

A Review of Data-Driven Risk Evaluation Models for Emerging Market Financial Institutions

OLAWOLE AKOMOLAFE¹, MICHAEL UZOMA AGU²

¹Nigeria Liability Insurance Pool, Lagos, Nigeria

²Shell Petroleum Development Company of Nigeria Limited

Abstract- *Emerging market financial institutions face heightened levels of uncertainty due to volatile macroeconomic conditions, weak regulatory environments, data scarcity, and structural market imperfections. Advances in data-driven modelling spanning statistical learning, machine learning, early warning systems, credit scoring, stress-testing frameworks, and macro-prudential forecasting offer significant potential to strengthen risk evaluation and supervisory oversight. This paper provides a comprehensive review of data-driven risk evaluation models applicable to banks, microfinance institutions, insurance firms, and capital-market intermediaries operating in emerging economies. Drawing solely on pre-2019 scholarship, the review evaluates the evolution, methodological foundations, strengths, and limitations of major quantitative approaches used in credit risk assessment, liquidity risk prediction, operational risk detection, and systemic vulnerability monitoring. The article identifies persistent challenges such as data quality issues, modelling instability, weak governance, and context adaptation gaps. A set of implications for institutional practice and future research is outlined.*

Keywords: *Data-Driven Modelling, Emerging Markets, Financial Institutions, Risk Evaluation, Machine Learning, Early Warning Systems.*

I. INTRODUCTION

Emerging market financial institutions operate within environments characterised by structural volatility, macroeconomic fragility, and substantial information asymmetries [1], [2], [3]. Unlike advanced economies where financial systems tend to exhibit high levels of formalisation, transparency, and regulatory maturity, emerging markets are frequently shaped by fragmented supervisory environments, limited institutional capacity, political uncertainty, and

uneven financial sector development [4], [5]. These conditions contribute to unique patterns of risk exposure for local banks, microfinance institutions, insurers, and capital-market intermediaries [6], [7]. As such, the effectiveness of risk evaluation practices becomes a foundational determinant of institutional resilience, investor confidence, and financial system stability [8], [9]. The last two decades have witnessed substantial interest in developing data-driven risk evaluation models capable of adapting to these environments, utilising advances in statistical learning, artificial intelligence, and computational techniques to augment traditional supervisory assessments [10], [11].

The adoption of data-driven approaches in risk modelling was historically constrained by limited data availability, poor data quality, and the absence of reliable digital infrastructures. Prior to the mid-2000s, many financial institutions in emerging markets relied on manual risk evaluation techniques with limited predictive ability. These included qualitative risk matrices, basic scoring systems, and supervisory checklists which, although useful for routine assessment, lacked the capacity to forecast extreme events or detect latent vulnerabilities [12], [13]. The weaknesses of such approaches were exposed during multiple financial disturbances affecting emerging economies, including currency crises, banking failures, commodity price shocks, sovereign debt defaults, and procyclical capital flows [14]. These events reinforced the need for methodologies capable of processing large data volumes, capturing nonlinear relationships, and providing early warnings of impending instability.

Data-driven modelling gained momentum as emerging markets underwent digital transformation in banking operations, capital markets, and regulatory reporting [15], [16]. The diffusion of electronic payments, mobile banking platforms, and credit information

infrastructures created unprecedented opportunities to collect granular transaction data, customer behaviour patterns, and real-time market indicators [17]. At the same time, advancements in computational power made it increasingly feasible to operationalise sophisticated modelling algorithms previously confined to academic research or advanced economy institutions [18], [19]. By 2018, a wide range of statistical and machine learning models had been adapted or proposed for emerging market risk evaluation, including logistic regression, survival analysis, panel-based early warning systems, support vector machines (SVMs), random forests, gradient boosting machines, Bayesian networks, neural networks, and hybrid ensemble approaches [20], [21].

However, emerging market conditions introduce several modelling challenges that differ from those encountered in advanced economies. These challenges include non-stationary macroeconomic environments, volatile exchange rate regimes, incomplete reporting, informal sector dominance, and limited availability of long-term historical data [22]. Such conditions frequently undermine model accuracy, impair feature extraction, and complicate parameter estimation [23], [24]. Furthermore, structural breaks such as sudden policy shifts, political disturbances, or commodity price fluctuations introduce significant instability into model performance over time [25], [26], [27]. As a result, risk evaluation models that perform well during stable periods may deteriorate rapidly during crises, rendering them unreliable for early warning or supervisory intervention.

Due to these complexities, emerging market research emphasises the integration of macro-financial variables, institutional factors, and market microstructure indicators into risk evaluation algorithms. For instance, credit risk models increasingly incorporate borrower demographics, transactional histories, behavioural indicators, and network relationships, rather than relying solely on financial statements which are often incomplete or outdated [28], [29]. Similarly, liquidity risk models utilise payment flows, interbank rate spreads, and real-time settlement data to predict liquidity shortages. Systemic risk evaluation frameworks integrate banking interlinkages, contagion channels, cross-holding patterns, and sectoral concentration indicators

[30], [31]. In insurance markets, advancements in nonlife claims modelling, lapse prediction, and reserving analytics have strengthened actuarial risk assessment, particularly where regulatory frameworks remain underdeveloped [32], [33].

Despite these advances, the adoption of data-driven modelling across emerging market institutions remains uneven. Many banks and insurers lack the technical capacity, data governance systems, or computational resources necessary to implement advanced analytics effectively [34]. Model risk stemming from misspecification, overfitting, or misinterpretation poses substantial regulatory concern, especially where governance oversight is weak [35]. Supervisory agencies in emerging economies have struggled to create regulatory guidelines for advanced modelling, resulting in fragmented or inconsistent practices. Moreover, cultural resistance and institutional inertia hinder transition from traditional risk evaluation paradigms to data-oriented models.

The global financial crisis of 2008 highlighted the limitations of traditional and data-driven models alike. In advanced economies, several high-profile risk models failed to capture systemic vulnerabilities due to improper assumptions, historical bias, or structural blind spots [36], [37]. For emerging markets, the crisis exposed gaps in supervisory reporting systems, market surveillance frameworks, and stress-testing practices. In response, international organisations including financial stability councils, multilateral development banks, and regional regulatory associations promoted adoption of early warning systems and macro-prudential modelling frameworks tailored to local conditions [38], [39]. These initiatives accelerated experimentation with data-driven methods, fostering hybrid approaches that combine machine learning with expert judgment, scenario simulation, and qualitative supervision [40], [41].

In microfinance and small-business lending, data-driven models such as mobile credit scoring, psychometric assessment tools, and transaction-based scoring systems have gained prominence in several emerging regions. These models mitigate information asymmetry by analysing behavioural data, digital footprints, consumption patterns, and mobile platform

activity [42], [43]. Although these approaches hold significant promise for financial inclusion, their adoption raises concerns related to data privacy, algorithmic bias, and model transparency. Similar tensions arise in insurance markets, where predictive analytics can improve risk classification but may amplify inequities or violate regulatory standards on consumer fairness [44], [45].

The technological revolution occurring across global financial systems by 2018 created fertile ground for the expansion of data-driven risk evaluation in emerging markets. Distributed computing, cloud-based analytics, and real-time data streaming facilitated model deployment at scale, particularly for larger banks and insurers. Advances in natural language processing (NLP) enabled extraction of information from financial news, social media, or regulatory disclosures, broadening the range of risk indicators available to analysts [46]. Graph-based approaches and network analytics helped identify contagion pathways between financial institutions. Meanwhile, artificial intelligence techniques though used cautiously showed potential in forecasting credit events, detecting anomalies, and analysing systemic linkages [47].

