# YASSER EL-MANZALAWY

Henry Hood Center for Health Research, 100 N. Academy Ave, Danville, PA 17822 | yelmanzalawi@geisinger.edu

## RESEARCH INTERESTS

My primary research methodology integrates domain knowledge, statistical methods, network science, and machine learning to develop predictive models for improving our understanding of the underlying mechanisms controlling different types of macromolecular interactions and complex diseases. My current research aims at leveraging EHR data and cutting-edge data science approaches to better understand the pathophysiology, treatment responses, clinical trajectories, and outcomes of critical diseases.

## EDUCATION

Iowa State University, Ames, IA, USA
**Ph.D. in Computer Science**                                                2004-2008
Thesis: "Machine Learning Approaches for Epitope Prediction"
Advisor: Professor Vasant Honavar
Honors: Research Excellence Award

Al-Azhar University, Cairo, Egypt
**M.S. in Systems and Computer Engineering**                                 1998-2000
Thesis: "Asynchronous Transfer Mode Systems Design"
Advisor: Professor Hani Harb

Al-Azhar University, Cairo, Egypt
**B.S. in Systems and Computer Engineering**                                 1990-1995
Honors: Excellence with Honor Degree

## HONORS AND AWARDS

Outstanding reviewer – Computational Biology and Chemistry, Elsevier      2015
Recognized reviewer – Computational Biology and Chemistry, Elsevier       2013
Iowa State University Research Excellence Award                            2008
Research assistantship funded by the ISU Graduate College                 2008
Egyptian Government Ph.D. Fellowship                                       2004 – 2008

## GRANTS

**Integrated trajectory data and machine learning approaches for precision obesity medicine.**
Anne Justice (PI), Yasser El-Manzalawy (Co-PI). PA DOH, $226,838 (2021-2025).

**Validation and Applications of EMR-derived Passive Digital Markers (PDMs) for Predicting Dementia and Alzheimer's Disease in Geisinger Population**. Yasser El-Manzalawy (PI). Eisai, $188,228 (2021-2022).

**Machine Learning Approach to Venous Thromboembolism Prediction in Newly Diagnosed Patients with Cancer.**
Benjamin Andrick (PI), Yasser El-Manzalawy (Co-PI). HOPA Research Fund Award, $100,000 (2021-2022).

**EAGER: Towards a Computational Infrastructure for Analysis of Sensitive Data**.
Vasant Honavar (PI), Yasser El-Manzalawy (Co-PI), John Yen (Co-PI). National Science Foundation, $231,578 (September 2015- August 2018).

