About Me



NAME

Tozammel Hossain

TITLE

Assistant Professor, Data Science

Director, MIDAS Lab

YEARS AT UNT

4 years

RESEARCH FOCUS AREAS

- Clinical predictive modeling from EHRs
- Population health & event forecasting
- Medical imaging AI (CNNs, ViTs)
- LLMs for trustworthy health information
- Single-cell genomics & comp. biology

Clinical predictive modeling from electronic health records

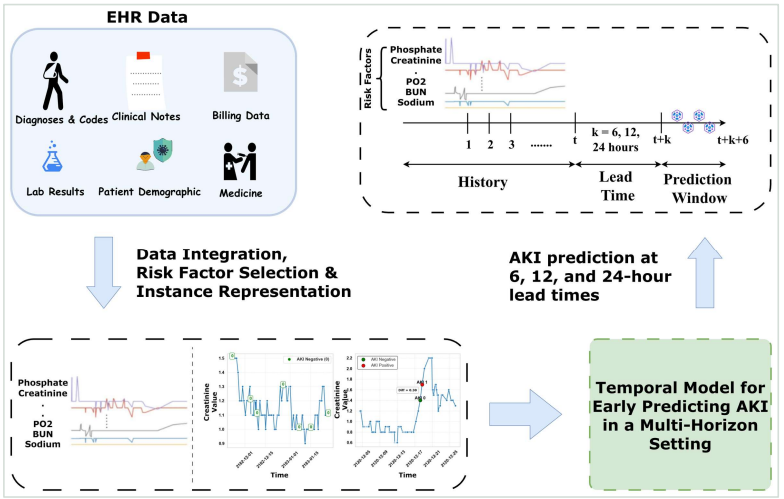
Clinicians need early, explainable warnings rather than after-the-fact predictions

Why
- ICU patients can decline fast
- Early warnings save lives

Problem
- Most models alert too late to act
- Clinicians can't see why a patient was flagged

How
- Hybrid Transformer + LSTM models (RenalTransLSTM)
- Temporal Concept Tracing for explanations
- An LLM writes the explanation in plain clinical language

Contributions
- Predicts kidney injury up to 24 hours ahead
- Shows which risk factors drive each alert
- Explanations take 0.19 s, not minutes



RenalTransLSTM. 48 h of ICU history → AKI risk 6, 12 and 24 h ahead

Selected papers: Badhon et al., AAAI FSS 2025 · Shuvo et al., IEEE MeMeA 2024 · Karanam et al., AMIA CIC 2022 · Mosa et al., J. Bus. Anal. 2021

Population health surveillance and event forecasting

From tracking outbreaks on social media to behavior-aware epidemic digital twins

Why

- Outbreaks spread faster than official reports
- Human behavior shapes how disease spreads

Problem

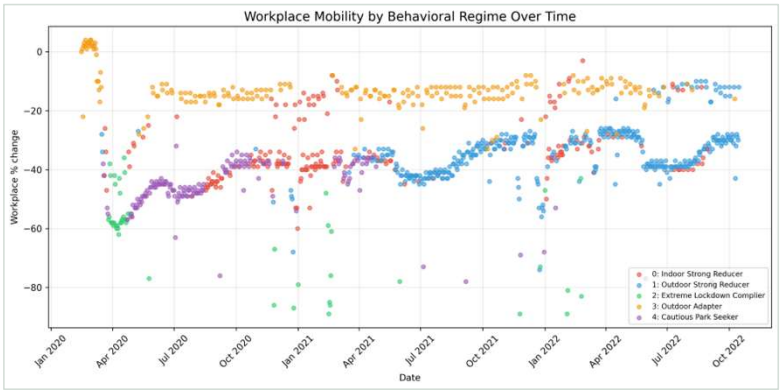
- Official surveillance data arrives too late
- Epidemic models treat behavior as fixed

How

- Social media signals for early detection
- Attention-based LSTMs for event forecasting
- LLM-driven epidemic digital twin (MobilityTwin)

Contributions

- Tracked flu from social media
- Detected epidemics earlier with social network sensors
- Simulated a whole county with behavior-aware agents
- 22% lower error than a standard epidemic model



MobilityTwin. Five mobility regimes in Denton County, 2020–22, drive LLM behavior decisions

Selected papers: Chen et al., ICDM 2014 · Chen et al., DAMI 2016 · Shao et al., KDD epiDAMIK 2018 · Hossain et al., Front. AI 2022

