**Айгуль Кайрулаевна Шайханова** – PhD, ассоциированный профессор, Евразийский национальный университет имени Л.Н. Гумилева, г. Астана, Казахстан; e-mail: shaikhanova_ak@enu.kz. ORCID: https://orcid.org/0000-0001-6006-4813.

**Information about the authors**

**Bauyrzhan Smakanov** – PhD, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan; e-mail: bauka.10.10@mail.ru. ORCID: https://orcid.org/0000-0002-4343-0463.

**Indira Uvaliyeva*** – PhD, Associate Professor, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan; e-mail: iuvalieva@ektu.kz. ORCID: https://orcid.org/0000-0002-2117-5390.

**Aigul Shaikhanova** – PhD, Associate Professor, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan; e-mail: shaikhanova_ak@enu.kz. ORCID: https://orcid.org/0000-0001-6006-4813.

*Редакцияға енуі 02.06.2026*
*Өңдеуден кейін түсуі 19.08.2026*
*Жариялауға қабылданды 25.08.2026*

https://doi.org/10.53360/2788-7995-2026-3(23)-21

IRSTI: 20.53.19

**I. Bainazarova[1], Zh. Umarova[1]*, Yu. Jusoh[2], Zh. Iztayev[1], S. Botayeva[3]**
[1]South Kazakhstan Research University, Tauke Khan Avenue, 5, Shymkent, Kazakhstan
[2]Universiti Putra Malaysia, Serdang, Selangor, Malaysia
[3]Tashenev University, Konayev Avenue, 16, Shymkent, Kazakhstan
*e-mail: Zhanat-u@mail.ru

# CONTEXT-DEPENDENT TRAJECTORY FORECASTING FOR AUTONOMOUS TRANSPORT SYSTEMS USING HYBRID GRAPH AND TRANSFORMER NEURAL NETWORKS

***Abstract:*** *Trajectory forecasting is a critical component of autonomous transport systems because it supports conflict assessment and safe motion planning in dynamic road environments. This study proposes a hybrid forecasting architecture that combines adaptive graph-based modeling of local agent interactions with a transformer module that encodes the global scene context. The model was evaluated on the Argoverse 2 Motion Forecasting and Waymo Open Motion Dataset benchmarks using ADE, FDE, Miss Rate, minADE, and minFDE, together with analyses of robustness, interaction density, and inference time. On Argoverse 2, the proposed model reduced ADE from 1.21 to 1.17, FDE from 2.81 to 2.69, and Miss Rate from 0.25 to 0.23 compared with Motion Transformer. On Waymo, the improvement was smaller but consistent: ADE decreased from 1.58 to 1.53 and FDE from 3.82 to 3.69. The largest gain was observed in high-density interaction scenes, where ADE decreased from 1.52 to 1.45. These results indicate that adaptive interaction weighting improves forecasting robustness in complex multi-agent road scenarios, although it increases computational cost.*

***Key words:*** *autonomous transport; trajectory forecasting; deep learning; graph neural networks; transformers; multi-agent interaction; computer vision; motion prediction.*

## Introduction

Autonomous transport systems operate in environments where the future trajectory of one road user is rarely independent of surrounding agents, road geometry, and potential motion conflicts. A forecasting error may propagate to the planning layer and lead to an unsafe maneuver. For this reason, trajectory forecasting remains a central component of autonomous driving. Recent methods have evolved from recurrent models and social pooling approaches to graph- and attention-based architectures that model inter-agent relationships and scene-level context [1-5]. A remaining problem is the context-dependent selection of relevant interactions. In many existing models, neighboring agents are processed either through a predefined graph structure or through a general attention mechanism. Such representations do not always distinguish between agents that genuinely affect the future trajectory and agents whose presence has little influence on the forecast. In practice, interaction relevance depends on several factors at once: distance, relative velocity, heading

©I. Bainazarova, Zh. Umarova, Yu. Jusoh, Zh. Iztayev, S. Botayeva, 2026

difference, scene density, and the likelihood of trajectory conflict. Insufficient differentiation among interactions may therefore reduce forecasting robustness in dense traffic, at intersections, and during late lane changes.

This study proposes a hybrid trajectory forecasting architecture in which local interactions are modeled through an adaptively weighted graph, while global scene context is encoded by a transformer module. Unlike Motion Transformer, which uses attention-based intention localization and motion refinement [5], the proposed model explicitly computes inter-agent interaction weights from distance, relative velocity, and heading difference. Also, unlike conventional graph-based models such as Social-STGCNN [2], the interaction graph does not use static edge weights: the contribution of neighboring agents changes with the current scene configuration.

The novelty of the study lies in the adaptive weighting of inter-agent interactions and its integration with transformer-based global context encoding. The contribution is not the mere combination of a GNN and a transformer, but the context-dependent identification of agents that exert the strongest influence on the target agent's future trajectory. This is particularly relevant in high-density traffic, late lane changes, and ambiguous maneuvers.

The experimental evaluation was conducted on Argoverse 2 Motion Forecasting and the Waymo Open Motion Dataset. Both datasets are designed for autonomous driving research and include multi-agent road scenes, trajectories of different road users, and map-related information. Compared with pedestrian-oriented datasets such as ETH/UCY, they are better aligned with forecasting tasks in autonomous driving. Other datasets, including nuScenes and INTERACTION, are also relevant for motion prediction; however, they were not used in this study in order to keep the evaluation focused on large-scale motion forecasting scenarios with high-definition (HD) maps and comparable validation protocols [6-9].

The aim of this study was to develop and evaluate a hybrid trajectory forecasting model for autonomous transport that combines adaptive graph-based modeling of local interactions with attention-based encoding of global scene context.

The study addressed the following objectives:

1. to develop an adaptive interaction weighting mechanism based on distance, relative velocity, and heading difference;
2. to design a hybrid architecture combining a graph module and a transformer-based context encoder;
3. to compare the proposed model with Social-LSTM, Social-STGCNN, VectorNet, and Motion Transformer on Argoverse 2 and the Waymo Open Motion Dataset;
4. to evaluate forecasting quality using ADE, FDE, MR, minADE, and minFDE;
5. to analyze robustness to input noise and the change in inference time as the number of agents in a scene increases.

