# Sandipan Choudhuri

[+1] 480-300-9385 | s.choudhuri@asu.edu | s-choudhuri.github.io | s-choudhuri | S-Choudhuri

## EDUCATION

**PHD** | COMPUTER SCIENCE                                    USA | Aug 2017 - Summer 2023 [tentative]
*Dissertation: "Domain Adaptation in Unconstrained Label Spaces."*                    CGPA: 4/4
Arizona State University

**MASTERS** | COMPUTER SCIENCE & ENGINEERING                          India | Aug 2013 - Jul 2015
*Thesis: "A Quality-Concordance Measure-Based Approach to Edge Detection."*                CGPA: 8.44/10
Jadavpur University

**BACHELORS** | COMPUTER SCIENCE & ENGINEERING                         India | Aug 2009 - Jul 2013
West Bengal University of Technology                                    CGPA: 8.60/10

## PROFESSIONAL EXPERIENCE

**ROVICARE** | MACHINE LEARNING INTERN                              USA | May 2022 – Aug 2022

- Leveraged **Azure Form-Recognizer** and **image-processing** routines to develop a framework for seamless extraction of requisite entities from scanned patient intake forms with **92%** accuracy, expediting the decision-making process for patient intake by **64%**.
- Engineered a word localizer module using **Azure Cognitive Service** to render scanned patient files searchable for keywords.
- Streamlined **HubSpot** operations for enhancing in-house productivity: built ticketing system add-ons to analyze support requests for actionable insights, automated contract filling through **PandaDoc** integration, and created workflows in deal pipelines.
- Deployed machine learning models as **Python RESTful APIs** on the **Azure App Services** with **Docker** runtime.

**INTERBIZ** | MACHINE LEARNING INTERN                              USA | May 2021 - Aug 2021

- Developed a batch-parser for poorly-digitized health insurance claim forms with **Google Cloud Vision API** and **region-proposal networks** with an extraction accuracy of **96%**, translating to an improvement in the turnaround time by **72%**.
- Reduced the processing time for auto insurance collision claims by **49%** through the development of category detectors and form-processors for the scanned estimates, using **Google Document AI**, **Google Cloud Vision API**, and **deep-siamese networks**.
- Built a **Google Cloud Web Application** with **OAuth 2.0 framework** authentication (conforming to HIPAA security and privacy standards) for secured access to healthcare documents from Google Drive and Google Cloud Storage.

**NETXT LAB** | RESEARCH ASSISTANT                                  USA | Aug 2019 – Present

- Addressed the data annotation overhead by proposing **transfer-learning frameworks** with realistic label-set overlap assumptions that yielded state-of-the-art performance in **67%** of the adaption challenges on benchmark datasets.
- Developed an engine for forecasting future overdose incidence volumes and strategically determining optimal locations to aid the AHCCCS[1] personnel in setting up MAT [2] facilities under budgetary constraints; an initiative to **combat state's opioid crisis**.
- Evaluated the effectiveness of Twitter data to monitor the opioid crisis through characterization using **RoBERTa** language model.

**CMATER LAB** | JUNIOR RESEARCH FELLOW                            India | Aug 2015 – Mar 2017

- Technical lead in a production-level **Breast Cancer Diagnostic System** development program on Fine-Needle Aspiration Cytology images. Duties involved conducting research and evaluation, mentoring the development team, communicating with the expert groups at diagnostic units for requirements gathering, data acquisition, and configuring access controls.
- Conducted exploratory analysis on challenges of **Object Localization, Contour Detection**, and **Image Segmentation**. Addressed them by proposing novel machine-learning frameworks for RGB and near-infrared image datasets and publishing findings.

## COURSEWORK & SKILLS

| | |
|---|---|
| COURSE[3] | Deep Learning • Machine Learning • Statistical Learning • Computer Vision • Data Mining • Data Visualization |
| RESEARCH | Domain Adaptation • Time-Series Analysis • Object Localization • Contour Detection • Image Segmentation |
| PLATFORM | Google Cloud Platform (GCP) • Microsoft Azure |
| LANGUAGE | Python • Java • C++ • C • SQL • JavaScript • HTML • CSS |
| TECHNOLOGY | Docker • Apache Spark • Tableau • Matlab • Git • ClickUp |
| LIBRARY | Pytorch • Keras • Pandas • Scikit-learn • OpenCV • NumPy • Apache Spark MLlib • Seaborn • Flask • D3.js |

[1] AHCCCS: Arizona Health Care Cost Containment System      [2] MAT: Medically Assisted Treatment      [3] Selected graduate AI courses

# ACADEMIC PROJECTS

**TIME-SERIES ANALYSIS**    Python, Statsmodels, Pandas, Scikit-Learn, Seaborn, Tableau

Developed a pipeline for predicting opioid overdose incidence volumes in a future time frame using time-series analysis and forecasting over *Arizona EMS data*, *opioid overdose records* and *demographic* information. Tasks involved time-series decomposition, stationarity testing using *unit-root tests*, transforming the time series to stationary, testing for seasonality, addressing missing values, computing partial autocorrelation and estimating the series forecastability, running *Granger Causality tests* to gauge the effectiveness of the time series in forecasting another, and developing forecasting models using *ARIMA*, *SARIMA*, *SARIMAX*, *VARMA*, and *VARMAX*.

**DEEPVID++**    Python, Keras, Flask, D3, JavaScript, Knowledge Distillation

The reliability of a deep-learning model is majorly dependent on the interpretability of its internal operations. Citing this, the *DeepVID* framework is improved to train a basic network using a complex deep-neural model such that its behavior is visually interpretable. This is accomplished by utilizing *variational information* to generate neighbors around data instances and training a linear network to replicate the complex network's behavior on the generated samples through *knowledge distillation*.

