A MULTIMODAL GENERATIVE AI FRAMEWORK FOR MECHANICAL DESIGN SYNTHESIS

Dr.R.Raveen[1] , C.Siva Muthiah[2], T.Surya Kumar[3], T.Ragul[4]

[1]Department Of Mechanical Engineering, Sri Ramakrishna Engineering College, Coimbatore
raveen.rangasamy@srec.ac.in

[2]Department Of Mechanical Engineering, Sri Ramakrishna Engineering College, Coimbatore
sivamuthiah.2204121@srec.ac.in

[3]Department Of Mechanical Engineering, Sri Ramakrishna Engineering College, Coimbatore
suryakumar.2204128@srec.ac.in

[4]Department Of Mechanical Engineering, Sri Ramakrishna Engineering College, Coimbatore
ragul.2204109@srec.ac.in

Abstract— Mechanical design automation continues to evolve with the integration of artificial intelligence into computer-aided design workflows. Conventional parametric modeling environments provide high geometric precision but remain dependent on manual feature construction, dimensional constraint assignment, and sequential modeling operations. This research presents a multimodal generative artificial intelligence framework that synthesizes parametric mechanical components directly from natural language descriptions and two-dimensional sketch images. The framework integrates structured natural language parsing, geometric reasoning, contour-based visual interpretation, and script-driven parametric CAD generation into a unified architecture. Unlike mesh-based generative models, the proposed system produces dimensionally consistent parametric geometry compatible with engineering modification and manufacturing workflows. Text-to-CAD conversion leverages structured semantic mapping techniques, while image-to-CAD conversion employs contour extraction and shape classification algorithms grounded in computer vision methodologies. The resulting components are rendered into STL format and visualized interactively within a web-based 3D environment. Experimental evaluation demonstrates substantial reduction in modeling time while preserving dimensional accuracy and editability. The study contributes toward bridging multimodal artificial intelligence research and practical mechanical design synthesis.

Keywords: Multimodal Artificial Intelligence, Parametric CAD, Generative Engineering Design, Mechanical Automation, Text-to-CAD

INTRODUCTION

Computer-aided design systems have fundamentally transformed mechanical engineering practice by enabling precise digital modeling, simulation, and manufacturability validation. Feature-based and parametric modeling techniques introduced associative geometry control, allowing dimensional relationships to be maintained automatically [3], [4], [18]. Despite these advancements, CAD modeling remains labor-intensive during early conceptual design phases. Engineers must manually translate conceptual ideas into structured geometric representations, often repeating similar modeling operations.

The emergence of artificial intelligence has introduced generative techniques capable of producing structured outputs from abstract inputs. Foundational research in deep learning has demonstrated that neural architectures can extract hierarchical features from unstructured data [6]. Geometric deep learning has further extended these capabilities to non-Euclidean domains such as point clouds and graph-structured representations [5], [17]. These developments provide the theoretical basis for automated shape understanding and reconstruction.

Recent works in generative modeling and shape synthesis have shown that probabilistic latent representations can reconstruct complex three-dimensional geometry from limited information [2], [8]. However, most existing approaches focus on mesh generation for visualization purposes rather than parametric feature-based modeling required in engineering practice. Furthermore, deep reconstruction systems often require extensive computational resources and large training datasets.

Multimodal learning approaches integrate linguistic and visual reasoning, enabling models to interpret multiple information sources simultaneously [17]. This paradigm offers a promising direction for engineering design automation, where textual specifications and sketches frequently represent early-stage design intent. However, there remains a gap between multimodal AI capabilities and parametric CAD generation systems.

This research proposes a structured multimodal framework that combines semantic parsing, contour-based image analysis, and rule-driven parametric CAD scripting. The goal is to transform conceptual mechanical design descriptions into editable parametric geometry while maintaining computational efficiency and engineering interpretability.

II. THEORETICAL FOUNDATION

A. Parametric and Feature-Based Modeling

Parametric CAD systems are built upon dimension-driven feature hierarchies that maintain geometric relationships through constraint propagation [3], [4], [18]. Unlike static mesh models, parametric systems allow dynamic modification of geometry through parameter adjustments. ISO 10303 STEP standards further emphasize structured product data exchange for maintaining geometric fidelity [19]. These principles form the foundation of the proposed generation engine.

B. Geometric Learning and Shape Understanding

Geometric deep learning techniques extend convolutional architectures to irregular domains such as graphs and point clouds [5]. Wang et al. introduced dynamic graph convolutional neural networks capable of extracting local geometric features from spatial data [1]. Such methodologies inform the theoretical basis for structured geometric reasoning.

