© June 2026 | IJIRT | Volume 13 Issue 1 | ISSN: 2349-6002

AI-Based Automatic Mapping of BIM Elements to Cost Items For 5D BIM Automation

Rushikesh D. Deshmukh[1], Rahul D. Shinde[2], Pratibha. M. Alandkar[3]
[1]*M.E. Student, RMD Sinhgad Technical Institutes Campus, Warje, Pune 411058*
[2]*Prof., RMD Sinhgad Technical Institutes Campus, Warje, Pune 411058*
[3]*Head of Dept., RMD Sinhgad Technical Institutes Campus, Warje, Pune 411058*

***Abstract*—Building Information Modelling (BIM) has improved quantity extraction and project visualization, but cost estimation in 5D BIM still relies heavily on manual mapping between BIM elements and cost database items. This study presents an AI-based semantic similarity framework that automatically maps BIM elements to corresponding DSR (Data Schedule of Rates) cost items. Using Natural Language Processing (NLP) and sentence embedding techniques, the system understands contextual similarities between BIM descriptions and DSR item descriptions, overcoming the limitations of traditional keyword-based approaches. A Python-based prototype was developed to process BIM data, perform semantic matching, and generate cost estimates automatically. The proposed framework reduces manual effort, improves consistency, and supports the advancement of intelligent 5D BIM cost estimation workflows.**

***Index Terms*—Building Information Modelling (BIM), 5D BIM, Artificial Intelligence, Natural Language Processing, Semantic Similarity, Cost Estimation, DSR, Construction Automation.**

I. INTRODUCTION

The construction industry is undergoing a significant digital transformation through the adoption of Building Information Modeling (BIM), which enables the creation and management of digital representations of physical and functional characteristics of buildings. BIM has become a widely accepted technology for improving project planning, design coordination, visualization, and quantity takeoff processes. By integrating project information within a single digital environment, BIM enhances collaboration among architects, engineers, contractors, and project managers throughout the project lifecycle.

The evolution of BIM has led to the development of multi-dimensional BIM applications. While 3D BIM focuses on geometric modeling and visualization, 4D BIM incorporates project scheduling information, and 5D BIM integrates cost information with the digital model. The integration of cost data enables stakeholders to perform quantity-based cost estimation, monitor project budgets, evaluate design alternatives, and support informed decision-making during the planning and construction phases. As a result, 5D BIM has become an important component of modern construction project management.

Despite the advantages of 5D BIM, one of the major challenges in its implementation is the mapping of BIM model elements to corresponding cost database items such as the Data Schedule of Rates (DSR). In conventional workflows, estimators manually identify and assign cost items to BIM elements such as beams, columns, slabs, and walls. This process is time-consuming, repetitive, and highly dependent on individual expertise. Furthermore, variations in naming conventions and engineering descriptions between BIM software and cost databases often lead to incorrect mappings and inconsistencies in cost estimation results.

Recent advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) have created new opportunities for automating information processing tasks within the construction industry. Semantic similarity techniques allow computers to understand contextual relationships between textual descriptions rather than relying solely on exact keyword matching. These capabilities can be utilized to establish intelligent relationships between BIM element descriptions and DSR cost-item descriptions, thereby reducing manual intervention and improving mapping accuracy.

This research proposes an AI-based semantic similarity framework for automatically mapping BIM elements to corresponding DSR cost items. BIM quantity information extracted from Tekla Structures is compared with DSR descriptions using sentence embedding models and semantic similarity algorithms. The proposed framework automates the BIM-to-cost mapping process and generates cost estimates based on the identified matches. By integrating BIM, Artificial Intelligence, and semantic analysis, the study aims to improve the efficiency, consistency, and reliability of 5D BIM cost estimation workflows.

The developed system contributes to the advancement of intelligent construction cost management by reducing manual effort and establishing a foundation for future AI-driven BIM applications. The proposed approach supports the broader objective of digital transformation within the Architecture, Engineering, and Construction (AEC) industry and demonstrates the potential of AI-assisted automation in BIM-based project delivery systems.