Yet the adoption of sophisticated techniques does not eliminate fundamental constraints. Data scarcity remains pervasive, especially among small institutions operating in low-income regions [48], [49]. In many cases, macroeconomic volatility undermines the stability of predictive relationships, requiring frequent recalibration or adaptive learning mechanisms [50], [51]. Regulatory capacity in emerging markets varies widely, making it difficult to standardise model approval processes, establish validation frameworks, or harmonise disclosure requirements. The cultural and organisational transformation necessary for widespread adoption of data-driven risk evaluation remains incomplete [52], [53].

These factors underscore the need for a comprehensive review of data-driven risk evaluation models applicable to emerging market financial institutions [54], [55]. Existing research is fragmented across disciplines such as econometrics, financial engineering, actuarial science, machine learning, regulatory economics, and development finance [56],

[57]. A unified review is therefore essential for synthesising methodological advances, identifying cross-cutting themes, and highlighting gaps in the literature [58], [59]. Such a review allows researchers and practitioners to assess the suitability of various models for emerging market conditions and inform future directions in financial risk management [60], [61].

The objectives of this review are threefold. First, the paper examines the evolution of data-driven risk evaluation methodologies prior to 2019, focusing on statistical, machine learning, and hybrid frameworks. Second, it reviews empirical applications of these models in emerging market banks, insurers, MFIs, and financial regulators. Third, it synthesises methodological challenges and contextual factors affecting model performance, providing insights for future research and institutional practice.

The remainder of the paper is structured as follows. Section 2 presents a detailed review of the literature on data-driven risk evaluation, covering credit, liquidity, operational, market, and systemic risk models. Section 3 synthesises cross-cutting insights, identifies research gaps, and highlights policy implications. Section 4 concludes the paper.

II. LITERATURE REVIEW

Data-driven risk evaluation has emerged as one of the most transformative themes in financial sector modelling, particularly within emerging markets where institutions face persistent uncertainty, limited data quality, and unstable macroeconomic conditions. Prior to 2019, scholarly and applied research advanced a wide spectrum of quantitative models for credit, liquidity, operational, insurance, and systemic risk assessment. This literature review synthesises the evolution, methodological foundations, and empirical applications of these models across emerging market financial institutions.

A major strand of early work drew from classical econometric modelling, where risk evaluation relied heavily on linear relationships, structured probability models, and statistically tractable assumptions. Logistic regression became widely used for credit default prediction due to its simplicity, interpretability, and computational efficiency [62], [63]. Emerging

market banks used logistic models to evaluate borrower characteristics, financial ratios, repayment histories, and macroeconomic indicators. While logistic regression provided baseline predictive capacity, its linearity often limited accuracy in environments where borrower behaviour exhibited nonlinear and volatile dynamics [64], [65]. Probit models represented an alternative but similarly encountered challenges with structural breaks and data discontinuities characteristic of developing economies [66], [67].

Survival analysis also gained traction for predicting time-to-default and loan deterioration events. Cox proportional hazard models were applied to microfinance portfolios, SME lending, and consumer credit lines in several emerging markets [68], [69]. These models offered insight into temporal patterns of borrower vulnerability but were sensitive to the proportionality assumption, which often failed during crisis periods or high-inflation episodes [70], [71]. Meanwhile, panel-based econometric frameworks allowed integration of cross-sectional and time-series data, supporting early-warning indicators for banking crises [72], [73]. These models incorporated variables such as credit growth, nonperforming loan ratios, exchange rate pressures, foreign reserve trends, and capital adequacy measures. Despite empirical usefulness, they were hampered by sparse data and rigid specifications.

Machine learning began gaining recognition for its flexibility and ability to capture nonlinear patterns in financial risk prediction. Support Vector Machines (SVMs) were among the earliest machine learning models applied to credit risk detection in emerging market institutions [74], [75]. Their margin-maximisation approach and capacity to handle sparse or high-dimensional data made them favourable for banking datasets with inconsistent reporting. However, SVMs required significant parameter tuning and lacked intuitive interpretability, limiting supervisory acceptance in less mature regulatory environments. Decision tree algorithms and random forests emerged as powerful alternatives capable of capturing interactions between financial ratios, behavioural variables, and external factors [76]. Random forests demonstrated robustness to noisy data, missing observations, and outlier's conditions

common in emerging markets but their ensemble nature complicated transparency [77].

Artificial neural networks (ANNs), particularly multilayer perceptron's, were widely applied to credit scoring, loan-approval systems, and SME financing models in developing countries [78]. ANNs excelled at modelling complex risk relationships, but they required extensive training data, making them difficult to apply in markets with small datasets or inconsistent data generation processes [79]. Hybrid models combining ANNs with genetic algorithms, fuzzy logic systems, or Bayesian optimisation were also explored [80], [81]. These models exhibited high predictive performance yet remained computationally intensive and difficult to integrate into conventional risk governance frameworks [82], [83].

Advancements in behavioural data collection enabled machine-learning-driven credit scoring in microfinance and financial inclusion initiatives. Research demonstrated that mobile-money transaction flows, telecom usage patterns, and digital wallet activity could serve as proxies for customer creditworthiness [84], [85]. Psychometric scoring systems introduced additional layers of behavioural prediction, analysing borrower traits, cognitive responses, and risk-taking preferences. By 2018, these systems were deployed across multiple African, Asian, and Latin American MFIs. However, concerns persisted regarding algorithmic fairness, data privacy, and the risk of reinforcing social biases [86].

Market-based data-driven approaches contributed significantly to liquidity and market-risk evaluation. Early work applied GARCH-type models to forecast volatility in emerging market bonds, equities, and currency markets [87], [88]. While powerful for short-term volatility modelling, GARCH frameworks struggled with sudden regime shifts, a recurrent phenomenon in volatile developing markets. More advanced stochastic volatility models, including Markov-switching GARCH variants, improved adaptability but demanded large datasets often unavailable locally [89], [90]. Machine learning models such as gradient boosting machines (GBM) and elastic-net regularisation expanded modelling capacity by integrating market microstructure variables such as order-book imbalances, bid-ask

spread dynamics, and sentiment indicators extracted from financial news [91], [92]. Nevertheless, market fragmentation and liquidity constraints in many emerging economies limited the reliability of such indicators.

Operational risk modelling evolved from basic loss-distribution approaches toward analytics-enabled frameworks capable of detecting fraudulent transactions, cyber intrusions, internal process failures, and human-error-induced incidents. Pre-2018 research on anomaly detection used unsupervised learning, clustering algorithms, and outlier-based scoring to identify irregular patterns in financial transactions [93], [94]. For instance, k-means clustering and density-based approaches were adopted for fraud detection in electronic payment systems prevalent in emerging markets [95]. Although useful, false-positive rates remained high due to unstable behavioural patterns and limited labelled fraud data.

Systemic-risk modelling became increasingly relevant after the 2008 financial crisis. Network theory emerged as a dominant analytical paradigm for assessing interbank linkages, contagion channels, cross-holding exposures, and financial interconnectedness in emerging markets [96], [97]. Graph-based models measured centrality, clustering, and vulnerability across banking networks, yielding insight into systemic fragility. Stress-testing frameworks incorporated network spillovers, enabling regulators to evaluate the resilience of financial institutions under macroeconomic shocks [98]. Furthermore, agent-based models were explored to simulate heterogeneous financial behaviours and institutional feedback loops. Despite conceptual richness, data limitations impeded widespread adoption.

