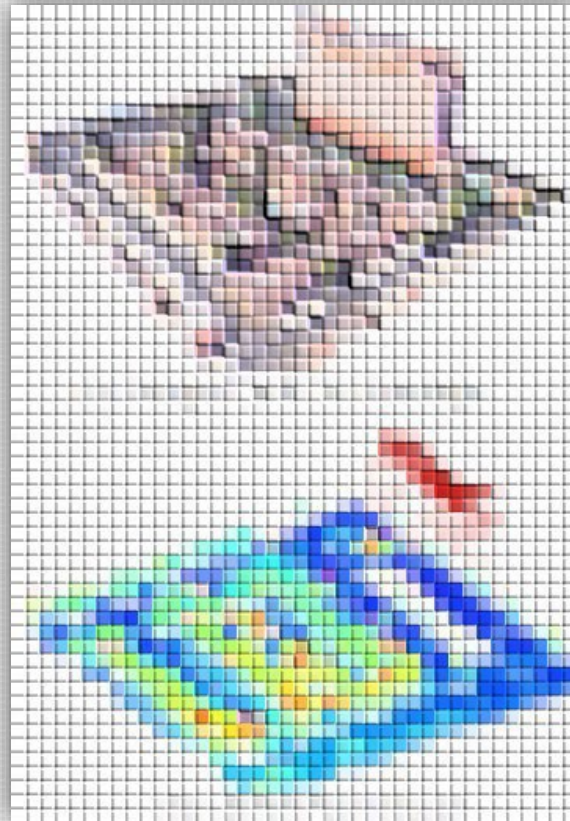# Big Data Research

UNLV | HOWARD R. HUGHES
COLLEGE OF ENGINEERING

Big Data Research

Dr. Rama Venkat
Dean, College of Engineering
Phone: (702) 895-1094
Email: Rama.Venkat@unlv.edu

Dr. Mohamed Trabia
Associate Dean, College of
Engineering
Phone: (702) 895-0957
Email: Mohamed.Trabia@unlv.edu

For more than a decade, UNLV researchers have been conducting a world-class effort in various aspects of big data research. This program has been funded by federal and state agencies, as well as many industrial partners. Our researchers have addressed questions related to many fields, including big data, relative to national security and health issues.

We would like to introduce you to some of our researchers. Please feel free to contact us if we can help with future collaboration.

Big Data
Research Areas of Expertise

- Unstructured data analysis
- Cloud computing
- Deep learning
- Document layout analysis
- Human-computer interaction
- Information retrieval
- Machine learning
- UAS and GIS integration
- Medical image analysis
- Data mining
- Physiological data sensing applications

Big Data Research

Why UNLV?

- UNLV is a leader among the state's public entities dedicated to advancing big data research in the region and beyond.

- UNLV is located centrally in the west and houses the high-performance computing center, also known as The National Supercomputing Institute.

- Additional data center space is located at Switch Communications, which houses the Intel Supercomputer "Cherry Creek" (26,000 cores) cluster. The site provides Federal and State researchers with state-of-the-art, highly-secure cloud computing.

UNLV | HOWARD R. HUGHES COLLEGE OF ENGINEERING

Big Data Research

Why UNLV?

- UNLV's outstanding achievements in big data research, its success in forging public/private partnerships, and its excellent academic programs place the university at the forefront of the field.

- UNLV was awarded funding from various agencies in the Big Data area.

UNLV | HOWARD R. HUGHES COLLEGE OF ENGINEERING

Faculty Involved in
Big Data Research

Dr. Mingon Kang

Assistant Professor, Department of Computer Science

Dr. Jeehee Lee

Assistant Professor, Department of Civil and Environmental Engineering and Construction

Dr. Beiyu Lin

Assistant Professor, Department of Computer Science

Dr. Fatma Nasoz

Associate Professor, Department of Computer Science
Director of Data Science, The Lincy Institute
Affiliate Faculty, Nevada Institute of Personalized Medicine (NIPM)

