

Predicting Mortality of Diabetic ICU Patients

by

Ian Wittler

Presented in partial fulfillment of the requirements

for

Departmental Honors

in the

Department of Computer Science

Hood College

April 2019

Predicting Mortality of Diabetic ICU Patients

Ian Wittler

imw2@hood.edu

Department of Computer Science

Hood College

Frederick, Maryland

ABSTRACT

Diabetes mellitus (DM) is a major public health concern that requires continuing medical care. It is also a leading cause of other serious health complications associated with longer hospital stays and increased mortality rates. Fluctuation of blood glucose levels are easy to monitor. Physicians manage patients' blood glucose to prevent or slow the progress of diabetes. In this paper, the MIMIC-III data set is used to develop and train multiple models that aim to predict the mortality of DM patients. Our deep learning model of convolutional neural network produced a 0.885 AUC score, above all baseline models we constructed, which include decision trees, random forests, and fully connected neural networks. The inputs for each model were comprised of admission type, age, Elixhauser comorbidity score, blood glucose measurements, and blood glucose range. The results obtained from these models are valuable for physicians, patients, and insurance companies. By analyzing the features that drive these models, care management for diabetic patients in an ICU setting can be improved resulting in lowered mortality rate.

CCS CONCEPTS

• **Computing methodologies** → **Supervised learning by classification**; **Neural networks**; *Feature selection*; *Visual analytics*.

KEYWORDS

data mining, diabetes, convolutional neural networks, supervised classification

1 INTRODUCTION

Diabetes mellitus affects more than 30 million Americans - roughly 10% of the U.S. population. Additionally, an estimated 80 million Americans suffer from prediabetes, a condition that if left untreated often leads to diabetes within 5 years [1]. What is diabetes and how does it affect those who have it? Simply explained, diabetes prevents the body from properly using energy obtained from food. Those with diabetes have trouble producing insulin in their pancreas. Without proper insulin, the body cannot maintain steady blood glucose levels, which if not monitored and treated properly can lead to numerous other health complications [2].

Analyzing large amounts of electronic health records (EHR) to extract useful information has the potential (and in many cases already has) to improve care in medical settings drastically [3]. One specific area of study involves using classification techniques to predict patient mortality. Generally these studies use generic features between all patients; however, this does not account for unique features that certain subgroups of the total population may potentially have. For example, it is required that diabetic patients

have their blood glucose levels monitored frequently. This requirement creates a unique feature that other non-diabetic patients may not have.

In this paper, the primary goal is to model diabetic patient mortality, in addition to utilizing the glucose measurement feature. The glucose measurements are recorded in mg/dL (milligrams per deciliter) and fluctuate up and down throughout the time of a patient's visit. Due to the time sequence aspect of the glucose measurements, a convolutional neural network (CNN) is the main modeling technique of focus. CNN's (commonly used for imaging task) are very good at detecting trends in data. For example, locating rapid changes in glucose measurements before death, and associating those measurement changes to an output. By incorporating blood glucose measurements along with a few additional features, it will be determined if glucose measurements are a beneficial feature to the modeling of mortality for diabetic patients.

2 RELATED WORK

The ability to analyze and explore health record data is important for improving medical knowledge and creating models that can predict different patient outcomes. There are many articles related to this idea. For example, one study aimed to predict unplanned admissions, length of stays, diagnoses, and mortality by using deep learning on raw patient records [4]. This study's method is unique from traditional modeling methods since it incorporates all data related to a patient without discarding information deemed unimportant. Other studies, including this one [5], also used deep learning to investigate similar prediction task. However, in the second study cited, a more structured input feature technique was used. It is important to note that this study recognized that lab results were not easy to process and therefore were non-included (a limitation). One aspect of modeling electronic health records that proves to be slightly difficult is processing medical notes. Data mining medical notes in an effective way is an entire branch of interdisciplinary research which is further explained in this referenced paper [6].

In my paper, there are two primary areas of interest that surround the modeling experiment - mortality and diabetic patients. Modeling mortality in an ICU is an important tool to help determine care options for patients and physicians. Most models that are currently in use to predict likelihood of mortality use a given number of generic input features. The goal of mortality prediction studies are to find both methods and features which can best classify a patient's outcome, in comparison to a benchmark model [7]. One study, done on real-time mortality prediction, attempted to predict a patient's outcome at different time intervals throughout their stay in order to provide physicians with a simple synthesis of patient acuity [8].

