Fall Risk Classification Among Seniors

(Technical Paper)

Big Data and Privacy: Finding the Balance in Distrust and Progress

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Ziyang Guo
Spring, 2020

Technical Project Team Members
Leyao Li

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Introduction

Successfully identifying the relationship between the cause and the outcome of a disease is paramount for treating the disease. Medical professionals and researchers have used many computational tools to achieve this identification goal. Artificial neural network (ANN) is a powerful machine learning technique, well-suited for data analysis and modeling relationships between variables for best prediction of an outcome (Zhang et al., 2018). ANN is almost analogous to the human brain. It takes in a collection of data, finds the relationships between each data point to the relevant output, and produces a prediction of outcome. This process is similar to teaching children to distinguish different animal faces. Each animal face can be labeled with its corresponding animal name and shown to the child. With enough sample size, the child will eventually notice the nuance of features, such as color, fur, eye distance, teeth, and facial shape, recognizing animal faces with high accuracy. However, training an ANN to be competent in doing pattern recognition and data analysis often requires tens of thousands of sample data. Therefore, there is a trend driving the development of big data technology (Rüping, 2015).

There are many illnesses that can be prevented or controlled by application of big data and ANN. For example, researchers developed an ANN algorithm that eliminated 75% of cardiac imaging diagnostic errors and medical malpractice claims that are related to physicians' cognitive factors (Dilsizian & Siegel, 2013). The promising result leads many scientists to believe that ANN and big data will transform the field of medicine and improve the current health system. Another recent focus has been fall risk analysis in senior patients. Falls in elderly people are the leading cause of visits to emergency departments and can lead to serious health problems (Fuller, 2000). Cognitive factors and high degree of variables make accurately predicting the likelihood of a person to fall almost impossible to a physician. However, an ANN with high degree of computational complexity and unbiased judgement can achieve the impossible (Hodas & Stinis, 2018). In this technical research, various ANN techniques will be explored to design an algorithm system that is capable of accurately predicting the potential fallers in a group of senior patients.

Although ANN seems to be the solution to many medical problems, it also raises many concerns over the use of big data. Between 2010 and 2013, data breaches reported by HIPAA-covered entities involved 29 million records (Liu, Musen, & Chou, 2015). In 2015, a safety failure in Anthem Insurance database single-handedly created the biggest medical data breach in American history, leaking 78.8 million customers' personal information, including names, social security numbers, street address, and income (Edwards, Hofmeyr, & Forrest, 2016). A study showed the general public has developed a deep distrust towards sharing personal data with private entities or hospitals (Beskow, 2016; Skloot, 2017). Over 70% of the American people expressed concern of their data security in the hands of their own insurance companies, and 10% of the Americans completely withhold any personal information from their health policy provider (V. Patel, Hughes, Savage, & Barker, 2015). The rampant medical data breaching problem did not exist prior to 2009 (Glenn & Monteith, 2014). There is clearly a need for better understanding of the current state of medical data security and the potential problem projection in the future. This STS research project will focus on investigating the safety of our medical data and the regulations that are still needed to lessen the distrust.

Technical Topic: Fall Risk Classification in Seniors

According to the U.S. Centers for Disease Control and Prevention, one in four Americans aged 65 and older falls each year, and 1,800 falls directly result in death. Falling is also the major factor that contributes to fracture risk; fractures of the hip, forearm, humerus, and pelvis usually result from the combined effect of falls and osteoporosis, which lead to further limitation of activity to the extent of loss of mobility (Prevention, Berg, & Cassells, 1992). Falling might also induce future fear of falling, which is a lasting concern that can cause an individual to avoid activities that he remains capable of performing (Adelsberg, Pitman, & Alexander, 1989). One strategy to preemptively prevent elderly patients from falling is to identify patients who are prone to falls and supply them with walking aid in advance (Oliver, Britton, Seed, Martin, & Hopper, 1997). Therefore, developing a method for predicting risk of falling to target high-risk individuals for preventive intervention is essential. However, due to difficulties in gathering gait data and connecting meaningful data in statistical analysis, current predictive methods remain inaccurate (Oakley et al., 1996). Hence, a better solution to the problem needs to be explored.

