

JUNHEE KIM

$$\pi(\theta | X) \propto p(X | \theta)\pi(\theta)$$

Phone

010-8543-0390

Email

kim8881472@gmail.com

Who am I

- **Education**

- 인하대학교 통계학과 & SCSC 연계전공(2019.02 ~ 2025.02)
- 인하대학교 통계데이터사이언스학과, 통계학 전공(2025.02 ~ 2027.02)

- **Experience**

- 학부연구생 - 인하대학교 컴퓨터공학과 HCI 연구실 (2023.08 ~ 2024.07)
- 학부연구생 - 인하대학교 통계학과 베이지안 추론 연구실 (2024.07 ~ 2025.02)
- 석사 연구원 - 인하대학교 통계학과 베이지안 추론 연구실 (2025.02 ~ Present)

- **Awards**

- 2025 인하 인공지능 챌린지 – 최우수상, 인하대학교 AI 융합 연구세터(2025.07)
- 2025 LG Aimers 온오프라인 해커톤 – 본선 진출(23/794), 최종 7위(7/27)
- 대조학습 기반 주행 패턴 이상 탐지 연구 – 우수 논문상, 2023 한국 ITS 학회 추계학술 대회(2024.08)
- 스마트폰 가속도 센서를 활용한 도로 상태 탐지 시스템 – 동상, 2023년 가을학기 VIP 발표대회(2023.12)

- **Skills**

- **Language:** Python, R, SQL
- **ML/DL Framework:** Pytorch, Scikit-learn, PyMC3, Pyro
- **Etc:** Git, Docker, Notion



- [CV](#)
- [Blog](#)

Who am I

- **데이터 속에서 문제의 본질을 찾아내는 사람**

복잡한 산업 데이터를 다루며, 현상의 표면이 아닌 근본적인 원인과 패턴을 탐구하는 사고력을 길러왔습니다.

- **예측보다 신뢰를 중시하는 연구자**

높은 정확도보다 예측의 신뢰성과 해석 가능성을 더 중요한 가치로 여기며, 모델이 만들어내는 결과의 의미를 통계적으로 증명하는 과정을 중시합니다.

- **끈기와 몰입으로 한계를 넘는 실행자**

새로운 기술을 짧은 시간안에 흡수하고 완성도 있는 결과로 연결하며, 끝까지 파고드는 과정 속에서 배움의 즐거움과 성장을 경험합니다.

- **스스로 문제를 정의하고 해답을 찾아가는 탐구자**

주어진 문제에 머물지 않고, 가설을 세우고 실험으로 검증하는 사고 과정을 즐깁니다. 데이터 경진대회(데이콘, 캐글)에 꾸준히 참가하며, 새로운 문제를 탐구하는 과정을 즐깁니다.



- [CV](#)
- [Blog](#)