Dr. Haroon Stephen

Professor, Department of Civil and Environmental Engineering and Construction

Dr. Kazem Taghva
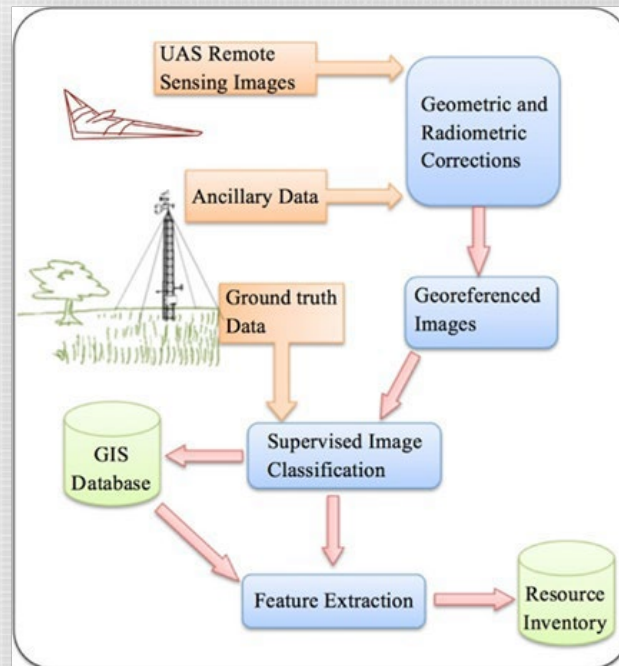
Professor and Chair, Department of Computer Science

Big Data Research

Additional Resources
www.NevadaCompareCare.Net

UNLV | HOWARD R. HUGHES COLLEGE OF ENGINEERING

Big Data

Research Highlights

UNLV | HOWARD R. HUGHES
COLLEGE OF ENGINEERING

Dr. Mingon Kang

Assistant Professor, Department of Computer Science

Director, DataX Lab (http://www.dataxlab.org)

Phone: (702) 895-4884

Email: mingon.kang@unlv.edu

- Expertise
  - Data science, machine learning, big data analytics
  - Deep learning:
    - Interpretable deep learning
    - Integrative deep learning
  - Integrative analysis of multiple types of data such as image and text data
  - Medical image analysis
  - Document layout analysis
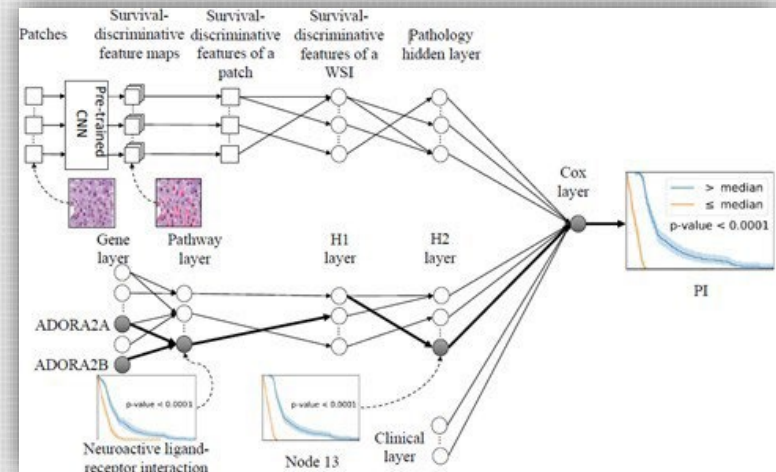
UNLV | HOWARD R. HUGHES COLLEGE OF ENGINEERING

