Publisher
Singh Publication
www.singhpublication.com

E-ISSN:2583-1747

Management Journal for Advanced Research

2026 Volume 6 Number 4 August

Research Article

Diabetes Intervention

Management of Hospital Readmission Intervention Program for Diabetes Patients: A Machine Learning Approach

Kim YS[1*], Kuang J[2]

DOI:10.54741/MJAR/6.4.2026.323

[1*] Yong Seog Kim, Professor, Data Analytics and Information Systems Department, Utah State University, USA.

[2] Jie Kuang, Department Analytics Manager, Western Governors University, USA.

This paper presents a simple intervention program to reduce the risk of hospital readmission for people with type 2 diabetes (T2D). To this end, we deployed machine learning models to identify T2D patients who are most likely to be readmitted and assessed them using both a set of numerical metrics and graphical tools. Then, we applied models to subgroups based on three demographic features---race, age, and gender---to fine tune subgroup specific predictive gains and insights on the way of profiling patients who need to be monitored carefully. Overall, we found that the proposed intervention program would be financially feasible with currently available prediction models and would help health administrators determine the scale of the proposed intervention program based on estimated benefits and costs.

Keywords: diabetes intervention, readmission classification, machine learning, lift chart, profit chart, financial feasibility

Corresponding Author

Yong Seog Kim, Professor, Data Analytics and Information Systems Department, Utah State University, USA.
Email: yong.kim@usu.edu

How to Cite this Article

Kim YS, Kuang J, Management of Hospital Readmission Intervention Program for Diabetes Patients: A Machine Learning Approach. Manag J Adv Res. 2026;6(4):5-14.
Available From
https://mjar.singhpublication.com/index.php/ojs/article/view/323

To Browse

Manuscript Received	Review Round 1	Review Round 2	Review Round 3	Accepted
2026-07-04	2026-07-20			2026-08-10

Conflict of Interest	Funding	Ethical Approval	Plagiarism X-checker	Note
None	Nil	Yes	4.11	

1. Introduction

According to International Diabetes Federation [1], type 2 diabetes (T2D) is a worldwide epidemic affecting 8.3% of the global population and the total number of people living with diabetes is projected to rise to643 million people by 2030and783 million people by 2045. To make it worse, almost half of people with diabetes are unaware that they are living with the condition and three in four adults with diabetes live in low- and middle-income countries. In U.S. alone, 38.4 million people (or 11.6% of the U.S. population) suffer from diabetes and its complications, and, of those 38.4 million U.S. individuals, 8.7 million (22.8%) are undiagnosed [2].

Note that diabetes can cause serious complications such as hyperglycemia, heart disease, blindness and even death if it is not treated in time and properly. However, it is very difficult to treat T2D in time because T2D-related complications often develop before T2D is diagnosed by up to 10 years [3]. Moreover, diabetes is a high-cost epidemic and hence individuals with diabetes spend an average annual medical cost of $19,735, which is 2.6 times higher than what individuals without diabetes pay.According to the Economic Report from the American Diabetes Association in 2022, the total annual cost of diabetes in U.S. for 2022 is $412.9 billion, including $306.6 billion in direct medical costs and $106.3 billion in indirect costs (i.e., disability, work loss or premature death).

It is also important to note that T2D is a chronic medical condition with a high risk of readmission even after proper treatment. Because of its difficulty to early detect and completely treat, the number of people with T2D has been increasing and it is projected that one of every three Americans will have diabetes in 2050.

Therefore, it will be of great value to identify biological, genetic, cultural, behavioral and demographic factors that increase the risk of developing T2D. Once such factors are identified, it is very possible to develop predictive models to pinpoint who are most likely to develop T2D even when they are very young. Then the next step is to develop a comprehensive T2D intervention program to carefully monitor T2D patients and people with a high risk of developing T2D through diabetes specialist teams during their hospitalization and after discharge.

Note that such prediction and intervention programs are highly pursued to not only minimize individual financial costs associated with T2D but nationally occurred indirect costs due to presenteeism (i.e., lost productivity at work) and absenteeism (i.e., unemployment or mortality). However, such programs should be validated in advance for their feasibility and effectiveness through large-scale studies from a multidisciplinary team of primary care providers, health care researchers, economists, and data analysts [4].

Therefore, it is important to identify key factors found in T2D patients who have been frequently readmitted after they were discharged from hospitals as a short-term solution. This kind of T2D-related readmission intervention program constantly tracks patients' biological and health conditions to detect and inform patients and doctors of abnormal changes in health conditions to take precautions before unwanted T2D incidents and readmission.

Based on afore-mentioned observations and needs, this paper intends to provide a proof of concept for a simple T2D-related readmission intervention program. To this end, this paper first develops prediction models to identify T2D patients who are most likely to be readmitted and extract factors commonly found in patients who are frequently readmitted. Note that the same predictive model can be used to predict who are most likely to develop T2D in their lifetime and extract key features to profile them. Then three demographic features---race, age, and gender---are used to divide data sets into subgroups and analyze whether they provide additional insights on profiling patients who are readmitted. The performance of predictive models will be assessed using both a set of numerical metrics and graphical tools. Finally, this paper intends to test the financial feasibility of a simple readmission intervention program for T2D patients that can be extended to a comprehensive intervention program consider pre- and after-admission strategies [5][6].