**Materials and Methods**

The experimental evaluation was conducted on two open datasets for motion forecasting in autonomous driving: Argoverse 2 Motion Forecasting and the Waymo Open Motion Dataset (Table 1). Both datasets contain multi-agent road scenes, agent trajectories, object categories, and map-related information. They were processed separately because they differ in temporal structure, data splits, and evaluation protocols.

Table 1 – Characteristics of the datasets used in the study

| Dataset | Split | Temporal horizon | Frequency | Object types | HD maps | Use in the study |
|---|---|---|---|---|---|---|
| Argoverse 2 Motion Forecasting | train: 199,908; validation: 24,988; test: 24,984 | 11 s; typical protocol: 5 s observation and 6 s prediction | 10 Hz | vehicles, pedestrians, cyclists, motorcyclists, buses, etc. | yes | urban scenes with diverse road geometry |
| Waymo Open Motion Dataset | train: 486,995; validation: 44,097; test: 44,920 | 10 historical steps + current step + 80 future steps; 11 observed states were used as input | 10 Hz | vehicles, pedestrians, cyclists | yes | dense multi-agent road scenes |

All locally computed metrics were obtained from the validation splits. The test splits were not used because the future trajectories in the Waymo test split are withheld for server-side evaluation.

For each scenario, agent coordinates, velocity, heading, object type, and map fragments were extracted. Coordinates were normalized with respect to the target agent: the scene was centered at the target agent's position in the last observed frame and rotated into the agent-centric coordinate frame. For scene-complexity analysis, the number of active agents, the mean nearest-neighbor distance, the share of agents within the interaction radius, and the presence of potentially conflicting motion directions were also computed.

The baseline models represented sequential, graph-based, map-aware, and attention-based approaches to trajectory forecasting (Table 2).

Table 2 – Architectural characteristics of the compared models

| Model | Architecture type | Input representation | Interaction modeling | Map use | Main mechanism |
|---|---|---|---|---|---|
| Social-LSTM [1] | RNN | coordinate sequences | social pooling | no | temporal dependency modeling |
| Social-STGCNN [2] | GNN | spatio-temporal graph | explicit graph-based interaction | no | graph convolutions |
| VectorNet [10] | hybrid architecture | vectorized trajectories and HD map | indirect | yes | hierarchical polyline encoding |
| Motion Transformer [5] | Transformer | tokenized scene representation | attention-based | yes | self-attention |
| Proposed model | Hybrid GNN + Transformer | interaction graph and spatial scene context | adaptive interaction weighting | yes | joint modeling of local and global context |

This set of models was used to examine whether adaptive interaction weighting provides an additional benefit over strong architectural baselines, rather than only over a simple sequential model.

Each road scene was represented as a directed graph, where nodes corresponded to road users and edges denoted potential interactions. Directed edges were added only between agents located within the specified interaction radius: 50 m for vehicles and 20 m for pedestrians and cyclists. Before computing edge weights, all input features were normalized to reduce dependence on measurement units.

The edge weight between agents (i) and (j) was computed as follows:

$$\boldsymbol{w_{ij} = \sigma(\beta_0 + \alpha_1 \widetilde{d_{ij}} + \alpha_2 \Delta\widetilde{v_{ij}} + \alpha_3 \widetilde{\theta_{ij}})} \quad \textbf{(1)}$$

where $w_{ij}$ is the interaction weight, σ(·) is the sigmoid function, $\beta_0$ is the bias term, and $\alpha_1$, $\alpha_2$, $\alpha_3$ are feature coefficients.

The normalized features were defined as

$$\widetilde{d_{ij}} = \exp\left(-\frac{d_{ij}}{R_c}\right), \Delta\widetilde{v_{ij}} = \min\left(\frac{\|\Delta v_{ij}\|}{v_{95}}, 1\right), \widetilde{\theta_{ij}} = \frac{1+\cos(\Delta\theta_{ij})}{2} \quad (2)$$

Here, $d_{ij}$ is the distance between agents, $R_c$ is the interaction radius, $\Delta v_{ij}$ is the relative speed magnitude, $v_{95}$ is the 95th percentile of relative speeds in the training split, and $\Delta v_{ij}$ is the heading difference. The coefficients $\beta_0$ and $\alpha_1$, $\alpha_2$, $\alpha_3$ were tuned by grid search on the validation split. Fifty configurations were evaluated, and the configuration that achieved the lowest minADE on the validation split was selected (Table 3).

Table 3 – Parameters of the interaction weighting function

| Parameter | Meaning | Setting |
|---|---|---|
| $\beta_0$ | baseline interaction level | grid search, 50 trials on validation split |
| $\alpha_1$ | distance contribution | grid search, 50 trials on validation split |
| $\alpha_2$ | relative velocity contribution | grid search, 50 trials on validation split |
| $\alpha_3$ | heading contribution | grid search, 50 trials on validation split |
| $R_c$ | interaction radius | 50 m for vehicles; 20 m for pedestrians and cyclists |
| $v_{95}$ | velocity normalization coefficient | 95th percentile of relative speeds in the training split |

The target agent is connected to surrounding vehicles through weighted graph edges. Edge weights $w_{ij}$ are computed from normalized distance, relative velocity, and heading difference according to Equation (1). Higher values indicate a stronger influence on the predicted future trajectory of the target agent (Figure 1).

