Blessy Antony

Blacksburg, Virginia, USA

+1 (570) 917 9513 • blessyantony@vt.edu • blessyantony9.github.io

Education

Virginia Polytechnic Institute and State University (Virginia Tech) **Blacksburg, USA**
Doctor of Philosophy (Ph.D.), GPA: 3.97/4.00 *Jan. 2021 - Dec. 2025*
Computer Science
Advisor: Dr. T. M. Murali, Co-advisor: Dr. Anuj Karpatne.

Virginia Polytechnic Institute and State University (Virginia Tech) **Blacksburg, USA**
Master of Science (M.S.) en route to Ph.D., GPA: 3.97/4.00 *Jan. 2021 - May 2023*
Computer Science

Fr. Conceicao Rodrigues College of Engineering (Fr. CRCE) **Mumbai, India**
Mumbai University
Bachelor of Engineering (B.E.), CGPA: 9.36/10.00 *June 2012 - May 2016*
Computer Engineering

Research Experience

Postdoctoral Research Fellow **Virginia Tech**
NSF COMPASS Center, Advisors: Dr. T. M. Murali, Dr. Anuj Karpatne *Jan. 2026 - present*

- Developing computational approaches to model organoids, three-dimensional multicellular entities derived from stem cells to mimic native organ functions.
  - Orchestrating several computational methods that predict cellular responses to perturbations to develop an agentic framework aimed at modeling organoid maturation.
  - Implementing *in silico* models of organoids to simulate different experimental conditions, thereby optimizing the time and resources required for exploratory wet-lab analyses.

Graduate Research Assistant **Virginia Tech**
Dr. T. M. Murali's Computational Systems Biology Research Group *May 2021 - Dec. 2025*

- Designed VILLA, a unique retrieval-augmented generation (RAG)-based framework to accurately retrieve information from scientific publications using large language models (LLMs).
  - Benchmarked and evaluated the performance of general-purpose and domain-specific LLMs, and RAG- and agent-based frameworks in curating datasets in virology.
  - Collaborated with virologists from the NSF COMPASS Center to obtain ground-truth datasets and conduct qualitative evaluation of LLMs.
- Developed a viral protein language model (pLM) called HAVEN that uses a hierarchical self-attention architecture. The pLM generates rich embeddings for viral protein sequences of any length.
  - Pretrained HAVEN on 1.2 million viral protein sequences using masked language modeling.
  - Curated a virus-host dataset containing viral protein sequences and the host from which they were sampled.
  - Fine-tuned HAVEN for virus-host prediction using the systematically curated dataset. The viral pLM accepts any protein sequence of any virus and predicts its host.
  - Assessed the generalizability of the model by integrating HAVEN with a few-shot learning framework to predict rare and unseen hosts and hosts of unseen viruses.
- Implemented machine learning models to predict post-acute sequelae of SARS-CoV-2 infection (PASC, or long COVID) based on information derived from the electronic health records (EHRs) of acute COVID-19 infection available in the National Clinical Cohort Collaborative (N3C).
  - Trained logistic regression and random forest models using features such as symptoms experienced by COVID-19 patients, drugs ordered or administered to them, measures of COVID-19 treatment, comorbidities, and demographic information.
  - Identified risk factors in COVID-19 patients using SHAP-based interpretation of predictions.
  - Performed a novel cross-site analysis to evaluate the generalizability and impact of different data sources on prediction models.

- Implemented reusable classification pipelines in the N3C Enclave for disease prognosis using EHRs.

Professional Experience

JP Morgan Chase & Co. **Mumbai, India**
Software Engineer - Associate *Jan. 2019 - Dec. 2020*
Software Engineer - Analyst *July 2016 - Jan. 2019*

- Developed a pipeline of containerized microservices that collate, parse, and enrich application logs to determine the availability of applications and reduce issue triaging time by 20%.
- Implemented a framework using JMeter, Elasticsearch-Logstash-Kibana, and Node.js to automate performance testing and enhance report generation. The system saved $\sim \$100,000$ annually by reducing manual effort by 70%.
- Revamped and migrated Post Trade Reporting services from legacy to new, regulated data centers, and containerized platform (Kubernetes), thus reducing the infrastructure cost by 50%.
- Re-implemented and converted monolithic applications to microservices to improve resiliency, stability, and efficiency.
- Conceptualized and materialized a prototype of a Virtual Reality system to analyze bank transactional data and establish customer expenditure patterns, during the global JP Morgan Chase & Co. Hackathon 2018 as part of a two-member team. The project was selected for the final round globally.