2. Literature Review

Many prior and recent studies have been undertaken to report various socio-demographic and socio-economic, genetic, or biological factors that are believed to contribute to the disparity of T2D incidents.

To this end, several studies employed statistical models to identify possible causal relationships between variables and T2D incidents from small data sets. For example, Wilson et al. [7] provided the prediction rules of T2D incidents for middle-aged white adults using a series of logistic regression models. They calibrated two models so that their simple statistical model included only limited personal information such as age, gender and parental history of diabetes, while their more complex model included additional information from clinical experiments such as measurement of insulin levels. Through their comparative experiments, they concluded that several variables including parental history of diabetes, obesity, elevated triglyceride levels, low levels of high-density lipoprotein cholesterol, and hypertension statistically are significant predictors of T2D incidence. In a study from [8], separate regression models were applied to each gender group to identify informative variables from a set of clinical, biological, and genetic information obtained from patients in 30–65 years of age. They found that several clinical (e.g., adiposity) and biological variables (e.g., baseline glucose) were statistically significant while genetic variables were only marginally predictive.

Other studies have utilized the rapid advancement of genotyping technologies to investigate predictors of incident diabetes. For example, a classical study showed that T2D is strongly heritable with estimates ranging from 30% to 70% based on their observations of concordance for T2D diabetes mellitus in male twins [9]. In addition, Saxena et al. [10] identified three additional diabetes-associated loci in multi-ethnic populations based on a large meta-analysis of T2D-candidate-gene association studies. In a study [11], 70 diabetes-associated loci have been identified although signals originated from such genes explained only 10% of the genetic variance of T2D risks. Similarly, another study was undertaken to locate T2D loci through a gene-centric meta-analysis [12]. Another popular direction of studies is to employ clinically proved evidence to reduce the prevalence of T2D patients and lower their readmission rates. For example, Strack et. al. [13] reported that measuring hyperglycemia, a clinically effective treatment of diabetes, of diabetes patients during the hospitalization can lower readmission rate.

Another popular direction of diabetes-related studies involves in multi-ethnic populations, environments, and socioeconomic position.

For example, Liu et al. [14] studied the prevalence, awareness, treatment and control of T2D for 16,413 individuals aged 18-74 years residents in rural areas of China. According to their study, there were significantly more people with T2D in China countryside were higher mainly due to inadequate awareness and limited treatment options available. In addition, Harris et. al. [15] conducted interview and standard laboratory procedures to estimate risk factors of diabetic retinopathy and found that the prevalence and severity of diabetic retinopathy is greater in non-Hispanic blacks and Mexican Americans with T2D in the U.S. population than in non-Hispanic whites. Similarly, Maty et. al. [16] reported that socioeconomic status (SES) variables such as education, income, and occupation were associated with increased diabetes risk. However, Brown et. al. [17] claimed that racial/ethnic and socioeconomic variation in managed-care settings were not consistently associated with worse or better outcomes of different race and SEP groups. Readers who are interested in this line of studies are recommended to read several strongly associated seminal studies [18][19][20]. A recent study used the national health and nutrition examination survey to report that, due to their behavioral lifestyles, minorities and those with low socioeconomic status (SES) are more likely to suffer from diabetes [21].

The latest summary data on the prevalence of diagnosed diabetes across socio-economic position, gender and race is also publicly available on the Internet. For example, according to the National Diabetes Statistics Report (2024), the prevalence of people with diagnosed diabetes varies significantly among adults in different race and ethnicity subgroups in US: American Indian (16%) followed by Black (12.5%), Hispanic (10.3%), Chinese (7.1%), Japanese (6.8%) and Korean (6.1%). It also shows that the estimates of diagnosed diabetes incidence vary across U.S. counties (ranging from 2.2 to 53.5 per 1,000 people in 2020) and by an indicator of socioeconomic status (13.1% of adults with less than a high school education level vs. 6.9% of those with more than a high school education).

In summary, most prior studies adopted statistical analyses (e.g., logistic regression) to identify statistically significant predictive variables from small data sets. While they have greatly contributed to the scientific understanding of the root causes of T2D, their prediction models were designed from the perspective of physicians,

biologists and geneticists, and hence did not provide financial/economic benefits and managerial insights that state and federal government officials needed to develop and initiate a comprehensive T2D intervention program. In particular, it is strongly recommended to develop data-driven prediction models from a large data collected and managed through data warehouse technology.

