Doctoral Dissertation and Research Experience

Masters Student **01/2016-06/2016**

Mentored by Michael Thompson and Jennifer Fowler at Argonne National Laboratory

Project: A Mobile Encryption Gateway (MEG) for Email Encryption

Starting January of 2016 I took a course aimed at improving graduate experiences in security research. Course directors paired me with a team from Argonne National Laboratory working on improving email security. To ensure the security of email we must ensure that the email cannot be read by malicious actors while an email is in transit to its destination, or is in storage waiting to be read. Another important attribute for any secure application is that it must be usable by average computer users. We desired to create a secure, and highly usable email encryption application. For this project, I built a comprehensive computing architecture, including a mobile application, a computing server, and an email plugin to encrypt emails so that their contents would be safe from hackers. We published one paper related to this work in an online conference held by the organizers of the course. Through this work I learned how to build distributed systems that could secure sensitive data and how to write publication worthy material. These experiences have served me well in my continued work in informatics especially given how research in medical informatics often uses sensitive patient data.

- **Rehm GB**, Thompson M, Busenius B, Fowler J. (2016, Sept.) "Mobile Encryption Gateway (MEG) for Email Encryption; Proceedings of the 2016 Information Security Research and Education (INSuRE) Conference (INSuRECon-16). [On-line]. Available: https://sites.google.com/a/uah.edu/insurecon16/proceedings [November 22, 2017].

Masters Student **10/2015-07/2016**

Mentored by Jason Adams and Nicholas Anderson at UC Davis

Project: Bedside to the Cloud and Back: Real-Time Data Analytics From Critical Care Instrumentation

Lack of access to high-frequency patient data has limited the use of these data for research and decision support. Also of significance, is the necessity to create high performance algorithms designed to detect anomalies in waveform data. We used mechanical ventilation and the collection and analysis of ventilator waveform data (VWD) as a use case of this problem. I designed a system for performing automated collection, aggregation, and analysis of data collected from mechanical ventilators within the University California (UC) Davis Medical Center. As of time of writing, our data collection platform has allowed us to gather a dataset of VWD for 502 patients.

After collection of waveform data we often need to analyze in order to perform anomaly detection. But first, we need to be able to create a ground truth dataset of waveforms linked with annotation of either anomalous observations or normal ones. So, we created a system for viewing VWD and performing annotations on it so we could use it for later algorithm development and validation. I further refined this system so that future informaticists can use it as an example for creating customized, collaborative, low cost programs that can output ground truth validation data. Through this work I learned how to build responsive web platforms, and found methods for reducing inter-rater disagreement on classifications.

Using the previously discussed components, we created high performance heuristic algorithms for detecting two types of patient-ventilator asynchrony (PVA) that could lead to dynamic hyperinflation. I assisted with improving the computational speed of these algorithms and set up software testing infrastructure so that we could more rapidly improve them. Computationally I learned how to perform performance testing and optimization of a software suite, a skill I have put to continued use on current projects. In medicine I learned many of the fundamental principles of mechanical ventilation including patient synchrony, and why patient ventilator asynchrony occurs. I also learned about various ventilator modes, the dangers of volutrauma, barotrauma, and atelectrauma for the lungs. I hope to take these lessons into my new work detecting ARDS and change in ARDS severity.

Finally, the informatics infrastructure we built allowed us to extract large amounts of VWD, but we were unable to obtain any ventilator configuration settings in the data. Instead, we had to rely on computationally deriving ventilator settings from VWD. Informative settings include positive end expiratory pressure (PEEP), inspiratory time (I-time), and importantly ventilator mode. To detect the type of ventilator mode used we built a high performance machine learning classifier that could discriminate between 4 types of ventilator modes: volume control, pressure control, pressure support,

and continuous positive airway pressure (CPAP). This project helped improve my skills with machine learning, model creation, model debugging, and also helped me to learn the more detailed characteristics of various ventilator modes. I am currently in process of preparing a manuscript for this work.