Coursework Projects

Evaluation of Code Representation Learning Models **Virginia Tech**
Course: CS 5814 Deep Learning *Spring 2022*

Designed a standardized pipeline of downstream tasks for evaluation of code representation learning models - CodeBERT (uses contextual information only) and GraphCodeBERT (uses both contextual and structural information).

- Used BigCloneBench and CodeSearchNet datasets to benchmark the performance of models in clone detection and code summarization tasks respectively.
- Fine-tuned pretrained models from Hugging Face Transformers, evaluated them across ten randomly sampled test datasets, and computed and compared performance metrics.
- Observed the importance of structural information in learning code embeddings with GraphCodeBERT yielding higher BLEU (code summarization), and AUPRC, F1, precision, and recall (clone detection) scores.

SARS-CoV-2 Interactors Identification using Network-Based Label Propagation **Virginia Tech**
Course: CS 5854 Computational Systems Biology *Spring 2021*

Implemented random walk with restart (RWR) algorithm to discover human proteins interacting with the SARS-CoV-2 virus proteins directly or indirectly.

- Propagated the labels from known human protein interactors in the STRING protein-protein interaction network and identified unknown virus interactors.
- Exercised the implementation within National Clinical Cohort Collaborative (N3C) Data Enclave and executed it on the knowledge graph, 'KG-COVID-19'. Enhanced the implementation using sparse matrices to improve performance and reduce memory footprint.
- Compared with external experimentally derived results using Fisher's Exact Test and Jaccard coefficient.

Social Media Sentiment Analyzer for COVID-19 **Virginia Tech**
Course: CS 5824 Machine Learning *Spring 2021*

Built a model that stratifies Tweets related to COVID-19 into three sentiment categories to identify posts propagating anti-vaccine misinformation through social media.

- Implemented logistic regression, support vector machine (SVM), and long short-term memory recurrent neural network (LSTM-RNN) multi-class classification models using the TweetsCOV19 dataset.
- Performed hyperparameter tuning and 5-fold cross-validation for all models, and evaluated them using accuracy, precision, recall, and F1 scores. Highest accuracy of 98.2% achieved by LSTM-RNN model.
- Added functionality to analyze and classify real-time data by streaming Tweets using Twitter API v2.

Cancer Diagnosis using Support Vector Machine (SVM) **Fr. CRCE**
B.E. Capstone Project *Fall 2015 - Spring 2016*

Constructed a breast cancer classification model using GNU Octave 4.0 to implement SVM with Gaussian RBF Kernel and improved the prediction accuracy to $\sim 97\%$ using feature selection.

- Utilized the Breast Cancer Wisconsin dataset to train and test the built classification model.
- Executed an algorithm to compute the impact of each attribute and determined an effective feature set.
- Implemented a mechanism to find optimal kernel configuration and data-splitting ratio for training and validation.

Aarogya – An Intelligent Multi-Agent Pediatric System **Fr. CRCE**
Funded by the Department of Science and Technology, Government of India *Spring 2015 - Fall 2016*

Developed an intelligent system offering pediatric services to counter the infant mortality rate caused by unavailability of pediatricians in rural India. Mentored by the Innovation and Entrepreneurship Development Center at Fr. CRCE.

- Implemented intelligent agents using Java Agent Development Framework v4.3.3 to gather information using an adaptive questionnaire, ascertain the disease, and prescribe treatment.
- Built an Android application as the user interface for patients requiring the services of the intelligent agents.

Publications

Peer-reviewed Publications

1. **Antony, B.**, Dutta, A., Aggarwal, S., Gatne, V., Gökdemir, O., Grimes, S., Lauring, A., Wasik, B., Karpatne, A., Murali, T. M., "VILLA: Versatile Information Retrieval from Scientific Literature Using Large Language Models." *In Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD)*. 2026.
2. **Antony, B.**, Blau, H., Casiraghi, E., Loomba, J. J., Callahan, T. J., Laraway, B. J., Wilkins, K. J., Antonescu, C. C., Valentini, G., Williams, A. E., Robinson, P. N., Reese, J. T., Murali, T. M., Chute, C. G., "Predictive models of long COVID." *eBioMedicine*. 2023.
3. Shirzadian, P., **Antony, B.**, Gattani, A. G., Tasnina, N., Heath, L. S. "A time evolving online social network generation algorithm." *Scientific Reports*. 2023.
4. Reese, J. T., Blau, H., Casiraghi, E., Bergquist, T., Loomba, J. J., Callahan, T. J., Laraway, B., Antonescu, C., Coleman, B., Gargano, M., Wilkins, K. J., Cappelletti, L., Fontana, T., Ammar, N., **Antony, B.**, Murali, T. M., Caufield, J. H., Karlebach, G., McMurry, J. A., Williams, A., Moffitt, R., Banerjee, J., Solomonides, A. E., Davis, H., Kostka, K., Valentini, G., Sahner, D., Chute, C. G., Madlock-Brown, C., Divers, J., Haendel, M. A., Robinson, P. N., N3C Consortium, RECOVER Consortium. "Generalisable long COVID subtypes: Findings from the NIH N3C and RECOVER programmes." *eBioMedicine*. 2023.
5. Chan, L. E., Casiraghi, E., Laraway, B., Coleman, B., Blau, H., Zaman, A., Harris, N. L., Wilkins, K., **Antony, B.**, Gargano, M., Valentini, G., Sahner, D., Haendel, M., Robinson, P. N., Bramante, C., Reese, J. "Metformin is associated with reduced COVID-19 severity in patients with prediabetes." *Diabetes Research and Clinical Practice*. 2022.