## PUBLICATIONS

### Refereed Journal and Conference Papers

1. Abbas, M, Soto, P, Ramalingam, L, **El-Manzalawy, Y**, Bensmail, H. and Moustaid-Moussa, N (2022) Sex Differences in Fish Oil and Olanzapine Effects on Gut Microbiota in Diet-Induced Obese Mice. Nutrients, 14(2), p.349.
2. Alruily, M, Ezz, M, Mostafa, M, Yanes, N, Abbas, M and **El-Manzalawy, Y** (2022) Prediction of covid-19 transmission in the united states using google search trends. Computers, Materials and Continua, 71(1), pp.1751-1768.
3. **El-Manzalawy, Y**, Abbas, M, Hoaglund, I, Cerna, A.U, Morland, T.B, Haggerty, C.M, Hall, E.S, Fornwalt, B.K (2021) OASIS+: leveraging machine learning to improve the prognostic accuracy of OASIS severity score for predicting in-hospital mortality. BMC medical informatics and decision making 21.1 (2021): 1-13.
4. Abbas, M, Morland, T.B, Hall, E.S, **El-Manzalawy, Y** (2021). Associations Between Google Search Trends for Symptoms and COVID-19 Confirmed and Death Cases in the United States. International journal of environmental research and public health 18.9 (2021): 4560.
5. Matta, J, Dobrino D, Howard S, Yeboah, D, Kopel J, **El-Manzalawy, Y**, Obafemi-ajayi, Y., 2021. A PheWAS Model of Autism Spectrum Disorder. 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 2110-2114)
6. Matta, J, Dobrino, D, Howard, S, Yeboah, D, Kopel, J, **El-Manzalawy, Y** and Obafemi-Ajayi, T (2021) November. A PheWAS Model of Autism Spectrum Disorder. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 2110-2114). IEEE.
7. Abbas M, **El-Manzalawy Y** (2020) Machine learning based refined differential gene expression analysis of pediatric sepsis." BMC medical genomics 13, no. 1: 1-10.
8. Abbas M, Matta J, Le T, Bensmail H, Obafemi-Ajayi T, Honavar V, and **El-Manzalawy Y** (2019) Biomarker discovery in inflammatory bowel diseases using network-based feature selection. PloS one 14, no. 11: e0225382.
9. Jung Y, **El-Manzalawy Y**, Dobbs D, Honavar V (2019) Partner-specific prediction of RNA-binding residues in proteins: A critical assessment. Proteins: Structure, Function, and Bioinformatics, 87(3), pp.198-211.
10. Abbas M, Le T, Bensmail H, Honavar V, **El-Manzalawy Y** (2018) Microbiomarkers Discovery in Inflammatory Bowel Diseases using Network-Based Feature Selection. Proceedings of the 9th ACM Conference on Bioinformatics, Computational Biology and Health Informatics (pp. 172-177).
11. **El-Manzalawy Y** (2018) CCA based multi-view feature selection for multi-omics data integration. Proceedings of IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) (pp. 1-8).
12. **El-Manzalawy Y**, Hsieh T-Y, Shivakumar M, Kim D, Honavar V (2018) Min-Redundancy and MaxRelevance Multi-view Feature Selection for Predicting Ovarian Cancer Survival using Multi-omics Data. BMC Medical Genomics 11 (Suppl 3) :71.
13. Hsieh T-Y, **El-Manzalawy Y**, Sun Y, Honavar V (2018) Compositional Stochastic Average Gradient for Machine Learning and Related Applications. Proceedings of the 19th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL) (pp. 740-752).
14. Khademi A, **El-Manzalawy Y**, Buxton O, Honavar V (2018) Toward Personalized Actigraphy Sleep/Wake Classification. IEEE International Conference on Biomedical and Health Informatics (BHI) (pp 414-417).

15. Abbas M, **El-Manzalawy Y** (2017) Predictive and Comparative Network Analysis of the Gut Microbiota in Type 2 Diabetes. Proceedings of the 8th ACM Conference on Bioinformatics, Computational Biology and Health Informatics: ACM. pp. 313-320.

16. **El-Manzalawy Y**, Hsieh T-Y, Shivakumar M, Kim D, Honavar V (2017) Min-Redundancy and Max-Relevance Multi-view Feature Selection for Predicting Ovarian Cancer Survival using Multi-omics Data. Presented at the 7th Annual Translational Bioinformatics Conference.

17. **El-Manzalawy Y**, Buxton O, Honavar V (2017) Sleep/wake state prediction and sleep parameter estimation using unsupervised classification via clustering. IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2017). pp. 718-723.

18. **El-Manzalawy Y**, Munoz EE, Lindner SE, Honavar V (2016) PlasmoSEP: Predicting surface-exposed proteins on the malaria parasite using semi-supervised self-training and expert-annotated data. Proteomics 16: 2967-2976.

19. **El-Manzalawy Y**, Abbas M, Malluhi Q, Honavar V (2016) FastRNABindR: Fast and Accurate Prediction of Protein-RNA Interface Residues. PloS one 11: e0158445.

20. Abbas M, Mohie-Eldin M, **El-Manzalawy Y** (2015) Assessing the effects of data selection and representation on the development of reliable E. coli sigma 70 promoter region predictors. PloS One 10: e0119721.

21. Xue LC, Jordan RA, **El-Manzalawy Y**, Dobbs D, Honavar V (2014) DockRank: ranking docked conformations using partner-specific sequence homology-based protein interface prediction. Proteins 82: 250-267.

22. Walia RR, Xue LC, Wilkins K, **El-Manzalawy Y**, Dobbs D, Honavar V (2014) RNABindRPlus: a predictor that combines machine learning and sequence homology-based methods to improve the reliability of predicted RNA-binding residues in proteins. PLoS One 9: e97725.