One potential issue with mortality models such as the ones described above is that they are intended for a general population

and do not always account for patients with specific conditions. Diabetic patients have particular predictors that are unique to their subgroup which may not apply to non-diabetic patients. Using simple demographic, socioeconomic, and biological factors, mortality can be predicted with fair reliability for diabetic people [9]. One study, which attempted to predict mortality (for an even smaller subgroup) used a point based model that focused on elderly-specific adults with diabetes [10]. The study did this because the research done in that area was still lacking.

As discussed above, mortality models can be useful as a tool to help make patient care decisions. While diabetes does not directly associated with increased mortality, it is known that critically ill diabetic patients in the ICU have an increased risk for the development of further complications [11], which make these patients susceptible to higher health care cost. This correlation hints that there may be benefit in predicting mortality of diabetic ICU patients. One study which did this using logistic regression achieved a good AUC value of 0.787 using a very simple model, which required way less variables than the other comparable benchmark models [12]. One of the features used in this model was the patient’s average blood glucose. In my paper, I aim to expand the potential of this feature by using multiple blood glucose measurements. Another area of potential improvement, suggested from this study, was the modeling technique. Overall, there seems to be little research into how glucose measurement volatility over time relates to different outcomes of diabetic patients. In my paper, the goal is to use glucose measurements as a unique feature for mortality modeling.

3 EXPERIMENT

The primary goal of the experiment in my research was to create a model that could predict mortality of diabetic patients in an ICU environment. With 20% (11820/58976) of all patient visits in the MIMIC III data set being diabetic patients, an interesting opportunity is presented. In a hospital setting, each patient who has diagnosed diabetes of any type has their glucose levels recorded each couple of hours. This monitoring allows for the experiment to have a unique extra feature that would not be present in a non-diabetic population. These glucose measurements will be a main concentration of the input features for the prediction model. The secondary goal of the experiment is to reduce the number of input features to a relatively small number. Many models currently exist that require 35 features or more for a risk assessment and prediction. By reducing the number of features required, it becomes cheaper and quicker to run these types of assessments [12].

3.1 Pre-processing

MIMIC III is a large, freely-available database comprised of de-identified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012 [13]. The database encompasses tables that include patient admission data, diagnoses, medical notes, and chart events (such as blood glucose levels). In whole, the database takes up between 40-50 GB of storage and contains thousands of potential features to observe.

For this experiment, PostgreSQL was used as the database management system due to the compatibility with the MIMIC III data

Table 1: Example ICD-9 Codes

ICD-9 Code	Meaning
250	Diabetes mellitus
401.9	Hypertension
584.9	Acute kidney failure

set. Each of the rows extracted, which will be described below, was done through the process of querying and exporting rows to CSV files. Once the proper data was collected into separately labeled CSV files, the data could then be manipulated and pre-processed with python. The first and most important table of interest in MIMIC III was the Chart Event table. Chart Events consist of all charted data during a patient’s stay. The data most important for this experiment in the Chart Events table was the blood glucose measurements. Due to multiple data collection labeling techniques, the blood glucose measurements were labeled with numerous ID’s including: 807, 811, 1529, 3744, 3745, 220621, 225664, and 226537. By querying only the rows that contained these ID’s, all patients with at least one blood glucose measurement were accounted for.

The next table of interest was the ICD Diagnoses table. ICD Diagnoses consist of all ICD-9 (International Classification of Diseases) codes individual patients were assigned during their visit. An example of some common ICD-9 codes can be found in table 1 above. The ICD-9 codes needed for this experiment were all codes greater than or equal to 250.00 and less than 251.00. By querying the rows that fit within the range 250-251, all patients who were given codes that signify diabetes mellitus (type 1 or 2) were accounted for. It should be noted that the ICD-9 system is used for insurance and medical billing purposes, and therefore is not always completely accurate. For more information into how the accuracy of diabetes-specific coding affected this experiment refer to Appendix A: Examining the Reliability of ICD9 Code Diagnosis.

After finding all patients that were coded to be diabetic, it is helpful to recognize that each patient had a variable amount of glucose measurements during their visit. The total amount varied from 1-78 measurements and is graphed in figure 1. The proportion of patients who lived or not is also located on the same graph. In order to combat this variability, if a patient had more than 30 glucose measurements, only the final 30 glucose measurements before death or discharge were used as input features. Additionally, a new feature - glucose range, was engineered from the glucose measurements to represent the volatility of each patient’s measurements during their stay.