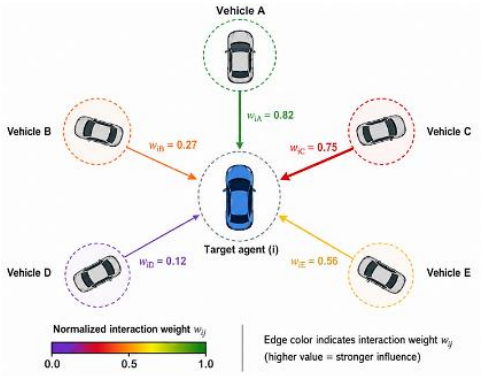



Figure 1 – Adaptive interaction weighting between neighboring agents

***Hybrid Model Architecture***

The model consisted of three components: a graph-based local encoder, a transformer-based global context module (Table 4), and a multimodal trajectory decoder. The graph module used the interaction weights $w_{ij}$ from Equation (1) and updated each node representation as

$$h_i^{(l+1)} = \phi(h_i^{(l)}, \sum_{j \in N(i)} w_{ij} * \psi(h_i^{(l)}, h_j^{(l)})) \quad (3)$$

The function $\psi$ $(\cdot)$ was implemented as a two-layer MLP for generating inter-agent messages, whereas $\phi(\cdot)$ was implemented as a two-layer MLP with a residual connection and LayerNorm to update the node states. After graph encoding, the features were passed to the transformer module:

$$\boldsymbol{Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V} \quad (4)$$

The decoder generated $K = 6$ future trajectory modes and their corresponding probabilities. The loss combined the error of the best-matching trajectory with a classification loss for mode selection:

$$\boldsymbol{L = \min_{k \in K} \frac{1}{T} \sum_{t=1}^{T} \left\| \hat{p}_t^k - p_t \right\|_2 + \lambda L_{cls}} \quad (5)$$

$$\boldsymbol{L_{cls} = -log \frac{\exp(s_{k^*})}{\sum_{k=1}^{K} \exp(s_k)}} \quad (6)$$

$$\boldsymbol{k^* = arg \min_{K \in \{1,..,K\}} \frac{1}{T} \sum_{t=1}^{T} \left\| \widehat{p_t^k} - p_t \right\|_2} \quad (7)$$

In the main configuration, $\lambda$ = 0.2. This value was selected on the validation split from {0.1, 0.2, 0.5, 1.0}.

Table 4 – Main parameters of the proposed model

| Parameter | Value |
|---|---|
| Number of graph layers | 3 |
| Hidden dimension $h_i$ | 128 |
| Implementation of ψ(·) | two-layer MLP, 128 hidden units, ReLU |
| Implementation of ϕ(·) | two-layer MLP + residual connection + LayerNorm |
| Number of transformer layers | 4 |
| Number of attention heads | 8 |
| Key dimension $d_k$ | 64 |
| Dropout | 0.1 |
| Number of trajectory modes *K* | 6 |
| Loss coefficient $\lambda$ | 0.2 |
| Decoder | MLP trajectory head + classification head |
| Optimizer | AdamW |

The architecture consists of the following sequence of blocks: Trajectory History → Scene Encoding → Interaction Graph Construction → Graph Neural Network → Transformer Context Module → Feature Fusion → Multi-Modal Trajectory Decoder → Predicted Future Trajectories (Figure 2).

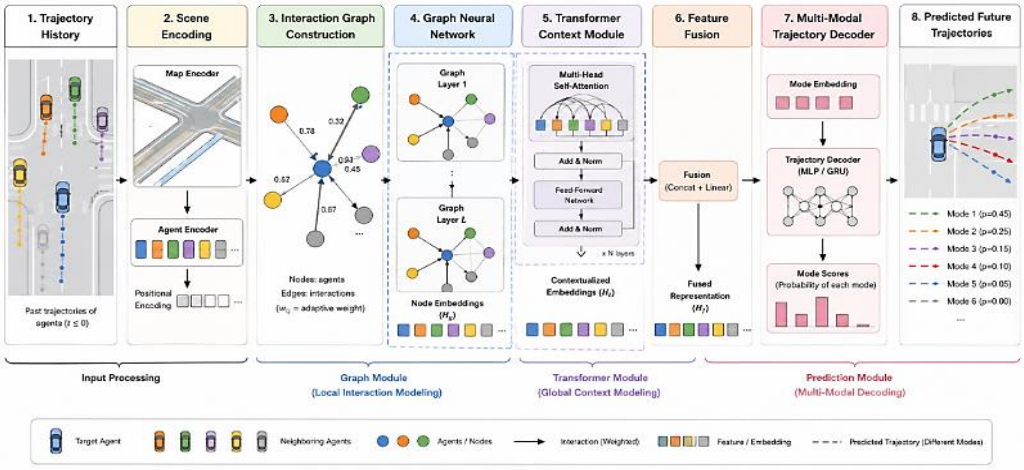



Figure 2 – Overall architecture of the hybrid trajectory forecasting system

Forecasting quality was evaluated using ADE, FDE, MR, minADE, and minFDE. Since the model generated multiple future trajectories, ADE and FDE were computed for the most probable mode, whereas minADE and minFDE were computed for the best-matching trajectory among all *K* modes.

$$k^{prob} = arg \max_{K\in\{1,..,K\}} s_k \quad (8)$$

$$ADE = \frac{1}{T}\sum_{t=1}^{T} \left\| \hat{p}_t^{prob} - p_t \right\|_2 \quad FDE = \left\| \hat{p}_t^{prob} - p_T \right\|_2 \quad (9)$$

$$\min ADE = \min_{k\in K} \frac{1}{T}\sum_{t=1}^{T} \left\| \hat{p}_t^{k} - p_t \right\|_2, \min FDE = \min_{k\in K} \frac{1}{T} \left\| \hat{p}_t^{k} - p_T \right\|_2 \quad (10)$$

Miss Rate was defined as the proportion of predictions whose final displacement error exceeded a fixed threshold:

$$MR = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(FDE_i > \tau) \quad (11)$$

For the Waymo Open Motion Dataset, MR was computed using the WOMD evaluator protocol, in which the thresholds depend on the agent type and evaluation horizon. Therefore, MR values were compared only within each dataset and were not interpreted as directly comparable between Argoverse 2 and Waymo.

Experiments were conducted separately on Argoverse 2 Motion Forecasting and the Waymo Open Motion Dataset. All reported numerical results were computed on the validation split. The test splits were not used for local evaluation, since future trajectories in the Waymo test split are hidden and used for server-side leaderboard evaluation (Table 5).

