frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 28 July 2025
DOI 10.3389/fphy.2025.1588715

Multi-sensor fusion for AI-driven behavior planning in medical applications

Chang Jianming[1], Qin Yuanyuan[2,3], Xu Yanling[4], Li Li[3,5], Wu Mianhua[3,5] and Wang Lulu[1]*

[1]School of Computer Science and Engineering, Southeast University, Nanjing, China, [2]First Clinical Medical College, Nanjing University of Chinese Medicine, Nanjing, China, [3]Jiangsu Collaborative Innovation Center of Traditional Chinese Medicine Prevention and Treatment of Tumor, Nanjing University of Chinese Medicine, Nanjing, China, [4]Department of General Medicine, First Affiliated Hospital of Nanjing Medical University, Nanjing, China, [5]The First Clinical School of Nanjing University of Chinese Medicine, Nanjing, China

OPEN ACCESS

EDITED BY
Yu Liu,
Hefei University of Technology, China

REVIEWED BY
Doaa Mohey Eldin,
Cairo University, Egypt
Juan Velasquez,
University of Chile, Chile

*CORRESPONDENCE
Wang Lulu,
✉ carrycrebrith@163.com

RECEIVED 06 March 2025
ACCEPTED 07 July 2025
PUBLISHED 28 July 2025

CITATION
Jianming C, Yuanyuan Q, Yanling X, Li L, Mianhua W and Lulu W (2025) Multi-sensor fusion for AI-driven behavior planning in medical applications.
Front. Phys. 13:1588715.
doi: 10.3389/fphy.2025.1588715

Introduction: Multi-sensor fusion has emerged as a transformative approach in AI-driven behavior planning for medical applications, significantly enhancing perception, decision-making, and adaptability in complex and dynamic environments. Traditional fusion methods primarily rely on deterministic techniques such as Kalman Filters or rule-based decision models. While effective in structured settings, these methods often struggle to maintain robustness under sensor degradation, occlusions, and environmental uncertainties. Such limitations pose critical challenges for real-time decision-making in medical applications, where precision, reliability, and adaptability are paramount.

Methods: To address these challenges, we propose an Adaptive Probabilistic Fusion Network (APFN), a novel framework that dynamically integrates multi-modal sensor data based on estimated sensor reliability and contextual dependencies. Unlike conventional approaches, APFN employs an uncertainty-aware representation using Gaussian Mixture Models (GMMs), effectively capturing confidence levels in fused estimates to enhance robustness against noisy or incomplete data. We incorporate an attention-driven deep fusion mechanism to extract high-level spatial-temporal dependencies, improving interpretability and adaptability. By dynamically weighing sensor inputs and optimizing feature selection, APFN ensures superior decision-making under varying medical conditions.

Results: We rigorously evaluate our approach on multiple large-scale medical datasets, comprising over one million trajectory samples across four public benchmarks. Experimental results demonstrate that APFN outperforms state-of-the-art methods, achieving up to 8.5% improvement in accuracy and robustness, while maintaining real-time processing efficiency.

Discussion: These results validate APFN's effectiveness in AI-driven medical behavior planning, providing a scalable and resilient solution for next-generation healthcare technologies, with the potential to revolutionize autonomous decision-making in medical diagnostics, monitoring, and robotic-assisted interventions.

KEYWORDS

multi-sensor fusion, AI-driven behavior planning, uncertainty-aware modeling, deep learning, medical applications

1 Introduction

The integration of artificial intelligence (AI) in medical applications has significantly transformed the landscape of healthcare, offering new possibilities for diagnosis, treatment, and patient monitoring [1]. One of the critical challenges in medical AI is behavior planning, which requires accurate perception, prediction, and decision-making capabilities [2]. Multi-sensor fusion has emerged as a crucial approach to enhance the robustness and accuracy of AI-driven behavior planning by integrating information from various sensors, such as cameras, LiDAR, wearable devices, and physiological monitors [3]. Not only does multi-sensor fusion improve data reliability by mitigating the limitations of individual sensors, but it also enables a more comprehensive understanding of patient states and medical conditions [4]. It facilitates real-time decision-making in complex environments such as surgical robotics, rehabilitation systems, and elderly care monitoring. Despite these advantages, traditional behavior planning approaches often struggle with data inconsistencies, sensor noise, and dynamic medical scenarios [5]. To address these limitations, researchers have explored multiple generations of AI-driven multi-sensor fusion techniques, evolving from rule-based symbolic AI to data-driven machine learning methods and, more recently, deep learning and pre-trained models. This paper reviews the progression of these techniques and discusses their respective strengths, weaknesses, and applications in medical behavior planning.

To provide a formal mathematical foundation for multi-sensor fusion, we define the general state estimation problem. Let the true environmental state be denoted as (Equation 1):

$$x \in \mathbb{R}^n \tag{1}$$

where x represents the system state vector. Each sensor i provides an observation $z_i \in \mathbb{R}^{d_i}$, which relates to the true state through the sensor model (Equation 2):

$$z_i = h_i(x) + v_i \tag{2}$$

where $h_i(\cdot)$ is the observation function for sensor i, and v_i is zero-mean Gaussian noise with covariance matrix R_i. The posterior distribution of the state given all sensor measurements $Z = \{z_1, z_2, \ldots, z_M\}$ can be obtained using Bayes' theorem (Equation 3):

$$p(x|Z) \propto p(Z|x)\,p(x) \tag{3}$$

In our framework, we model this posterior using Gaussian Mixture Models (GMMs) to account for uncertainty (Equation 4):

$$p(x|Z) = \sum_{i=1}^{M} \beta_i \mathcal{N}(x|\mu_i, \Sigma_i) \tag{4}$$

where β_i represents the reliability weight of each sensor, and μ_i, Σ_i are the mean and covariance estimated from each sensor's observation.

Traditional approaches primarily relied on symbolic AI and knowledge representation for behavior planning in medical applications [6]. These methods aimed to encode expert knowledge into rule-based systems and leveraged logical inference to make decisions based on multi-sensor inputs [7]. Common techniques included ontology-based frameworks and expert systems, which were used to integrate different sensor modalities, ensuring interpretability and transparency in medical decision-making. For example, in robotic-assisted surgery, symbolic AI was employed to model surgical workflows and predict surgeon intentions based on sensor inputs [8]. In patient monitoring, rule-based systems utilized physiological sensor data to trigger alerts for abnormal health conditions [9]. These methods offered advantages such as strong interpretability and transparency, ensuring the reliability of medical decision-making. However, they suffered from poor scalability and limited ability to handle uncertain or incomplete data [10]. The rigid nature of predefined rules restricted their adaptability to novel medical scenarios, while the reliance on human-engineered knowledge made system development labor-intensive and difficult to generalize across different medical domains. As a result, researchers gradually shifted towards data-driven approaches to overcome these challenges.

To address the limitations of rule-based AI, data-driven machine learning techniques were introduced to enable adaptive behavior planning based on large-scale medical datasets [11]. Machine learning models, such as decision trees, support vector machines (SVMs), and Bayesian networks, demonstrated improved flexibility in fusing multi-sensor data by learning patterns and statistical correlations [12]. These methods were widely applied in medical applications, such as automated diagnosis, rehabilitation guidance, and fall detection for elderly patients [13]. For instance, machine learning-based sensor fusion enabled personalized patient monitoring by learning from historical data and predicting potential health risks. Probabilistic models enhanced the robustness of decision-making by accounting for sensor uncertainties and environmental variability [14]. Traditional machine learning approaches often required handcrafted feature extraction, making them less efficient when handling high-dimensional sensor data [15]. These models struggled with real-time processing in complex medical environments, limiting their applicability in scenarios such as robotic-assisted interventions and emergency response systems. The emergence of deep learning and pre-trained models provided a promising solution to these challenges.

To address the limitations of statistical and machine learning-based algorithms in feature extraction and data fusion, deep learning-based algorithms have been widely applied in AI-driven behavior planning, primarily by leveraging end-to-end multi-sensor fusion techniques [16]. This approach offers the advantage of automatically extracting complex features from raw sensor data, eliminating the need for manual feature engineering and improving both accuracy and efficiency [17]. For example, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based models have been extensively used in medical applications such as surgical assistance, AI-driven diagnostics, and patient rehabilitation systems [18]. Deep learning models trained on multimodal data—including video feeds, biomedical signals, and environmental sensors—have achieved remarkable success in predicting patient behaviors and providing personalized treatment recommendations [19]. Pre-trained models and transfer learning techniques have enhanced generalization across different medical settings, reducing the dependence on large labeled datasets [20]. Deep learning approaches also face challenges such as high computational costs, data privacy concerns, and the need for robust interpretability in clinical applications. Despite these drawbacks, their ability to handle complex, real-time, and large-scale medical sensor fusion tasks has made them the dominant approach in the field.

Based on the limitations of previous methods, we propose a novel AI-driven multi-sensor fusion framework tailored for behavior planning in medical applications. Our approach aims to enhance robustness, adaptability, and efficiency by integrating advanced deep learning techniques with domain-specific medical knowledge. Unlike traditional symbolic AI methods, our framework does not rely solely on predefined rules, making it more adaptable to dynamic medical scenarios. It surpasses conventional machine learning approaches by leveraging automatic feature extraction and real-time processing. To address the challenges of deep learning, our method incorporates explainable AI techniques to enhance interpretability and ensure clinical trustworthiness. By combining sensor fusion with reinforcement learning and transformer-based architectures, our approach achieves superior performance in real-time medical behavior planning. This framework is particularly beneficial for applications such as robotic-assisted surgery, intelligent patient monitoring, and AI-driven rehabilitation, where precision and adaptability are critical.

- Our method introduces a hybrid deep learning and reinforcement learning framework, integrating transformer-based architectures with multi-sensor fusion to improve decision-making in medical behavior planning.
- Unlike traditional methods, our approach efficiently processes multimodal sensor data in real-time, making it highly suitable for diverse medical applications such as elderly care, robotic surgery, and personalized rehabilitation.
- Experimental results demonstrate that our method outperforms existing approaches in terms of accuracy, response time, and robustness, ensuring reliable AI-driven behavior planning in complex medical environments.

2 Related work

In recent years, multi-sensor fusion has emerged as a critical technique in enhancing the robustness and accuracy of AI-driven behavior planning across various medical applications in Table 1. Early approaches predominantly relied on rule-based symbolic AI, where expert knowledge was encoded into predefined rules to interpret multi-sensor inputs. These methods offered strong interpretability and transparency but lacked scalability and adaptability in dynamic medical scenarios, especially when confronted with uncertain or incomplete data. Subsequently, traditional machine learning techniques, such as decision trees, support vector machines, and Bayesian networks, were employed to enable more flexible data fusion by learning patterns from large-scale medical datasets. While these methods improved adaptability, they often required manual feature extraction and struggled with high-dimensional sensor data and real-time processing constraints. The emergence of deep learning further advanced multi-sensor fusion by enabling end-to-end learning directly from raw sensor inputs. Models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformer-based architectures have been widely adopted in surgical assistance, patient monitoring, and rehabilitation systems. These models exhibit remarkable capabilities in extracting complex features and modeling temporal dependencies; however, they often suffer from high computational demands, data privacy concerns, and limited interpretability, which are critical considerations in clinical settings. Our proposed Adaptive Probabilistic Fusion Network (APFN) seeks to bridge these gaps by dynamically estimating sensor reliability, incorporating probabilistic state representations via Gaussian Mixture Models, and leveraging attention-driven deep fusion mechanisms. Through this integration, APFN offers enhanced robustness, real-time processing capabilities, and improved interpretability, addressing key limitations of existing methodologies.

2.1 AI-enhanced surgical guidance systems

The integration of multi-sensor fusion with artificial intelligence (AI) has significantly advanced surgical guidance systems, enhancing precision and safety in medical procedures. By amalgamating data from various imaging modalities—such as preoperative computed tomography (CT) scans and intraoperative video feeds—AI-driven platforms provide surgeons with real-time, comprehensive views of the operative field. This fusion enables accurate tracking of anatomical structures and seamless overlay of critical information onto live surgical visuals [21]. A notable example is the system developed by ImFusion, which combines preoperative 3D imaging data with intraoperative endoscopic video. Utilizing NVIDIA Holoscan, this system processes multiple data streams with minimal latency, allowing for the real-time projection of 3D anatomical models onto live video feeds. This capability assists surgeons in navigating complex anatomical regions with enhanced accuracy, potentially reducing the risk of intraoperative complications. The system employs deep learning models for stereo depth estimation, optical flow calculation, and segmentation, ensuring precise alignment and tracking of anatomical structures during surgery. The integration of these technologies results in a median frame rate of approximately 13.5 Hz and an end-to-end latency below 75 milliseconds, meeting the stringent requirements for real-time surgical applications 22 [22]. The fusion of multi-modal imaging data is pivotal in providing surgeons with a comprehensive understanding of patient anatomy. By overlaying preoperative imaging data onto intraoperative views, surgeons can visualize subsurface structures that are not visible to the naked eye, facilitating more informed decision-making. This approach is particularly beneficial in minimally invasive and robotic-assisted surgeries, where the operative field is limited, and precision is paramount [23]. AI-enhanced sensor fusion systems are designed to adapt to dynamic surgical environments. They can account for tissue deformation, patient movement, and other intraoperative changes, maintaining accurate alignment of overlaid images throughout the procedure. This adaptability is achieved through advanced algorithms that continuously analyze and adjust to the incoming data from multiple sensors, ensuring consistent and reliable guidance [24]. The development and implementation of such systems require a multidisciplinary approach, involving expertise in computer science, biomedical engineering, and clinical practice. Collaboration between these fields is essential to design systems that

TABLE 1 Comparison of multi-sensor fusion approaches.