- **Rehm GB**, Kuhn BK, Delplanque JP, Guo E, Lieng M, Nguyen, J., Anderson, N., Adams, J. Development of a Research Oriented System for Collecting Mechanical Waveform Data; *Journal of the American Medical Informatics Association*. 2017 Oct 28
- Adams JY, Lieng M, Kuhn BK, **Rehm GB**, Guo E, Taylor S, Delplanque JP, Anderson N. Development and Validation of a Multi-Algorithm Analytic Platform to Detect Off-Target Mechanical Ventilation; Scientific Reports. 2017 Nov 3;7(1):14980
- **Rehm GB**, Kuhn BK, Lieng M, Guo E, Delplanque JP, Anderson NR, Adams JY. Analytic Platform for Rapid Visualization, Description, and Annotation of Continuous Clinical Waveform Data. (manuscript in preparation)
- **Rehm GB**, Kuhn BK, Delplanque JP, Anderson NR, Adams JY. A Machine Learned Classifier for Determining Ventilator Mode From Ventilator Waveform Data. (manuscript in preparation)

Masters Student **07/2016-06/2017**

Mentored by Chen-Nee Chuah and Jason Adams at UC Davis

Project: Leveraging Big Data Analytics for Precision Medicine in Critical Care -- Development of Statistical & Machine Learning Methods for Preventing Acute Respiratory Failure Requiring Mechanical Ventilation

It has been hypothesized that PVA plays a role in worsening the severity of acute respiratory distress syndrome (ARDS). To provide evidence for this hypothesis, we aimed to utilize patient VWD collected by our lab to determine whether PVA could be linked to events such as clinical decompensation. Other goals were to advance our previous work by developing machine learning-based analytics to improve the accuracy of classifying off-target mechanical ventilation, and to create a phone application that could remotely alert clinicians if a deterioration in patient state was detected. We created a machine learning (ML) classifier that could detect two harmful types of PVA with sensitivity/specificity of over 0.94/0.95. We also created an analytic system able to create data linkages between electronic medical record (EMR) data and VWD. In addition I created a web server for the processing of real-time VWD so that the information could be sent to a mobile device. Through this project, I learned the fundamentals of performing research in machine learning, and the importance of being embedded in a clinical team that could offer rapid feedback on model development. I also learned how to more effectively create algorithms for linking disparate sources of medical informatics data. Work for predicting clinical decompensation and development of a phone application for visualizing ventilation data is currently being spearheaded by other students in the laboratory and I am assisting with the work. We anticipate publishing an additional 1 or 2 journal articles from the results of this research.

- **Rehm GB**, Han J, Kuhn BT, Delplanque JP, Anderson NR, Adams JY, Chuah CN. Development of a Research Oriented System for Collecting Mechanical Waveform Data; (manuscript under review with minor revisions)

ORISE Fellow **07/2017-Present**

Mentored by Jason Adams and Michael Johnson at UC Davis/Travis Air Force Base

Project: Resuscitation and Ventilation Asynchrony Monitor

When providing manual ventilation via bag valve mask it is vital to ensure that a patient is receiving necessary volumes of air to sustain healthy blood oxygenation levels. Frequently however, providers are unable to determine whether they are providing the appropriate volume of air due to a variety of factors including exhaustion or distraction. To assist providers in dispensing adequate manual ventilation, we designed a chamber that could record flow and pressure observations during inhalation and exhalation, and display it on an external device, and then send an alert if inadequate ventilation was being provided. We combined this functionality with algorithms developed to detect patient-ventilator asynchrony during automated ventilation to create an integrated hardware and software platform aimed at providing ventilation decision support in both manual and automated ventilation settings. For this project I learned embedded software development on Arduino and FreeRTOS devices and I improved my skills in the C++ programming language. I also improved my skills in coding Bluetooth capable communication platforms. For our work the University of California is in process of submitting a patent for our invention.

- Johnson MA, Adams JY, Delplanque JP, Koos J, **Rehm GB**; Resuscitation and Ventilation Asynchrony Monitor, U.S. Application No. 62/420,943. November, 2017

Computer Science PhD Student **09/2017-Present**

Mentored by Jason Adams and Nicholas Anderson at UC Davis

Project: **A Real-Time Computational System for Detecting ARDS Using Ventilator Waveform Data**

For my doctorate, I have hypothesized we can detect if a patient has ARDS by performing analysis of VWD and EMR data. To this end we have already built a machine learning model using retrospective patient VWD and existing patient diagnoses. We are interested in expanding this model so that it utilizes additional Berlin criteria like PaO2:FiO2 measurements and other unbiased information extracted from the EMR. To advance the state of the art in computer science, we intend to create a generalizable framework that scientists can use for predicting patient state transitions. To this end, we will compare the accuracy of Hidden Markov Models and deep learning in the use case of predicting worsening of the ARDS disease in intubated patients. As ubiquitous monitoring and data-driven diagnosis procedures become the norm in medicine, our framework may become invaluable to modelers for creating the future of clinical decision support and predictive analytic systems.