**DOMAIN ADAPTATION FRAMEWORKS**    Python, Pytorch, Keras, Adversarial Learning

Developed frameworks for cross-domain image translation, data augmentation and classification tasks under a domain adaptation setup, with *constrained* and *unconstrained* label-space assumptions between the source and target domains (*closed-set*, *partial* and *open-set* scenarios); built adaptation models for scenarios where access to source data is prohibited during the adaptation process (*source-free* scenario).

**OBJECT LOCALIZATION**    Python, Keras, OpenCV, Matlab, Computer Vision

Formulated an approach for object localization and semantic image segmentation in natural scenes, using image *super-pixelization* and *deep supervised classification*. Context-based information is captured through patches formed by the first and second neighbor super-pixels (atomic units of an image formed by a perceptual grouping of pixels). These are then mapped to the corresponding object labels using deep classifier networks. A consensus-based labeling strategy is subsequently employed by ensembling the output probabilities at different scales through multiple voting routines.

**IMAGE SEGMENTATION**    Python, Pytorch, Open-CV, Bio-medical Image Processing

Developed a framework for polyp segmentation in optical colonoscopic video sequences by coupling *deep-siamese networks* with a clustering policy using *balanced entropy-based random walks* to identify informative frames and segment out existing polyps (precursors to colon cancer).

**SEMANTIC CONTOUR DETECTION**    Python, Keras, C++, OpenCV, Game Theory, Computer Vision

A multi-scale feature-based contour detection approach is presented here from a game-theoretic standpoint. A mixed strategy Nash Equilibrium is searched in a two-player game (*texture* and *color*-based feature extractors acting as players, with strategies involving *extraction at different scales*). The objective centers around finding a weighted balance of these strategies to minimize the false positives (highlighting spurious edges) and false negatives (missing salient contours).

# PUBLICATIONS

**Asilomar**
2022

Domain-Invariant Feature Alignment Using Variational Inference For Partial Domain Adaptation
**S.Choudhuri**, S. Adeniye, H Venkateswara, A.Sen                                        (accepted)

**AdvML@Kdd**
2022

Coupling Adversarial Learning with Selective Voting Strategy for Distribution Alignment in Partial Domain Adaptation
**S.Choudhuri**, H Venkateswara, A. Sen

**Globecom**
2021

Optimal Cost Network Design for Bounded Delay Data Transfer from PMU to Control Center
A. Sen, S. Roy, K. Basu, S. Adeniye, **S. Choudhuri**, A. Pal

**Drcn**
2020

Structural Dependency Aware Service Chain Mapping for Network Function Virtualization
A Sen, **S Choudhuri**, K Basu

**Asilomar**
2020

Partial Domain Adaptation Using Selective Representation Learning For Class-Weight Computation
**S.Choudhuri**\*, R. Paul\*, A.Sen, B.Li, H Venkateswara

**Rdmlda**
2019

Combining Multilevel Contexts of Superpixel using Convolutional Neural Networks to Perform Natural Scene Labeling
A Das, S Ghosh, R Sarkhel, **S Choudhuri**, N Das, M Nasipuri

**epiDamik@Kdd**
2018

Identification of At-Risk Groups for Opioid Addiction Through Web Data Analysis
K Basu, **S Choudhuri**, A Sen, A Majumdar, D Dey

| | |
|---|---|
| Mмтc 2018 | User Satisfaction-Driven Bandwidth Allocation for Image Transmission in a Crowded Environment<br>**S Choudhuri**, K Basu, A Sen |
| Ijprai 2018 | Object Localization on Natural Scenes: A Survey<br>**S Choudhuri**, N Das, R Sarkhel, M Nasipuri |
| Ficta 2017 | A Quality-Concordance Metric Based Contour Detection by Utilizing Composite-Cue Information and Particle Swarm Optimisation<br>**S Choudhuri**, N Das, M Nasipuri |
| Icacci 2016 | A Multi-Cue Information Based Approach to Contour Detection by Utilizing Superpixel Segmentation<br>**S Choudhuri**, N Das, S Ghosh, M Nasipuri |

## TEACHING

Foundation of Algorithms  (TA[1], Spring & Fall 2018-2021)  •  Data Mining  (TA[1], Fall 2019)  •  Object-Oriented Programming and Data Structures  (TA[1], Fall 2017 & Spring 2018)  •  Principles of Programming  (TA[1], Fall 2017)

## AWARDS & SERVICES

- Recipient of the **Engineering Graduate Fellowship** for *"strong academic work and research progress."*
- Recipient of **SCAI Ph.D. Conference Fellowships**.
- Served as the **Web chair** for the *International Workshop on Network Science for Quantum Communication Networks, 2022* (**NETSCIQ-COM - INFOCOM**). Duties included designing, updating, and stress-testing the conference website, managing submissions, providing tech support, and reviewing workshop live streams.
- **Co-wrote research grants** and was awarded **AHCCCS Grant** for the *Arizona State Opioid Response Data Project*.
- **Teaching** and **Research Awards reviewer** for the *Graduate & Professional Student Association* at Arizona State University.
- **Top 0.5%** of *0.25 million applicants* who took the *Graduate Aptitude Test in Engineering (GATE)-2013* in Computer Science (conducted by the Indian Ministry of Human Resource Department).

---

[1]  TA: Teaching Assistant