The two input features, age and admission type, were both extracted from the Admission table. Age, by default, is an integer and needed no further manipulation. Admission type had to be encoded from a string to an integer (1 -> "elective", 0 -> "emergency", etc.). The final feature, which also had to be engineered, was the Elixhauser score. The Elixhauser score (an integer value) is used for measuring patient comorbidity and is created from an algorithm which uses ICD-9 codes as the input with the Elixhauser score being the output [14]. After the above feature extraction, all input features were collected (see figure 2) and ready to be tested for each model.

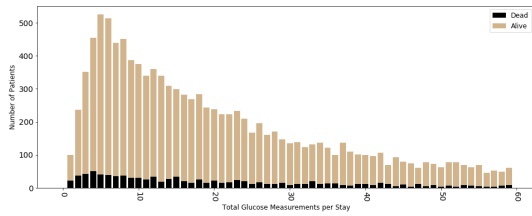


Figure 1: Glucose Measurements per Stay vs. Number of Patients Distribution.

Expire Flag	Admission Type	Age	Elixhauser Score	Glucose Measurements(mg/dL)	Glucose Range
0	0	0	82	12	94,98,98,...
1	0	0	72	0	167,151,137,...
2	1	1	63	19	56,146,126,...

Figure 2: Example Post-Processing Dataframe.

3.2 Data Visualization

The use of data mining techniques including neural networks has vastly improved how computers can process large amounts of data to find patterns and trends. Yet as humans, observing large amounts of raw data is mostly rather pointless if the data cannot be comprised into a more visually organized method. By using visualization for large data, decision-makers can feel more comfortable about discovering patterns, comprehending information, and forming an opinion [15].

In many instances, data visualization can help to alleviate the black box effect. The black box effect is the idea that with different modeling methods (especially neural networks), there are inputs and outputs, but in the middle is a mess of numerical weights and connections that provide no direct evidence to relationships between the inputs and outputs [16]. This black box makes it hard for science and engineering fields to accept neural networks since they cannot easily be understood or validated. In very high-stake applications, using techniques that are considered black boxes are unlikely. Therefore, to help better understand the data used in this experiment, multiple data visualization techniques were used.

One example of data visualization is of patient glucose levels. By graphing different patient's charts (both those who are alive and not alive) it is easier for us to understand and locate potential data trends. Using HoloViews, an open source Python library that concentrates on data analysis and visualization [17], a graph for each patient's glucose measurements were plotted. The patients were separated by green, which signals an alive status, and red, which signals an expired status. An example of both an alive and expired graph can be observed in figure 3.

3.3 Feature Analysis

Developing a good understanding of the input features used in any modeling experiment is critical. One way to determine a feature's value for a given model is to see how much a model's performance decreases when that feature is removed. The more a feature is used to make key decisions in a model, the higher relative importance that feature has. Feature importance helps to simplify what exactly

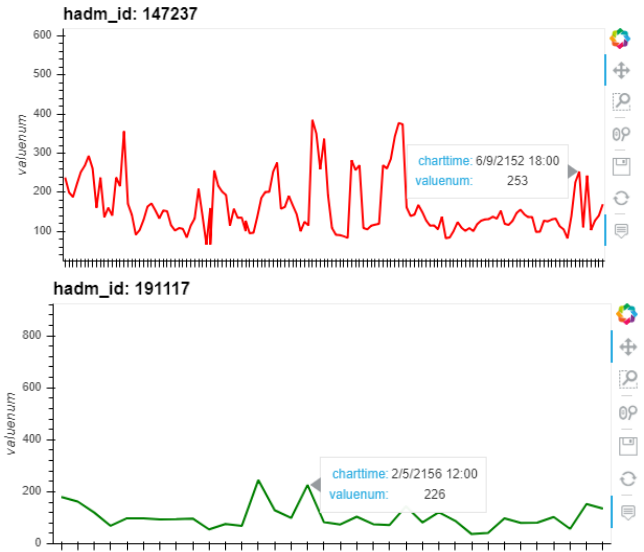


Figure 3: Patient Glucose Measurements Graphs - Expired (top) and Alive (bottom).