Insurance-sector risk evaluation models also experienced data-driven transformation. Nonlife actuarial risk modelling employed generalized linear models (GLMs), generalized additive models (GAMs), and machine learning methods such as random forests and boosted trees for claims prediction, reserve estimation, and fraud detection [99]. Life insurers experimented with lapse modelling, mortality forecasting, and policy-holder behaviour prediction using survival models and ANNs [100]. Emerging

markets posed challenges such as low insurance penetration, sparse historical data, and limited actuarial expertise. Nonetheless, advances in mobile-based micro-insurance platforms yielded new data sources for predictive analytics [101], [102].

Macroeconomic early-warning systems (EWS) constituted another major research domain. EWS models typically integrated macro-financial variables such as interest-rate spreads, credit-to-GDP gaps, foreign reserve fluctuations, asset-price misalignments, banking profitability indicators, and external vulnerability indices [103], [104]. Composite indicator frameworks and signal extraction models were applied across Asia, Africa, and Latin America to anticipate currency crises, banking distress, and sovereign debt defaults [105]. Machine learning enhancements improved predictive capability by capturing nonlinear interactions between global commodity prices, capital flows, and domestic credit cycles [106]. However, emerging markets' exposure to external shocks often reduced model robustness.

Big-data analytics played a growing role in data-driven risk evaluation, supported by increased adoption of cloud computing, distributed databases, and digital financial services infrastructure. Techniques such as text mining, natural language processing (NLP), and social media analysis were used to assess investor sentiment, detect reputational risk, and forecast market reactions [107]. Mobile-network metadata, satellite imagery, and alternative data sources enhanced risk intelligence for agricultural lending, disaster insurance, and supply-chain-linked financing [108]. Constraints remained in model validation, regulatory guidance, and data governance frameworks.

Another influential stream examined the governance dimensions of data-driven modelling. Scholars highlighted that risk model performance was highly dependent on institutional capacity, board oversight, and data governance quality [109]. Weak governance contributed to model misuse, poor calibration, and overreliance on automated systems without adequate human judgment. Regulatory agencies in emerging markets struggled to evaluate model risk, enforce validation standards, and monitor algorithmic opacity in machine-learning-based tools [110]. This

necessitated hybrid supervisory frameworks combining quantitative risk scores with qualitative assessments.

The literature also emphasised the distinction between micro-level risk evaluation at the institutional level and macro-level systemic-risk evaluation. Micro-level models focused on borrower creditworthiness, institutional liquidity, capital adequacy, operational vulnerabilities, and insurance risk exposure. Macro-level models addressed cross-institution linkages, asset–liability mismatches, contagion pathways, and global spillover effects. Many scholars argued for integrated modelling approaches combining these perspectives, although implementation remained limited in emerging markets due to data fragmentation [111].

In summary, the literature demonstrates substantial progress in applying data-driven methods to emerging market risk evaluation across banks, insurers, MFIs, and capital-market intermediaries. However, pre-2019 research consistently notes that modelling constraints including weak data infrastructure, macroeconomic instability, regulatory gaps, and cultural resistance continue to hinder widespread adoption. The review shows that while machine learning and hybrid approaches significantly enhance predictive accuracy, practical implementation must address governance, transparency, and data-quality challenges. These findings underscore the need for risk modelling frameworks that balance predictive performance with interpretability, regulatory compatibility, and contextual adaptability.

III. DISCUSSION

The evolution of data-driven risk evaluation models reflects an ongoing transformation in how financial institutions particularly those operating in emerging markets conceptualise, measure, and respond to financial vulnerabilities. The literature reviewed demonstrates that emerging market conditions fundamentally shape the types of models that can be effectively deployed, the quality and quantity of data available, and the institutional capacity required to sustain sophisticated analytical systems. This discussion synthesises the key insights drawn from the preceding review and critically reflects on the

implications for practitioners, policymakers, and researchers.

A central theme emerging from the literature is that data-driven models have significantly enhanced predictive capability compared to traditional rule-based or judgment-based methods, yet they remain constrained by contextual limitations inherent in emerging economies. Statistical approaches such as logistic regression, probit models, and survival analysis provided foundational analytic structures for early modelling efforts. Their appeal lay in simplicity, transparency, ease of interpretation, and low data requirements. However, these methods were frequently insufficient for capturing nonlinear patterns, behavioural complexities, or rapid structural changes in macroeconomic environments. Emerging markets are prone to volatile inflation rates, political instability, commodity price shocks, and irregular credit cycles, which reduce the stability of traditional econometric models. As a result, their predictive accuracy deteriorates during periods of stress precisely when institutions require reliable assessments of risk.

The rise of machine learning methods offered solutions to several of these challenges. Models such as random forests, support vector machines, neural networks, and gradient boosting algorithms gained prominence due to their ability to capture nonlinear interactions and process high-dimensional data [112]. Their adoption in credit scoring, fraud detection, claims forecasting, and market-risk prediction represented a substantial advancement in modelling sophistication. Yet machine learning models introduce a critical trade-off between predictive power and interpretability. In emerging market regulatory environments often characterised by limited technical expertise and underdeveloped supervisory frameworks this opacity can limit institutional adoption. Regulators may hesitate to approve models whose internal mechanics are difficult to explain or audit. This tension highlights the continued relevance of hybrid modelling approaches that combine machine learning techniques with transparent statistical components or expert judgment frameworks.

An equally important finding is that emerging market financial institutions face persistent data-related constraints that affect model performance. Weak

reporting infrastructures, incomplete borrower histories, and inconsistent regulatory filings undermine model calibration and back-testing. For smaller institutions, particularly MFIs and community-level lenders, limited digitisation further restricts access to structured datasets. In these contexts, mobile-money transaction data, telecom metadata, psychometric assessments, and alternative data sources have emerged as partial solutions. These new data streams compensate for absent financial records, although they generate new challenges relating to privacy, algorithmic bias, and data ownership. Ethical considerations surrounding the use of behavioural and psychological data for credit scoring remain an area of active debate and require careful regulatory oversight.

Liquidity, market, operational, and systemic risk modelling in emerging markets also face significant methodological challenges. Market-based models such as GARCH variants, stochastic volatility models, and sentiment-driven analytics perform well in liquid markets but struggle in fragmented or thinly traded environments [113]. Liquidity constraints amplify price distortions, undermine assumptions of market efficiency, and reduce reliability of market microstructure indicators. Operational risk models, particularly anomaly detection and clustering algorithms, are sensitive to high noise levels and irregular behavioural patterns. Fraud patterns in emerging markets evolve rapidly, requiring continuous model retraining and adaptation. Systemic risk modelling, while theoretically powerful, is hindered by limited availability of interbank exposure data and the absence of centralised data repositories in many jurisdictions. These limitations underscore the importance of developing hybrid and adaptive models capable of functioning effectively within imperfect information environments.

