

SIDDHANT SRIVASTAVA

Final Year Student (B.Tech-M.Tech dual degree programme)

(+91)7388617414 ◊ siddhant.srivastava11@gmail.com

<https://in.linkedin.com/in/siddhantsrivastava1> ◊ <https://github.com/sidsriv>

EDUCATION

ABV Indian Institute of Information Technology & Management, Gwalior

Integrated B.Tech-M.Tech degree in Information Technology

July 2014 - Present

Final year Graduate

Overall GPA: 8.24/10.0

Cathedral Sr. Sec. School, Lucknow

Intermediate (C.B.S.E.)

Score: 94.4%, Year:2014

High School (C.B.S.E.)

Score: 9/10, Year:2012

RESEARCH INTERESTS

- **Data Science:** Machine Learning, Statistics, Artificial Intelligence, Natural Language Processing, Deep Learning, Nature Inspired Algorithms.
- **Bioinformatics:** Structural Biology, Genomic Data Analysis, Mathematical modelling of Biological Systems.

CONFERENCES AND PUBLICATIONS

- A.Trivedi, S.Srivastava, A.Mishra, A.Shukla, R.Tiwari, "Hybrid evolutionary approach for Devanagari handwritten numeral recognition using Convolutional neural network", presented in **Sixth International Conference on Smart Computing and Communications**, Kurukshetra, India on 07-08 December 2017.

RESEARCH EXPERIENCE

Institute for Bioinformatics and Applied Biotechnology

May'16 - July'16

Construction and analysis of protein-protein interaction network in *Leishmania donovani*
Supervisor: Shubhada Hegde

- *Research objective* : Comparison of cross study gene expression data (Agilent Microarray) for drug sensitive and drug resistant *L.donovani*.
- *Individual Responsibility*: Cross-platform normalized gene expressions from different experiments. Calculated Pearson correlation, mutual information score for gene pairs and implemented Principal component analysis, K-means clustering algorithms on normalized datasets.
- *Accomplishments*: Performed correlation and mutual information analysis on gene pairs to study pattern similarity in gene expression and further employed cluster algorithms. Normalized 22 gene expression samples from 6 different experiments available in NCBI GEO database.
- *Tools used* : Python Scientific stack (Numpy, Pandas, Scikit-learn, Matplotlib), R and Rstudio (Bioconductor).

ABV Indian Institute of Information Technology & Management Gwalior

May'17 - July'17

Hybrid evolutionary approach for Devanagari handwritten numeral recognition using Convolutional neural network

Supervisor: Anupam Shukla, Ritu Tiwari

- *Research objective* : To develop a novel approach for training Convolutional Neural Network using initial weights based on Genetic Algorithms for fast optimization.

- *Individual Responsibility: Pre-processed 25,000 images of isolated handwritten Hindi characters using binarization, blurring, cropping, size normalization. Implemented vanilla CNN architecture with integrated genetic algorithm optimization using python.*
- *Accomplishments: Achieved state-of-the-art accuracy of 96.83% on test data using our model and compared our results with standard classification algorithms like logistic regression, random forest and support vector machine.*
- *Tools used : Python 3.5, Numpy, Matplotlib, Scikit-learn.*

Indian Institute of Technology Guwahati

May'18 - July'18

Mining fine-grained relations from Wikipedia Infoboxes

Supervisor: Ashish Anand

- *Research objective : To extract fine-grained relations among entity mentions from wikipedia infobox data.*
- *Individual Responsibility: Developed script to extract infobox data from article entries present in Wikipedia multistream XML dump. Developed script to scrape vital article list from wikipedia metadata page. Pre-processed relations and attributes present in infobox entries for person, organisation and location infobox types.*
- *Tools used : Python 3.5, regex, BeautifulSoup, pandas, xml.*

Indian Institute of Technology Guwahati

May'18 - July'18

Self-Attention based BLSTM model for named entity recognition in Drug, Disease and Clinical data.

Supervisor: Ashish Anand

- *Research objective : To develop a novel Long short term memory (LSTM) model with multi-head attention to recognise named entities in Biomedical and Clinical data.*
- *Individual Responsibility: Developed novel architecture consisting of multi-head attention from transformer network cascaded with Bi-LSTM and conditional Random Fields (CRF). Used pre-trained word embeddings from PubMed data and tested model on NCBI Disease NER, i2b2-2010 Clinical NER and Drugbank+MedLine Drug NER Datasets.*
- *Accomplishments: Achieved result better than state-of-the-art model on Disease dataset (precision:83.27%,recall:84.9% score:84.08%) and Clinical dataset (precision:84.63%,recall:83.06%,F-score:83.83%).*
- *Tools used : Python 3.5, Tensorflow, Numpy, Scikit-Learn.*

CERTIFICATIONS

- **Specializations**

1. Python for Everybody, a 5-course specialization by University of Michigan on Coursera. Specialization Certificate earned on April 6, 2016.
2. Genomic Data Science, an 8-course specialization by Johns Hopkins University on Coursera. Specialization Certificate earned on October 4, 2016.
3. Data Science , a 10-course specialization by Johns Hopkins University on Coursera. Specialization Certificate earned on November 18, 2016.
4. Machine Learning, a 4-course specialization by University of Washington on Coursera. Specialization Certificate earned on September 16, 2016.
5. Deep Learning, a 5-course specialization by deeplearning.ai on Coursera. Specialization Certificate earned on April 3, 2018.
6. Data Mining, a 5-course specialization by University of Illinois on Coursera. Specialization Certificate earned on June 7, 2017.

• Relevant Courses

1. Bioinformatic Methods I by University of Toronto on Coursera. Certificate earned on December 18, 2015.
2. Bioinformatic Methods II by University of Toronto on Coursera. Certificate earned on January 17, 2016.
3. Data Analytics and Statistical Inference by Duke University on Coursera. Certificate earned on November 27, 2015.
4. Simulation and modeling of natural processes by University of Geneva on Coursera. Certificate earned on August 26, 2017.
5. Computational Neuroscience by University of Washington on Coursera. Certificate earned on October 25, 2016.

TECHNICAL SKILLS

Computer Languages	C/C++, Python, R, MATLAB, MySQL, Julia.
Software & Tools	Shell, L ^A T _E X, Anaconda, Rstudio.
API's and packages	Numpy, Scipy, pandas, Sci-kit learn, matplotlib, Tensorflow, Keras.

ACADEMIC ACHIEVEMENTS

A.I.S.S.C.E 2014 Topper, Cathedral Sen. Sec. School
Secured 99% marks in Chemistry, A.I.S.S.C.E 2014

POSITION OF RESPONSIBILITY

Event Organizer	October 2016 - November 2016
<i>Otaku - Anime Quiz</i>	<i>Infotsav 2016</i>

- Organized Manga trivia quiz at Infotsav, the annual technical festival at ABV-IIITM Gwalior

EXTRA-CURRICULAR

1. Active Member of Coursera's Beta Tester Community, my task is to review upcoming courses before making it available to students world wide.
2. Active Member of Coursera's Data Science Community, a network of data science professionals and enthusiasts where we discuss and keep up to date with latest news, trends and technologies in data science.

RECOMMENDATIONS

Anupam Shukla	Ritu Tiwari	Ashish Anand
Professor	Assoc. Professor	Assoc. Professor
ABV-IIITM Gwalior	ABV-IIITM Gwalior	IIT Guwahati
anupamshukla@iiitm.ac.in	ritutiwari@iiitm.ac.in	ashish.anand@iitg.ernet.in