

A CONCEPTUAL FRAMEWORK FOR PREDICTING CIDA CONSTRUCTION COST INDICES USING BIG DATA ANALYTICS

ANISTAN ROMILAS M.¹, JAYASENA H.S.², MANNAPPERUMA M.M.N.³ & S. PIRANAVAN S.^{4*}

^{1,2,3,4} Department of Building Economics, University of Moratuwa, Sri Lanka

¹romianistan@gmail.com, ²suranga@uom.lk, ³nwjmp@gmail.com, ⁴sivasothypiranavan27@gmail.com

Abstract: The Sri Lankan construction industry is significantly affected by financial risks stemming from volatile material costs, which are inadequately addressed by the static and often delayed CIDA price indices. This paper addresses this challenge by exploring the predictability of CIDA indices using Big Data Analytics (BDA). Adopting a mixed-method approach, which included an analysis of historical quantitative data and in-depth expert interviews, the study synthesizes findings to produce its main contribution: the development of a conceptual BDA framework designed to improve forecasting accuracy. This framework integrates diverse, real-time data sources and leverages advanced analytical techniques such as machine learning. The findings suggest that a data-driven approach can transform cost forecasting from a reactive to a proactive process, significantly enhancing the accuracy and responsiveness of CIDA indices and providing a robust tool for financial risk management.

Keywords: *CIDA Indices, Big Data Analysis, Predictive Modelling, Cost Forecasting, Construction Industry*

1. Introduction

The Construction industry occupies a pivotal role in Sri Lanka's economy, contributing approximately 7% to the national GDP and functioning as a key driver of employment creation, and infrastructure development (Weerakoon et al., 2023). Despite its significance, the sector continues to face persistent vulnerabilities related to cost volatility especially during period of 2020-2023 driven by macro economic instability, import restrictions, currency depreciation, supply chain disruption, and material shortages (Perera et al., 2021, Malkanthi et al., 2023). These conditions create financial uncertainty and significantly undermine the reliability of cost planning, contract administration, and risk allocation across projects (Janardana et al., 2021; Hajjath and Rathnayake, 2019). One of the principal mechanisms used in Sri Lanka to manage price fluctuation is construction cost indices published by Construction Industry Development Authority (CIDA). Historically known as ICTAD indices and often referenced in both names, this paper uses CIDA for consistency. These indices form the basis of nationally used price fluctuation formula, yet several studies have highlighted their limitations including delayed publication, inadequate geographical coverage, limited supplier participation, and restricted representation of actual market behaviour (Jayalath & perera, 2019; Nissanka & Wijesinghe, 2022, Janardana et al., 2021). These gaps are particularly problematic during periods of rapid price volatility when monthly or bi-monthly index updates cannot capture sudden shifts in cement, steel, fuel, or imported material prices.

Globally, diverse forecasting techniques including econometric models, hybrid statistical methods, and machine learning approaches, have been used to predict construction cost indices with varying levels of accuracy (AlTalhoni et al., 2024; Lederer, 2024; Shahandasti & Ashuri, 2015). However such models rely on large, well structured, and centralised datasets, which contrast sharply with Sri Lanka's fragmented and often incomplete data landscape. For example, CIDA index calculations are based predominantly on data from Colombo and Gampaha districts, leading to reduced accuracy when applied nationwide (Jayalath & Perera, 2019). Therefore, while global methods provide valuable insights, they cannot be directly transferred to Sri Lankan context without adopting to the country's unique data constraints, market structures, and economic uncertainties. Mustafa et al. (2024) argue that the sheer volume and volatility of modern construction data necessitate the adoption of Big Data Analytics (BDA) over traditional estimation methods, which struggle to handle such uncertainty. Furthermore, data fragmentation remains a critical barrier in developing contexts; as noted by Ahmed et al. (2018), the siloed nature of industry data often hinders the direct application of standard global forecasting models.

These challenges highlight the need for a more responsive and data-rich forecasting approach. Big Data Analytics (BDA), which enables the integration of diverse structured and unstructured datasets along with advanced machine learning techniques, offering a promising pathway to overcome current limitations. BDA has demonstrated string predictive capability in markets characterised by volatility, complex interactions, and non-linear patterns, including commodity pricing, logistics, and financial forecasting (Chen et al., 2014; Cao & Ashuri, 2020; Gong & Guan, 2024). In Sri Lanka, where material prices, fuel costs, and exchange rates fluctuate unpredictably and real-time data are dispersed across multiple institutions, BDA is particularly suited for consolidating information from suppliers, customs records, online procurement

*Corresponding author: Tel: +94 760849867 Email Address: sivasothypiranavan27@gmail.com
DOI: <https://doi.org/10.31705/FARU.2025.77>

platforms, macroeconomic datasets and global commodity trends. Despite its potential, no study have yet examined the use of BDA for forecasting construction cost indices, representing a clear and significant gap in local research and industry practice.

1.1 PROBLEM STATEMENT

Although CIDA indices are central to the Sri Lanka's price fluctuation mechanism, the construction industry currently lacks any predictive capability to estimate future index movements. This absence forces practitioners to rely solely on retrospective monthly index releases which provide limited support in tendering, budgeting and escalation assessments during volatile market conditions. Existing Sri Lankan research has focused primarily on identifying limitations in current indexation process such as delayed publication, narrow geographical coverage rather than developing forecasting models (Janardana et al., 2019; Jayalath & Perera, 2019; Nissanka & Wijesinghe, 2022). Although international research has demonstrated effectiveness of econometric and machine learning approaches, they rely on high quality datasets which is not available in Sri Lanka (Lederer, 2024; AlTalhoni et al., 2024, David et al., 2025). Research Gap and Proposed solution As explored above, the traditional econometric methods have been challenged by their inability to capture the frequent fluctuations in construction material prices (Liu et al., 2024). As such, the advanced cost management practices ranging from lifecycle costing and value engineering to integration of digital tools such as Building Information Modelling, Digital Twin, Blockchain, Big Data and others enables collaborative financial risk mitigation within the construction project environments. The study of the application of big data technology in cost prediction and control management of construction projects can provide an effective cost control management mode for the construction industry (Sun, 2024). While Big Data Analysis (BDA) has emerged as a powerful tool for prediction and decision-making across various sectors (Liu, 2014), its application in forecasting construction price indices in Sri Lanka has not been explored. Therefore, there is a clear gap identified in pressing need to investigate the feasibility and effectiveness of using Big Data to develop a dynamic, data-driven predictive framework for CIDA indices.