Dr. Mingon Kang

Assistant Professor, Department of Computer Science

Relevant Publications

- M. Kang, E. Ko, T. Mersha, "A Roadmap for Multi-Omics Data Integration using Deep Learning", *Briefings in Bioinformatics*, 2022
- H. Lee, M. Kang, Y. Li, D. Seo, D. Kim, "Epidemic Vulnerability Index for Effective Vaccine Distribution against Pandemic", *ISBRA*, 2021
- J. H. Oh, W. Choi, E. Ko, M. Kang, A. Tannenbaum, and J. O. Deasy, "PathCNN: Interpretable convolutional neural networks for survival prediction and pathway analysis applied to glioblastoma", ppi443-i450, *Bioinformatics*, 2021
- S. Kim, S. Yang, K. Lim, E. Ko, H. Jang, M. Kang, P. Suh, and J. Joo, "Prediction of Alzheimer's disease-specific phospholipase c gamma-1 SNV by deep learning-based approach for high-throughput screening", Proceedings of the National Academy of Sciences of the United States of America (PNAS), 2021
- J. Hao, S. Kosaraju, N. Tsaku, D. H. Song, and M. Kang, "PAGE-Net: Interpretable and Integrative Deep Learning for Survival Analysis Using Histopathological Images and Genomic Data", Pacific Symposium on Biocomputing (PSB), 2019
- T. Mallavarapu, J. Hao, Y. Kim, J.H. Oh, M. Kang, "Pathway-based Deep Clustering for Molecular Subtyping of Cancer," *Methods*, 2019
- S. Kosaraju, M. Masum, N. Tsaku, P. Patel, T. Bayramoglu, G. Modgil, and M. Kang, "DoT-Net: Document Layout Classification Using Texture-based CNN," International Conference on Document Analysis and Recognition (ICDAR), 2019
- Y. Kim, J. Hao, T. Mallavarapu, J. Park, and M. Kang, "Hi-LASSO: High-Dimensional LASSO," *IEEE Access*, 2019, pp 44562-44573, 2019
- J. Hao, M. Masum, J. H. Oh, and M. Kang, "Gene- and Pathway-based Deep Neural Network for Multi-omics Data Integration to Predict Cancer Survival Outcomes," 2019 International Symposium on Bioinformatics Research and Applications (ISBRA), 2019
- J. Hao, Y. Kim, T. Mallavarapu, J.H. Oh, and M. Kang, "Cox-PASNet: Pathway-based Sparse Deep Neural Network for Survival Analysis", Proceedings of IEEE International Conference on Bioinformatics & Biomedicine (IEEE BIBM), 2018
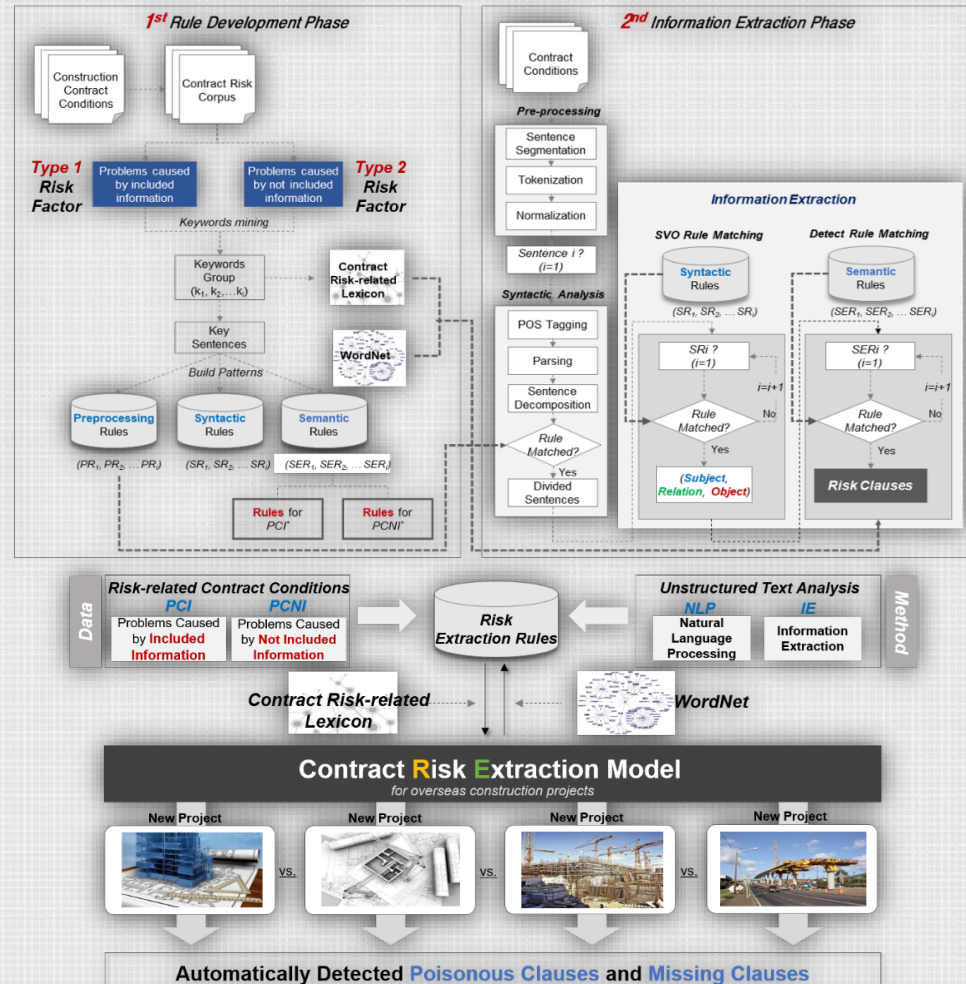
Dr. Jeehee Lee

