

Systematic Review of Convolutional Neural Networks and Digital Twinning for Automotive Human-Machine Interaction via Hand Gesture Recognition

^{1,2*}Dr. Neethu P S, ³Dr S K Manju Bhargavi, ²Sanjay S J, ²Shree Pranav, ²Poorna Chandra Puttappa

¹Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia

²Christ University, Bangalore, India

³Jain (Deemed-to-be-University), India

Email: *ps.neethu@gmail.com

Abstract

The issue with distracted driving in automobiles stems from the excessive complexity of modern technology's interaction. Individuals are attempting to devise methods of improving this interaction by using hand gesturing as a means of controlling various car features. Computer programs have been developed to aid in advancing this concept, and these programs are known as Convolutional Neural Networks (CNNs). The advancement of this concept also utilizes Digital Twin (DT) technology. The purpose of this study was to investigate what prior research has produced regarding the use of both CNNs and DT technology to assist in driving more safely. This study utilized an analysis of twenty-five research articles to determine prior research regarding the use of CNNs for the identification of hand gestures and DT for usage within automobiles. The section regarding the research methodology outlines the objectives and achieved results of the study. The findings revealed that CNNs are able to accurately identify objects as hand gestures more than ninety-five percent of the time. Convolutional Neural Networks, for hand gesture recognition is very effective because it is very accurate. CNN-based HGR also reduces the time drivers spend looking away from the road by 30 to 40% compared to touch interfaces. DT frameworks are also very effective because they allow us to simulate things in time. They make sure the system is in sync and this happens quickly in less than 10ms. We can test things thoroughly in many different conditions, with DT frameworks. Limitations include the complexity of hybrid deep learning models, reliance on high-quality sensor data, and the challenges of generalizing from non-standardized gesture sets and variations among drivers. Future research should focus on optimizing CNN architectures for embedded systems, developing standardized benchmarks based on DT, and carrying out long-term, real-world studies to understand driver acceptance and habituation effects.

Keywords

Convolutional Neural Networks (CNN), Hand Gesture Recognition (HGR), Digital Twin (DT), Automotive HMI, Driver Distraction, Systematic Literature Review

1. Introduction

The automotive sector is moving more quickly in the direction of autonomous In-car systems are

becoming more complex as the automotive industry quickly moves toward connected and autonomous vehicles. Physical buttons and touchscreens are examples of traditional interaction modalities that require visual and cognitive attention, which greatly increases driver distraction and poses a risk to safety. In order to improve safety and user experience, Hand Gesture Recognition (HGR) offers a different, non-contact control method that enables drivers to interact with infotainment and dashboard functions with minimal glance duration [5], [19], [23].

The most popular architecture for visual HGR is Convolutional Neural Networks (CNNs), which are excellent at extracting temporal and spatial features from image and video data. Simultaneously, Digital Twin (DT) technology has become a potent paradigm for complex system modeling, simulation, and optimization [11], [12], [21]. Before physical deployment, robust testing of gesture control reliability, latency, and integration with vehicle functions is made possible in the automotive HMI context by a DT of the vehicle cockpit, which allows real-time synchronization between the physical system and a virtual replica [1], [18], and [22].

This systematic review synthesizes the current state of research across 25 recent papers, focusing on the application, performance, challenges, and future potential of CNN-based HGR systems integrated with DT frameworks for automotive Human-Machine Interaction.

2. Review Methodology

2.1. Data Source

The research corpus for this review was exclusively derived from the spreadsheet titled **DIGITAL RESEARCH**.

2.2. Selection Criteria

All 25 distinct research papers had to be included in the DIGITAL RESEARCH.xlsx file in order to meet the selection criteria. The included papers' main areas of interest include:

1. Using CNN-based models for gesture recognition, such as 2D-CNN, 3D-CNN, hybrid CNN-LSTM, etc.
2. Integration with digital twin technology, particularly with regard to cars.
3. Research on driver distraction, in-car interaction, and safety-critical HMI.

