

AI Symposium 2025 - Poster Presentations

Applying Behavior Pattern Ontology for Social Media Analysis, Early Warning Systems, and Cross-Study Integration

Student Researcher: Maria Baloch

Advisor: Hande K. McGinty, Computer Science

Abstract

Understanding how behavior patterns emerge and evolve within psychological contexts is a central concern of behavioral science. Establishing a behavior pattern ontology is therefore crucial, as standardized vocabularies enable researchers to work with a consistent and unified framework. In this study, we apply behavior pattern ontology to practical domains, including the annotation of social media posts for building social media analysis dashboards that visualize emerging trends. The ontology also supports the integration of behavioral data across multiple studies and powers intelligent systems that provide early warnings for high-risk behaviors. In particular, ontology-guided features enable machine learning and large language models (LLMs) to improve the classification of risky versus resilient behaviors while making predictions more explainable. The ontology provides a foundation for integrating behavioral insights across domains, supporting reproducible research and fostering cross-disciplinary collaboration.

Keywords

Behavior Pattern, Ontology, Social Media Analysis, Semantic Web

AutoPK: A Hybrid LLM and Similarity Metric Framework for Extracting Pharmacokinetic Data from Complex Tables and Text

Student Researcher: Hossein Sholehrasa

Advisor: Doina Caragea, Majid Jaber-Douraki, Computer Science

Abstract

Pharmacokinetics (PK) plays a pivotal role in drug development and regulatory decision-making across both human and veterinary medicine, directly influencing public health through assessments of drug safety and efficacy. Yet, PK data are frequently embedded in complex, heterogeneous tables with variable structures and inconsistent terminologies, creating major challenges for automated retrieval and standardization. We present AutoPK, a novel two-stage framework for accurate and scalable extraction of PK data from scientific tables. In the first stage, AutoPK identifies PK parameter variants using a combination of

large language models (LLMs), a hybrid similarity metric, and LLM-based validation. In the second stage, it filters relevant rows, converts tables into key–value text, and employs an LLM to reconstruct standardized, machine-readable outputs. Evaluated on 605 annotated PK tables, including captions and footnotes, AutoPK achieved significant gains in precision and recall compared with direct LLM baselines. For example, AutoPK with LLaMA 3.1-70B reached F1-scores of 0.92 for half-life and 0.91 for clearance, outperforming direct LLaMA 3.1-70B usage by 0.10 and 0.21, respectively. Smaller models such as Gemma 3-27B and Phi 3-12B achieved 2–7 fold F1 improvements, while Gemma’s hallucination rates dropped from 60–95% to just 8–14%. Notably, AutoPK enabled open-source models like Gemma 3-27B to surpass commercial systems such as GPT-4.1 Mini on several PK parameters. By enabling efficient, scalable, and high-confidence PK data extraction, AutoPK offers strong potential for advancing veterinary pharmacology, drug safety monitoring, and public health decision-making.

Keywords

Pharmacokinetic Data Extraction, Automated Information Retrieval (IR), Table Information Extraction, Large Language Models (LLMs)

Building Knowledge Graphs for Aging Biomarker Research

Student Researcher: Srikar Reddy Gadusu

Advisor: Hande K. McGinty, Computer Science

Abstract

Aging biomarkers play a crucial role in uncovering the biological mechanisms behind aging and in developing strategies to support healthy aging. However, the search for reliable aging biomarkers is particularly challenging due to the intricate and multifactorial nature of the aging process. Furthermore, biomarker names and categories are not well-standardized in the current literature. While, a formal definition of a biomarker is nonexistent in the current literature, formally defining biomarkers and standardizing the vocabulary for biomarkers can help accelerate AI research around this concept which can lead to better, faster and more accurate analyses of the existing data and literature. Thus, in this work, we generated Knowledge Graphs that can help us define and standardize biomarkers. We present our Knowledge Graphs (KGs) generated using both an LLM and expert-curated datasets. We compare both KGs to understand why systematic integration between these two models is needed. The integration of Knowledge Graphs (KGs) and Large Language Models (LLMs) presents a promising approach to advancing aging biomarker research through the inherent structured and standardized nature of ontology schemas in knowledge graphs. We showcase that the accuracy of LLM-generated KGs remains questionable but systematic methods such as KNARM can help us with the accuracy of these efforts. In future work, we will propose a synergistic framework where KGs and LLMs interact iteratively to improve both the comprehensiveness and accuracy of aging biomarker information.