II. LITERATURE REVIEW

Eastman et al. [1] introduced Building Information Modeling (BIM) as a digital representation of a facility's physical and functional characteristics and demonstrated its effectiveness in improving quantity takeoff accuracy and project visualization. However, the integration of BIM with cost estimation systems remained dependent on manual mapping between BIM elements and cost database items. Abanda et al. [2] proposed a BIM-based ontology framework for construction cost estimation that improved interoperability between BIM models and cost databases through semantic representation of construction information. Anireddy [3] investigated BIM integration in cost estimation and reported improvements in estimation consistency and project visualization while highlighting the continued dependency on manual cost-item assignment.

Banihashemi et al. [4] developed a machine learning-integrated 5D BIM framework for construction cost management and demonstrated the potential of Artificial Intelligence in automating cost-related processes. Park and Yun [5] applied deep learning techniques to BIM-based cost prediction and reported improved estimation accuracy during the early design stages. Darko et al. [6] reviewed Artificial Intelligence applications in the Architecture, Engineering, and Construction (AEC) industry and identified significant opportunities for integrating machine learning with BIM-enabled project management systems.

Pauwels et al. [7] explored the application of semantic web technologies in the AEC industry and emphasized their importance in improving interoperability and information exchange among construction systems. Cer et al. [8] introduced the Universal Sentence Encoder for semantic text representation and demonstrated its effectiveness in understanding contextual relationships between textual descriptions. Devlin et al. [9] developed BERT, a transformer-based language model capable of capturing contextual meaning in textual data, while Reimers and Gurevych [10] proposed Sentence-BERT for semantic similarity analysis and achieved significant improvements in sentence-level matching tasks. Avogaro et al. [11] reviewed the intersection of BIM, Artificial Intelligence, and cost management and highlighted the growing importance of intelligent cost estimation systems. Bilal et al. [12] discussed the role of data-driven technologies in the construction industry and emphasized the need for intelligent systems capable of processing large volumes of project information.

Research Gap

The reviewed literature indicates that BIM has successfully automated quantity extraction and improved project visualization. Although Artificial Intelligence and Natural Language Processing techniques have demonstrated strong capabilities in semantic understanding and information processing, limited research has focused on applying semantic similarity methods for automatic BIM-to-DSR mapping. Existing BIM-based cost estimation systems still rely heavily on manual mapping or predefined rule-based approaches, creating a need for intelligent frameworks capable of automatically associating BIM elements with corresponding cost database items for efficient 5D BIM cost estimation.

III. METHODOLOGY

The proposed framework aims to automate the process of mapping BIM elements to corresponding DSR cost items using Artificial Intelligence and semantic similarity techniques. The workflow consists of BIM data extraction, DSR database processing, text preprocessing, semantic similarity analysis, and automated cost estimation. The overall methodology is shown in Fig. 1.

A. BIM Data Collection and Quantity Extraction

A Building Information Model developed in Tekla Structures was used as the primary source of project information. BIM enables the extraction of detailed quantities directly from the digital model, eliminating manual quantity takeoff procedures. The Bill of Materials (BOM) generated from Tekla Structures was exported in Excel format and contained information such as element names, material types, volumes, weights, and quantities.

The extracted BIM data served as the input dataset for the proposed system. Since BIM elements may have different naming conventions than those used in DSR databases, direct matching is often difficult and requires intelligent processing techniques.