1.2 AIM AND CONTRIBUTION

The aim of this paper is to examine the predictability of CIDA indices using BDA. The established research gap in terms of CIDA indices would be addressed through this aim by developing an integrated framework. Addressing research gaps by "developing an integrated framework" is a recognized method for providing comprehensive approaches that inform strategies and mitigate financial setbacks (Jezzini et al., 2025). Meanwhile, the novelty of this paper lies in integration of BDA as a tool in this specific context as it navigates the complexity of synthesizing interdisciplinary knowledge from construction economics and data science to create a practical roadmap for modernization.

1.3 PAPER STRUCTURE

This paper is structured in 6 sections as follows: section 1 provided a brief introduction, section 2 outlines the theoretical foundations underpinning the framework, section 3 describes the methodology used for the framework's development, section 4 presents the proposed Big Data framework in detail, and section 5 discusses its validity, implications, and implementation challenges. Finally, Section 6 provides conclusions and directions for future research.

2. Theoretical foundations

2.1 LIMITATIONS OF TRADITIONAL COST INDEXATION

Construction cost indices are important parameters that assist stakeholders in estimating price sensitivity across different locations and time periods (AlTalhoni et al., 2019). In the Sri Lankan construction industry, the CIDA indices play a vital role by providing essential data for cost management and project planning (Perera & Rodrigo, 2018). For example, if a client disputes a price escalation claim, a contractor can use the CIDA indices as an authoritative reference to justify price adjustments (Malkanathi et al., 2023). Furthermore, government agencies and major construction firms utilize these indices to forecast future trends in the industry (Perera et al., 2021). However, there are well-documented weaknesses in applying these indices within the ICTAD price fluctuation formulae.

Globally, most indices rely on retrospective and aggregated data, making them reactive rather than proactive tools. A major challenge lies in the accuracy of the indices due to inconsistencies in data collection from multiple suppliers. For instance, prices of materials such as bitumen fluctuate rapidly, and the problem is intensified when suppliers fail to provide data or cease production. Labor wage data, often collected from contractors, also pose difficulties because of wide variations in pay and unreliable reporting (Janardana et al., 2021). Another limitation is that CIDA uses a composite price index. Many of these indices are derived from survey data collected from a small number of suppliers or contractors, mainly concentrated in urban centres. As a result, they fail to adequately reflect regional price variations (Jayalath & Perera, 2019). In practice, the indices are determined solely using price data from the Gampaha and Colombo districts, which means they are based on limited sources. Applying these location-dependent indices across the entire country often leads to inaccurate cost estimations. To improve accuracy, the sample size should be expanded and distributed island wide. Additionally, while CIDA indices provide a framework for cost recovery, they may not fully account for all variables influencing construction costs (Hajjath & Rathnayake, 2019; Jayalath & Wickramasinghe, 2021). The formula's design has also been criticized for its inability to adapt to sudden and significant changes in material prices. This limitation has caused severe financial strain on projects, particularly in the context of the COVID-19 pandemic (Nissanka & Wijesinghe, 2022).

2.2 THE POTENTIAL OF BIG DATA ANALYTICS IN FORECASTING

"Big Data" refers to massive datasets generated at high velocity and from varied sources and come in various formats, including structured, semi-structured, and unstructured data. (Moorthy et al., 2015). It is typically defined by the 3Vs – Volume (magnitude), Velocity (speed), and Variety (diversity) – with more recent literature expanding its dimensions up to 10Vs including Veracity (reliability), Value (utility), Validity (accuracy), Volatility (stability), Variability (inconsistency), Visualization (representation), and Vulnerability (security) (Chu et al., 2021). In addition, Big Data accounts largest growing data sets under 3 major types as structured, semi-structured, and unstructured data (El-Astal et al., 2025). Traditional data management tools often struggle to process data with all these various dimensions. As a result, Big Data Analytics (BDA) has emerged combining advanced computational methods, cloud platforms, and artificial intelligence to transform raw, fragmented data sets into meaningful insights by identifying trends, patterns and creating value (Liu et al., 2019).

The relevance of BDA for forecasting stems from its ability to move beyond static, retrospective analysis towards dynamic, real-time predictions. By integrating multiple data streams – ranging from structured financial records to unstructured sources like web data and social media sentiment – BDA enables the identification of patterns and trends that conventional econometric models to often overlook. This allows for anticipated future developments rather than merely responding to past outcomes. Zhai and Ma (2025) demonstrate that dynamic modelling in Big Data environments significantly outperforms static cost engineering methods by capturing real-time spatiotemporal variations. In sectors such as healthcare, finance, and logistics, predictive frameworks powered by Big Data have already demonstrated superior accuracy and timeliness compared to traditional forecasting methods (Chen et al., 2014; Zong & Guan, 2024). Similarly, Serugga (2025) highlights that data-driven models, such as digital twins, can reduce lifecycle cost variability by integrating real-time data acquisition, a capability lacking in retrospective indexation.

In construction industry, the potential for BDA is especially significant. In Sri Lanka, CIDA indices are critical in providing basis for price adjustments, contract administration and project planning. Yet, these indices are constrained by geographic and temporal limitations, as they are compiled retrospectively from data in only a few districts (Malkanathi et al., 2023). By contrast, BDA can incorporate real-time supplier quotations, exchange rates, inflation data, and even transaction level purchase orders to produce forecasts that better reflect current market conditions. For instance, integrating macroeconomic data from the Central Bank with digital procurement records or IoT-driven logistics information could improve responsiveness to sudden market shocks.