Recently, ANN has become a popular and helpful strategy for classification, clustering, pattern recognition, and prediction in many disciplines (Abiodun et al., 2018). ANNs are a type of machine learning algorithm that contains neuron-like nodes. Each node has a weight and a bias values associated. With all the different combinations of weights and biases within the network, a relatively small ANN can gather millions of degrees of freedom in the model. Allowing ANNs to simultaneously depict the relations between multiple factors and outcomes, high degree of freedom makes ANNs superior compared to conventional regression and statistical models (Abiodun et al., 2018).

ANNs have been widely applied in many medical fields including processing genetic information, electronic signal analysis, and medical image analysis and radiology (J. L. Patel & Goyal, 2007). However, limited research has been done on motion biomechanics, especially fall risk analysis in senior patients. Although previous studies, including analyses of local dynamic stability and forces of balance, used computational models to simulate human walking gait features and analyze various elements that contributed to falling, they did not utilize ANN to predict the patient's likelihood of falling (Ihlen, Weiss, Helbostad, & Hausdorff, 2015; Santos, Fukuchi, Fukuchi, & Duarte, 2017). In addition, previous methods often involved only one or few aspects of the falling elements; these methods failed to incorporate several contributors into a systematically comprehensive model. Furthermore, previous methods were proven inaccurate: a study shows conventional statistical modeling designed to identify fallers in patients failed to produce repeatable and accurate predictions (da Costa, Rutjes, Mendy, Freund-Heritage, & Vieira, 2012). This proposal highlights a novel application of ANN in elderly patient fall risk classification and analysis. With integration of various contributing factors identified from the previous studies and implementation of high-degree-of-freedom ANN algorithms, this approach aims to achieve high accuracy predictions of patient’s likelihood of falling.

This proposal, co-authored by Ziyang Guo and Leyao Li of department of biomedical engineering, includes two separate systems of algorithm. Research, development, and implementation of the two algorithm systems will span throughout two semesters: from fall of 2019 to spring of 2020. Each phase will take approximately a third of the total time. Two previously described algorithm systems are convolutional neural network (CNN) and long short term memory neural network (LSTM). CNN is often used for visual pattern recognition and object detection because its ability to distinguish spatial relations (Ouyang et al., 2015; Phung & Bouzerdoum, 2007). *Figure 1* demonstrates how a CNN recognizes a car. In this proposal, forces, center of mass, gait features, and other potential contributing factors will be treated as high-dimensional images. These three-dimensional images are constructed by adding time axis and three spatial planes. These images will then be fed into the CNN system to map correlations between the parameters and the likelihood of falling. LSTM, on the other hand, is often used for time series prediction (Connor, Martin, & Atlas, 1994). LSTM, trained to recognize patterns in a series of past events, is able to foresee the future events. Due to its unique capability, LSTM neural networks are often implemented for weather forecasting and speech recognition (Min Han, Jianhui Xi, Shiguo Xu, & Fu-Liang Yin, 2004). This proposal underlines a promising application of processing patient's time-dependent walking data through an LSTM algorithm to spot the potential fallers in the future. Currently, no similar applications of these two algorithms have been reported. This proposal can establish a brand new field.