Table 5 – Main experimental protocol

| Parameter | Value |
|---|---|
| Datasets | Argoverse 2 Motion Forecasting; Waymo Open Motion Dataset |
| Argoverse 2 split | train: 199,908; validation: 24,988; test: 24,984 |
| Waymo Open Motion split | train: 486,995; validation: 44,097; test: 44,920 |
| Basis of numerical results | validation split for both datasets |
| Test split | not used for local evaluation; Waymo test has hidden future ground truth |
| Data frequency | 10 Hz |
| Main metrics | ADE, FDE, MR, minADE, minFDE |
| Miss Rate threshold | Argoverse 2: $\tau$ = 2.0 m; Waymo: WOMD evaluator protocol |
| Random seeds | 13, 21, 42, 77, 101 |
| Optimizer | AdamW |
| Learning rate range | $1 \times 10^{-4}$ to $5 \times 10^{-4}$ |
| Batch size | 16–64, depending on model and GPU memory |
| Early stopping | validation minADE; (patience = 10) epochs |
| Hyperparameter budget | 50 trials per model |
| Unit of analysis | scene / target agent |
| Additional analyzes | noise, interaction density, inference time |

Each model used the same hyperparameter budget of 50 trials. Training was stopped if validation minADE did not improve for 10 epochs. Hyperparameters were not tuned on the test split.

Robustness was evaluated by adding Gaussian noise only to the observed part of the trajectory; the future ground truth was unchanged. The main robustness indicator was the change in ADE relative to the no-noise condition.

Computational efficiency was assessed by inference time per scene and peak GPU memory usage. All measurements were performed on a single NVIDIA A100 40GB GPU with batch size = 1 and float32 precision, using PyTorch 2.1 and CUDA 12.1. Inference time excluded data loading and scene preprocessing.

For each model, the mean, standard deviation, median, interquartile range, and 95% confidence interval were computed. Comparisons were conducted separately for each dataset on the validation split.

When error distributions departed from normality, the Friedman test for related samples was used; each scene served as a block across the compared models. Kendall's W was calculated to estimate the consistency of model differences. Post hoc pairwise comparisons were adjusted for multiple testing. In addition, the distribution tails and high-error scenes were analyzed, because rare large errors may be more critical in autonomous driving than small improvements in mean error.

**Results**

Model performance was evaluated on the validation splits of Argoverse 2 Motion Forecasting and the Waymo Open Motion Dataset using ADE, FDE, MR, minADE, and minFDE. Lower values indicate better forecasting performance (Table 6). The reported results should not be directly compared with public leaderboard scores because all models were trained and evaluated under a unified local protocol without server-side submission or leaderboard-specific optimization.

Table 6 – Performance comparison on the Argoverse 2 Motion Forecasting validation split

| Model | ADE ↓ | FDE ↓ | MR ↓ | minADE ↓ | minFDE ↓ |
|---|---|---|---|---|---|
| Social-LSTM | 1.93 ± 0.07 | 4.37 ± 0.16 | 0.41 ± 0.02 | 1.76 ± 0.06 | 3.92 ± 0.15 |
| Social-STGCNN | 1.61 ± 0.06 | 3.72 ± 0.14 | 0.35 ± 0.02 | 1.48 ± 0.05 | 3.28 ± 0.12 |
| VectorNet | 1.34 ± 0.05 | 3.08 ± 0.11 | 0.29 ± 0.02 | 1.22 ± 0.04 | 2.71 ± 0.10 |
| Motion Transformer | 1.21 ± 0.04 | 2.81 ± 0.10 | 0.25 ± 0.01 | 1.08 ± 0.04 | 2.46 ± 0.09 |
| Proposed Model | 1.17 ± 0.04 | 2.69 ± 0.09 | 0.23 ± 0.01 | 1.03 ± 0.03 | 2.31 ± 0.08 |

*Note: Values are reported as mean ± standard deviation across five independent runs.*

On Argoverse 2, the proposed model achieved the lowest error across all evaluation metrics. Compared with Motion Transformer, ADE decreased from 1.21 to 1.17 m, FDE from 2.81 to 2.69 m, and MR from 0.25 to 0.23. Although the improvement was moderate, it remained consistent across all metrics (Table 7).

Table 7 – Performance comparison on the Waymo Open Motion Dataset validation split

| Model | ADE ↓ | FDE ↓ | MR ↓ | minADE ↓ | minFDE ↓ |
|---|---|---|---|---|---|
| Social-LSTM | 2.48 ± 0.09 | 5.94 ± 0.21 | 0.49 ± 0.02 | 2.22 ± 0.08 | 5.21 ± 0.19 |
| Social-STGCNN | 2.11 ± 0.08 | 5.07 ± 0.18 | 0.44 ± 0.02 | 1.89 ± 0.07 | 4.58 ± 0.17 |
| VectorNet | 1.76 ± 0.06 | 4.13 ± 0.15 | 0.37 ± 0.02 | 1.55 ± 0.06 | 3.66 ± 0.13 |
| Motion Transformer | 1.58 ± 0.05 | 3.82 ± 0.13 | 0.33 ± 0.01 | 1.39 ± 0.05 | 3.37 ± 0.12 |
| Proposed Model | 1.53 ± 0.05 | 3.69 ± 0.12 | 0.31 ± 0.01 | 1.34 ± 0.05 | 3.24 ± 0.11 |

Errors were generally higher on Waymo, which may reflect differences in scene complexity, traffic density, and uncertainty in future motion. Nevertheless, the proposed model maintained a consistent advantage over Motion Transformer, reducing ADE from 1.58 to 1.53 m, FDE from 3.82 to 3.69 m, and MR from 0.33 to 0.31.

The Friedman test was applied to compare ADE distributions across models for each dataset separately (Table 8).

Table 8 – Statistical comparison of models based on ADE

| Dataset | Friedman $\chi^2$ | p-value | Kendall's W |
|---|---|---|---|
| Argoverse 2 | 41.82 | < 0.001 | 0.23 |
| Waymo | 37.15 | < 0.001 | 0.19 |

Although the overall differences were statistically significant, Kendall's W values below 0.30 indicate a small overall effect of model choice across scenes. Consequently, the results should be interpreted as differences in error distributions rather than evidence that one architecture is consistently superior across all scenarios.

Interaction density (Table 9) was defined by the number of neighboring agents within the interaction radius of the target agent: low density (≤ 5 agents), medium density (6-15 agents), and high density (> 15 agents).