PROJECTS

- 산업용 회전 장비 이상 진동 탐지 프로젝트

- 대조 학습을 이용한 주행 패턴 이상 탐지

- 스마트폰 가속도 센서를 이용한 도로 노면 탐지 시스템

- 강화학습을 이용한 조난자 이동 경로 예측

- 난임 환자 시술 데이터 기반 임신 성공률 예측

- GAT 임베딩 기반 베이지안 커널 릿지를 이용한 분자 예측

- 언어 정보 기반 색채화

- 베이지안 딥커널 머신을 이용한 양식 낚치 성장 예측

산업용 회전 장비 이상 진동 탐지 프로젝트

2023-2학기 Edge Computing 기말프로젝트

- 하드웨어 설계, 데이터 분석, 모델링 전 과정을 주도적으로 수행하여 실시간 이상 탐지 시스템을 구현함

72

에지컴퓨팅 특론

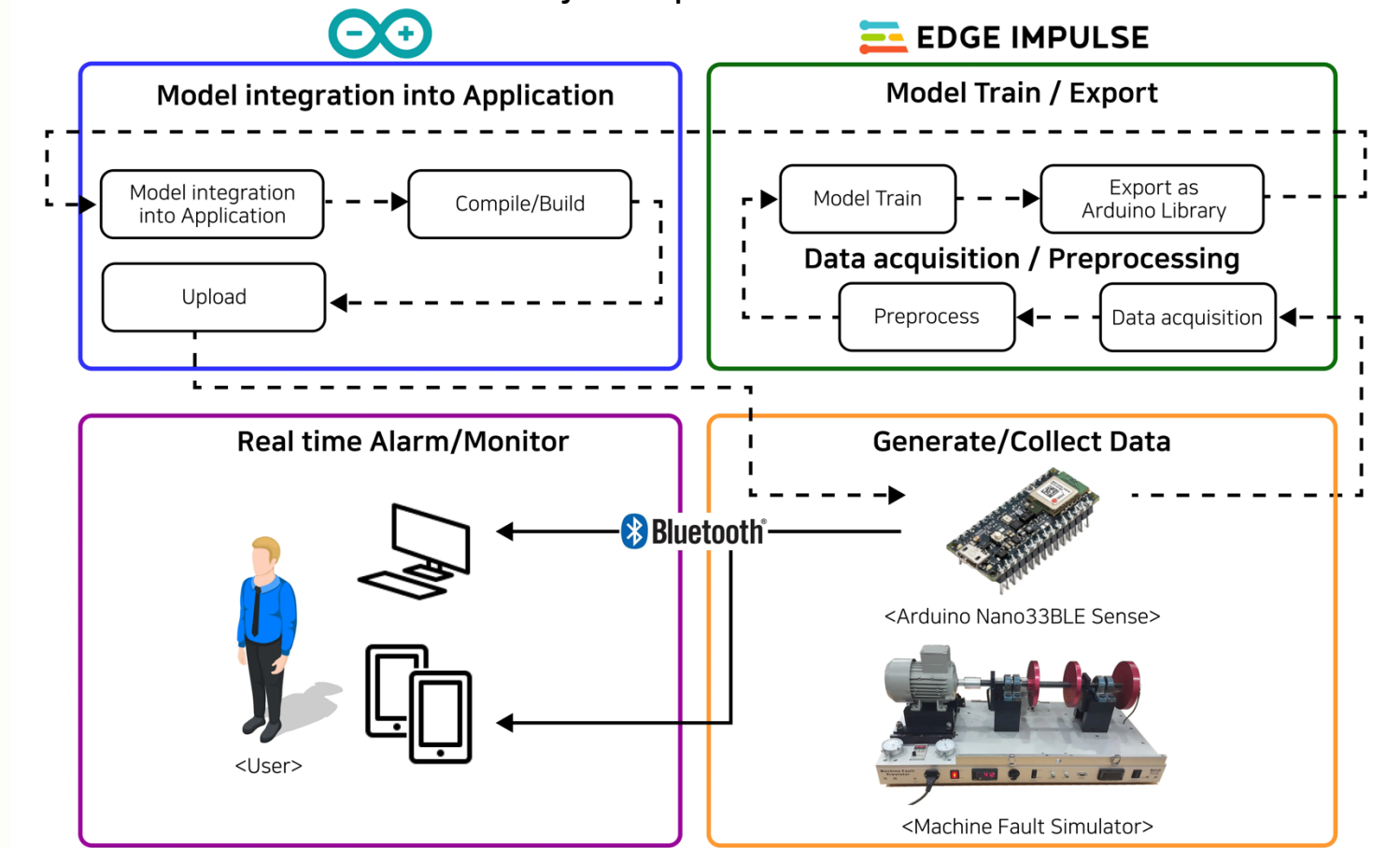
3

A+

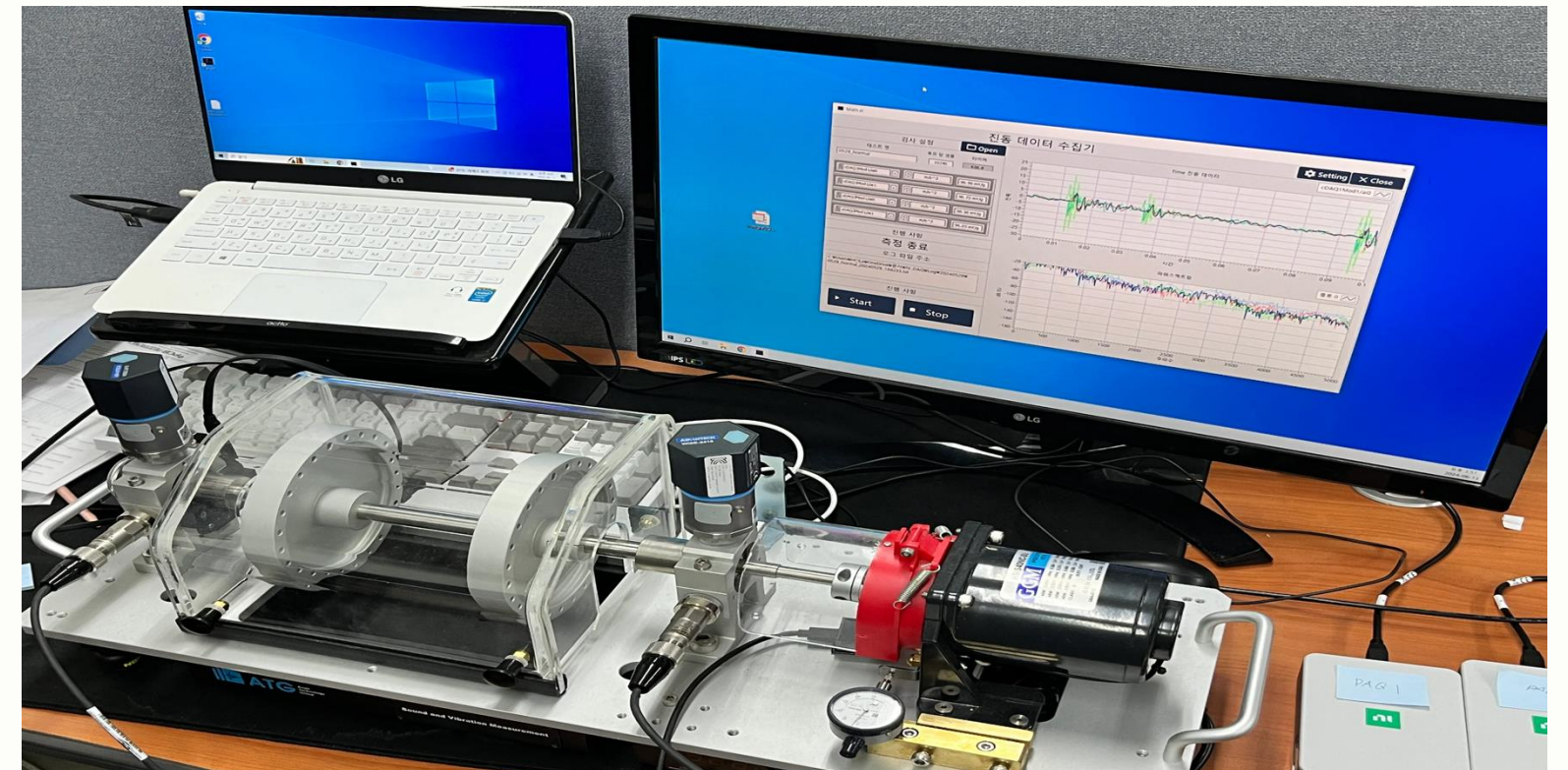
Project

- 회전 설비의 미세 진동 데이터를 엣지 디바이스를 통해 실시간으로 수집·분석하는 시스템 구축
- Normal / Imbalance / Misalignment 세 가지 상태를 분류하는 모델 개발
- 데이터 수집 과정에서 Arduino Nano 33 BLE 를 활용하고, 회전 모사 설비에 부착을 위해 3D 프린터를 사용하여 케이스를 직접 설계 및 제작
- 하드웨어부터 데이터 분석까지 전 과정을 경험하며, 문제 해결을 위한 빠른 기술 습득 능력 강화
- 분류 정확도 96.8% 달성
- 데이터 수집, 하드웨어 설계, 분석 및 모델링까지 수행하며 문제를 끝까지 밀고 가는 실행력과 기술 적응력을 키움
- Github** : <https://github.com/joon0390/Edge-Computing>
- Keywords** : Edge Device / Arduino IDE / Vibration Analysis / 3D Printing

Project Pipeline



회전 모사 설비에서 데이터를 수집하는 모습



대조학습을 이용한 주행 패턴 이상 탐지

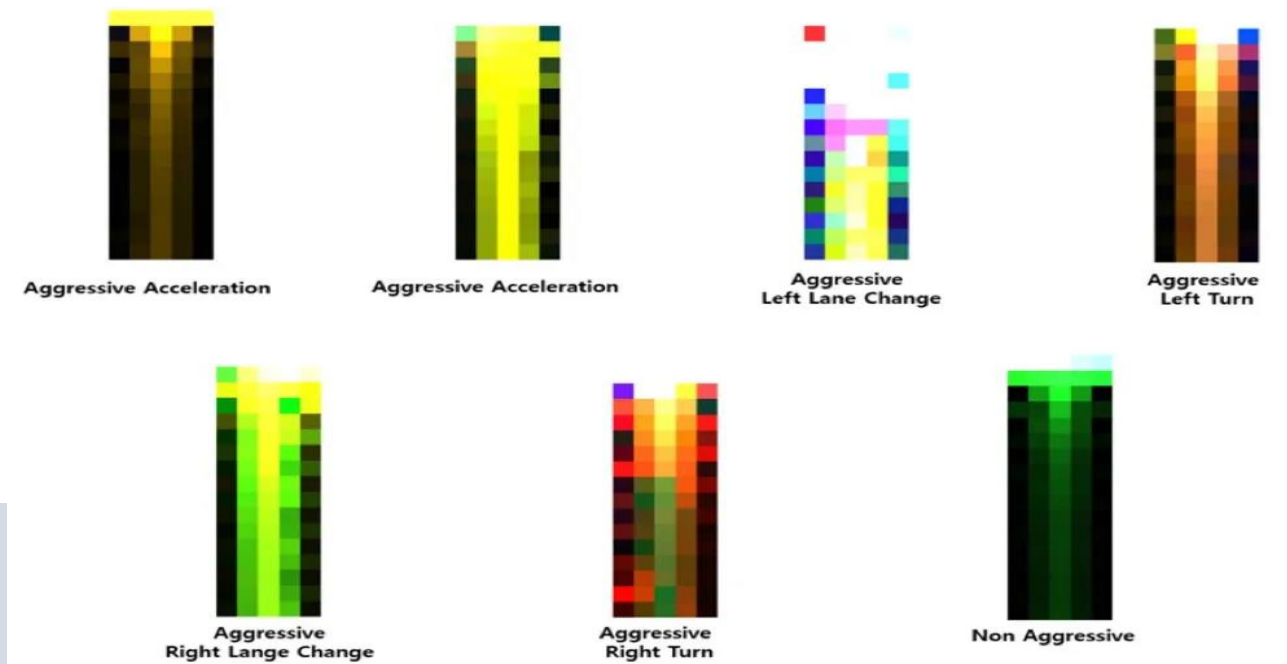
2023 한국 ITS 학회 추계 학술대회

- 레이블링 데이터가 부족한 환경에서도 대조학습을 적용하여 높은 정밀도의 주행패턴 이상 탐지 모델을 구축하고, 체계적인 실험설계를 통해 성능을 검증함

