# Sohail Ahmed Khan

✉ sohailahmedkhan173@gmail.com https://github.com/sohailahmedkhan173

in linkedin.com/in/sohail-ahmed-khan-97794b175 ☎ +447938893771

## EDUCATION

| | |
|---|---|
| **University of Sheffield** | **Sheffield, United Kingdom** |
| *MSc Cybersecurity and Artificial Intelligence* | *2018-2020* |
| **Comsats University, Islamabad** | **Islamabad, Pakistan** |
| *Bachelor of Science in Computer Science* | *2013-2017* |

## Research Interests

I have recently graduated from the University of Sheffield, United Kingdom with an MSc in Cybersecurity and Artificial Intelligence. My research interests are at the intersection of AI with fields, such as, Computer Vision, Cybersecurity, Natural Language Processing. I am interested in intersecting AI with above mentioned fields to solve real world problems in different domains such e.g. IoT, smart cities, health, finance, education etc. Following are some of the topics I am interested to explore and carry out my research:

- Deep Learning for Computer Vision (e.g object detection, re-identification and tracking, image segmentation, 3d image reconstruction, countering deepfake media etc)
- Machine learning / Deep learning for Natural Language Processng Tasks (e.g. natural language understanding, natural language generation etc)
- Security and Privacy in Machine Learning/Deep Learning (e.g. homomorphic encryption, differential privacy)

I have practical knowledge of deep learning and machine learning libraries and frameworks, such as, PyTorch, TensorFlow, Scikit Learn, Keras, FastAi etc. I have used these libraries during my MSc studies as well as during my MSc dissertation research. I also have technical knowledge of a number of tools I used in my MSc for example, NLTK, Spacy, FastText and OpenCV. I am currently working under the guidance of my mentor Dr. Wasiq Khan, who is a senior researcher in AI at the Liverpool John Moores University, Uk. We are currently working on (1) detection of Covid-19 in X-Ray images, (2) demographic analysis of covid-19 data obtained from different contries in 7 continents, (3) DeepFake media detection, and plan on submitting a comprehensive journal in coming months.

## Research Papers

- **Phishing Attacks and Websites Classification Using Machine Learning and Multiple Datasets (A Comparative Analysis)** - Accepted at **ICIC2020**(International Conference on Intelligent Computing), to be held in October, in Bari, Italy and will be published by Springer. The conference submission is funded by Liverpool John Moores University, UK.
- **Demographic Analysis of COVID-19** - In progress: working in collaboration with, Dr. Wasiq Khan, senior lecturer and researcher at Liverpool John Moores University, UK.

## Experience and Certifications

- **Microsoft Virtual Internship** August, 2020 - November, 2020.
- **Internship Experience UK** July 2020.
- **Cybersecurity Fundamentals** Certified by IBM.

## SELECTED PROJECTS

**Phishing Webstes Classification Using DL/ML (Dissertation)**
In my MSc dissertation, I did a very detailed survey of the previously proposed techniques and also, developed a new phishing websites classification system, which uses Neural Networks as well as some traditional Machine Learning algorithms, such as, Decision Trees, Support Vector Machines, K-Neighbors Algorithm. I used three different feature selection methods along with PCA for dimensionality reduction, to compress feature space. Besides PCA, I have employed Deep Autoencoder for dimensionality reduction, which is an unsupervised neural network able to learn efficient data encodings. This project basically aims to present an in-depth comparison on how well traditional machine learning algorithms perform, in comparison with the deep learning based algorithms. This research also studies how well the algorithms perform on selected set of features and how well does the feature selection methods work in comparison to dimensionality reduction. In this project, I used three different phishing websites datasets to train my models.

I used KERAS deep learning library to implement a feed-forward neural network and a deep autoencoder. I used Scikit Learn to implement feature selection, PCA for dimensionality reduction and 4 machine learning algorithms.

**Flood Detection from Images using Deep Learning.**
I employed Deep learning library KERAS and fine-tuned its MobileNet CNN for binary image classification task. The model accuracy was more than 98% after being trained on about 300 images for 10 epochs. Code is available at my Github portfolio.

**Adversarial Attack on Fine-Tuned Flood Detection Model**
This is basically an implementation of FGSM (Fast Gradient Sign Method) attack on my fine-tuned MobileNet architecture trained for flood detection in images. Code is available at my Github portfolio.

**Grouping Semantically Relevant News Headlines using Word Embeddings**
In this NLP based project, I used SPACY's word embeddings to calculate sentence embeddings of news headlines and then used the calculated sentence embeddings to find cosine similarities between sentences, then I grouped the sentences together based on high cosine similarity scores. I also used Facebook's FastText to calculate my own Sentence Embeddings.

**Named Enitity Recognizer**
The goal of this script is to learn a named entity recogniser (NER) using the structured perceptron algorithm. For each word in a sequence, the named entity recogniser should predict one of the following labels:

- O: not part of a named entity
- PER: part of a person's name
- LOC: part of a location's name
- ORG: part of an organisation's name
- MISC: part of a name of a different type (miscellaneous, e.g. not person, location or organisation)

I also implemented dynamic programming methods such as Viterbi, Beam Search, which the script uses to accelerate the process of training the Perceptron.

**XSS / SQL Fuzzer in Python**
A very basic Fuzzing tool, to fuzz web applications with some predefined Static and then Dynamically generated Payloads. The tool is tested on a particular web application, and has statically defined endpoints to test, but it can be updated to crawl for the potential endpoints with a little effort.

For more projects, please visit my GitHub portfolio at: **https://github.com/sohailahmedkhan173**. I also write articles related to Machine Learning, Computer Vision, Cybersecurity at Medium.com. Link: **https://medium.com/@sohailahmedkhan173**.

## TOOLS & SKILLS

Some Tools I am familiar with:

- TensorFlow
- Keras
- PyTorch
- OpenCV
- FastAI
- Sci-kit Learn
- Python
- Apache Spark
- NLTK
- Spacy
- FastText
- Latex

## References

- **Dr Achim Brucker** (Professor in Cybersecurity/Head of the Cybersecurity Group at the University of Exeter, UK)
- **Dr Nesrine Kaaniche** (Lecturer in Cybersecurity, University of Sheffield, UK)
- **Dr Wasiq Khan** (Senior Researcher in AI, Liverpool John Moores University, UK)