Assistant Professor, Department of Civil and Environmental Engineering and Construction

Email: jeehee.lee@unlv.edu

Expertise

- Unstructured data analysis: natural language process (NLP), text mining, information retrieval, information extraction
- Physiological data sensing applications in smart building and smart city
- Interpretable machine learning

Dr. Jeehee Lee
Assistant Professor, Department of Civil and Environmental Engineering and Construction

Relevant Publications

- Lee, J. and Ham, Y. (2020). "Physiological sensing-driven personal thermal comfort modeling in consideration of occupant activity variations." *Building Research & Information* (Taylor & Francis).
- Lee, J., Ham, Y., Yi, J., and Son, J. (2020). "Effective risk positioning through automated identification of missing contract conditions from the contractor's perspective: based on the FIDIC contract cases". *Journal of Management in Engineering* (ASCE).
- Lee, J. and Ham, Y. (2020). "Investigating the effect of reflecting human activity information in physiological sensing- driven occupant thermal comfort modeling." ASCE Construction Research Congress, Tempe, AZ.
- Lee, J., Yi, J., and Son, J. (2018). "Development of an automatic-extraction model of poisonous clauses in international construction contracts using rule-based NLP." *Journal of Computing in Civil Engineering* (ASCE). 33(3).
- Lee, J. and Yi, J. (2017). "Predicting project's uncertainty risk in the bidding process by integrating unstructured text data and structured numerical data using text mining." *Applied Sciences* (MDPI). 7(11).
- Sung, Y., Lee, J., Yi, J., and Son, J. (2017). "Establishment of growth strategies for international construction firms by exploring diversification-related determinants and their effects." *Journal of Management in Engineering* (ASCE). 33(5).
- Lee, J., Yi, J., Son, J., and Jang, Y. (2017). "Pre-bid clarification for construction project risk identification using unstructured text data analysis." Proceedings of the joint conference on computing in construction (JC3), Heraklion, Greece.
- Lee, J., Yi, J., and Son, J. (2016). "Unstructured construction data analytics using R programming-focused on overseas construction adjudication cases." *Journal of the Architectural Institute of Korea Structure & Construction*. 32(5).
- Lee, J., Yi, J., and Son, J. (2016). "Text analytics on construction tender documents for project-oriented risk mining." Proceedings of the CIB World Building Congress 2016, Tampere, Finland.
- Lee, J., Son, J., and Yi, J. (2015). "Multilevel project-oriented risk-mining approach for overseas construction project's preemptive action." Proceedings of the 32nd CIB W78 Conference, Eindhoven, The Netherlands.

Dr. Beiyu Lin

Assistant Professor, Department of Computer Science

Phone: (702) 895-4707

Email: beiyu.lin@unlv.edu

- Expertise
  - Data science
  - Big data analytics
  - Machine learning
  - Integrative analysis of multiple types of data
  - Adaptive systems

UNLV | HOWARD R. HUGHES
COLLEGE OF ENGINEERING

Dr. Beiyu Lin

Assistant Professor, Department of Computer Science

Relevant Publications

- Jia, X., Lin, B., Zwart, J., Sadler, J., Appling, A., Oliver, S. and Read, J., 2021. "Graph-based Reinforcement Learning for Active Learning in Real Time: An Application in Modeling River Networks." 2021 SIAM International Conference on Data Mining (pp. 621-629). Society for Industrial and Applied Mathematics.

- Ek, S., Curci, M., Yang, X., Lin, B., Liu, P. and Xu, H., 2021. "Sentiment Analysis of Long-term Social Data during the COVID-19 Pandemic." Int'l Conf on Internet Computing and Internet of Things, 2021. Harrington A, Vo V, Tillett RL, Chang CL, Gerrity D, Oh EC. Urban monitoring of antimicrobial resistance during a COVID-19 surge through the analysis of sewage. bioRxiv. Preprint. 2021.

- Lin, B. and Cook, D.J., 2020. "Analyzing Sensor-Based Individual and Population Behavior Patterns via Inverse Reinforcement Learning." *Sensors*, 2020(18), p.5207.