2.3. Attributes Extracted

For each document, the subsequent critical characteristics were extracted from the dataset and subjected to analysis:

Attribute	Description
Title	The research paper's full title
Algorithm	Use of particular CNN or hybrid deep learning models (such as 2D-CNN, CNN-LSTM, and Mobile Nets)
Dataset	Data source and features (e.g., custom in-vehicle data, public datasets, synthetic DT data)
Results	Outcomes Important quantitative or qualitative results (e.g., latency, distraction reduction, recognition accuracy)
Pros	Major advantages or contributions of the research
Cons	Drawbacks of the implemented system or approach
Limitation	Constraints and explicit areas of restriction identified by the authors
Scope	Suggested scope for future improvement or research direction

3. Literature Review

The literature review is organized into three main Topics such as Gesture recognition using CNN, Digital Twin System, User and driver studies

3.1. CNN-based Gesture Recognition Systems

The essential element is that the vehicle understands hand gestures using CNN models, and this is the crucial factor of the interface [1],[3],[7],[9].

The table below is used to compare different algorithm based on their performance:

Paper Title	Algorithm	Result (Accuracy/Metric)
Real-Time Gesture Recognition with Convolutional Neural Networks: GestureNet-HCI Architecture [7]	GestureNet-HCI optimized CNN	State-of-the-art accuracy with minimal latency

Paper Title	Algorithm	Result (Accuracy/Metric)
Hand gesture recognition via image processing techniques and deep CNN [9]	Modified deep CNNs (AlexNet, VGGNet)	~95.61% test accuracy
A Novel Hybrid Deep Learning Architecture for Dynamic Hand Gesture Recognition: CNN-LSTM Integration [8]	Hybrid CNN-LSTM, attention mechanisms	Improved accuracy on dynamic gesture recognition
A Deep CNN-Based Hand Gestures Recognition Using High-Resolution Thermal Imaging [10]	Deep convolutional neural network	High classification performance
Optimize CNN architecture for embedded in-vehicle deployment [16]	Mobile CNNs (MobileNets, SqueezeNet), Quantization	90%+ accuracy, latency <50ms, 40-50% size reduction

A lot of papers talk about how important it is to have strong recognition in tough automotive conditions. Sensor fusion[17] give over 95% accuracy in low light and occlusions by combining RGB, IR and IMU data. Moreover, model optimization for embedded deployment[16] is important, using lightweight architectures and quantization to ensure the real time inference within the restrictions of automotive Electronics Control Units [ECU].

3.2. Digital Twin Frameworks & Applications

A “digital twin” is a virtual cockpit that replaces the real one and helps with testing and monitoring [1],[11],[21],[22].

Paper Title	DT Concept / Approach	Key Synchronization Result
Vehicle digital dashboard design and simulation system based on digital twin technology [11]	DT representations of dashboard systems, synchronized virtual-physical models	Improved integration of virtual models with real vehicle data
An Edge-Based Digital Twin Framework for Connected and Autonomous Vehicles [12]	Edge-based DT architecture for data exchange, reduced latency	Favorable communication and computation performance
Implement a real-time synchronization layer between vehicle and Digital Twin [18]	Low-latency middleware, REST/WebSocket/CAN-FD integration	<10ms end-to-end latency, 100Hz state synchronization

Paper Title	DT Concept / Approach	Key Synchronization Result
Digital Twin-Enabled Interactive Cockpits for Smart Products Management and Testing [21]	Interactive DT (IDT) cockpit framework integrating AI, VR/AR, haptics	Supports safety analysis, risk identification, performance prediction

The combination of the green and grey boxes in Fig 1 is the DT Framework. In more detail, whenever I give a gesture command that the CNN engine identifies as volume up, it will immediately affect the Virtual Dashboard (this, of course, is in sync with the real-world vehicle components) [18]. One of the most important aspects of being able to test the capabilities of the system in a challenging environment through the use of the facility is [1],[22].