The predictive power of BDA is further enhanced through machine learning (ML) and Deep Learning (DL) techniques. ML algorithms detect hidden patterns with vast datasets, while DL architectures such as Long Short-Term Memory (LSTM) networks are highly effective at modelling sequential data like construction cost indices (Bandara et al., 2021). Studies demonstrate that hybrid frameworks combining statistical models with ML and DL methods consistently achieve higher forecasting accuracy than traditional techniques alone (Ngo et al., 2020; Kuster et al., 2017). Overall adoption of Big Data driven forecasting frameworks represents a shift from reactive cost adjustments based on delayed indices to proactive cost management that anticipates market dynamics. This transition is particularly relevant for Sri Lanka's construction sector in its post-crisis context, where volatility in material prices, fuel costs, and exchange rates require robust forward-looking decision-support systems.

2.3 ADVANCED ANALYTICAL TECHNIQUES FOR TIME-SERIES DATA

Forecasting economic and construction related indices has traditionally relied on statistical models such as Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA). The ARIMA model is a foundational approach to time series forecasting, widely used for its ability to capture temporal interdependence in data (Li et al., 2023). However, these models are constrained by structural assumptions, including linearity and stationarity, which are rarely consistent with the complexity of the real world datasets (Kim et al., 2022). When applied to volatile variables such as material prices, labour costs and currency fluctuations, recent studies indicate that ARIMA and related approaches often underperform, as they cannot adequately capture non-linear dynamics or long-range dependencies (Hung & Lee, 2022). For this reason, more advanced analytical techniques are needed to strengthen forecasting accuracy in construction cost indices.

ML approaches offer more robust alternatives for multi-variable, non-linear time series forecasting. Among these, Support Vector Machines (SVM) and Random Forests have been successfully applied to predict cost indices with higher accuracy, particularly where data exhibits non-linear dependencies (Ngo et al., 2020). However, the real breakthrough has come from DL techniques in handling sequential and multivariate data. Long Short-Term Memory (LSTM) networks are specifically designed to capture sequential dependencies and can handle multiple seasonal patterns. Hua et al. (2019) and Mathew & Abdulla (2020) showed that LSTMs consistently outperform ARIMA in multi-series datasets by learning complex temporal structures that classical methods cannot detect. This is particularly relevant for developing economies; Alzara et al. (2025) found that LSTM and Gated Recurrent Unit (GRU) models significantly outperformed traditional regression in predicting construction cost indices in volatile markets like Egypt. International studies provide strong empirical support: an LSTM-based framework was found to deliver more accurate predictions than time series models for highly volatile cost indices (Cao & Ashuri, 2020), while another computational LSTM model significantly outperformed a linear statistical model in economic forecasting (Assaad & El-Adaway, 2021).

More advanced models such as DeepAR leverage probabilistic approaches to not only forecast future values but also quantify uncertainty, which is essential for risk-sensitive sectors like construction (Salinas et al., 2020). Hybrid model which combine statistical approaches (e.g., ARIMA) with ML (e.g., LSTM or Gradient Boosted Trees), have consistently delivered superior performance compared to individual models. Azami et al. (2022) showed that hybrid models incorporating feature selection and domain-specific variables can significantly improve accuracy in cost forecasting. Liu et al. (2024) further stressed their robustness in volatile and data-rich environments. The integration of Big Data infrastructure also enhances the efficiency of timeseries modelling. Platforms such as Hadoop and Spark allow for parallel processing of massive datasets, enabling forecasters to process thousands of series in real time (Arputhamary & Arockiam, 2016; Zhang, 2021). This scalability ensures that models can update continuously as new data arrives, making forecasts dynamic and adaptive. The convergence of these advanced analytical techniques with Big Data technologies is what allows for construction of predictive frameworks that are not only accurate but also timely, scalable, and actionable for construction industry stakeholders. This hybrid approach is especially promising for construction cost forecasting in Sri Lanka, where traditional indices are hampered by geographic and temporal limitations, and accurate, real-time forecasting tools are urgently required.

3. Methodology for Framework Development

This study adopts a conceptual research design to develop a BDA-driven forecasting framework suitable for predicting CIDA construction cost indices in Sri Lanka. A conceptual methodology is appropriate when existing knowledge is fragmented and

3.1 RESEARCH APPROACH

This research adopted a mixed method combining both qualitative and quantitative techniques (Creswell & Creswell, 2018) to investigate the feasibility of using Big Data analysis for predicting CIDA indices. The qualitative component involved expert interviews to explore industry needs and validate the framework's design, while the quantitative component involved the analysis of historical CIDA data and economic indicators to understand key trends and relationships. This study is defined to be a conceptual and a theory building research. The objective is not to test a pre-existing hypothesis but to construct a new theoretical model grounded in established principles and validated by expert knowledge (Meredith, 1993).

3.2 DATA COLLECTION

The study began with the collection of secondary quantitative data to analyse historical trends and inform the initial development of the framework. This included historical CIDA index bulletins, alongside macroeconomic data such as inflation rates and exchange rates from the Central Bank of Sri Lanka (CBSL), and fuel price data. Following this initial analysis, primary data was collected through semi-structured interviews to refine and validate the framework (Saunders et al., 2019). This involved discussions with 12 experts, including quantity surveyors, construction economists, and data analysts, to explore current challenges and validate the framework's practicality and relevance. The Table 1 provides the profile details and the exposure areas of the interviewed experts (Guest et al., 2006).