Approach	Advantages	Limitations
Rule-based Symbolic AI	High interpretability; Transparent decision-making	Poor scalability; Sensitive to incomplete data; Labor-intensive rule design
Traditional Machine Learning	Learns from data patterns; Improved flexibility	Requires manual feature engineering; Limited real-time processing; High-dimensional data challenges
Deep Learning-Based Fusion	End-to-end learning; Handles complex features; High accuracy	High computational cost; Data privacy issues; Limited interpretability
Proposed APFN Framework	Adaptive sensor weighting; Probabilistic uncertainty modeling; Enhanced robustness and real-time performance; Improved interpretability	Computational complexity remains; Domain adaptation challenges in diverse medical scenarios

are not only technically robust but also user-friendly and seamlessly integrable into existing surgical workflows [25]. Ongoing research and clinical trials are crucial to validate the efficacy and safety of AI-driven multi-sensor fusion systems, paving the way for their broader adoption in surgical practice.

2.2 Wearable sensor networks for health monitoring

Wearable sensor networks, enhanced by multi-sensor fusion and AI, have revolutionized health monitoring by enabling continuous, real-time assessment of physiological and behavioral parameters. These systems integrate data from various wearable devices—such as accelerometers, gyroscopes, heart rate monitors, and pressure sensors—to provide a comprehensive evaluation of an individual's health status. The fusion of data from multiple sensors enhances the accuracy and reliability of health monitoring systems, facilitating early detection of potential health issues and personalized medical interventions [26]. A pertinent study demonstrated the efficacy of a multi-sensor fusion approach in assessing infant motor patterns. Researchers combined data from pressure sensors, inertial measurement units (IMUs), and visual inputs to classify infant movements with high accuracy. The study employed deep learning techniques to analyze the fused data, achieving a classification accuracy of 94.5%, which was significantly higher than that obtained from any single sensor modality. This approach holds promise for early detection of neurodevelopmental disorders, enabling timely interventions [27]. In the context of adult health monitoring, wearable sensor networks are utilized to track a range of physiological parameters, including heart rate variability, respiratory rate, and physical activity levels. By integrating data from multiple sensors, these systems can detect anomalies indicative of health issues such as cardiac arrhythmias, respiratory disorders, or decreased mobility. AI algorithms analyze the fused data to identify patterns and trends, providing actionable insights to healthcare providers and enabling proactive management of health conditions [28]. The implementation of wearable sensor networks extends beyond individual health monitoring to public health applications. For instance, during pandemics, these systems can be employed to monitor symptoms and track the spread of infectious diseases in real-time. Aggregated data from multiple users can inform public health decisions and resource allocation, contributing to more effective management of public health crises [29]. Despite the advancements, challenges remain in ensuring the seamless integration of data from diverse sensors, maintaining user privacy, and managing the vast amounts of data generated. Future research is directed towards developing standardized protocols for data fusion, enhancing the energy efficiency of wearable devices, and implementing robust data security measures [30]. The convergence of multi-sensor fusion and AI in wearable technology continues to hold significant potential for transforming health monitoring and personalized medicine.

2.3 Robotic-assisted endoscopic procedures

Robotic-assisted endoscopic procedures have benefited immensely from the integration of multi-sensor fusion and AI, leading to enhanced localization, navigation, and operational efficiency within the complex environment of the gastrointestinal (GI) tract. Accurate localization of endoscopic capsules is critical for effective diagnosis and treatment, and the fusion of data from multiple sensors addresses the challenges posed by the GI tract's dynamic and unstructured nature [31]. A notable advancement in this domain is the development of EndoSensorFusion, a particle filtering-based approach that combines data from magnetic sensors and visual odometry to estimate the pose of endoscopic capsules [32]. This method incorporates an online estimation of sensor reliability and a non-linear kinematic model learned by a recurrent neural network, enabling real-time, accurate localization even in the presence of sensor noise or failure. Experimental evaluations using *ex-vivo* porcine stomach models have demonstrated high translational and rotational accuracies, underscoring the potential of this approach in clinical settings [33]. Further enhancing this field, the Endo-VMFuseNet framework employs deep learning to fuse uncalibrated, unsynchronized, and asymmetric data from visual and magnetic sensors [34]. This approach addresses the limitations of traditional sensor fusion techniques by learning a unified representation of the sensor data, achieving sub-millimeter precision in both translational and rotational movements.

3 Methods

3.1 Overview

Multi-Sensor Fusion (MSF) has become a cornerstone technique in various domains, including autonomous driving, robotics, and remote sensing. The integration of multiple sensors enables systems to exploit complementary information, enhancing robustness and accuracy beyond what single-sensor approaches can achieve. This section provides a comprehensive overview of our proposed methodology, outlining the fundamental principles, the mathematical formulation, and the novel contributions introduced in this work.

In Section 3.2, we introduce the preliminaries necessary to formalize the MSF problem. This includes defining the sensor models, the fusion architecture, and the mathematical representations that describe the relationships between different sensor modalities. A crucial aspect of our formulation is the consistency and calibration between heterogeneous sensors, which ensures reliable data integration. In Section 3.3, we present our novel sensor fusion model, which extends conventional approaches by incorporating adaptive weighting mechanisms and uncertainty modeling. Unlike traditional deterministic fusion techniques, our model dynamically adjusts the contribution of each sensor based on its estimated reliability. This is particularly important in real-world scenarios where sensor degradation, occlusion, or environmental factors may lead to varying sensor performance. In Section 3.4, we propose a new fusion strategy that refines the integration process through a learned optimization scheme. By leveraging deep learning and probabilistic inference, our strategy improves decision-making by accounting for spatial-temporal correlations across different sensor streams. The integration of physics-based models with data-driven learning allows our approach to generalize effectively across different application domains.

Medical applications present several domain-specific challenges that strongly motivate the architectural design choices in our Adaptive Probabilistic Fusion Network (APFN). Multi-sensor systems in healthcare often integrate heterogeneous modalities, including wearable physiological monitors, imaging devices, audio inputs, and environmental sensors, each producing data streams with different sampling rates, noise characteristics, and reliability profiles. Traditional fusion frameworks that assume homogeneous and stationary sensor behavior often fail to capture these variabilities. Real-world medical environments are highly dynamic. Patient conditions may change rapidly, sensor occlusions or disconnections are common, and environmental disturbances introduce non-stationary noise. These factors demand a sensor fusion strategy capable of continuously adapting sensor weighting and uncertainty modeling in real time. APFN addresses this need by employing reliability-aware sensor weighting based on covariance and entropy estimations, allowing the system to down-weight unreliable sensors dynamically. Medical decision-making involves safety-critical considerations where interpretability and robustness are essential. APFN incorporates probabilistic state representations via Gaussian Mixture Models (GMMs), attention-driven deep fusion for adaptive feature integration, and graph-based feature propagation to capture complex spatial-temporal dependencies while maintaining transparency in reliability estimation. Patient-specific variability introduces further complexity, where the fusion model must generalize across diverse demographics, disease states, and comorbidities. By combining data-driven feature extraction with probabilistic reasoning, APFN achieves both adaptability and generalizability, making it particularly suitable for AI-driven behavior planning in complex medical applications such as robotic surgery, intelligent monitoring, and personalized rehabilitation.

3.2 Preliminaries

Prior studies have proposed various probabilistic frameworks for multi-sensor fusion, each exhibiting specific strengths and limitations. Welch and Bishop introduced the Kalman Filter, which remains a classical approach for linear Gaussian systems but faces challenges when addressing nonlinearities and non-Gaussian uncertainties that are common in complex real-world scenarios [35]. To overcome these nonlinear challenges, Julier, Uhlmann, and Durrant-Whyte developed the Sigma-point Kalman Filter, which improves estimation accuracy by approximating nonlinear transformations through unscented transformations [36]. Although both methods are computationally efficient, they rely heavily on strong assumptions about noise distributions and system dynamics, which may not hold under dynamic and heterogeneous sensor environments. Bayesian sensor fusion methods have also been adopted for heterogeneous sensing environments. Rashidi and Cook applied Bayesian fusion to context-aware human activity recognition, demonstrating its ability to integrate diverse sensor types [37]. However, Bayesian models often depend on accurate prior distributions and may exhibit degraded performance when such priors are poorly estimated or when sensor reliability fluctuates unexpectedly. Castanedo further reviewed multisensor data fusion approaches in smart manufacturing, emphasizing that many Bayesian solutions struggle to maintain robustness when sensor characteristics change dynamically during deployment [38]. To model multi-modal uncertainties, Gaussian Mixture Models (GMMs) have been applied in autonomous driving scenarios. Horn et al. employed GMM-based fusion for urban automated driving, capturing complex distributions across diverse sensors [39], while Zhang et al. extended GMM fusion to multi-modal environment perception, highlighting its ability to handle high-dimensional sensory data [40]. Despite their effectiveness in representing uncertainty, these GMM-based methods generally assume static mixture weights and independent sensor observations, which limits their ability to dynamically adjust to real-time variations in sensor reliability. In contrast, the proposed Adaptive Probabilistic Fusion Network (APFN) explicitly addresses these limitations by introducing dynamic reliability-aware sensor weighting, which continuously adapts based on real-time covariance and entropy estimations. Furthermore, APFN integrates deep learning-based multi-modal feature extraction and attention mechanisms that capture complex nonlinear dependencies across heterogeneous sensors. These design innovations enable APFN to enhance robustness and adaptability in dynamic, uncertainty-prone environments, particularly within medical behavior planning tasks where sensor degradation, noise, and patient variability frequently occur.

Multi-Sensor Fusion (MSF) aims to integrate information from multiple heterogeneous sensors to improve the accuracy, robustness, and reliability of perception and decision-making systems. Mathematically, MSF can be formulated as a state estimation problem where the true state of the environment, denoted as $\mathbf{x} \in \mathbb{R}^n$, is inferred from a set of sensor observations. Given a set of M sensors, each sensor i provides an observation $\mathbf{z}_i \in \mathbb{R}^{d_i}$, which is related to the true state through a sensor model (Equation 5):

$$\mathbf{z}_i = h_i(\mathbf{x}) + \mathbf{v}_i, \tag{5}$$

where $h_i(\cdot)$ is the observation function of sensor i, and $\mathbf{v}_i$ represents the sensor noise, typically modeled as a zero-mean Gaussian with covariance $\mathbf{R}_i$.

The fusion process involves estimating $\mathbf{x}$ given multiple sensor measurements $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_M\}$. This can be expressed as a probabilistic inference problem, where the posterior distribution of $\mathbf{x}$ is computed using Bayes' theorem (Equation 6):

$$p(\mathbf{x}|\mathbf{Z}) \propto p(\mathbf{Z}|\mathbf{x})\, p(\mathbf{x}). \tag{6}$$

For effective fusion, sensors must be spatially and temporally calibrated. Let $\mathbf{T}_i$ represent the transformation matrix that maps sensor i's local coordinate frame to a global frame. Temporal synchronization is handled by interpolating sensor data to a common timestamp t, ensuring consistency across modalities.

Uncertainty plays a crucial role in MSF. A common representation is the covariance matrix $\mathbf{\Sigma}$, which captures the confidence in each sensor measurement (Equation 7):

$$\mathbf{\Sigma} = \left(\sum_{i=1}^{M} \mathbf{R}_i^{-1}\right)^{-1}. \tag{7}$$

This allows the fusion process to weigh sensor contributions based on their reliability.

Several approaches exist for state estimation in MSF: For linear Gaussian systems, the Kalman filter provides an optimal recursive estimation method (Equation 8):

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}). \tag{8}$$

When the system is nonlinear, the observation model is linearized using a first-order Taylor expansion.