Keywords

Enhancing Aging Biomarker Research through Large Language Models and Knowledge Graphs

Drop to Data: Neural Networks for Water and Ice Droplet Analysis for Space Applications

Student Researcher: Aryan Dalal and Jane Turner

Advisor: Hande K. McGinty and Amy Betz, Department of Computer Science and Department of Mechanical and Nuclear Engineering

Abstract

Condensation from moist air occurs when the surface temperature is below the dewpoint temperature. Condensation and subsequent frosting impact many processes and systems on Earth, including heating and refrigeration systems, agriculture, aviation, desalinization, and electronics. We are investigating, in partnership with BioServe and the International Space Station through the National Science Foundation, droplet nucleation and droplet dynamics in microgravity and gravity environments. To process the volume of data yielded by these experiments, a knowledge graph that describes the condensation process and a Neuro-symbolic AI (NAI) model are being developed to quantify and understand droplet growth rates and frequency of coalescence events and identify phase change and freezing mechanisms. This knowledge graph and NAI model will help provide fundamental new insights into the role of gravity and heat transfer on droplet nucleation, growth, and droplet shapes.

Keywords

neuro-symbolic ai, knowledge graph, water ice, droplets, freezing, condensation

Facial Imaging classification and Automated Machine Learning to predict outcome in Feedyard Cattle

Student Researcher: Jordana A. R. Zimmermann

Advisor: Eduarda M. Bortoluzzi, Anatomy and Physiology

Abstract

Background Use of machine learning (ML) and predictive models to identify Bovine Respiratory Disease (BRD) demonstrated promising results for feedyard cattle. Facial images have been valuable in other species; however, to our knowledge cattle facial images at treatment have yet to be used to predict feedyard cattle outcomes. Objective This study aimed to predict feedyard cattle outcomes (Recovered or Did-not-finish (DNF)) using facial

images at time of BRD treatment. **Methods** Facial images (n=923) were collected in a cross-sectional observational study evaluating commercial feedyard cattle at time of BRD treatment from July to December 2023. Cattle outcomes were determined 60 days post-enrollment and classified as Recovered: animals that at the end of the 60-day post-enrollment period were still alive or DNF: animals that died or were culled within the 60-day enrollment. Microsoft Azure Machine Learning Studio was used to classify images using 'multi-class' function and data were imported to automated ML feature to model facial image classification. **Results** Images (n=155) were removed due to duplication or poor picture quality. The remaining 768 images, 547 were labeled as Recovered (71%) and 221 were labeled as DNF (29%). 'Resnet50' was elected best algorithm based on accuracy. Facial images classified in best model showed a sensitivity of 54%, specificity of 82%, positive predicted value of 57%, and negative predicted value of 79%. **Conclusions** Results demonstrate potential of facial imaging-based models for predicting cattle outcomes. Integrating these models with other diagnostic modalities could enhance feedyard management strategies. Note: This abstract/poster was previously presented at the Symposium of Artificial Intelligence in Veterinary Medicine

Keywords

Automated machine learning, bovine respiratory disease, facial image, feedyard cattle, image classification, predictive model.

Integrating Satellite and Weather Data into Machine Learning Models for Yield Prediction in Long-Term Corn-Soybean Fertility Research

Student Researcher: Matias Federico

Advisor: Deepak Joshi, Agronomy

Abstract

Satellite remote sensing, when combined with Machine Learning (ML) models, can support farmers in decision-making by providing more accurate yield prediction under different management practices. To test this hypothesis, the last 6 years of data have been used from a long-term soil fertility study that has been conducted for 43 years in eastern Kansas with corn-soybean rotation with five different nitrogen rates (0, 89.7, 134.5, 179.3, and 224.2 kg/ha), three phosphorus rates (0, 33.6, and 67.3 kg/ha), and two potassium rates (0 and 168.1 kg/ha). The main objective of this project was to develop an ML-based framework to predict corn and soybean yield based on long-term nutrient management, weather, and remote sensing data. High spatial-temporal resolution PlanetScope imagery, collected from early vegetative growth through late maturity, was used to derive vegetation indices (VIs) that represent crop performance across the growing season. These data were integrated into four different ML models, Random Forest (RF), XGBoost (XGB), Lasso (LS), and Ridge (RG), and their performance was compared to determine the most effective model for yield

estimations. Preliminary results showed good correlations between yield, satellite imagery-derived VIs, and nutrient application rates, particularly when analyzing peak and cumulative vegetation index values. These findings highlight the potential of remote sensing-based regression models to guide nutrient management decisions, contributing to improved crop profitability while promoting sustainable soil, nutrient, and water management practices.