Table 1 Profile of the interviewees

Interviewees	Discipline	Experience	Exposure to	
			CIDA Indices	Big Data
R1	Director- Consultancy Firm (Ch. Quantity Surveyor)	30 years	Yes	Yes
R2	Director- Consultancy Firm (Ch. Quantity Surveyor)	35 years	Yes	Yes
R3	Director -CIDA	22 years	Yes	Partially
R4	District Coordinating Officer- CIDA	8 years	Yes	Partially
R5	Quantity Surveyor- Consultancy Firm	13 years	Yes	Partially
R6	Lecturer- Leading University (Computer Science Engineer)	5 years	No	Yes
R7	Lecturer- Leading University (Computer Science Engineer)	6 years	Partially	Yes
R8	Data Engineer	11 years	No	Yes
R9	Director- IT Consultancy Firm	19 years	No	Yes
R10	Sales Data Analyzer	4 years	Yes	Yes
R11	Civil Engineer	10 years	Yes	Yes
R12	Civil Engineer	7 years	Yes	Yes

3.3 DATA SYNTHESIS

The collected data were synthesized systematically to produce a conceptual framework analysis procedure outlined by Jabareen (2009) to predict cost indices using BDA. This process involved identifying theoretical gaps from the literature, gathering primary insights on variables and data sources from experts, and then integrating these findings through methodological triangulation (Carter et al., 2014) to develop the framework's architecture and components. The synthesis of identified gaps from the literature and the qualitative insights from expert interviews regarding key variables, data sources, and practical industry needs are collectively done and a predictive framework was developed as an outcome.

4.The Proposed Big Data Framework for Predicting CIDA Indices

This section presents the framework developed for predicting CIDA indices using BDA, which was designed based on the synthesis of the literature review and expert consultations.

4.1 FOUNDATIONAL COMPONENTS

The framework is comprised of the following foundational components as elaborated in this sub-section.

4.1.1 Key Predictive Variables

Respondents R1, R2, and R5, all experienced chartered quantity surveyors, noted that materials typically account for 55-70% of the total construction cost in Sri Lankan projects, especially in infrastructure and high-rise developments. As one chartered quantity surveyor (R2) noted, *"If cement or steel goes up 10%, it hits the BOQ like a shockwave, it immediately distorts everything from foundations to finishes."* Another expert (R10) highlighted the sensitivity to currency, stating, *"A sudden 10% drop in the LKR immediately inflates the cost of imported items and that inflation reflects in project BOQs within weeks."* The cascading effect of fuel was also noted by a quantity surveyor (R5): *"Fuel is one of those variables that doesn't just change its own price, it changes the cost of everything else connected to it."*

The Table 2 provides the summary of the data analysed from the interviewees regarding the key variables affecting the Construction cost indices especially CIDA indices in Sri Lankan context.

Table 2 Key Variables affecting the indices

Key Variables	Description / Use	Respondents
Material Prices	Most critical cost component accounting for 55% - 70% of construction cost. Frequent fluctuations in cement, steel, aggregates; essential for index modelling.	R1, R2, R3, R4, R5, R10, R11, R12
Inflation Rate	Influences labour costs, overheads, and general escalation trends. It is useful for forecasting macroeconomic trends.	R3, R4, R6, R7, R9, R10, R12
Exchange Rate	Vital for estimating the cost of imported materials and fuel; highly volatile in the Sri Lankan context.	R2, R10, R12
Fuel Prices	Affects transportation, equipment operation, and delivery costs; key for estimating logistics-related inflation.	R2, R6, R7, R9, R10, R12

4.1.2 Required Data Ecosystem

There are several types of data sources identified for the variables affecting the indices. The data used in these indices can be categorised into **structured, semi-structured, and unstructured**, each of which plays a unique role in shaping cost assessments and economic projections. While structured data provides numerical consistency, semi-structured and unstructured data offer deeper insights into market dynamics, and real-time data ensures timely responsiveness to economic fluctuations.

Respondents, particularly from the quantity surveying discipline (R1, R2, and R5), highlighted supplier quotations, invoices, and purchase orders as highly valuable sources of real-time pricing data. In addition, respondents R10 (sales data analyst), R11, and R12 (civil engineers) emphasised the significance of point-of-sale records and contractor-level purchase orders. As one civil engineer (R12) stated, *"CIDA data is important, but supplier quotations and site-level purchase records are more up-to-date and reflect real conditions on the ground."* These real time data can be included under semi -structured data. Experts R10 and R12 advocated for the inclusion of macroeconomic indicators specifically, data from the Central Bank of Sri Lanka (CBSL) and the Department of Census and Statistics (DCS). Moreover, Respondents R5, R11, and R12 specifically referenced data from the Ceylon Petroleum Corporation (CPC) and Lanka IOC, which regularly release fuel pricing updates. These are structured data regularly updated from the reliable sources. Furthermore, experts R6 and R7, both specialising in data analysis and computer science, introduced the value of unconventional data sources which are unstructured. They include import or export statistics from the Sri Lanka Customs Department, that can reveal material availability and international price trends and digital data from social media platforms (e.g., Twitter, Facebook) and e-commerce sites.

4.1.3 Framework Architecture

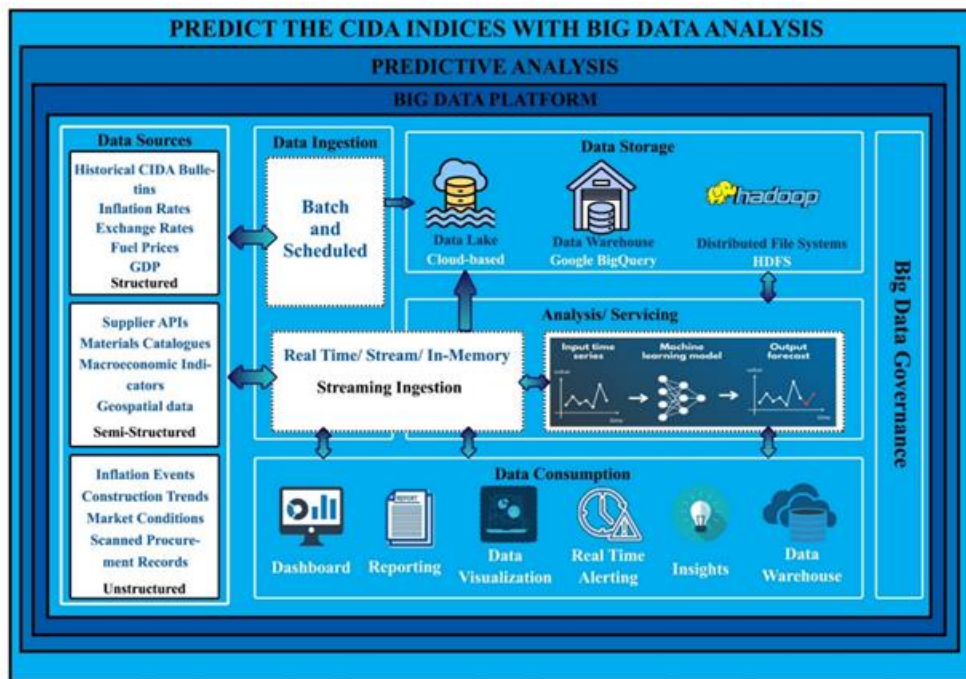


Figure 1 Framework System Architecture

The system architecture is presented in the **Error! Reference source not found.**. The proposed framework integrates historical and real-time data sources to forecast CIDA indices more accurately and responsively. It incorporates:

