

Dipkamal Bhusal

✉ db1702@rit.edu 🔗 Google Scholar 🐙 Github 🌐 Webpage in LinkedIn

Overview

I am dedicated to advancing the field of reliable AI that improves trust and understanding of deep learning models. My research focuses on understanding deep learning models with tools and techniques from explainable machine learning and adversarial machine learning. I have extensively worked in computer vision and cybersecurity models, with recent exploration of reliability of large language models (LLMs).

Education

- Ph.D. Rochester Institute of Technology**, Rochester, NY, USA,
2021–Present *Concentration – Computing and Information Science*
Advisor – Dr. Nidhi Rastogi .
- M.Sc. Tribhuvan University, Institute of Engineering, Pulchowk Campus** Lalitpur, Nepal,
2019–2021 *Concentration – Information and Communication Engineering*
Advisor – Dr. Sanjeeb Prasad Panday
Thesis – Multi-label classification of thoracic diseases using DenseNet on chest radiographs .
- B.E. Tribhuvan University, Institute of Engineering, Pulchowk Campus** Lalitpur, Nepal,
2012–2016 *Electronics and Communication Engineering*
Advisor – Dr. Nanda Bikram Adhikari
Thesis – Prototyping of a voice command based object recognizing robot using speech and image feature extraction .

Publications

- [1] A Mehrotra, **Dipkamal Bhusal***, N Rastogi, "Hessian Sets: Uncovering Feature Interactions in Image Classification" at Attributing Model Behavior at Scale (ATTRIB), NeurIPS 2024. ***Mentorship**.
- [2] MT Alam*, **Dipkamal Bhusal***, L Nguyen, N Rastogi, "CTIBench: A Benchmark for Evaluating LLMs in Cyber Threat Intelligence" at NeurIPS 2024 (Spotlight paper. Top 2-3% of accepted papers. ***Equal Contribution**).
- [3] **Dipkamal Bhusal***, MT Alam*, L Nguyen, A Mahara, Z Lightcap, R Frazier, R Fieblinger, GL Torales, N Rastogi "SECURE: Benchmarking Generative Large Language Models for Cybersecurity Advisory" at 40th Annual Computer Security Applications Conference (ACSAC). ***Equal Contribution**.
- [4] **Dipkamal Bhusal**, MT Alam, MK Veerabhadran, M Clifford, S Rampazzi, N Rastogi. "PASA: Attack Agnostic Unsupervised Adversarial Detection using Prediction & Attribution Sensitivity Analysis" at 9th IEEE European Symposium on Security and Privacy (2024)
- [5] **Dipkamal Bhusal**, R Shin, AA Shewale, MK Veerabhadran, M Clifford, S Rampazzi, N Rastogi. "SoK: Modeling Explainability in Security Analytics for Interpretability, Trustworthiness, and Usability." at 18th International Conference on Availability, Reliability and Security (2023)
- [6] MT Alam, **Dipkamal Bhusal**, Y Park, N Rastogi, "Looking Beyond IoCs: Automatically Extracting Attack Patterns from CTI". 26th International Symposium on Research in Attacks, Intrusions and Defenses (2023)
- [7] S Kasarapu, **Dipkamal Bhusal**, N Rastogi, SM Pudukotai Dinakarrao, "Comprehensive Analysis of Consistency and Robustness of Machine Learning Models in Malware Detection" Great Lakes Symposium on VLSI 2024.
- [8] **Dipkamal Bhusal**, N Rastogi, "Adversarial Patterns: Building Robust Android Malware Classifiers", ACM Computing Surveys 2025
- Under Review [1] **Dipkamal Bhusal**, MK Veerabhadran, M Clifford, S Rampazzi and N Rastogi, "On the connection between model robustness and saliency map interpretability".
- [2] MT Alam, **Dipkamal Bhusal**, N Rastogi, "Revisiting Static Feature-Based Android Malware Detection".
- arXiv [1] **Dipkamal Bhusal**, SP Panday, "Multi-Label Classification of Thoracic Diseases using Dense Convolutional Network on Chest Radiographs"
- [2] MT Alam, **Dipkamal Bhusal**, Y Park, N Rastogi, "CyNER: A Python Library for Cybersecurity Named Entity Recognition". on arXiv

Skills

Languages	Python
Frameworks	PyTorch, Keras
Libraries	NumPy, Pandas, Scikit-learn, OpenCV, Matplotlib
Utilities	Jupyter Notebook, Visual Studio, Git, Latex
Electronics	PCB Design, Circuit Simulation, 8 bit microcontroller programming
Industry	Teaching, Project Management, Public Speaking, Business Strategy and Development