C. Computer Vision for Shape Extraction

Computer vision techniques provide deterministic alternatives to deep reconstruction models. OpenCV-based image processing pipelines support grayscale conversion, thresholding, and contour extraction [10], [23]. Edge detection and segmentation strategies form the backbone of efficient shape classification methods. These lightweight algorithms enable real-time geometric approximation without computationally intensive training.

D. Generative Design and Engineering Automation

Generative design research emphasizes algorithmic exploration of geometry under performance constraints [15], [16], [21]. However, most generative design systems operate on predefined geometry rather than translating conceptual input into base models. Human-AI collaborative frameworks suggest that intelligent systems should augment designers rather than replace them [20]

III. PROPOSED ARCHITECTURE

The proposed framework integrates frontend interaction, multimodal processing, and parametric CAD scripting within a modular pipeline. The backend architecture leverages FastAPI-based communication protocols [12] to handle structured requests. The frontend visualization environment is built using modern WebGL frameworks such as Three.js and React-based rendering engines [11], [13].

Textual prompts undergo lexical tokenization and structured parameter extraction. Inspired by semantic parsing techniques in AI-assisted design frameworks [25], keywords corresponding to component types trigger predefined parametric templates. Numerical values extracted from text map directly to geometric dimensions within OpenSCAD scripts [9].

Image inputs are processed using contour detection techniques grounded in classical computer vision methodologies [10], [23]. The largest detected contour is classified based on aspect ratio and curvature analysis. Rectangular contours generate cuboidal primitives, circular contours generate cylindrical primitives, and irregular contours result in extruded plates. efficiency.

IV. TEXT-TO-CAD GENERATION

Text-based synthesis translates engineering descriptions into structured geometry. Inspired by knowledge-based engineering automation systems [24], the framework maintains template-driven definitions for bolts, nuts, springs, and flange couplings.

For instance, a bolt model integrates a cylindrical shaft and head union, while a spring utilizes helical extrusion parameters derived from user-defined coil diameter and pitch. The reliance on parametric definitions ensures compatibility with conventional CAD modification workflows [18].

International *Journal on Applications in Engineering and Technology*
Volume 12 : Issue 3 : March 2026, pp 1 – 5 www.aetsjournal.com ISSN (Online) : 2455 - 0523

The system architecture reflects principles of explainable AI in engineering design, avoiding opaque neural decision-making while retaining automation benefits [20].

Fig. 1 Text-to-CAD generation module demonstrating automated flange coupling model creation.

Fig. 2 Text-to-CAD generation module demonstrating automated Nut model creation

V. IMAGE-TO-CAD RECONSTRUCTION

The image-to-CAD module follows a structured computer vision pipeline instead of relying on deep neural reconstruction. The input image is first converted to grayscale and processed using thresholding to isolate the primary object from the background. Contour detection is then applied to extract the dominant boundary of the shape using established OpenCV-based techniques [10], [23]. Unlike probabilistic mesh-generation methods [8], the proposed approach focuses on identifying simple engineering-relevant primitives such as circles and rectangles, ensuring computational efficiency and interpretability

Once the dominant contour is identified, geometric parameters such as diameter, width, and height are estimated using bounding techniques. Circular contours are approximated using diameter calculation, while rectangular contours are measured using bounding box dimensions. These extracted parameters are mapped into predefined parametric CAD templates consistent with feature-based modeling principles [3], [18]. This deterministic mapping ensures that the generated 3D models remain editable, dimensionally consistent, and compatible with standard engineering workflows.

Fig. 3 Circle silhouette converted into a solid 3D model using the image-to-CAD module.

Fig. 4 Rectangle image converted into a 3D solid model using the image-to-CAD module.

VI. EXPERIMENTAL ANALYSIS

The experimental evaluation was conducted by testing the framework with multiple textual prompts and 2D silhouette images representing basic mechanical components such as flanges, bolts, circular plates, and rectangular profiles. For text-to-CAD generation, dimensional inputs including diameter, thickness, hole count, and pitch circle diameter were verified against the generated 3D models to assess accuracy. For image-to-CAD reconstruction, high-contrast circular and rectangular silhouettes were used to evaluate contour detection and primitive classification performance. In all test cases, the system successfully extracted geometric parameters and generated corresponding parametric models within a few seconds, demonstrating consistent real-time performance.

The results indicate that the proposed deterministic, template-based approach maintains dimensional consistency while significantly reducing modeling time compared to manual CAD construction.

Generated models remained fully editable due to their parametric structure, ensuring compatibility with standard engineering workflows. Unlike mesh-based generative systems, which may introduce geometric irregularities, the structured primitive mapping ensured clean and manufacturable geometry. Overall, the experimental analysis confirms that the framework achieves reliable automation with computational efficiency and practical engineering usability..