3. Data Sets & Prediction Tasks

The diabetes dataset used in this study consists of 101,766 inpatient encounter records along with demographic (e.g., race, gender, age, and weight) and clinical/medical test features (e.g., primary diagnosis, admission type, number of lab procedures, number of medications and number of emergency visits, and others). This data is a preprocessed version of the Health Facts data collected from 130 US hospitals and integrated delivery networks over 10 years for a prior study [13], which is available at UCI Machine Learning Repository [22]. For our further analysis, we removed three attributes—weight (97%), payer_code (52%), and medical_specialty (53%) from our analysis due to high rate of missing values (>50%), resulting in 14 nominal and eight numerical input variables in addition to the class variable, "readmitted."

The classification task in this study is to predict one of three possible values of readmitted variable for each record: "<30" for a readmitted patient within 30 days after being discharged from a hospital, ">30" for a readmitted patient after 30 days from discharge, and "No" for a patient who is not readmitted at all. Three most representative data mining models (artificial neural network (ANN), decision tree (DT), and Naïve Bayes (NB)) were built with the default parameter setting of Analysis Server in Microsoft SQL Server 2014. We randomly select 70% (21,000 records) of records for calibrating each model and the remaining 30% (9,000 records) for testing purpose. In the final data set, patients with "<30", ">30", and "No" class variable values consist of 11.56%, 32.97% and 55.47% of records, respectively.

Note that, in this study, it is not our main goal to develop a completely new prediction algorithm or fine tune any existing algorithm to obtain a better performing algorithm.

Instead, we intend to estimate and visualize financial and economic impact of an imaginary T2D monitoring and intervention program. It is deeply noted that there will be greater (but intangible) benefits for T2D patients such as extended and healthy lifetime than just financial and economic savings once such a program is successfully initiated and maintained. Even so, our financial and economic analysis will provide a starting point for state and federal government officials to initiate such programs nationwide. In addition, our financial and economic analysis assumes that highly accurate prediction models are available to identify patients who will be readmitted due to worsen T2D or its complications. Therefore, it is still necessary to evaluate the performance of prediction models.

To this end, prediction models are compared in terms of accuracy, specificity, sensitivity, and precision. To define these metrics, it is necessary to simplify our multi-class problems into a bi-class classification problem, in which prediction models are trained to identify patients who are readmitted within 30 days (positive records) or those who are not (negative records). When models return their predictions for records in a test data, four possible cases are tabulated: the number of true positive (TP) cases when a prediction model correctly classifies a positive record as positive, and the number of false positive (FP) cases when it misclassifies a negative record as positive, the number of true positive (TN) cases when it a prediction model correctly classifies a negative record as negative, and the number of false negative (FN) cases when it misclassifies a positive record as negative. Based on tabulated numbers of four cases, accuracy, sensitivity, precision, and specificity are calculated as follows:

$$\text{accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$\text{sensitivity} = TP / (TP + FN) \quad (2)$$

$$\text{specificity} = TN / (TN + FP) \quad (3)$$

$$\text{precision} = TP / (TP + FP) \quad (4)$$

In addition to four numerical metrics, a lift chart is also adopted to visualize the predictive power of a prediction model. In a lift chart, *y*-axis represents sensitivity (or true positive rate or hit rate) for a corresponding *x*-axis value, a top *x*% of an entire population in terms of the estimated probability for a record to be positive.

By definition, a lift chart for a random model is always a diagonal line because it will statistically correctly identify *x*% of positive records when top *x*% of population are classified.

4. Model Selection & Analysis

4.1 Model Selection

Based on values of four numerical metrics, we immediately removed NB from further consideration mainly because its low accuracy and specificity value. We also consider the fact that while many data sets contain numerical variables but NB cannot process them appropriately by its nature. While ANN and DT are comparable in terms of accuracy (68.77% vs. 69.35%), we preferred ANN due to its significantly higher sensitivity value (10.00% vs. 3.98%). Note that sensitivity indicates a model's ability to correctly detect patients who are readmitted (or, in general, who have a disease) and is preferred to specificity in a medical domain. Therefore, from now on, visual outcomes from ANN only will be presented due to the limited space. However, our analysis structure is generic and hence outputs from DT can be presented and discussed in the same manner.

Using ANN as a prediction model, three lift charts can be obtained depending on the prediction task. For example, the first and the second lift chart in Figure 1-a and 1-b represents cases in which a prediction task is to identify which patients are readmitted within 30 days (i.e., records with "<30" class value are regarded as positive while others negative) or after 30 days (i.e., records with ">30" class value are regarded as positive while others negative), respectively. In these figures, a green line represents the lift chart of ANN while a red and a blue line represents the lift chart of an ideal and a random model, respectively.

According to Figure 1-a, when top 10% of samples are chosen for a prediction task, ANN identifies about 35% of positive samples while an ideal and a random model identifies 100% and 10% of positive samples, respectively. Therefore, the lift (or improvement) of ANN over a random model is calculated as 35%/10% = 3.5. According to Figure 1-a, for ANN to identify 80% of patients who are readmitted within 30 days, it will require 42% of samples.