Experience

August 2021–Present	Graduate Research Assistant, Rochester Institute of Technology. As a research assistant in the lab "AI for Security," I work at the intersection of machine learning and security. Primarily, my research focuses on developing reliable and actionable explainable AI, but I have also been involved in knowledge graphs and adversarial machine learning research. My research goals and activities can be enlisted in three major points: study of existing explanation methods in deep learning to understand their shortcomings in various applications, propose design changes to existing explanation methods to improve the reliability of explanations and propose a new framework that handles the shortcomings of existing methods and provides reliable and actionable explanations.
Dec 2016–June 2021	Co-founder/Software Engineer, Paaila Technology. I co-founded an AI startup in December 2016 (Kathmandu) to work on robotics and AI projects. From December 2016 to November 2018, I worked as a developer and project manager for the manufacture of service robots, Pari, Pari 2.0, and Ginger. All of these robots were designed and manufactured in Nepal. I also contributed as a machine learning engineer for developing ML models for specific tasks of the robot like speech synthesis, and recognition. I took the position of Managing Director in December 2018, was involved in handling planning and managerial activities, and was responsible for the overall growth of the startup. I resigned in July 2021 to pursue PhD in the USA.
Sept 2020–Aug 2021	Lecturer, IIMS College. As a lecturer in the Computer Science department, I taught two BSc. IT undergraduate courses: Introduction to Python and Machine Learning. In addition to teaching, I organized various data science camps and Python training sessions.

Teaching

Oct 2024–Dec 2024	Lecturer of Explainable Artificial Intelligence (DSCI 789) at RIT.
-------------------	--

Mentorship and Supervision

2023-2025	<i>Ayushi Mehrotra, Troy High School.</i> Research on interactive feature attribution method for image classifiers..
2024-2025	<i>Sanish Suwal, RIT Masters in Computer Science.</i> Supervised his final year capstone project on study of explanation methods on different model training strategies..
2024-	<i>Sayali Rajesh Kale and Achyut Sridhar Kulkarni, RIT Masters in Data Science.</i> Supervising their data science capstone project on concept-based explanations..

Peer-Reviewer

2024	2nd Workshop on Attributing Model Behavior at Scale, NeurIPS 2024.
------	--

Honors and achievements

Scholarship	Financial Support for Ph.D. in Computer Science at RIT, 2021-Present.
Training	Conducted two-week training in Python and Data Science at IIMS College, Kathmandu, 2021.
Award	National ICT Innovation Award by Ministry of Communication and Information Technology (Nepal Government) for Paaila Technology, 2019 .
Scholarship	Full scholarship for Masters in Information Engineering at Pulchowk Campus, Tribhuvan University, 2019.

- Award Most Creative Business of Nepal by Antarprena. Represented Team Nepal at global finals in Copenhagen, Denmark, 2018.
- Award Best Startup of Nepal by ICT Magazine, 2017.
- Contest Winner of Object Oriented Programming Competition by FlipKarma at Pulchowk Campus, 2014.
- Volunteer Worked with the youth team of Needleweave society to install prefabricated houses for victims of Gorkha earthquake.
- Arts Established a music group Silver Strings for writing and publishing songs on different themes.
- Award Secured third position at national level quiz competition Quizmania broadcasted on national television (2012).
- Scholarship Full scholarship for bachelor in engineering at Pulchowk Campus, Tribhuvan University, 2012.
- Award Secured second position at inter-college oratory competition organized at the 25th anniversary of Balkumari College, 2011.
- Scholarship Full Scholarship at Balkumari College, Chitwan (10+2 college equivalent to junior and senior high school level in US) + College Topper + First Position at final examinations in the whole district.

Graduate Courses

Quantitative Foundations, Deep Learning, Statistical Machine Learning, Foundation of Algorithms, Software Engineering, Neural Network, Image Processing, Big Data.

Selected news media

Republica Network: Using AI for better customer experience.

Digital Trend: On waiter robots of Nepal.

AFP News Agency: Nepal's first robot waiter ready for orders.

Kantipur (Nepali): Paaila Technology.

NDTV: In A First, Nepal's Restaurant Uses Robots As Waiters.

India Today: On waiter robots of Nepal.

South China Morning Post: Meet Ginger: Nepal's first robot waiter is ready to take your order .

New Business Age Magazine: Embracing the Age of AI, Robotics and ML in Nepal.

Republica Network: Ginger Robot of Nepal.

Kathmandu Post: Emergency Ventilators.

Selected Certifications

1. AI for Medical Treatment, Medical Diagnosis, and Medical Prognosis by deeplearning.ai on Coursera. Certificate earned at June, 2020.
2. Deep Learning Specialization by deeplearning.ai on Coursera. Certificate earned in June 2020. Courses include Neural Networks and Deep Learning, Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization, Structuring Machine Learning Projects, Convolutional Neural Networks and Sequence Models..
3. Project Management Principles and Practices Specialization by University of California, Irvine-The Paul Merage School of Business on Coursera. Certificate earned at May 2020.
4. Mathematics for Machine Learning: Linear Algebra and Multivariate Calculus by Imperial College London on Coursera. Certificate earned at April 2020..