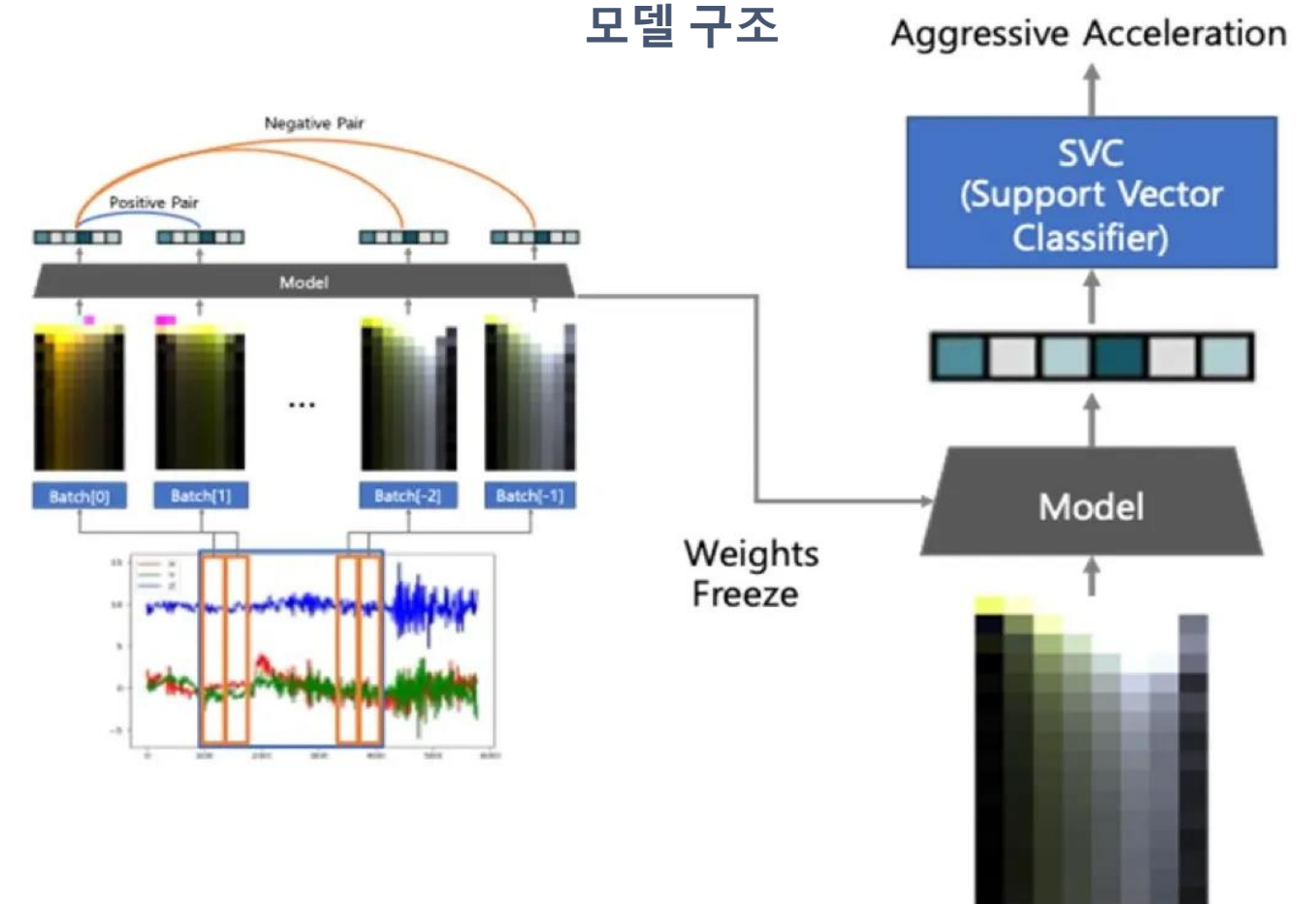
Project

- 라벨이 부족한 운전 패턴 데이터를 활용하여, Contrastive Learning 기반의 이상 탐지 시스템을 설계함
- 시계열 데이터를 STFT 변환 후, 시간 일관성 가정에 따라 인접 구간 간 유사도를 학습하는 프레임워크를 구현함.
 - (Time-Consistency 가설: 시간적으로 가까운 구간은 유사한 상태를 가질 가능성이 높다는 가정)
- 제한된 데이터 환경에서도 정밀도 0.92, F1 Score 1.00을 달성하여, 제안한 접근의 효율성과 강건성을 입증함.
- Head 구조 실험 : Backbone 고정 후, 여러 Head(SqueezeNet, RegNet 등)를 비교해 최적 모델 구성
- 데이터 파이프라인 구축 : 비정규 시계열 센서 데이터의 보간, 정규화, STFT 변환 과정을 자동화한 전처리 파이프라인 설계
- 모델 성능을 체계적으로 벤치마킹하고, 재현 가능한 실험 설계 역량을 키움
- Awards : 한국 ITS 학회 추계 학술대회 우수논문상
- Keywords : STFT / Contrastive Learning / Time Series Classification

