

SHASHWAT SHAHI

Boston, Massachusetts | shahi.sh@northeastern.edu | 8574235879 | [LinkedIn](#)

EDUCATION

Northeastern University, Boston

May 2025 (Expected)

Master of Science in Computer Software Engineering

Relevant Coursework: Object Oriented Design, Web Design and UX Engineering, Program Structures and Algorithms, Software Engineering

Visvesvaraya Technological University, Bengaluru, India

Aug 2021

Bachelor of Engineering in Computer Science and Engineering

Relevant Coursework: Data Structures and Applications, Design and Analysis of Algorithms, Unix and Shell Programming, Software Engineering, Database Management Systems, Operating Systems, Machine Learning

SKILLS

Languages: Python, C++, HTML, CSS, JavaScript, React, SQL, NoSQL, Cypher Query Language (CQL), Spark

Tools: Version Control (Git), Neo4J, Snowflake, Hadoop, TensorFlow, PyTorch, AWS, Azure, Jenkins, GitHub actions, Airflow, Docker, Tableau

Technologies: Relational and Graph Databases, Statistical Modeling, Data Analysis, Machine Learning (ML), Deep Learning (DL), Reinforcement Learning (RL), Neural Networks, Natural Language Processing (NLP), Image Processing (OpenCV), Multimodal ML, LLM, Generative AI, CI/CD

EXPERIENCE

Tata Consultancy Services Ltd.

Bengaluru, India

Software Developer – ML (Research, and Innovation)

Aug 2021 – Aug 2023

- Worked on building innovative machine learning solutions to solve challenging problem statements
- Invented Domain Knowledge Infused Adaptive Neuro Fuzzy Inference System (DKI-ANFIS), an architecture, for assessing welding joint quality from images, achieving a state-of-the-art accuracy of 82.4% and generating revenue of ~\$100k
- Developed Priority Weighted Graph Neural Networks (PW-GNN), a novel architecture, which can fetch semantic search results from knowledge graph databases and can be customized to domain specific needs with accuracy > 90%
- Designed a module for automating the prediction of ICD-10 disease diagnostic codes (codes used by health insurance companies for disbursement) from a patient's medical records using NLP techniques and Neo4J's KG, with ~90% accuracy
- Demonstrated exceptional problem-solving skills by deriving critical insights to reduce carbon emissions in the aspirational net-zero emission project

PROJECTS

CareerCraft AI (Advanced Career Pathway & Skill Analysis)

[GitHub Link](#)

Tech Stack: Python, NLTK, Spacy, Neo4J, GNN, Aura DB, NumPy, Pandas, AWS (EC2, RDS, S3), GitHub Actions, Docker

- **Knowledge Graph & Skill Mapping:** Built a Python-based system to construct a knowledge graph of roles and skills using data extraction techniques and employed graph algorithms to semantically map user's current skills extracted from resumes to the knowledge graph, identifying skill gaps.
- **Custom NER & Resume Analysis:** Utilized SpaCy and custom NER models to extract user details and skills from resumes with various data preprocessing techniques like noise filtering, ROI segmentation etc. to enhance NER performance, ensuring accurate mapping to the knowledge graph.
- **Personalized Recommendations:** Utilized collaborative filtering and content-based recommendation systems to suggest Udemy courses tailored to individual skill gaps and keyword matching algorithms to suggest job postings relevant to user aspirations and current skills
- **Deployment:** Deployed the application on AWS using Docker containers for scalability

A System for Brain Tumor Segmentation and Classification from MRI Images

Tech Stack: Python, Flask, OpenCV, Pytorch, TensorFlow, ResNet-50, CNN, Image Segmentation, Transfer Learning, NumPy

[GitHub Link](#)

- Designed a high-precision medical tool that segments and classifies brain tumors from MRI images, utilizing cutting-edge machine learning techniques.
- Achieved over 95% accuracy by employing transfer learning methods with **ResNet-50** as the foundational model architecture.
- Leveraged Flask for creating a user-friendly web interface that enables easy uploading and analysis of MRI images.
- Utilized **OpenCV** for advanced image processing, enhancing the system's capability to handle and preprocess medical imaging data.
- Integrated PyTorch's deep learning framework to effectively train and deploy the neural network models.
- Applied NumPy for high-performance numerical computing, aiding in the manipulation and analysis of large imaging datasets.
- **Other Features:** Real-time analysis, secure image upload, and user-centric design for straightforward navigation.

PATENTS/PUBLICATIONS

- A multimodal learning approach towards effective patent similarity search by leveraging knowledge graphs with Priority Weighted Graph Neural Networks (Patent No.: 20231051755) (To be published in ICDMAI 2024)
- An Unsupervised Image Processing Approach for Weld Quality Inspection (Patent No.: 202221033807) [\(Patent's Link\)](#) [\(Paper's Link\)](#)
- Deep Recurrent Neural Network based audio speech recognition system [\(Paper's Link\)](#)
- Data Analytics of COVID-19 outbreak using Linear Regression [\(Paper's Link\)](#)

Awards/Achievements

- **Special Achievement Award (Jul 2023) – Tata Consultancy Services:** Received this in appreciation for outstanding performance throughout the year which resulted in the acquisition of new business for the organization
- **Innovation Spark Award (Jun 2022) - Tata Consultancy Services:** Received this award for designing and developing and filing patent for an IP based solution which won a ~100k\$ contract for the organization from a major manufacturing client