Insurance-sector risk evaluation models face additional constraints distinct from the banking sector. Sparse historical claims data, low insurance penetration, high rates of policy lapsation, and limited actuarial capacity hinder predictive analytics in many emerging economies. While GLMs, GAMs, and random forests have strengthened nonlife actuarial modelling their performance is constrained by inconsistent reporting and limited digitalisation of

claims workflows. Life insurance modelling for mortality and lapse prediction benefits from survival analysis and neural networks, but these require substantial data not always available. Micro-insurance platforms have improved data collection and risk pooling but remain vulnerable to idiosyncratic claims shocks, natural disasters, and climate-related events [114].

A recurring insight across the literature is that institutional capacity including governance structures, risk culture, and technical expertise plays a decisive role in determining the success of data-driven risk evaluation models. Weak governance contributes to model misuse, insufficient oversight, poor validation practices, and overreliance on automated outputs without proper contextual interpretation [115]. Many emerging market regulators lack the capacity to independently assess sophisticated machine learning models, complicating the approval process for advanced algorithms. The absence of established model risk management frameworks increases vulnerability to overfitting, calibration drift, and misspecification. These governance gaps illustrate that improvements in modelling must be accompanied by parallel investments in supervisory capability, institutional training, and data governance.

The discussion further reveals that macro-level and micro-level risk modelling approaches have often evolved independently. Micro-level models address borrower repayment risk, institutional liquidity pressures, operational inefficiencies, and internal vulnerabilities, whereas macro-level models examine broader systemic linkages, capital flow dynamics, and aggregate financial stability. The reviewed literature emphasises the potential for integrated modelling frameworks that combine institution-specific data with macro-financial indicators. Despite conceptual appeal, practical implementation remains limited due to data fragmentation and institutional silos. Nevertheless, the growing availability of digital financial data, alternative datasets, and cloud-based computing environments indicates potential for future integration.

Another important implication is that emerging market financial institutions require risk evaluation models that remain robust under regime shifts, political instability, and externally induced shocks. Traditional

models assume relatively stable structural relationships, but emerging markets experience frequent and abrupt changes in regulatory policies, exchange rate regimes, and fiscal environments. These dynamics often necessitate frequent recalibration and stress testing. Techniques such as Markov-switching approaches, adaptive learning algorithms, and scenario-based evaluation frameworks can partially address these challenges, but they remain underutilised due to technical complexity and limited data availability.

Despite methodological advances, it is evident that data-driven models must complement not replace human judgment in risk evaluation. Model outputs require careful interpretation within local contexts, integrating institutional experience, market knowledge, and qualitative insights. Hybrid human-machine systems, where risk analysts critically evaluate model outcomes and adjust for contextual nuances, represent a pragmatic approach for emerging market institutions. This aligns with governance frameworks emphasising accountability, model validation, and supervisory review.

In conclusion, the literature indicates that data-driven risk evaluation models offer substantial promise for enhancing predictability, improving institutional decision-making, and supporting financial stability in emerging markets. However, effective implementation requires attention to contextual factors such as data quality, regulatory capacity, institutional governance, and market structure. Robust model risk management, ethical data usage, and integration of human judgment remain essential. The insights from this literature review form the basis for refining risk evaluation frameworks and guiding future research in emerging market financial analytics.

IV. CONCLUSION

The purpose of this review was to synthesise the development and application of data-driven risk evaluation models for emerging market financial institutions prior to 2019. The analysis shows that while substantial progress has been made in adopting quantitative, computational, and machine-learning-based risk modelling techniques, the effective translation of these innovations into emerging market contexts remains uneven. The key challenge lies not in

the absence of methodological sophistication but in navigating the structural, institutional, and data-related limitations that shape financial systems in developing regions.

A central insight emerging from the literature is that traditional econometric models such as logistic regression, probit analysis, hazard models, and early-warning panel frameworks continue to provide an important baseline for risk assessment due to their interpretability, transparency, and relative simplicity. However, their performance is constrained in environments marked by macroeconomic instability, structural breaks, and incomplete data. Machine learning models, including SVMs, random forests, neural networks, and ensemble algorithms, offer substantial predictive advantages by capturing nonlinearities and processing high-dimensional inputs. Yet these advantages come at the cost of interpretability and governance complexity, posing challenges for regulatory acceptance and model risk management in emerging markets.

The review further demonstrates that the contextual realities of emerging markets such as fragmented datasets, limited historical depth, inconsistent regulatory reporting, and the prevalence of informal-sector financial activity significantly influence model applicability. Data-driven models must therefore be adapted to accommodate data sparsity, high noise levels, and sudden regime shifts. Methods relying on large, stable datasets frequently underperform, while those designed to operate under uncertainty and structural volatility demonstrate greater resilience. As financial digitalisation accelerates, new data sources such as mobile-money platforms, telecom metadata, psychometric assessments, and alternative datasets provide opportunities to strengthen risk evaluation, though they introduce ethical, regulatory, and governance considerations that must be addressed carefully.

The literature also highlights the importance of institutional capacity in shaping model performance. Strong governance structures, robust data management systems, and skilled analytical personnel play decisive roles in determining whether sophisticated models can be implemented effectively. Weak regulatory oversight, insufficient validation

frameworks, and limited technical expertise generate risk of model misuse or misinterpretation, particularly when deploying black-box machine learning systems. In this context, hybrid modelling approaches combining statistical modelling, machine learning, and expert judgment emerge as a practical strategy for balancing predictive accuracy with interpretability and regulatory compliance.

Another key conclusion of the review is that micro-level and macro-level risk assessment approaches are often treated separately, limiting their effectiveness in identifying system-wide vulnerabilities. While micro-level models provide insights into borrower risk, institutional liquidity, and operational vulnerabilities, macro-level models emphasise systemic risk propagation, contagion pathways, and financial stability indicators. Integrating these perspectives remains a major research and policy challenge but offers substantial potential for improving early-warning capabilities in emerging market financial systems.

Overall, this review underscores that data-driven risk evaluation models offer transformative potential for enhancing risk management practices, supervisory oversight, and financial stability in emerging markets. Their successful adoption requires contextual adaptation, strong governance, ethical safeguards, and regulatory frameworks capable of managing model risk. Future research must explore robust methods for modelling under structural uncertainty, integrating heterogeneous data sources, and combining machine learning with domain expertise. Practitioners and policymakers should prioritise capacity development, data infrastructure investment, and the establishment of model validation standards suited to emerging market realities.

In summary, data-driven risk evaluation represents both an opportunity and a challenge for emerging market financial institutions. When appropriately designed and implemented, these models can strengthen financial resilience, improve resource allocation, and support inclusive growth. However, their potential will only be realised through sustained investment in data quality, institutional capability, and regulatory alignment. This review contributes to ongoing academic and policy discussions by providing

a coherent synthesis of pre-2019 developments and outlining directions for future inquiry aimed at advancing risk evaluation in contexts where it is needed most.

