Roop Harshit Vankayalapati

roopharshit@gmail.com | (979) 946-0163 | linkedin.com/in/roop-harshit-v | https://github.com/roopharshit

EDUCATION

Texas A&M University, College Station, Texas *Aug 2022 - ongoing*
Ph.D. and Master of Science, Industrial Engineering (Data Science)

- Academic excellence award "Outstanding Master of Science Thesis Student" (2024)
- Relevant Coursework: Applied Bayesian Machine Learning, Stochastic Dynamic Programming and Reinforcement Learning, Deep Learning, Time Series Analysis, Advanced Quality Control, Production System Planning & Control, Lean Engineering

Indian Institute of Technology, Delhi, India. *Aug 2017 – May 2021*
Bachelor of Technology, Industrial and Production Engineering

TECHNICAL SKILLS

Programming Languages: C/C++, C#, Python, R, SQL, GraphQL, MATLAB
Frameworks: PyTorch, TensorFlow, SciKit-Learn, Tableau, PowerBI, Pandas, NumPy
Software & Technologies: Git, Docker, IDEF0, Minitab, Simio, Visio, High Performance Computing

WORK EXPERIENCE

Research Assistant, *Texas A&M University*

Multi-Scale Information Fusion Transformer for Defect Pattern Recognition *Oct 2023 – May 2024*

- Engineered MSF-Trans, a hybrid CNN – transformer framework for high dimensional data processing in semiconductor defect pattern recognition
- Implemented a multi-scale feature extraction technique that achieved 97.17% accuracy in identifying 38 complex mixed-type semiconductor wafer map defect patterns
- Effectively addressed the challenge of overlapping defect patterns by integrating CNN's local feature extraction with transformer's global contextual analysis for precise defect detection in advanced semiconductor manufacturing
- Conducted extensive trials to validate MSF-Trans's superior performance in precision and recall metrics against established deep learning and machine learning models

Bi-directional Digital Twin of a Smart Manufacturing Unit *May 2023 – Dec 2023*

- Engineered a novel bi-directional digital twin with real-time data transfer from an industry 4.0 manufacturing environment utilizing big data technologies and algorithms to enable data-driven insights and analytics
- Captured digital shadow of smart sensors and pneumatic system of the physical twin using Squeaks data collection software
- Successfully developed a virtual Siemens PLC in TIA portal and integrated into the digital twin using NetToPLCSIM framework
- Established communication between virtual Siemens PLC and Allen Bradley PLC using OPC UA communication protocol

PROJECT WORK

Dual – phase image classification using Vision Transformer, *Texas A&M University* *Apr 2024 – May 2024*

- Implemented a Vision Transformer (ViT) model using PyTorch on the CIFAR-10 dataset, achieving image classification into 10 distinct categories through self-attention mechanisms with 96.5% accuracy
- Employed a dual-phase training approach with 45,000 images for feature extraction, fine-tuning on 5,000 images to optimize model generalization
- Applied data augmentation techniques, including random cropping, horizontal flipping, rotation, colour jittering, and random affine transformations to increase dataset variability and robustness of the model
- Utilized dropout layers and One Cycle learning rate scheduling for effective regularization, leveraging TAMU HPRC clusters for efficient GPU usage

Enhanced image classification using Deep Residual Networks, *Texas A&M University* *Feb 2024 – Mar 2024*

- Leveraged PyTorch to build a sophisticated Residual Network (ResNet) model on CIFAR-10 dataset with 91.4% accuracy
- Implemented data augmentation techniques using NumPy to significantly enhance model training and accuracy
- Executed ResNet by incorporating full pre-activation residual blocks with bottleneck designs which led to a notable increase in model accuracy and efficiency through strategic hyper-parameter tuning

Complex Predictive Modelling of Mask-Wearing Behaviour in Individuals, *Texas A&M University* *Mar 2023 – May 2023*

- Implemented innovative Multivariate Imputation by Chained Equation (MICE) imputation method for processing and enhancing data quality from the data of a big survey for analytics
- Explored 10+ statistical learning models and employed Lasso regression for feature selection, explaining 95% variability in predicting individuals' mask wearing behaviour
- Applied Random Forest model, fine tuning different parameters to achieve a highly interpretable model
- Achieved a predictive accuracy of 91% and verified robustness through cross-validation and analysis of diagnostic plots

Time Series Modelling of a Stock Return Employing ARMA+GARCH model, *Texas A&M University* *Oct 2022 – Dec 2022*

- Performed data cleaning and data analysis on the IBM common stock to gauge the extent of correlation among different lags
- Implemented various model selection strategies to identify unique models to better capture the data with minimal parameters
- Employed the ARMA (0,1) + GARCH (1,1) model to fit time series data of stock prices, precisely modelling price volatility
- Forecasted the price of the stock for a 3- year period with a coverage of 0.96 in 95% prediction interval