Keywords

"machine learning", "planetscope", "long-term fertility", "yield prediction", "corn-soybean rotation".

Inverse Reinforcement Learning for Rat Locomotion Data

Student Researcher: Mobina Golmohammadi

Advisor: William Hsu and Majid Jaber-Douraki, Computer Science

Abstract

Problem Statement. This work introduces an approach to modeling rat locomotion in platform-mediated avoidance (PMA) experiments, aiming to build policy-based representations of neural fear and avoidance circuits that predict both observed trajectories and latent behavioral mediators. A baseline is the descriptive occupancy map, which discretizes locomotion into grid-based ground truth. By training predictive policy models using reinforcement learning, we seek to estimate these maps and other PMA measures; because a reward function is required, we use inverse reinforcement learning (IRL) to infer it from experimental data. **Background.** IRL addresses the problem of learning hidden reward functions $r(a, s)$ from sequential behavioral data, enabling predictive forward RL policies in a simulation environment. **Method.** We developed a simulative PMA model that integrates occupancy grid data with time-stamped stimuli from the experiment logs. This defines a Markov Decision Process (MDP) for rat behavior, expressed as value and policy gradients with observed state-action data and unknown rewards, estimated via maximum entropy IRL. **Results.** Preliminary analyses show that IRL and policy-gradient methods recover spatial gradients and localizations of inferred reward. Difference maps indicate higher estimation error in regions predicted to have lower reward. **Summary and Continuing Work.** IRL can recover interpretable reward landscapes from observed trajectories while revealing model failures. Future work will examine policy generalization, search limitations, and comparisons with alternative neural circuit modeling approaches.

Keywords

inverse reinforcement learning, platform-mediated avoidance (PMA), Rats behaviors

Pharma in Review: A NLP Framework for Pharmacological Study Reviews

Student Researcher: Daniel Robertson

Advisor: Hande K. McGinty, Computer Science

Abstract

Pharmacological studies depends on the capacity of researchers to monitor new treatments, trials, and compounds as they emerge. Large language models (LLMs) have shown some promise with summarizing texts and performing reviews, and topic modeling has been used to assist text analytics for systematic reviews. Recently, research is investigating where LLMs might increase the efficiency of text summary and review tasks. This research proposes a framework for taking advantage of LLM prompt pipelines to modify and expand the utility of text analytics for reviews of pharmacological literature. Using a knowledge graph, topic modeling, and LLMs together, we hypothesize that systematic reviews can become faster and more useful requiring less time. At this stage, we are presenting preliminary approaches for pipelines and preliminary results for the current approaches we are using.

Keywords

pharmacology, ontology, LLM, NLP, clinical trials, PTSD, knowledge graph

PHYSICAL HALLUCINATIONS: Prompt, Image, Translate, Make. AI-generated images made physical.

Student Researcher: Brian K Lee

Advisor: Matt Knox, Architecture

Abstract

Brian K. Lee Assistant Professor of Architecture Kansas State University This poster abstract presents the results of a fabrication-based seminar course based on an AI-assisted working method focused on translating AI hallucinations of material conditions into physical material, with the basic operating premise of Prompt. Image. Translate. Make. In this course, students used generative artificial intelligence programs like Mid-Journey to generate images of architectural details or material conditions that they then translated into built physical materiality. Students were encouraged to look for something that focused on detail, was potentially impossible to build, and not something they would design. Students explored a variety of digital and analog fabrication processes to make their physical hallucinations. The course concluded with a reflection paper requiring the students to reflect on the course and respond to 3 questions relating to the process presented in the course and their thoughts on the impact of AI on their future careers. 1. The

architect's/designer's role in the design process relative to AI. 2. The connection between making and design relative to AI. 3. How did the premise and process of this course inform your thoughts on these topics? Students unanimously responded that they felt their careers were safe from AI at least for now. Pointing out AI's shortcomings in the understanding of physics materiality and assembly. The reflections expressed their lack of confidence in AI to understand materiality, physical forces, and assemblies. Even with AI-assisted design, an understanding of materials, assemblies, and physical materiality is required to execute designs in physical reality.