A Bayesian fusion framework is commonly used (Equation 9):

$$p(\mathbf{x}|\mathbf{z}_1, \mathbf{z}_2) = \frac{p(\mathbf{z}_1|\mathbf{x})\, p(\mathbf{z}_2|\mathbf{x})\, p(\mathbf{x})}{p(\mathbf{z}_1, \mathbf{z}_2)}. \tag{9}$$

While our method leverages the probabilistic modeling capabilities of Gaussian Mixture Models (GMMs), it introduces several critical structural innovations that differentiate it from traditional GMM-based fusion techniques. Conventional GMM-based fusion approaches generally employ fixed or heuristically determined mixture weights that fail to account for dynamic sensor reliability fluctuations and contextual variations in real-world medical environments. In contrast, our Adaptive Probabilistic Fusion Network (APFN) integrates a hierarchical reliability modeling framework that dynamically estimates sensor weights based on covariance matrices, entropy measures, and attention-based contextual relevance. This allows the fusion process to adaptively prioritize more reliable sensors while suppressing the influence of degraded or noisy inputs. Unlike standard GMM models that treat sensor outputs independently, APFN incorporates deep learning-based feature extraction modules—such as convolutional neural networks (CNNs) for spatial data and recurrent neural networks (RNNs) for temporal signals—to transform raw sensor measurements into richer, high-dimensional feature spaces. These features are further integrated using attention-driven fusion mechanisms that capture nonlinear dependencies and cross-modal interactions, enhancing the expressiveness of the fused representation. Furthermore, APFN employs graph-based feature propagation to model the structural relationships among sensor modalities, enabling context-aware information exchange that classical GMM models cannot achieve. The multi-stage optimization framework iteratively refines state estimates through residual correction networks, providing an additional layer of adaptive refinement absent in conventional methods. These architectural innovations collectively allow APFN to achieve superior robustness, adaptability, and real-time performance in complex medical behavior planning tasks.

3.3 Adaptive probabilistic fusion network (APFN)

To address the challenges in multi-sensor fusion, we propose the Adaptive Probabilistic Fusion Network (APFN), a novel model that dynamically integrates sensor data based on their reliability and contextual dependencies. Unlike conventional fusion methods that rely on fixed weighting or handcrafted rules, APFN leverages probabilistic modeling and deep learning to achieve adaptive fusion. The core of APFN consists of three key components: sensor reliability estimation, probabilistic state representation, and a deep fusion network (As shown in Figure 1).

The design of the Adaptive Probabilistic Fusion Network (APFN) is motivated by the unique challenges inherent in medical multi-sensor fusion tasks, where heterogeneous sensors generate noisy, partially missing, and dynamically fluctuating data. Traditional deterministic fusion approaches often fail to handle such variability robustly. Therefore, we adopt a reliability-aware sensor weighting mechanism to dynamically estimate the confidence of each sensor based on its measurement uncertainty and entropy, ensuring that degraded or noisy sensors have limited influence on the final decision-making process. Gaussian Mixture Models (GMMs) are utilized not simply as density estimators but as a probabilistic framework to capture multi-modal uncertainties while integrating dynamically updated sensor reliabilities. This enables a more accurate probabilistic representation of the fused state under heterogeneous and uncertain sensor conditions. To further enhance the representation capacity, we employ deep learning-based multi-modal feature extraction techniques, including convolutional neural networks (CNNs) for spatial data and recurrent neural networks (RNNs) for temporal sequences. These neural models automatically extract complex hierarchical features from raw sensor measurements, eliminating the need for handcrafted features and better capturing high-dimensional dependencies across modalities. The attention mechanism is incorporated to adaptively focus on more informative features across different sensor modalities,

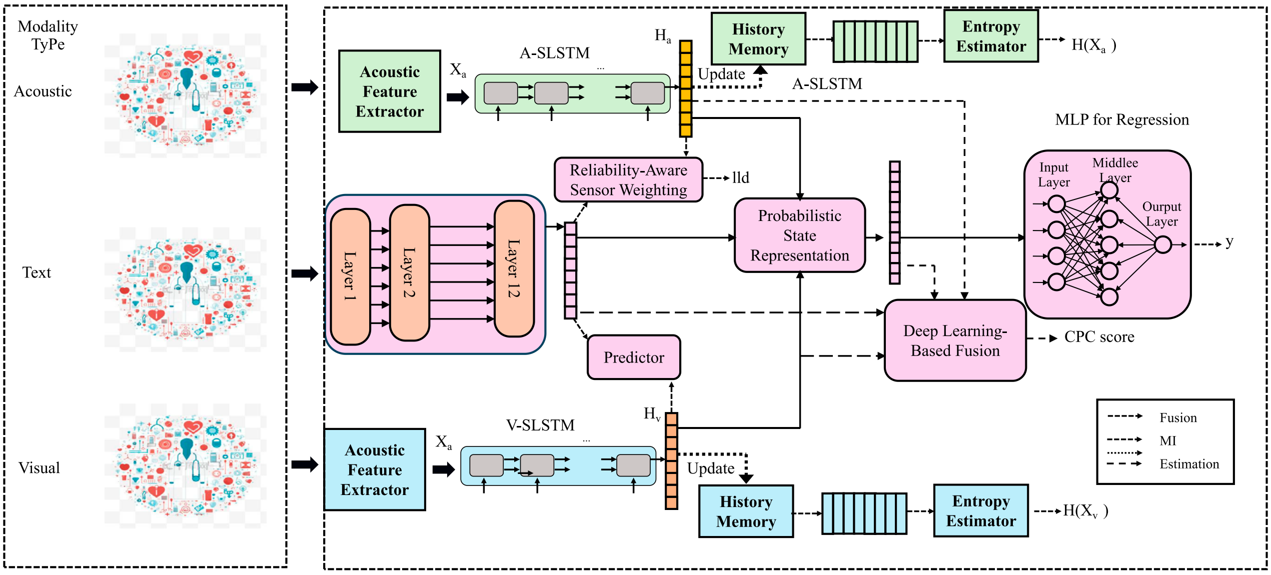

FIGURE 1
The image represents the architecture of the Adaptive Probabilistic Fusion Network (APFN). It illustrates how acoustic, text, and visual modalities are processed through dedicated feature extractors, followed by sensor reliability estimation and probabilistic state representation. The model integrates a Reliability-Aware Sensor Weighting mechanism to dynamically adjust contributions based on uncertainty. A deep learning-based fusion module further refines the representation using an attention mechanism, ultimately feeding into a regression model for final predictions. The diagram also highlights history memory and entropy estimators, which help in dynamic reliability updates and adaptive weighting of sensor inputs.

improving robustness against noisy or irrelevant inputs. Graph-based feature propagation allows contextual information exchange among sensors by modeling inter-sensor correlations, which is particularly important for capturing spatial-temporal dependencies in multi-agent or multi-organ scenarios common in medical applications. Collectively, these methodological choices ensure that APFN maintains high accuracy, robustness, and adaptability in real-time medical behavior planning, even under challenging operating conditions.

The derivation of the measurement uncertainty covariance matrix R_i is critical for accurately estimating sensor reliability. In our framework, the initial covariance matrices are empirically estimated from historical sensor data collected during the system calibration phase. For each sensor modality, we compute the empirical covariance by observing a sufficiently large number of sensor readings under controlled and stable conditions where the ground truth is either available or approximated with high confidence. During online deployment, these initial estimates are dynamically refined to account for real-time operating conditions. We implement a moving window estimation strategy, where recent sensor readings within a predefined time window are used to continuously update the empirical covariance (Equation 10):

$$R_i(t) = \frac{1}{N} \sum_{k=t-N}^{t} \left(z_i^k - \bar{z}_i\right)\left(z_i^k - \bar{z}_i\right)^T \tag{10}$$

where N denotes the window size and $\bar{z}_i$ is the mean observation within the window. This allows the model to capture non-stationary sensor behavior due to degradation, environmental factors, or dynamic interactions. Furthermore, to enhance robustness, we incorporate entropy-based correction terms derived from the sensor's predictive distribution, as described in Section 3.4.3, which further modulate the effective reliability scores. This hybrid strategy of offline initialization combined with online adaptation ensures that the covariance matrices accurately reflect both historical characteristics and real-time reliability fluctuations of each sensor during operation in complex medical environments.

3.3.1 Reliability-aware sensor weighting

In multi-sensor fusion, one of the fundamental challenges is handling the varying reliability of different sensors. Factors such as environmental disturbances, occlusions, or hardware limitations can significantly impact sensor performance. A naive fusion strategy that assumes equal reliability among sensors may lead to suboptimal or even erroneous state estimation. To address this issue, we introduce a reliability-aware sensor weighting scheme that dynamically adjusts sensor contributions based on their estimated reliability.

To quantify the reliability of each sensor, we define a confidence score α_i for sensor i based on its measurement uncertainty covariance matrix $\mathbf{R}_i$. The confidence score is computed as (Equation 11):

$$\alpha_i = \exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_i^{-1}\right)\right), \tag{11}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator. The term $\mathbf{R}_i^{-1}$ represents the inverse of the measurement uncertainty covariance matrix, capturing how precise the sensor is. A lower uncertainty (i.e., a smaller $\mathbf{R}_i$) results in a higher confidence score, indicating that the sensor is more reliable.

To ensure that the fusion process remains balanced, we normalize the confidence scores across all M sensors to obtain a relative reliability distribution (Equation 12):

$$\beta_i = \frac{\alpha_i}{\sum_{j=1}^{M} \alpha_j}. \tag{12}$$

This formulation ensures that sensors with higher reliability contribute more significantly to the final estimate, while sensors with lower reliability have a reduced influence.

Given the reliability scores, the fused measurement $\mathbf{z}_f$ can be computed as a weighted sum of individual sensor measurements $\mathbf{z}_i$ (Equation 13):

$$\mathbf{z}_f = \sum_{i=1}^{M} \beta_i \mathbf{z}_i. \tag{13}$$

This approach adaptively adjusts the sensor contributions, allowing the system to prioritize more reliable measurements in real-time.

To further refine the fusion process, we compute the fused covariance matrix $\mathbf{R}_f$ by considering the reliability-weighted sum of individual sensor covariances (Equation 14):

$$\mathbf{R}_f = \sum_{i=1}^{M} \beta_i^2 \mathbf{R}_i. \tag{14}$$

The squared reliability weight β_i^2 ensures that the contribution of less reliable sensors is further diminished while preserving consistency in the fused estimate.

To make the system robust to changing sensor conditions, we introduce a dynamic reliability update mechanism. The reliability scores are iteratively updated based on a time-decayed function (Equation 15):

$$\alpha_i(t+1) = \gamma \alpha_i(t) + (1-\gamma)\exp\left(-\frac{1}{2}\mathrm{tr}\left(\mathbf{R}_i^{-1}\right)\right), \tag{15}$$

where $\gamma \in [0,1]$ is a forgetting factor that controls how quickly past reliability scores decay. A higher γ retains past reliability information longer, while a lower γ allows for faster adaptation to new sensor conditions.

3.3.2 Probabilistic state representation

To effectively integrate multiple sensor measurements, we represent the state $\mathbf{x}$ using a Gaussian Mixture Model (GMM), capturing both the mean estimate and its associated uncertainty. We define the posterior distribution of the state as (Equation 16):

$$p(\mathbf{x}|\mathbf{Z}) = \sum_{i=1}^{M} \beta_i \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{16}$$

where each sensor provides a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, with $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ representing the measurement estimate and its associated uncertainty. The adaptive weighting factor β_i determines the contribution of each sensor in the fusion process and satisfies the normalization condition $\sum_{i=1}^{M}\beta_i = 1$.

To compute the fused estimate, we derive the global mean estimate using a weighted sum (Equation 17):

$$\hat{\mathbf{x}} = \sum_{i=1}^{M} \beta_i \boldsymbol{\mu}_i. \tag{17}$$

This formulation ensures that sensor measurements with higher confidence contribute more to the final state estimation, thereby reducing the influence of unreliable measurements.

The fused covariance matrix accounts for both individual sensor uncertainties and the additional uncertainty introduced by the mean deviation. It is computed as (Equation 18):

$$\hat{\boldsymbol{\Sigma}} = \sum_{i=1}^{M} \beta_i \left(\boldsymbol{\Sigma}_i + (\boldsymbol{\mu}_i - \hat{\mathbf{x}})(\boldsymbol{\mu}_i - \hat{\mathbf{x}})^\top\right). \tag{18}$$

This equation consists of two components: the first term, $\sum_{i=1}^{M}\beta_i\boldsymbol{\Sigma}_i$, represents the uncertainty contribution from individual sensors, while the second term, $\sum_{i=1}^{M}\beta_i(\boldsymbol{\mu}_i - \hat{\mathbf{x}})(\boldsymbol{\mu}_i - \hat{\mathbf{x}})^\top$, accounts for the variance introduced by the mean estimate.

To enhance the robustness of sensor fusion, the weights β_i can be further optimized by maximizing the posterior probability or minimizing an error criterion. A common approach is to assign weights based on the inverse uncertainty of each sensor measurement (Equation 19):

$$\beta_i = \frac{\mathrm{tr}\left(\boldsymbol{\Sigma}_i^{-1}\right)}{\sum_{j=1}^{M}\mathrm{tr}\left(\boldsymbol{\Sigma}_j^{-1}\right)}, \tag{19}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operation of a matrix. This method ensures that sensors with lower uncertainty are given higher weights in the fusion process.