The approach in this proposal will offer unparalleled advantages compared to previous regression and statistical modeling in multiple ways. If successful, the program will efficiently

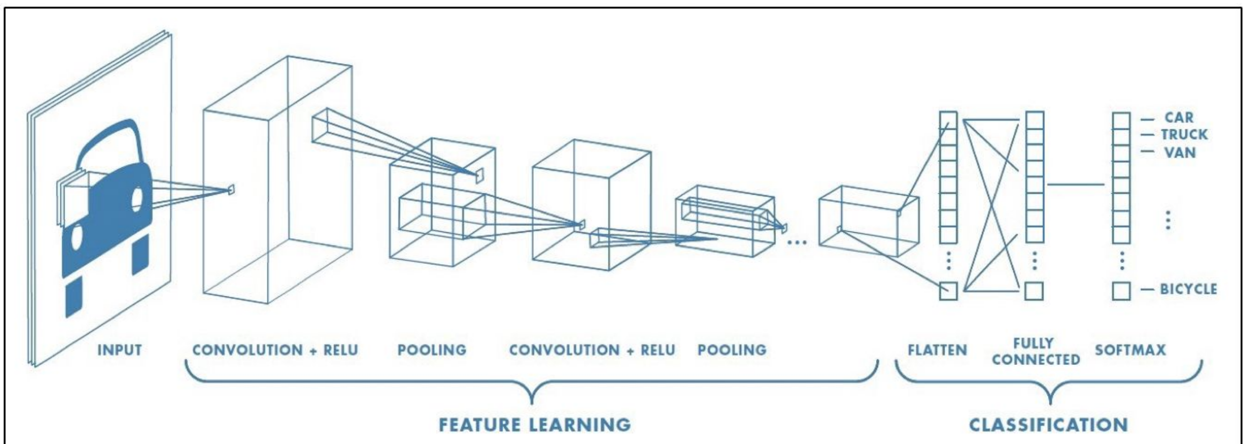

Figure 1: Convolutional Neural Network. This figure shows how convolutional neural network recognize a car. The network takes in pixels in batches, which are spatially related, and processes them into different features. (Saha, 2018)

revise the process in taking measurements of gait features, simplifying and formalizing the in-lab procedures. The computational program can provide insights into the contribution of different factors and identify most critical clinical characteristics. The two-algorithm system will also be appropriate to operate in a clinical setting: physicians can perform tests without having advanced statistical knowledge of a professional mathematician and patients will be able to receive their results in minutes. In addition, the proposed approach requires minimal processing of data by the physicians.. If successful, the proposed program will become a reliable tool for clinical practices. Elderly patients can take a measurement in the clinics by walking on a designed path and the program will generate results indicating whether the patient is a potential faller in the near future within a short period of time. Once a patient is identified as a faller, the clinicians can prescribe medications or aids accordingly to prevent falls from happening, and thus reduce numbers of deaths and injuries resulted from incidents.

STS Topic: Big Data in Medicine: Finding the Balance in Distrust and Progress

Information theft and data breaches in the medical field are serious problems that affect the majority of the population. ANN uses a massive amount of patient data to pinpoint the correlations between multiple variables, encouraging companies and individuals to collect patients' data illicitly. In the first half of 2019, there have been 32 million patient records breached (Davis, 2019). Data breaches are also prevalent in other fields. In 2013, 3 billion Yahoo users account information was hacked and released (Cheng, Liu, & Yao, 2018). In the same year, global surveillance disclosures revealed operational details about the United States National Security Agency (NSA) and its international partners' global surveillance of both foreign nationals and U.S. citizens (Snowden, 2019). In 2015, over 37.5 million records that contain

Figure 2: Cartoon Satire of Privacy Issue ("GDPR Jokes," n.d.)

personally identifiable information was leaked from Anthem (Browning & Tuma, 2015). In late 2018, Facebook had 50 million accounts worth of information stolen due to a security issue (Cadwalladr & Graham-Harrison, 2018). All of these incidents have eroded public trust in data security. Jokes (*Figure 2*) about insecurity in data privacy can be seen in every corner of the internet.