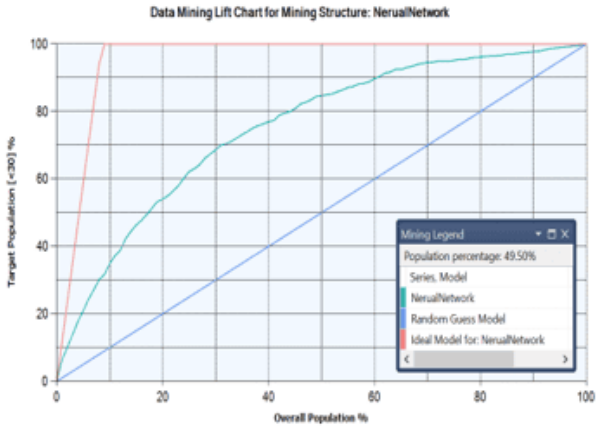

Figure 1-a: Lift Chart for "< 30"

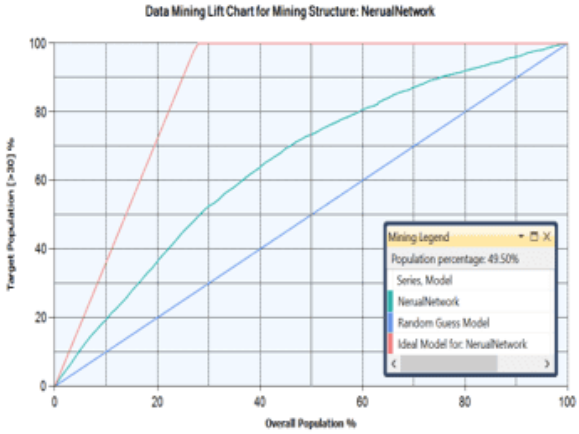

Figure 1-b: Lift Chart for "> 30"

In a similar way, according to Figure 1-b, when top 30% of samples are chosen for a prediction task, ANN identifies about 52% of positive samples while an ideal and a random model identifies 100% and 30% of positive samples, respectively. Overall, ANN model performs much better than a random model and hence we feel comfortable to use it for further analysis. Due to the limited space, we will present additional results only for a prediction task of identifying patients who are readmitted within 30 days.

4.2 Gender Group Analyses

For further analysis, the data set is divided into two subgroups based on gender to see if two subgroups reveal different factors that affect their readmission within 30 days. The data consists of female and female approximately equally (53.47% vs 46.73%) and lift charts of female and male groups are shown in Figure 2.

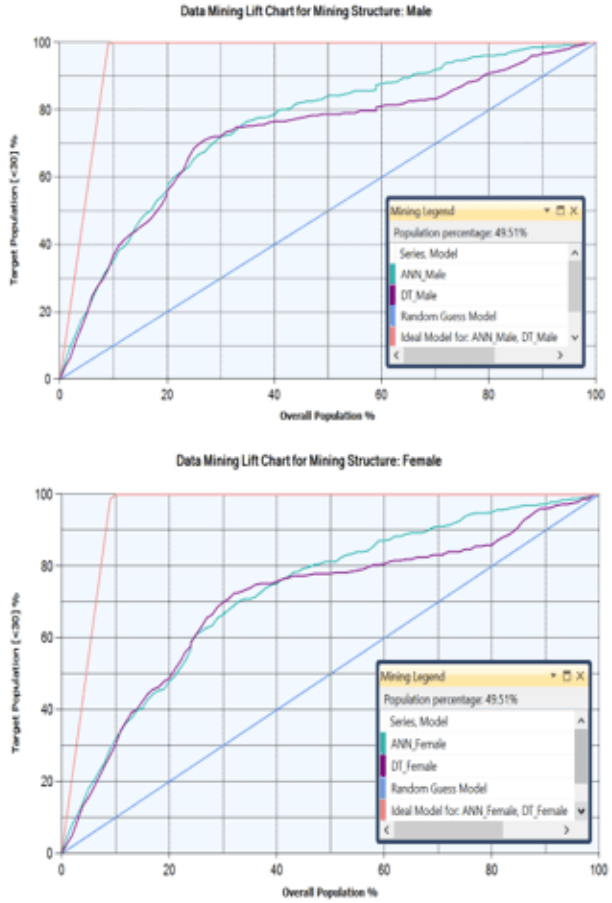

Figure 2: Lift Chart for Male (top) & Female

Two prediction models (ANN as a green line and DT as a purple line) along with an ideal (a red line) and a random model (a blue line) are shown together for a reference purpose. First of all, in terms of predictive performance, ANN performs better than DT. Then we contrasted patients in each subgroup against a baseline group, patients with the same gender but who were not readmitted. We find that female patients who take more Acarbose medicine are likely to readmitted within 30 days (with a probability = 50.47%). Female patients associated with up dose of Chlorpropamide, down dose of Pioglitazone, and steady dose of Chlorpropamide are also likely to be readmitted within 30 days. In contrast, from male patient group, we note that patients who are associated with steady dose of Tolazamide is most likely to be readmitted within 30 days with a probability of 75.27%. Male patients who keep steady dose of Troglitazone and Tolbutamide or down dose of Glimepiride are also likely to be readmitted within 30 days.