Keywords

fabrication, digital fabrication, physical materiality, material process, design process

Predicting Harmful Algal Bloom Likelihood with Supervised Machine Learning

Student Researcher: Rose Lauren Taylor

Advisor: Cogan Shimizu, Computer Science

Abstract

Harmful algal blooms (HABs) are ecological phenomena that disrupt aquatic ecosystems by degrading water quality and reducing biodiversity. Observations in various regions suggest that bloom frequency may be associated with nutrient enrichment and climatic variability. This study aims to develop a machine learning algorithm that serves as an early warning system to estimate the likelihood of HAB occurrences using real time water quality data. Supervised learning algorithms, including Random Forest, Support Vector Machine, and Gradient boosting, will be used to model associations between historical bloom events and parameters such as temperature, pH, dissolved oxygen (DO), turbidity, and nutrient concentrations. Determining HAB likelihood can help implement timely mitigation strategies, reducing ecological damage and supporting at-risk aquatic ecosystems.

Keywords

Machine Learning, Environmental Science, Algae, Algal Blooms

Psychedelic Ontology for PTSD Treatment: A Preliminary Framework for Standardized AI-Driven Insights

Student Researcher: yinglun zhang

Advisor: Hande K. McGinty, Computer Science

Abstract

Research into psychedelic-assisted therapy for Post-Traumatic Stress Disorder (PTSD) is expanding rapidly, yet progress is hindered by the absence of a standardized, machine-readable vocabulary to connect findings across neuroscience, psychology, clinical medicine, and regulatory science. To address this gap, we are developing the Psychedelic Ontology, a formal knowledge framework built using the Web Ontology Language (OWL). Our approach combines curated CSV templates with automated OWL generation via the ROBOT tool, followed by expert-guided refinement in Protégé. This workflow balances scalability with accuracy, enabling systematic modeling while incorporating domain expertise. The current version defines ten top-level classes and captures essential relationships, such as mechanisms of action at specific binding neuroreceptors, outcome measurements using validated assessments, and regulatory status of substances. By standardizing terminology, the ontology improves interoperability and reproducibility across studies. It also establishes a foundation for advanced computational applications, such as knowledge graph construction, automated literature mining, and clinical decision support. While preliminary, this work represents one of the first efforts to formalize a shared ontology for psychedelics in PTSD treatment.

Keywords

Post-Traumatic Stress Disorder (PTSD), Knowledge Representation, Ontology

Reconstruction of Molecular Imaging Data Using Neural Networks

Student Researcher: Amirhossein

Advisor: Doina Caragea, Computer Science

Abstract

We study the CEI inverse problem of inferring initial atomic positions $\mathbf{x}$ from final fragment velocities $\mathbf{y}$ under $\mathbf{y} = \mathbf{f}(\mathbf{x}) + \epsilon$. Large synthetic datasets are generated from priors (Random-in-sphere, Normal, Wigner) using a CE forward simulator. We evaluate generative models (VAE, normalizing flow/Real NVP, NPE, MDN, diffusion) and physics-informed predictors (GINN, PIRR, PINN) under a unified protocol. A coordinate rotation with velocity-component constraints and origin-fixing constraints, followed by rigid alignment, removes nuisance degrees of freedom and standardizes evaluation. At scale (e.g., one million oriented pairs), the generative models outperform the physics-informed predictors, with Real NVP strongest among evaluated methods. An energy-consistency term improves small-data regimes. Normal and Wigner priors yield high accuracy and transfer across each other; models trained on Random-in-sphere do not transfer to structured tests. Under leave-one-isomer-out, predictions remain accurate, with flows most robust. All results are reported with explicit metrics, ablations, and definitions to support replication.