23. **El-Manzalawy Y**, Dobbs D, Honavar V (2012) Predicting protective bacterial antigens using random forest classifiers. Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB). pp. 426-433.

24. Walia RR, Caragea C, Lewis BA, Towfic F, Terribilini M, **El-Manzalawy Y**, Dobbs D, Honavar V (2012) Protein-RNA interface residue prediction using machine learning: an assessment of the state of the art. BMC Bioinformatics 13: 89.

25. Jordan RA, **El-Manzalawy Y**, Dobbs D, Honavar V (2012) Predicting protein-protein interface residues using local surface structural similarity. BMC Bioinformatics 13: 41.

26. **El-Manzalawy Y**, Dobbs D, Honavar V (2012) Predicting protective bacterial antigens using random forest classifiers. Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB). pp. 426-433.

27. **El-Manzalawy Y**, Dobbs D, Honavar V (2011) Predicting MHC-II binding affinity using multiple instance regression. IEEE/ACM Trans Comput Biol Bioinform 8: 1067-1079.

28. Xue LC, Walia R, **El-Manzalawy Y**, Dobbs D, Honavar V (2011) Improving protein-RNA interface prediction by combining sequence homology based method with a naive Bayes classifier: preliminary results. Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine. pp. 556-558.

29. Xue LC, Jordan RA, **El-Manzalawy Y**, Dobbs D, Honavar V (2011) Ranking docked models of protein-protein complexes using predicted partner-specific protein-protein interfaces: a preliminary study. Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine. pp. 441-445.

30. **El-Manzalawy Y**, Honavar V (2010) Recent advances in B-cell epitope prediction methods. Immunome Res 6 Suppl 2: S2.

31. **El-Manzalawy Y**, Honavar V (2010) A framework for developing epitope prediction tools. Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology. pp. 660-662.

32. **El-Manzalawy Y**, Honavar V (2009) MICCLLR: Multiple-Instance Learning Using Class Conditional Log Likelihood Ratio. Discovery Science: Springer. pp. 80-91.
33. **El-Manzalawy Y**, Dobbs D, Honavar V (2008) On evaluating MHC-II binding peptide prediction methods. PLoS One 3: e3268.
34. **El-Manzalawy Y**, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. J Mol Recognit 21: 243-255.
35. **El-Manzalawy Y**, Dobbs D, Honavar V (2008) Predicting flexible length linear B-cell epitopes. Comput Syst Bioinformatics Conf. pp. 121-132.
36. **El-Manzalawy Y**, Dobbs D, Honavar V (2008) Predicting protective linear B-cell epitopes using evolutionary information. IEEE International Conference on Bioinformatics and Biomedicine (BIBM'08). pp. 289-292.

**Invited Articles/Book Chapters**

37. **El-Manzalawy Y**, Dobbs D, Honavar V (2017) In silico prediction of linear B-cell epitopes on proteins. In: Zhou Y, Faraggi E, Kloczkowski A, Yang Y, editors. Methods Mol Biol: Springer.
38. Walia R, **El-Manzalawy Y**, Dobbs D, Honavar V (2017) Sequence-based Prediction of RNA-binding Residues in Proteins. In: Zhou Y, Faraggi E, Kloczkowski A, Yang Y, editors. Methods Mol Biol: Springer.
39. **El-Manzalawy Y**, Honavar V (2014) Building classifier ensembles for B-cell epitope prediction. Methods Mol Biol. pp. 285-294.
40. **El-Manzalawy Y**, Honavar V (2013) B Cell Epitope Prediction. Encyclopedia of Systems Biology: Springer. pp. 63-67.
41. **El-Manzalawy Y**, Honavar V (2013) Major Histocompatibility Complex (MHC) Binder Prediction. Encyclopedia of Systems Biology: Springer. pp. 1162-1166.