STFT 변환



모델 구조



스마트폰 가속도 센서를 이용한 도로 노면 탐지 시스템

2023-2학기 알파프로젝트

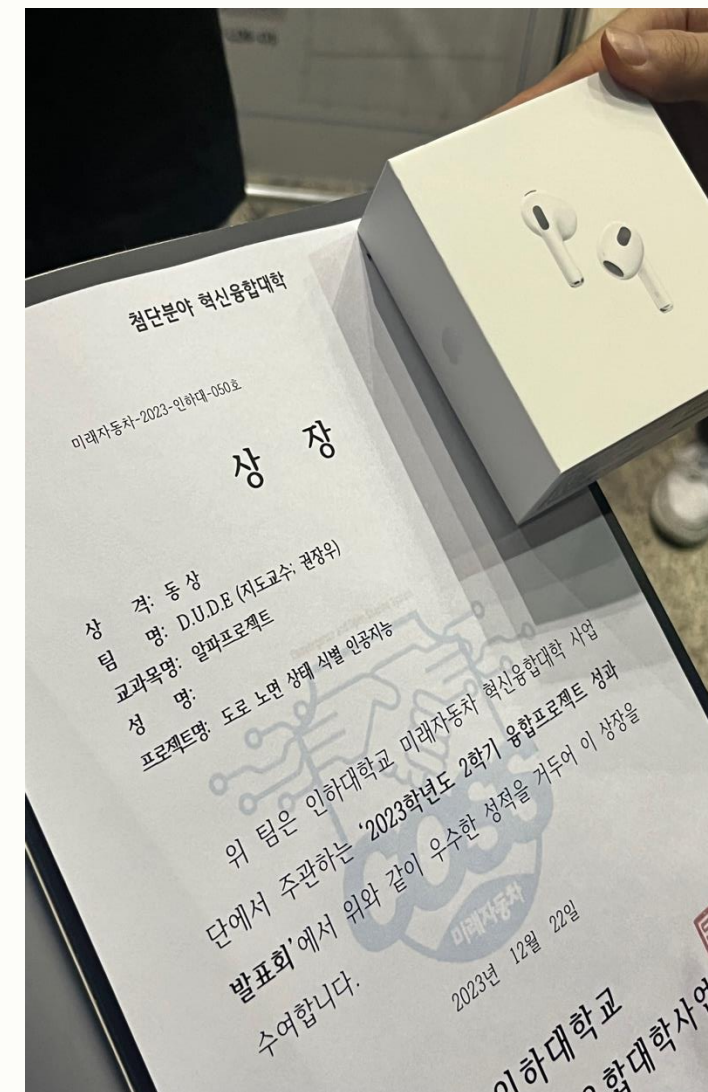
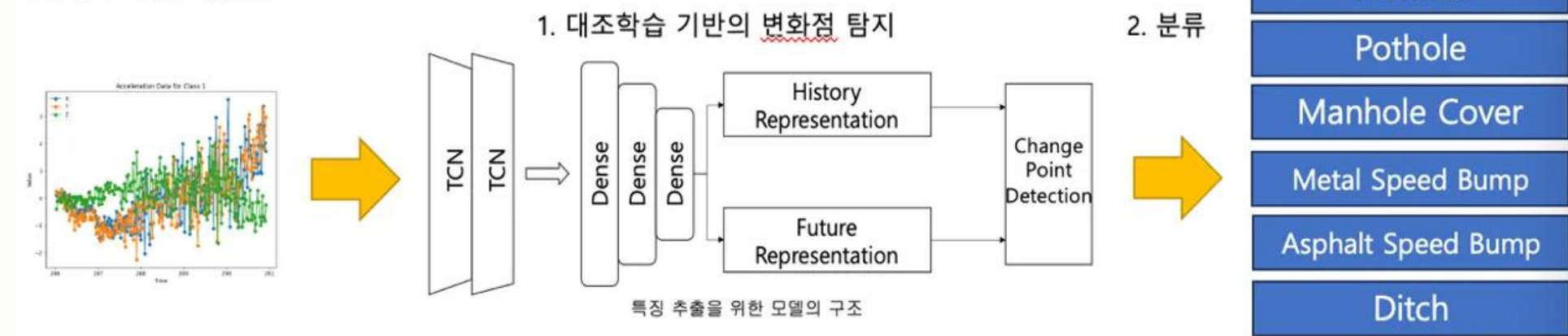
- 레이블이 부족한 환경에서 자기지도학습과 지도학습을 결합하여, 스마트폰 가속도 센서 기반 도로 노면 이상 탐지 시스템을 구현함

Project

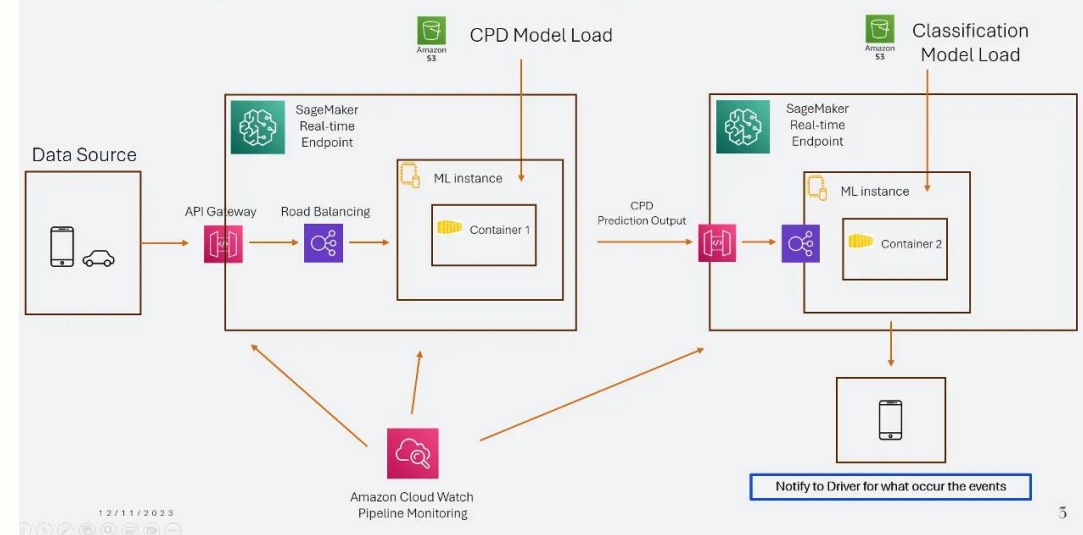
알파프로젝트 1 1.0 A+ 일선

- 스마트폰의 3축 가속도 데이터를 활용하여 요철 / 포트홀 등 도로 노면의 이상 상태를 실시간으로 탐지·분류하는 시스템을 개발함
- 1단계에서 대조학습 기반 변화점 탐지 모델을 통해 충격 신호의 이상 구간을 자동 검출하고,
- 2단계에서 LSTM-CNN 하이브리드 모델을 사용해 이상 상태의 종류를 분류함
- 데이터 수집부터 AWS SageMaker를 이용한 클라우드 학습·추론 파이프라인을 구축함
- ”주행 패턴 이상 탐지”의 연구를 응용하며, 데이터 중심의 연구가 현실의 문제 해결로 이어지는 경험을 통해 연구와 응용의 연결을 경험함
- Awards** : 2023-2학기 Vertically Integrated Projects 발표대회 동상
- Keywords** : Contrastive Learning / Edge & Cloud Integration / LSTM-CNN

전체 모델 개요



Real-time prediction model Serving



<표 2> 도로 노후 현황

(단위 : km)

구분	전체							
		개통	미개통	고속국도	일반국도	특별광역시도	지방도	시군구도
2022년	114,314	105,563	8,752	4,939	14,200	5,264	18,316	71,596
1992년	58,847	58,377	470	1,600	12,079	13,082	10,689	21,398
30년 이상 비중	51.5%	55.3%	5.4%	32.4%	85.1%	-	58.4%	-

주 : 특별광역시도의 도로는 구 소관으로 이관에 따라 변동성이 커 특별광역시도와 시군구도의 비중은 산정하지 않음.
자료 : 국토교통부, 2021년 도로현황조사.

강화학습을 이용한 조난자 이동 경로 예측

2024 The 12th IASC-ARS Conference, Taipei, Taiwan – Poster Session

- 복잡한 지형 환경에서 조난자의 이동 경로를 예측하기 위해 강화학습 기반의 DQN 모델을 설계하고 웹 애플리케이션으로 구현함

Project

- 실종자 수색 및 구조 작전의 효율을 높이기 위해 강화학습을 이용한 이동 경로 예측 모델을 개발함
- 수치 표고 모델(DEM)과 도로·하천 정보를 포함한 실제 지리 데이터를 활용하여, 에이전트가 지형 특성을 학습하며 최적 경로를 탐색하도록 설계함
- 기존의 Vanilla DQN과 Attention 기반 Pointer Network DQN을 비교하여, 후자가 평균 보상 +10.2, 성공률 94.7%로 우수한 성능을 보임
- Python Streamlit 라이브러리를 사용하여 사용자 인터페이스를 구축하여, 예측된 경로를 지도 위에서 실시간 시각화함
- 조난자의 복잡한 의사결정 문제를 강화학습으로 모델링하고, 실제 수색 환경에 적용가능한 수준의 모델로 발전시킴
- 단순한 성능 지표를 넘어 문제의 본질과 목적을 알고리즘에 반영하기 위한 노력을 통해 수학적 사고력을 기름
- Keywords** : Reinforcement Learning / DQN / Geospatial Data / Streamlit



인하대학교

Predicting Movement Paths in Search and Rescue Operations Using RL

Junhee Kim, Heedam Kwon, Seongil Jo, Jaehoh Kim
Inha University



INHA UNIVERSITY

Abstract

In search and rescue operations, achieving operational success is challenging even when substantial resources are deployed. This study develops a predictive model for missing person movement paths using reinforcement learning (RL) techniques based on deep neural networks.

