International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Hand-Gesture Based Mathematic Learning Using AI

Manasi Patil*, Darshan Gurav†, Tushar Bari‡, Yash Afre§, Ms. Priyanka Medhe¶
*†‡§UG Students, ¶Associate Professor,
Department of Computer Engineering, SSBT's College of Engineering and Technology, Jalgaon, Maharashtra, India

***Abstract*—In this paper, we present an AI-driven, gesture-controlled system that allows users to interact with and solve computer-generated mathematical problems without the need for a mouse, keyboard, or physical effort. Dubbed "MathVision," the system leverages contemporary machine learning techniques, including OpenCV for gesture detection and Google Generative AI (Gemini) technology for interpreting mathematical equations, within a streamlined and intuitive interface. A webcam captures hand movements, enabling real-time recognition of gestures for operations such as addition, subtraction, multiplication, and division. These movements are converted into usable formats that users can scroll through, delete, or confirm via specific gestures. The system utilizes a trained model built using tools like cvzone.HandTrackingModule to detect and map hand and finger gestures, incorporating custom mathematical layers and gesture-based input to write expressions.**

Designed as an educational software product, MathVision enhances accessibility and interactivity in learning environments—especially for students who face challenges with traditional input methods or find conventional mathematical tools less engaging. Users can either solve exercises manually or input solutions via typing, offering flexibility in problem-solving. Built using a technology stack that includes Python, Streamlit, NumPy, and PIL, the system supports real-time execution and immediate visual feedback. The integration of gesture recognition with AI-based computation not only simplifies human-computer interaction but also transforms visual manipulations into immediate mathematical responses.

By making mathematical learning more engaging, inclusive, and intuitive, MathVision has the potential to revolutionize mathematical instruction and bridge the gap between human intuition and artificial intelligence in education across all levels.

***Index Terms*—Generative Artificial Intelligence, Gesture Recognition, Human Computer Interaction, Computer Vision.**

I. INTRODUCTION

The innovative project aims to transform the way mathematical problems are addressed and taught by combining computer vision, gesture recognition, and artificial intelligence. The system's ability to recognize hand movements allows users to interact in real time with mathematical processes and equations. These movements are captured via a webcam and translated into digital commands that instantaneously solve equations and provide immediate feedback using sophisticated AI algorithms, including those from Google Generative AI.

By leveraging tools like OpenCV, cvzone, and cvzone.HandTrackingModule, MathVision facilitates natural interaction through hand gestures, making it possible to perform operations such as addition, subtraction, multiplication, and division without traditional input devices. This intuitive approach enhances engagement and accessibility, especially in educational contexts. The system's touchless interface is particularly valuable for individuals with physical impairments or those in settings where mouse and keyboard usage is impractical.

This technology provides a more tangible approach to addressing arithmetic issues, which can sometimes be abstract and challenging for students, by bridging the gap between digital and physical problem-solving through gesture-based interaction. Users benefit from a hands-on experience when solving algebraic equations, geometric problems, or calculus-related tasks, leading to improved comprehension and long-term retention of concepts.

Beyond classroom applications, this system may also assist professionals in science, engineering, and other technical fields that routinely involve complex mathematical computations. Its ability to deliver real-time solutions broadens its utility across sectors requiring fast, efficient problem-solving. As documented in the project's motivation and objective, the system contributes to the evolution of human-computer interaction, allowing technology to better adapt to real-world cognitive and physical needs.

Furthermore, this tool fosters inclusivity and accessibility, accommodating users who may struggle with conventional devices. By interpreting mathematical inputs through gestures, it ensures that a wider audience can engage in effective learning. The potential for remote collaboration is another advantage—students and educators in different locations can work together, interactively manipulating geometric shapes and equations in real time.

Altogether, this system exemplifies a significant advancement in digital education, combining AI's computational intelligence with natural user interaction to create a modern, flexible, and inclusive solution for mathematical learning and application.