## PROFESSIONAL EXPERIENCE

- Assistant Professor                                                           2019 – present
  Translational Data Science and Informatics
    Geisinger Clinic
- Assistant Research Professor                                          2018 – 2019
  College of Information Sciences and Technology
    Pennsylvania State University
- Research Associate                                                          2014 – 2017
    Center for Big Data Analytics and Discovery Informatics
    Pennsylvania State University
- Assistant Professor                                                         2009 – 2010, 2012-2014
    Systems and Computer Engineering
    Al-Azhar University, Cairo, Egypt
- Research Associate                                                         2010 –2012
    Center for Computational Intelligence, Learning, & Discovery
    Iowa State University
- Postdoctoral Research Associate                                   2008 –2009
    Center for Computational Intelligence, Learning, & Discovery
    Iowa State University
- Research Assistant                                                          2004 –2008
    Artificial Intelligent Research Lab
    Iowa State University
- Freelance Technical Writer                                            2004

[www.developer.com](www.developer.com)
Published tutorials on java language, aspect-oriented programming, and application frameworks
(Articles are available at [http://www.developer.com/author/Yasser-EL-Manzalawy-81410.htm](http://www.developer.com/author/Yasser-EL-Manzalawy-81410.htm)).

- Teaching and Lecturer Assistant                                                         1997 –2004
  Systems and Computer Engineering
  Al-Azhar University, Cairo, Egypt
- Programming Language Instructor                                                     1997 –2004
  IBM authorized education centers, Computek Integrated Systems, TRI-TECH
  C/C++, Visula C++, Visual Basic, Java

## TEACHING EXPERIENCE

- Data Integration & Fusion, College of Information Sciences & Technology, PSU, Fall 2018, Fall 2019
- Compilers – Systems & Computer Engineering, Al-Azhar University, Fall 2009, Fall 2012, Fall 2013
- Digital Logic Design – Systems & Computer Engineering, Al-Azhar University, Spring 2009
- Systems Analysis and Design – Systems & Computer Engineering, Al-Azhar University, Spring 2012
- Data Structures – Higher Technical Institute, Egypt, Fall 2013
- Introduction to Probability Theory – Systems & Computer Engineering, Al-Azhar University, Spring 2014
- Bioinformatics – Systems & Computer Engineering, Al-Azhar University, Fall 2012

### Invited Lectures

- Machine learning in bioinformatics (3 lectures), Introduction to Bioinformatics (BCB 544), Iowa State University, Spring 2011
- Evaluating classifiers, Principles of Artificial Intelligence (IST 597F), Penn State University, Fall 2014, Fall 2015, Fall 2016
- Introduction to WEKA, Principles of Machine Learning (IST 597K), Penn State University, Spring 2015, Spring 2016
- Multiple-Instance learning, Principles of Machine Learning (IST 597K), Penn State University, Spring 2015, Spring 2016
- Machine learning for Big data (2 lectures), Principles of Machine Learning (IST 597K), Penn State University, Spring 2015, Spring 2016

## PROFESSIONAL ACTIVITIES

### Organizing Committee

- Workshop on Microbiomics, Metagenomics, and Metabolomics, ACM-BCB 2019
- Immunoinformatics and Computational Immunology Workshops held in conjunction with ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB) (Years: 2010, 2011, 2012, 2013)

### Program Committee Member

- 33rd Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-21)
- 8th International Conference on Bioinformatics Research and Applications (ICBRA 2021)
- 6th IEEE Conference on Data Science and Machine Learning Applications (CDMA2020)
- IEEE International Conference on Machine Learning and Applications (ICMLA 2018)
- Midwest Artificial Intelligence and Cognitive Science Conference (MAICS)
      (Years: 2012, 2013, 2016, 2017)
- Workshop on Informatics Applications in Therapeutics, in conjunction with IEEE BIBM2011
- IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2009)

**Journal Reviews**

- Bioinformatics
- BMC Bioinformatics
- PLoS Computational Biology
- PLoS ONE
- Amino Acids
- Nucleic Acids Research
- Computational Biology and Chemistry
- Molecular Immunology
- PeerJ
- Journal of Clinical Medicine
- Immunome Research
- Journal of Immunological Methods
- Journal of Biomedicine and Biotechnology
- Journal of Theoretical Biology
- IEEE TCBB
- Cognitive Systems Research
- Electronic Commerce Research and Applications
- Sleep Health
- Biomedical Signal Processing and Control
- Journal of Machine Learning and Knowledge Extraction