1. **Data Ingestion Layer:** This layer is responsible for collecting data from the diverse ecosystem of sources. It uses both batch processing for historical, structured data (e.g., monthly CIDA bulletins) and stream processing for real-time feeds (e.g., live exchange rate APIs, fuel price updates).
2. **Processing Layer:** Raw data is often inconsistent and noisy. This layer utilizes powerful Big Data technologies like Apache Spark to perform large-scale data cleaning, transformation, normalization, and feature engineering, preparing the data for analysis.
3. **Analysis Layer:** This is the predictive "engine" of the framework. It employs Machine Learning and Deep Learning models, such as LSTM networks, to analyse the processed time-series data and generate forecasts for the CIDA indices. The models are continuously trained and updated as new data is ingested.
4. **Visualisation Layer:** The output of the analysis layer is made accessible to end-users through this user-facing interface. Interactive dashboards, reports, and real-time alerts provide stakeholders (e.g., quantity surveyors, project managers, policymakers) with actionable insights for budgeting, bidding, and risk management.

5. Discussion

This section presents a discussion of the theoretical and practical aspects of the framework on how the predictability of CIDA indices is relevant and applicable in real contexts.

5.1 THEORETICAL VALIDITY AND PRACTICAL RELEVANCE

Experts R1, R2, and R5, all with over a decade of experience in cost consultancy and construction economics, affirmed that the framework offers significant improvements over the current CIDA indexation process. By integrating structured, semi-structured, and unstructured data, the system increases both the accuracy and responsiveness of index forecasting. This consensus aligns with Zhai and Ma (2025), who demonstrated that integrating dynamic modelling with Big Data environments significantly improves prediction accuracy by addressing the static nature of conventional cost engineering. Experts appreciated this structure, especially the use of deep learning algorithms such as LSTM for learning from long-term economic trends and market patterns. The experts' preference for LSTM models in volatile markets is supported by Alzara et al. (2025), whose study on construction cost forecasting in developing economies confirmed that Deep Learning models significantly outperform traditional methods in handling macroeconomic fluctuations. This makes the prediction more accurate with the help of economic data published by institutions like Central Bank and Department of Statistics. R6 stated, "With this kind of pipeline, combining LSTM or ensemble models with Spark processing and HDFS, you can build a model that

continuously learns and adapts." This supports the findings of Mustafa et al. (2024), who argue that traditional cost estimation methods are increasingly inadequate for handling the volume and uncertainty of modern construction data, necessitating such adaptive BDA frameworks.

The framework incorporates streaming ingestion pipelines that enable continuous data processing from a variety of sources. These include currency exchange APIs from the Central Bank of Sri Lanka (CBSL), fuel price updates from Lanka IOC, inflation and economic indicators from the Department of Census and Statistics (DCS), and live market activity captured through supplier APIs, digital procurement platforms, and trade sentiment analysis. The framework's reliance on real-time data streams mirrors the Digital Twin approach proposed by Serugga (2025), which utilizes continuous data acquisition to reduce lifecycle cost variability and support adaptive decision-making. According to R2 and R12, this stream-based architecture allows the forecasting engine to react to market fluctuations in near real time, effectively eliminating the delays commonly associated with the static, monthly updates provided by traditional CIDA bulletins. Respondents R5 and R12 specifically noted that access to frequently updated indices allows contractors and consultants to make timely, proactive decisions, reducing the likelihood of financial disputes and keeping projects within budget. The integration of real-time supplier data in this framework is further justified by Aziz et al. (2024), whose findings indicate that digital supply chain tracking directly correlates with reduced lead times and improved cost efficiency. A senior quantity surveyor (R2) stated, *"If we can get real-time updates fed into one system and have it benchmark CIDA index predictions, which would be a game-changer for cost control and tender planning."*

Experts R5 and R10 highlighted that sudden policy interventions such as import restrictions, tax reforms, or subsidy removals can distort material prices and supply chain behaviour overnight. This highlights the impact of sudden price shocks; similarly, Shiha et al. (2025) argue that integrating advanced forecasting with volatility quantification is essential for managing financial risks associated with material price surges. R10 stated, *"Even the best model can't predict a presidential decree or a budget announcement that bans imports or changes taxes."* A civil engineer (R12) added, *"Real-time alerting through dashboards and notifications will empower quantity surveyors and contractors to take immediate action rather than reacting too late."* R2, R10, and R12 stated that while exchange rate APIs (e.g., CBSL, commercial banks) are included in the framework, rapid and irregular depreciation patterns still challenge real-time model calibration. However, the use of LSTM networks in this context is validated by Harbaoui (2025), who demonstrated that LSTM models achieved significantly higher accuracy than traditional ARIMA models during periods of high market disruption. Furthermore, the technical architecture was also validated, with a data engineer (R8) noting, *"The layered approach is ideal. It separates data cleaning, processing, and model training in a way that can scale as your data grows."* The expert review validated the predictive viability of the proposed Big Data framework for forecasting CIDA indices. It was deemed both technically robust and industry-aligned, with strong potential to transform how cost planning and risk management are conducted in Sri Lankan construction.