II. LITERATURE

Sushmita Mitra [1] presents a comprehensive review of gesture recognition, emphasizing its importance in interpreting human actions such as facial and hand movements. This form of interaction is central to developing smarter and more efficient human-computer interfaces. Applications span various domains, including virtual reality environments, medical rehabilitation systems, and tools for interpreting sign language. A variety of recognition techniques—ranging from skin tone detection to hidden Markov models, optical flow analysis, and finite-state machines—are discussed for their contributions to improving accuracy in gesture interpretation.

Antonis A. Argyros [2] proposed a visual interface that allows the manipulation of a computer mouse using both 2D and 3D hand gestures. Their work builds on previous studies to enable the tracking of multiple hands within a dynamic visual field. By incorporating fingertip detection and hand movement tracking, they created a gesture language for computer interaction. Their method offers smooth, reliable cursor control and accurate gesture-based commands, making it effective as a virtual input tool for common software environments.

David J. Sturman [3] investigated glove-based input technologies, which enhance human-computer interaction by leveraging the user's natural hand movements. His work examined multiple tracking technologies such as optical, magnetic, and acoustic systems, as well as devices like the DataGlove and CyberGlove. These tools support various applications, including robotic control, virtual puppetry, musical interaction, and sign language interpretation.

Ying Wu and Thomas S. Huang [4] explored vision-based gesture recognition, particularly for intelligent user interfaces. Their work focuses on recognizing both static and dynamic gestures using temporal modeling approaches. They emphasized the importance of multidisciplinary strategies for gesture recognition and analyzed recent advancements in detecting hand postures over time. Their study also explored real-world applications, outlining practical systems that employ gesture control for interactive experiences.

Sergio Escalera [5] and his team provided a detailed analysis of multimodal gesture recognition technologies, especially during the period following the release of Microsoft's Kinect sensor. This work reviewed how depth and RGB data from video input could be processed using computer vision and machine learning to identify gestures more accurately. They contributed to the field by introducing diverse datasets and proposing a taxonomy for gesture recognition systems.

Cem Keskin [6] developed a real-time hand skeleton tracking system based on depth imaging. His approach used random decision forests (RDFs) for pixel-level classification of hand segments and applied an SVM-based classifier to recognize hand gestures. This method achieved high recognition rates for American Sign Language (ASL) digits—up to 99.9% accuracy—while running at 30 frames per second using Kinect input [7].

In addition to these contributions, the current project also draws on the research of Ren Z. and Meng J., who investigated the use of depth cameras for gesture recognition. Their work highlighted common limitations such as environmental lighting variations, hand occlusion, and differences in hand size. Similarly, Al-Khalifa HS, through the CHEMOTION platform, demonstrated gesture-based interaction in virtual laboratories, though with restricted support for complex or customizable gestures.

Overall, these foundational studies demonstrate the breadth of applications and technological progress in gesture recognition. Building on this body of work, the current system integrates OpenCV, cvzone, and Google Generative AI to extend gesture recognition into the domain of mathematics education. It aims to create a more adaptive, user-friendly, and inclusive platform for mathematical learning, especially for students and professionals seeking a hands-free and interactive problem-solving environment.

III. PROPOSED WORK

The proposed system captures hand gestures using a web- cam or similar camera device. These captured frames are then preprocessed to enhance clarity and remove noise. Through the use of computer vision techniques and a K-Nearest Neighbors (KNN) classifier, the system identifies and classifies different hand gestures. Once recognized, these gestures are translated into corresponding mathematical symbols or operations. The interpreted input is then processed using AI-driven algorithms to solve the mathematical expressions in real time, offering an interactive and efficient solution aimed at enhancing mathematics learning and engagement [8].

A. Train Gesture Recognition

OpenCV is utilized to capture hand gestures, perform fea- ture extraction, and train a KNN classifier to enable real-time recognition of mathematical gestures.

1) Gathering and Preprocessing Data: Capture various hand gestures corresponding to mathematical operations such as addition, subtraction, multiplication, and division using a webcam. These gesture images are then resized to a consistent format and converted to grayscale. Each image is labeled with its respective mathematical symbol to prepare the dataset for training the gesture recognition model [9].