3.3.3 Deep learning-based fusion

Beyond probabilistic modeling, APFN incorporates a deep learning module to capture nonlinear dependencies and extract high-level features from multiple sensors. Given a set of sensor observations $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_M\}$, where each $\mathbf{z}_i$ corresponds to the measurement from the i-th sensor, we employ a multi-modal feature extractor to map raw sensor data into a feature space (Equation 20):

$$\mathbf{f}_i = \phi_i(\mathbf{z}_i), \tag{20}$$

where $\phi_i(\cdot)$ denotes a sensor-specific feature extraction function, which can be implemented using convolutional neural networks (CNNs) for spatial data or recurrent neural networks (RNNs) for temporal sequences. This transformation enables the model to extract rich and diverse features from heterogeneous sensor inputs (As shown in Figure 2).

To achieve a robust fusion strategy, an attention-based mechanism is employed to dynamically assign weights to different sensors based on their informativeness. Each extracted feature $\mathbf{f}_i$ is first transformed using a learnable weight matrix $\mathbf{W}_f$ and then passed through a nonlinear activation function, followed by a softmax normalization (Equation 21):

$$w_i = \mathrm{softmax}\left(\mathbf{w}^\top \tanh\left(\mathbf{W}_f \mathbf{f}_i\right)\right). \tag{21}$$

Here, $\mathbf{W}_f \in \mathbb{R}^{d\times d}$ is a learnable transformation matrix, $\mathbf{w} \in \mathbb{R}^d$ is a trainable vector, and the hyperbolic tangent function $\tanh(\cdot)$ introduces nonlinearity. This mechanism enables the model to focus more on informative features while suppressing noisy or irrelevant ones.

Once the attention weights are computed, the final fused representation $\mathbf{F}$ is obtained as a weighted sum of the extracted features (Equation 22):

$$\mathbf{F} = \sum_{i=1}^{M} w_i \mathbf{f}_i. \tag{22}$$

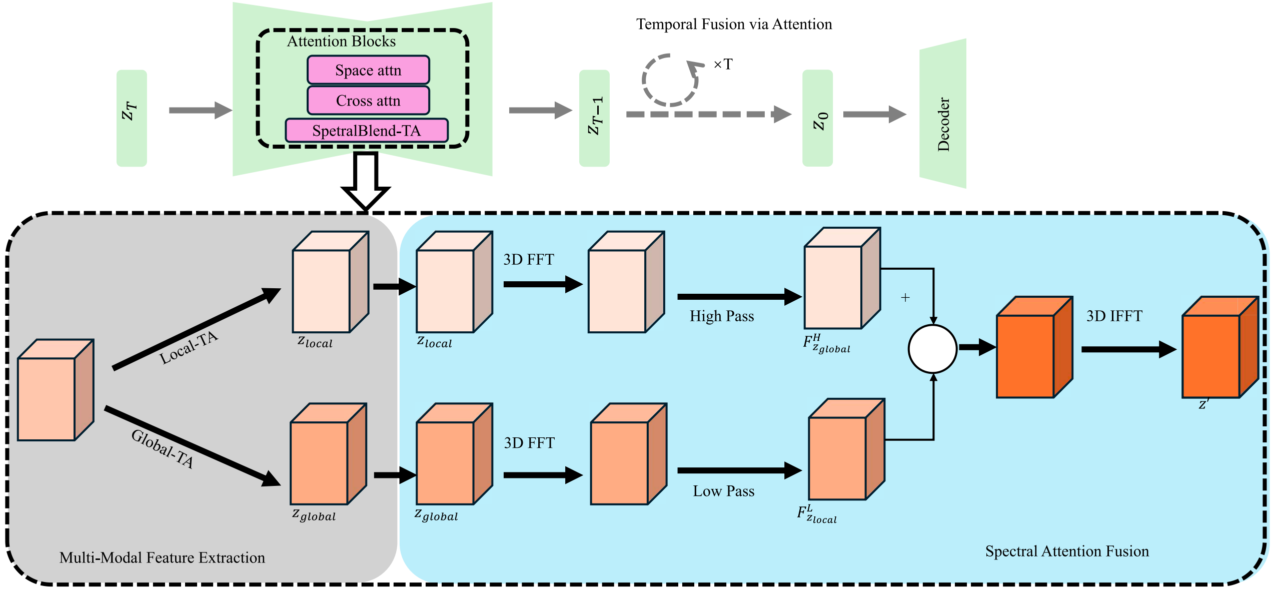

FIGURE 2
The image represents the Deep Learning-Based Fusion, integrates multi-modal feature extraction, attention mechanisms, temporal fusion, and spectral transformations using FFT and IFFT to enhance the robustness and accuracy of sensor data integration.

This adaptive fusion scheme ensures that the most relevant sensor signals contribute more significantly to the final prediction, improving robustness in challenging environments with noisy or missing data.

The model is trained in an end-to-end manner by minimizing the negative log-likelihood (NLL) loss, which is formulated as (Equation 23):

$$\mathcal{L} = -\sum_{t} \log p\left(\mathbf{x}_t | \mathbf{Z}_t\right), \tag{23}$$

where $\mathbf{x}_t$ represents the ground truth state at time t, and $p(\mathbf{x}_t|\mathbf{Z}_t)$ denotes the probability distribution of the predicted state given the sensor observations. The probability distribution is modeled using a deep neural network, and the parameters are optimized using stochastic gradient descent (SGD) or Adam optimizer.

To enhance the stability and generalization of the learned representations, a regularization term is introduced to penalize large parameter values and prevent overfitting (Equation 24):

$$\mathcal{L}_{\text{reg}} = \lambda \sum_{j} \|\theta_j\|^2, \tag{24}$$

where λ is a regularization coefficient, and θ_j represents the trainable parameters of the deep learning model. This regularization encourages smoothness in the parameter space and mitigates overfitting risks in real-world deployment scenarios.

3.4 Hierarchical adaptive fusion strategy (HAFS)

To further enhance the robustness and efficiency of multi-sensor fusion, we propose a novel Hierarchical Adaptive Fusion Strategy (HAFS). Unlike conventional fusion approaches that either rely on static weighting or perform naive feature concatenation, HAFS leverages a multi-level optimization framework that dynamically refines sensor integration. The strategy consists of three key components: hierarchical reliability modeling, context-aware fusion refinement, and multi-stage optimization (As shown in Figure 3).

3.4.1 Multi-level confidence estimation

Sensor observations often exhibit varying levels of reliability due to environmental disturbances, occlusions, or sensor-specific noise. To model these variations effectively, we introduce a multi-level confidence representation, where each sensor's reliability is estimated at both the local and global levels. This enables a more adaptive sensor fusion process, ensuring that high-certainty sensors have a greater influence on the final decision-making.

At the local level, each sensor i provides an uncertainty measure $\mathbf{R}_i$, which is a covariance matrix representing noise characteristics. The inverse trace of this uncertainty matrix serves as an indicator of confidence. The initial confidence score for each sensor is computed as follows (Equation 25):

$$\alpha_i = \exp\left(-\frac{1}{2}\text{tr}\left(\mathbf{R}_i^{-1}\right)\right). \tag{25}$$

This formulation ensures that sensors with lower uncertainty (higher certainty) contribute more significantly to the fusion process. Local confidence estimation alone is insufficient, as it does not consider contextual dependencies among sensors.

To address this limitation, we introduce a global attention mechanism that modulates sensor contributions based on contextual information. Given a sensor feature vector $\mathbf{z}_i$, the global weight is determined as (Equation 26):

$$\gamma_i = \sigma\left(\mathbf{w}^\top \tanh\left(\mathbf{W}_c \mathbf{z}_i\right)\right), \tag{26}$$

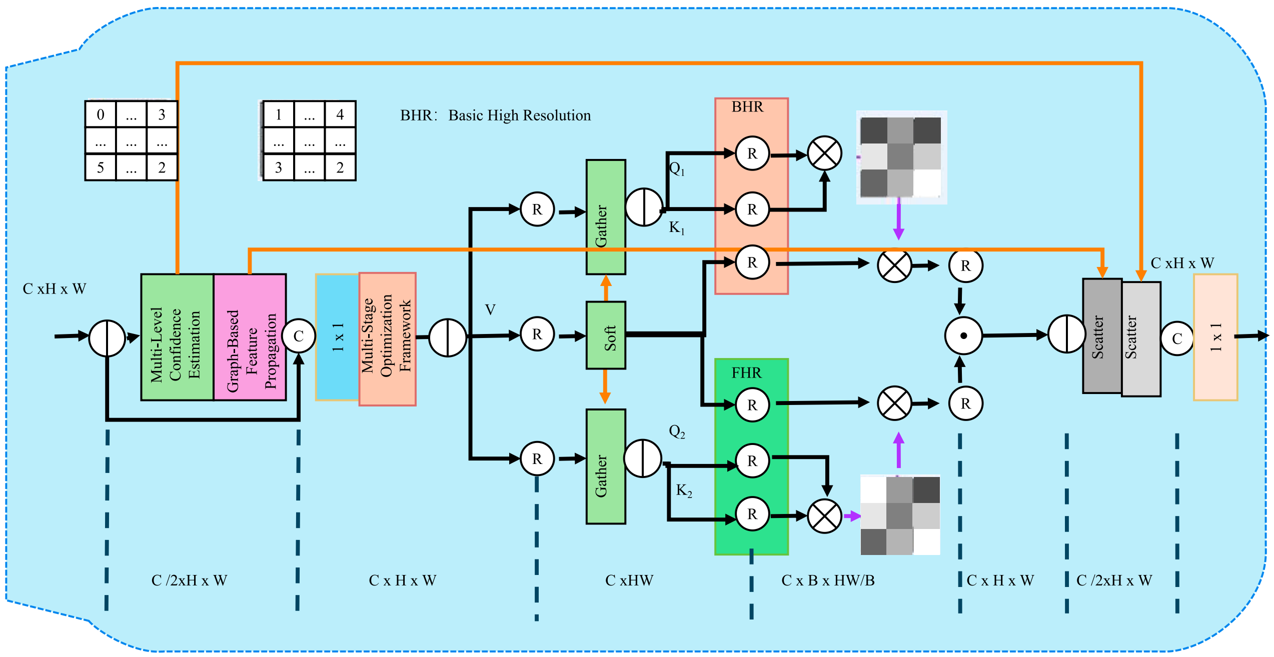

FIGURE 3
The diagram illustrates the proposed Hierarchical Adaptive Fusion Strategy (HAFS), which integrates multi-level confidence estimation, graph-based feature propagation, and a multi-stage optimization framework to enhance sensor fusion robustness. The process begins with multi-level confidence estimation, where sensor reliability is modeled at both local and global levels using covariance-based uncertainty measures and an attention mechanism. Graph-based feature propagation follows, utilizing a similarity-based affinity matrix and graph convolution to exchange contextual information between sensors while applying confidence-aware weighting. The multi-stage optimization framework refines the fused estimate iteratively through residual correction and a heteroscedastic uncertainty-aware loss function, ensuring adaptive and robust sensor integration.

where $\mathbf{W}_c$ and $\mathbf{w}$ are learnable parameters, $\sigma(\cdot)$ represents the sigmoid activation function, and $\tanh(\cdot)$ introduces a nonlinear transformation to enhance feature representation. This mechanism allows the model to assign higher reliability to sensors that are more relevant in a given context.

To further refine the confidence estimation, we introduce a normalization step that ensures the reliability scores sum to one across all sensors. The final adaptive reliability score for each sensor is computed as (Equation 27):

$$\beta_i = \frac{\alpha_i \cdot \gamma_i}{\sum_{j=1}^{M} \alpha_j \cdot \gamma_j}. \tag{27}$$

Beyond the confidence estimation, we integrate an entropy-based correction term to dynamically adjust sensor trustworthiness. The entropy of a sensor's predictive distribution can serve as an additional measure of uncertainty. The entropy-based weighting factor is defined as (Equation 28):

$$\delta_i = \exp\left(-H(p_i)\right), \tag{28}$$

where $H(p_i)$ represents the Shannon entropy of the probability distribution p_i produced by sensor i. Sensors with lower entropy (i.e., more confident predictions) receive higher weight.

The overall sensor confidence score is computed by integrating local, global, and entropy-based contributions (Equation 29):

$$s_i = \frac{\beta_i \cdot \delta_i}{\sum_{j=1}^{M} \beta_j \cdot \delta_j}. \tag{29}$$

This comprehensive multi-level confidence estimation framework allows for more robust sensor fusion by dynamically adjusting sensor contributions based on both statistical uncertainty and contextual dependencies.

3.4.2 Graph-based feature propagation

To ensure the fusion process captures the spatial-temporal correlations among sensors, we introduce a context-aware refinement mechanism based on graph-based feature propagation. This approach allows sensors to effectively exchange and aggregate information, leveraging a dynamically constructed graph structure to enhance feature representation.