Preprints

1. **Antony, B.**, Haghani, M., Lauring, A., Karpatne, A., Murali, T. M., "HAVEN: Hierarchical Attention for Viral protEin-based host iNference.", *bioRxiv*, 2025.

Under Review

1. Nikeghbal, P., **Antony, B.**, Murali, T. M., Rajagopalan, P., "Human Liver Organoids for Investigating Viral Infections", 2026.

Presentations

Conference Presentations

- **Antony, B.**, Dutta, A., Aggarwal, S., Gatne, V., Gökdemir, O., Grimes, S., Lauring, A., Wasik, B., Karpatne, A., Murali, T. M., "VILLA: Versatile Information Retrieval from Scientific Literature Using Large Language Models", *AI for Sciences track at the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '26)*. Jeju, Korea. August 2026. [Peer-reviewed Poster Presentation]
- **Antony, B.**, Dutta, A., Aggarwal, S., Gatne, V., Gökdemir, O., Grimes, S., Lauring, A., Wasik, B., Karpatne, A., Murali, T. M., "VILLA: Versatile Information Retrieval from Scientific Literature Using Large Language Models", *2026 NCI Summer School on Algorithmic Cancer Biology.* National Institutes of Health (NIH), USA. July 2026. [Peer-reviewed Oral and Poster Presentation]
- **Antony, B.**, Haghani, M., Lauring, A., Karpatne, A., Murali, T. M., "VirProBERT: A Sequence Language Model for Predicting Viral Hosts." *Artificial Intelligence for Nucleic Acids (AI4NA) workshop at International Conference on Learning Representations (ICLR).* Singapore. April 2025. [Peer-reviewed Poster Presentation]

- **Antony, B.**, Bu, J., Chan, A., Karpatne, A., Murali, T. M. "Zoonosis Prediction using Language Models." *Intelligent Systems for Molecular Biology (ISMB)/European Conference on Computational Biology (ECCB)*. Lyon, France. July 2023. [Peer-reviewed Oral and Poster Presentation]

Invited Talks

- "Using Language Models to Predict Zoonosis." *Center for Emerging, Zoonotic, and Arthropod-borne Pathogens (CeZAP)* Social - Computational Biology and Disease Modeling. Virginia Tech, USA. March 2023.

Mentorship and Teaching Experience

Mentored four graduate and six undergraduate students in Murali Research Group, Virginia Tech. *2023 - present*

- Systematic and standardized benchmarking of virus taxonomy prediction methods.
- Evaluating the generalization limits of protein language models (pLMs).
- Web interface to evaluate responses of LLMs in retrieving information from scientific publications.
- Case studies on extending and evaluating a protein language model in predicting virus-hosts using dissimilar viral protein sequences.
- Zero-shot evaluation of virus-host predictor on out-of-distribution hosts.

Invited guest lecture: CS 6824: Adv. Topics in Comp. Biology & Bioinformatics, Virginia Tech. *Spring 2026*

- Delivered a lecture on 'Predicting Viral Hosts using Machine Learning' to a graduate class of ~ 30 students.
- Mentored a Ph.D. student for course project on how the composition of pretraining dataset affects viral pLMs.

Co-mentor in Google Summer of Code (GSoC). *Summer 2024*

- Project in National Resource for Network Biology (NRNB) organization in GSoC: Inferring gene regulatory networks using foundation models.
- Co-authored the abstract for the accepted poster at the International Conference on Research in Computational Molecular Biology (RECOMB)/International Society for Computational Biology (ISCB) on Regulatory & Systems Genomics with DREAM Challenges, 2024.