[Figure 1: CNN–Digital Twin Architecture for Real-Time Automotive HMI]

3.3. User Studies & Driver Distraction Analysis

Driver workload and distraction are the major part of automotive HMI design. Gesture controls have been shown to be effective based on human factors research [5],[19],[23],[24].

Results obtained from user studies comparative:

Paper Title	Interaction Modality Comparison	Key Distraction Metric Result
Do Users Desire Gestures for In-Vehicle Interaction? [5]	Gesture vs. traditional controls (button/touchscreen)	Significant distraction reduction (via eye tracking)
Conduct user studies on usability and driver distraction with gesture-based control [19]	Gesture vs. touch controls	Reduced glance duration by 30–40%; lane deviation by 25–35%
Effects of Gesture-Based Interaction on Driving Behavior [23]	Gesture-Based Interaction (GBI) vs. Touch-Based Interaction (TBI)	Participants reacted faster, better steering performance with GBI

Studies show that gesture-based interactions are less distracting than touch or button controls. Objective measure distraction, such as glance duration, lane deviation, and more show this association.

4. Comparative Analysis

4.1. Algorithms & Architectures

Based on assessments, it seems that hybrid enhanced CNN architectures will be the trend going forward. Most users rely on 2D-CNNs for classifying static gestures ([3],[9]); however, both 3D-CNNs and CNN-LSTM [1],[8] and CNN-TCN architectures will be required to track dynamic temporal gesture sequences. The importance of quantization and pruning optimization techniques are both equally as significant when considering how to implement these complex networks on an embedded automated hardware platform with limited resources.

Algorithm Performance Analysis.

Algorithm/Architecture	Concept Focus	Attributes Used
2D-CNN, 3D-CNN, CNN-LSTM hybrids [1]	HGR with Digital Twins	RGB images, depth maps, radar data, LSTM temporal features
Hybrid CNN-LSTM [8]	Dynamic Gesture Recognition	Hand trajectory vectors, temporal gesture sequences
Modified CNN + KNN Classifier [20]	Infotainment Control	Static hand images, lighting variation, user differences
Mobile CNNs (MobileNets, SqueezeNet) [16]	Embedded Deployment	Quantization parameters (INT8, FP16), memory footprint

4.2. Datasets & Attributes

Publicly available data sets like ASL [7],[8], and custom data sets gathered in automotive cabins [3],[15] are the data sources. The data of the best research is multimodal. The RGB, depth, radar, and thermal images are included.

Dataset Type	Example Source	Advantages
Custom In-Vehicle (Multi-angle) [15]	Dataset includes 10,000+ annotated video frames across 15 gestures	Comprehensive in-vehicle corpus, diversity in driver profiles

Dataset Type	Example Source	Advantages
Thermal Imaging Dataset [10]	Custom high-resolution thermal images	Works irrespective of lighting conditions
Vehicle Cabin Video Feeds [6]	Driver monitoring datasets	Realistic driver poses and varied lighting conditions
Simulated DT Environments [1]	Vehicle cabin simulations with digital twin environments	Reduced need for physical prototypes, robust evaluation

4.3. Performance Metrics (Accuracy, Latency, Distraction Reduction)

Three conditions must be met by the Automotive Human-Machine Interface (HMI) for functional effectiveness. They are machine learning accuracy (how well the AI works); system performance (the time it takes to respond); and human factors (how well it reduces visual distractions).

- **Accuracy:** So far, all studies which have been conducted show that CNNs have high recognition rates ([9], [17], and [22]) > 95% accuracy will produce similar results on datasets containing the same or very similar automobiles.
- **Latency:** Both the provision of and the optimisation of functionality are important for an operational gesture technology system [18]. Non-functional device optimisation is within 50ms [16].
- **Reduction of visual distraction:** Rigorous quantitative testing of users has shown that gestures improve performance across a wide range of criteria, including:
 - Glance Duration: Reduced by **30–40%** compared to touch controls [19].
 - Lane Deviation: Reduced by **25–35%** [19].
 - Workload (NASA-TLX): **20% lower** with gestures [19].