the driving factors are with decision making in prediction models. Figure 4, located below, graphs the importance percentage of each feature for this experiment. What each number located on the x-axis means is explained in the following list:

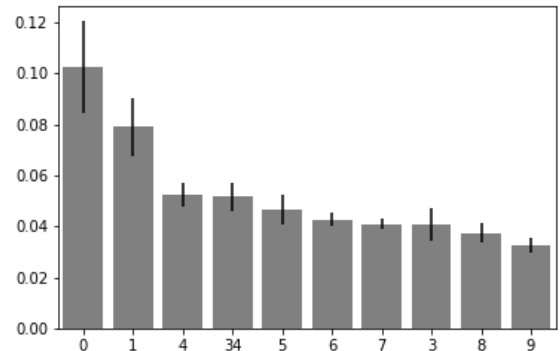


Figure 4: Feature Importance Graph.

- Feature 0 (Elixhauser Score) : proved to be the most important feature with a 10.0% gain.
- Feature 1 (Age) : second highest gain of 7.4%.
- Feature 34 (Glucose range) : gain of 5.6%.
- Feature 3-9 (Individual Glucose Measurements) : It should be noted that features 3-33 are the 30 glucose measurements taken. The final measurement being feature 3 and the first being feature 33. More notable is the fact that feature 2 (Admission Type) is not in the top 10.

While determining feature importance is one useful way to analyze input features, another useful method is creating heatmaps. Heatmaps are beneficial in determining how correlated different features are between each other. With the features used for the simplified model in this experiment a few notable points can be made (seen in figure 5). Overall, the individual glucose measurement features are each closely correlated with each other. Yet, as two glucose measurement get farther apart from each other they become less and less correlated. This is a logical trend in the sense that many patient's glucose levels will be relatively close from hour to hour. If glucose levels are not correlated heavily for an individual patient, then that could signify a large glucose range. Additionally, in figure 5 it can be seen that the Elixhauser score feature slightly correlates to both age and admission type. Again this logically makes sense since older people tend to have multiple health complications and emergency admissions. By observing the input features through various data visualizations, it becomes more clear how the different features interact with both each other and the overall model.

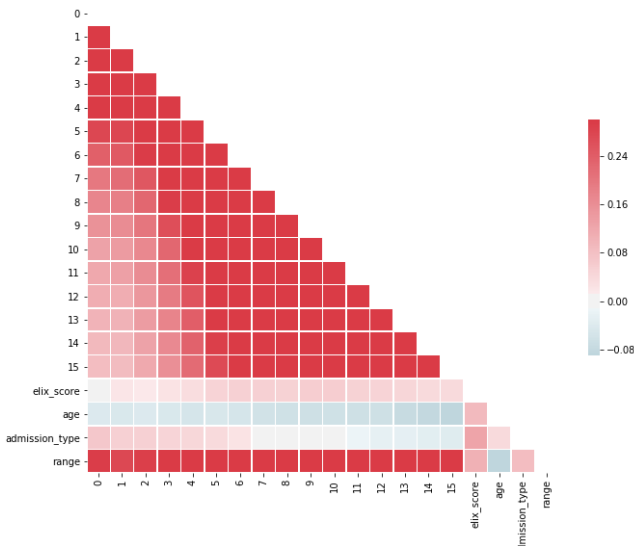


Figure 5: Heatmap of All Input Features.

3.4 Modeling

The goal in this experiment is to best classify the binary variable labeled expire flag which signals mortality. For the purpose of comparison, multiple techniques were used as baseline models, including a trained decision tree model, a trained random forest model, and a feed forward neural network model. In the below paragraphs, the baseline models will be further explained as well as the convolutional neural network model (the main model of interest for this experiment).

After pre-processing and feature exploration was completed, a few techniques were used to prepare the data for the models. The first concern was to address the imbalance of patients that died versus patients that lived. Seen in figure 6, roughly 90 percent of patients lived (12217 patients), while only 10 percent did not

(1443 patients). There are a few possible methods to address this imbalance. One method is to use under-sampling. This method is done by reducing the amount of patients who lived to a more even level of patients who did not. This method works best when the quantity of data is large. Another method is to use over-sampling, which is the idea of boosting the amount of minority observations to equal a number closer to the majority observations. For this experiment, a combination of these two methods were used, where the majority was under-sampled to 5000 and the minority was over-sampled to 4000. By performing these sampling techniques, the class imbalance concern was resolved. The second concern before the data was ready for modeling was how the data would be split. For this experiment, the data was split two ways - 70 percent for training and 30 percent for testing and validation.