Keywords

Neural networks; Deep learning; Molecular imaging; Image reconstruction; Generative models; Inverse problems

RESISTOME BASED DISCRIMINATION OF POULTRY PRODUCTION SYSTEMS USING MACHINE LEARNING: PRELIMINARY FRAMEWORK

Student Researcher: Emanoelli Aparecida Rodrigues dos Santos

Advisor: Valentina Trinetta, Department of Food, Nutrition, Health and Dietetics

Abstract

Antimicrobial resistance (AMR) is a challenge for animal production and public health. Advances in DNA sequencing have enabled resistome profiling at high resolution, yet integration of such data into predictive frameworks remains limited. In this project, machine learning (ML) was applied to evaluate whether resistome profiles from poultry production systems (farm, slaughterhouse, environmental, and human sources) could distinguish between conventional (CS) and antibiotic-free systems (AFS). METHODS: Data preprocessing included normalization, sample-wise pseudo count replacement of zeros ($\epsilon = 1e-6 \times \text{row sum}$), and Centered Log-Ratio transformation (CLR). To address dimensionality, feature selection was applied for informative genes and classes. Models (Regularized Logistic Regression, linear SVM, Random Forest) were trained in AMR classes (27 features) and genes (1140 features) using stratified fivefold cross-validation with iterative train/test splits (80/20) and internal validation sets to prevent overfitting. Performance was assessed by ROC-AUC, F1, and balanced accuracy, reporting mean $\pm$ SD across folds. RESULTS AND CONCLUSIONS: Random Forest achieved the best performance (class level: ROC-AUC = 0.80 $\pm$ 0.10, balanced accuracy = 0.74; gene level: ROC-AUC = 0.82, balanced accuracy = 0.75), followed by logistic regression (=0.70) and SVM (=0.69). Low risk of overfitting, simpler architectures with tuned hyperparameters, and enhanced generalization were obtained. Interpretation highlighted bleomycin, aminoglycosides, and macrolide–lincosamide–streptogramin classes, and genes such as *sul2*, *blaEC*, *aadA1*, and *blaCTX-M-8* as contributors to discrimination. ML could identify system-specific resistome signatures and reveal potential biomarkers of AMR in poultry production. This study shows the possibility of combining Next-Generation Sequencing and ML for AMR monitoring in food systems. KEYWORDS: Antimicrobial resistance, Metagenomic, Predictive modeling ACKNOWLEDGMENTS: FAPESP Process nº 23/01185-6 and 23/01195-1.

Keywords

Antimicrobial resistance, Metagenomic, Predictive modeling

Self-supervised Component Segmentation To Improve Object Detection and Classification For Bumblebee Identification

Student Researcher: Jahid Chowdhury Choton

Advisor: William H. Hsu, Computer Science

Abstract

The performance of computer vision (CV) models for object detection and classification is heavily influenced by the number of classes and quality of input images, particularly in biological applications such as species-level identification of bumblebees. Bee identification is time-consuming, costly, and requires specialized taxonomic training. Different deep learning based computer vision models have been proven to overcome this methodological bottleneck through automated identification of bee species from captured images. However, accurate identification of bee species poses significant challenges due to ambiguity, poor image quality, and noisy backgrounds. Moreover, training effective CV models only on bumblebees is more challenging than training on various insect groups, because many bumblebee species have similar visual features and are very closely identical. Existing pipelines (baselines) primarily rely on object detection to crop bees from images and classify the species for each cropped instance. This approach is limited by the inclusion of noisy backgrounds, low resolution, and poor image quality. To address these limitations, we propose an enhanced pipeline that integrates object detection with segmentation to generate body masks for bees and remove background noise. This process is complemented by a classification model that identifies the top-k species for each masked image. The proposed methodology significantly improves both detection and classification performance in most cases, demonstrating its potential to advance automated identification of bee species in complex image datasets. For the cases where the baselines performed much better, we investigated using a state-of-the-art explainable AI model (Grad-CAM) to explain the reason.

Keywords

pollinator health, segmentation, machine learning, deep learning, computer vision

State-Dependent Conformal Perception Bounds for Neuro-Symbolic Verification of Autonomous Systems

Student Researcher: Thomas Waite

Advisor: Radoslav Ivanov, Computer Science

Abstract

It remains a challenge to provide safety guarantees for autonomous systems with neural perception and control. A typical approach obtains symbolic bounds on perception error

(e.g., using conformal prediction) and performs verification under these bounds. However, these bounds can lead to drastic conservatism in the resulting end-to-end safety guarantee. This paper proposes an approach to synthesize symbolic perception error bounds that serve as an optimal interface between perception performance and control verification. The key idea is to consider our error bounds to be heteroskedastic with respect to the system's state --- not time like in previous approaches. These bounds can be obtained with two gradient-free optimization algorithms. We demonstrate that our bounds lead to tighter safety guarantees than the state-of-the-art in a case study on a mountain car.