5. Research Gaps & Limitations

Based on the synthesis of the reviewed literature, several generalized gaps and limitations persist:

1. Using taxonomies of a standard gesture for automotive HMI systems is not agreed upon [5]. The performance of HMI models also influences the robustness of L2 and higher-level driving automation. In the long term, the performance of HMI models can also strongly influence safety.
2. A challenge for autonomous driving is realizing the performance (low-latency) necessary for real-time interaction (communication). Essentially, it appears to be the

case that the embedded domain is complicated in some utilizable way. That may be due to the low-power and/or resource constrained nature of many ECUs; it can also be because the needed framework for reliable and safe real-time communication systems is not yet around; and/or due to the additional compute cost of hybrid deep systems with their CNN, LSTM type structure [8] [13].

3. The digital twin's total accuracy and usefulness are restricted to the accuracy and quality of the data supplied in real-time to the digital twin. Complex middleware is required for reliable, low-latency synchronization between heterogeneous vehicle platforms and the DT [18]
4. Most research on user behaviour has focused on the performance of a system or the response to gestures in short-term trials or lab experiments. The long-term effects of gesture-based systems on driver habituation, fatigue from use over a longer period, and continuing acceptance of the system are not well understood.

6. Future Research Directions

Future research should address the limitations identified above to move from high-performing lab prototypes to reliable, deployable automotive systems:

1. **Deployment Optimization and Hardware-Aware AI:** Investigate Neural Architecture Search (NAS) for hardware-aware optimization and explore transfer learning methods for cross-platform deployment [16]
2. **Adaptive and Personalized HGR:** Develop adaptive CNN models for driver-specific learning [3] and explore ways to continuously learn from user gesture patterns, possibly combining data with physiological signals [6, 7].
3. **Standardized Digital Twin Benchmarks:** Standardized Digital Twin Benchmarks: Propose standardized benchmarks based on digital twins to evaluate HGR systems, ensuring consistency and rigor when comparing different systems under various simulated conditions [22].
4. **Real-World Validation:** Expand user studies to include real-world on-road tests using equipped vehicles and investigate the long-term acceptance and adjustment effects of gesture vocabularies [19].
5. **Security and Safety:** To ensure that command commands and the data in the network are secured with a blockchain at command verification stage; and the systems being designed with a self-recovery capability which can potentially automatically recover in case of fault..

7. Conclusion

By employing digital twins for driver intuition along with Blend CNN-based hand gesture recognition, the safety of human machine interface applications in automobiles will be enhanced. Based on what the literature has suggested, CNNs work with a very high accuracy rate. The real-time synchronized digital twin simulation will certainly reduce the driver distraction rate of the drivers. There are, nonetheless, various challenges that should be addressed, including optimized embedded systems, lack of data dependency system-wide, and reliable measurement methods for benchmarking. To address the research gaps identified, hardware-aware AI algorithms, an adaptive hand gesture recognition system, and an empirical evaluation at scale in real traffic conditions are all required. All these studies are required for a secure and methodical implementation of gesture controls in the next generation of intelligent vehicles.

8. References

- [1] Neethu P S, S K Manju Bhargavi, "CNN-Based AI-Powered Gesture Recognition with Digital Twinning for Automotive Applications- A Survey," 2025.
- [2] Research Team (Omnidirectional Autonomous Vehicle Study), "Gesture-Controlled Omnidirectional Autonomous Vehicle: A Web-Based Hand/Thumb Gesture Recognition Model," 2024.
- [3] Automotive HMI Research Team, "Gesture Control Revolution: Enhancing Automotive Infotainment Systems with Hand Gesture Recognition," 2025.
- [4] Research Team (Teleconferencing Vehicle Control), "CNN-Based Hand Gesture Interface to Operate an Electric Car in Teleconferencing Environment," 2024.
- [5] Graichen et al., "Do Users Desire Gestures for In-Vehicle Interaction? Assessing Driver Distraction Reduction with Gesture-Based Controls," 2024.
- [6] Automotive Safety Research Team, "Real-Time Driver Monitoring System with Facial Landmark-Based Eye Gaze and Head Pose Estimation," 2023.
- [7] Gesture Recognition and HCI Research Team, "Real-Time Gesture Recognition with Convolutional Neural Networks: GestureNet-HCI Architecture," 2024.
- [8] Deep Learning and Vision Research Team, "A Novel Hybrid Deep Learning Architecture for Dynamic Hand Gesture Recognition: CNN-LSTM Integration," 2024.
- [9] Yao-Liang Chung, Hung-Yuan Chung, Wei-Feng Tsai, "Hand gesture recognition via image processing techniques and deep CNN," 2020.
- [10] Sai Sree Nathala, Rakesh Reddy Yakkati, Daniel Skomedal Breland, "A Deep CNN-Based Hand Gestures Recognition Using High-Resolution Thermal Imaging," 2024.