Data privacy laws are often described as lacking, or non-existent, in the past few years. Many industries, including the medical field, often self-regulate consumer data privacy and security (Listokin, 2015). A 2015 survey from Health Information Technology showed that 75% of Americans concerned about their health data being used by private companies, and 10% of all Americans decided to withhold all information from their health care providers to avoid private data breaches (V. Patel et al., 2015). The same survey also showed that the distrust that Americans hold towards data security in healthcare industry has only grown in recent years. This distrust also fuels another problem: the reluctance of individuals participating in medical-related research or giving consent to such studies. Many research groups repeatedly reported research failures due to data scarcity (Kuo, Leung, & Graham, 2012). To solve this cascade of problems, lack of proper legal boundaries must be addressed. However, before bluntly proposing any solution, one must closely examine the interactions between all stakeholders (legislation, hospitals, companies, consumers, and patients). Two theories, actor network theory (ANT) and technological momentum, will be employed to describe the situation and identifying the problems (Hughes, 1994; Latour, 1996).

Technological momentum is a theory that claims society can be heavily influenced by the arrival of new technology (Hughes, 1994). Technological momentum is often used to describe unprecedented social impact brought by novel technology. The current medical data privacy problem can be explained by technological momentum. The use of big data and ANN is a relatively new technology. Due to the lack of regulation, exploitation of big data such as data mining has gathered its momentum and caused billions of dollars damage via data breaching. Meanwhile, big data technology also has shaped the public views towards personal data privacy. People became more aware of the value of their personal data, and the society as a whole started to rethink how private data should be protected. However, technological momentum has its limitations at comprehensively describing the current situation of medical data privacy and big data technology. This framework mainly focuses on the period of time that the technology gained its most traction, and it fails to incorporate the aftermath. For instance, the framework does not account for the fact that laws and regulations have been proposed to restrict the rapid growth of big data technology.

ANT is a better theory to illustrate the cause and effect of every stakeholder has on each other. ANT is a constructivist approach to describe a group of unspecified relationships among entities (Latour, 1996). These entities in ANT are called actors. ANT will identify the relationship between the problems and the stakeholders, mapping the way that medical data privacy issues propagate throughout policymakers, technology, and society. However, ANT also has limitations of its own. ANT tries to identify all of the heterogeneous associations between the human and nonhuman actors in the network, which might lead to exaggeration over minor issues (Walsham, 1997).

Research Question and Methods:

What is the current state of medical data security and how is big data technology affecting it?

To investigate the safety of medical data, HIPAA (Health Insurance Portability and Accountability Act of 1996) regulations, NIH (National Institute of Health) standard procedures, and other applicable privacy laws will be compared before and after each major data breach incident. Historical cases such as Anthem and Premera Blue Cross breaches will be studies through HIPAA annual reports and scientific reviews (Khan & Hoque, 2016). A comprehensive report will be provided at the end of the research to illustrate how privacy regulations and novel technology have influenced medical data safety. In addition, social and political reactions to privacy regulations will be studied healthcare related surveys to demonstrate the negative and positive effects of technological impact. To formulate a proposal for better guarding vulnerable medical data, recent breaches will be examined to pinpoint the loopholes of the system. 35 million medical data breach incidents of 2019 will be studied by reviewing the incident report from the responsible entities (Jiang & Bai, 2019). The research will be conducted mainly via examining and analyzing currently available literature and statistical data. This research project ultimately aims to evaluate current medical data safety, especially under the context of technological shift. Furthermore, this research project also intends to provide suggestions on protecting the vulnerable medical data against illegal data mining.

Conclusion

The technical research project in fall risk classification of senior patients, which consists of development of CNN and LSTM, if completely successfully, will be able to accurately identify the latent fallers in elderly patients to prevent future injuries and shed light on how ANN technology can be utilized in clinical settings. This project will alleviate the stress of current technological scarcity of fall risk classification and aid physicians in diagnosing dynamic instability. The anticipated accuracy rate of prediction for this project is above 90%.

The STS research project will analyze the impact of big data technology and evaluate security of current medical data systems. If completed successfully, this project will formulate a proposal of potential improvements to the current medical data system. Additionally, this project is anticipated to equip the public with the knowledge of present medical data status while alleviating the profound public distrust towards big data technology.