REFERENCES

- [1] J. Gupta and A. Gregoriou, "Impact of market-based finance on SMEs failure," *Econ Model*, vol. 69, pp. 13–25, Jan. 2018, doi: 10.1016/J.ECONMOD.2017.09.004.
- [2] W. J. Henisz, B. A. Zelner, and M. F. Guillén, "The worldwide diffusion of market-oriented infrastructure reform, 1977-1999," *Am Sociol Rev*, vol. 70, no. 6, pp. 871–897, 2005, doi: 10.1177/000312240507000601.
- [3] C. Capps, D. Dranove, and M. Satterthwaite, "Competition and Market Power in Option Demand Markets," *Rand J Econ*, vol. 34, no. 4, p. 737, Winter 2003, doi: 10.2307/1593786.
- [4] Y. Akbar, B. Balboni, G. Bortoluzzi, D. Dikova, and A. Tracogna, "Disentangling resource and mode escalation in the context of emerging markets. Evidence from a sample of manufacturing SMEs," *Journal of International Management*, vol. 24, no. 3, pp. 257–270, Sep. 2018, doi: 10.1016/J.INTMAN.2018.01.003.
- [5] I. Hasan, K. Jackowicz, O. Kowalewski, and Ł. Kozłowski, "Do local banking market structures matter for SME financing and performance? New evidence from an emerging economy," *J Bank Financ*, vol. 79, pp. 142–158, Jun. 2017, doi: 10.1016/J.JBANKFIN.2017.03.009.
- [6] P. H. Bloch, "Product design and marketing: Reflections after fifteen years," *Journal of Product Innovation Management*, vol. 28, no. 3, pp. 378–380, May 2011, doi: 10.1111/J.1540-5885.2011.00805.X.
- [7] M. Chaffai and I. Medhioub, "Herding behavior in Islamic GCC stock market: a daily analysis," *International Journal of Islamic and Middle Eastern Finance and Management*, vol.

- 11, no. 2, pp. 182–193, Jun. 2018, doi: 10.1108/IMEFM-08-2017-0220.
- [8] I. H. Cheng and W. Xiong, “Financialization of commodity markets,” *Annual Review of Financial Economics*, vol. 6, pp. 419–441, Dec. 2014, doi: 10.1146/ANNUREV-FINANCIAL-110613-034432.
- [9] P. Bond, A. Edmans, and I. Goldstein, “The real effects of financial markets,” *Annual Review of Financial Economics*, vol. 4, pp. 339–360, Oct. 2012, doi: 10.1146/ANNUREV-FINANCIAL-110311-101826.
- [10] F. Caccioli, P. Barucca, and T. Kobayashi, “Network Models of Financial Systemic Risk: A Review,” *SSRN Electronic Journal*, Nov. 2017, doi: 10.2139/SSRN.3066722.
- [11] C. Hoffman and M. Mora Rodríguez, “Digitizing Financial Reports – Issues and Insights: A Viewpoint,” *The International Journal of Digital Accounting Research*, vol. 13, 2013, doi: 10.4192/1577-817/1577-8517-V13_3.
- [12] S. Mishra, “Financial management and forecasting using business intelligence and big data analytic tools,” <https://doi.org/10.1142/S2424786318500111>, vol. 05, no. 02, p. 1850011, Jul. 2018, doi: 10.1142/S2424786318500111.
- [13] S. O’Riain, E. Curry, and A. Harth, “XBRL and open data for global financial ecosystems: A linked data approach,” *International Journal of Accounting Information Systems*, vol. 13, no. 2, pp. 141–162, Jun. 2012, doi: 10.1016/J.ACCINF.2012.02.002.
- [14] H. Cheng, Y. C. Lu, and C. Sheu, “An ontology-based business intelligence application in a financial knowledge management system,” *Expert Syst Appl*, vol. 36, no. 2 PART 2, pp. 3614–3622, 2009, doi: 10.1016/J.ESWA.2008.02.047.
- [15] P. Kadlec, B. Gabrys, and S. Strandt, “Data-driven Soft Sensors in the process industry,” *Comput Chem Eng*, vol. 33, no. 4, pp. 795–814, Apr. 2009, doi: 10.1016/j.compchemeng.2008.12.012.
- [16] S. Fosso Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, “How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study,” *Int J Prod Econ*, vol. 165, pp. 234–246, Jul. 2015, doi: 10.1016/J.IJPE.2014.12.031.
- [17] M. A. Waller and S. E. Fawcett, “Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management,” *Journal of Business Logistics*, vol. 34, no. 2, pp. 77–84, 2013, doi: 10.1111/JBL.12010.
- [18] P. Kumar Jain, K. Mandli, I. Hoteit, O. Knio, and C. Dawson, “Dynamically adaptive data-driven simulation of extreme hydrological flows,” *Ocean Model (Oxf)*, vol. 122, pp. 85–103, Feb. 2018, doi: 10.1016/J.OCEMOD.2017.12.004.
- [19] A. Geissbuhler *et al.*, “Trustworthy reuse of health data: A transnational perspective,” *Int J Med Inform*, vol. 82, no. 1, pp. 1–9, Jan. 2013, doi: 10.1016/J.IJMEDINF.2012.11.003.
- [20] J. L. L. T. T Wang, “Data-driven fast moving consumer goods supply chain model and application,” *Int J Control Autom*, vol. 11, no. 5, pp. 125–138, 2018.
- [21] J. A. Burkell, “Remembering me: big data, individual identity, and the psychological necessity of forgetting,” *Ethics Inf Technol*, vol. 18, no. 1, pp. 17–23, Mar. 2016, doi: 10.1007/S10676-016-9393-1.
- [22] J. P. A. Ioannidis, “Informed Consent, Big Data, and the Oxymoron of Research That Is Not Research,” *American Journal of Bioethics*, vol. 13, no. 4, pp. 40–42, Apr. 2013, doi: 10.1080/15265161.2013.768864.
- [23] C. Meng, S. S. Nageshwaraniyer, A. Maghsoudi, Y. J. Son, and S. Dessureault, “Data-driven modeling and simulation framework for material handling systems in coal mines,” *Comput Ind Eng*, vol. 64, no. 3,