- Lin, B., Cook, D.J. and Schmitter-Edgecombe, M., 2020. "Using continuous sensor data to formalize a model of in-home activity patterns." *Journal of Ambient Intelligence and Smart Environments*, 12(3).

- Huangfu, Y., Lima, N.M., O'Keeffe, P.T., Kirk, W.M., Lamb, B.K., Pressley, S.N., Lin, B., Cook, D.J., Walden, V.P. and Jobson, B.T. "Diel variation of formaldehyde levels and other VOCs in homes driven by temperature dependent infiltration and emission rates." *Building and Environment*, 2019, 159, p.106153.

- Ghods, A., Caffrey, K., Lin, B., Fraga, K., Fritz, R., Schmitter-Edgecombe, M., Hundhausen, C. and Cook, D.J. "Iterative design of visual analytics for a clinician-in-the-loop smart home." *IEEE journal of biomedical and health informatics*, 2018, 23(4), pp.1742-1748.

- Lin, B., Huangfu, Y., Lima, N., Jobson, B., Kirk, M., O'Keeffe, P., Pressley, S.N., Walden, V., Lamb, B. and Cook, D.J. "Analyzing the relationship between human behavior and indoor air quality." *Journal of Sensor and Actuator Networks*, 2017, 6(3), p.13.

Dr. Fatma Nasoz

Associate Professor, Department of Computer Science

Director of Data Science, The Lincy Institute

Affiliate Faculty, Nevada Institute of Personalized Medicine (NIPM)

Phone: (702) 895-0097

Email: fatma.nasoz@unlv.edu

- Expertise
    - Machine learning and deep learning
    - Data analysis and visualization
    - Human-computer interaction
    - Cloud computing

UNLV | HOWARD R. HUGHES
COLLEGE OF ENGINEERING

Dr. Fatma Nasoz

Associate Professor, Department of Computer Science

Affiliate Faculty, Nevada Institute of Personalized Medicine (NIPM)

Relevant Publications

- Z Fitzhugh, MR Schiller Ph D, F Nasoz (2021). "Predicting Variant Pathogenicity with Machine Learning." *UNLV Undergraduate Research Symposium Posters.* 46.
- Wu, Q., Nasoz, F., Jung, J., Bhattarai, B., Han, M.V., Greenes, R.A., and Saag, K.G. (2021). "Machine learning approaches for the prediction of bone mineral density by using genomic and phenotypic data of 5130 older men." *Scientific Reports.* 11(1):4482, doi: 10.1038/s41598-021-83828-3
- Wu, Q., Nasoz F., Jung, J., Bhattarai, B, Han, MV. (2020). "Machine learning approaches for fracture risk assessment: a comparative analysis of genomic and phenotypic data in 5,130 older men." *Calcified Tissue International.* 107 (4): 353–361, doi: 10.1007/s00223-020-00734-y
- Singh, S. and Nasoz, F. (2020). "Facial expression recognition with convolutional neural networks." *10th Annual Computing and Communication Workshop and Conference (CCWC),* Las Vegas, NV, USA, pp. 0324-0328, doi: 10.1109/CCWC47524.2020.9031283.
- Shrestha, N. and Nasoz, F. (2019). "Deep Learning Sentiment Analysis of Amazon.com Reviews and Ratings." *International Journal on Soft Computing, Artificial Intelligence, and Applications*, Vol. 8(1): 1-15.
- Nasoz F. and Shrestha C. (2017) "A Web-Based User Interface for Machine Learning Analysis". In: Yamamoto S. (eds) Human Interface and the Management of Information: Supporting Learning, Decision-Making and Collaboration. HIMI 2017. *Lecture Notes in Computer Science*, vol. 10274. Springer, Cham. doi: 10.1007/978-3-319-58524-6_35

Dr. Haroon Stephen

Professor, Department of Civil and Environmental Engineering

Director, GIS and Remote Sensing Core Lab

Phone: (702) 774-1463

Email: haroon.stephen@unlv.edu

- Expertise
  - Sensors aboard UAS for remote sensing: testing and applications
  - UAS and GIS integration
  - Applications of UAS to resource mapping, monitoring, and management
  - Data fusion and data mining of multi-scale sources: UAS, aerial, and spaceborne sensors