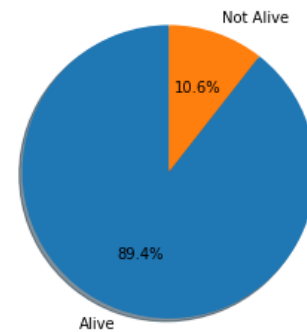


Figure 6: Distribution of Mortality Classes.

3.4.1 *Baseline Models.* The three baseline models used for this experiment included a trained decision tree, a random forest model, and a feed forward neural network. All of the models used in this experiment were supervised binary classification models. Decision trees make good baseline models due to their simplicity and easy to visualize nature. Figure 7 demonstrates this idea by showing the top two layers for the decision tree model, which uses a gini impurity algorithm. In this decision tree, it can be seen that X0 (Elixhauser score) is the first decision node and patients with a value of 14.5 and higher go left while a value lower continues right. Further down the feature X1 (Age) is also used to make decisions. Many times the features with the highest importance tend to be closer to the root node (seen here with the Elixhauser score and age).

The random forest model, which utilizes multiple decision trees, served as an intermediary model for the neural networks used. The baseline neural network, a feed forward network, was chosen for its similarity to the CNN. By comparing the feed forward to the convolutional network, it is easy to notice the performance improvements that the CNN provides.

3.4.2 *CNN Model.* Convolutional neural networks are commonly used in imaging classification tasks due to their ability to detect features and map them to specific outputs [18]. These networks

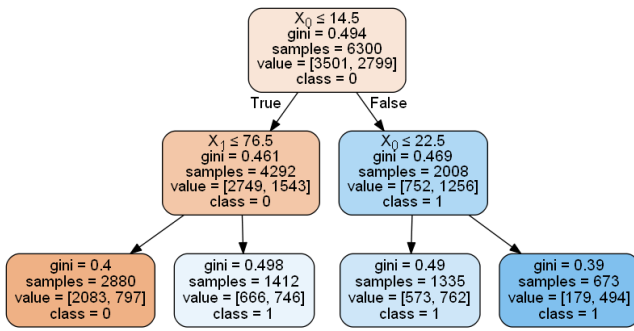


Figure 7: First Two Layers of Decision Tree Model.

are also beneficial since they can handle input features with little engineering done to the features and still detect trends. In this experiment, part of the input feature matrix is the time series glucose measurements. By using a convolutional neural network, detecting trends in the changes of glucose measurements should be improved. The summary used for this model can be found in figure 8. In addition to the model summary, the tuned hyper-parameters are as followed:

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 28, 100)	900
max_pooling1d_1 (MaxPooling1)	(None, 14, 100)	0
flatten_1 (Flatten)	(None, 1400)	0
dense_1 (Dense)	(None, 50)	70050
dense_2 (Dense)	(None, 1)	51
Total params: 71,001		
Trainable params: 71,001		
Non-trainable params: 0		

Figure 8: CNN Model Hyper-parameters.

- Number of Filters: 100
- Kernel size: 8
- Pool Size: 2
- Activation Functions: Relu
- Optimizer and Loss: Adam and MSE
- Epochs: 75
- Batch Size: 10

The convolutional network was created using keras, an open-source neural-network library, ran on top of tensorflow. The programming language python was used with the application Jupyter Notebook and tested with the XSEDE supercomputer. For more specific information on any of the code created for these modes, refer to the following Github link: <https://github.com/imw2/Diabetes-Diagnosis-Prediction>

4 RESULTS

The most important evaluation metric used for these binary classification models was Area Under the Curve (AUC), which in this case is preferred over accuracy. Observing the true positive rate (TPR) and false positive rate (FPR) better demonstrates how the model predicts both cases of mortality when compared to accuracy which only factors the overall percentage of true and false predictions.

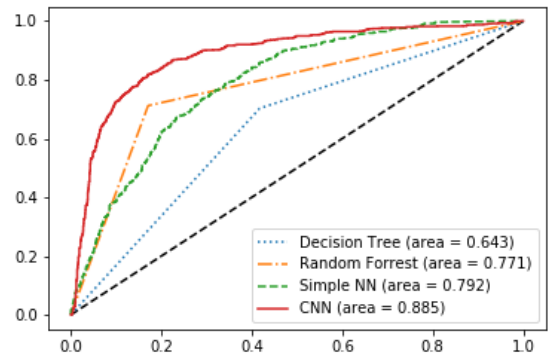


Figure 9: AUROC (Area Under the Receiver Operating Characteristics).