The research compares Vanilla Deep Q-Network (DQN) and Pointer Network DQN, demonstrating that the Pointer Network DQN achieves more accurate and consistent path predictions in complex terrain conditions.

Data

1. Geospatial Data:

- Digital Elevation Model (DEM):** Represents terrain height.
 - Shapefiles:** Includes forest roads, rivers, hiking trails, and accident points.
- #### 2. Data Preprocessing:
- Rasterization of geospatial data to grid format.
 - Binaryization to express the presence or absence of a particular terrain
 - Calculate the DEM to extract additional topographic features such as slope, roughness, etc

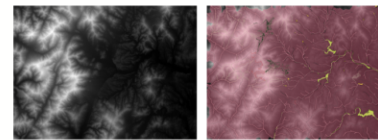


Figure 1: Geospatial Data Visualization: DEM and Shapefiles Example

Model Architectures

1. Vanilla DQN:

- Simple feedforward neural network with two hidden layers.
- Outputs Q-values for eight possible movement directions.

2. Pointer Network DQN:

- Encoder-decoder LSTM architecture with attention mechanism.
- Designed to handle complex terrain features and predict optimal paths.



Figure 2: Vanilla DQN & Pointer DQN Architecture

Model Inputs:

- The state is represented as an 8-dimensional feature vector:
 - `self.dem[y, x]`: Elevation at the current position.
 - `self.slope[y, x]`: Slope at the current position.
 - `np.mean(dem_patch)`: Average elevation in the surrounding patch.
 - `np.mean(slope_patch)`: Average slope in the surrounding patch.
 - `np.max(road_patch)`: Maximum road presence indicator in the patch.
 - `np.max(forestroad_patch)`: Maximum forest road presence indicator in the patch.
 - `np.max(climb_patch)`: Maximum climbing path presence indicator in the patch.
 - `np.std(dem_patch)`: Standard deviation of elevation values in the patch.

Actions(Model Output):

- 8 probabilities for 8 actions in a particular state are returned.

Reward Function Design

Designing the Reward Function for Optimal Pathfinding:

- Rewards are crafted to incentivize optimal behaviors and penalize inefficiencies.
- Each component of the reward function is calibrated for effective learning in complex terrains.

Reward/Penalty	Description
Path following reward	Base reward (+0.5) for staying on the predefined path.
Water proximity reward	Additional reward (+0.15) for being near water without entering it.
Directional consistency	Reward (+0.1) for maintaining the same direction as previous movement.
Exploration reward	Encourages new locations by adding a small reward (+0.02) for not revisiting previous positions.
Diversity penalty	Penalty (-0.1) for repetitive back-and-forth movements within a small area.
Stationary penalty	Penalty (-0.2) for not moving from the current position.
Revisit penalty	Penalty proportional to revisit count (-0.05 × number of visits) for the same position.
Slope penalty	Penalty (-0.05) for traversing areas with a slope greater than 30 degrees.
Water penalty	Large penalty (-0.5) for entering water regions.
Boundary reward	Bonus reward (+0.03) for staying within the predefined search area.
Proximity reward	Reward (+0.3 / distance) for moving closer to the predefined path during exploration mode.

Training Process

1. Data Preparation

- Processing DEM and shapefile data through rasterization and binarization
- Converting environmental data into 8D state vectors containing elevation, slope, trail presence, water proximity, and terrain features
- Normalizing all features to [0, 1] range for consistent network input

2. Training Setup

- Experience buffer size: 10,000 samples
- Epsilon-greedy exploration: 0.99 → 0.1 with scheduler
- Network updates: Adam optimizer (lr = 0.001)
- Target network synchronization: Every 1000 steps
- Maximum steps per episode: 1000
- Discount factor (γ): 0.99 for future reward consideration

3. Model Checkpointing

- Best model preservation based on reward metrics
- Regular checkpoints every 10 episodes
- Early stopping when performance plateaus
- Training statistics logged for performance analysis

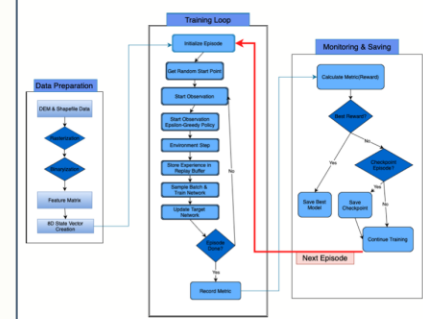


Figure 6: Training process workflow showing data preparation, training loop, and monitoring stages

Results

Performance Metrics

- Vanilla DQN:
 - Average Reward: +8.8 ± 1.2
 - Success Rate(Find Road Rate): 81.3%
- Pointer Network DQN:
 - Average Reward: +10.2 ± 0.8
 - Success Rate(Find Road Rate): 94.7%

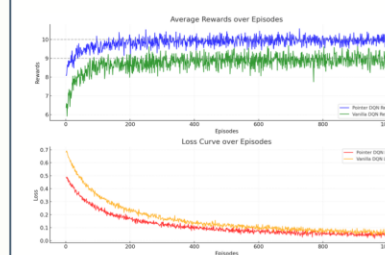


Figure 3: Average Reward & Loss Comparison

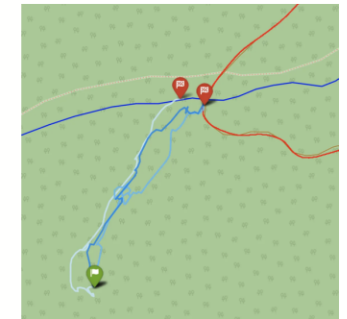


Figure 4: Simulated Path of Agent

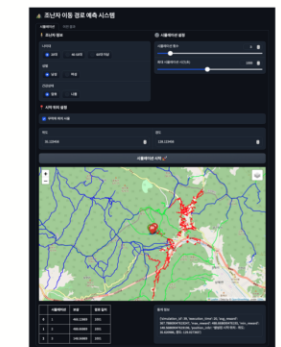


Figure 5: Gradio Interface for Displaying Simulation Path

Conclusion & Future Work

The Pointer Network DQN demonstrates significant improvements compared to the Vanilla DQN in terms of reward and loss, achieving much faster convergence. Additionally, it exhibited the ability to understand and adapt to complex and diverse terrain relationships through the Attention Mechanism. Although the model considered various types of terrain, it was limited by the sparsity of geographic data such as water-related features and trail information. For future research, alternative reinforcement learning methodologies such as Proximal Policy Optimization (PPO) or Soft Actor-Critic (SAC) could be explored. These methods offer potential advantages in terms of stability and sample efficiency. PPO's conservative policy updates could help maintain consistent performance across different terrain types, while SAC's entropy maximization approach might lead to more robust exploration strategies in complex environments.