Keywords

Neural network verification, conformal prediction, gradient-free optimization

**T-TEXTS (Teaching Text Expansion for Teacher Scaffolding):
Enhancing Text Selection in High School Literature through
Knowledge Graph-Based Recommendation**

Student Researcher: Nirmal Gelal

Advisor: Hande K. McGinty, Computer Science

Abstract

The selection of diverse, thematically aligned literature is a significant challenge for high school English teachers due to limited time and resources. This study presents T-TEXTS, a knowledge graph-based recommendation system designed to address this problem by scaffolding educators in their text selection process. We constructed a domain-specific ontology using KNowledge Acquisition and Representation Methodology (KNARM), which was then instantiated into a knowledge graph to power the recommendation engine. Our core contribution is a comparative analysis of two graph embedding paradigms: shallow methods (DeepWalk, Biased Random Walk, and Hybrid model) and a deep method (Relational Graph Convolutional Network, R-GCN). The models were evaluated on both link prediction and recommendation ranking tasks. While the shallow DeepWalk model achieved the highest AUC for link prediction (0.9739), the deep R-GCN model proved superior for the primary tasks of recommendation ranking, outperforming other models on metrics such as Hits@10, MRR, and nDCG@10. This finding supports our hypothesis that deep embedding approaches, by capturing richer relational semantics, are better suited for recommendation tasks on knowledge-augmented datasets. The results demonstrate that T-TEXTS provides an effective, ontology-driven solution to assist educators in making more informed and inclusive curricular decisions.

Keywords

Graph Representation Learning, Graph Neural Network, Knowledge Graph, Recommendation System

TASK-MATS

Student Researcher: Michael McCain

Advisor: Cogan Shimizu, Computer Science and Engineering

Abstract

We aim to develop a generalized framework that supports an AI Orchestrator (AO) for a multi-agent team (MAT) to accomplish tasks in digital manufacturing. Ultimately, we intend for the AO to: (a) grow an understanding of itself and the assigned team, (b) use the team and available tools to learn about the current environment, and (c) seek to accomplish assigned tasks. It will achieve this through the use of a semantically-enabled MAT, a novel Common-Sense Modular Ontology Design Library (CS-MODL), and a Knowledge Graph (KG). CS-MODL is based on the MODL architecture for curating Ontology Design Patterns (ODPs). In particular, this means we will create a set of patterns that tend to describe items, events, and other entities, which are typically commonly occurring in scenarios. However, we will need to fine-tune our notion of common sense to fit the domain of manufacturing. The KG will act as the knowledge backbone for the MATS, facilitating collaboration and coordination among team members for task completion. The KG will act as a framework that has preseeded knowledge, but will be capable of incorporating populated patterns taken from CS-MODL. Thus, we are choosing to use the Modular Ontology Modeling (MOMo) methodology, a pattern-based method, meaning that it will easily integrate with new patterns. To evaluate the ontology itself, we will check it against the best practices proposed in the paper MOMo. As for the MATS, we will use SOTA simulation tests like ARIAC and other physical verification methods.

Keywords

ontology, multi-agent teaming, manufacturing

Tracking Emerging Drug Trends with Multimodal Information Extraction

Student Researcher: Yihong Theis

Advisor: William Hsu, Computer science

Abstract

The widespread circulation of drug-related content on social media poses serious challenges for automated monitoring, requiring methods that can handle diverse modalities, evolving slang, and ambiguous visual cues. We present a multimodal information extraction pipeline for large-scale detection and analysis of drug-misuse references in short-form video platforms. The system integrates automatic speech recognition, optical character recognition, and video captioning with packaging and logo detection to capture both

linguistic and visual signals. A curated lexicon of substances, slang variants, and emojis supports robust normalization of cues across heterogeneous inputs. Extracted mentions are mapped into synsets and ontology-aligned entities, then linked as triples in a knowledge graph. The resulting knowledge structures enable downstream analytics in a Postgres-backed environment, supporting dashboards that differentiate contexts such as promotion, humor, recovery, or education. Our approach demonstrates how combining multimodal evidence with domain-specific ontologies enhances coverage and precision in monitoring emerging drug trends online. This framework establishes a scalable testbed for advancing research at the intersection of natural language processing, computer vision, and public health informatics