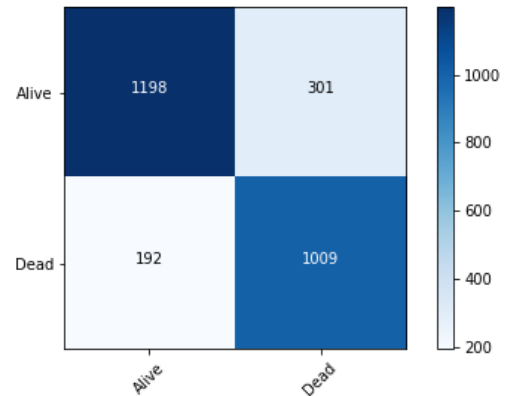


Figure 10: CNN Confusion Matrix.

The final results for each model can be seen in figure 9 above, where the CNN (area = 0.885) performed the best, the feed forward neural network (area = 0.792) second best, then random forest (area = 0.771) and decision tree (area = 0.643) being the remaining scores respectively. These results show that by using the convolutional network, the initial idea that it could better understand the changes in glucose measurements is correct. This proves that there is benefit to incorporating blood glucose measurements into the models. It also demonstrates that a patient's mean glucose measurement does

not do have the same impact as incorporating multiple glucose measurements. It can be concluded that using individual glucose measurements seems to improve mortality modeling of diabetic patients.

Another way to visualize binary classification is to use a confusion matrix, which shows how many times each possible outcome was predicted versus the true outcome in the validation set of data. This matrix makes it easier to see TPR and FPR which are used to calculate the above graph. Figure 10 shows a confusion matrix for the CNN model of this experiment.

5 CONCLUSION

It is known that monitoring and keeping healthy blood glucose levels is a necessary event for healthy diabetic people. In this paper, the goal was to determine if multiple blood glucose measurements are beneficial features for predicting the mortality of diabetic patients. From the results, it is seen that by utilizing and modeling the blood glucose measurements, an improved mortality model can be achieved. The successful use of supervised learning, a convolutional neural network, and proper computing resources allowed for the strides to be made in this paper. Additionally, throughout the research other aspects concerning diabetic patients were investigated, for example determining how the reliability of ICD9 coding effected the mortality model. Overall, there is still much research to be done regarding diabetic patients; however, the research done in this paper serves as a deep investigation into using modeling techniques to extract useful information from electronic health record data.

6 FUTURE WORK

During the research, some limitation were encountered. One of the challenges presented in this experiment was determining how to produce a model with both high evaluation metrics and a low total of input features. It should be noted that more features could be added to the existing models which would have improved the evaluation metrics, yet would also increase the complexity of the model. Another potential area of further investigation is determining which patients this model should apply to. It is possible that other patients with similar diseases to diabetes may also perform well using these models.

Moving forward from this research, there are two main focuses for future work. The first focus involves associating diabetic patient care methods to specific outcomes. Another study done investigated loose blood glucose control versus strict blood glucose control in an ICU setting [19]. The study concluded that the verdict on loose versus strict blood glucose control is not a simple task and often depends on other factors such as if the patient is neurological or surgical. In order to conduct research that better defines how different blood glucose control methods effect mortality (and other outcomes), more investigation into how the type of control is represented in electronic health records would be necessary.

The second focus involves finding patterns between blood glucose measurements and additional medical events. Examples of medical events include surgeries, unit transfers, or medication intakes. This focus would require new classification techniques, but could produce informative results. As medical records become more

connected and processed, personalized patient care has the ability to improve overall health.

7 ACKNOWLEDGMENTS

I would like to express my very great appreciation to Drs. Aijuan Dong and Xinlian Liu for their expertise in the field and advising to this research. I would also like to thank Dr. Miranda Darby of the Bioinformatics Program, Dr. Andrew Campbell of the Counseling Program, Dr. Linda Kennedy of the Nursing Program, and Dr. Silvia Crivelli of the Lawrence Berkeley National Laboratory for their invaluable insights and constructive criticisms.

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) [20], which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system [21], which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC) through an educational allocation TG-ASC190003.