4.3 Racial Group Analysis

We also divided data sets into five subgroups—Caucasian (72.79%), African American (20.92%), Hispanic (2.30%), Asian (0.49%) and other races (1.27%)—to see different patterns of each race subgroup against a baseline group. Lift charts were presented only for two largest groups (Caucasian and African American) in Figure 3.

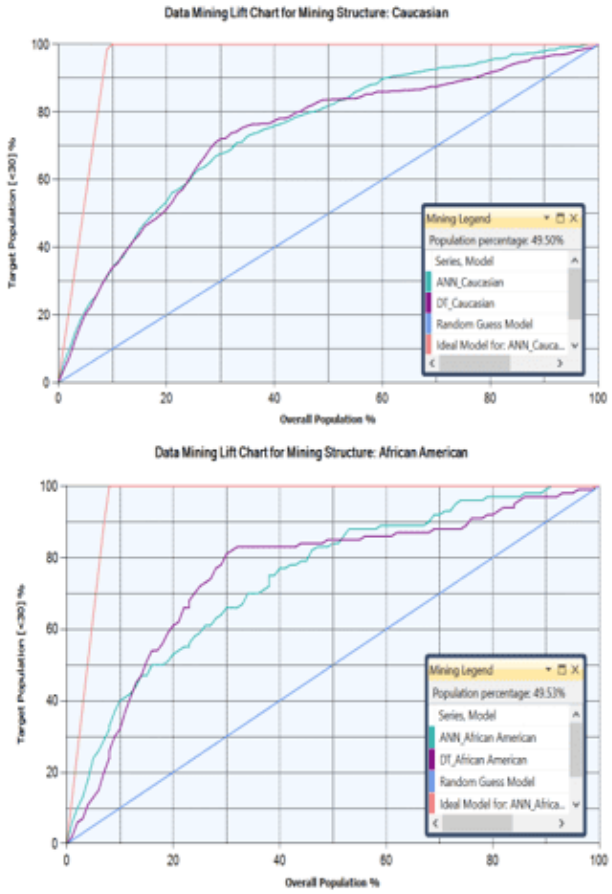

Figure 3: Lift Chart for Caucasian (top) & African American

In the Lift charts of Caucasian group, ANN and DT performed comparably up to top 25% of patients. However, DT outperformed ANN in a range of between top 25% and 50% of patients and ANN outperformed when more than top 50% of patients are considered. In the Lift charts of African American group, ANN performed better for the first top 15% of patients, but DT outperformed ANN in a range of between top 15% and top 50% of patients though ANN outperformed DT again over top 50% patients or higher. The most significant medicine factor that affected the readmission of Caucasian within 30 days was Troglitazone. Caucasian patients who were treated with steady dose of Troglitazone are more likely to be readmitted within 30 days. Steady administration of Tolbutamide, increased dose of Chlorpropamide, or decreased dose of Repaglinide medicine were also associated with the readmission of Caucasian within 30 days.

In contrast, for African American group, the most significant medicine indicator is Chlorpropamide: African American patients who have been administered with the steady amount of Chlorpropamide are far more likely to be readmitted within 30 days. Note that two medicines, Chlorpropamide and Repaglinide, are most prevalent medications administered to treat T2D.

4.4 Age Group Analysis

It is also speculated that patient subgroups based on their age would reveal different readmission patterns. To test this, we divided the data set into three subgroups based on age: [0-50] (18.49%), [51-70] (39.53%) and [71-100] (41.98%). According to Figure 4-a, DT outperformed ANN in a readmission prediction of patients in [0-50] age group over a range of between top 10% and top 35% of patients, but ANN performed better in other ranges. However, ANN was superior to DT across almost all ranges for age groups of [51-70] (Figure 4-b) and [71-100] (similar to Figure 4-b, hence omitted due to the limited space). Across all age groups, the number of times that patients have been admitted before is one of the strongest (but too obvious) indicators for a task of predicting whether or not patients will be readmitted within 30 days.

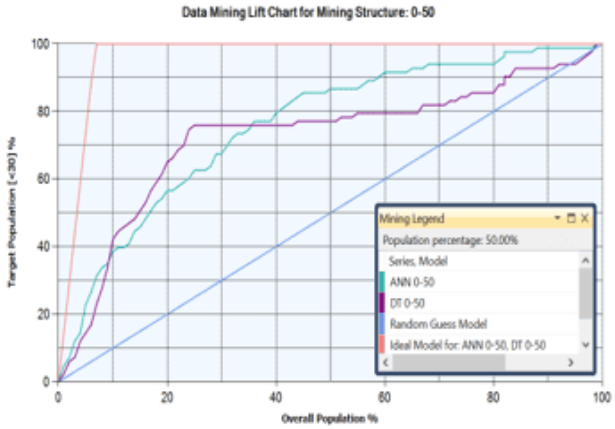

Figure 4-a: Lift Chart for Age Group of [0-50]