- pp. 766–779, 2013, doi: 10.1016/J.CIE.2012.12.017.
- [24] B. D. Mittelstadt and L. Floridi, “The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts,” *Sci Eng Ethics*, vol. 22, no. 2, pp. 303–341, Apr. 2016, doi: 10.1007/S11948-015-9652-2.
- [25] A. Castleberry and A. Nolen, “Thematic analysis of qualitative research data: Is it as easy as it sounds?,” *Curr Pharm Teach Learn*, vol. 10, no. 6, pp. 807–815, Jun. 2018, doi: 10.1016/J.CPTL.2018.03.019.
- [26] M. Bar-Sinai, L. Sweeney, and M. Crosas, “DataTags, Data Handling Policy Spaces and the Tags Language,” *Proceedings - 2016 IEEE Symposium on Security and Privacy Workshops, SPW 2016*, pp. 1–8, Aug. 2016, doi: 10.1109/SPW.2016.11.
- [27] S. Barocas and H. Nissenbaum, “Big data’s end run around procedural privacy protections,” *Commun ACM*, vol. 57, no. 11, pp. 31–33, Nov. 2014, doi: 10.1145/2668897.
- [28] P. Li *et al.*, “Promoting secondary analysis of electronic medical records in china: Summary of the plagh-mit critical data conference and health datathon,” *JMIR Med Inform*, vol. 5, no. 4, Oct. 2017, doi: 10.2196/MEDINFORM.7380.
- [29] J. W. Y Zhang, “Data-driven modeling and scientific computing,” *Appl Mech Rev*, vol. 68, no. 5, pp. 050801–051013, 2016.
- [30] J. N. K. SL Brunton, “Data-driven versus physics-based modeling,” *Annu Rev Fluid Mech*, vol. 50, pp. 645–668, 2018.
- [31] J. Hemerly, “Public policy considerations for data-driven innovation,” *Computer (Long Beach Calif)*, vol. 46, no. 6, pp. 25–31, 2013, doi: 10.1109/MC.2013.186.
- [32] J. Sandefur and A. Glassman, “The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics,” *Journal of Development Studies*, vol. 51, no. 2, pp. 116–132, Feb. 2015, doi: 10.1080/00220388.2014.968138.
- [33] K. Witkowski, “Internet of Things, Big Data, Industry 4.0 - Innovative Solutions in Logistics and Supply Chains Management,” *Procedia Eng*, vol. 182, pp. 763–769, 2017, doi: 10.1016/j.proeng.2017.03.197.
- [34] D. Li, W. Daamen, and R. M. P. Goverde, “Estimation of train dwell time at short stops based on track occupation event data: A study at a Dutch railway station,” *J Adv Transp*, vol. 50, no. 5, pp. 877–896, Aug. 2016, doi: 10.1002/ATR.1380.
- [35] H. Li, L. Xiong, L. Zhang, and X. Jiang, “DPSynthesizer: Differentially private data synthesizer for privacy preserving data sharing,” *Proceedings of the VLDB Endowment*, vol. 7, no. 13, pp. 1677–1680, 2014, doi: 10.14778/2733004.2733059.
- [36] A. Halevy, P. Norvig, and F. Pereira, “The unreasonable effectiveness of data,” *IEEE Intell Syst*, vol. 24, no. 2, pp. 8–12, 2009, doi: 10.1109/MIS.2009.36.
- [37] D. Donoho, “50 Years of Data Science,” *Journal of Computational and Graphical Statistics*, vol. 26, no. 4, pp. 745–766, Oct. 2017, doi: 10.1080/10618600.2017.1384734.
- [38] A. Aswani, Z. J. M. Shen, and A. Siddiq, “Inverse optimization with noisy data,” *Oper Res*, vol. 66, no. 3, pp. 870–892, May 2018, doi: 10.1287/OPRE.2017.1705.
- [39] J. Fan, F. Han, and H. Liu, “Challenges of Big Data analysis,” *Natl Sci Rev*, vol. 1, no. 2, pp. 293–314, Jun. 2014, doi: 10.1093/NSR/NWT032.
- [40] C. Allen *et al.*, “Data Governance and Data Sharing Agreements for Community-Wide Health Information Exchange: Lessons from the Beacon Communities,” *EGEMS*, vol. 2, no. 1, p. 1057, Apr. 2014, doi: 10.13063/2327-9214.1057.

- [41] N. Park, M. Mohammadi, K. Gorde, S. Jajodia, H. Park, and Y. Kim, "Data synthesis based on generative adversarial networks," *Proceedings of the VLDB Endowment*, vol. 11, no. 10, pp. 1071–1083, 2018, doi: 10.14778/3231751.3231757.
- [42] E. T. Cheah, W. L. Chan, and C. L. L. Chieng, "The corporate social responsibility of pharmaceutical product recalls: An empirical examination of U.S. and U.K. markets," *Journal of Business Ethics*, vol. 76, no. 4, pp. 427–449, Dec. 2007, doi: 10.1007/S10551-006-9292-1.
- [43] C. I. Enyinda and D. Tolliver, "Taking counterfeits out of the pharmaceutical supply chain in Nigeria: Leveraging multilayer mitigation approach," *Journal of African Business*, vol. 10, no. 2, pp. 218–234, 2009, doi: 10.1080/15228910903187957.
- [44] S. A. Yawar and S. Seuring, "Management of Social Issues in Supply Chains: A Literature Review Exploring Social Issues, Actions and Performance Outcomes," *Journal of Business Ethics*, vol. 141, no. 3, pp. 621–643, Mar. 2017, doi: 10.1007/S10551-015-2719-9.
- [45] P. G. Pilar, A. P. Marta, and A. Antonio, "Profit efficiency and its determinants in small and medium-sized enterprises in Spain," *BRQ Business Research Quarterly*, vol. 21, no. 4, pp. 238–250, Oct. 2018, doi: 10.1016/J.BRQ.2018.08.003.
- [46] J. Louis and P. S. Dunston, "Integrating IoT into operational workflows for real-time and automated decision-making in repetitive construction operations," *Autom Constr*, vol. 94, pp. 317–327, Oct. 2018, doi: 10.1016/J.AUTCON.2018.07.005.
- [47] P. M. Hartmann, M. Zaki, N. Feldmann, and A. Neely, "Capturing value from big data—a taxonomy of data-driven business models used by start-up firms," *International Journal of Operations & Production Management*, vol. 36, no. 10, pp. 1382–1406, 2016, doi: 10.1108/ijopm-02-2014-0098.
- [48] S. Rosenbaum, "Data governance and stewardship: Designing data stewardship entities and advancing data access," *Health Serv Res*, vol. 45, no. 5 PART 2, pp. 1442–1455, Oct. 2010, doi: 10.1111/J.1475-6773.2010.01140.X.
- [49] L. Edwards, "Privacy, Security and Data Protection in Smart Cities:," *European Data Protection Law Review*, vol. 2, no. 1, pp. 28–58, Feb. 2017, doi: 10.21552/EDPL/2016/1/6.
- [50] P. Voigt and A. von dem Bussche, "The EU General Data Protection Regulation (GDPR)," 2017, doi: 10.1007/978-3-319-57959-7.
- [51] L. Cheng, F. Liu, and D. D. Yao, "Enterprise data breach: causes, challenges, prevention, and future directions," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 7, no. 5, Sep. 2017, doi: 10.1002/WIDM.1211.
- [52] Ana. Azevedo and M. Filipe. Santos, "Integration of data mining in business intelligence systems," p. 314, 2015.
- [53] R. Schroeder, "Big Data and the brave new world of social media research," *Big Data Soc*, vol. 1, no. 2, Jul. 2014, doi: 10.1177/2053951714563194.
- [54] A. O’Cathain, E. Murphy, and J. Nicholl, "Three techniques for integrating data in mixed methods studies," *BMJ*, vol. 341, no. 7783, pp. 1147–1150, Nov. 2010, doi: 10.1136/bmj.c4587.
- [55] Y. Gong, "Data consistency in a voluntary medical incident reporting system.," *J Med Syst*, vol. 35, no. 4, pp. 609–615, Aug. 2011, doi: 10.1007/s10916-009-9398-y.
- [56] L. Curren and J. Kaye, "Revoking consent: A ‘blind spot’ in data protection law?," *Computer Law and Security Review*, vol. 26, no. 3, pp. 273–283, May 2010, doi: 10.1016/J.CLSR.2010.03.001.
- [57] B. A. Wilbanks and P. A. Langford, "A review of dashboards for data analytics in nursing," *CIN - Computers Informatics Nursing*, vol. 32,