Keywords

Multimedia, machine learning, large language models

Unsupervised model-informed deep learning algorithm to solve inverse source problem

Student Researcher: Nhung Nguyen, Aravinth Krishnan Ravi

Advisor: Dinh-Liem Nguyen, Mathematics

Abstract

We propose a model-informed deep learning algorithm for the inverse source problem, in which the objective is to reconstruct the location, shape, and contrast of a source from measurements of the waves it radiates. This problem is fundamental in a variety of applications, including medical imaging, antenna synthesis, and remote sensing, where accurate source characterization is critical. Our algorithm employs a two-step unsupervised approach: first, a stable imaging function, derived from the underlying physical model, provides an initial approximation of the source distribution; second, this information is combined with a truncated Fourier series capturing the dominant modal behavior and utilized as input to a deep neural network. By incorporating both physics-informed and Fourier-based features, the proposed model-informed algorithm achieves enhanced efficiency and robustness, effectively mitigating challenges associated with high dimensionality and limited training data.

Keywords

deep learning, inverse source problem, Fourier transform, inverse problems, physics informed neural networks

ViTs for action classification in videos: An approach to risky tackle detection in American football

Student Researcher: Syed Ahsan Masud Zaidi

Advisor: William Hsu, Computer Science

Abstract

Player safety in American football can be ensured by identifying hazardous tackle mechanics early, especially during the training of early-career players. We study the automated detection of risky tackles from practice videos to support timely interventions. Modern computer vision makes this feasible: transformer-based video models such as ViViT capture long-range temporal dependencies and multiscale motion cues that are difficult for earlier 3D CNNs to represent. However, real-world training data are small, highly imbalanced, and noisy conditions under which existing datasets and models struggle to recover the rare but safety-critical risky class. We extend prior work with a 558-video corpus in which a single player tackles a dummy, each clip labeled Risky/Safe via SATT scores and manually temporally localized at the tackle event (70:30 Safe:Risky). Our method trains ViViT with Focal Loss to counter class imbalance and goes on further to balance the classes by exploring robustness through controlled perturbations (noise, brightness, rotations). Rather than exhaustively test all 54 augmentation combinations, we employ a Taguchi orthogonal array, reducing the search to 18 settings evaluated under 5-fold cross-validation. The best configuration achieves risky recall 0.756, risky F1 0.713, safe recall 0.789, accuracy 0.777, and balanced accuracy 0.773 in the complete set. Compared with a published C3D baseline reported on a smaller, temporally and spatially localized subset (balanced accuracy 0.668, recall 0.556), our approach yields higher balanced accuracy and markedly better risky-class recall. These results show that coupling ViViT with imbalance-aware learning and a principled augmentation design provides reliable detection of rare, safety-relevant tackle mechanics, offering a practical path toward injury-prevention tools for players

Keywords

American football, risky tackle detection, player Safety, video action recognition, vision transformers, ViViT, imbalanced learning, data augmentation, Taguchi orthogonal arrays

VRI-YOLO11: An Optimized Object Detection Model for HRW Wheat Quality Assessment

Student Researcher: Olusola Olagunju

Advisor: Yonghui Li, Grain Science and Industry

Abstract

Research Purpose: In compliance with the U.S. Grain Standards Act, Hard Red Winter (HRW) wheat intended for export must be officially graded by Federal Grain Inspection Service officers. The grading is based on specific criteria, including kernel weight and the presence level of defects such as damaged, shrunken, and/or broken grains, and foreign materials. However, manual visual inspection is repetitive and prone to fatigue-induced inconsistencies. To improve efficiency and accuracy, this study introduces Visual-Reference-Image-YOLO11 (VRI-YOLO11), an enhanced object detection model based on YOLO11.