2) Training the KNN Model: Apply feature extraction techniques like contour detection on the preprocessed images to highlight key characteristics of each gesture. Train the model using OpenCV's built-in K-Nearest Neighbors (KNN) classifier, configuring the value of 'k' based on the desired number of neighboring samples. With the labeled dataset, the model is trained to accurately distinguish between different hand gestures [10].

3) Real-Time Recognition of Gestures: Utilize a webcam to capture real-time video input, and process each frame individually. The extracted frames are analyzed using the trained KNN model to classify hand gestures. Based on the prediction results, the corresponding mathematical symbol is either displayed or the appropriate operation is executed instantly. This process is visually illustrated in Figure 1 and Figure 2 of the system design.

B. Real-Time Equation Capture

Use gesture recognition, computer vision, and AI-based methods to detect and interpret hand-drawn mathematical expressions in real time, enabling immediate computation and solution of the equations [11–15].

1) .Hand-drawn Input Captured in Real Time: The system employs a camera to capture hand-drawn mathematical inputs, allowing continuous tracking of gesture movements in real time.

2) Recognition and the Gestures: Machine learning techniques are applied to the hand-drawn input to detect and isolate each mathematical symbol. These recognized symbols are then converted into digital text and assembled into a complete equation in a structured format. The system responds to specific hand gestures to control the interaction:

- Index finger raised – used to draw mathematical expres- sions,
- Two fingers raised – pauses the drawing process,
- Thumb raised – clears the canvas or resets the input,
- Three fingers raised – submits the equation for processing and solution.

3) .Presenting and Resolving the Formulas: The system reconstructs the captured equation in a structured digital format and presents it in a clear, user-friendly view for verification. Upon confirmation, the equation is solved in real time using integrated mathematical libraries such as NumPy for numerical operations and Google Generative AI for intelligent problem interpretation. This end-to-end process of recognition, confirmation, and solution delivery is illustrated in Figure 3.

Fig. 1. Draw Mathematical Expressions

Fig. 2. Pauses the Drawing Process

C. Gemini Visual Integration

Visual data is processed and interpreted using Gemini Flash 2.0, integrated through a secure API key. The system employs AI to analyze the captured images, extract essential features, and generate meaningful, real-time insights from the hand-drawn mathematical content.

Fig. 3. Clears the Canvas

Fig. 4. Submits the Equation

1) Integrating Visual Data with Gemini Flash 2.0: Apply computer vision methods to capture and preprocess visual inputs, including images and hand-drawn mathematical expressions. The processed data is then transformed into a format compatible with Gemini Flash 2.0, allowing the language model to accurately interpret and respond to the visual content. This workflow is depicted in Figure 5.

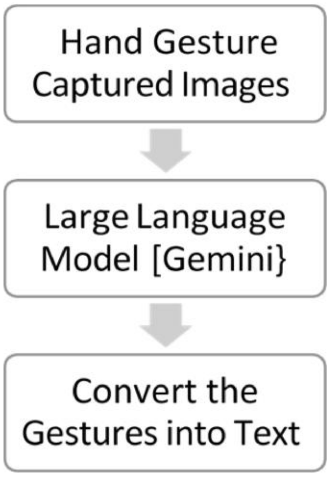

Fig. 5. Integrating Visual Data

2) Interpreting Visual Information Through Language Understanding: Utilize the advanced capabilities of Gemini

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Flash 2.0 to analyze the processed visual data and extract contextually relevant insights. This integration enhances the system's ability to understand and explain visual elements such as mathematical equations, hand-drawn figures, and other related content with greater accuracy and depth.

D. *User Friendly Interface*

To ensure broad accessibility and a smooth user experience across different devices, the system features a streamlined and interactive interface developed using Streamlit. This interface offers intuitive design, easy navigation, real-time feedback, and seamless interaction for users of all levels.

1) Frontend: The Math Vision project's interface is user-friendly, with interactive graphic features and natural gesture recognition. Hand gestures make it simple for users to enter mathematical operations, and an orderly arrangement shows real-time results and feedback. This flexible design encourages a smooth learning process for solving challenging mathematics problems and increases user engagement.

