–149.
22. Kokina J, Davenport TH. The emergence of artificial intelligence: how automation is changing auditing. *J Emerg Technol Account*. 2017;14(1):115–122.
23. Deloitte. *Deloitte Wins 2020 'Audit Innovation of the Year' at the Digital Accountancy Forum & Awards*. 2020.
24. Vasarhelyi MA, Halper FB. The continuous audit of online systems. *Auditing-J Pract Th*. 1991;10(1):110–125.
25. Issa H. *Exceptional Exceptions*. Rutgers University-Graduate School-Newark; 2013.
26. Appelbaum D, Nehmer RA. Using drones in internal and external audits: an exploratory framework. *J Emerg Technol Account*. 2017;14(1):99–113.
27. Byrnes P, Brennan CG, Vasarhelyi M, Moon D, Ghosh S. *Managing risk and the audit process in a world of instantaneous change*. In: *Chapter 6 in AICPA, Audit Analytics and Continuous Audit: Looking toward the Future*. 2015.
28. American Institute of Certified Public Accountants (AICPA). *The Data-Driven Audit: How Automation and AI are Changing the Audit and the Role of the Auditor*. 2020.