- no. 11, pp. 545–549, 2014, doi: 10.1097/CIN.0000000000000106.
- [58] B. Baesens, R. Bapna, J. R. Marsden, J. Vanthienen, and J. L. Zhao, “Transformational issues of big data and analytics in networked business,” *MIS Quarterly*, vol. 40, no. 4, pp. 807–818, Dec. 2016, doi: 10.25300/misq/2016/40:4.03.
- [59] Y. Sun and S. Upadhyaya, “Secure and privacy preserving data processing support for active authentication,” *Information Systems Frontiers*, vol. 17, no. 5, pp. 1007–1015, Oct. 2015, doi: 10.1007/s10796-015-9587-9.
- [60] W. Wang, M. Winner, and C. R. Burgert-Brucker, “Limited service availability, readiness, and use of facility-based delivery care in Haiti: A study linking health facility data and population data,” *Glob Health Sci Pract*, vol. 5, no. 2, pp. 244–261, Jun. 2017, doi: 10.9745/GHSP-D-16-00311.
- [61] E. Baccarelli, N. Cordeschi, A. Mei, M. Panella, M. Shojafar, and J. Stefa, “Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: Review, challenges, and a case study,” *IEEE Netw*, 2016.
- [62] J. A. Potts *et al.*, “Classification of dengue illness based on readily available laboratory data,” *Am J Trop Med Hyg*, vol. 83, no. 4, pp. 781–788, Oct. 2010, doi: 10.4269/ajtmh.2010.10-0135.
- [63] E. Naydenova, A. Tsanas, S. Howie, C. Casals-Pascual, and M. De Vos, “The power of data mining in diagnosis of childhood pneumonia,” *J R Soc Interface*, vol. 13, no. 120, Jul. 2016, doi: 10.1098/RSIF.2016.0266;WGROU:STRING :PUBLICATION.
- [64] X. Liu, P. V. Singh, and K. Srinivasan, “A structured analysis of unstructured big data by leveraging cloud computing,” *Marketing Science*, vol. 35, no. 3, pp. 363–388, May 2016, doi: 10.1287/MKSC.2015.0972.
- [65] M. A. Gianfrancesco, S. Tamang, J. Yazdany, and G. Schmajuk, “Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data,” *JAMA Intern Med*, vol. 178, no. 11, pp. 1544–1547, Nov. 2018, doi: 10.1001/JAMAINTERNMED.2018.3763.
- [66] A. Gunasekaran *et al.*, “Big data and predictive analytics for supply chain and organizational performance,” *J Bus Res*, vol. 70, pp. 308–317, Jan. 2017, doi: 10.1016/j.jbusres.2016.08.004.
- [67] C. S. Kruse, R. Goswamy, Y. Raval, and S. Marawi, “Challenges and opportunities of big data in health care: A systematic review,” *JMIR Med Inform*, vol. 4, no. 4, Oct. 2016, doi: 10.2196/MEDINFORM.5359.
- [68] E. W. K. See-To and E. W. T. Ngai, “Customer reviews for demand distribution and sales nowcasting: a big data approach,” *Ann Oper Res*, vol. 270, no. 1–2, pp. 415–431, Nov. 2018, doi: 10.1007/S10479-016-2296-Z.
- [69] N. Khalil Zadeh, M. M. Sepehri, and H. Farvaresh, “Intelligent sales prediction for pharmaceutical distribution companies: A data mining based approach,” *Math Probl Eng*, vol. 2014, 2014, doi: 10.1155/2014/420310.
- [70] J. Kaye, “The tension between data sharing and the protection of privacy in genomics research,” *Annu Rev Genomics Hum Genet*, vol. 13, pp. 415–431, Sep. 2012, doi: 10.1146/ANNUREV-GENOM-082410-101454.
- [71] J. C. LB Moses, “Using big data for legal and law enforcement decisions: testing the new tools,” *Univ. New South Wales Law J.*, vol. 37, no. 2, pp. 643–678, 2014.
- [72] M. Song, Q. Du, and Q. Zhu, “A theoretical method of environmental performance evaluation in the context of big data,” *Production Planning and Control*, vol. 28, no. 11–12, pp. 976–984, 2017, doi: 10.1080/09537287.2017.1336801.

- [73] H. Baars and H. G. Kemper, "Management support with structured and unstructured data - An integrated business intelligence framework," *Information Systems Management*, vol. 25, no. 2, pp. 132–148, Mar. 2008, doi: 10.1080/10580530801941058.
- [74] A. Moser and I. Korstjens, "Series: Practical guidance to qualitative research. Part 3: Sampling, data collection and analysis," *European Journal of General Practice*, vol. 24, no. 1, pp. 9–18, Jan. 2018, doi: 10.1080/13814788.2017.1375091.
- [75] G. Y. Song, Y. Cheon, K. Lee, H. Lim, K. Y. Chung, and H. C. Rim, "Multiple categorizations of products: Cognitive modeling of customers through social media data mining," *Pers Ubiquitous Comput*, vol. 18, no. 6, pp. 1387–1403, 2014, doi: 10.1007/S00779-013-0740-5.
- [76] K. W. Pang and H. L. Chan, "Data mining-based algorithm for storage location assignment in a randomised warehouse," *Int J Prod Res*, vol. 55, no. 14, pp. 4035–4052, Jul. 2017, doi: 10.1080/00207543.2016.1244615.
- [77] S. Spiekermann and A. Novotny, "A vision for global privacy bridges: Technical and legal measures for international data markets," *Computer Law and Security Review*, vol. 31, no. 2, pp. 181–200, Apr. 2015, doi: 10.1016/J.CLSR.2015.01.009.
- [78] G. Rosano, F. Pelliccia, C. Gaudio, and A. J. Coats, "The challenge of performing effective medical research in the era of healthcare data protection," *Int J Cardiol*, vol. 177, no. 2, pp. 510–511, Dec. 2014, doi: 10.1016/J.IJCARD.2014.08.077.
- [79] S. Tiwari, H. M. Wee, and Y. Daryanto, "Big data analytics in supply chain management between 2010 and 2016: Insights to industries," *Comput Ind Eng*, vol. 115, pp. 319–330, Jan. 2018, doi: 10.1016/j.cie.2017.11.017.
- [80] Y. H. G. X. H Zhang, "Big data analytics for fast-moving consumer goods supply chain management: a review," *J Ind Inf Integr*, vol. 9, pp. 52–60, 2018.
- [81] S. Son, S. Na, and K. Kim, "Product data quality validation system for product development processes in high-tech industry," *Int J Prod Res*, vol. 49, no. 12, pp. 3751–3766, Jun. 2011, doi: 10.1080/00207543.2010.486906.
- [82] N. G. Weiskopf and C. Weng, "Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research," *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 144–151, 2013, doi: 10.1136/AMIAJNL-2011-000681.
- [83] P. E. Mbondji, D. Kebede, E. W. Soumbey-Alley, C. Zielinski, W. Kouvidila, and P. S. Lusamba-Dikassa, "Health information systems in Africa: Descriptive analysis of data sources, information products and health statistics," *J R Soc Med*, vol. 107, pp. 34–45, 2014, doi: 10.1177/0141076814531750.
- [84] A. Sebaa, F. Chikh, A. Nouicer, and A. K. Tari, "Medical big data warehouse: Architecture and system design, a case study: Improving healthcare resources distribution," *J. of Medical Systems*, vol. 42, no. 4, p. 59, Apr. 2018, doi: 10.1007/s10916-018-0894-9.
- [85] P. De Hert and V. Papakonstantinou, "The new General Data Protection Regulation: Still a sound system for the protection of individuals?," *Computer Law and Security Review*, vol. 32, no. 2, pp. 179–194, Apr. 2016, doi: 10.1016/J.CLSR.2016.02.006.
- [86] C. Bartolini and L. Siry, "The right to be forgotten in the light of the consent of the data subject," *Computer Law and Security Review*, vol. 32, no. 2, pp. 218–237, Apr. 2016, doi: 10.1016/J.CLSR.2016.01.005.
- [87] R. Zhao, Y. Liu, N. Zhang, and T. Huang, "An optimization model for green supply chain management by using a big data analytic approach," *J Clean Prod*, vol. 142, pp. 1085–