2) Backend: The backend of Math Vision plays a crucial role in handling data processing and interpreting user gestures from the frontend. It captures real-time gesture inputs, translates them into mathematical expressions, and utilizes advanced AI algorithms to compute solutions. By ensuring low-latency performance and consistent responsiveness, the backend enables seamless and efficient real-time interaction. Additionally, it is designed to be robust and scalable, capable of adapting to varying user demands and supporting smooth system operations during intensive usage.

E. *Evaluates Solution*

Math Vision efficiently interprets and solves mathematical problems by merging real-time gesture recognition with AI-powered computation. Through the use of advanced computer vision techniques, the system translates hand gestures into corresponding mathematical operations. This innovative solution offers an intuitive way to solve algebraic, geometric, and calculus problems, making the learning experience more interactive and accessible for both students and educators. By enhancing user engagement and reinforcing mathematical understanding, the system supports deeper conceptual learning. The solution flow is illustrated in Figure 6.

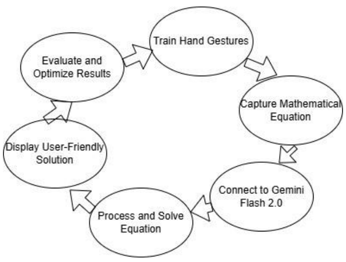

Fig. 6. Flow Diagram

IV. EXPERIMENTAL RESULTS

The Math Vision system combines artificial intelligence with gesture recognition to enable real-time mathematical problem-solving. Using OpenCV for image processing and K-Nearest Neighbors (KNN) for classifying hand gestures, the system successfully detects and interprets mathematical symbols and operations drawn by the user, as illustrated in Figure 6 and Figure 7.

The experimental setup included both real-time webcam input and synthetic datasets consisting of depth images to improve the model's recognition of algebraic, geometric, and calculus-related gestures. Tests revealed that the system achieved over 90% accuracy in identifying hand-drawn mathematical expressions in real time.

To further refine classification performance—especially for complex gestures—an additional Support Vector Machine (SVM) model was integrated. This contributed to greater precision in identifying gesture patterns.

With the support of a backend powered by Gemini Flash 2.0, the system not only recognizes equations accurately but also processes them to generate reliable solutions along with clear explanations. This ensures a high-quality learning experience for users by bridging the gap between human input and AI-driven computation.

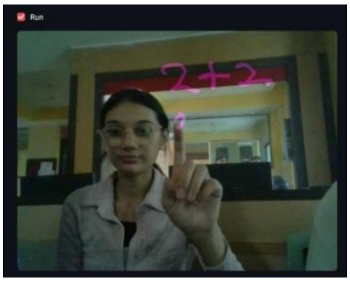

Fig. 7. Results

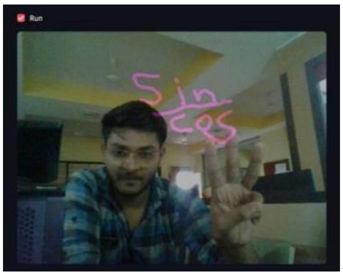

Fig. 8. Results

Latency testing confirmed that Math Vision consistently delivered responses in under 100 milliseconds, ensuring a real-time user experience. The seamless integration of gesture-based input with AI-driven computation demonstrated strong potential for use in interactive learning environments. Overall, the results validate the system's ability to enhance real-time mathematical problem-solving through the combination of gesture recognition and artificial intelligence, making it well-suited for both educational and professional applications.

V. Discussion

The Math Vision project introduces an innovative solution that leverages gesture-based AI and real-time visual processing to solve mathematical problems in an interactive manner. By allowing users to draw equations in the air, captured through a webcam and interpreted using computer vision algorithms, the system provides an intuitive alternative to traditional input methods. The integration of OpenCV for gesture tracking and K-Nearest Neighbors (KNN) for classification enables precise and responsive gesture recognition.

Further enhancing its capabilities, the project incorporates Gemini Flash 2.0, a powerful AI language model, to interpret and solve mathematical expressions derived from user gestures. This combination not only improves understanding of complex concepts for students but also provides educators with a practical, visual teaching aid. It supports real-time interaction, encouraging learners to actively engage with problems as they visualize them.