Reference

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11), e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Adelsberg, S., Pitman, M., & Alexander, H. (1989). Lower extremity fractures: Relationship to reaction time and coordination time. *Archives of Physical Medicine and Rehabilitation*, *70*(10), 737–739.

Beskow, L. M. (2016). Lessons from HeLa Cells: The Ethics and Policy of Biospecimens. *Annual Review of Genomics and Human Genetics*, *17*, 395–417. https://doi.org/10.1146/annurev-genom-083115-022536

Browning, J. G., & Tuma, S. (2015). If Your Heart Skips a Beat, It May Have Been Hacked: Cybersecurity Concerns with Implanted Medical Devices. *South Carolina Law Review*, (3), [i]-678.

Cadwalladr, C., & Graham-Harrison, E. (2018). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. 5.

Cheng, L., Liu, F., & Yao, D. (Daphne). (2018). Enterprise data breach: Causes, challenges, prevention, and future directions. *Communications on Pure and Applied Mathematics*, e1211. https://doi.org/10.1002/widm.1211@10.1002/(ISSN)1097-0312.India-Mathematics-Day

Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, *5*(2), 240–254. https://doi.org/10.1109/72.279188

da Costa, B. R., Rutjes, A. W. S., Mendy, A., Freund-Heritage, R., & Vieira, E. R. (2012). Can Falls Risk Prediction Tools Correctly Identify Fall-Prone Elderly Rehabilitation Inpatients?

A Systematic Review and Meta-Analysis. *PLoS ONE*, *7*(7). https://doi.org/10.1371/journal.pone.0041061

Davis, J. (2019, August 1). 32M Patient Records Breached in First Half of 2019, 88% Caused by Hacking. Retrieved October 23, 2019, from HealthITSecurity website: https://healthitsecurity.com/news/32m-patient-records-breached-in-first-half-of-2019-88-caused-by-hacking

Dilsizian, S. E., & Siegel, E. L. (2013). Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment. *Current Cardiology Reports*, *16*(1), 441. https://doi.org/10.1007/s11886-013-0441-8

Edwards, B., Hofmeyr, S., & Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, *2*(1), 3–14. https://doi.org/10.1093/cybsec/tyw003

Fuller, G. F. (2000). Falls in the elderly. *American Family Physician*, *61*(7), 2159–2168, 2173–2174.

GDPR Jokes. (n.d.). Retrieved October 31, 2019, from Exonar website: https://www.exonar.com/gdpr-jokes/

Glenn, T., & Monteith, S. (2014). Privacy in the Digital World: Medical and Health Data Outside of HIPAA Protections. *Current Psychiatry Reports*, *16*(11), 494. https://doi.org/10.1007/s11920-014-0494-4

Hodas, N. O., & Stinis, P. (2018). Doing the Impossible: Why Neural Networks Can Be Trained at All. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01185

Hughes, T. (1994). *Technological Momentum*.

Ihlen, E. A. F., Weiss, A., Helbostad, J. L., & Hausdorff, J. M. (2015). The Discriminant Value of Phase-Dependent Local Dynamic Stability of Daily Life Walking in Older Adult

Community-Dwelling Fallers and Nonfallers [Research article]. https://doi.org/10.1155/2015/402596

Jiang, J. (Xuefeng), & Bai, G. (2019). Evaluation of Causes of Protected Health Information Breaches. *JAMA Internal Medicine*, *179*(2), 265–267. https://doi.org/10.1001/jamainternmed.2018.5295

Khan, S. I., & Hoque, A. S. L. (2016). *Digital Health Data: A Comprehensive Review of Privacy and Security Risks and Some Recommendations*. 20.

Kuo, Y.-H., Leung, J. M. Y., & Graham, C. A. (2012). Simulation with data scarcity: Developing a simulation model of a hospital emergency department. *Proceedings of the 2012 Winter Simulation Conference (WSC)*, 1–12. https://doi.org/10.1109/WSC.2012.6465061

Latour, B. (1996). On actor-network theory: A few clarifications. *Soziale Welt*, *47*(4), 369–381. Retrieved from JSTOR.