난임 환자 시술 데이터 기반 임신 성공률 예측

LG Aimers 6th 온·오프라인 해커톤

• 온라인 해커톤

- 온라인 예선은 난임 환자의 시술 데이터를 기반으로 임신 성공 “여부”를 분류하는 과제
- 의료적 변수 간의 상관관계를 분석하고 배아이식 비율, 수정 성공률 등 추가 파생 변수를 생성하여 예측 정확도를 향상시킴

• 오프라인 해커톤

- 본선은 과제의 형태가 이진 분류에서 확률 예측(Weighted Brier Score + F1 Score)로 변경됨
- 회귀로 전환하지 않고, 분류기의 확률 출력을 직접 학습 대상으로 삼음
- 이를 위해 임신 성공 케이스에 대한 가중치를 부여하는 가중치 함수를 설계하여 예측 확률 정밀도와 클래스 불균형을 동시에 개선함
- 또한 Stacking 앙상블(Catboost → LightGBM 잔차보정)을 적용하여 확률 보정을 강화함
- 결과적으로 확률 기반 평가지표 모두에서 향상된 성능을 기록하고 오프라인 본선 7위를 달성함
- 이 경험을 통해, 단순히 모델을 바꾸는 것보다 문제 정의와 평가 지표에 맞는 최적화 전략을 세우는 것의 중요성을 배움
- Github : <https://github.com/joon0390/joon0390/blob/main/files/LGaimers.pdf>

난임 환자 대상 임신 성공 여부 예측 AI 온라인 해커톤

LG Aimers | 채용 | 알고리즘 | 의료 | 바이오 | 정형 | 분류 | ROC-AUC

₩ 상금 : 본선 진출

🕒 2025.02.01 ~ 2025.02.27 09:59 [+ Google Calendar](#)

👤 1,568명 📅 마감

난임 환자 대상 임신 성공 확률 예측 AI 오프라인 해커톤

LG Aimers | 채용 | 알고리즘 | 의료 | 바이오 | 정형 | 회귀 | Brier Score

₩ 상금 : 1,000만 원

🕒 2025.04.05 ~ 2025.04.06 10:59 [+ Google Calendar](#)

👤 101명 📅 마감

$$Score = 0.5 \times \left(1 - \frac{\sum w_i (y_i - \hat{y}_i)^2}{\sum w_i} \right) + 0.5 \times \frac{2TP}{2TP + FP + FN}$$

$$w_i = 1 + 4y_i + (0.5 - y_i)^2$$

$$\hat{y}_i^{(bin)} = \begin{cases} 1, & \text{if } \hat{y}_i > 0.5 \\ 0, & \text{otherwise} \end{cases}$$

GAT 임베딩 기반 베이지안 커널 릿지를 활용한 분자 예측

2025 BayesComp, Singapore – Poster Session

- 그래프 어텐션 네트워크로 분자 구조를 임베딩하고, 베이지안 커널 릿지를 결합하여 예측의 신뢰도를 정량화한 하이브리드 모델을 제안함

Project

- 분자 구조 정보를 그래프로 표현하고, GAT를 이용하여 각 분자를 고차원 임베딩 벡터로 인코딩함
- 생성된 임베딩을 입력으로 베이지안 커널 릿지 회귀(BKRR)를 적용하여, 분자의 물성을 예측함과 동시에 예측 불확실성을 정량화함
- 소규모 데이터셋 환경에서 딥러닝 모델의 오버피팅을 방지하기 위해 커널 트릭 기반의 비선형 학습 구조를 채택함
- GAT의 표현력과 베이지안 회귀의 통계적 안정성을 결합하여 정확성과 신뢰성을 모두 고려한 예측 모델을 구축
- 딥러닝과 통계 모델의 장점을 결합하며 모델의 신뢰성과 해석력의 중요성을 다시금 인식하게 됨
- 이를 통해 단순히 높은 정확도보다, 신뢰할 수 있는 예측을 만드는 AI의 방향성을 고민하게 되어 Conformal Prediction 등 불확실성 정량화 기법에 대한 관심을 갖고 연구를 확장함
- Keywords** : Bayesian Kernel Ridge Regression / Graph Attention Network / Molecular Property Prediction



인하대학교

Bayesian Kernel Ridge on GAT Embeddings for Molecular Prediction

Junhee Kim, Seongil Jo
INHA University



INHA UNIVERSITY

Abstract

Accurate prediction of aqueous solubility is crucial for early-stage drug discovery and environmental chemistry. We present a systematic evaluation of a predictive framework integrating Graph Attention Networks (GATs) with various Bayesian regression models to deliver both point estimates and principled uncertainty quantification. A GAT encoder constructs rich molecular embeddings from graph topology. These embeddings, along with traditional RD-Kit descriptors, are used to train Ridge, Bayesian Ridge (BR), and Bayesian Kernel Ridge (BKR) models. By comparing these combinations, we analyze the independent and synergistic contributions of learned graph features and probabilistic regression techniques, aiming to identify the most robust and reliable framework for solubility prediction.

Keywords: Solubility Prediction, Graph Attention Networks, Bayesian Ridge, Bayesian Kernel Ridge, Uncertainty Quantification

Dataset & Preprocessing

- Dataset:** A benchmark collection of 9,914 molecules with experimentally measured aqueous solubility (logS).
- Feature Sets:** Two distinct feature types were prepared for model comparison: **RDKit Descriptors** (over 70 traditional features for baseline models) and **Molecular Graphs** (graph representations for GNN models).
- Atom Features:** Each atom (node) is featured with an 8D vector including element type (H/C/N/O/F), degree, H-count, and aromaticity.
- Target Transformation:** A Yeo-Johnson power transform is applied to the logS target to reduce distribution skewness, stabilizing variance and aiding model convergence.
- Data Split:** An 80% train, 20% test split was performed, stratified by logS quantiles to ensure a balanced distribution for reliable evaluation.

Dataset Visualization

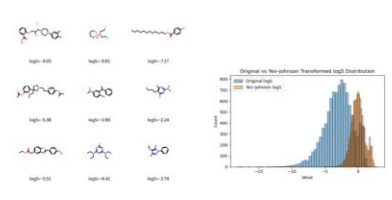


Figure 1: Left: Sample molecular structures are converted into graph representations for the model. Right: The Yeo-Johnson transformation mitigates the heavy skewness of the original logS distribution, resulting in a more symmetric, Gaussian-like shape that is more suitable for model training.

Model Architecture

- Graph Feature Extractor (GAT):** A two-layer Graph Attention Network (GAT) with GraphNorm and ELU activation processes molecular graphs. It uses multi-head attention and concatenates mean, max, and sum pooling to generate a fixed-size graph embedding for each molecule.
- Regression Heads:** The extracted features (either GAT embeddings or RDKit descriptors) are fed into one of several regression models:
 - Ridge:** A standard deterministic linear model.
 - Bayesian Ridge (BR):** A probabilistic linear model that provides uncertainty.
 - Bayesian Kernel Ridge (BKR):** A probabilistic non-linear model using kernels to capture complex relationships and provide uncertainty.

Model Architecture Diagram



Figure 2: The GAT workflow: SMILES strings are converted to molecular graphs. The GAT Encoder processes these graphs to create learned embeddings, which are then used by various regression heads for prediction.

Methodology

- Feature Sets:** Two sets of features were compared: traditional RDKit descriptors and learned embeddings from a Graph Attention Network (GAT).
- GAT Architecture:** The encoder uses two GATConv layers with multi-head attention. Node features h_i are updated by aggregating neighbor features h_j weighted by attention coefficients α_j :

$$h'_i = \sigma \left(\sum_{j \in \mathcal{N}_i \cup \{i\}} \alpha_j \mathbf{W} h_j \right)$$

- Regression Models:** A **non-linear Bayesian Kernel Ridge (BKR) model** was benchmarked against BR and standard Ridge. All share the Ridge L2-regularization principle, which minimizes the following objective function:

$$\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_2^2$$

- Standardized Evaluation:** All models were trained and evaluated on the Yeo-Johnson transformed logS target to ensure a fair and consistent comparison of performance.

