# Umar Khan

✆ (+92) 335 3429383
✉ umar.khan1@seecs.edu.pk
🖥 http://umarky.com
*Note: Some of my personal commercial projects are given below.*

## Experience

**October '20** **Associate Researcher**,
**Present** *DLL NCAI*, Islamabad, Pakistan.

Working as a member of the primary research group under the supervision of Prof. Dr. Faisal Shafait. My role at DLL NCAI has been oriented around two aspects of same challenge to solve challenging problems with novel ideas and converting those ideas into generalized research publications.

- TabAUG: A novel tabular **Data Augmentation** technique for improving data efficiency in **table structure recognition** in document images. **Accepted in ICDAR2021**
- Achieved state of the art results on a custom dataset with the novdel idea of Word pool based **Text Colorization** for template classification in **Insurance Document Images** for Automated Extraction. **Collaboration with PhelixAI**
- Specie based max n-count calculation using **Fish Tracking and Specie Classification** for biodiversity estimation. **Collaboration with Australian Institute of Marine Sciences**

**June '19** **Machine Learning Engineer**,
**September '20** *Dcube Technologies*, Islamabad, Pakistan.

My main role in the company revolved around Engineering scalable solutions for our multitude of clients by implementing and deploying Deep Learning based architectures to the cloud infrastructure.

- Successfully deployed an end-to-end **Tabular Data Extraction** pipeline. Main challenge involved dealing with data scarcity and data complexity. **Collaboration with CloudSDS**
- Designed and Implemented and a novel **Data Augmentation** technique for a small dataset allowing us to train and deploy the **Table Structure Extraction** algorithm based on GNN. This project was further developed into TabAUG publication **Accepted in ICDAR2021**
- We deployed a cloud based **Recommendation System** that was auto inference on changes in the google analytic data for users and items. **Collaboration with AARZ**
- Built a python package **PyTabulate** for generating tables **Metadata** allowing for visualizing and understanding table datasets. This package also formed the basis for the statistical driven augmentations in **TabAUG**.
- Custom **Data Labeling Tool** for of tagging **Vertical and Horizontal** tables. This reduced labeling time from an average 7 seconds per image to 2 seconds per image.

## Publications

**2021** **TabAug: Data Driven Augmentation for Enhanced Table Structure Recognition**,
*U. Khan*, *S. Zahid, M.A. Ali, A. Ul-Hasan and F. Shafait*,
(Accepted in ICDAR 2021), Arxiv.

**2021** **Content-Driven Document Colorization for Table Detection**,
*M.A. Ali, S. Zahid, U. Khan, A. Ul-Hasan and F. Shafait*,
Pre-print available, Pre-Print.

## Education

**2015–2019** **Bachelor's of Engineering in Software Engineering**,
*SEECS-NUST (National University of Sciences and Technology)*, Islamabad, Pakistan.

**Supervisor:** *Prof. Dr. Faisal Shafait*
**CGPA:** *3.46 / 4.0*

**Final Year Thesis:** `Map-less & Autonomous Navigation for Quadcopters`
We worked on developing an autonomous system of maneuverability for navigating a quadcopter to a relative target location in a controlled environment. We used a semi-supervised CNN for depth map estimation using the monocular sensor. Using depth map with the optical flow sensor on board the quadcopter a relative flight path was estimated for reaching the target.

## Internships

**Jul 2017 — Sep 2017** **Requirements Engineering**,
*AAWAZ, Islamabad*, .

AAWAZ is an application that has a focus on providing AI and visual driven tools to help accelerate and improve learning in Autistic children to give them a chance to learn and develop learning skills as close as possible to normal children. My role was designed around Requirements Engineering for the product in developed. I was one of the payoneers who joined the project and my ideas went into the final product. the product also won IBM startup cup. The product is still in development and you can learn more about it here:
AAWAZ

**Key Responsibilities:**
- Stakeholder Surveys to stake out different Requirements
- Stakeholder Interviews for resolving conflicting Requirements
- Feasibility Matchup for Requirements and Resource Availability
- Application Concept Development
- Prototype Development in Unity 3D

**Jun 2017 — Aug 2017** **Research Assistant**,
*Dr. Seemab Latif, NUST*, .

I conducted my research under the supervision of Dr.Seemab Latif who is a PhD in Natural Language Processing where we looked into the medical aspects of challenges faced by children with Autism in their ability to read, write and speak and develop a strategy to tackle the challenges with the use of NLP. Our support partner was AFIRM (Armed Forces Institute of Rehabilitation Medicine) where we collected our data for developing strategies for an application which was further adopted in AAWAZ to develop the application which showed tremendous success in its beta release.

**Key Responsibilities:**
- Data collection on learning the limitations faced by Autistic Children
- Feedback on aspects of proposed solutions from Center for Austitic Children Pakistan (CACP)
- Brainstorming the potential of NLP in our proposed solutions
- Concept Design for the solution, adapted by AAWAZ

## Teaching Assistant

**Fall 2018**  **Operating Systems**,
*Assistant Professor Dr. Muhammad Ali Tahir, SEECS-NUST*, .

Taught PintOS (open source operating system) implementation to student in labs, helped perform checking and evaluation of their performance throughout the course. In addition I also evaluated several academic test (quizzes and assignments) through the entire course.

**Spring 2019**  **Computer Networks**,
*Assistant Professor Dr. Arsalan Ahmad, SEECS-NUST*, .

I helped in designing a feasible course project for the students documenting project requirements and deliverable, I also helped in evaluating the course project grades. In addition I also evaluated several academic test (quizzes and assignments) through the entire course.

**Spring 2019**  **Design and Analysis of Algorithms**,
*Assistant Professor Dr. Muhammad Ali Tahir, SEECS-NUST*, .

Due to several students facing difficulties in the course I was asked by Dr Ali Tahir to help students with the concepts, conducting several additional class meetups after lectures. In addition I also evaluated several academic test (quizzes and assignments) through the entire course.

## Personal Commercial Projects

**2019-2020**  **Buyer Behavioral Clustering**, *I developed a prototype for behavioral clustering in buyer behavior and the products they would likely buy. This project was for Manager in SalesForce Germany for a company pitch. I have been endorsed by Nikolas Makris on Linked In and would be happy to provide more details if required.*.

**2019-2020**  **Light Transfer**, *Developed a commercial tool for Vactor Manufacturing Inc (parent company Federal Signal Corporation) where highlights in source document images were detected and transferred to the target Document Images reducing their processing time from an average 16 minutes per document to 12 seconds per document with highly reliable system. I have been endorsed by Christopher Kip Ketner on Linked In and would be happy to provide more details if required.*.

**2019-2020**  **Pneumonia Detection**, *Developed a web application for pneumonia detection for a doctor presenting the use of machine learning in medical imaging for a medical convention, our method of pneumonia detection achieved a 92% validation accuracy, the web application is still live and can be found here*.

## Skills

- **Computer Vision**
- **Data Extraction Automation**
- **System Architecture Design**
- Deep Learning Cloud Deployment
- Requirement Engineering

- PyTorch, Tensorflow
- OpenCV, Numpy, NLTK
- Python, C++
- Flask, Django, REST API
- Docker, Kubernetes