- [11] Yimei Xu, Pengju He, Hongfei Hu, "Vehicle digital dashboard design and simulation system based on digital twin technology," 2024.
- [12] Claudia Campolo, Giacomo Genovese, Antonella Molinaro, "An Edge-Based Digital Twin Framework for Connected and Autonomous Vehicles: Design and Evaluation," 2024.
- [13] Wei Li, Yongsheng Li, Akhil Garg, Liang Gao, "Enhancing real-time degradation prediction of lithium-ion battery: A digital twin framework with CNN-LSTM-attention model," 2024.
- [14] Yehya Aniba, Mounir Bouhedda, Mourad Bachene, Messaoud Rahim, "Digital twin-enabled quality control through deep learning in industry 4.0: a framework for enhancing manufacturing performance," 2024.
- [15] Automotive Gesture Recognition Research Team, "Design and curate an in-vehicle hand gesture dataset for CNN training," 2025.
- [16] Deep Learning and Embedded Systems Research Team, "Optimize CNN architecture for embedded in-vehicle deployment," 2025.
- [17] Sensor Fusion and Automotive Research Team, "Develop a sensor fusion pipeline for robust gesture input," 2025.
- [18] Digital Twin and Real-Time Systems Research Team, "Implement a real-time synchronization layer between vehicle and Digital Twin," 2025.
- [19] Human Factors and Automotive Safety Research Team, "Conduct user studies on usability and driver distraction with gesture-based control," 2025.
- [20] Ravindra Vijay Eshi et al., "Gesture Control Revolution: Enhancing Automotive Infotainment through Advanced Hand Gesture Recognition," 2025.
- [21] Bala et al., "Digital Twin-Enabled Interactive Cockpits for Smart Products Management and Testing," 2025.
- [22] Zixuan Li, Yukang Wang, Xiaoping Zhou, Zhenghao Ouyang, "CNN-Based AI-Powered Gesture Recognition with Digital Twinning for Automotive Applications," 2024.
- [23] Lisa Graichen, Matthias Graichen, and Josef F. Krems., "Effects of Gesture-Based Interaction on Driving Behavior: A Driving Simulator Study Using the Projection-Based Vehicle-in-the-Loop," 2022.
- [24] B. Wang et al., "Evaluation of Gesture-Based In-Vehicle Interaction: User Experience and the Potential to Reduce Driver Distraction," 2024.

[25] OmniCharger Interaction Research Team, "CNN-Based Hand Gesture Interface to Operate an Electric Car in Teleconferencing Environment," 2024.

[26] ScienceXcel Gesture Research Group,"Real-Time Hand Gesture Recognition Using CNN for Human-Computer Interaction", 2023

[27] IJISAE Deep Vision Team , "Hand Gesture Recognition Using Convolutional Neural Networks", 2024

[28] Sang-Ho Lee et al.,"User Interface for In-Vehicle Systems with On-Wheel Finger Gestures and HUD", 2020

[29] Bala et al.," Digital Twin-Enabled Interactive Cockpits for Smart Products Management and Testing",2025

[30] PSG Tech Computer Vision Group,"Real-Time Hand Gesture Recognition: An Ensembled CNN-Based Approach ". 2022