Listokin, S. (2015). Industry Self-Regulation of Consumer Data Privacy and Security. *John Marshall Journal of Information Technology and Privacy Law*, (1), 15–32.

Liu, V., Musen, M. A., & Chou, T. (2015). Data Breaches of Protected Health Information in the United States. *JAMA*, *313*(14), 1471–1473. https://doi.org/10.1001/jama.2015.2252

Min Han, Jianhui Xi, Shiguo Xu, & Fu-Liang Yin. (2004). Prediction of chaotic time series based on the recurrent predictor neural network. *IEEE Transactions on Signal Processing*, *52*(12), 3409–3416. https://doi.org/10.1109/TSP.2004.837418

Oakley, A., Dawson, M. F., Holland, J., Arnold, S., Cryer, C., Doyle, Y., … Eastwood, A. (1996). Preventing falls and subsequent injury in older people. *Quality in Health Care*, *5*(4), 243–249.

Oliver, D., Britton, M., Seed, P., Martin, F. C., & Hopper, A. H. (1997). Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: Case-control and cohort studies. *BMJ*, *315*(7115), 1049–1053. https://doi.org/10.1136/bmj.315.7115.1049

Ouyang, W., Wang, X., Zeng, X., Qiu, S., Luo, P., Tian, Y., … Tang, X. (2015). *DeepID-Net: Deformable Deep Convolutional Neural Networks for Object Detection*. 2403–2412. Retrieved from https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Ouyang_DeepID-Net_Deformable_Deep_2015_CVPR_paper.html

Patel, J. L., & Goyal, R. K. (2007). Applications of artificial neural networks in medical science. *Current Clinical Pharmacology*, *2*(3), 217–226.

Patel, V., Hughes, P., Savage, L., & Barker, W. (2015). *Individuals' Perceptions of the Privacy and Security of Medical Records*. 9.

Phung, S. L., & Bouzerdoum, A. (2007). A Pyramidal Neural Network For Visual Pattern Recognition. *IEEE Transactions on Neural Networks*, *18*(2), 329–343. https://doi.org/10.1109/TNN.2006.884677

Prevention, I. of M. (US) D. of H. P. and D., Berg, R. L., & Cassells, J. S. (1992). *Falls in Older Persons: Risk Factors and Prevention*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK235613/

Rüping, S. (2015). Big Data in Medizin und Gesundheitswesen. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, *58*(8), 794–798. https://doi.org/10.1007/s00103-015-2181-y

Saha, S. (2018, December 17). A Comprehensive Guide to Convolutional Neural Networks—The ELI5 way. Retrieved October 31, 2019, from Medium website: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Santos, D. A. dos, Fukuchi, C. A., Fukuchi, R. K., & Duarte, M. (2017). A data set with kinematic and ground reaction forces of human balance. *PeerJ*, *5*, e3626. https://doi.org/10.7717/peerj.3626

Skloot, R. (2017). *The Immortal Life of Henrietta Lacks*. Broadway Books.

Snowden, E. (2019). *Permanent Record*. Henry Holt and Company.

Walsham, G. (1997). Actor-Network Theory and IS Research: Current Status and Future Prospects. In A. S. Lee, J. Liebenau, & J. I. DeGross (Eds.), *Information Systems and Qualitative Research: Proceedings of the IFIP TC8 WG 8.2 International Conference on Information Systems and Qualitative Research, 31st May–3rd June 1997, Philadelphia, Pennsylvania, USA* (pp. 466–480). https://doi.org/10.1007/978-0-387-35309-8_23

Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018). Opening the black box of neural networks: Methods for interpreting neural network models in clinical applications. *Annals of Translational Medicine*, *6*(11). https://doi.org/10.21037/atm.2018.05.32