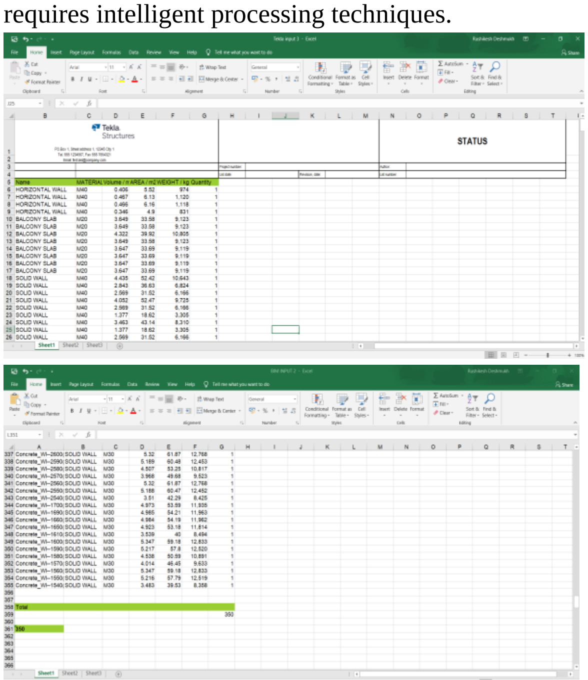

Fig. 1. BIM Quantity Data Extracted from Tekla Structures

B. DSR Database Collection

The Data Schedule of Rates (DSR) database was collected from government-published schedule of rates documents available in PDF format. The database contained item descriptions, item codes, units of measurement, and corresponding rates for various construction activities.

Python-based PDF extraction techniques were used to convert the DSR data into a structured format. The extracted information was stored in tabular form to facilitate comparison with BIM element descriptions. Since DSR descriptions are often lengthy and unstructured, preprocessing was required before performing semantic analysis.

Additionally, the extracted DSR data was examined for inconsistencies such as duplicate entries, varying terminology, abbreviations, and differences in description formats. These inconsistencies can significantly affect the accuracy of semantic matching between BIM elements and cost items. Therefore, preprocessing techniques were applied to improve data quality and create a standardized dataset suitable for Artificial Intelligence-based analysis. This step plays a crucial role in enhancing the reliability and effectiveness of the proposed BIM-to-DSR mapping framework.

Sr. No.	Chapter	SSR Item No.	Reference No.	Description of the item	Additional Specification	Unit	Completed Rate for 2022-23 excluding GST In Rs.	Labour Rate for 2022-23 excluding GST In Rs.
SSR 2022-23 :- Reinforced Cement Concrete								
836	Reinforced Cement Concrete	25.11	BDF 3	Providing and laying in situ /Ready Mix cement concrete M-20 of trap / granite /quartzite/ gneiss metal for R.C.C. work in foundations like raft, strip foundations, grillage and footings of R.C.C. columns and steel stanchions etc. including bailing out water, Steel centering formwork, laying/pumping cover blocks, compaction and curing roughening the surface if special finish is to be provided (Excluding reinforcement and structural steel) etc. complete, with fully automatic micro processor based PLC with SCADA enabled reversible Drum Type mixer/ concrete Batch mix plant (Pan mixer) etc. complete. With fine aggregate (Crushed sand VSI Grade)	Bd.F.3 Page One Number 298 and B.7, Page One Number38	One Cubic Metre	7104	2110
837	Reinforced Cement Concrete	25.13	BDF 3	Providing and laying Cast in situ/Ready Mix cement concrete M-25 of trap / granite/quartzite/ gneiss metal for R.C.C. work in foundations like raft, strip foundations, grillage and footings of R.C.C. columns and steel stanchions etc. including bailing out water, Steel centering, formwork ,cover blocks, laying/pumping, compaction and curing roughening the surface if special finish is to be provided (Excluding reinforcement and structural steel) etc. complete, with fully automatic micro processor based PLC with SCADA enabled reversible Drum Type mixer/ concrete Batch mix plant (Pan mixer) etc. complete. With fine aggregate (Crushed sand VSI Grade)	Bd.F.3 Page One Number 298 and B.7, Page One Number38	One Cubic Metre	7287	2110
838	Reinforced Cement Concrete	25.15	BDF 3	Providing and laying Cast in situ/Ready Mix cement concrete M-30 of trap / granite/quartzite/ gneiss metal for R.C.C. work in foundations like raft, strip foundations, grillage and footings of R.C.C. columns and steel stanchions etc. columns as per detailed designs and drawing or as directed including Steel centering formwork, cover blocks laying/pumping, compaction finishing the formed surfaces with cement mortar 1:3 of sufficient minimum thickness to give a smooth and even surface or roughening if special finish is to be provided and curing etc. complete. (Excluding reinforcement and structural steel). with fully automatic micro processor based PLC with SCADA enabled reversible Drum Type mixer/ concrete Batch mix plant (Pan mixer) etc. complete. With fine aggregate (Crushed sand VSI Grade)	Bd.F.3 Page One Number 298 and B.7, Page One Number38	One Cubic Metre	7471	2110
839	Reinforced Cement Concrete	25.17	BDF 3	Providing and laying Cast in situ/Ready Mix cement concrete M-35 of trap / granite/quartzite/ gneiss metal for R.C.C. work in foundations like raft, strip foundations, grillage and footings of R.C.C. columns and steel stanchions etc. columns as per detailed designs and drawing or as directed including Steel centering, formwork, cover blocks, laying/pumping, compaction finishing the formed surfaces with cement mortar 1:3 of sufficient minimum thickness to give a smooth and even surface or roughening if special finish is to be provided and curing etc. complete.(Excluding reinforcement and structural steel).with fully automatic micro processor based PLC with SCADA enabled reversible Drum Type mixer/ concrete Batch mix plant (Pan mixer) etc. complete. With fine aggregate (Crushed sand VSI Grade)	Bd.F.3 Page One Number 298 and B.7, Page One Number38	One Cubic Metre	7562	2110
840	Reinforced Cement Concrete	25.19	BDF 3	Providing and laying Cast in situ/Ready Mix cement concrete M-40 of trap / granite/quartzite/ gneiss metal for R.C.C. work in foundations like raft, strip foundations, grillage and footings of R.C.C. columns and steel stanchions etc. columns as per detailed designs and drawing or as directed including steel centering, formwork, cover blocks, laying/pumping, compaction finishing the formed surfaces with cement mortar 1:3 of sufficient minimum thickness to give a smooth and even surface or roughening if special finish is to be provided and curing etc. complete. (Excluding reinforcement and structural steel).with fully automatic micro processor based PLC with SCADA enabled reversible Drum Type mixer/ concrete Batch mix plant (Pan mixer) etc. complete. With fine aggregate (Crushed sand VSI Grade)	Bd.F.3 Page One Number 298 and B.7, Page One Number38	One Cubic Metre	7801	2110
841	Reinforced Cement Concrete	25.31	BDF 5	Providing and laying Cast in situ/Ready Mix cement concrete M-20 of trap / granite /quartzite/ gneiss metal for R.C.C. columns as per detailed designs and drawing or as directed including steel centering, formwork, cover blocks, laying/pumping, compaction finishing the formed surfaces with cement mortar 1:3 of sufficient minimum thickness to give a smooth and even surface or roughening if special finish is to be provided and curing etc. complete.(Excluding reinforcement and structural steel).with fully automatic micro processor based PLC with SCADA enabled reversible Drum Type mixer/ concrete Batch mix plant (Pan mixer) etc. complete. With fine aggregate (Crushed sand VSI Grade)	Bd.F.5 Page One Number 300 and B.7, Page.One Number 38	One Cubic Metre	13783	5252