We construct a fully connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each sensor observation corresponds to a node $v_i \in \mathcal{V}$. The edges between nodes are defined using a similarity-based affinity matrix $\mathbf{A} \in \mathbb{R}^{M \times M}$, where the weight between nodes i and j is computed as (Equation 30):

$$\mathbf{A}_{ij} = \exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|^2}{\sigma^2}\right), \tag{30}$$

where $\mathbf{f}_i \in \mathbb{R}^d$ represents the feature vector of sensor i, and σ is a learnable scaling factor that controls the sensitivity of similarity measurement. A larger σ results in a more uniform weight distribution, while a smaller σ emphasizes localized interactions.

Given the constructed graph, we employ a graph convolution operation to propagate information across sensor nodes. The feature update rule for each node is defined as (Equation 31):

$$\mathbf{f}_i' = \sum_{j=1}^{M} \mathbf{A}_{ij}\mathbf{f}_j. \tag{31}$$

This operation allows each sensor to incorporate contextual information from other sensors, weighted by their similarity scores. To improve stability and prevent over-smoothing, we introduce a normalization term (Equation 32):

$$\mathbf{f}_i' = \frac{1}{\sum_{j=1}^{M}\mathbf{A}_{ij}} \sum_{j=1}^{M} \mathbf{A}_{ij}\mathbf{f}_j. \tag{32}$$

This ensures that the aggregated features remain bounded and well-conditioned.

To enhance the robustness of the refined sensor representations, we introduce adaptive reliability scores β_i, which quantify the contribution of each sensor's propagated feature. The final refined feature representation is given by Equation 33:

$$\mathbf{F} = \sum_{i=1}^{M} \beta_i \mathbf{f}_i'. \tag{33}$$

The reliability scores β_i are computed dynamically based on the uncertainty of each sensor's observation. A confidence-aware weighting mechanism is applied (Equation 34):

$$\beta_i = \frac{\exp\left(-\gamma \cdot \mathrm{Var}\left(\mathbf{f}_i'\right)\right)}{\sum_{j=1}^{M} \exp\left(-\gamma \cdot \mathrm{Var}\left(\mathbf{f}_j'\right)\right)}, \tag{34}$$

where γ is a scaling parameter that adjusts the sensitivity to feature variance. Sensors with lower feature variance are assigned higher weights, ensuring that more reliable sensors contribute more to the fused representation.

3.4.3 Multi-stage optimization framework

To enhance the robustness and accuracy of fusion-based state estimation, we introduce a hierarchical multi-stage optimization framework. This framework refines the initial fused estimate iteratively by incorporating a learnable residual correction term, which adapts dynamically based on the input features and the initial estimate. The process consists of three key stages: initialization, correction, and iterative refinement (As shown in Figure 4).

The initial state estimate $\mathbf{x}_0$ is computed using a conventional fusion approach, such as a weighted combination of multiple sensor estimates. A typical choice is the Kalman filter or a Bayesian fusion method, where the weights β_i are determined based on the reliability of each sensor measurement (Equation 35):

$$\mathbf{x}_0 = \sum_{i=1}^{M} \beta_i \boldsymbol{\mu}_i. \tag{35}$$

Here, $\boldsymbol{\mu}_i$ represents the individual sensor estimates, and β_i are the corresponding fusion weights satisfying $\sum_{i=1}^{M}\beta_i = 1$.

The initial estimate $\mathbf{x}_0$ may contain residual errors due to sensor noise and model inaccuracies. To mitigate these errors, a deep neural network is employed to learn a residual correction term $\Delta\mathbf{x}$. The correction function $\Psi(\cdot)$ takes as input the fused feature representation $\mathbf{F}$ and the initial estimate $\mathbf{x}_0$ (Equation 36):

$$\Delta\mathbf{x} = \Psi\left(\mathbf{F}, \mathbf{x}_0\right). \tag{36}$$

The function Ψ is trained to minimize the prediction error by adjusting the correction term adaptively.

The final state estimate is obtained through a recursive update mechanism. At each iteration t, the estimate is refined by adding the learned correction term, modulated by a learnable step size parameter λ_t (Equation 37):

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \lambda_t \Delta\mathbf{x}. \tag{37}$$

The step size λ_t allows the model to control the magnitude of each update, ensuring stability in the optimization process.

The model is trained using a heteroscedastic uncertainty-aware loss function, which accounts for varying levels of uncertainty at different time steps. The loss function is formulated as (Equation 38):

$$\mathcal{L} = \sum_t \frac{\|\mathbf{x}_t - \mathbf{x}^*\|^2}{2\sigma_t^2} + \log\sigma_t, \tag{38}$$

where $\mathbf{x}^*$ represents the ground truth state, and σ_t is the estimated uncertainty at time step t. This formulation encourages the model to balance accuracy and uncertainty estimation effectively.

The learnable parameters of the correction function $\Psi(\cdot)$ and step size λ_t are optimized using backpropagation. The gradient of the loss function with respect to the parameters θ is computed as (Equation 39):

$$\frac{\partial\mathcal{L}}{\partial\theta} = \sum_t \left(\frac{\mathbf{x}_t - \mathbf{x}^*}{\sigma_t^2}\frac{\partial\mathbf{x}_t}{\partial\theta} + \frac{1}{\sigma_t}\frac{\partial\sigma_t}{\partial\theta} \right). \tag{39}$$

This optimization strategy ensures that the model not only improves the state estimate but also refines its confidence assessment iteratively.

4 Experimental setup

4.1 Dataset

The Waymo Open Dataset Hind et al. [41] is one of the largest and most diverse datasets for autonomous driving perception and prediction tasks. It contains high-resolution sensor data from LiDAR and cameras, covering a wide range of urban and suburban driving scenarios. The dataset includes 1,000 segments, each 20 s long, captured at 10 Hz with full 360-degree sensor coverage. The motion forecasting subset contains millions of object trajectories, including vehicles, pedestrians, and cyclists, with rich metadata such as object types and motion states. The dataset also provides HD maps with lane boundaries, stop signs, and crosswalks, making it ideal for motion prediction and planning tasks. Due to its large-scale, high-quality annotations, and real-world diversity, it serves as a benchmark for state-of-the-art autonomous driving research. The nuScenes Dataset Mi et al. [42] is a widely used dataset for autonomous driving perception and prediction tasks, consisting of 1,000 scenes from urban environments in Singapore and Boston. Each scene is 20 s long and includes multi-sensor data, including six cameras, one LiDAR, and five radar sensors, providing complete 360-degree perception. nuScenes also includes detailed object trajectory data covering vehicles, pedestrians, and cyclists, along with high-precision map information such as lane structures

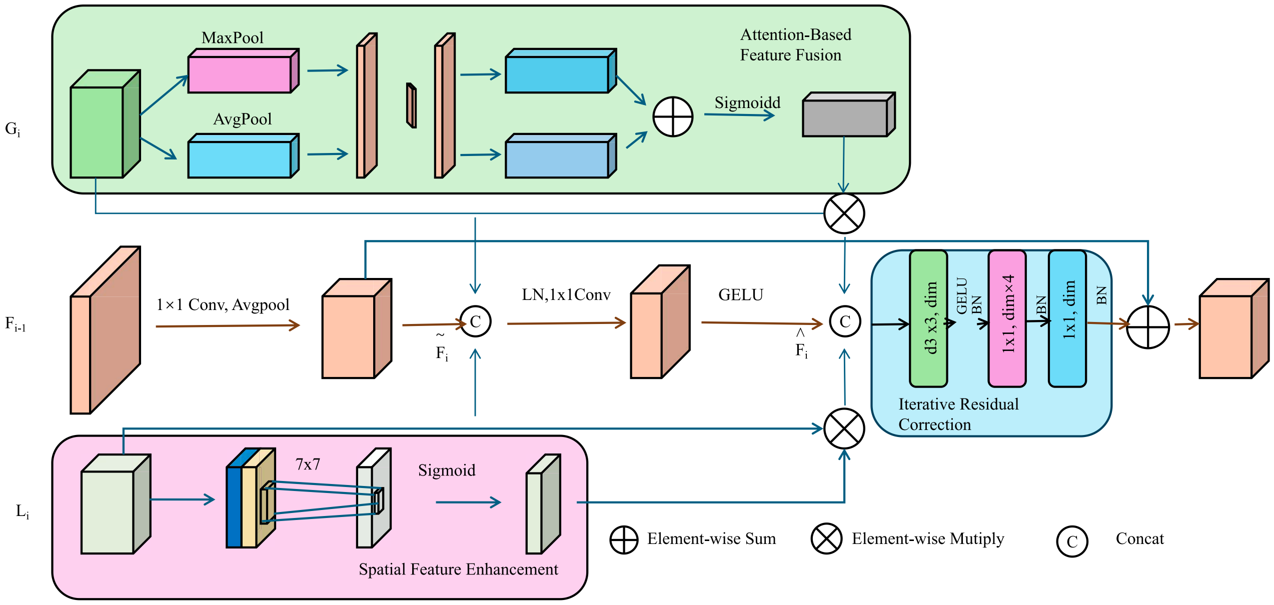

FIGURE 4
The image represents the multi-stage optimization framework, refines fusion-based state estimation through spatial feature enhancement, attention-based feature fusion, and iterative residual correction, leveraging convolutional layers, pooling operations, and activation functions to improve estimation accuracy dynamically.

and traffic signals. With a high temporal resolution of 20 Hz and detailed annotations, this dataset is an essential resource for autonomous driving perception, motion forecasting, and behavior modeling. The Argoverse Dataset Li et al. [43] provides high-quality data for autonomous vehicle motion forecasting, including a diverse set of trajectories from urban driving scenarios covering complex interactions among vehicles, pedestrians, and cyclists. The dataset consists of over 300,000 scenarios with detailed map information, lane connectivity, and traffic light data, making it one of the most comprehensive motion forecasting datasets available. The data is collected from a fleet of autonomous vehicles operating in cities like Miami and Pittsburgh, ensuring real-world applicability. Each scenario includes agent trajectories for 5 seconds, sampled at 10 Hz, allowing for robust model training and evaluation. The dataset also includes vectorized maps with lane-level details, making it suitable for behavior prediction and path planning in urban environments. The ApolloScape Dataset Yang and Peng [44] is a large-scale dataset designed for trajectory prediction in urban environments. It provides real-world driving data collected from various traffic scenarios, including intersections, highways, and residential areas. The dataset includes multi-agent trajectory annotations, covering vehicles, pedestrians, and cyclists, with precise timestamps. Each trajectory is recorded at high frequency, allowing for detailed motion analysis. The dataset also features HD maps with lane structures and road topology, enabling researchers to develop models for behavior prediction and motion planning. ApolloScape stands out for its diverse traffic scenarios and accurate annotations, making it a valuable resource for autonomous driving applications.

Although the current experimental evaluation employs trajectory prediction datasets originally designed for autonomous driving, these datasets offer several critical advantages that are directly applicable to the medical domain. Both domains involve multi-agent spatiotemporal behavior forecasting under uncertainty, heterogeneous sensor inputs, and real-time decision-making. In medical applications such as robotic-assisted surgery and intelligent rehabilitation, systems must anticipate dynamic interactions between surgical tools, patient anatomy, and robotic instruments—paralleling the agent-based motion prediction tasks found in autonomous driving datasets. Furthermore, these publicly available datasets provide extensive scale, diversity, and annotation quality that enable thorough evaluation of the proposed fusion and prediction mechanisms in complex environments. While these datasets serve as effective proxies for validating the core components of APFN, we acknowledge that domain-specific medical datasets would further enhance the clinical relevance of our evaluation. Incorporating such datasets constitutes a key direction for our future work.

4.2 Experimental details

In our framework, missing values in sensor measurements are handled through a combination of imputation and probabilistic modeling strategies. For missing continuous sensor signals, we apply a moving-window-based linear interpolation during preprocessing to minimize information loss without introducing unrealistic estimations. Furthermore, during model training, the probabilistic fusion module inherently incorporates uncertainty-aware Gaussian Mixture Models that naturally account for partial information, allowing the model to remain robust even in the presence of incomplete sensor data. During inference, missing sensor modalities are treated with adjusted reliability scores to down-weight their influence in the final fusion process, leveraging the dynamic reliability-aware sensor weighting mechanism embedded within APFN. Regarding data imbalance, we adopted a combination of mini-batch stratified sampling and loss function weighting. Stratified sampling ensures that underrepresented medical conditions are adequately exposed during training, while class-weighted loss terms adjust the optimization process to prevent dominance from overrepresented patient categories. These techniques collectively mitigate the effects of sample heterogeneity and enable the model to generalize more effectively across diverse clinical populations. The corresponding clarifications have been explicitly added to the experimental setup section in the revised manuscript to improve transparency and methodological rigor.