## PUBLICALLY AVAILABLE SOFTWARE

- **BCPREDS** (http://ailab.ist.psu.edu/bcpred/): A resource for linear B-cell epitope prediction tools. Current implementation supports three different methods for predicting linear B-cell epitopes and the benchmark datasets used to build the classifiers.
- **MHCMIR** (http://ailab.ist.psu.edu/mhcmir/): A server for predicting the binding affinity of MHC-II peptides using multiple instance regression. The current version supports 13 HLA-DR alleles and three mouse H2-IA alleles.
- **Epitopes Toolkit (EpiT)** (http://ailab.ist.psu.edu/epit/): A framework for developing epitope prediction tools. An EpiT developer can distribute his predictor as a serialized Java object (model file). This allows other researchers to use the developed predictor on their own machines, rebuild the predictor on other data sets, or combine the predictor with other predictors to obtain a customized hybrid or consensus predictor.
- **WEKA LibSVM (WLSVM)** (https://weka.wikispaces.com/LibSVM):  A wrapper class for libsvm library by Chih-Chung Chang and Chih-Jen Lin. WLSVM is part of WEKA since version 3.5.2.
- **Predictor of Interface Residues using Structural Elements (PrISE)** (http://ailab1.ist.psu.edu/prise/index.py): A novel predictor of protein-protein interface residues based on local structural similarity.
- **DockRank** (http://ailab1.ist.psu.edu/DockRank/): A tool to rank protein-protein docked models based on predicted partner-specific protein interface residues.
- **RNABindR v2** (http://ailab1.ist.psu.edu/RNABindR/): An improved version of RNABindR, a method for predicting protein-RNA interface residues in proteins, using sequence alignment profiles and Support Vector Machine (SVM) classification.
- **RNABindRPlus** (http://ailab.ist.psu.edu/RNABindRPlus/): A server for predicting protein-RNA interface residues using a hybrid method that combines RNABindR v2 and a sequence homology based predictor (HomPRI).
- **FastRNABindR** (http://ailab.ist.psu.edu/FastRNABindR/):  A tool for accurate and fast prediction of protein-RNA interface residues. To the best of our knowledge, this is the first genome-scale protein-RNA interface prediction server.
- **Genome Annotate Toolkit (Gennotate)** (http://ailab.ist.psu.edu/Gennotate/): A framework for sharing data representations, predictors, and machine learning algorithms for a broad range of gene structure prediction tasks.
- **Python Libraries under-development**
  - **D-Fuse**: Python library and framework for data fusion and multi-view feature selection.
  - **Network Similarity Library (NSL)**: Python library and evaluation scripts for static/temporal graph and vertex similarity.

## MENTORED STUDENTS

- **Rafael Jordan** (Ph.D.), Protein-Protein Interface Residue Prediction, Iowa State University
- **Rasna Walia** (Ph.D.), Protein-RNA Interaction Prediction, Iowa State University
- **Li Xue** (Ph.D.), Docking Scoring Functions, Iowa State University
- **Usha Muppirala** (Ph.D.), Data Driven Docking, Iowa State University
- **Katie Wilkins** (Under graduate), Protein-RNA Interaction Prediction, Iowa State University
- **Krishnakumar Sridharan** (Ph.D.), Core Promoter Prediction, Iowa State University
- **Khaled Hassan** (M.S.), Virulent Protein Prediction, Al-Azhar University
- **Mostafa Abbas** (Ph.D.), E. coli Promoter Region Prediction, Al-Azhar University
- **Thanh Le** (Ph.D.), Protein-DNA Interaction Prediction, Sentiment Classification, Penn State University
- **Yong Jung** (Ph.D.), Protein-RNA Interaction Prediction, Penn State University
- **Shlomit Gur** (Ph.D.), Feature Selection, Sentiment Classification, Penn State University
- **Aria Khademi** (Ph.D.), m-Health, Metagenomics, Penn State University
- **Tsung Yu Hsieh** (Ph.D.), Multi-view Learning, Compositional Optimization, Penn State University
- **YiWei Sun** (Ph.D.), Deep Learning for Biological Sequence Classification, Penn State University
- **Ben Chen** (M.Sc.), Dynamic Network Analysis, Penn State University
- **Morteza Jaderyan** (Ph.D.), Hashtag Recommendation Systems, Penn State University
- **Hoang Nguyen** (Ph.D.) Cardiac Data Mining, University of Missouri at Kansas