Fig. 2. Sample DSR Database Extract

C. Data Preprocessing

Data preprocessing was performed to improve the quality and consistency of textual information obtained from BIM and DSR datasets.

The preprocessing stage included:

- Conversion of all text to lowercase.

- Removal of special characters, punctuation marks, and unnecessary symbols.
- Elimination of redundant words and irrelevant information.
- Standardization of engineering terminology and material descriptions.
- Normalization of textual data to improve similarity matching.

This process reduced inconsistencies between BIM element names and DSR descriptions and improved the effectiveness of semantic similarity calculations.

D. Semantic Similarity-Based Mapping

The core component of the proposed framework is the semantic similarity engine. Traditional keyword-based matching techniques rely on exact word matching and often fail when BIM and DSR descriptions use different terminology. To overcome this limitation, the proposed system utilizes Natural Language Processing (NLP) and sentence embedding techniques.

Sentence Transformer (all-MiniLM-L6-v2) was used to generate vector representations of BIM element descriptions and DSR item descriptions. These embeddings capture the contextual meaning of sentences rather than individual keywords. Cosine Similarity was then applied to measure the similarity between BIM and DSR embeddings.

The DSR item with the highest similarity score was automatically selected as the most relevant cost item for the BIM element. This approach enables the system to identify contextual relationships even when the wording differs significantly.