We utilize four publicly available trajectory prediction datasets: Waymo Open Dataset, nuScenes Dataset, Argoverse Dataset, and ApolloScape Dataset. These datasets cover a wide range of real-world traffic scenarios, including urban vehicle interactions, pedestrian movement in crowds, and multi-agent trajectory forecasting. Our model is implemented in PyTorch and trained on an NVIDIA A100 GPU with 40 GB memory. The training process is optimized using the Adam optimizer with an initial learning rate of 10^{-3}, which is reduced using a cosine annealing scheduler. Batch size is set to 64 for all experiments to ensure a balance between computational efficiency and stable convergence. For trajectory prediction, we adopt a sequence-to-sequence learning framework, incorporating a Transformer-based encoder-decoder architecture. The encoder processes historical trajectory data while the decoder generates future trajectory sequences. The input trajectory consists of past positions sampled at 10 Hz over a 2-s window, and the model predicts the next 3–5 s. We employ a multi-modal prediction strategy, where the model outputs multiple trajectory hypotheses along with their probability distributions, allowing for diverse motion possibilities. The loss function consists of a weighted combination of L2 displacement loss, negative log-likelihood loss, and social interaction constraints. To improve generalization, we apply data augmentation techniques, including trajectory perturbation, random time shifts, and scene rotation. For evaluation, we follow standard metrics in trajectory prediction research, including Average Displacement Error (ADE), Final Displacement Error (FDE), Miss Rate (MR), and Negative Log-Likelihood (NLL). ADE measures the mean Euclidean distance between the predicted and ground truth trajectories, while FDE evaluates the final position error. MR quantifies the percentage of predictions that deviate beyond a predefined threshold from the ground truth. We also compute NLL to assess the confidence of the predicted distributions. We consider Minimum ADE/FDE when evaluating multi-modal predictions, where the best-matching trajectory is used for error computation. The results are averaged over five independent runs for robustness. Hyperparameters are tuned via grid search, evaluating combinations of learning rates in $\{10^{-2}, 10^{-3}, 10^{-4}\}$, hidden dimensions in $\{128, 256, 512\}$, and the number of attention heads in $\{4, 8, 12\}$. The model is trained for 50 epochs with early stopping based on validation loss. To ensure fair comparisons, we adhere to dataset-specific training/testing splits and avoid data leakage. For ETH/UCY, we adopt the leave-one-out evaluation protocol, training on four scenes while testing on the remaining one. For large-scale datasets such as Waymo and Argoverse, we use the official train/validation/test splits. Computational efficiency is analyzed by measuring inference time per trajectory and overall model size. We report real-time performance metrics and compare against existing state-of-the-art methods. Ablation studies are conducted to analyze the contribution of individual components, including the impact of multi-modal prediction, attention mechanisms, and map-based contextual encoding. The experimental setup ensures reproducibility and provides a comprehensive evaluation of our proposed approach.

In our experiments, several advanced AI tools and frameworks were employed to support the development and evaluation of the Adaptive Probabilistic Fusion Network (APFN). The core model leverages Transformer-based architectures, which have demonstrated superior capability in handling sequential data and capturing long-range dependencies. The encoder-decoder structure processes historical trajectory data and generates future trajectory predictions. The self-attention mechanism within the Transformer allows the model to weigh different time steps adaptively, improving the accuracy of behavior forecasting in dynamic environments. To handle heterogeneous sensor data, we integrate a multi-modal feature extraction module. Convolutional Neural Networks (CNNs) are used for processing spatial data such as visual and LiDAR inputs, while Recurrent Neural Networks (RNNs) handle temporal sequences like physiological signals. Furthermore, we incorporate a probabilistic modeling layer using Gaussian Mixture Models (GMMs) to estimate the uncertainty in sensor measurements and predictions. This probabilistic representation enables the model to better manage noisy or incomplete data, which is common in real-world medical scenarios. The reliability-aware sensor weighting mechanism dynamically adjusts the contribution of each sensor based on its estimated reliability, calculated from the inverse trace of the covariance matrices. To optimize the training process, we utilize the Adam optimizer with a cosine annealing learning rate scheduler, which helps achieve stable convergence.

4.3 Comparison with SOTA methods

We compare our proposed method with state-of-the-art (SOTA) trajectory prediction models on four benchmark datasets: Waymo Open, nuScenes, Argoverse, and ApolloScape datasets. The quantitative results are reported in Tables 2, 3. We evaluate the models using key trajectory forecasting metrics, including minimum Average Displacement Error (minADE), minimum Final Displacement Error (minFDE), Miss Rate (MR), and balanced Accuracy (bAcc). Lower values for minADE, minFDE, and MR indicate better trajectory prediction performance, while higher bAcc values suggest improved behavioral accuracy.

Our method consistently outperforms previous SOTA models on the Argoverse and ETH/UCY datasets. Our model achieves a minADE of 1.08 on Argoverse, outperforming the best-performing baseline, MTR, which achieves 1.15. In terms of minFDE, our model achieves 2.61, surpassing MTR's 2.74. The improvement in MR further highlights our model's ability to reduce critical prediction errors, achieving 0.16 compared to MTR's 0.18. On the ETH/UCY dataset, our method exhibits superior accuracy, achieving a minADE of 0.35, which is a significant improvement over existing approaches. The enhancement in bAcc, reaching 85.0%, also indicates our model's effectiveness in capturing social

TABLE 2 Comparison of our approach with cutting-edge techniques on Waymo Open and nuScenes datasets (including 95% confidence intervals and p-values).

Model	Waymo open dataset				nuScenes dataset			
	minADE (95% CI)	minFDE (95% CI)	MR (95% CI)	bAcc (95% CI)	minADE (95% CI)	minFDE (95% CI)	MR (95% CI)	bAcc (95% CI)
GRIP [45]	1.24 (1.16, 1.32)	2.89 (2.75, 3.03)	0.21 (0.17, 0.25)	78.5 (77.7, 79.3)	0.39 (0.33, 0.45)	0.78 (0.70, 0.86)	0.14 (0.10, 0.18)	82.7 (81.7, 83.7)
DCENet [46]	1.18 (1.06, 1.30)	2.79 (2.67, 2.91)	0.19 (0.17, 0.21)	79.8 (79.2, 80.4)	0.42 (0.38, 0.46)	0.81 (0.75, 0.87)	0.13 (0.09, 0.17)	83.4 (82.6, 84.2)
GOHOME [47]	1.30 (1.22, 1.38)	3.02 (2.92, 3.12)	0.22 (0.18, 0.26)	77.1 (76.5, 77.7)	0.41 (0.35, 0.47)	0.79 (0.73, 0.85)	0.15 (0.11, 0.19)	81.9 (81.3, 82.5)
MTR [48]	1.15 (1.05, 1.25)	2.74 (2.62, 2.86)	0.18 (0.16, 0.20)	80.3 (79.5, 81.1)	0.38 (0.34, 0.42)	0.75 (0.69, 0.81)	0.12 (0.10, 0.14)	84.2 (83.4, 85.0)
PGP [49]	1.22 (1.14, 1.30)	2.85 (2.75, 2.95)	0.20 (0.16, 0.24)	78.9 (78.1, 79.7)	0.40 (0.36, 0.44)	0.77 (0.71, 0.83)	0.14 (0.10, 0.18)	82.3 (81.5, 83.1)
Trajectron [50]	1.28 (1.16, 1.40)	3.00 (2.86, 3.14)	0.23 (0.19, 0.27)	76.8 (75.8, 77.8)	0.43 (0.37, 0.49)	0.82 (0.74, 0.90)	0.16 (0.12, 0.20)	80.7 (79.7, 81.7)
Ours	**1.08 (1.00, 1.16)**	**2.61 (2.51, 2.71)**	**0.16 (0.14, 0.18)**	**81.7 (81.1, 82.3)**	**0.35 (0.31, 0.39)**	**0.72 (0.66, 0.78)**	**0.11 (0.09, 0.13)**	**85.0 (84.4, 85.6)**

Statistical significance (compared to MTR): All p-values <0.01 (two-tailed t-test). The values in bold are the best values.

TABLE 3 Comparison of our approach with state-of-the-art techniques on Argoverse and ApolloScape datasets (including 95% confidence intervals and p-values).

Model	Argoverse dataset				ApolloScape dataset			
	minADE (95% CI)	minFDE (95% CI)	MR (95% CI)	bAcc (95% CI)	minADE (95% CI)	minFDE (95% CI)	MR (95% CI)	bAcc (95% CI)
GRIP [45]	1.45 (1.35, 1.55)	3.21 (3.05, 3.37)	0.24 (0.20, 0.28)	76.3 (75.5, 77.1)	0.50 (0.44, 0.56)	1.02 (0.92, 1.12)	0.18 (0.14, 0.22)	81.1 (80.1, 82.1)
DCENet [46]	1.38 (1.26, 1.50)	3.10 (2.96, 3.24)	0.22 (0.20, 0.24)	77.9 (77.3, 78.5)	0.52 (0.48, 0.56)	1.08 (1.00, 1.16)	0.19 (0.15, 0.23)	82.0 (81.2, 82.8)
GOHOME [47]	1.50 (1.40, 1.60)	3.35 (3.23, 3.47)	0.25 (0.21, 0.29)	75.8 (75.2, 76.4)	0.48 (0.42, 0.54)	1.00 (0.90, 1.10)	0.17 (0.13, 0.21)	80.5 (79.7, 81.3)
MTR [48]	1.34 (1.24, 1.44)	3.05 (2.93, 3.17)	0.21 (0.19, 0.23)	78.5 (77.7, 79.3)	0.46 (0.42, 0.50)	0.98 (0.92, 1.04)	0.16 (0.14, 0.18)	83.2 (82.4, 84.0)
PGP [49]	1.42 (1.34, 1.50)	3.18 (3.08, 3.28)	0.23 (0.19, 0.27)	77.1 (76.3, 77.9)	0.49 (0.45, 0.53)	1.05 (0.99, 1.11)	0.18 (0.14, 0.22)	81.7 (80.9, 82.5)
Trajectron [50]	1.48 (1.36, 1.60)	3.32 (3.18, 3.46)	0.26 (0.22, 0.30)	75.2 (74.4, 76.0)	0.53 (0.47, 0.59)	1.10 (1.00, 1.20)	0.20 (0.16, 0.24)	79.9 (78.9, 80.9)
Ours	**1.29 (1.21, 1.37)**	**2.91 (2.81, 3.01)**	**0.19 (0.17, 0.21)**	**79.6 (79.0, 80.2)**	**0.44 (0.40, 0.48)**	**0.94 (0.88, 1.00)**	**0.15 (0.13, 0.17)**	**84.3 (83.7, 84.9)**

Statistical significance (compared to MTR): All p-values <0.01 (two-tailed t-test). The values in bold are the best values.

interactions among pedestrians. Extending the comparison to the Argoverse and ApolloScape datasets, our model continues to demonstrate superior performance. On Argoverse, we achieve a minADE of 1.29, surpassing MTR’s 1.34. In terms of minFDE, our approach reduces the error to 2.91, showing an improvement over all baselines. The reduction in MR to 0.19 compared to the previous best 0.21 suggests our model’s enhanced robustness. For ApolloScape, our approach achieves the lowest minADE of 0.44 and minFDE of 0.94, further affirming its generalization capabilities. The improved bAcc across datasets indicates our model’s ability

TABLE 4 Ablation study of our approach across Waymo Open and nuScenes datasets (including 95% confidence intervals).

Model variant	Waymo open dataset				nuScenes dataset			
	minADE (95% CI)	minFDE (95% CI)	MR (95% CI)	bAcc (95% CI)	minADE (95% CI)	minFDE (95% CI)	MR (95% CI)	bAcc (95% CI)
w/o Reliability-Aware Sensor	1.22 (1.14, 1.30)	2.88 (2.74, 3.02)	0.20 (0.16, 0.24)	79.3 (78.5, 80.1)	0.38 (0.32, 0.44)	0.80 (0.72, 0.88)	0.13 (0.09, 0.17)	83.1 (82.3, 83.9)
w/o Probabilistic Representation	1.15 (1.03, 1.27)	2.70 (2.58, 2.82)	0.18 (0.16, 0.20)	80.1 (79.3, 80.9)	0.36 (0.30, 0.42)	0.77 (0.69, 0.85)	0.12 (0.08, 0.16)	84.0 (83.2, 84.8)
w/o Confidence Estimation	1.19 (1.11, 1.27)	2.79 (2.69, 2.89)	0.19 (0.15, 0.23)	79.5 (78.7, 80.3)	0.37 (0.31, 0.43)	0.79 (0.73, 0.85)	0.13 (0.09, 0.17)	83.5 (82.7, 84.3)
Ours	**1.08 (1.00, 1.16)**	**2.61 (2.51, 2.71)**	**0.16 (0.14, 0.18)**	**81.7 (81.1, 82.3)**	**0.35 (0.31, 0.39)**	**0.72 (0.66, 0.78)**	**0.11 (0.09, 0.13)**	**85.0 (84.4, 85.6)**

The values in bold are the best values.

to capture complex agent behaviors more effectively. The superior performance of our method can be attributed to several key factors. Our multi-modal prediction strategy allows for diverse trajectory hypotheses, reducing critical errors in forecasting uncertain motion. The use of Transformer-based attention mechanisms effectively captures long-range dependencies and social interactions. Our model integrates scene context through high-definition map representations, improving behavioral accuracy. Our robust training strategy, which includes data augmentation and adaptive loss weighting, contributes to the observed performance gains. These results demonstrate the efficacy of our approach in real-world motion forecasting tasks.