Training Goals and Objective

My career goal is to be a leading scientist in the field of computational and translational medicine. Specifically I would like to orient my research toward solving impactful analytic problems, and set myself at the forefront of applying computational and machine learning techniques to ICU and critical care environments. Upon completing my PhD, I eventually I envision myself as a professor at an academic medical center, where I will continue to perform multidisciplinary team science at the crossroads or medicine, computer science, and informatics. This career path will straddle multiple different disciplines and will require a diverse skillset. In particular my priorities for the F31 are: improving my professional network, broadening my fluency with analytic algorithms, becoming an expert informaticist, and a better leader and communicator. In cooperation with Dr. Jason Adams and Dr. Nicholas Anderson we have developed a training plan to facilitate these goals.

Expand network of clinical collaborators: Having been embedded in a clinical team for the past 2 years I understand that for my own work to be effective I need to have tight collaborative relationships with clinicians that take into account fundamental medicine surrounding a problem. As such, I hope to foster relationships through attending conferences with other forward-thinking physicians who are interested in utilizing translational data-driven science in their research. I will present the results of my work at local conferences like the UC Davis Lung Day symposium, and I will also present my research at national conferences such as the American Medical Informatics Association (AMIA) symposium and the annual meeting of the American Thoracic Society (ATS). I will also join several of the student networks of AMIA related to machine learning and clinical decision support systems. I also plan to join ATS subcommittees related to ML and data science.

Strengthen underlying skills in machine learning: I currently have a paper in the process of review that utilizes machine learning (ML) to solve the problem of detecting patient ventilator asynchrony. I have also completed a high-performance machine learning classifier that is able to classify ventilator mode from VWD. In order to perform these projects successfully, I gained skills in the fundamentals of machine learning research and developed a firm understanding of traditional ML algorithms such as random forests, neural networks, and support vector machines. I still have more work to do to improve my skills however, and the field of ML is progressing rapidly.

Specifically, I need to learn and become proficient in deep learning, a type of ML that is able to extract patterns and usable information from large amounts of unstructured data. Deep learning has been applied to image recognition in the past and shows promise for analyzing waveform data as well. Before I can apply deep learning to medical data I need to improve my skills in it. I will do this by working with Dr. Yong-Jae Lee, who is a member on my PhD committee and an expert in deep learning. I will also take a ML course taught by Dr. Lee starting January 2018 that focuses primarily on deep learning. For additional reinforcement, Aim 2 of my proposal has a sub aim for developing deep learning models. Aim 2 of my proposal will also improve my skills with Bayesian inference models (e.g. Hidden Markov Models) which have found wide use in predictive analytics. I will also continue seeking the assistance of Dr. Chuah of the Electrical and

Computer Engineering Department. With the help of Dr. Chuah I am improving my skills in machine learning, statistical learning, and linear algebra techniques.

Become an expert informaticist: I will improve my skills in informatics by taking courses offered to me by the Graduate Group in Informatics at UC Davis. In particular I will focus on improving my skills in study design and building clinical support decision systems. As part of my proposal I will also be collaborating with members of the clinical information technology group to create systems that can more rapidly gather EMR data. My co-mentor, Dr. Anderson is also Director of Informatics Research at UC Davis, and I will continue collaborating with him in my research.

Improve leadership skills: Interdisciplinary research cannot be conducted by a single researcher, so being a good scientist will require that I become a better leader as well. I currently am supervising the implementation of one of my project goals in conjunction with a highly motivated respiratory therapist. I will undertake more leadership training that UC Davis offers, including "Leaders for the Future" to help improve my skills in these areas. I plan to complete an individual development plan (IDP) yearly until graduation. In these I will lay out skills I wish to learn towards becoming a better leader, and I will convey them to my mentorship team to seek feedback on how to actualize my goals. I also currently provide mentorship for a computer science masters student in our lab, and I plan to mentor additional computer science students in the future. Through this training, I hope to gain the skills necessary to lead and train the next generation of scientists.