5.2 CHALLENGES AND BARRIERS TO IMPLEMENTATION

One of the most significant constraints identified by experts R6, R7 and R8, was the lack of structured, centralised, and labelled datasets in the local construction domain. These challenges are not unique to Sri Lanka; studies in Malaysia have identified the *"reliability and validity of CCI"* as a major implementation challenge (Saar et al., 2019), and research in China has specifically developed models to handle "poor sample data" (Xie & Fang, 2018). Data fragmentation remains a primary hurdle; Ahmed et al. (2018) identify the siloed nature of IT systems and the lack of clearly defined business goals as significant barriers to implementing data mining in the AEC sector, paralleling the issues found in the Sri Lankan context. A data science expert (R6) addressed this directly: *"We can't train deep learning models properly if most of the data is still in scanned documents or manually recorded Excel sheets."* Another expert (R7) warned, *"Model accuracy is highly dependent on data consistency, a critical issue in Sri Lanka, where standardised and real-time construction data is still limited."*

A key barrier noted by experts R2 and R9 is the skills gap between construction professionals and data scientists. An IT director (R9) emphasized the need for collaboration, stating, *"Without quantity surveyors defining the right variables, and without IT people building the systems to process them, the model will never reach operational maturity."* Experts such as R10 and R12 warned that ML models may struggle to accommodate sudden policy changes, natural disasters, or unexpected geopolitical events, factors that dramatically influence material prices, exchange rates, and fuel costs in Sri Lanka. While not a major theme, R7 and R9 also raised the importance of data privacy, particularly when using commercial supplier data or scraping content from social media and e-commerce platforms.

5.3 IMPLICATIONS FOR THE CONSTRUCTION INDUSTRY

The potential impact of implementing this framework is significant for all key stakeholders. Accurate cost forecasting can *"provide valuable reference for the budget management and bidding decision-making"* (Li et al., 2021) and helps *"cost engineers and capital planners prepare more accurate bids, cost estimates, and budgets"* (Shahandashti & Ashuri, 2016).

- **For Contractors:** It enables more accurate bidding, better management of financial risks associated with price volatility, and more robust contractual claims. This is supported by Aziz et al. (2024), whose findings indicate that digital supply chain tracking directly correlates with reduced lead times and improved cost efficiency.

- **For Clients/Owners:** It allows for more reliable budget forecasting and contingency planning, reducing the likelihood of project cost overruns.
- **For CIDA:** It offers a clear pathway to modernize its services, providing the industry with a more valuable and credible tool for economic management.

6. Conclusions and Further research

The concluding section of this article is structured for clarity and impact. The summary of study and remarks are combined into single subsection titled conclusion followed by Directions for Future Research subsection to present specific and future-looking research that arises out of this study.

6.1 CONCLUSION

This research examined the predictability of CIDA construction cost indices using Big Data Analytics and concludes that a data-driven approach is not only feasible but offers a significant improvement over traditional methods. The study systematically addressed its objectives by evaluating potential real-time and historical data sources, identifying advanced machine learning and deep learning techniques like LSTM as highly suitable for forecasting, and identifying the most influential predictive variables, including material prices, fuel costs, and exchange rate fluctuations.

The main contribution of this paper is the proposal and conceptual validation of a novel Big Data framework tailored to the Sri Lankan construction industry. This framework provides a clear roadmap for integrating diverse data streams to produce more accurate, timely, and reliable cost forecasts. This research contributes to the groundwork for future empirical model development and offers a foundation for data-driven decision-making in Sri Lanka's construction sector. By enabling a shift from reactive to proactive cost management, the framework has the potential to enhance financial stability and strategic planning for all industry stakeholders.

6.2 DIRECTIONS FOR FUTURE RESEARCH

Future studies can build on the conceptual groundwork established here by addressing current limitations, testing the model in live environments, and exploring broader data dimensions.

The critical next steps should include:

1. **Empirical grounding of the framework:** Analysing historical quantitative data to validate the identified variables.
2. **Advanced Model Development:** explore the use of more advanced hybrid models that combine statistical methods (e.g., ARIMA) with AI-driven techniques to potentially yield superior forecasting accuracy.
3. **Broader Data Integration:** Integrate additional unstructured data sources such as social media sentiment, climatic data, and international commodity prices to strengthen the model.
4. **Examine Governance:** Examine data governance, privacy, and ethical considerations related to using supplier and financial data in forecasting systems.

7. References

- Abdul, El-Adaway, I. H., & Assaad, R. H. (2024). Investigating Price Fluctuation Transmission among Construction Materials. *Construction Research Congress 2024, CRC 2024*, 3, 339-349. <https://doi.org/10.1061/9780784485286.035>
- Ahmed, V., Aziz, Z., Tezel, A., & Riaz, Z. (2018). Challenges and drivers for data mining in the AEC sector. *Engineering, Construction and Architectural Management*, 25(11), 1436-1453. <https://doi.org/10.1108/ECAM-01-2018-0035>
- AlTalhoni, A., Liu, H., & Abudayyeh, O. (2024). Forecasting Construction Cost Indices: methods, trends, and influential factors. *Buildings*, 14(10), 3272. <https://doi.org/10.3390/buildings14103272>
- Alzara, M., Gihad, N., Abdou, H. M., Soltan, A., Hilal, A. H., & Ehab, A. (2025). Macroeconomic-aware forecasting of construction costs in developing countries: Using gated recurrent unit and long short-term memory deep learning framework. *PLOS ONE*, 20(10), e0333189. <https://doi.org/10.1371/journal.pone.0333189>
- Arputhamary, & Arockiam. (2016). Parallel Implementation of Holt-Winters Algorithm for Big Data using MapReduce Programming Model. *International Journal of Engineering Research And*, V5(02). <https://doi.org/10.17577/IJERTV5IS020557>
- Assaad, R., & El-Adaway, I. H. (2021). Determining critical combinations of safety fatality causes using spectral clustering and computational data mining algorithms. *Journal of Construction Engineering and Management*, 147(5). [https://doi.org/10.1061/\(asce\)co.1943-7862.0002040](https://doi.org/10.1061/(asce)co.1943-7862.0002040)
- Azami, R., Lei, Z., Hermann, U., & Zubick, T. (2022). A Predictive Analytics Framework for Mobile Crane Configuration Selection in Heavy Industrial Construction Projects. *Buildings*, 12(7), 960. <https://doi.org/10.3390/buildings12070960>
- Aziz, S., Kumar, P., & Khan, A. B. (2024). Assessing the Impact of Digital Supply Chain Management on the Sustainability of Construction Projects. *Review of Business and Economics Studies*, 12(3), 60-73. <https://doi.org/10.26794/2308-944X-2024-12-3-60-73>
- Bandara, K., Bergmeir, C., & Hewamalage, H. (2021). LSTM-MSNet: Leveraging Forecasts on Sets of Related Time Series with Multiple Seasonal Patterns. *IEEE Transactions on Neural Networks and Learning Systems*, 32(4), 1586-1599. <https://doi.org/10.1109/TNNLS.2020.2985720>
- Burtonshaw, S. A. (2017). *Risk and Financial Management in Construction*. Routledge. <https://doi.org/10.4324/9781315244112>
- Cao, Y., & Ashuri, B. (2020). Predicting the volatility of highway construction cost Index using Long Short-Term Memory. *Journal of Management in Engineering*, 36(4). [https://doi.org/10.1061/\(asce\)me.1943-5479.0000784](https://doi.org/10.1061/(asce)me.1943-5479.0000784)

- Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology nursing forum*, 41(5). <https://doi.org/10.1188/14.ONF.545-547>
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171-209. <https://doi.org/10.1007/s11036-013-0489-0>
- Chu, X., Nazir, S., Wang, K., Leng, Z., & Khalil, W. (2021). Big Data and Its V's with IoT to Develop Sustainability. *Scientific Programming*, 2021. <https://doi.org/10.1155/2021/3780594>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage publications.
- El-Astal, M., Allaymoun, M., & Mustafa, A. Y. (2025). What is big data? Building a fuller understanding of the term. *International Journal of Innovative Research and Scientific Studies*, 8(1), 2610–2617. <https://doi.org/10.53894/ijirss.v8i1.5035>
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59-82. <https://doi.org/10.1177/1525822X05279903>
- Hajjath, A.K.M. and Rathnayake, M.D., 2019. Appropriateness of CIDA price fluctuation formula for road construction in Sri Lanka. In: Sandanayake, Y.G., Gunatilake, S. and Waidyasekara, A. (eds). Proceedings of the 8th World Construction Symposium, Colombo, Sri Lanka, 8-10 November 2019, pp. 107-115. DOI: doi.org/10.31705/WCS.2019.11. Available at: <https://2019.ciobwcs.com/papers>
- Harbaoui, A. (2025). Forecasting Prices of Strategic Food Products: Time Series Models in the Context of Post-Crisis Market Disruption. 2025 International Conference on Artificial Intelligence and Smart Cities (ICAISC). <https://doi.org/10.1109/ICAISC64594.2025.10959395>
- Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z., & Zhang, H. (2019). Deep Learning with Long Short-Term Memory for Time Series Prediction. *IEEE Communications Magazine*, 57(6), 114-119. <https://doi.org/10.1109/MCOM.2019.1800155>
- Hung, R., & Lee, S. (2022). Analyzing the Effectiveness of COVID-19 Mitigation Policies Using ARIMA Forecasting. *Journal of Student Research*, 11(3). <https://doi.org/10.47611/jsrhs.v11i3.2998>
- Jabareen, Y. (2009). Building a conceptual framework: philosophy, definitions, and procedure. *International Journal of Qualitative Methods*, 8(4), 49-62. <https://doi.org/10.1177/160940690900800406>
- Janardana, J. A. B., Samaraweera, A., & Jayasena, H. S. (2021). Suitability of ICTAD formula and CIDA price indices to calculate the amount of price escalated in construction projects. *World Construction Symposium*, 505–515. <https://doi.org/10.31705/WCS.2021.44>
- Jayalath, C., & Perera, B. A. K. S. (2019). Policy gaps that deter fostering sustainable construction in Sri Lanka. *Proceedings of the 8th World Construction Symposium*, 484–492. <https://doi.org/10.31705/WCS.2019.48>
- Jayalath, C., & Wickramasinghe, W. (2021). Limitations and Hypothesis Underlying the Application of CIDA Price Fluctuation Formula in Sri Lanka. *Proceedings of International Conference on Business Management*, 17. <https://doi.org/10.31357/icbm.v17.5168>
- Jezzini, Y., Assaad, R. H., El-Adaway, I. H., & Nabi, M. A. (2024). An Accurate-Pricing Estimate Game-Theoretic Model for Determining Price Escalations in Construction Projects during Economic Uncertainties. *Construction Research Congress 2022*, 1170–1180. <https://doi.org/10.1061/9780784485262.119>
- Kim, S., Lee, P., Lee, M., Kim, J., & Na, W. (2021). Improved State-of-health prediction based on auto-regressive integrated moving average with exogenous variables model in overcoming battery degradation-dependent internal parameter variation. *Journal of Energy Storage*, 46, 103888. <https://doi.org/10.1016/j.est.2021.103888>
- Kuster, C., Rezgui, Y., & Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*, 35, 257-270. <https://doi.org/10.1016/j.scs.2017.08.009>
- Lederer, L., Ellingerová, H., Ďubek, S., Bočkaj, J., & Ďubek, M. (2024). Construction Price Forecasting Models in the Construction Industry: A Comparative analysis. *Buildings*, 14(5), 1325. <https://doi.org/10.3390/buildings14051325>
- Lei Sun. (2024, July 19). *Application of big data optimal selection algorithm in project bidding scheme selection*. In *International Conference on Image Processing and Artificial Intelligence (ICIPAI) 2024*, Proc. SPIE, 13213, Article 132131I. <https://doi.org/10.1117/12.3035377>
- Li, B., Xin, Q., & Zhang, L. (2022). Engineering Cost Prediction Model based on DNN. *Scientific Programming*, 2022, 1–8. <https://doi.org/10.1155/2022/3257856>
- Li, Y., Cheng, Z., Wang, W., & Deng, Y. (2023). Big data analysis model for water property forecasting. In X. Li (Ed.), *Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022)* (p. 120). SPIE. <https://doi.org/10.1117/12.2671219>
- Liu, A., Du, X., & Wang, N. (2019). Unstructured Text Resource Access Control Attribute Mining Technology Based on Convolutional Neural Network. *IEEE Access*, 7, 43031-43041. <https://doi.org/10.1109/ACCESS.2019.2907815>
- Liu, Z., Deng, L., Wang, F., Xiong, W., Wu, T., Demian, P., & Osmani, M. (2025). Building information modeling and big data in sustainable building management: research developments and thematic trends via data visualization analysis. *Systems*, 13(7), 595. <https://doi.org/10.3390/systems13070595>
- Liu. (2014). Research on the Data Model and the Approaches to Data Mining in the Semi-Structured Data. *Applied Mechanics and Materials*, 513-517, 663–666. <https://doi.org/10.4028/www.scientific.net/AMM.513-517.663>
- Malkanthi, S. N., Dharmaratne, P. D., & Galabada, G. H. (2023). Impact of Using Price Fluctuation Related Conditions on Construction Projects. *Engineer: Journal of the Institution of Engineers, Sri Lanka*, 56(1), 43-49. <https://doi.org/10.4038/engineer.v56i1.7559>
- Mathew, E., & Abdulla, S. (2020). The LSTM technique for demand forecasting of e-procurement in the hospitality industry in the UAE. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 9(4), 757. <https://doi.org/10.11591/ijai.v9i4.pp757-765>
- Meredith, J. (1993). Theory building through conceptual methods. *International Journal of Operations & Production Management*, 13(10), 3-11. <https://doi.org/10.1108/01443579310028120>
- Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., & Ghosh, P. (2015). Big Data: Prospects and Challenges. *Vikalpa*, 40(1), 74-96. <https://doi.org/10.1177/0256090915575450>
- Mustafa, M. H., Mohd-Rahim, F. A. M., Jamaludin, A. F., & Hin, K. Y. (2024). Big Data Analytics (BDA) Framework for Construction Cost Estimation in Malaysia. *Planning Malaysia*, 22(3), 273-287. <https://doi.org/10.21837/pm.v22i32.1507>
- Ngo, J., Hwang, B. G., & Zhang, C. (2020). Factor-based big data and predictive analytics capability assessment tool for the construction industry. *Automation in Construction*, 110. <https://doi.org/10.1016/j.autcon.2019.103042>
- Nissanka, N., & Wijesinghe, T. (2022). Regional relevancy of the CIDA price indices under the restrictions urged by the covid-19 pandemic. *10th World Construction Symposium*, 144–155. <https://doi.org/10.54389/DCGT7296>
- Perera, K. K. S., & Rodrigo, M. N. N. (2018). *Outsourcing Quantity Surveying Services in Construction Projects in Sri Lanka: Client/Consultant's Perspective*. [Conference Paper]. <https://www.researchgate.net/publication/322700379>