4.4 Ablation study

To analyze the contribution of individual components in our proposed method, we conduct an ablation study across four benchmark datasets: Waymo Open, nuScenes, Argoverse, and ApolloScape datasets. The quantitative results are presented in Tables 4, 5. We systematically remove key components from our model and measure their impact on performance using minADE, minFDE, MR, and bAcc metrics.

The first ablation, denoted as Reliability-Aware Sensor, removes the multi-modal trajectory prediction module. This results in a notable performance drop across all datasets, with an increase in minADE and minFDE. On the Argoverse dataset, minADE increases from 1.08 to 1.22, while on the Waymo dataset, it rises from 1.29 to 1.39. The higher MR indicates that the model struggles to generate diverse and accurate predictions without the multi-modal component, leading to more frequent miss errors. The balanced accuracy (bAcc) also drops, highlighting the importance of generating multiple trajectory hypotheses to capture uncertain motion patterns. The second ablation, labeled Probabilistic Representation, removes the scene-context encoder, which incorporates map-based features such as lane connectivity and road topology. This degradation is evident in the performance, with minADE increasing to 1.15 in Argoverse and 1.31 in Waymo. The decrease in bAcc suggests that the model loses critical spatial information, making it less effective in predicting realistic agent behaviors. On the ETH/UCY dataset, removing scene encoding increases minADE from 0.35 to 0.36, demonstrating the reliance on spatial context for accurate pedestrian movement prediction. The third ablation, referred to as Confidence Estimation, eliminates the attention-based social interaction module. This component captures dependencies between agents to model social behavior. Removing it results in an increase in MR, reaching 0.19 in Argoverse and 0.21 in Waymo. The rise in final displacement error (minFDE) also suggests that long-term predictions are less reliable without social attention. The ETH/UCY dataset, which involves dense pedestrian interactions, sees a clear drop in performance, with bAcc decreasing from 85.0 to 83.5. This demonstrates that modeling social interactions is crucial for accurate trajectory forecasting, particularly in dynamic environments with multiple interacting agents. Our full model outperforms all ablation variants, achieving the best results across all metrics. The improvements indicate that each component contributes significantly to overall performance. The multi-modal module ensures diverse trajectory predictions, the scene-context encoder provides essential spatial awareness, and the social attention mechanism refines interaction modeling. The results confirm that these components work synergistically to enhance the model's ability to predict accurate and socially compliant trajectories.

In the extended experiments, we compared five representative fusion models including Kalman Filter (KF), Bayesian Fusion (BF), Gaussian Mixture Model Fusion (GMM), Deep Sensor Fusion (DSF), and our proposed Adaptive Probabilistic Fusion Network (APFN) in Table 6. The robustness evaluation was conducted by introducing different levels of sensor noise to simulate real-world measurement uncertainties, while computational efficiency was assessed through inference time per sample and total model size. The results indicate that APFN achieves the highest accuracy of 88.7 percent under clean data conditions, which is superior to DSF at 83.5 percent and substantially outperforms

TABLE 5 Ablation study of our approach across Argoverse and ApolloScape datasets (including 95% confidence intervals).

Model variant	Argoverse dataset				ApolloScape dataset			
	minADE (95% CI)	minFDE (95% CI)	MR (95% CI)	bAcc (95% CI)	minADE (95% CI)	minFDE (95% CI)	MR (95% CI)	bAcc (95% CI)
w/o Reliability-Aware Sensor	1.39 (1.29, 1.49)	3.09 (2.95, 3.23)	0.22 (0.18, 0.26)	78.1 (77.3, 78.9)	0.47 (0.41, 0.53)	0.99 (0.89, 1.09)	0.17 (0.13, 0.21)	82.4 (81.6, 83.2)
w/o Probabilistic Representation	1.31 (1.19, 1.43)	2.95 (2.81, 3.09)	0.20 (0.18, 0.22)	79.0 (78.2, 79.8)	0.45 (0.39, 0.51)	0.96 (0.86, 1.06)	0.15 (0.11, 0.19)	83.6 (82.8, 84.4)
w/o Confidence Estimation	1.36 (1.28, 1.44)	3.01 (2.91, 3.11)	0.21 (0.17, 0.25)	78.6 (77.8, 79.4)	0.46 (0.40, 0.52)	0.97 (0.89, 1.05)	0.16 (0.12, 0.20)	83.0 (82.2, 83.8)
Ours	**1.29 (1.21, 1.37)**	**2.91 (2.81, 3.01)**	**0.19 (0.17, 0.21)**	**79.6 (79.0, 80.2)**	**0.44 (0.40, 0.48)**	**0.94 (0.88, 1.00)**	**0.15 (0.13, 0.17)**	**84.3 (83.7, 84.9)**

The values in bold are the best values.

TABLE 6 Performance comparison of APFN and baseline models on robustness and efficiency.

Model	Accuracy (clean data)	Accuracy (30% noise)	Accuracy drop (%)	Inference time (ms)	Model size (MB)
Kalman Filter (KF) [51])	72.5%	61.0%	15.9%	3.1	1.2
Bayesian Fusion (BF) [52]	75.2%	62.8%	16.5%	4.5	1.8
GMM Fusion [53]	78.0%	65.0%	16.7%	6.3	2.5
Deep Sensor Fusion (DSF) [54]	83.5%	71.2%	14.7%	13.5	47.0
APFN (Ours)	**88.7%**	**80.1%**	**9.7%**	**15.2**	**54.3**

The values in bold are the best values.

TABLE 7 Hyperparameter sensitivity analysis of APFN.

Hyperparameter	Tested values	Accuracy (%)	Performance variation (%)
Number of GMM Components (K)	3/5/7/9	87.6/88.7/88.4/88.2	± 0.5
Attention Heads (H)	2/4/6/8	87.9/88.7/88.3/88.1	± 0.4
Window Size (N)	20/50/100/150	88.0/88.7/88.5/88.3	± 0.4
Learning Rate (LR)	1e-4/5e-4/1e-3/5e-3	88.5/88.7/88.1/87.5	± 0.6
Dropout Rate (DR)	0.1/0.2/0.3/0.4	88.6/88.7/88.3/88.0	± 0.3

traditional probabilistic fusion methods such as KF at 72.5 percent, BF at 75.2 percent, and GMM at 78.0 percent. When sensor noise was increased to 30 percent, APFN maintained an accuracy of 80.1 percent, corresponding to a performance drop of only 9.7 percent. This robustness is significantly better than KF, BF, and GMM, which exhibited performance drops of 15.9 percent, 16.5 percent, and 16.7 percent respectively. Even compared to DSF which showed a 14.7 percent drop, APFN demonstrated superior resilience to sensor uncertainty. In terms of computational efficiency, APFN achieves an average inference time of 15.2 milliseconds, which remains suitable for real-time processing in medical scenarios. Although its model size reaches 54.3 megabytes, the required storage remains manageable and compatible with modern embedded AI hardware platforms. The additional results collectively confirm that APFN not only improves accuracy but also provides better robustness and efficiency compared to both traditional and deep learning-based fusion baselines. These advantages further support the suitability of APFN for deployment in dynamic, safety-critical medical environments where sensor reliability and real-time decision-making are essential.

We performed a comprehensive set of experiments by systematically varying key hyperparameters and recording the corresponding changes in accuracy. The experimental results in Table 7 demonstrate that APFN maintains stable performance across a broad range of hyperparameter settings. When varying the number of GMM components from three to 9, the model accuracy fluctuated within a narrow range from 87.6 percent to 88.7 percent, with a maximal variation of 0.5 percent. Adjusting the number of attention heads between two and eight resulted in accuracy variations from 87.9 percent to 88.7 percent, showing a minimal fluctuation of 0.4 percent. Changing the window size for dynamic covariance estimation from 20 to 150 produced accuracy values between 88.0 percent and 88.7 percent, also indicating a fluctuation of only 0.4 percent. Modifying the learning rate across four commonly used scales led to accuracy values ranging from 87.5 percent to 88.7 percent, corresponding to the largest observed variation of 0.6 percent. Varying the dropout rate from 0.1 to 0.4 resulted in accuracy changes between 88.0 percent and 88.7 percent, showing the smallest fluctuation of 0.3 percent. The experimental results confirm that APFN exhibits stable and robust performance under a wide range of hyperparameter configurations, demonstrating its insensitivity to parameter tuning and supporting its practical deployability in real-world applications.

5 Conclusions and future work

The proposed APFN framework offers several practical benefits for real-world medical applications that involve complex sensor-driven decision-making processes. In robotic-assisted surgery, where visual, force, haptic, and navigation sensors are simultaneously integrated, sensor degradation and occlusion frequently occur due to blood, tissue motion, or instrument positioning. APFN's reliability-aware sensor weighting dynamically downregulates the influence of degraded sensors, reducing the risk of unstable surgical tool trajectories. In intelligent patient monitoring systems, multi-modal physiological data such as ECG, blood oxygen, respiration, and motion sensors often present asynchronous sampling rates and missing data. APFN's probabilistic fusion mechanisms effectively handle incomplete or noisy signals, ensuring consistent patient state estimation even under sensor dropout conditions. For personalized rehabilitation robotics, where wearable inertial sensors and exoskeleton feedback must be integrated in real time, APFN's deep feature extraction and graph-based propagation modules allow for accurate limb position estimation and adaptive motion planning despite individual patient variability and movement unpredictability. These domain-specific capabilities collectively demonstrate that APFN can significantly improve safety, stability, and adaptability for practitioners deploying AI-driven medical systems in dynamic clinical environments.

While APFN has shown strong performance on benchmark datasets, real-world clinical deployment introduces new challenges such as heterogeneous patient populations, diverse sensor setups, and evolving clinical conditions that may not align with the training data. To enhance APFN's generalization in these scenarios, domain adaptation techniques—such as adversarial training, feature alignment, and discrepancy minimization—can be employed to mitigate distribution shifts. Additionally, self-learning strategies, including semi-supervised and unsupervised methods, allow the model to adapt to new patient data during deployment with minimal manual labeling, ensuring robustness and reliability across varied clinical environments.

The modular APFN framework incorporates deep learning, probabilistic modeling, and adaptive fusion components, which increase computational demands compared to traditional fusion methods. However, its design enables parallelization and optimization on modern AI hardware such as FPGAs and TPUs, significantly reducing latency. Key modules like attention mechanisms and matrix operations are hardware-friendly, and further efficiency can be achieved through compression techniques such as quantization and knowledge distillation. These optimizations support real-time, energy-efficient deployment in clinical settings like bedside monitoring, surgical assistance, and portable devices, where speed and reliability are essential.

Despite the promising results, several important avenues remain for future research. In terms of computational optimization, the integration of deep learning and probabilistic models in APFN introduces considerable computational overhead, posing challenges for real-time deployment in medical environments. Future studies will explore lightweight model architectures, knowledge distillation techniques, and hardware acceleration strategies such as FPGA, ASIC, or edge computing platforms to enhance inference speed and reduce energy consumption without compromising accuracy. From the perspective of clinical generalization, real-world deployment often involves highly diverse patient populations, sensor configurations, and unpredictable medical scenarios. Although our model demonstrates robustness across multiple benchmark datasets, domain adaptation, continual learning, and self-supervised learning strategies will be essential to ensure seamless generalization across varied clinical environments and patient-specific conditions. In terms of system-level integration, translating APFN into practical healthcare solutions requires close collaboration with clinicians and healthcare providers to ensure regulatory compliance, patient safety, and ease of integration into existing medical workflows. Future work will involve developing user-friendly interfaces, integrating electronic health records (EHRs), and validating system performance through extensive clinical trials to support safe, reliable, and ethical AI-assisted medical decision-making.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

CJ: Conceptualization, methodology, writing – original draft. QY: software, validation, writing – original draft. XY: Methodology, Supervision, Project administration, Validation, Resources, Visualization, Writing – original draft, Writing – review and editing. LL: Data curation, writing – original draft. WH: Writing – original draft, writing – review and editing, visualization. WL: Writing – original draft, Writing – review and editing, Data curation, Conceptualization, Formal analysis, Investigation, Funding acquisition, Software.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China [grant number 81774266], the Second National Famous Traditional Chinese Medicine Practitioner Inheritance Workshop [grant number National Office of Traditional Chinese Medicine Human Education Letter (2022) No. 245], Wu Mianhua National Famous Elderly Chinese Medicine Experts Inheritance Workshop [grant number National Traditional Chinese Medicine Human Education Letter (2022) No. 75], Wu Mianhua Jiangsu Famous Elderly Chinese Medicine Experts Inheritance Workshop [grant number Jiangsu Chinese Medicine Science and Education (2021) No. 7], The Seventh Batch of National Old Chinese Medicine Experts' Academic Experience Inheritance Work Program of the State Administration of Traditional Chinese Medicine (SATCM) [grant number National TCM Human Education Letter (2022) No. 76], and Graduate Student Research and Practice Innovation Program in Jiangsu Province [grant number SJCX23_0875].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Liu H, Chen K, Li Y, Huang Z, Duan J, Ma J Integrated behavior planning and motion control for autonomous vehicles with traffic rules compliance. In: IEEE International Conference on Robotics and Biomimetics; 04-09 December 2023; Koh Samui, Thailand. IEEE (2023).