VII. CONCLUSION

This research presented a multimodal generative AI framework for automated mechanical design synthesis from textual descriptions and 2D sketch images. By integrating structured natural language parsing, contour-based image processing, and template-driven parametric CAD generation, the system enables rapid creation of dimensionally consistent and editable 3D models. Unlike mesh-based generative approaches, the proposed deterministic method ensures compatibility with conventional engineering workflows while maintaining computational efficiency. Experimental results demonstrate that the framework significantly reduces modeling time while preserving geometric accuracy and parametric control.

Overall, the study establishes that structured multimodal AI can effectively bridge conceptual design intent and executable CAD geometry. The combination of semantic interpretation and primitive-based reconstruction provides a practical and interpretable alternative to data-intensive neural reconstruction systems. The framework supports engineering usability through editable parametric outputs and integration with standard design environments. Future work will focus on extending the system to more complex components, assembly-level generation, and integration with optimization and simulation tools for enhanced industrial application.

REFERENCES

[1] Y. Wang, Y. Sun, Z. Liu, S. Sarma, M. Bronstein and J. Solomon, "Dynamic Graph CNN for Learning on Point Clouds," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 9, pp. 2995–3008, 2021.

[2] T. Müller, A. Evans, C. Schied and A. Keller, "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding," ACM Transactions on Graphics (SIGGRAPH), 2022.

[3] J. Shah and M. Mäntylä, Parametric and Feature-Based CAD/CAM: Concepts, Techniques and Applications, Wiley-Interscience, 1995.

[4] S. Krishnamurti and W. Wiese, "Feature-Based Modeling: Concepts and Applications," Computer-Aided Design Journal, Elsevier.

[5] M. Bronstein, J. Bruna, T. Cohen and P. Veličković, "Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges," IEEE Signal Processing Magazine, 2017.

[6] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.

[7] O. Sorkine-Hornung and M. Alexa, "As-Rigid-As-Possible Surface Modeling," Eurographics Symposium, 2007.

[8] J. Wu, C. Zhang, T. Xue, W. Freeman and J. Tenenbaum, "Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling," NeurIPS, 2016.

[9] A. Dai, C. Ruizhongtai Qi and M. Nießner, "Shape Completion Using 3D Encoder-Predictor CNNs," CVPR, 2017.

[10] OpenCV Development Team, OpenCV: Open Source Computer Vision Library Documentation, 2024.

[11] Three.js Authors, Three.js JavaScript 3D Library Documentation, 2024.

[12] S. Ramírez, FastAPI Framework Documentation, 2024.

[13] Meta Open Source Team, React JavaScript Library Documentation, 2024.

[14] OpenJS Foundation, Node.js Runtime Environment Documentation, 2024.

[15] Autodesk Research, "Generative Design in Mechanical Engineering," Autodesk University Publications.

[16] CIRP Society, "Intelligent Manufacturing Systems," CIRP Annals – Manufacturing Technology.

[17] Baltrušaitis, Ahuja and Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.

[18] R. Bidarra and W. Bronsvoort, "Semantic Feature Modeling," Computer-Aided Design, Elsevier.
Available:

[19] ISO, ISO 10303 STEP Standard for Product Data Representation, International Organization for Standardization.

[20] Amershi et al., "Guidelines for Human-AI Interaction," ACM CHI Conference, 2019.

[21] J. Lee, B. Bagheri and H. Kao, "A Cyber-Physical Systems Architecture for Industry 4.0 Manufacturing Systems," Manufacturing Letters, Elsevier.

Available:

[22] Siemens Digital Industries, Digital Engineering and Industry 4.0 Platforms.

[23] R. Szeliski, Computer Vision: Algorithms and Applications, Springer, 2022 Edition.

[24] T. La Rocca, “Knowledge-Based Engineering: Between AI and CAD,” Advanced Engineering Informatics, Elsevier.

[25] J. Cagan and K. Shea, “Generative Design and Computational Creativity,” Artificial Intelligence for Engineering Design, Analysis and Manufacturing, Cambridge University Press.

[26] OpenSCAD Community, OpenSCAD Programmer’s Solid 3D CAD Modeller Documentation, 2024.

[27] Khronos Group, WebGL Specification Documentation, 2024.

[28] Vite Development Team, Vite Frontend Build Tool Documentation, 2024.

[29] PyPI Community, Python Package Index Documentation, 2024.

[30] Russell and Norvig, Artificial Intelligence: A Modern Approach, Pearson Education.