- Perera, P. A., Senanayake, G. P. D. P., Perera, B. A. K. S., & Dissaratna, V. (2021). Price Fluctuations and the impact of the CIDA formula on the profits earned by the contractors of Civil Engineering Construction Projects implemented in Sri Lanka. *Bhumi, The Planning Research Journal*, 8(2), 12. <https://doi.org/10.4038/bhumi.v8i2.78>
- Saar, C. C., Chuing, L. S., Yusof, A. M., Zakaria, R., & Chuan, T. M. (2019). Construction cost index: a case study in Malaysia. *IOP Conference Series Materials Science and Engineering*, 620(1), 012059. <https://doi.org/10.1088/1757-899x/620/1/012059>
- Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181-1191. <https://doi.org/10.1016/i.ijforecast.2019.07.001>
- Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students* (8th ed.). Pearson.
- Serugga, J. (2025). Digital Twins and AI Decision Models: Advancing Cost Modelling in Off-Site Construction. *Eng*, 6(2), 22. <https://doi.org/10.3390/eng6020022>
- Shahandashti, S. M., & Ashuri, B. (2015). Highway construction cost forecasting using vector error correction models. *Journal of Management in Engineering*, 32(2). [https://doi.org/10.1061/\(asce\)me.1943-5479.0000404](https://doi.org/10.1061/(asce)me.1943-5479.0000404)
- Shiha, A., Elazhary, M., & El-Adaway, I. H. (2025). Price Adjustment Clauses: Enhanced Price Volatility Forecasting for Construction Materials. *Journal of Construction Engineering and Management*, 151(9), 04025121. <https://doi.org/10.1061/JCEMD4.COENG-15920>
- Weerakoon, T. G., Wimalasena, S., & Fedotova, K. (2023). Economic Crisis Adaptation in Sri Lankan Construction Industry: Pathway to Prosperity. *Baltic Journal of Real Estate Economics and Construction Management*, 11(1), 240-256. <https://doi.org/10.2478/bjreecm-2023-0015>
- Xie, S., & Fang, J. (2018). Prediction of construction cost index based on multi variable grey neural network model. *International Journal of Information Systems and Change Management*, 10(3), 209. <https://doi.org/10.1504/ijiscm.2018.096770>
- Zhai, H., & Ma, J. (2025). Study on forecasting method of power engineering cost based on BIM and DynGCN. *PLOS ONE*, 20(5), e0322202. <https://doi.org/10.1371/journal.pone.0322202>
- Zhang. (2021). Overview of Construction and Application of Big Data Analysis Platform. *Journal of Electronic Research and Application*, 4(6). <https://doi.org/10.26689/jera.v4i6.1750>
- Zong, Z., & Guan, Y. (2024). AI-Driven Intelligent Data Analytics and Predictive Analysis in Industry 4.0: Transforming Knowledge, Innovation, and Efficiency. *Journal of the Knowledge Economy*, 16(1), 864-903. <https://doi.org/10.1007/s13132-024-02001-z>