(a) Google Earth 3D Buildings
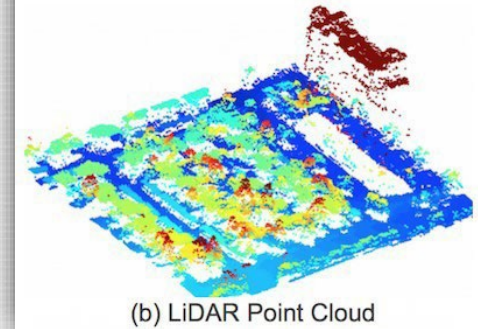
(b) LiDAR Point Cloud

Example of 3D view of buildings and land cover

Dr. Haroon Stephen

Professor, Department of Civil and Environmental Engineering and Construction

Relevant Publications

- Abedin, S. J. H., & Stephen, H. (2019). GIS Framework for Spatiotemporal Mapping of Urban Flooding. Geosciences, 9(2), 77.

- Stephen, H. (2018). Trend Analysis of Las Vegas Land Cover and Temperature Using Remote Sensing. Land, 7(4), 135.

- Lachniet, M. S., Lawson, D. E., Stephen, H., Sloat, A. R., & Patterson, W. P. (2016). Isoscapes of $\delta18O$ and $\delta2H$ reveal climatic forcings on Alaska and Yukon precipitation. Water Resources Research, 52(8), 6575-6586.

- Black, A., & Stephen, H. (2014). Relating temperature trends to the normalized difference vegetation index in Las Vegas. GIScience & Remote Sensing, 51(4), 468-482. doi: 10.1080/15481603.2014.940695.

- Acharya, A., Piechota, T. C., Stephen, H., & Tootle, G. (2011). Modeled streamflow response under cloud seeding in the North Platte River watershed. Journal of Hydrology, 409(1), 305-314, doi:10.1016/j.jhydrol.2011.08.027.

- Puri, S., Stephen, H., & Ahmad, S. (2011). Relating TRMM Precipitation Radar backscatter to water stage in wetlands. Journal of Hydrology, 401(3), 240-249, doi:10.1016/j.jhydrol.2011.02.026.

- Ahmad, S., Kalra, A., & Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. Advances in Water Resources, 33(1), 69-80, doi:10.1016/j.advwatres.2009.10.008.

UNLV | HOWARD R. HUGHES COLLEGE OF ENGINEERING
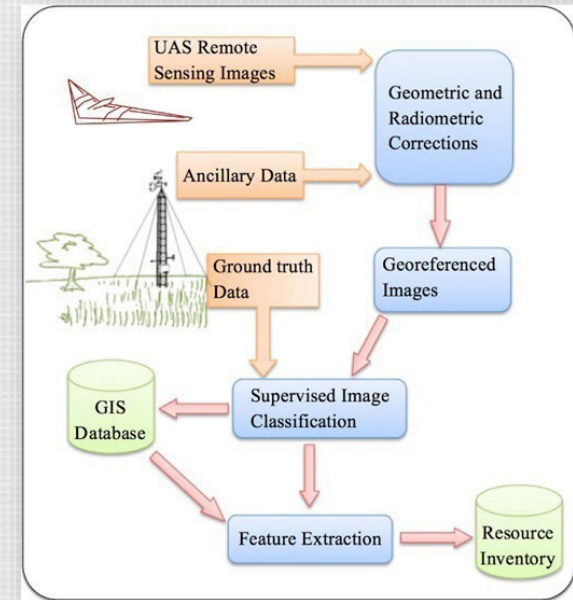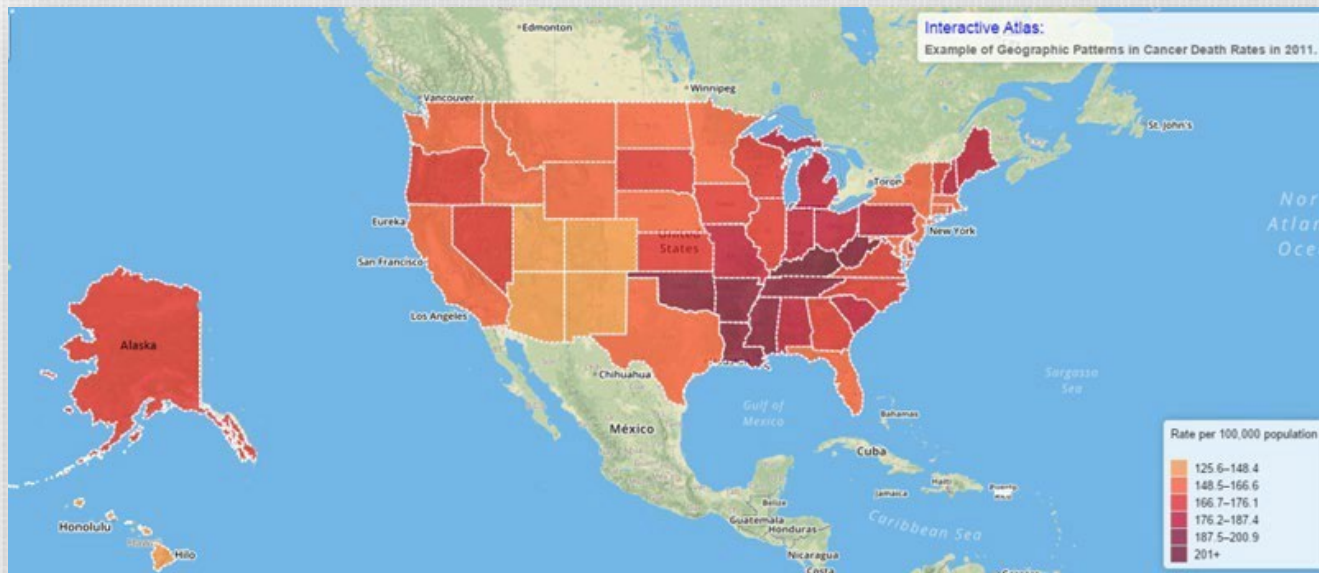
Dr. Kazem Taghva