Table 9 – ADE as a function of interaction density

| Interaction Density | Social-LSTM | Social-STGCNN | VectorNet | Motion Transformer | Proposed Model |
|---|---|---|---|---|---|
| Low (≤ 5 agents) | 1.47 | 1.32 | 1.18 | 1.12 | 1.10 |
| Medium (6-15 agents) | 1.89 | 1.61 | 1.36 | 1.23 | 1.18 |
| High (> 15 agents) | 2.54 | 2.03 | 1.69 | 1.52 | 1.45 |

Differences between Motion Transformer and the proposed model were minimal in sparse scenes. As interaction density increased, the performance gap widened, suggesting that adaptive graph-based interaction modeling is particularly beneficial when multiple neighboring agents simultaneously influence the target trajectory.

Gaussian noise was added only to the observed trajectory segment, while future ground-truth trajectories remained unchanged (Table 10).

Table 10 – Robustness to input noise measured by ADE

| Noise Level | Social-LSTM | Social-STGCNN | VectorNet | Motion Transformer | Proposed Model |
|---|---|---|---|---|---|
| 0% | 1.93 | 1.61 | 1.34 | 1.21 | 1.17 |
| 5% | 2.18 | 1.79 | 1.47 | 1.32 | 1.27 |
| 10% | 2.51 | 2.03 | 1.68 | 1.49 | 1.42 |
| 15% | 2.94 | 2.37 | 1.95 | 1.71 | 1.63 |

Performance deteriorated for all models as noise increased. Social-LSTM exhibited the fastest degradation, whereas Motion Transformer and the proposed model remained more robust. Across all noise levels, the proposed architecture consistently achieved the lowest ADE (Figure 3).

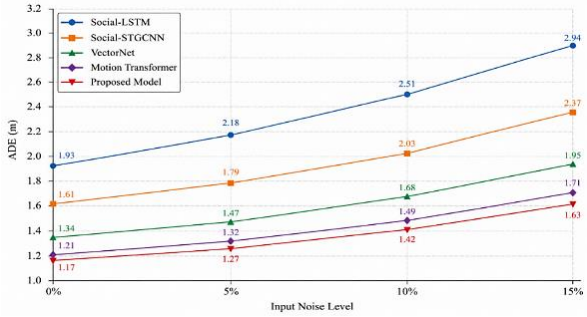



Figure 3 – Effect of input noise on ADE

Curves illustrate the increase in ADE at noise levels of 0%, 5%, 10%, and 15% for all evaluated models.

Inference benchmarks were conducted on an NVIDIA A100 GPU with 40 GB of memory using batch size of 1, float32 precision, PyTorch 2.1, and CUDA 12.1. Reported inference times exclude data loading and preprocessing (Table 11).

Table 11 – Inference time per scene (ms)

| Number of Agents | Social-LSTM | Social-STGCNN | VectorNet | Motion Transformer | Proposed Model |
|---|---|---|---|---|---|
| ≤10 | 4.8 | 7.6 | 9.2 | 11.4 | 12.7 |
| 11–25 | 6.5 | 11.8 | 13.7 | 18.2 | 19.6 |
| 26–50 | 9.4 | 19.3 | 21.5 | 31.8 | 33.7 |
| >50 | 14.7 | 33.5 | 37.4 | 58.1 | 61.8 |

Social-LSTM was the fastest model but also the least accurate. The proposed model had the longest inference time; in scenes containing more than 50 agents, its inference time exceeded Motion Transformer by 3.7 ms (Figure 4).

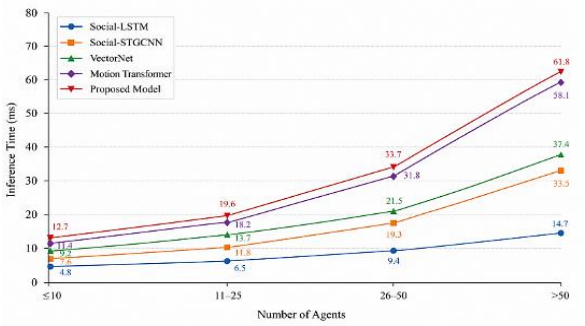



Figure 4 – Inference time as a function of the number of agents

The figure shows how computational cost increases with scene complexity for all evaluated models (Table 12).

Table 12 – Peak GPU memory consumption (GB)

| Model | ≤10 Agents | 11-25 | 26-50 | >50 |
|---|---|---|---|---|
| Social-LSTM | 1.3 | 1.5 | 1.7 | 2.0 |
| Social-STGCNN | 2.1 | 2.8 | 3.6 | 4.8 |
| VectorNet | 2.7 | 3.4 | 4.5 | 5.9 |
| Motion Transformer | 3.4 | 4.7 | 6.3 | 8.6 |
| Proposed Model | 3.8 | 5.2 | 7.1 | 9.4 |

GPU memory consumption increased most markedly for the transformer-based and hybrid architectures. This may limit the deployment of the proposed model in high-density scenes without additional inference optimization or neighbour-selection strategies.

Average metrics do not fully capture rare but potentially critical forecasting failures. Therefore, ADE distributions and representative trajectory predictions were analyzed qualitatively (Figure 5).

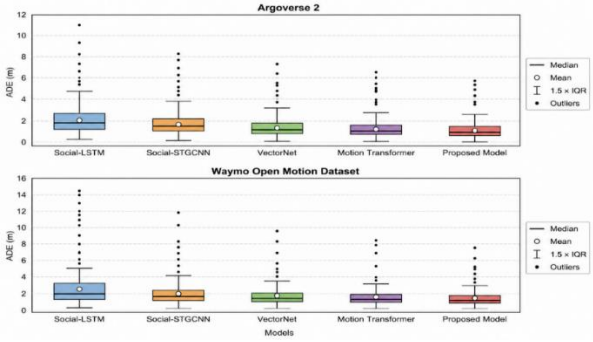



Figure 5 – ADE distributions across models

Boxplots illustrate medians, interquartile ranges, and outliers separately for Argoverse 2 and the Waymo Open Motion Dataset (Figure 6).