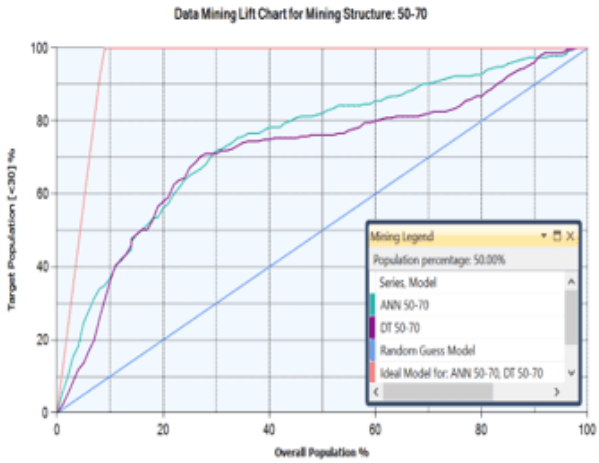

Figure 4-b: Lift Chart for Age Group of [51-70]

The most influential medicine indicator for [0-50] year-old group, however, was Pioglitazone. Patients who are treated with decreased dose of Pioglitazone (along with steady dose of Chlorpropamide, decreased dose of Metformin, increased dose of Glimepiride) were more likely to be readmitted within 30 days. The most influential medicine indicator of patient readmission in [51-70] year-old group was Miglitol with steady dose, while it was with increased dose of Chlorpropamide (along with steady dose of Troglitazone and Glyburide-metformin, and decreased dose of Pioglitazone) in [71-100] year-old group.

5. Feasibility Analysis

One of many reasons to develop prediction models and profile patients who are most likely to develop T2D or who are most likely to be readmitted is to initiate a T2D monitoring and intervention program. For example, state or federal government may initiate a diabetes intervention program targeted at pre-diabetic individuals to delay or prevent the onset of T2D. Another kind of intervention program may include intensive lifestyle changes advice for patients who are most likely to be readmitted so that they can avoid readmission or at least reduce the number of unexpected readmissions. Such intervention programs can save not only lives of patients but also direct and indirect costs due to T2D prevalence. However, implementing such programs will necessarily incur various costs and hence it is necessary to test the feasibility of programs. To this end, we consider a simple T2D monitoring and intervention program that incorporates a prediction model (i.e., ANN) to actively identify T2D patients who will be readmitted.

We also consider that health administrators will be interested in knowing an optimal magnitude of such a readmission intervention program in terms of how many T2D patients to be included in the program to make it financially feasible.

To estimate the financial feasibility of such a readmission intervention program, we first set potential target (diabetes) population to 38.4 million (Centers for Disease Control and Prevention, 2024) and fixed cost (i.e., initial set up costs) to $100 million based on the latest statistics. Note that using different values for the fixed cost does not necessarily affect the implications of the feasibility analysis. We set variable cost (individual cost in Analysis Server) to $300 per individual, an estimate close to the price of a smart watch/health tracking device that each patient in the program receive and use to track various health related (e.g., heartbeat and blood pressure) and lifestyle information (e.g., daily diet, exercise log).

Finally, we consider two different revenue (or cost saving) settings per each correctly predicted patient to be readmitted: $3,000 ($412.9 billion / 38.4 million * 25% ≈ $2,688) and $6,000 $412.9 billion / 38.4 million * 50% ≈ $5,376). That is, we assume that the readmission prevention program will result in the cost saving of direct and indirect costs by correctly identifying and monitoring who are most likely to be readmitted before they suffer from severe complications and readmitted for a long period of times.

These revenue settings were derived from assumptions that a new program may reduce the total costs of $412.9 billion among 38.4 million T2D patients ($306.6 billion for direct medical costs and $106.3 billion for indirect costs, Centers for Disease Control and Prevention, 2024) by 25% (Scenario 1) or 50% (Scenario 2). We also consult another study [23] in which interventions with intensive lifestyle changes could reduce the risk of developing T2D by 58%.

With these setting values, we obtained two profit charts presented in Figure 5-a (Scenario 1) and 5-b (Scenario 2). In Scenario 1 where the intervention program is assumed to reduce the cost by 25%, we find that it is profitable if at least top 5% of patients participate for the program and its profit (or cost saving) is maximized when 30% of T2D patients are included in the program.

However, the profit of the program decreases as more patients with lower probability of developing T2D or being readmitted participate and finally turns into negative as more than 65% of patients are included in the program.

Figure 5-b depicts the profit chart when the program aims to reduce the cost by 50%. In Scenario 2 where the intervention program is assumed to reduce the cost by 50%, the profit steadily increases at a decreasing rate and is maximized when 50% of patients are involved.

