

BINAY DAHAL



binaydahal.com

Senior Data Scientist | NLP, CV, Healthcare, SCADA system

Minneapolis, MN (702) 372-9126 dabinay@gmail.com [binay-dahal](https://www.linkedin.com/in/binay-dahal)

EXECUTIVE HIGHLIGHTS

Phd with research focus on AI and its applications on NLP and Computer Vision, and neural architecture search using Neuroevolution.

- 10 years experience in the field, 5+ years of professional experience.
- Authored 10+ peer-reviewed publications, 150+ citations.
- As the first data scientist in the division at Met Council, introduced AI in wastewater treatment processes. Did multiple AI deployments helping in forecasting and process optimization. Played a key role in transitioning from on-prem to cloud-based data pipeline.

PROFESSIONAL TRANSITION

Research Engineer, Logpoint, 2015-2016 Research Assistant, UNLV, 2018-2020 Senior Applied Scientist, R1 RCM, 2022-2023 Senior Data Scientist, Met Council, Twin Cities, 2024-

KEY DEPLOYMENTS

2024 Influent Flow Forecasting: Train, test, deployment of LSTM models to forecast influent flow at the treatment plant for the next 24 hours.

- Improves the 90 minutes forecast span of current non-ML models.
- 90% CI models trained to provide a range of flow every hour.
- Crucial during flow spikes for staff and equipment planning.

Agentic Data Platform: Databricks app-based modular and agentic platform to let stakeholders talk to data, perform analytics and consume data science results using english language.

- Multi-agent system to understand user queries, generate required code, correct errors, and present results.
- Let data scientists create their modules and deploy. Agents bind modules together. Division-wide integrated analytics platform.

Microscopic Image analysis: Vision transformer(ViT) based model trained on microscopic slide images to detect microbial health of waste water that is being treated.

- Detects the growth of flocs, bacteria and microorganisms.
- About 80% time saved for lab personnel on manual report creation on wastewater microbiology.

Workforce forecasting: Applied stock-flow model to employee churn to forecast annual hiring number of wastewater plant operators over next 10 years.

- Monte Carlo simulation over bayesian for each churn factor combined with target optimal number. Provides a statistical basis of hiring needs.

2022 Mischarges Identification: Xgboost based classification model to detect missing charges on submitted claims from health provider to insurer.

2023 → Detects 90% of the false positives of the existing rules based system with almost perfect precision, thereby saving potentially huge manual detection cost.

PUBLISHED RESEARCH

- **Learn to ask what you don't know.** Assessing question asking capabilities of pretrained transformers.(2022)
- **Effective mutation and recombination for evolving convolutional networks.** Proposed novel neuroevolution based algorithm for architecture search of CNNs.(2020)
- **USRRM: Pairwise ranking and scoring images using its aesthetic quality.** Developed an objective function that trains deep CNN to assess image aesthetic.(2019)
- Other NLP research on paraphrase identification, acronym detection, etc.
- For full published research: [Google Scholar](https://scholar.google.com/citations?user=...)

EDUCATION

PhD in Computer Science(2021)
University of Nevada, Las Vegas
Dissertation: From language comprehension towards general AI

MS in Computer Science(2018)
University of Nevada, Las Vegas

BE in Computer Science(2013)
Kathmandu University, Nepal

DATA & AI SKILLS

Data Pipelines
ML and Deep learning
Time-series forecasting
Predictive Modeling
LLMs
Finetuning
RAG
Agents
MCP

ENGINEERING STACK

Python: Tensorflow, Pytorch, keras, pandas, sklearn, pandas, streamlit, mlflow, pyspark, matplotlib, seaborn

Query language: MSSQL, SparkSQL

Ecosystem: Databricks, Azure, Huggingface

Visualization: PowerBI

Project tracking: Git, Azure devops

OTHER WORKS

Databricks connect excel addin:
Excel addin to fetch data from databricks without writing queries.

Data pipeline optimization:

Optimized existing azure to delta lake scada pipeline saving 90% of the compute cost.