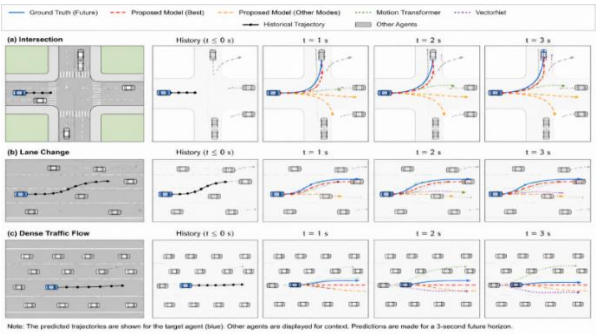



Figure 6 – Examples of predicted trajectories in different traffic scenarios

Three representative cases are shown: (a) intersection crossing, (b) lane-change maneuver, and (c) dense traffic flow.

The largest forecasting errors occurred at complex intersections, during late lane changes, and in scenarios involving overlapping trajectories of multiple agents. In such cases, the observed motion history was often insufficient to determine a unique future maneuver, even when a multimodal decoder was employed.

**Discussion**

The results show that the adaptive graph module provides a modest but consistent improvement over a strong attention-based baseline. This gain appears to be associated with the explicit selection and weighting of local interactions rather than with architectural complexity alone. Although Motion Transformer captures global scene context, it does not explicitly assign interaction weights based on distance, relative velocity, and heading difference. The proposed graph module addresses this by passing a context-weighted local representation to the transformer.

This finding is consistent with prior work on explicit inter-agent modeling [1-3]. Unlike fixed graph structures, the proposed model varies the influence of neighboring agents according to distance, relative velocity, and heading difference. Its advantage therefore becomes more visible in dense traffic, where several agents may affect the target trajectory at the same time.

The results do not imply that one architecture is universally superior. Low Kendall's W values indicate that model rankings changed across scenes, and in some cases Motion Transformer and VectorNet produced comparable predictions. The proposed model should therefore be interpreted as improving robustness in difficult scenarios rather than dominating across all traffic conditions.

The largest errors occurred at complex intersections, during late lane changes, and in dense traffic with partially overlapping trajectories. In these cases, the observed history often does not contain enough information to determine a single future maneuver. Even multimodal prediction cannot fully remove this uncertainty. Computational cost remains a limitation. The proposed model required more memory and longer inference times than the baseline models, especially in dense scenes. Since forecasting is only one part of an autonomous-driving pipeline, practical deployment would require optimization of the graph module, neighbor selection, or inference acceleration. The study was also limited to evaluation on the validation splits of two datasets. Ffuture work should include official benchmark submissions, additional datasets, and richer semantic cues such as intent and traffic-rule context.

**Conclusion**

This study presented a hybrid trajectory forecasting model, combining adaptive graph-based interaction modeling with transformer-based encoding of the global scene context. Experiments on Argoverse 2 Motion Forecasting and the Waymo Open Motion Dataset showed consistent, moderate improvements over Motion Transformer, especially in dense interaction scenarios. The findings suggest that context-dependent weighting of neighboring agents is most useful when multiple road users influence the target trajectory. In simpler scenes, differences between strong architectures remained small.

The proposed model should not be viewed as universally optimal: performance varied across scenes, and accuracy gains increased computational cost. Future research should focus on evaluation through official benchmark submissions, broader dataset coverage, and reductions in inference latency and memory consumption without compromising forecasting accuracy.

**References**

1. Social LSTM: Human trajectory prediction in crowded spaces / A. Alahi et al // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – Las Vegas, USA, 2016. – P. 961-971. https://doi.org/10.1109/CVPR.2016.110.

2. Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction / A. Mohamed et al // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – Seattle, USA, 2020. – P. 14424-14432. https://doi.org/10.1109/CVPR42600.2020.01444.

3. Learning lane graph representations for motion forecasting / M. Liang et al // Proceedings of the European Conference on Computer Vision (ECCV). – Glasgow, United Kingdom, 2020. – P. 541-556. https://doi.org/10.1007/978-3-030-58536-5_32.

4. AgentFormer: Agent-aware transformers for socio-temporal multi-agent forecasting / Y. Yuan et al // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). – Montreal, Canada, 2021. – P. 9813-9823. https://doi.org/10.1109/ICCV48922.2021.00967.
5. Motion Transformer with global intention localization and local movement refinement / S. Shi et al // Advances in Neural Information Processing Systems. – 2022. – Vol. 35. – New York: Curran Associates. – P. 6531-6543. https://doi.org/10.48550/arXiv.2209.13508.
6. nuScenes: A multimodal dataset for autonomous driving / H. Caesar et al // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – Seattle, USA, 2020. – P. 11618-11628. https://doi.org/10.1109/CVPR42600.2020.01164.
7. Zhan W. et al. INTERACTION dataset: An INTERnational, adversarial and cooperative moTION dataset in interactive driving scenarios with semantic maps [The electron. resource]. – 2019. https://arxiv.org/abs/1910.03088 (date of request: 17.05.2025). https://doi.org/10.48550/arXiv. 1910.03088.
8. Argoverse 2: Next generation datasets for self-driving perception and forecasting / B. Wilson et al // Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks. – 2023. https://arxiv.org/abs/2301.00493 (date of request: 17.05.2025). https://doi.org/10.48550/ arXiv.2301.00493.
9. Large scale interactive motion forecasting for autonomous driving: The Waymo Open Motion Dataset / S. Ettinger et al // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). – Montreal, Canada, 2021. – P. 9710-9719. https://doi.org/10.1109/ ICCV48922.2021.00957.
10. VectorNet: Encoding HD maps and agent dynamics from vectorized representation / J. Gao et al // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – Seattle, USA, 2020. – P. 11525-11533. https://doi.org/10.1109/CVPR42600.2020.01154.

