

# SHUBHANKAR JOSHI

+1 (541)-250-1125 ✉ [joshishu16@gmail.com](mailto:joshishu16@gmail.com) [in s-joshi16](https://www.linkedin.com/in/s-joshi16) [portfolio](#) [shubhu163](#)

## EDUCATION

Oregon State University, Corvallis, OR.

Sept 2023 – Sept 2025

Master of Science in Computer Science - GPA – 4.0

Relevant Coursework: Machine Learning, Deep Learning, DBMS, Optimization for AI, Algorithms, Cyber Security, & NLP.

Savitribai Phule Pune University, Pune, India.

Aug 2019 – May 2023

Bachelor of Technology in Instrumentation & Control - CGPA - 9.26 / 10

Relevant Coursework: Data Structures, Web Technology, Project Management, Data Science, Mathematics & Statistics.

## EXPERIENCE

Research Assistant, Oregon State University, Corvallis, OR.

Sept 2023 – Present

- Analyzed **50M+ ribosomal sequences** from 9 diverse human tissues, performing transcriptome alignment and extracting key insights on protein translation using **Bowtie and Pysam**, enhancing **Data Warehousing** capabilities.
- Designed and implemented automated test cases for pipeline validation, ensuring a 15% reduction in manual efforts.
- Developed an experimental design for hybrid **LLM** with transformer and CNN layers for ribosomal count predictions.
- Automated project pipeline with **PyTorch Lightning**, enabling efficient model training and testing.
- Achieved a Pearson correlation coefficient of **0.837** between predicted and true ribosomal coverage in the **test set**, demonstrating strong **Predictive Analytics**.
- Developed testing frameworks for hybrid LLMs using PyTorch Lightning, integrating unit and integration tests for seamless deployment.
- Integrated loss functions like **Poisson and multinomial NLL**, leading to a **5%** increase in model convergence speed.
- Optimized model parameters using **grid search** and **Optuna**, leading to a **10%** increase in model performance metrics.
- Automated **loss metric logging** and visualization using **Wandb**, resulting reduction of manual validation efforts.
- Generated **correlation heatmaps** for shape and count levels across tissues, identifying **key patterns with Power BI**.

Data Analyst Intern, Landmark Techedge Pvt Ltd, Pune, India.

Aug 2022 – May 2023

- Extracted & managed **5GB** daily production data from **MySQL, HDFS, and Excel** to create **Dashboards on Tableau**.
- Pre-processed data with **Python** and **SQL queries** & analyzed it through **Pandas** to calculate **10+** key performance indicators (**KPIs**) for decision-making.
- Collaborated in **agile teams** to update real-time **data visualization dashboards** which increased **sales by 19%**.
- Performed API testing to validate data integrity between **MySQL, HDFS, and Tableau** dashboards.

## PROJECTS

[Bird Classification Robustness](#) 🚩 | Python, PyTorch, Torchattacks.

Jan 2024 – March 2024

- Developed a robust **bird species classifier using a ResNet** with a model **precision** of **98.53%** and **F1 score** of **98.14%**.
- Data Cleaning, Feature Engineering, Feature Selection, Model Training, Model Evaluation & Deployment.
- Utilized predictive analysis and mathematical modeling to enhance classifier precision

[Car Insurance Sales Dashboard](#) 🚩 | SQL, Excel, Tableau.

June 2023 – Aug 2023

- Designed a e-commerce Tableau dashboard handling **30,000+** insurance records, and established **10+ pivotal KPIs**.
- Integrated **Slicers, filters, and parameters** for intuitive data exploration in **Tableau**.

[Ocular Disease Recognition](#) 🚩 | PyTorch, OpenCV, Torchvision, Gradio.

Feb 2023 – May 2023

- Built an ensemble of **4 CNN** models for classifying ocular diseases **from 10K+ fundus** images, achieving **85% accuracy**.
- Applied **image preprocessing** and **feature extraction**; deployed an interactive model with **Gradio**.

## SKILLS

**Data Science & Analytics:** Statistical Analysis, Predictive Modeling, Operations Research, Analytical Frameworks

**Programming & Tools:** Python, R, Java, C++, C#, SQL, Git, Docker, CI/CD Pipelines, Bash

**QA Skills:** Test Automation, Regression Testing, API Testing, Integration Testing

**Data Visualization:** Tableau, Power BI, Matplotlib, Seaborn, Dashboard Creation

**Libraries & Frameworks:** Pandas, NumPy, Matplotlib, Scikit-learn, PyTorch, TensorFlow, Keras, Seaborn.

**Big Data & Cloud:** AWS, Azure, Google Cloud, Apache Spark, Hadoop

**Data Management:** MySQL, MS SQL Server, ETL Pipelines, Data Warehousing

**Soft Skills:** Leadership, Communication skills, Presentation skills, Cross-Functional Collaboration, Problem-Solving.

## EXTRACURRICULAR & PUBLICATIONS

Led a 20-member team as Aaroh Music Club Head, organized events, and performed as a singer. ([sing.shubhankar.](#))

Ocular Disease Recognition using Ensemble Techniques: Published on IEEE Xplore in ICCUBEA - 2023.

User Survey of Music Streaming Apps: Selected in HWWE 2023 and currently in the process of publication.