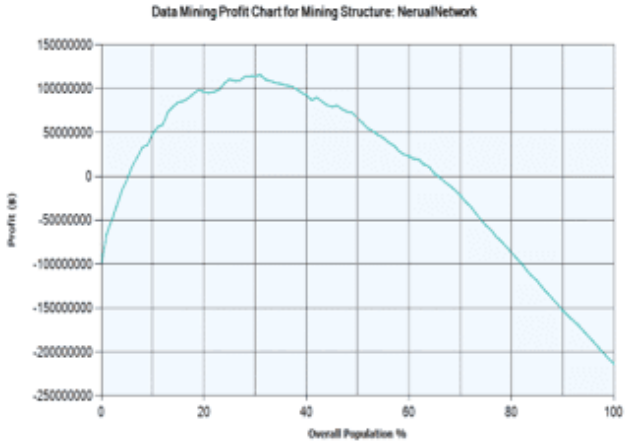

Figure 5-a: Profit Chart (25% Cost Reduction)

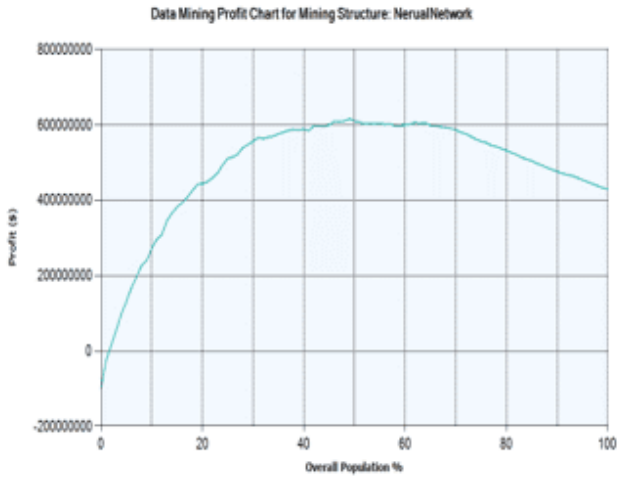

Figure 5-b: Profit Chart (50% Cost Reduction)

Note that while the financial estimates itself from both scenarios is meaningful for health administrators, a more compelling insight from this study is that, with the help of even simple machine learning models, it is readily possible to identify T2D patients who are most likely to be readmitted and health administrators can easily decide how many T2D patients should be included in a readmission prevention program when targeting all T2D patients is not financially feasible.

Therefore, findings in this section are particularly useful because various findings from many prior studies may provide scientific insights or medicine indicators related to repeated hospital admission of T2D patients from the perspective of medical professionals and biologists, but they do not provide financial and managerial insights from the perspective of local and federal health administrators who would initiate a readmission intervention program for T2D patients.

6. Conclusion

This paper considers a simple intervention program to minimize the readmission of T2D patients and tests its financial feasibility. To this end, various prediction models were first calibrated to identify who are most likely to be readmitted and extract influential factors on readmission rates. Through subgroup analyses of gender, race and age, we found that patients in each subgroup show distinct responses to clinical medicines doses during their hospitalization, which may impact the probability of readmission after discharge. In addition, we also found that this simple T2D intervention program can be financially feasible as long as highly predictive models are used to select T2D patients who are likely to be readmitted and program administrators determine the number of T2D patients targeted for the prevention program.

In future, we will consider a more realistic and comprehensive T2D intervention program and test its economic feasibility with better estimates of costs and cost savings. In particular, T2D intervention program should be considered and tested for various socio-economic, demographic, environmental, and genetic factors. In addition, such intervention program should be updated and monitored dynamically.

Therefore, the proposed readmission intervention program with clinically effective treatment of inpatients may be implemented as an initiative project of a comprehensive T2D prevention program. Then it will incorporate multiple approaches such as pre-arranged follow-up appointments, individualized discharge planning, patient education and communication with the primary care provider.

References

1. International Diabetes Federation (IDF). (2021). *The diabetes atlas*. (10th ed.). https://diabetesatlas.org/atlas/tenth-edition/

2. Centers for Disease Control and Prevention (CDC). (2024). *National diabetes statistics report*. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept. of Health and Human Services. Available at: https://stacks.cdc.gov/

3. Keating, B. J. (2015). Advances in risk prediction of type 2 diabetes: Integrating genetic scores with Framingham risk models. *Diabetes*, *64*(5), 1495-1497.

4. Bannuru,R.R. (2024). Economic costs of diabetes in the U.S. in 2022. *Diabetes Care*, *47*(1), 26-43.

5. Jack, B.W., Chetty, V.K., Anthony, D., Greenwald, J.L., Sanchez, G.M., Johnson, A.E., Forsythe, S.R., O'Donnell, J.K., Paasche-Orlow, M.K., Manasseh, C., Martin, S., & Culpepper, L. (2009). A reengineered hospital discharge program to decrease rehospitalization: A randomized trial. *Annals of Internal Medicine*, *150*(3), 178–187.

6. Dedhia, P., Kravet, S., Bulger, J., Hinson, T., Sridharan, A., Kolodner, K., Wright, S., & Howell, E. (2009). A quality improvement intervention to facilitate the transition of older adults from three hospitals back to their homes. *Journal of the American Geriatrics Society*, *57*(9), 1540–1546.