Interpretable, label-efficient AI for medical imaging

Expert labels are scarce, and clinicians need to see what the model sees

Why

- Many clinics lack imaging specialists
- Early screening prevents blindness and cancer

Problem

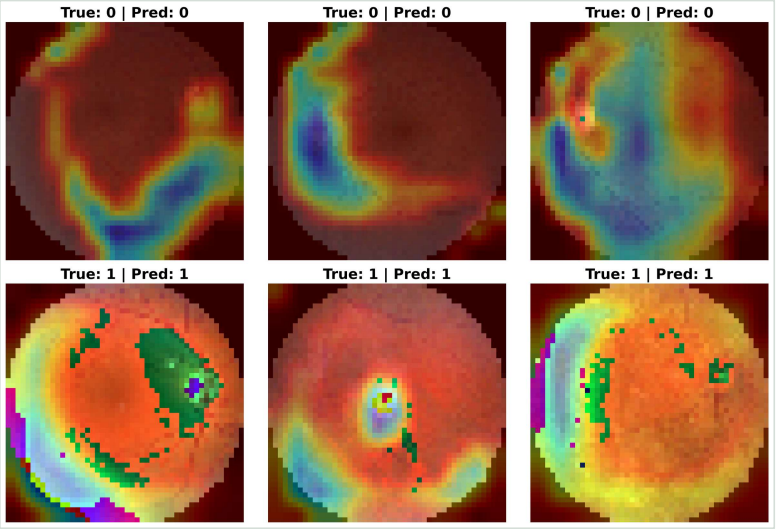
- Expert-labeled images are scarce and costly
- Clinicians distrust black-box models

How

- Semi-supervised vision transformers (AUM-ST)
- Lightweight multi-structural CNNs (CerviXpert)
- Grad-CAM shows where the model looks

Contributions

- 88% accuracy from 5 labeled images per class
- Beats standard CNN baselines
- Highlights the lesions behind each prediction
- Light enough for low-resource clinics



Grad-CAM. Healthy eyes (top) vs. diabetic retinopathy (bottom): attention lands on lesions

Selected papers: Rikka et al., IJCNN 2026 · Badhon et al., Digital Health 2025 · Akash et al., Digital Health 2024

Large language models for trustworthy biomedical information

Grounding LLMs in evidence so that health information can be trusted

Why

- Health misinformation harms public health
- People need answers they can trust

Problem

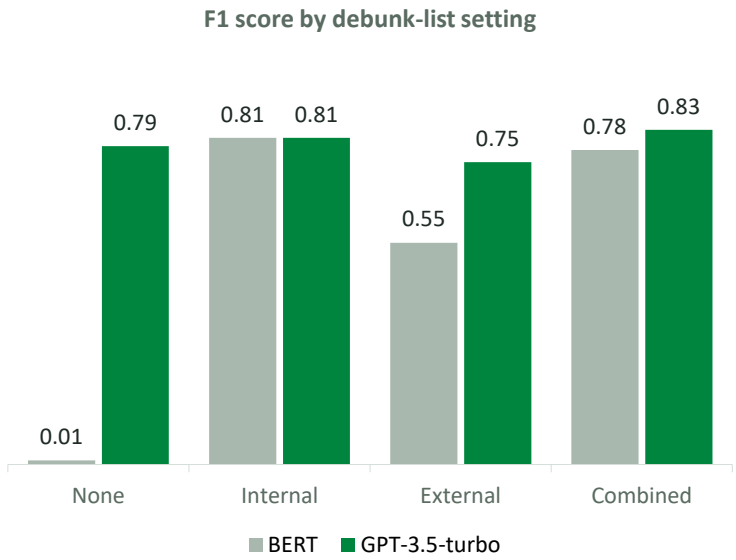
- Labeled data is scarce in a new crisis
- LLMs can sound confident but be wrong

How

- LLMs grounded in debunk lists
- Retrieval-augmented generation (OPEN-RAG)
- Multimodal RAG assistants over EHR data (CareLens)

Contributions

- Detects misinformation without labeled training data
- Better multi-hop reasoning in open-source LLMs
- Clinical assistants that answer from patient records



COVID-19 misinformation. F1 score of BERT vs. GPT-3.5-turbo across debunk-list settings

Selected papers: Rostami et al., CHASE 2025 · Islam et al., EMNLP Findings 2024 · Vuppaladhadiam et al., AMIA 2026