E. Automated Cost Estimation

After successful mapping, the quantity extracted from the BIM model was linked with the corresponding DSR rate. The cost of each construction element was calculated using:

Cost = Quantity × Rate

The individual costs of all mapped elements were aggregated to obtain the total project cost. The system automatically generated a cost estimation report containing BIM element descriptions, matched DSR items, rates, quantities, similarity scores, and total costs.

F. System Implementation

The proposed framework was implemented as a web-based application using Python and Django. Various libraries such as Pandas, PDFPlumber, Sentence Transformers, Scikit-Learn, and OpenPyXL were used for data processing, semantic similarity analysis, and report generation.

The developed application allows users to upload BIM quantity files and DSR databases, perform automatic mapping, and generate cost estimation reports through a simple and user-friendly interface. This workflow significantly reduces manual effort and supports efficient implementation of 5D BIM cost estimation.

IV. RESULTS AND DISCUSSION

The proposed AI-based semantic similarity framework was successfully implemented and tested using BIM quantity data extracted from Tekla Structures and DSR cost data obtained from Schedule of Rates documents. The developed system was capable of automatically extracting BIM element information, processing DSR descriptions, performing semantic matching, and generating cost estimation reports.

The semantic similarity model effectively identified contextual relationships between BIM element descriptions and DSR item descriptions, reducing the dependency on exact keyword matching. The developed tool successfully automated the BIM-to-DSR mapping process and generated element-wise cost estimates with minimal user intervention.

The generated output included BIM element descriptions, matched DSR items, similarity scores, quantities, rates, and estimated costs. The results demonstrated that semantic matching techniques can significantly improve consistency and efficiency in BIM-based cost estimation workflows.

During implementation, it was observed that the quality and structure of DSR descriptions influence mapping performance. Variations in terminology, abbreviations, and formatting occasionally affected the matching accuracy. Therefore, ongoing work focuses on data cleaning, standardization, and optimization of the DSR database to further improve the reliability of the system.

Overall, the developed framework demonstrated the feasibility of integrating Artificial Intelligence, Natural Language Processing, and BIM technologies for automated 5D BIM cost estimation.

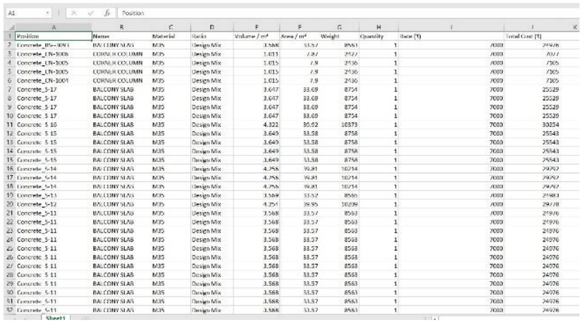

Fig. 3. Semantic Matching Results Generated by the Proposed System

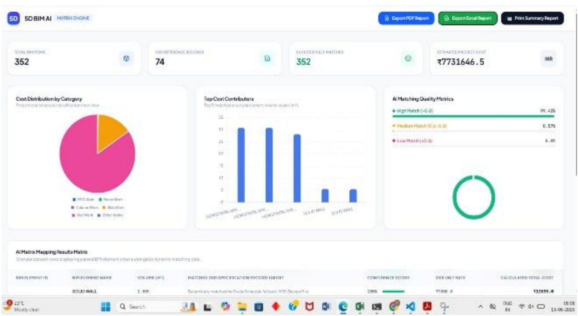

Fig. 4. Cost Estimation Output Generated by the Proposed System

BIM Element	Matched DSR Item	Similarity Score
RCC Beam	RCC Beam M25	0.92
RCC Column	RCC Column M25	0.90
RCC Slab	RCC Slab M25	0.89

Fig. 5 Sample BIM Element to DSR Mapping Results

V. CONCLUSION

This study presented an AI-based semantic similarity framework for automatically mapping BIM elements to corresponding DSR cost items. The proposed system integrates BIM quantity extraction, DSR processing, semantic similarity analysis, and automated cost estimation within a unified workflow.

The developed framework successfully reduced the manual effort involved in BIM-to-cost mapping and improved the efficiency of 5D BIM cost estimation. By utilizing Natural Language Processing and Sentence Transformer embeddings, the system was able to identify contextual similarities between BIM element descriptions and DSR item descriptions more effectively than traditional keyword-based approaches.