**И. Байназарова[1], Ж. Умарова[*], Ю. Джусох[2], Ж. Изтаев[1], С. Ботаева[3]**
[1]Южно-Казахстанский исследовательский университет,
Республика Казахстан, г. Шымкент, проспект Тауке Хана, 5
[2]Университет Путра, Малайзия, Серданг, Селангор
[3]Университет Ташенова, Республика Казахстан, г. Шымкент, проспект Конаева, 16
[*]e-mail: Zhanat-u@mail.ru

## КОНТЕКСТНО-ЗАВИСИМОЕ ПРОГНОЗИРОВАНИЕ ТРАЕКТОРИЙ УЧАСТНИКОВ ДОРОЖНОГО ДВИЖЕНИЯ ДЛЯ СИСТЕМ АВТОНОМНОГО ТРАНСПОРТА НА ОСНОВЕ ГИБРИДНЫХ ГРАФОВЫХ И ТРАНСФОРМЕРНЫХ НЕЙРОННЫХ СЕТЕЙ

*Прогнозирование траекторий участников дорожного движения является одним из ключевых компонентов систем автономного транспорта, поскольку обеспечивает оценку потенциальных конфликтов и поддержку безопасного планирования движения в динамической дорожной среде. В работе предложена гибридная архитектура прогнозирования, объединяющая адаптивное графовое моделирование локальных взаимодействий между участниками движения и трансформерный модуль для кодирования глобального контекста сцены. Оценка модели проводилась на наборах данных Argoverse 2 Motion Forecasting и Waymo Open Motion Dataset с использованием метрик ADE, FDE, Miss Rate, minADE и minFDE, а также анализа устойчивости, плотности взаимодействий и времени инференса. На наборе Argoverse 2 предложенная модель снизила значение ADE с 1,21 до 1,17, FDE – с 2,81 до 2,69, а Miss Rate – с 0,25 до 0,23 по сравнению с Motion Transformer. На наборе Waymo улучшение оказалось менее выраженным, однако сохранялось по всем основным показателям: ADE уменьшилась с 1,58 до 1,53, а FDE – с 3,82 до 3,69. Наиболее заметное преимущество наблюдалось в сценах с высокой плотностью взаимодействий, где значение ADE снизилось с 1,52 до 1,45. Полученные результаты показывают, что адаптивное взвешивание взаимодействий повышает устойчивость прогнозирования в сложных многоагентных дорожных сценариях, хотя сопровождается увеличением вычислительных затрат.*

***Ключевые слова:*** *автономный транспорт; прогнозирование траекторий; глубокое обучение; графовые нейронные сети; трансформеры; многоагентные взаимодействия; компьютерное зрение; прогнозирование движения.*

**И. Байназарова[1], Ж. Умарова[1*], Ю. Джусох[2], Ж. Изтаев[1], С. Ботаева[3]**
[1]М. Әуезов атындағы Оңтүстік Қазақстан университеті,
Қазақстан Республикасы, Шымкент қаласы, Тәуке хан даңғылы, 5
[2]Путра Малайзия университеті, Малайзия, Серданг, Селангор
[3]Тәшенев университеті, Қазақстан Республикасы, Шымкент қаласы, Қонаев даңғылы, 16

# ГИБРИДТІ ГРАФТЫҚ ЖӘНЕ ТРАНСФОРМЕРЛІК НЕЙРОНДЫҚ ЖЕЛІЛЕР НЕГІЗІНДЕ АВТОНОМДЫ КӨЛІК ЖҮЙЕЛЕРІ ҮШІН ЖОЛ ҚОЗҒАЛЫСЫНА ҚАТЫСУШЫЛАР ТРАЕКТОРИЯЛАРЫН КОНТЕКСТКЕ ТӘУЕЛДІ БОЛЖАУ

Жол қозғалысына қатысушылардың траекторияларын болжау автономды көлік жүйелерінің маңызды құрамдас бөліктерінің бірі болып табылады, өйткені ол ықтимал қақтығыстарды бағалауға және динамикалық жол жағдайында қауіпсіз қозғалысты жоспарлауға мүмкіндік береді. Бұл жұмыста жол қозғалысына қатысушылар арасындағы жергілікті өзара әрекеттесулерді адаптивті графтық модельдеуді және көріністің жаһандық контекстін кодтауға арналған трансформер модулін біріктіретін гибридті болжау архитектурасы ұсынылған. Модель Argoverse 2 Motion Forecasting және Waymo Open Motion Dataset деректер жиынтықтарында ADE, FDE, Miss Rate, minADE және minFDE метрикалары, сондай-ақ орнықтылықты, өзара әрекеттесу тығыздығын және инференс уақытын талдау негізінде бағаланды. Argoverse 2 жиынтығында ұсынылған модель Motion Transformer моделіне қарағанда ADE көрсеткішін 1,21-ден 1,17-ге, FDE көрсеткішін 2,81-ден 2,69-ға және Miss Rate көрсеткішін 0,25-тен 0,23-ке дейін төмендетті. Waymo деректер жиынтығында жақсару деңгейі төменірек болғанымен, барлық негізгі көрсеткіштер бойынша сақталды: ADE 1,58-ден 1,53-ке, ал FDE 3,82-ден 3,69-ға дейін азайды. Ең айқын артықшылық өзара әрекеттесу тығыздығы жоғары сценарийлерде байқалды, мұнда ADE мәні 1,52-ден 1,45-ке дейін төмендеді. Алынған нәтижелер өзара әрекеттесулерді адаптивті түрде салмақтау күрделі көпагентті жол сценарийлерінде болжаудың орнықтылығын арттыратынын көрсетті, алайда бұл есептеу шығындарының өсуімен қатар жүреді.

**Түйін сөздер:** автономды көлік, траекторияларды болжау, терең оқыту, графтық нейрондық желілер, трансформерлер, көпагентті өзара әрекеттесу, компьютерлік көру, қозғалысты болжау.

**Information about authors**

**Indira Bainazarova** – PhD student, Department of Automation and Control, M. Auezov South Kazakhstan University, Shymkent, Kazakhstan; e-mail: B.k.indira@mail.ru. ORCID: https://orcid.org/0009-0009-1553-7090.

**Zhanat Umarova*** – PhD, Associate Professor, Department of Information Systems and Modeling, M. Auezov South Kazakhstan University, Shymkent, Kazakhstan; e-mail: Zhanat-u@mail.ru. ORCID: https://orcid.org/0000-0002-0257-4417.

**Yusmadi Jusoh** – Associate Professor, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia; e-mail: yusmadi@upm.edu.my. ORCID: https://orcid.org/0000-0002-7767-5001.