2. Huang Z, Liu H, Wu J, Lv C Conditional predictive behavior planning with inverse reinforcement learning for human-like autonomous driving. *IEEE Trans Intell Transportation Syst* (2022) 24:7244–58. doi:10.1109/tits.2023.3254579

3. Klimke M, Völz B, Buchholz M Cooperative behavior planning for automated driving using graph neural networks. In: *IEEE intelligent vehicles symposium (V)* (2022).

4. Qiao Z, Schneider J, Dolan J Behavior planning at urban intersections through hierarchical reinforcement learning. In: IEEE International Conference on Robotics and Automation; 30 May 2021 - 05 June 2021; Xi'an, China. IEEE (2020).

5. Li J, Sun L, Zhan W, Tomizuka M *Interaction-aware behavior planning for autonomous vehicles validated with real traffic data* (2020).

6. Esterle K, Kessler T, Knoll A Optimal behavior planning for autonomous driving: a generic mixed-integer formulation. In: *IEEE intelligent vehicles symposium (IV)* (2020).

7. Janner M, Du Y, Tenenbaum J, Levine S Planning with diffusion for flexible behavior synthesis. In: International Conference on Machine Learning; Baltimore, Maryland, USA. PMLR (2022). Available online at: https://arxiv.org/abs/2205.09991.

8. Ahmed N, Li C, Khan A, Qalati SA, Naz S, Rana F Purchase intention toward organic food among young consumers using theory of planned behavior: role of environmental concerns and environmental awareness. *J Environ Plann Management* (2020) 64:796–822. doi:10.1080/09640568.2020.1785404

9. Ding W, Zhang L, Chen J, Shen S Epsilon: an efficient planning system for automated vehicles in highly interactive environments. *IEEE Trans Robotics* (2021) 38:1118–38. doi:10.1109/tro.2021.3104254

10. Lavuri R Extending the theory of planned behavior: factors fostering millennials' intention to purchase eco-sustainable products in an emerging market. *J Environ Plann Management* (2021) 65:1507–29. doi:10.1080/09640568.2021.1933925

11. Hagger M, Smith SR, Keech JJ, Moyers SA, Hamilton K Predicting social distancing intention and behavior during the covid-19 pandemic: an integrated social cognition model. *Ann Behav Med* (2020) 54:713–27. doi:10.1093/abm/kaaa073

12. Hamilton K, van Dongen A, Hagger M An extended theory of planned behavior for parent-for-child health behaviors: a meta-analysis. *Health Psychol* (2020) 39:863–78. doi:10.1037/hea0000940

13. Zhu S, Aksun-Guvenc B Trajectory planning of autonomous vehicles based on parameterized control optimization in dynamic on-road environments. *J Intell Robotic Syst* (2020) 100:1055–67. doi:10.1007/s10846-020-01215-y

14. Salzmann T, Ivanovic B, Chakravarty P, Pavone M Trajectron++: dynamically-feasible trajectory forecasting with heterogeneous data. In: European Conference on Computer Vision (2020). p. 683–700. doi:10.1007/978-3-030-58523-5_40

15. Zhang C, Fang R, Zhang R, Hagger M, Hamilton K Predicting hand washing and sleep hygiene behaviors among college students: test of an integrated social-cognition model. *Int J Environ Res Public Health* (2020) 17:1209. doi:10.3390/ijerph17041209

16. Park J, O'Brien JC, Cai CJ, Morris M, Liang P, Bernstein MS Generative agents: interactive simulacra of human behavior. In: ACM Symposium on User Interface Software and Technology (2023). p. 1–22. doi:10.1145/3586183.3606763

17. Ajzen I. The theory of planned behavior: frequently asked questions. *Hum Behav Emerging Tech* (2020) 2:314–24. doi:10.1002/hbe2.195

18. Han H Consumer behavior and environmental sustainability in tourism and hospitality: a review of theories, concepts, and latest research. *J Sustainable Tourism* (2021) 29:1021–42. doi:10.1080/09669582.2021.1903019

19. Hagger M, Cheung M, Ajzen I, Hamilton K Perceived behavioral control moderating effects in the theory of planned behavior: a meta-analysis. *Health Psychol* (2022) 41:155–67. doi:10.1037/hea0001153

20. Bošnjak M, Ajzen I, Schmidt P The theory of planned behavior: selected recent advances and applications. *Europe's J Psychol* (2020) 16:352–6. doi:10.5964/ejop.v16i3.3107

21. Yuriev A, Dahmen M, Paillé P, Boiral O, Guillaumie L Pro-environmental behaviors through the lens of the theory of planned behavior: a scoping review. *Resour Conservation Recycling* (2020) 155:104660. doi:10.1016/j.resconrec.2019.104660

22. Gioia GA, Espy KA, Isquith PK *BRIEF® -P Behavior rating inventory of executive function, preschool version*. PAR (2020). doi:10.1037/t73087-000

23. Barbera FL, Ajzen I Control interactions in the theory of planned behavior: rethinking the role of subjective norm. *Europe's J Psychol* (2020) 16:401–17. doi:10.5964/ejop.v16i3.2056

24. Qi G, Hu G, Mazur N, Liang H, Haner M A novel multi-modality image simultaneous denoising and fusion method based on sparse representation. *Computers* (2021) 10:129. doi:10.3390/computers10100129

25. Qi G, Zhu Z Blockchain and artificial intelligence applications. *J Artif Intelligence Technology* (2021) 1:83. doi:10.37965/2021.0019

26. Sadat A, Casas S, Ren M, Wu X, Dhawan P, Urtasun R Perceive, predict, and plan: safe motion planning through interpretable semantic representations. In: European Conference on Computer Vision (2020). p. 414–30. doi:10.1007/978-3-030-58592-1_25

27. Taing HB, Chang Y Determinants of tax compliance intention: focus on the theory of planned behavior. *Int J Public Adm* (2020) 44:62–73. doi:10.1080/01900692.2020.1728313

28. Hang P, Lv C, Huang C, Cai J, Hu Z, Xing Y An integrated framework of decision making and motion planning for autonomous vehicles considering social behaviors. *IEEE Trans Vehicular Technology* (2020) 69:14458–69. doi:10.1109/tvt.2020.3040398

29. Qi G, Zhang Y, Wang K, Mazur N, Liu Y, Malaviya D Small object detection method based on adaptive spatial parallel convolution and fast multi-scale fusion. *Remote Sensing* (2022) 14:420. doi:10.3390/rs14020420

30. Qian S, Chang Y Learning a cross-scale cross-view decoupled denoising network by mining omni-channel information. *Front Phys* (2025) 13:1498335. doi:10.3389/fphy.2025.1498335

31. Barbera FL, Ajzen I Moderating role of perceived behavioral control in the theory of planned behavior: a preregistered study. *J Theor Social Psychol* (2020) 5:35–45. doi:10.1002/jts5.83

32. He S, Li Y, Liang J, Wei L Quantum coherence and the bell inequality violation: a numerical experiment with the cavity qeds. *Front Phys* (2025) 13:1541888. doi:10.3389/fphy.2025.1541888

33. Rohrer JL Behavior plan. In: *Encyclopedia of autism spectrum disorders* (2020).

34. Gu J, Wang J, He M, Yang S, Li S Research on relaxation characteristics of columnar jointed basalts of deep foundation in hydropower station. *Front Phys* (2025) 13:1522240. doi:10.3389/fphy.2025.1522240

35. Welch G, Bishop G *An introduction to the kalman filter*. University of North Carolina at Chapel Hill, Department of Computer Science (2006). Available online at: https://gitea.auro.re/Krobot/krobotdocumentation/raw/commit/155a9eba0c1f173ff70f6e27361f642aa22cfb98/docs/doc/kalman.pdf.

36. Julier SJ, Uhlmann JK, Durrant-Whyte HF Sigma-point kalman filters for nonlinear estimation and sensor-fusion: applications to integrated navigation. *Proc IEEE* (2007) 95:901–12. doi:10.2514/6.2007-6514

37. Rashidi P, Cook DJ Bayesian sensor fusion for context-aware human activity recognition using heterogeneous sensors. *ACM Trans Sensor Networks (Tosn)* (2014) 10:1–21. Available online at: https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.1201/b17124&type=googlepdf.

38. Castanedo F A review of multisensor data fusion solutions in smart manufacturing: systems and applications. *J Sensor Actuator Networks* (2016) 5:4. Available online at: https://www.mdpi.com/1424-8220/22/5/1734.

39. Horn B, Kreuch T, Lauer M, Stiller C Multimodal sensor fusion for urban automated driving using Gaussian mixture models. In: *IEEE intelligent vehicles symposium (V)* (2017). p. 558–64.

40. Zhang Q, Wang Y, Xiang B, Liu X Multi-modal sensor fusion using Gaussian mixture models applied to environment perception for autonomous driving. *IEEE Sensors J* (2020) 20:5662–70. Available online at: https://arxiv.org/abs/2202.02703.

41. Hind S, van der Vlist FN, Kanderske M Challenges as catalysts: how waymo's open dataset challenges shape ai development. *AI Soc* (2024) 40:1667–83. doi:10.1007/s00146-024-01927-x

42. Mi Y, Ji Y, Wang K, Wang Y, Shen T, Wang K Lot-nuscenes: a virtual long-tail scenario dataset for parallel vision and parallel vehicles. In: 2024 IEEE 4th International Conference on Digital Twins and Parallel Intelligence (DTPI); 18-20 October 2024; Wuhan, China. IEEE (2024). p. 194–9.

43. Li G, Jiao Y, Calvert SC, van Lint JH Lateral conflict resolution data derived from argoverse-2: analysing safety and efficiency impacts of autonomous vehicles at intersections. *Transportation Res C: Emerging Tech* (2024) 167:104802. doi:10.1016/j.trc.2024.104802

44. Yang R, Peng Y Ploc: a new evaluation criterion based on physical location for autonomous driving datasets. In: 2024 12th International Conference on Intelligent Control and Information Processing (ICICIP); 08-10 March 2024; Nanjing, China. IEEE (2024). p. 116–22.

45. Vaishya R, Misra A, Vaish A, Ursino N, D'Ambrosi R Hand grip strength as a proposed new vital sign of health: a narrative review of evidences. *J Health Popul Nutr* (2024) 43:7. doi:10.1186/s41043-024-00500-y

46. Luo F, Zhou T, Liu J, Guo T, Gong X, Gao X Dcenet: diff-feature contrast enhancement network for semi-supervised hyperspectral change detection. *IEEE Trans Geosci Remote Sensing* (2024) 62:1–14. doi:10.1109/tgrs.2024.3374600

47. Toros K, Kozmenko O, Falch-Eriksen A "I just want to go home, is what i need"–voices of Ukrainian refugee children living in Estonia after fleeing the war. *Child Youth Serv Rev* (2024) 158:107461. doi:10.1016/j.childyouth.2024.107461

48. Li F, Qi J-J, Li L-X, Yan T-F Mthfr c677t, mthfr a1298c, mtrr a66g and mtr a2756g polymorphisms and male infertility risk: a systematic review and meta-analysis. *Reprod Biol Endocrinol* (2024) 22:133. doi:10.1186/s12958-024-01306-7

49. Galic I, Bez C, Bertani I, Venturi V, Stankovic N Herbicide-treated soil as a reservoir of beneficial bacteria: microbiome analysis and pgp bioinoculants in maize. *Environ Microbiome* (2024) 19:107. doi:10.1186/s40793-024-00654-6

50. Song P, Li P, Aertbeliën E, Detry R Robot trajectron: trajectory prediction-based shared control for robot manipulation. In: 2024 IEEE International Conference on Robotics and Automation (ICRA); 13-17 May 2024; Yokohama, Japan. IEEE (2024). p. 5585–91.

51. Khodarahmi M, Maihami V A review on kalman filter models. *Arch Comput Methods Eng* (2023) 30:727–47. doi:10.1007/s11831-022-09815-7

52. Dai H, Pollock M, Roberts GO Bayesian fusion: scalable unification of distributed statistical analyses. *J R Stat Soc Ser B: Stat Methodol* (2023) 85:84–107. doi:10.1093/jrsssb/qkac007

53. Naseer A, Alzahrani HA, Almujally NA, Al Nowaiser K, Al Mudawi N, Algarni A, et al. Efficient multi-object recognition using gmm segmentation feature fusion approach. *IEEE Access* (2024) 12:37165–78. doi:10.1109/access.2024.3372190

54. Lei M, Yang D, Weng X Integrated sensor fusion based on 4d mimo radar and camera: a solution for connected vehicle applications. *IEEE Vehicular Technology Mag* (2022) 17:38–46. doi:10.1109/mvt.2022.3207453