Graduate teaching assistant: CS 2506 Computer Organization II, Virginia Tech. *Spring 2021*

- Graded assignments/homework and held weekly office hours for a class of ~ 250 students.

Undergraduate teaching assignment: Advanced Algorithms, Fr. CRCE, Mumbai University. *2016*

- Delivered a lecture on 'Graph Theory - Vertex Cover Problem' to an undergrad class of ~ 75 students.

Honors and Awards

Won the Pandemic Science Nutshell Games. *July 2026*

- Challenge of explaining research in 90 seconds and judged by a panel of community members.
- Presented research titled "Can a virus infect a train or an airplane?" at Conor O'Neill's in Ann Arbor, Michigan.
- Organized by the Center for Communicating Science at Virginia Tech.

Won 2^{nd} place in Methods (the "How") category at Flip the Fair. *Sept. 2025*

- Presented research titled "Viral Language: Teaching Computers the Virus Alphabet" in science-fair-style on tri-fold poster boards to elementary school students at Roanoke Public Library, Virginia.
- Judged by 450 fifth graders from public schools in Roanoke, Virginia.
- Organized by Center for Communicating Science at Virginia Tech.

Scholarship to attend Grace Hopper Celebration of Women in Computing (GHC) *2021*

- Awarded by the Department of Computer Science, Virginia Tech.

Quarterly Global Excellence Award at JP Morgan Chase & Co. *Dec. 2020*

- For strategic and impactful improvement in software architecture.

Excellence in Performance Award at JP Morgan Chase & Co. *Apr. 2017*

- Across all job levels in the department.

Tata Consultancy Services - Best Student of Fr. CRCE Award *2016*

- Across all departments for excellence in academics and co-curricular activities.

Dr. S. S. S. Award for securing 1st rank in Applied Mathematics *2016*

- Across all departments through all 4 years of B.E.

Tata Trust Scholarship *2013, 2014*

- For excellence in academics in freshman and sophomore years of undergraduate program.

Computer Skills

Programming: Python, Java, C, JavaScript, GNU Octave, XML, SQL, HTML-CSS

Platforms: Linux - RedHat and Ubuntu, Windows 11, Mac OS X

Frameworks: PyTorch, NumPy, Pandas, scikit-learn, DSPy, NLTK, NetworkX, Spring, SpringBoot, Node.js, Express.js

Concepts: Cloud Computing (Cloud Foundry), Containerization (Docker, Kubernetes, Apache Mesos), Message Queue (Apache Kafka)

Miscellaneous: Git, LaTeX, Microsoft Office Suite, MySQL, Jenkins, Joget, Firebase

Service

Professional Associations

ISCB Student Council (ISCB-SC) Executive Team: Secretary *2026 - present*

ISCB Student Council Symposium at ISMB 2026: Co-chair of the symposium *2026*

ISCB-SC: Chair of web committee *2024 - present*

ISCB Student Council Symposium at ISMB/ECCB 2025: Member of organizing committee *2025*

ISCB-SC: Member of web committee *2023 - 2024*

IEEE Women in Engineering (IEEE - WIE): Student representative in Fr. CRCE student chapter *2013 - 2014*

Reviewer

ACM SIGKDD Conference on Knowledge Discovery and Data Mining *2026*

JAMA Network Open *2025*

ISCB-Asia Student Council Symposium *2025*

IEEE International Conference on Data Mining (ICDM) *2025*

ISCB Student Council Symposium *2025 - present*

AI4NA workshop at ICLR *2025*

Google Summer of Code *2024*

Asia & Pacific Bioinformatics Joint Conference (APBJC) *2024*

Intelligent Systems for Molecular Biology (ISMB) Conference *2023 - present*

European Conference on Computational Biology (ECCB) *2023 - present*

IEEE International Conference on Bioinformatics and Biomedicine (BIBM) *2021, 2022*

ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB) *2021, 2022*

Virginia Tech

Facilitated the National Dialogue on Ethics in Pandemic Science organized by the NSF COMPASS Center. *2026*

Reviewed proposals for the Graduate and Professional Student Senate (GPSS) - Travel Fund Program (TFP). *2023*

Reviewed proposals for the GPSS - Graduate Research Development Program (GRDP). *2023*

Volunteered as a mentor for the Mentorship Program organized by the Computer Science Graduate Council. *2024*

JP Morgan Chase & Co.

Participated in the 'Force for Good' program to build software applications for nonprofit organizations. *2017, 2020*

Subject Matter Expert in the 'Code for Good' hackathon to guide undergraduate students. *2018, 2019*

Editor-in-Chief of the internal Newsletter Committee for India and Philippines locations. *2017 - 2018*

Last Updated: November 4, 2026