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 *2018-2019*

Comsats University, Islamabad **Islamabad, Pakistan**
Bachelor of Science in Computer Science *2013-2017*

Research Interests

I have an MSc in Cybersecurity and Artificial Intelligence from the University of Sheffield, United Kingdom. My research interests lie in the area of artificial intelligence, specifically in image forensics, computer vision. I am interested in employing machine learning/deep learning techniques 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, adversarial machine learning, 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)

My current research intersects image forensics and deep learning. I am working on deepfake media detection and trying to address the issue of poor generalization capability of current deepfake detection systems. I will also work on adversarial attacks against deepfake systems and their defences.

Research Papers

1. **Phishing Attacks and Websites Classification Using Machine Learning and Multiple Datasets (A Comparative Analysis):** *Accepted at International Conference on Intelligent Computing, Oct 2020*
   DOI: https://doi.org/10.1007/978-3-030-60796-8_26

2. **Analysing the Impact of Global Demographic Characteristics over the COVID-19 Spread Using Class Rule Mining and Pattern Matching:** *Published in Royal Society Open Science Journal, Jan 2021*
   DOI: https://doi.org/10.1098/rsos.201823

3. **Adversarially Robust DeepFake Media Detection Using Fused Deep Convolutional Neural Network Predictions:** *Under Review at International Conference on Image Processing (ICIP)*

Experience and Certifications

Research Assistant at 3d Vision Lab, Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE, *December 2020 - Present*:

-> I am currently working under the supervision of **Dr Hang Dai**. Our current work intersects image forensics and deep learning.

Research Assistant (Intern) at DeepCamera Research Lab, CYENS Centre of Excellence, Nicosia, Cyprus, *September 2020 - December 2020*:

-> Worked under the supervision of **Dr. Alessandro Artusi** on deepfake media detection and adversarial machine learning.

Internship Experience UK, *July 2020*:

-> Fortune 500 companies such as, Amazon, Goldman Sachs, Bloomberg, Google, Vodafone, GSK, Accenture etc participated in the program. The purpose of the internship was to experience what it is like to work in a fortune 500 company, and what daily tasks might look like.

Cybersecurity Fundamentals: Certified by IBM, *August 2020*:

-> This certification certifies that the earner of this certification demonstrates a foundational understanding of cybersecurity concepts, objectives, and practices. This includes cyber threat groups, types of attacks, social engineering, case studies, overall security strategies, cryptography, and common approaches that organizations take to prevent, detect, and respond to cyber attacks. This also includes an awareness of the job market.

Completed Projects

Phishing Webstes Classification Using DL/ML (MSc 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.

-> Project Link: *https://github.com/sohailahmedkhan173/Phishing-Websites-Classification-using-Deep-Learning*

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.

-> Project Link: *https://github.com/sohailahmedkhan173/Flood-Detection-from-Images-using-Deep-Learning*

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.

-> Project Link: *https://github.com/sohailahmedkhan173/Adversarial-Attack-on-Fine-Tuned-Flood-Detection-Model*

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.

-> Project Link: *https://github.com/sohailahmedkhan173/Efficient-Named-Entity-Recognizer-using-Structured-Perceptron-Viterbi-and-Beam-Search-Python*

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.

-> Project Link: *https://github.com/sohailahmedkhan173/Simple-SQL-XSS-Fuzzing-Tool-PYTHON-*

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)

References will be provided upon request.