Although challenges related to DSR data quality and standardization remain, the results demonstrate that AI-driven semantic matching can serve as a practical solution for intelligent cost estimation in BIM environments. The proposed framework provides a strong foundation for future research and development of advanced BIM-integrated cost management systems.

VI. FUTURE SCOPE

The proposed framework establishes a foundation for intelligent BIM-based cost estimation; however, several opportunities exist for further enhancement and industrial implementation.

- Integration of the developed system with BIM software such as Tekla Structures and Autodesk Revit through APIs to enable real-time cost estimation during model development.
- Development of a dedicated Tekla Structures plugin that allows automatic BIM-to-DSR mapping and cost estimation directly within the BIM environment.
- Incorporation of contractor-specific and private cost databases in addition to government-issued DSR/SSR databases, improving the applicability of the system across different construction sectors.
- Implementation of advanced transformer-based language models and feedback-learning mechanisms to improve semantic matching accuracy and continuously enhance system performance.
- Development of an intelligent standalone software capable of processing PDF, Excel, and other cost database formats while automatically cleaning and structuring unorganized data.
- Extension of the framework toward complete 5D BIM and 6D BIM integration by combining cost estimation, project scheduling, and lifecycle management within a unified platform.
- Deployment of the system on cloud-based platforms to support multi-user collaboration, centralized data management, and remote accessibility.
- Development of automated data-cleaning and standardization modules to improve the quality of DSR databases and further enhance mapping reliability and estimation accuracy.

REFERENCES

[1] C. Eastman, P. Teicholz, R. Sacks, and K. Liston, *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*, 2nd ed. Hoboken, NJ, USA: John Wiley & Sons, 2011.

[2] F. H. Abanda, B. Kamsu-Foguem, and J. H. M. Tah, "BIM–New Rules of Measurement Ontology for Construction Cost Estimation," Engineering Science and Technology, an International Journal, vol. 20, no. 2, pp. 443–459, 2017.

[3] A. R. Anireddy, "Integrating Building Information Modeling (BIM) in Cost Estimation: Assessing the Benefits and Challenges," European Journal of Advances in Engineering and Technology, vol. 7, no. 5, pp. 125–129, 2020.

[4] S. Banihashemi, S. Khalili, M. Sheikhkhoshkar, and A. Fazeli, "Machine Learning-Integrated 5D BIM Informatics: Building Materials Costs Data Classification and Prototype Development," Innovative Infrastructure Solutions, vol. 7, no. 1, 2022.

[5] D. Park and S. Yun, "Construction Cost Prediction Using Deep Learning with BIM Properties in the Schematic Design Phase," Applied Sciences, vol. 13, no. 12, 2023.

[6] A. Darko, A. P. C. Chan, M. A. Adabre, D. J. Edwards, M. R. Hosseini, and E. E. Ameyaw, "Artificial Intelligence in the AEC Industry: Scientometric Analysis and Visualization of Research Activities," Automation in Construction, vol. 112, Art. no. 103081, 2020.

[7] P. Pauwels, S. Zhang, and Y. C. Lee, "Semantic Web Technologies in AEC Industry: A Literature Overview," Automation in Construction, vol. 73, pp. 145–165, 2017.

[8] D. Cer et al., "Universal Sentence Encoder," arXiv preprint arXiv:1803.11175, 2018.

[9] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," in Proc. North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2019, pp. 4171–4186.

[10] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks," in Proc. Conf. Empirical Methods in Natural Language Processing and Int. Joint Conf. Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3982–3992.

[11] D. Avogaro, J. Cassandro, E. Dall'Anese, C. Dori, A. Farina, and E. Laurini, "Mapping Cost Intersection Through LCC, BIM, and AI: A Systematic Literature Review for Future Opportunities," Buildings, vol. 15, no. 18, 2025.

[12] M. Bilal, L. O. Oyedele, J. Qadir, K. Munir, S. O. Ajayi, O. O. Akinade, M. Pasha, and S. A. Bello, "Big Data in the Construction Industry: A Review of Present Status, Opportunities, and Future Trends," Advanced Engineering Informatics, vol. 30, no. 3, pp. 500–521, 2016.