Methodology: Three major architectural modifications were introduced. The P5 detection head was removed to focus detection on small and medium-sized objects via P3 and P4 heads. The C3K2 module was redesigned by incorporating Ghost and Bottleneck structures to improve computational efficiency. The C2PSA layer was replaced with the CSMM-SEAM module to enhance spatial awareness and feature refinement. Both the default Yolo11 and the improved VRI-YOLO 11 were trained on 300 images of HRW wheat samples containing damaged, shrunken, and broken wheat, contrasting wheat classes, and foreign objects.

Results: Ablation results confirmed that each modification contributed to improved grading performance. Adjusting the detection layer increased precision and mean average precision (mAP) by 2.2%. The redesigned C3K2 and CSMM-SEAM modules reduced Floating Point Operations per second (FLOPs) by 9.52% and 15.9%, respectively. The optimized model achieved a processing speed of 625 FPS. **Conclusion:** The proposed VRI-YOLO11 model achieved high performance in HRW wheat grade assessment, with precision of 0.95, recall of 0.91, and mAP of 0.95. Additionally, the VRI-YOLO11 reduced computational demand and image processing time, offering a robust solution for automated grain inspection.

Keywords

Hard red wheat, Yolo 11, object detection, wheat quality assessment, architectural modification

Index of Posters

Applying Behavior Pattern Ontology for Social Media Analysis, Early Warning Systems, and Cross-Study Integration — Maria Baloch (Hande K. McGinty)

AutoPK: A Hybrid LLM and Similarity Metric Framework for Extracting Pharmacokinetic Data from Complex Tables and Text — Hossein Sholehrasa (Doina Caragea, Majid Jaber-Douraki)

Building Knowledge Graphs for Aging Biomarker Research — Srikar Reddy Gadusu (Hande K. McGinty)

Drop to Data: Neural Networks for Water and Ice Droplet Analysis for Space Applications — Aryan Dalal and Jane Turner (Hande K. McGinty and Amy Betz)

Facial Imaging classification and Automated Machine Learning to predict outcome in Feedyard Cattle — Jordana A. R. Zimmermann (Eduarda M. Bortoluzzi)

Integrating Satellite and Weather Data into Machine Learning Models for Yield Prediction in Long-Term Corn-Soybean Fertility Research — Matias Federico (Deepak Joshi)

Inverse Reinforcement Learning for Rat Locomotion Data — Mobina Golmohammadi (William Hsu and Majid Jaber-Douraki)

Pharma in Review: A NLP Framework for Pharmacological Study Reviews — Daniel Robertson (Hande K. McGinty)

PHYSICAL HALLUCINATIONS: Prompt, Image, Translate, Make. AI-generated images made physical. — Brian K Lee (Matt Knox)

Predicting Harmful Algal Bloom Likelihood with Supervised Machine Learning — Rose Lauren Taylor (Cogan Shimizu)

Psychedelic Ontology for PTSD Treatment: A Preliminary Framework for Standardized AI-Driven Insights — Yinglun Zhang (Hande K. McGinty)

Reconstruction of Molecular Imaging Data Using Neural Networks — Amirhossein Ghanaatian (Doina Caragea)

RESISTOME BASED DISCRIMINATION OF POULTRY PRODUCTION SYSTEMS USING MACHINE LEARNING: PRELIMINARY FRAMEWORK — Emanoelli Aparecida Rodrigues dos Santos (Valentina Trinetta)

Self-supervised Component Segmentation To Improve Object Detection and Classification For Bumblebee Identification — Jahid Chowdhury Choton (William H. Hsu)

State-Dependent Conformal Perception Bounds for Neuro-Symbolic Verification of Autonomous Systems — Thomas Waite (Radoslav Ivanov)

T-TEXTS (Teaching Text Expansion for Teacher Scaffolding): Enhancing Text Selection in High School Literature through Knowledge Graph-Based Recommendation — Nirmal Gelal (Hande K. McGinty)

TASK-MATS — Michael McCain (Cogan Shimizu)

Tracking Emerging Drug Trends with Multimodal Information Extraction — Yihong Theis (William Hsu)

Unsupervised model-informed deep learning algorithm to solve inverse source problem — Nhung Nguyen, Aravindh Krishnan Ravi (Dinh-Liem Nguyen)

ViTs for action classification in videos: An approach to risky tackle detection in American football — Syed Ahsan Masud Zaidi (William Hsu)

VRI-YOLO11: An Optimized Object Detection Model for HRW Wheat Quality Assessment — Olusola Olagunju (Yonghui Li)