Experiments & Results

This experiment systematically evaluates the impact of feature representation and model choice on solubility prediction. We designed a comprehensive comparison to isolate the effects of GAT embeddings, Bayesian methods, and kernel-based non-linearity. The key results from the five benchmarked models are presented below:

Model	RMSE (YJ)	MAE (YJ)	R2 (YJ)	Coverage (%)
Ridge (RDKit Features)	0.8531	0.6582	0.6981	N/A
Bayesian Ridge (RDKit Features)	0.8015	0.6104	0.7315	94.1%
Bayesian Kernel Ridge (RDKit Feat.)	0.6949	0.5533	0.7923	94.5%
GAT + Bayesian Ridge	0.6521	0.5118	0.8111	95.0%
GAT + Bayesian Kernel Ridge	0.6288	0.4875	0.8289	95.2%

- The combination of **GAT embeddings** and a **Bayesian Kernel Ridge (BKR)** head achieves the best performance ($R^2 = 0.8289$), demonstrating the synergistic effect of learned graph features and non-linear probabilistic modeling.
- GAT embeddings consistently outperform RDKit descriptors.** This is evident across all model types, with RMSE dropping from 0.8015 to 0.6521 for Bayesian Ridge, and from 0.6949 to 0.6288 for Bayesian Kernel Ridge. This highlights the superiority of learning features directly from molecular topology.
- Within each feature set, more complex probabilistic models yielded better results (**BKR > BR > Ridge**), confirming the benefits of both Bayesian uncertainty modeling and capturing non-linear relationships.
- All Bayesian models produced reliable 95% credible intervals with empirical coverage rates near the nominal level (94.1% - 95.2%), which validates their uncertainty estimates.

Training Loss Curve

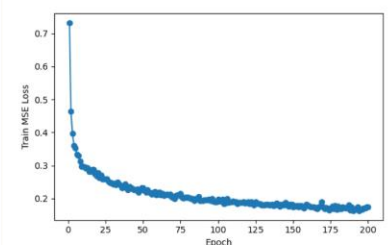


Figure 3: The training curve for the GAT encoder plots MSE loss on the Yeo-Johnson transformed data per epoch. The smooth convergence indicates effective and stable training.

Uncertainty Visualization

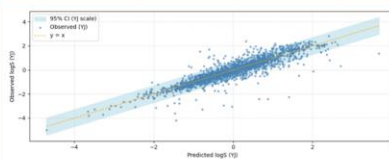


Figure 5: Predicted vs. observed values for the GAT+BKR model. Each point's vertical bar represents its 95% credible interval. Most intervals capture the true value, visually confirming the high coverage rate.

Observed vs Predicted

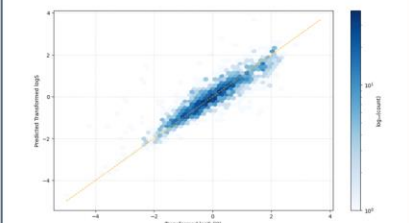


Figure 4: This hexbin plot for our best model (GAT+BKR) shows a high density of predictions along the diagonal (y=x) line, indicating excellent model accuracy and calibration on the transformed scale.

Conclusion

This systematic evaluation of predictive frameworks for aqueous solubility yielded three key findings:

- Superior Features:** GAT-derived embeddings consistently and significantly outperform traditional RDKit descriptors in predictive accuracy.
- Optimal Model:** The combination of GAT embeddings with a non-linear Bayesian Kernel Ridge (BKR) model yielded the best results by effectively capturing complex data relationships.
- Reliable Uncertainty:** The framework's Bayesian models provide well-calibrated uncertainty estimates, which are critical for risk assessment in drug discovery.

Future work will focus on applying this robust framework to predict other critical ADMET properties, such as metabolism and toxicity.

언어 정보 기반 색채화

2025 인하 인공지능 챌린지

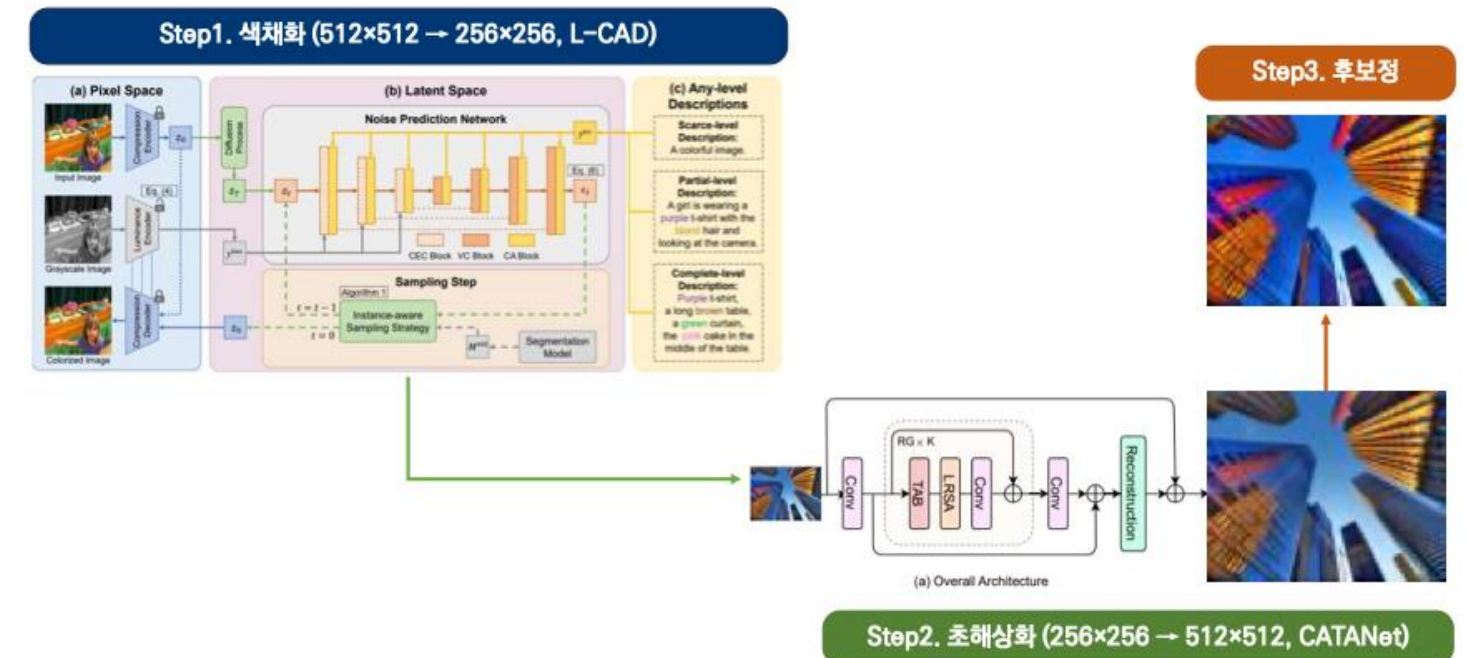
- 교내 인공지능 챌린지에서, 제한된 연산 자원 속에서 사전학습 디퓨전 모델을 최적 활용하고, 초해상화 및 후보정을 결합한 파이프라인으로 고품질 색채화 결과를 만들어냄