7. Wilson, P.W., Meigs, J.B., Sullivan, L., Fox, C.S., Nathan, D.M., and D'Agostino, R.B. (2007). Prediction of incident diabetes mellitus in middle-aged adults: The framingham offspring study. *Archives of Internal Medicine*, *167*(10), 1068-1074.

8. Balkau, B., Lange, C., Fezeu, L., Tichet, J., de Lauzon-Guillain, B., Czernichow, F., Froguel, P., Vaxillaire, M., Cauchi, S., Ducimetière, P., & Eschwège, E. (2008). Predicting diabetes: Clinical, biological, and genetic approaches data from the epidemiological study on the insulin resistance syndrome (DESIR). *Diabetes Care, 31*(10), 2056-2061.

9. Newman, B., Selby, J.V., King, M.C., Slemenda, C., Fabsitz, R., & Friedman, G.D. (1987). Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. *Diabetologia*, *30*(10), 763-768.

10. Saxena, R., Elbers, C.C., Guo, Y., et al. (2012). Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. *American Journal of Human Genetics*, *90*(4), 410–425.

11. Keating, B.J. (2015. Advances in risk prediction of type 2 diabetes: Integrating genetic scores with Framingham risk models. *Diabetes*, *64*(5), 1495-1497.

12. Asselbergs, F.W., Guo, Y., van Iperen, E.P., Sivapalaratnam, S., Tragante, V., Lanktree, M.B., ... and Baumert, J. (2012). Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *American Journal of Human Genetics*, *91*(5), 823-838.

13. Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., and Clore, J. N. (2014). Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. *BioMed Research International.* doi:10.1155/2014/781670.

14. Liu, X., Li, Y., Li, L., Zhang, L., Ren, Y., Zhou, H., ... & Wang, C. (2016). Prevalence, awareness, treatment, control of type 2 diabetes mellitus and risk factors in Chinese rural population: The RuralDiab study. *Scientific Reports*, *6*, 31426.

15. Harris, M. I., Klein, R., Cowie, C.C., Rowland, M., & Byrd-Holt, D. D. (1998). Is the risk of diabetic retinopathy greater in non-Hispanic blacks and Mexican Americans than in non-Hispanic whites with type 2 diabetes? A US population study. *Diabetes Care*, *21*(8), 1230-1235.

16. Maty, S.C., Everson-Rose, S.A., Haan, M.N., Raghunathan, T.E., & Kaplan G.A. (2005). Education, income, occupation, and the 34-year incidence (1965-99) of Type 2 diabetes in the Alameda County Study. *International Journal of Epidemiology*, *34*(6). 1274-1281. doi:10.1093/ije/dyi167.

17. Brown, A.F., Gregg, E.W., Stevens, M.R., Karter, A.J., Weinberger, M., Safford, M.M., ... & Beckles, G.L. (2005). Race, Ethnicity, Socioeconomic Position, and Quality of Care for Adults with Diabetes Enrolled in Managed Care the Translating Research into Action for Diabetes (TRIAD) study. *Diabetes Care*, *28*(12), 2864-2870.

18. Maty, S.C., James, S.A., & Kaplan, G.A. (2010). Life-course socioeconomic position and incidence of diabetes mellitus among blacks and whites: The Alameda County Study, 1965-1999. *American Journal of Public Health*, *100*(1), 137-145. doi:10.2105/AJPH.2008.133892.

19. Maty, S.C., Lynch, J.W., Raghunathan, T.E., & Kaplan, G.A. (2008). Childhood socioeconomic position, gender, adult body mass index, and incidence of type 2 diabetes mellitus over 34 years in the Alameda County Study. *American Journal of Public Health*, *98*(8), 1486-94. doi:10.2105/AJPH.2007.123653.

20. Baltrus P.T., Everson-Rose S.A., Lynch J.W., Raghunathan, T.E., & Kaplan, G.A. (2007). Socioeconomic position in childhood and adulthood and weight gain over 34 years: The Alameda county study. *Annals of Epidemiology*, *17*(8), 608-614. doi:10.1016/j.annepidem.2007.03.007.

21. Liu, C., He, L., Li, Y., Yang, A., Zhang, K., & Luo, B. (2023). Diabetes risk among US adults with different socioeconomic status and behavioral lifestyles: Evidence from the National Health and Nutrition Examination Survey. *Front Public Health*, *22*(11). doi:10.3389/fpubh.2023.1197947.

22. Frank, A., & Asuncion, A. (2010). *UCI machine learning repository*.

23. Palmer, A. J., Roze, S., Valentine, W. J., Spinas, G. A., Shaw, J. E., & Zimmet, P. Z. (2004). Intensive lifestyle changes or metformin in patients with impaired glucose tolerance: Modeling the long-term health economic implications of the diabetes prevention program in Australia, France, Germany, Switzerland, and the United Kingdom. *Clinical Therapeutics*, *26*(2), 304-321.