8 APPENDICES

There is one appendix in connection to this paper: Examining the Reliability of ICD9 Code Diagnosis. The content of this appendix was discovered during research as it was noticed that the ICD9 coding had some interesting patterns in relation to diabetes. Some figures and explanation can be found below, but for more information on this topic along with other topics explored see the Github repository: <https://github.com/imw2/Diabetes-Diagnosis-Prediction>.

REFERENCES

- [1] New cdc report: More than 100 million americans have diabetes or prediabetes | cdc online newsroom | cdc.
- [2] Diabetes mellitus: An overview.
- [3] Alistair E.W. Johnson, Tom J. Pollard, Leo A. Celi, and Roger G. Mark. Analyzing the eicu collaborative research database. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, ACM-BCB '17*, pages 631–631, New York, NY, USA, 2017. ACM.
- [4] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Peter J. Liu, Xiaobing Liu, Mimi Sun, Patrik Sundberg, Hector Yee, Kun Zhang, Gavin E. Duggan, Gerardo Flores, Michaela Hardt, Jamie Irvine, Quoc Le, Kurt Litsch, Jake Marcus, Alexander Mossin, Justin Tansuwan, De Wang, James Wexler, Jimbo Wilson, Dana Ludwig, Samuel L. Volchenboum, Katherine Chou, Michael Pearson, Srinivasan Madabushi, Nigam H. Shah, Atul J. Butte, Michael Howell, Claire Cui, Greg Corrado, and Jeff Dean. Scalable and accurate deep learning for electronic health records. 2018.
- [5] Riccardo Miotto, Li Li, Brian A. Kidd, and Joel T. Dudley. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Scientific Reports*, 2016.
- [6] Luis Pereira, Rui Rijo, Catarina Silva, and Ricardo Martinho. Text Mining Applied to Electronic Medical Records. *International Journal of E-Health and Medical Communications*, 2015.
- [7] Angela K. M. Lipshutz, John R. Feiner, Barbara Grimes, and Michael A. Gropper. Predicting mortality in the intensive care unit: a comparison of the university health consortium expected probability of mortality and the mortality prediction model iii. *Journal of Intensive Care*, 4(1):35, May 2016.
- [8] Alistair E W Johnson and Roger G Mark. Real-time mortality prediction in the Intensive Care Unit. Technical report.
- [9] Laura N. McEwen, Andrew J. Karter, Beth E. Waitzfelder, Jesse C. Crosson, David G. Marrero, Carol M. Mangione, and William H. Herman. Predictors of mortality over 8 years in type 2 diabetic patients. *Diabetes Care*, 35(6):1301–1309, 2012.
- [10] Y. K. Chang, L. F. Huang, S. J. Shin, K. D. Lin, K. Chong, F. S. Yen, H. Y. Chang, S. Y. Chuang, T. J. Hsieh, C. A. Hsiung, and C. C. Hsu. A Point-based Mortality Prediction System for Older Adults with Diabetes. *Scientific Reports*, 7(1):1–10, 2017.
- [11] Sarah E. Siegelhaar, Maartje Hickmann, Joost B.L. Hoekstra, Frits Holleman, and J. Hans DeVries. The effect of diabetes on mortality in critically ill patients: A systematic review and meta-analysis. *Critical Care*, 15(5):R205, 2011.

- [12] Rajsavi S Anand, Paul Stey, Sukrit Jain, Dustin R Biron, Harikrishna Bhatt, Kristina Monteiro, Edward Feller, Megan L Ranney, Indra Neil Sarkar, and Elizabeth S Chen. Predicting Mortality in Diabetic ICU Patients Using Machine Learning and Severity Indices. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*, 2017:310–319, 2018.
- [13] Iii critical care database.
- [14] David J. Stone Leo A. Celi Johnson, Alistair EW and Tom J. Pollard. The mimic code repository: enabling reproducibility in critical care research. *Journal of the American Medical Informatics Association*, 2017.
- [15] George K. Knopf and Archana Sangole. Data Visualization. *Wiley Encyclopedia of Biomedical Engineering*, (December), 2006.
- [16] Fan-Yin Tzeng and Kwan-Liu Ma. Opening the Black Box - Data Driven Visualization of Neural Network. *IEEE Visualization 2005 - (VIS'05)*, pages 49–49, 2005.
- [17] PyViz developers. 1.11.3.
- [18] An intuitive explanation of convolutional neural networks.
- [19] Livier Josefina Silva-Perez, Mario Alberto Benitez-Lopez, Joseph Varon, and Salim Surani. Management of critically ill patients with diabetes. *World Journal of Diabetes*, 8(3):89, 2017.
- [20] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr. Xsede: Accelerating scientific discovery. *Computing in Science & Engineering*, 16(5):62–74, Sept.-Oct. 2014.
- [21] Nicholas A. Nystrom, Michael J. Levine, Ralph Z. Roskies, and J. Ray Scott. Bridges: A uniquely flexible hpc resource for new communities and data analytics. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, XSEDE '15, pages 30:1–30:8, New York, NY, USA, 2015. ACM.
- [22] Icd-9 cm to icd-10 cm: Implementation issues and challenges.