Project

- 2025 인하 인공지능 챌린지에서, 텍스트 프롬프트를 기반으로 흑백 이미지를 색채화하는 대회에 참여함
- GPU 자원이 제한이 있는 환경에서 학습 대신, 유사한 학습 데이터 형태로 사전학습된 LCAD 모델의 추론 성능을 극대화하는 전략을 수립함.
- 이미지 품질 향상을 위해 초해상화(CATANet)과 후보정을 결합한 3단계 파이프라인을 설계함
- Awards** : 2025 인하 인공지능 챌린지 대학원생 부문 최우수상(Team Posterior)
- Keywords** : Diffusion Model / Colorization / Super Resolution / Image Processing

II. Method

2.0 전체 파이프라인



III. Conclusion

3.2 최종 파이프라인 정성적 결과



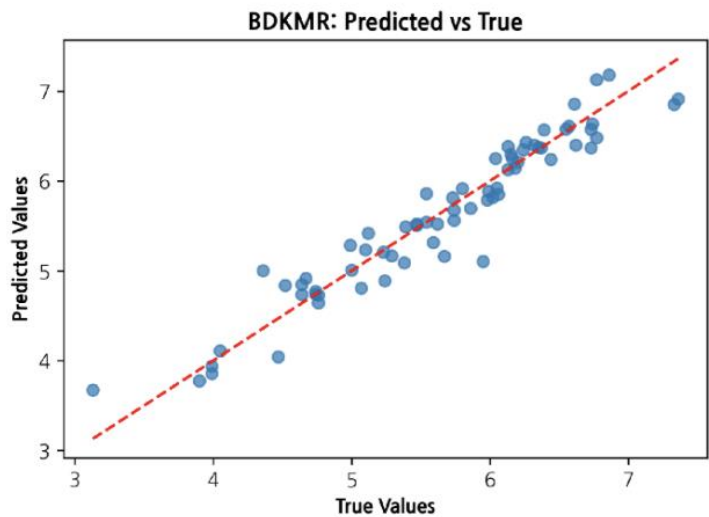
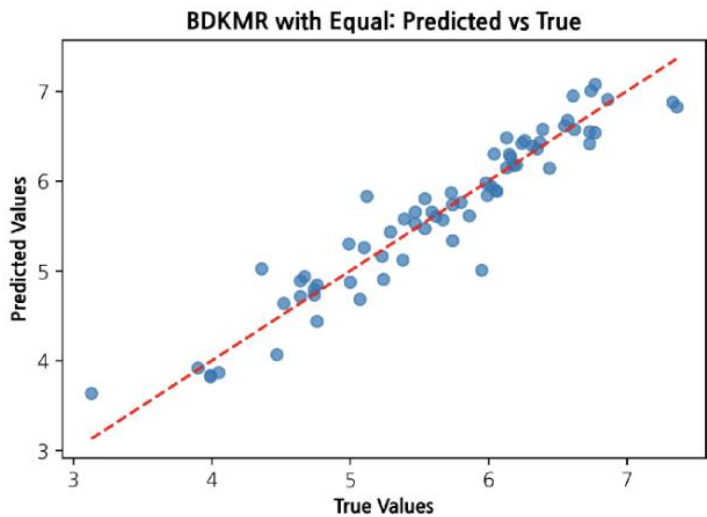
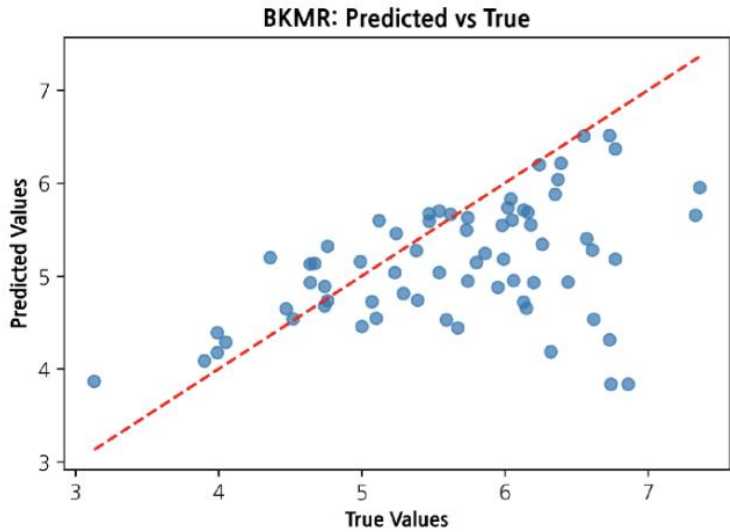
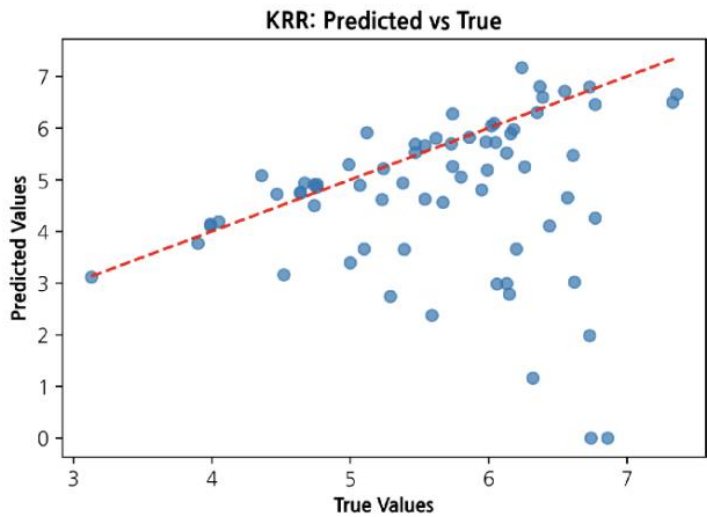
베이지안 딥커널 머신을 이용한 양식 넙치 성장 예측

Paper: Predicting Flatfish Growth in Aquaculture Using Bayesian Deep Kernel Machines

- 딥러닝 비선형 표현 학습과 베이지안 커널 머신의 불확실성 정량화를 결합한 BDKMR 모델을 제안하고, 실제 양식장 데이터를 통해 검증함

Project

- 국내 주요 양식 어종인 넙치(Olive Flounder)의 성장률을 예측하기 위해, 환경 요인(수온, 용존 산소량, 사료량 등)의 복합적 영향을 학습하는 베이지안 딥커널 머신 회귀 모델을 개발함
- 심층 신경망(DNN)을 통해 환경 변수 간의 비선형 상호작용을 표현 벡터(Latent Feature)로 추출하고, 이를 가우시안 프로세스 기반 베이지안 커널 머신에 결합하여 성장률을 예측함
- 기존 KRR 및 BKMR 모델 대비 MAE 0.1895, MSE 0.0629로 성능이 향상되었으며, 예측 불확실성을 함께 산출하여 신뢰 가능한 예측을 제공함
- 실제 양식장 데이터를 이용해 LOOCV 방식으로 평가를 수행하고, 모델의 안정성과 일반화 성능을 입증함
- Publication : Applied Sciences, Vol.15, No.9, Article 9487 (2025)
- Keywords : Bayesian Deep Kernel Machine Regression / Gaussian Process



Model	MAE	MSE
KRR	1.1141	3.5665
BKMR	0.6977	0.9447
BDKMR (Equal)	0.2006	0.0721
BDKMR	0.1895	0.0629

데이터가 쌓여 더 좋은 모델이 되듯이, 끊임없는 배움을 통해 성장하겠습니다.

$$\pi(\theta | X) \propto p(X | \theta)\pi(\theta)$$