Improve communication and presentation skills: Presenting translational research to multiple audiences can be a difficult challenge. To effectively convey my research with the broader scientific community, I will work to improve my scientific writing and oral communication skills. In writing, I will continue preparing and submitting manuscripts to peer-reviewed journals, and I will assist with additional grant writing required by the lab. Dr. Brooks Kuhn and I also participate in an informal writing seminar that is held weekly for an hour. In addition, UC Davis holds biweekly writers retreats where writers are helped to improve their skills through help of an experienced graduate student, and I plan to attend these events when I must complete or edit my manuscripts. I will practice my oral communication at local and national conferences, at lab meetings, and at other university organized events. To learn how to better explain my research to others, I will compete in "Grad Slam," an event held at UC Davis where contestants compete to see who can best convey complex research topics in 3 minutes or less. This will help improve my communication skills, and this opportunity will give me the chance to network within the wider UC Davis community. As mentioned above, ATS and AMIA symposiums will also prove to be excellent locations to improve my oral scientific communication skills. I will also seek the mentorship of Dr. Anderson in this area, who is respected in the informatics community and may understand nuances of conveying translational research in effective ways.

Find preliminary data for later studies: Many computational systems are developed by researchers in the hospital, but are not tested and validated in multi-center studies. Instead, I'd like my ideas to be tested and put into practice. I'd like to continue my PhD work into my postdoctoral studies by implementing resulting models for ARDS detection in a clinical trial. The work that I will perform in professional networking will be critical to help me find physicians who are also interested in performing this trial. Furthermore, the research that I will perform on predicting worsening of ARDS severity has potential to be translated to many other areas of research that deal with patient state changes.

Activities Planned

Percentage time spent in each category

	12/2017-11/2018	12/2018-11/2019	12/2019-12/2020
Research	64	80	74
Coursework	16	10	6
Outside work with US Air Force	10	10	10
Qualifying Examination and Preparations	10		
Teaching Assistantship			10

The majority of my time will be dedicated to the proposed research project during the funding period of this fellowship. I will also engage in the following activities to supplement my training:

I still need to satisfy an additional course requirement for the Graduate Group in Computer Science, and I plan to take a course in machine learning, two additional courses in Informatics, and a course on responsible conduct in research (RCR). I need to complete my qualifying examination in Spring of 2018, and finally I plan to continue my work with Dr. Johnson as an ORISE Fellow working as a data scientist with the US Air Force, albeit on a reduced basis.

I plan to attend and present my work at conferences to stay informed about scientific progress in bioinformatics and biomedicine, and to develop my communication and presentation skills. I plan to present at these forums either in form of a poster presentation or in conference workshops. In particular, I will prioritize attending the Annual Symposium of American Medical Informatics Association in November and the UC Davis Lung Day retreat.

Course timeline

	Winter 2018	**Spring 2018**	**Fall 2018**	**Winter 2019**	**Spring 2019**	**Fall 2019**	**Winter 2020**
ECS 289G Special Topics in Deep learning	4 hr.						
ECS 220 Theory of Computation	4 hr.						
CLH 204 Responsible Conduct in Research			1 hr.	1 hr.	1 hr.		
MHI 207 Medical Decision Support			4 hr.				
EPI 205 Clinical Epidemiology and Study Design							4 hr.

*All courses in the Graduate Group in Computer Science at UC Davis are set on the quarter system.

Timeline of research

12/2017-11/2018	**12/2018-11/2019**	**12/2019-12/2020**
- Complete qualifying examination in April or May - Through coursework and assistance from Dr. Yong Jae Lee improve machine learning skills and develop skills in deep learning. - Finish enrollment of 30 additional patients for use in validating respiratory compliance and airway resistance calculations. Then incorporate this information into existing ARDS detection model. (Aim 1A) - Attend and present results, or preliminary results of sub aim 1A at either the Annual Symposium of American Medical Informatics Association or the annual meeting of the American Thoracic Society - Attend and present results of sub aim 1A at UC Davis Lung Day in June	- Improve skills in informatics by taking courses on clinical decision support systems - Gather relevant EMR data in collaboration with members of clinical IT and utilize it in machine learning model to detect ARDS. (Aim 1B) - Build Hidden Markov Models for predicting worsening of patient ARDS severity. (Aim 2A) - Attend and present results, of Aim 1 at the Annual Symposium of American Medical Informatics Association or the annual meeting of the American Thoracic Society - Attend and present results of Aim 1 at UC Davis Lung Day in June	- Improve skills in informatics by taking courses on study design methodology - Fulfill teaching assistantship requirement of Graduate Group in Computer Science - Build deep learning and recurrent neural network models for predicting worsening of patient ARDS severity (Aim 2B) - Attend and present results, or preliminary results of Aim 2 at the Annual Symposium of American Medical Informatics Association or the annual meeting of the American Thoracic Society - Attend and present results of Aim 2 at UC Davis Lung Day in June - Write and submit dissertation - Hold PhD exit seminar.