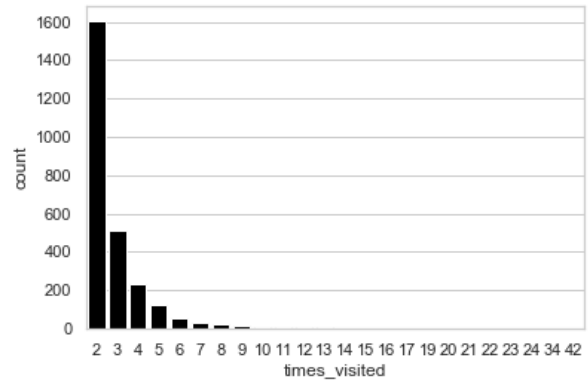


Figure 11: Multiple Visit Distribution.

this represents a likely missed diagnosis. Demonstrated by this small snippet of patients, this trend happens rather regularly.

A EXAMINING THE RELIABILITY OF ICD9 CODE DIAGNOSIS

ICD9 coding was developed by the World Health Organization to set a standard practice of disease coding that could be used for statistical compilation of data reimbursement purposes, and an array of other reasons [22]. Since the implementation of the ICD9 system, hospitals have transitioned to an ICD10 system, however for the data in this research, it was collected while the ICD9 system was still in place. In this paper, ICD9 codes were used for finding which patients in the data set were diabetic, in addition to the calculation of comorbidity scores.

While analyzing the number of diabetic patients, it was discovered that many of these patients had multiple ICU visits. In fact, figure 11 shows the exact distribution of patients who had more than one visit and were diagnosed with diabetes. As seen in the graph, two visits is the most frequent occurrence, however some patients have as many as 42 visits.

It could be assumed that diabetes is a disease that should be coded each visit, since each diabetic patient requires glucose monitoring that non-diabetic patients may not require. With this logic, if a patient has ever been diagnosed at a hospital, each subsequent visit should also be coded with diabetes. For example, in a scenario where a patient is coded with diabetes on their first visit out of three, they should have a diabetes code on each of the next two visits. To put this logic to the test, patients were graphed based on whether they were coded with diabetes. An example of this can be seen in figure 12, where a green dot indicates no diabetes code and a red dot indicates a diabetes code. After graphing each patient, an unexpected pattern was found. In many instances, the logic described earlier did not hold true. To demonstrate this, observe the first row of figure 12, which is a graph of patients with four total visits. In the first, third, and fourth visit, the patient was diagnosed with diabetes; however, the second visit they were not. Logically,

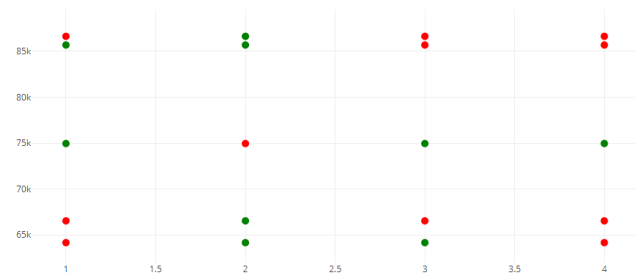


Figure 12: Scatter Plot of Patient vs. Visit Number(Red - Diabetes Code Present, Green - No Diabetes Code Present).

It can be concluded from this trend that the ICD9 coding should be used with a small amount of caution. For this paper, the reliability was not a major concern, but the discovery seemed significant. The observation from this appendix is a positive result from performing data exploration.

I authorize Hood College to lend this thesis, or reproductions of it, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.