- 1097, Jan. 2017, doi: 10.1016/j.jclepro.2016.03.006.
- [88] G. Wang, A. Gunasekaran, E. W. T. Ngai, and T. Papadopoulos, "Big data analytics in logistics and supply chain management: Certain investigations for research and applications," *Int J Prod Econ*, vol. 176, pp. 98–110, Jun. 2016, doi: 10.1016/j.ijpe.2016.03.014.
- [89] L. Barabesi, A. Cerasa, D. Perrotta, and A. Cerioli, "Modeling international trade data with the Tweedie distribution for anti-fraud and policy support," *Eur J Oper Res*, vol. 248, no. 3, pp. 1031–1043, Feb. 2016, doi: 10.1016/J.EJOR.2015.08.042.
- [90] P. Quinn, A. K. Habbig, E. Mantovani, and P. De Hert, "The Data protection and Medical Device Frameworks - Obstacles to the deployment of mHealth across Europe?," *Eur J Health Law*, vol. 20, no. 2, pp. 185–204, Apr. 2013, doi: 10.1163/15718093-12341267.
- [91] R. Ramanathan, U. Ramanathan, and Y. Zhang, "Linking operations, marketing and environmental capabilities and diversification to hotel performance: A data envelopment analysis approach," *Int J Prod Econ*, vol. 176, pp. 111–122, Jun. 2016, doi: 10.1016/j.ijpe.2016.03.010.
- [92] E. Raguseo and C. Vitari, "Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects," *Int J Prod Res*, vol. 56, no. 15, pp. 5206–5221, Aug. 2018, doi: 10.1080/00207543.2018.1427900.
- [93] J. Wang and H. Yue, "Food safety pre-warning system based on data mining for a sustainable food supply chain," *Food Control*, vol. 73, pp. 223–229, Mar. 2017, doi: 10.1016/J.FOODCONT.2016.09.048.
- [94] R. Y. Zhong, S. T. Newman, G. Q. Huang, and S. Lan, "Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives," *Comput Ind Eng*, vol. 101, pp. 572–591, Nov. 2016, doi: 10.1016/j.cie.2016.07.013.
- [95] H. Demirkan and D. Delen, "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud," *Decis Support Syst*, vol. 55, no. 1, pp. 412–421, Apr. 2013, doi: 10.1016/j.dss.2012.05.048.
- [96] J. Nickerson, S. Chackungal, L. Knowlton, K. McQueen, and FM. Burkle, "Surgical care during humanitarian crises: a systematic review of published surgical caseload data from foreign medical teams.," *Prehosp Disaster Med*, vol. 27, 2012.
- [97] A. Mantelero, "The EU Proposal for a General Data Protection Regulation and the roots of the right to be forgotten," *Computer Law and Security Review*, vol. 29, no. 3, pp. 229–235, Jun. 2013, doi: 10.1016/J.CLSR.2013.03.010.
- [98] J. P. Belaud, S. Negny, F. Dupros, D. Michéa, and B. Vautrin, "Collaborative simulation and scientific big data analysis: Illustration for sustainability in natural hazards management and chemical process engineering," *Comput Ind*, vol. 65, no. 3, pp. 521–535, 2014, doi: 10.1016/j.compind.2014.01.009.
- [99] L. Sudhof *et al.*, "Local use of geographic information systems to improve data utilisation and health services: mapping caesarean section coverage in rural Rwanda," *Trop Med Int Health*, vol. 18, no. 1, pp. 18–26, Jan. 2013, doi: 10.1111/tmi.12016.
- [100] T. Sustrova, "A Suitable Artificial Intelligence Model for Inventory Level Optimization," *Trends Economics and Management*, vol. 10, no. 25, p. 48, May 2016, doi: 10.13164/TRENDS.2016.25.48.
- [101] G. Khoshhsima, "A strategic model for measuring agility with fuzzy logic," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*

- Bioinformatics*), vol. 5370 LNCS, pp. 258–268, 2008, doi: 10.1007/978-3-540-92137-0_29.
- [102] B. Alarie, A. Niblett, and A. H. Yoon, “How artificial intelligence will affect the practice of law,” *Univ. Tor. Law J.*, vol. 68, no. supplement 1, pp. 106–124, Jan. 2018, doi: 10.3138/utlj.2017-0052.
- [103] L. Tesfatsion, “Agent-based computational economics: Modeling economies as complex adaptive systems,” *Inf Sci (N Y)*, vol. 149, no. 4, pp. 262–268, 2003, doi: 10.1016/S0020-0255(02)00280-3.
- [104] D. Strömberg, “Natural disasters, economic development, and humanitarian aid,” *Journal of Economic Perspectives*, vol. 21, pp. 199–222, 2007.
- [105] “Impact of SMEs in economic growth in Albania,” *European Journal of Sustainable Development*, vol. 5, no. 3, Oct. 2015, doi: 10.14207/EJSD.2016.V5N3P151.
- [106] D. Rodrik, A. Subramanian, and F. Trebbi, “Institutions rule: The primacy of institutions over geography and integration in economic development,” *Journal of Economic Growth*, vol. 9, no. 2, pp. 131–165, Jun. 2004, doi: 10.1023/B:JOEG.0000031425.72248.85.
- [107] O. A. IO Ogunrinola, “Health and economic implications of waste dumpsites in cities: The case of Lagos, Nigeria,” *Int J Econ Finance*, vol. 4, no. 4, p. 239, 2012.
- [108] G. Benito, B. Grøgaard, and R. Narula, “Environmental influences on MNE subsidiary roles: Economic integration and the Nordic countries,” *J Int Bus Stud*, vol. 34, no. 5, pp. 443–456, 2003, doi: 10.1057/PALGRAVE.JIBS.8400047.
- [109] V. Khatri and C. V. Brown, “Designing data governance,” *Commun ACM*, vol. 53, no. 1, pp. 148–152, Jan. 2010, doi: 10.1145/1629175.1629210.
- [110] N. Wilkinson and P. Coetzee, “Internal audit assurance or consulting services rendered on governance: How does one decide?,” *Journal of Governance and Regulation*, vol. 4, no. 1, pp. 186–200, 2015, doi: 10.22495/JGR_V4_I1_C2_P3.
- [111] M. Elhelaly, “Related party transactions, corporate governance and accounting quality in Greece,” Jul. 2014.
- [112] E. Nicol, D. Bradshaw, J. Uwimana-Nicol, and L. Dudley, “Perceptions about data-informed decisions: An assessment of information-use in high HIV-prevalence settings in South Africa,” *BMC Health Serv Res*, vol. 17, Dec. 2017, doi: 10.1186/S12913-017-2641-1.
- [113] U. Flick, “Doing Qualitative Data Collection – Charting the Routes,” *The SAGE Handbook of Qualitative Data Collection*, pp. 3–16, Jan. 2018, doi: 10.4135/9781526416070.N1.
- [114] C. Pope, S. Ziebland, and N. Mays, “Qualitative research in health care: Analysing qualitative data,” *BMJ*, vol. 320, no. 7227, pp. 114–116, Jan. 2000, doi: 10.1136/bmj.320.7227.114.
- [115] T. B. Murdoch and A. S. Detsky, “The inevitable application of big data to health care,” *JAMA*, vol. 309, no. 13, pp. 1351–1352, Apr. 2013, doi: 10.1001/JAMA.2013.393.