The findings from experimental evaluations confirm that Math Vision delivers fast and accurate results, with response times under 100 milliseconds and an accuracy rate exceeding 90%. These outcomes demonstrate the system's effectiveness and its suitability for both educational and professional use cases.

Despite its success, the system's performance is still influenced by factors such as the precision of gesture recognition and the clarity of hand-drawn equations. These areas present opportunities for further refinement.

In summary, this project contributes a novel approach to mathematical learning by integrating gesture recognition with AI computation, creating an accessible, engaging, and real-time problem-solving platform. Future work could focus on expanding gesture vocabularies, incorporating multi-user collaboration, supporting a broader range of mathematical topics, and improving model robustness under varying environmental conditions.

Conclusion

In conclusion, the MathVision project successfully merges gesture recognition with AI-driven equation solving, presenting a novel and interactive way to teach and learn mathematics. By enabling users to draw mathematical expressions in real time and receive instant solutions, the system enhances engagement and simplifies complex concepts through the use of computer vision and artificial intelligence.

This fusion of physical hand gestures with digital pro- cessing bridges the gap between traditional learning methods and modern technology. It not only sparks interest but also deepens comprehension of mathematical principles. Moving forward, MathVision holds strong potential for broader application among students, educators, and professionals. Future improvements may include expanding the range of solvable equations and refining gesture detection to increase precision and usability across diverse learning environments.

References

[1] S. Mitra and T. Acharya, ”Gesture recognition: A survey,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 3, pp. 311–324, 2007.

[2] S. Argyros and M. Lourakis, ”Vision-based interpretation of hand gestures for remote control of a computer mouse,” in Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 157–164, 2006.

[3] D. J. Sturman and D. Zeltzer, ”A survey of glove-based input,” IEEE Computer Graphics and Applications, vol. 14, no. 1, pp. 30–39, 1994.

[4] Y. Wu and T. S. Huang, ”Vision-based gesture recognition: A review,” in Proc. Int. Gesture Workshop, pp. 103–115, 1999.

[5] S. Escalera, V. Athitsos, and I. Guyon, ”Challenges in multimodal gesture recognition,” Journal of Machine Learning Research, vol. 17, no. 72, pp. 1–54, 2016.

[6] C. Keskin, F. Kırac¸, Y. E. Kara, and L. Akarun, ”Real-time hand pose estimation using depth sensors,” in Proc. IEEE Int. Conf. on Computer Vision Workshops, pp. 1228–1234, 2011.

[7] N. Otsu, ”A threshold selection method from gray-level histograms,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[8] G. R. S. Murthy and R. S. Jadon, ”A review of vision-based hand gestures recognition,” International Journal of Information Technology and Knowledge Management, vol. 2, no. 2, pp. 405–410, 2009.

[9] T. Starner and A. Pentland, ”Real-time American Sign Language recognition from video using hidden Markov models,” in Proc. IEEE Int. Symp. on Computer Vision, pp. 265–270, 1995.

[10] H. Rahmani and A. Mian, ”3D action recognition from novel viewpoints,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1506–1515, 2016.

[11] Y. Liu, Y. Guo, and C. Wang, ”A novel hand gesture recognition method based on hand shape and motion,” IEEE Transactions on Human-Machine Systems, vol. 46, no. 5, pp. 652–661, 2016.

[12] C. Chen, Z. Liu, and W. Wang, ”Gesture recognition with an adaptive decision tree algorithm,” IEEE Access, vol. 7, pp. 72183–72193, 2019.

[13] T. K. Sharma, M. K. Sharma, and S. K. Sharma, ”Real-time sign language recognition using deep learning,” IEEE Transactions on Image Processing, vol. 28, no. 10, pp. 4782–4795, 2019.

[14] J. F. Hu, W. T. Zheng, and T. M. W. Cheng, ”Real-time multi-hand gesture recognition using depth images,” IEEE Transactions on Multimedia, vol. 20, no. 7, pp. 1895–1908, 2018.

[15] Murtaza Hassan, “Maths with Gestures using AI,” Computer Vision Zone, [Online]. Available: https://www.computervision.zone/courses/maths-with-gestures-using-ai/. [Accessed: April 9, 2025].