**Zhalgasbek Iztayev** – Candidate of pedagogucal sceinces, Associate Professor, Department of Information Systems and Modeling, M. Auezov South Kazakhstan University, Shymkent, Kazakhstan; e-mail: zhalgasbek71@mail.ru. ORCID: https://orcid.org/0000-0002-3210-2963.

**Saule Botayeva** – Candidate of technical sceinces, Department of IT and Digital Technologies, Tashenev University, Shymkent, Kazakhstan; e-mail: saule_bb@mail.ru. ORCID: https://orcid.org/0000-0002-4995-7730.

**Сведения об авторах**

**Индира Байназарова** – докторант, кафедра Автоматизации и управления, Южно-Казахстанский университет имени М. Ауэзова, Шымкент, Казахстан; e-mail: B.k.indira@mail.ru. ORCID: https://orcid.org/0009-0009-1553-7090.

**Жанат Умарова*** – PhD, ассоциированный профессор, кафедра Информационных систем и моделирования, Южно-Казахстанский университет имени М. Ауэзова, Шымкент, Казахстан; e-mail: Zhanat-u@mail.ru. ORCID: https://orcid.org/0000-0002-0257-4417.

**Юсмади Джусох** – ассоциированный профессор, факультет Компьютерных наук и информационных технологий, Университет Путра Малайзия, Малайзия; e-mail: yusmadi@upm.edu.my. ORCID: https://orcid.org/0000-0002-7767-5001.

**Жалгасбек Изтаев** – кандидат педагогических наук, ассоциированный профессор, кафедра Информационных систем и моделирования, Южно-Казахстанский университет имени М. Ауэзова, Шымкент, Казахстан; e-mail: zhalgasbek71@mail.ru. ORCID: https://orcid.org/0000-0002-3210-2963.

**Сауле Ботаева** – кандидат технических наук, кафедра IT и цифровых технологий, Университет Ташенова, Шымкент, Казахстан; e-mail: saule_bb@mail.ru. ORCID: https://orcid.org/0000-0002-4995-7730.

**Авторлар туралы мәліметтер**

**Индира Байназарова** – PhD докторант, Автоматтандыру және басқару кафедрасы, М. Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент, Қазақстан; e-mail: B.k.indira@mail.ru. ORCID: https://orcid.org/0009-0009-1553-7090.

**Жанат Умарова*** – PhD, қауымдастырылған профессор, Ақпараттық жүйелер және модельдеу кафедрасы, М. Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент, Қазақстан; e-mail: Zhanat-u@mail.ru. ORCID: https://orcid.org/0000-0002-0257-4417.

**Юсмади Джусох** – қауымдастырылған профессор, Компьютерлік ғылымдар және ақпараттық технологиялар факультеті, Путра Малайзия университеті, Малайзия; e-mail: yusmadi@upm.edu.my. ORCID: https://orcid.org/0000-0002-7767-5001.

**Жалғасбек Изтаев** – педагогика ғылымдарының кандидаты, қауымдастырылған профессор, Ақпараттық жүйелер және модельдеу кафедрасы, М. Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент, Қазақстан; e-mail: zhalgasbek71@mail.ru. ORCID: https://orcid.org/0000-0002-3210-2963.

**Сәуле Ботаева** – техника ғылымдарының кандидаты, IT және цифрлық технологиялар кафедрасы, Тәшенев университеті, Шымкент, Қазақстан; e-mail: saule_bb@mail.ru. ORCID: https://orcid.org/0000-0002-4995-7730.

*Received 01.06.2026*
*Revised 18.08.2026*
*Accepted 19.08.2026*

https://doi.org/10.53360/2788-7995-2026-3(23)-22

МРНТИ: 28.23.15

**Г.К. Абишева, Б.Ш. Разахова, Т.А. Айдынов*, М.К. Сапарбек**
Евразийский национальный университет имени Л. Н.Гумилева,
10000, Республика Казахстан, г. Астана, ул. Сатпаева, 2
*e-mail: tolegen.ch@gmail.com

# ИССЛЕДОВАНИЕ ПУБЛИЧНЫХ ДАТАСЕТОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ЖЕЛЕЗНОДОРОЖНЫХ ПРИЛОЖЕНИЙ

***Аннотация:*** *Развитие технологий искусственного интеллекта (ИИ) в железнодорожной отрасли напрямую зависит от наличия качественных и репрезентативных публичных датасетов. В статье представлен систематический анализ открытых наборов данных, применяемых для задач обнаружения дефектов рельсов, диагностики контактной сети, мониторинга подвижного состава и обеспечения безопасности железнодорожной инфраструктуры. Анализ выполнен в соответствии с методологией PRISMA: поиск проводился в базах IEEE Xplore, Scopus и Web of Science, а также в репозиториях Kaggle, GitHub и Zenodo. Проанализировано 30 ключевых публикаций и 13 открытых датасетов, которые классифицированы по типу сенсоров (RGB, LiDAR, вибрационные, мультисенсорные), задачам машинного обучения (детекция, сегментация, классификация), объёму данных и формату аннотаций. Установлено, что большинство наборов данных основаны на RGB-изображениях, тогда как доля мультимодальных датасетов постепенно увеличивается. Выявлены ключевые ограничения существующих наборов данных -малый объём выборок, дисбаланс классов, ограниченная географическая вариативность и отсутствие унифицированных benchmark-протоколов, и определены направления дальнейших исследований. Полученные результаты подчёркивают необходимость создания крупных мультимодальных наборов данных и разработки единых стандартов аннотации и оценки моделей ИИ.*

***Ключевые слова.*** *железная дорога; искусственный интеллект; датасет; компьютерное зрение; дефект рельса; сегментация; YOLO; U-Net.*

**Введение.** Железнодорожный транспорт является одной из ключевых составляющих современной транспортной инфраструктуры и играет важную роль в обеспечении устойчивого развития экономики. Повышение интенсивности перевозок и увеличение скорости движения поездов предъявляют высокие требования к надёжности и безопасности железнодорожной инфраструктуры. В связи с этим особое значение приобретает своевременное обнаружение

©Г.К. Абишева, Б.Ш. Разахова, Т.А. Айдынов, М.К. Сапарбек, 2026