Chair and Professor, Department of Computer Science

Phone: (702) 895-0873

Email: kazem.taghva@unlv.edu

- Expertise
    - Databases
    - Machine learning
    - Information retrieval

Dr. Kazem Taghva

Chair and Professor, Department of Computer Engineering

Relevant Publications

- S Choi, P Puranik, B Dahal, K Taghva. "How to generate data for acronym detection and expansion". *Advances in Computational Intelligence* 2 (2), 1-8, 2022.

- J Cummings, G Lee, K Zhong, J Fonseca, K Taghva. "Alzheimer's disease drug development pipeline: 2021" *Alzheimer's & Dementia: Translational Research & Clinical Interventions vol.* 7 issue 1, e12179, 2021.

- Pouyan Nahed; Mina E. Zahed Nojoo Kambar; Jorge Ramón Fonseca Cacho, Garam Lee; Jeffrey Cummings; Kazem Taghva. "Clinical Text Classification of Alzheimer's Drugs' Mechanism of Action using BERT.", Proceedings of Sixth International Congress on Information and Communication Technology pp 513–521, 2021.

- Cacho, Jorge Ramón Fonseca, Ben Cisneros, and Kazem Taghva. "Building a Wikipedia N-GRAM Corpus." *Proceedings of SAI Intelligent Systems Conference.* Springer, Cham, 2020.

- Cacho, Jorge Ramón Fonseca, and Kazem Taghva. "Aligning ground truth text with OCR degraded text." *Intelligent Computing-Proceedings of the Computing Conference.* Springer, Cham, 2019.

- Fonseca Cacho J.R., Taghva K. "Reproducible Research in Document Analysis and Recognition." *Advances in Intelligent Systems and Computing*, vol. 738, pp. 389-395, 2018, Springer.

- Kazem Taghva, "Name identification and extraction with formal concept analysis." *International Journal of Machine Learning and Cybernetics*, 2017, Volume 8, Number 1, Page 171.

- Bozorgi, Mandana, Taghva, Kazem, and Singh, Ashok, "Cancer Survivability with Logistic Regression", Proceedings of Computing Conference 2017, 416- 420.

- Taghva, Kazem. "Name identification and extraction with formal concept analysis." *International Journal of Machine Learning and Cybernetics,* 8.1, 171-178 (2017).

- Kazem Taghva, "Contextual Analysis for Middle Eastern Languages with Hidden Markov Models," *International Journal on Natural Language Computing*, vol. 4, No. 4, August 2015, pp. 1-11.

UNLV | HOWARD R. HUGHES COLLEGE OF ENGINEERING